Springer Proceedings in Mathematics & Statistics

Maria João Alves
João Paulo Almeida
José Fernando Oliveira
Alberto Adrego Pinto *Editors*

Operational Research

IO 2018, Aveiro, Portugal, September 5–7

Springer

Springer Proceedings in Mathematics & Statistics

Volume 278

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Maria João Alves · João Paulo Almeida ·
José Fernando Oliveira ·
Alberto Adrego Pinto
Editors

Operational Research

IO 2018, Aveiro, Portugal, September 5–7

Springer

Editors
Maria João Alves
Faculty of Economics/CeBER
University of Coimbra
Coimbra, Portugal

José Fernando Oliveira
Department of Industrial Engineering
and Management/INESC TEC
University of Porto
Porto, Portugal

João Paulo Almeida
Department of Mathematics/CeDRI-IPB
and LIAAD-INESC TEC
Polytechnic Institute of Bragança
Bragança, Portugal

Alberto Adrego Pinto
Department of Mathematics/
LIAAD-INESC TEC
University of Porto
Porto, Portugal

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-030-10730-7 ISBN 978-3-030-10731-4 (eBook)
https://doi.org/10.1007/978-3-030-10731-4

Library of Congress Control Number: 2018967413

Mathematics Subject Classification (2010): 90Bxx, 97M40

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

To Zé Manel, Zé and Manel

Maria João Alves

To Ana Rita, Ana Mafalda and Cristina

João Paulo Almeida

To São, Beatriz and Mariana

José Fernando Oliveira

To Maria Barreira Pinto, Ana Barreira and Maria Guiomar Pinto

Alberto Adrego Pinto

Foreword

It is always a pleasure to congratulate an active scientific society on the occasion of its 40th anniversary; however, it is a bit arbitrary, since the Babylonians would have celebrated it four years ago…

What's more important is to witness the success of a thriving association, i.e. its members. Although the important Portuguese personalities in Operations Research were always well connected and enjoyed an excellent reputation in the international community, the newly emerging generation of Portuguese Operations Researchers is continuing on that path, as can be seen in the impressive visibility and recognition of their scientific achievements, coupled with a rapidly growing network at the global level. An important testimony to this fact is the sheer number of well-received international scientific meetings that are co-organized by the APDIO and/or its active members every year—which is itself an excellent cause for celebration!

In this vein, I would like to wish the APDIO and its members, including the generations of researchers to come, the same success in their endeavours. IO 2018 has already contributed significantly to this goal by providing a forum for fruitful exchange and the development of fresh ideas for the benefit of all its delegates, Portuguese and international, and for the field of operational research as a whole. In your hands, you hold the best proof of my humble opinion.

Vienna, Austria
September 2018

Immanuel M. Bomze

Acknowledgements

The editors would like to thank all authors of this volume for having shared their vision along the contributed chapters. We also are very grateful to the anonymous referees, members of the Programme Committee, whose work and effort much contributed to the improvement of the quality of the book.

We thank Immanuel M. Bomze for contributing the foreword of the book, for his kind words and support.

We thank the Executive Editor for Mathematics, Computational Science and Engineering at Springer-Verlag, Martin Peters, for his valuable suggestions and advice throughout this project. We thank Ruth Allewelt at Springer-Verlag for her assistance throughout this project.

Alberto Adrego Pinto and João Paulo Almeida would also like to acknowledge the support from LIAAD-INESC TEC.

This project was partially financed by the ERDF—European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation—COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT—Fundação para a Ciência e a Tecnologia within projects:

- MAnAGER (POCI-01-0145-FEDER-028040);
- POCI-01-0145-FEDER-029609 and POCI-01-0145-FEDER-029279;
- Dynamics, optimization and modelling (PTDC/MAT-NAN/6890/2014);
- Modelling, Dynamics and Games (PTDC/MAT-APL/31753/2017).

Contents

Towards an Integrated Framework for Aerospace Supply Chain Sustainability

Cátia Barbosa, Nuno Falcão e Cunha, Carlos Malarranha, Telmo Pinto, Ana Carvalho, Pedro Amorim, M. Sameiro Carvalho, Américo Azevedo, Susana Relvas, Tânia Pinto-Varela, Ana Cristina Barros, Filipe Alvelos, Cláudio Alves, Jorge Pinho de Sousa, Bernardo Almada-Lobo, José Valério de Carvalho and Ana Barbosa-Póvoa

Abstract Supply chains have become one of the most important strategic themes in the aerospace industry in recent years as globalization and deep technological changes have altered the industry at many levels, creating new dynamics and strategies. In this setting, sustainability at the supply chain level is an emerging research topic, whose contributions aim to support businesses into the future. To do so the development of new products and the response to new industry requirements, while incorporating new materials appears as a path to follow, which require more resilient and agile supply chains, while guaranteeing their sustainability. Such supply chains will be better prepared for the future complex challenges and risks faced by the

C. Barbosa · N. Falcão e Cunha · P. Amorim · A. Azevedo · J. P. de Sousa · B. Almada-Lobo
INESC TEC and Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
e-mail: catia.r.barbosa@inesctec.pt

N. Falcão e Cunha
e-mail: nuno.cunha@fe.up.pt

P. Amorim
e-mail: pamorim@fe.up.pt

A. Azevedo
e-mail: ala@fe.up.pt

J. P. de Sousa
e-mail: jorge.p.sousa@fe.up.pt

B. Almada-Lobo
e-mail: almada.lobo@fe.up.pt

C. Malarranha · A. Carvalho · S. Relvas · T. Pinto-Varela · A. Barbosa-Póvoa (✉)
CEG-IST, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
e-mail: apovoa@tecnico.ulisboa.pt

C. Malarranha
e-mail: carlos.malarranha@tecnico.ulisboa.pt

A. Carvalho
e-mail: anacarvalho@tecnico.ulisboa.pt

S. Relvas
e-mail: susana.relvas@tecnico.ulisboa.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_1

aerospace companies. Such challenges are addressed in this work, where an integrated framework is proposed to contribute to the resilience and sustainability of aerospace supply chains. Using different analysis methods, the framework addresses four important challenges in the context of aerospace supply chain sustainability: evolution and new trends, performance assessment, supplier selection, and supply chain design and planning.

Keywords Aerospace industry · Supply chain management · Sustainability · Integrated framework

1 Introduction

Over the years, major aircraft manufacturers have evolved into integrators of complex sets of parts, systems and large modules manufactured by third-party companies [1]. This sets many challenges for the industry, with the introduction of new materials playing a critical role [2]. Understanding the implications at supply chain level brought by the introduction of these materials, and their impacts in terms of sustainability is fundamental [3]. The time and high costs associated with these changes to the aerospace industry are one of the major open issues for aerospace manufacturers and their respective supply network. Clearly, alternative methods to exploit new materials in a more efficient way must be developed [4]. Future supply chains, in addition to being sustainable, must be resilient and agile in responding to new industry requirements, especially when dealing with new materials.

The IAMAT project (Introduction of Advanced Materials Technologies into New Product Development for the Mobility Industries) was launched to address these industry challenges and involves four universities, two companies and different fields of study. Its main goal is to develop an integrated methodology that facilitates the

T. Pinto-Varela
e-mail: tania.pinto.varela@tecnico.ulisboa.pt

T. Pinto · M. S. Carvalho · F. Alvelos · C. Alves · J. V. de Carvalho
Centro ALGORITMI, Escola de Engenharia, Universidade do Minho, Braga, Portugal
e-mail: telmo@dps.uminho.pt

M. S. Carvalho
e-mail: sameiro@dps.uminho.pt

F. Alvelos
e-mail: falvelos@dps.uminho.pt

C. Alves
e-mail: claudio@dps.uminho.pt

J. V. de Carvalho
e-mail: vc@dps.uminho.pt

A. C. Barros
INESC TEC, Porto, Portugal
e-mail: acbarros@inesctec.pt

introduction of new materials in the aerospace industry. This paper explores one of the main research lines of the overall project concerning the development of decision support tools to improve supply chain systems aiming at ensuring their sustainability and resilience in the presence of uncertainty.

In the context of this project, we intend to develop a framework integrating various qualitative and quantitative decision support tools to analyse different problems in the context of aerospace supply chain resilience and sustainability. This research paper describes each building block of the framework within the overarching IAMAT project. The remaining sections include the underlying motivation (Sect. 2), and, in Sect. 3, the description of the integrated framework, as well as the methodologies for each of the tools within it. Section 4 highlights the projects main contributions and how the aerospace industry stands to benefit from it. Main conclusions and future research directions are also depicted in this section.

2 Motivation

As mentioned in Sect. 1, major aircraft manufacturers have evolved into system integrators, outsourcing most aircraft parts as large integrated modules from key supply partners. As a result of the markets pull for more efficient aircraft, original equipment manufacturers (OEMs) had to rethink their products which, in turn, led to a revolution within the incurring supply chains. To boost the effectiveness of R&D efforts, OEMs delegated considerable product design, development and manufacturing responsibilities to their suppliers. The other main reason behind this approach was to share the non-recurring costs in new aircraft programs, thus reducing the exposure of the OEM. These emerging super tier 1 suppliers entered risk sharing partnerships with OEMs and only began to receive returns on their investments after the aircraft was being sold. The biggest step in this direction was taken by Boeing when 70% of 787 Dreamliners production was delegated to 50 strategic partners, which were called Integrators [2]. This paradigm has proven challenging to suppliers who have had to quickly develop their technical and managerial capabilities.

Concurrently, growing sustainability concerns in the aerospace industry, along with pressure from airlines, have been pushing OEMs to improve the efficiency of their aircraft and of their supply chains [5]. While attempting to reduce aircraft operating costs, and CO_2 emissions from worldwide air traffic, aerospace OEMs have been engaging in New Product Development (NPD) programmes. They have prioritized the weight reduction of aircraft by using advanced materials in aero structures and the improvement of engine efficiency [2]. Our goal is to provide OEMs with decision support tools that seek to minimize the impact on sustainability performance associated with the introduction of new advanced materials and technologies.

The general structure of the aerospace supply chain can be visualised in Fig. 1, and it has been derived from the works presented in [6, 7]. It consists of four supplier levels and the OEM as the responsible for the final assembly. The tier 4 suppliers are responsible for supplying the materials to be processed throughout the chain. The

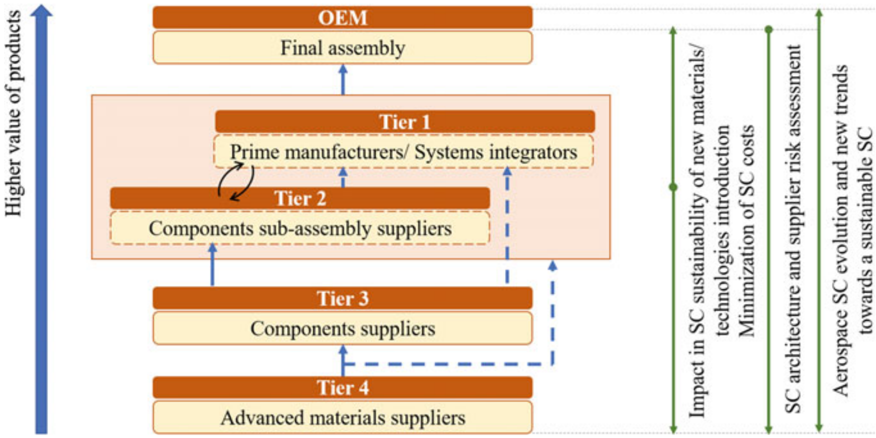


Fig. 1 General structure of an aerospace supply chain

components suppliers, tier 3, use the raw materials to produce simple components that are assembled with other components or integrated in major aerostructures at tier 2 and tier 1 levels. Tier 1 and tier 2 suppliers can be classified according to the type of structures being supplied, assuming an interchangeable role. While for integrated structures, suppliers assume a tier 1 role, for sub-assemblies, suppliers assume a tier 2 role. In an attempt to improve the aircraft efficiency, these suppliers often engage in adopting new manufacturing processes and using new materials. The OEM is responsible for the assembly of the aircraft, the highest value product in the whole chain. However, this is not the only responsibility of the OEM, it must also control and set the supply chain architecture considering the supplier risks. The sustainability of the aerospace supply chain can only be globally addressed through the cross-tier supplier analysis and assessment of industry evolution and trends. Figure 1 also shows the scope of each of the tools within the framework.

3 Integrated Framework for the Improvement of Supply Chain Sustainability

As identified in the introduction section, the aerospace supply chain is currently facing many sustainability challenges. This calls for the integration of solution approaches that can respond to the different problem requirements, while in an integrated way provides a pathway for improving the overall aerospace supply chain sustainability, particularly during NPD initiatives with the introduction of advanced materials/manufacturing processes.

Figure 2 presents the proposed integrated framework that builds on the sustainability challenges and considers a Triple Bottom Line approach (TBL) as the trigger

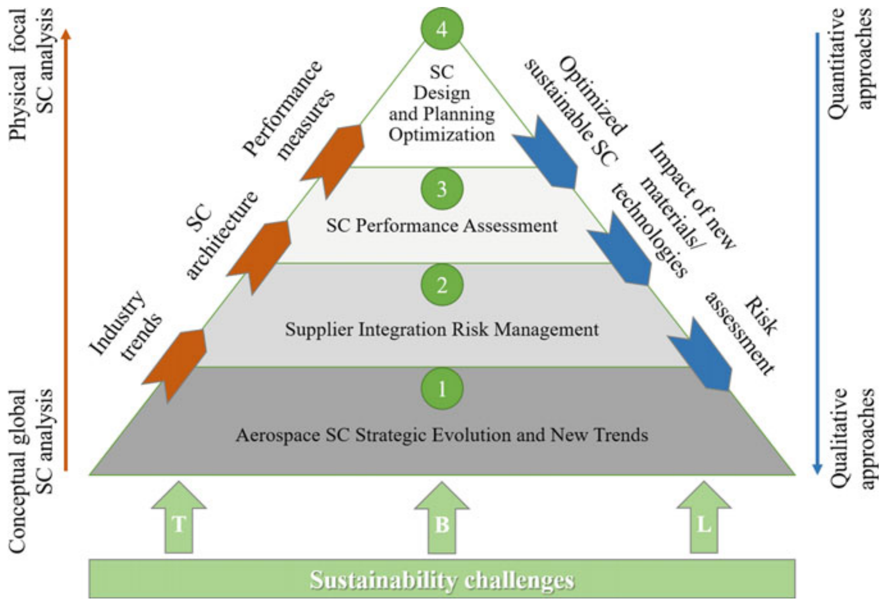


Fig. 2 The integrated framework

for the improvement of the business value. The bottom of the pyramid (1) frames both the evolution and the new trends of aerospace supply chain. Additionally, they aim to identify and characterize the supply chain stakeholders, while mapping the connections among them and assessing their engagement and main priorities towards sustainability. These supply chain strategic analytical tools based on the new trends are the input to the higher levels of the framework that consider the Supply Chain Design (SCD). Qualitative research methods have been used for a global and conceptual supply chain analysis, such as systems thinking and content analysis.

Rising in the framework levels, appears the second group of tools, (2), that proposes concurrent product and SCD considering the Supplier Integration Risk (SIR). This tool uses both qualitative and quantitative approaches by introducing risk management practices into the process of selecting suppliers and using Robust Optimization (RO) techniques. This tool allows managers to assess the danger of disruptions caused by the introduction of new technologies, and their outsourcing practices. While this tool gives an overview of a selected supply chain architecture for the physical supply chain analysis performed by the upper levels of the framework, it gives important risk assessment inputs to the lower framework level. The integration of the risk assessment enhances the development of the sustainable supply chain concepts at the lower level.

The third group of tools, (3), deepens the working principles of the aerospace supply chain, by integrating the supply chain architecture proposed by the lower level tool. Using a simulation approach to achieve a dynamic analysis of the supply chain behaviour, the hybrid simulation models, by integrating different levels of analysis at the supply chain nodes, allow extracting relevant performance measures predictions for comparing the impact of using different advanced materials and manufacturing processes. The relevant measures assessed serve as input to be integrated in an optimization tool at the upper framework tool, and can be used to narrow the number of possible supply chain participants to be considered at the concurrent product and SCD level. The hybrid simulation models make use of quantitative methods for the model development and integrate a trade-off assessment of the relevant metrics to be considered.

At the top of the framework lie the Mixed Integer Programming (MIP) models (4), that are quantitative tools for the design of a sustainable supply chain. It builds on the mapped performance measures from the hybrid simulation tool to minimize the overall supply chain costs and environmental impact, while maximizing the social value created. This tool assesses the performance of final supply chain configurations, and provides the optimized design, given the constraints arising from the lower framework levels.

From the top to the bottom of the proposed framework, the scope of the aerospace supply chain sustainability analysis is broadened. Different levels of analysis are targeted, providing a cross-functional decision support tool that tackles the prominent needs of the aerospace supply chain and contributes for the analysis and setting of policies towards a more sustainable development. The following subsections provide a more detailed description of how each of these tools addresses current issues in the aerospace industry.

3.1 Aerospace SC Strategic Evolution and New Trends

The competitive context of commercial aerospace industry has been changing rapidly over the last two decades, and supply chains have become an important driver for NPD, where sustainability concerns need to be guaranteed. OEMs and aerospace supply chains have faced various challenges and have developed several strategic responses in the context of NPD [2, 8]. New challenges and responses will emerge in the future and will continue to have a high impact on the various stakeholders involved and in the sustainability of aerospace supply chains [9, 10]. Thus, it is important to know how aerospace supply chain has evolved and what the new trends towards a sustainable supply chain are. It is also mandatory to know the most relevant stakeholders towards sustainability, and how to promote their engagement in this purpose.

To answer the question related to the evolution and new trends three mixed methodologies were used: systems thinking, content analysis and comparative analysis (Fig. 3). Systems thinking approach is used to structure the problem, while content analysis helps to understand how authors have been exploring the questions carried above. Finally, the comparative analysis is carried out involving the four most important aerospace companies worldwide, in a new product development situation. Based

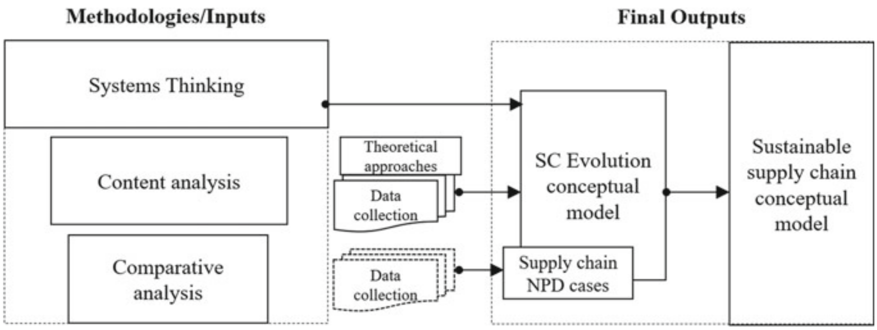


Fig. 3 Strategic evolution and new trends methodology

on this analysis stage, two conceptual models were defined. The first proposed conceptual model will allow companies to understand the evolution of aerospace supply chains in a NPD context and explore the informations output. The second explores the integration of the critical characteristics of the aerospace supply chain, such as collaboration and system integration with their new trends towards the development of new products accounting for economic, environmental and social objectives.

In this context, sustainable supply chains have been recognized as a key element of organizations [11], encouraging the interest of several stakeholders in recent years. Building a sustainable supply chain and be recognized as a sustainable industry are however major challenges to the commercial aerospace industry. Aerospace supply chains are continuously evolving concerning strategies making the process unstable and uncertain and dependent of multiple and powerful stakeholders with its own views. Aircraft new product development cycles and the new development approaches are crucial in this field. The development of an airplane has become so complex that has led to an increasingly dependency on a network of a large supply chain that involves customers, suppliers, scientific communities, regulators, governments and many others. The subjects related with stakeholder engagement had therefore become critical to evolving and building the sustainability strategies [12]. In this context, and as stated above, there is a need to identify the main stakeholders in the aerospace supply chain and their priorities and engagement towards a sustainable supply chain. For this purpose, a multimethodology was explored to allow an integrated stakeholders analysis leading to a conceptual model that helps to frame stakeholders engagement. The multimethodology includes the application of three tasks: scope definition, stakeholders identification and stakeholders prioritization, which are supported by five methods: literature review, brainstorming, snowball sampling, survey and statistics. Finally, the proposed conceptual models enables a complete stakeholders analysis, recognizing sustainability engagement flows among the most important stakeholders while improving the collaboration processes to derive a strategy towards a sustainable supply chain.

3.2 Robust Concurrent Product and Supply Chain Design Under Supplier Integration Uncertainty

Adapting the methodology for concurrent product and supply chain design (CP-SCD) described in Gan et al. [13], we propose a model that seeks an efficient and robust supply base with minimum number of supplier modules. Supplier modules hold the supplier, or group of suppliers, from which a set of parts is sourced, as an integrated component module. In recent years, OEMs have delegated onto their tier 1 suppliers the development and integration of large component modules. With the introduction of new materials and processes, suppliers are sometimes incapable of complying with the target specifications, which is represented in the new model as the SIR. One of the main extensions proposed in this model is the consideration of each suppliers integration risk (IR) in SCD. This uncertainty will be represented in the model as a risk factor. The IR is an input of the model and should result of an assessment of the suppliers technical expertise, as well as own supply chain management capabilities. The aim of the model is to mitigate supplier modules that exceed certain risk thresholds, defined by the decision maker. Another particular aspect of the model is the selective sourcing flexibility for each module, which is a decision variable within the problem.

The first necessary input for the model is the Part-Supplier Matrix (PSM) which matches the complete pool of supplier candidates to the set of parts required for the aircraft program. The PSM matches the two sets via a binary relation, and indicates whether the part can be sourced from a certain potential supplier. To each of these connections corresponds an IR for the supplier to integrate that part. The value associated with this risk is compiled in an Integration Risk Matrix (IRM). The total risk in a module will be the sum of the IR of each component integrated by the suppliers assigned to it. The IRM is another input for the model and contains evaluations of the risk according to 3 risk-levels: low, medium or high. OEMs can perform their own estimations for these risks based on their knowledge of the suppliers operations, as well as during the supplier development programs, which are already common practice ahead of new aircraft programs [14]. As seen in Fig. 4, if there were no constraints associated to supplier risk, the algorithm simply seeks to select the minimum number of suppliers, each of them delivering the largest possible number of integrated parts. While this may be the desired outcome for a risk neutral decision maker, this solution presents the greatest possible density of risk at each supplier module. Therefore, a limit is imposed on the IR within each module.

To further test the resilience of the solutions proposed by the model, we build on the work by Bertsimas et al. [15] and Alem et al. [16] on Robust optimization (RO). RO is used to investigate possible mitigation strategies of the modules IR. As part of this methodology, the values in the IRM are increased by an uncertain parameter. These parameters become iteratively larger, and their impact on the suggested SCD is recorded. This simulates the decision makers level of conservativeness by admitting that some values in the IRM may have been underestimated. The right-hand side of Fig. 4 illustrates how increasing levels of risk may stand to affect the size of supplier

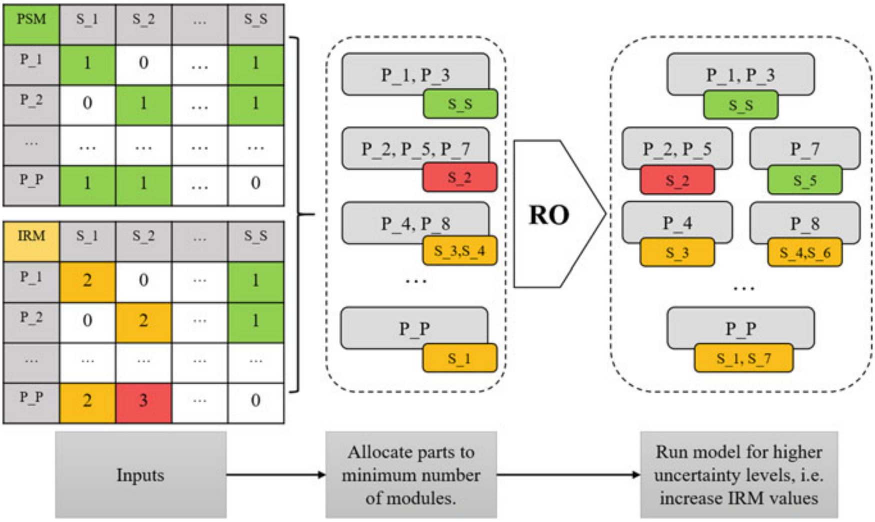


Fig. 4 Robust CP-SCD model under supplier integration risk

modules, as well as the selected sourcing flexibility. SCDs produced through higher uncertainty levels represent more pessimistic scenarios, meaning that the suppliers performance deviated considerably from the original assessment.

3.3 *Supply Chain Sustainability Performance Assessment Model: Hybrid Simulation Models*

Given the supply chain structure defined at the framework level 2, the performance assessment model intends to evaluate the impact on the supply chain sustainability of the different manufacturing processes/advanced materials being considered within the IAMAT project. Following the TBL approach, different supply chain sustainability metrics, including CO_2 emissions, energy use, supply chain costs, and workload are used for a trade-off assessment.

Building on the idea proposed by Schieritz and Größler [17], each supply chain actor is modelled as an agent. Agents locations are real and positioned in a Geographic Information System (GIS) map. A GIS map allows extracting important information, as the distance between the agents, during the model run time. Additionally, the transportation modes (ship, airplane, truck) between the supply chain actors have also been modelled as agents. The internal behaviour of the transportation modes is defined by a state-chart, while the internal behaviour of the supply chain actors depends on their role.

The model was built considering the perspective of a tier 1 aerospace supplier, servicing directly the OEM. System Dynamics (SD) and business rules are used to

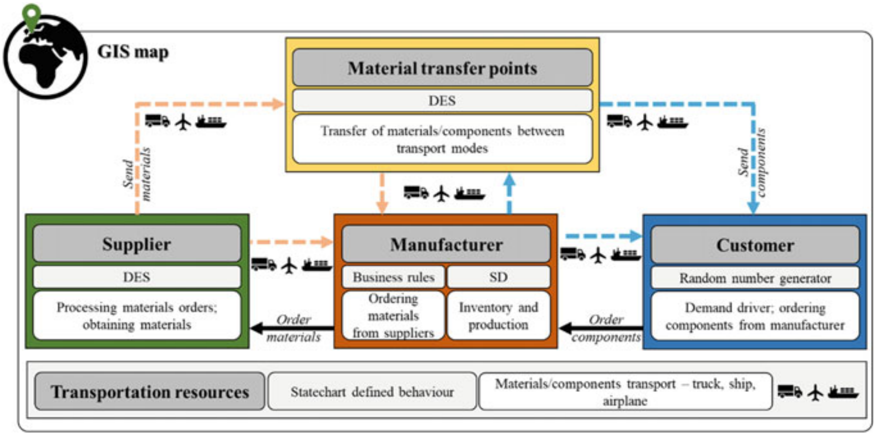


Fig. 5 Hybrid supply chain simulation model structure

represent the manufacturers structure of inventory and manufacturing. Also, Discrete Event Simulation (DES) is used for simulating the transport of materials between the supply chain agents, the transfer of material between transport modes at the airports and seaports, and the delay for obtaining the materials at the suppliers. The overall model structure is represented in Fig. 5.

Five agent types can be identified in the model: customer, manufacturer, supplier, material transfer points, and transportation resources. The customer agent, the OEM, is responsible for setting the demand in the model, placing orders to the manufacturer and receiving the modules supplied by the manufacturer (tier 1 supplier). Internally, the manufacturer has the most complex structure. The manufacturers behaviour is given by the intertwined action of a SD model and business rules. While the implemented business rules are used for establishing when and how much to order materials, the SD model, an adapted version of the model proposed by Sterman [18], for the policy structure of inventory and production, allows simulating the production activities, and the use of materials in stock. The flow rates in the SD model and the quantity of materials used to obtain the final components depend on the type of manufacturing process being considered. Within the manufacturer agent, there is a DES transportation module used to send components and receive materials.

The material transfer points correspond to the airports and sea ports where materials and final products are transferred between transportation resources. Finally, the transportation resources, that exist inside the suppliers, manufacturer, seaports, and airports, can present five internal states: when are not being used, at their owner, when loading the materials/components, going to the target location, unloading the materials/components, and returning to their owner. Each type of transportation resource has different associated monetary cost and environmental impact.

3.4 Mixed Integer Programming Models in Sustainable Supply Chains

The SCD and planning optimization referred to in Fig. 2 claims for mathematical models that can integrate the three dimensions of sustainability in the supply chain. We focus now our attention on the solution techniques for model-based quantitative research, and more precisely, in mixed integer programming (MIP) models addressing sustainability in supply chain.

Recently, some efforts have been made in this direction. However, quantitative modelling based approaches are still rare, as stressed in [19]. The author presented a comprehensive survey in this topic, and concluded that the environmental dimension plays a major role and the social dimension is often neglected. This can be explained by the challenging process of modelling this component. The integration of the three dimensions of the sustainability (economic, environmental and social) was also rare since, on that date, only two contributions integrated the three dimensions of sustainability. The author pointed out that the integration of social dimension with both the economic and environmental dimensions as a future research direction.

Some efforts were done to tackle this gap. A multi-objective mixed integer linear programming model for the design and planning of sustainable supply chains is addressed in [20]. This model integrates the three dimensions of sustainability. The economic dimension is assessed through the investment, salaries, acquisition, production, transportation, storage and disposal costs. The environmental impact is addressed using the life cycle impact assessment method ReCiPe 2008 [21]. As stated by the authors, this methodology is not widely used in supply chain optimization. The social dimension is assessed through the number of jobs created in less developed regions, namely the less populated regions.

A multi-objective mixed integer linear programming model approach for sustainable supply chain management is also addressed in [22]. Different technology constraints are also considered, including capacity and installation constraints. The social performance is assessed taking into account the gross domestic product combined with the number of created jobs.

In the scope of the integrated framework, the hybrid simulation models can provide a set of inputs to the MIP model such as the environmental metrics of these processes, among other parameters. Additionally, uncertainty can be embedded to evaluate the impact of different scenarios. For instance, demand uncertainty, delays or noncompliance with some orders may be considered in scenario analysis.

On the other hand, the computational time spent solving this type of MIP models tends to be very large. Therefore, efficient solution methods are mandatory to achieve good quality solutions in a reasonable computational time. Some of those methods include exact algorithms, meta-heuristic approaches or hybridization of both methods. Besides the use of solution methods, some reformulations of the original model can reduce the computational time within the use of a commercial solver. Figure 6 depicts the methodology in the MIP development.

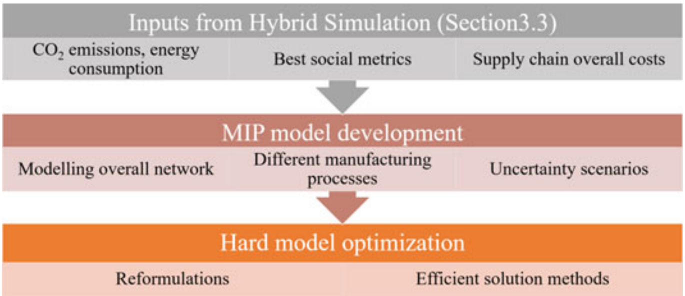


Fig. 6 MIP model development methodology

Using the MIP multi-objective model for supply chain design and planning in Mota et al. [20] as a starting point, changes adopted include the replacement of some decision variables, the exclusion of some non-effective constraints and other actions that strengthen the model. Besides the real-world context of the case study, the contribution of this type of reformulations is clear, leading to stronger models that are able to provide solutions in acceptable computational time.

4 Conclusion

In this paper, we address an integrated framework for the overall aerospace supply chain in order to ensure sustainability, flexibility, resilience, and robustness. The focus is given to the development of a sustainable supply chain that enhances the introduction of new materials/manufacturing processes or new industry requirements.

A competitive decision tool to understand the evolution and new trends of aerospace supply chains is developed, along with a multimethodology for identifying and characterizing stakeholders, and mapping the connections between them and their engagement and priorities towards sustainability. The CP-SCD model will yield a set of policies for resilient and robust supply chains combining efficient allocation of resources with risk management to define the modularity level of aircraft components and allocate modules to suppliers. The hybrid simulation performance assessment tool aids in the new material/manufacturing process selection through a trade-off analysis between different sustainability metrics. These metrics are used as input for the MIP model. Some reformulations and high efficient methods are used in order to achieve good solutions in a reasonable computational time.

The proposed framework provides a cross-functional decision support system that is able to assist decision makers in the development of various dimensions of the supply chain design and planning for new aircraft programs.

As new trends in the aerospace industry emerge, the framework must evolve to consider new scenarios and uncertainty contexts. This will keep the proposed approach updated and adaptable for both OEMs and suppliers.

Acknowledgements This work has been supported by project MITP-TB/PFM/0005/2013 in the frame of MIT-Portugal program, by COMPETE: POCI-01-0145-FEDER-007043 and FCT - Fundação para a Ciência e a Tecnologia within the project UID/CEC/00319/2013, and by the grant PD/BD/105988/2014.

References

1. Beelaerts van Blokland, W.W., Santema, S., Curran, R.: Lean supply chain management in aerospace. Encyclopedia of Aerospace Engineering (2010)
2. Tang, C.S., Zimmerman, J.D., Nelson, J.I.: Managing new product development and supply chain risks: the boeing 787 case. Supply Chain Forum Int. J. **10**(2), 74–86 (2009)
3. Sustainability Accounting Standards Board. Aerospace & Defense Research Brief (2015)
4. Slayton, R., Spinardi, G.: Radical innovation in scaling up: Boeings dreamliner and the challenge of socio-technical transitions. Technovation **47**, 47–58 (2016)
5. IATA. Iata technology roadmap
6. International, ICF. Global Aerospace Sector M&A update. Catalyst Corporate Finance (2012)
7. Mocenco, D.: Supply chain features of the aerospace industry particular case airbus and boeing. Sci. Bull.-Econ. Sci. **14**(2), 17–25 (2015)
8. Tang, C., Yeh, B., Zimmerman, J.: Boeing's 787 dreamliner: a dream or a nightmare? (2013). http://blogs.anderson.ucla.edu/global-supply-chain/2013/05/boeings-787-dreamliner-a-dream-or-a-nightmare-by-christopher-tang-based-on-work-with-brian-yeh-pwc-advisory-and-joshua.html
9. Wyman, O.: Challenges for European Aerospace Suppliers (2015)
10. Berger, R.: Aerospace industry: turning point ahead? (2016)
11. United Nations: Supply Chain Sustainability - A Practical Guide for Continuous Improvement - Second Edition (2015)
12. Bombardier stakeholder engagement. https://www.bombardier.com/en/sustainability/stakeholder-engagement.html. Accessed 09 March 2018
13. Gan, T., Steffan, M., Grunow, M., Akkerman, R.: Concurrent design of product and supply chain architectures: method and application for modularity and flexibility
14. Airbus be an airbus supplier. http://www.airbus.com/be-an-airbus-supplier.html. Accessed 22 Feb 2018
15. Bertsimas, D., Sim, M.: The price of robustness. Oper. Res. **52**(1), 35–53 (2004)
16. Alem, D., Curcio, E., Amorim, P., Almada-Lobo, B.: A computational study of the general lot-sizing and scheduling model under demand uncertainty via robust and stochastic approaches. Comput. Oper. Res. **90**, 125–141 (2018)
17. Schieritz, N., Größler, A.: Emergent structures in supply chains - a study integrating agent-based and system dynamics modeling. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences (2013)
18. Sterman, J.D.: Business dynamics: systems thinking and modeling for a complex world
19. Seuring, S.: A review of modeling approaches for sustainable supply chain management. Decis. Support Syst. **54**(4), 1513–1520 (2013)
20. Mota, B., Gomes, M.I., Carvalho, A., Barbosa-Póvoa, A.P.: Towards supply chain sustainability: economic, environmental and social design and planning. J. Clean. Prod. **105**, 14–27 (2015)
21. Goedkoop, M., Heijungs, R., Huijbregts, M., De Schryver, A., Struijs, J., Van Zelm, R.: Recipe 2008: a life cycle impact assessment method which comprises harmonised category indicators at the midpoint and the endpoint level **1** (2009)
22. Mota, B., Gomes, M.I., Carvalho, A., Barbosa-Póvoa, A.P.: Sustainable supply chains: an integrated modeling approach under uncertainty. Omega **77**, 32–57 (2018)

Critical Node Detection with Connectivity Based on Bounded Path Lengths

Fábio Barbosa, Agostinho Agra and Amaro de Sousa

Abstract For a given graph representing a transparent optical network, a given weight associated to each node pair and a given positive integer c, the Critical Node Detection problem variant addressed here is the determination of the set of c nodes that, if removed from the graph, minimizes the total weight of the node pairs that remain connected. In the context of transparent optical networks, a node pair is considered connected only if the surviving network provides it with a shortest path not higher than a given positive value T representing the optical transparent reach of the network. Moreover, the length of a path depends both on the length of its links and on its number of intermediate nodes. A path-based Integer Linear Programming model is presented together with a row generation approach to solve it. We present computational results for a real-world network topology with 50 nodes and 88 links and for $c = 2$ up to 6. The optimal results are compared with node centrality based heuristics showing that such approaches provide solutions which are far from optimal.

Keywords Critical node detection · Transparent optical networks · Path model · Decomposition approach

1 Introduction

For a given network, Critical Node Detection (CND) problems aim to optimally remove a subset of nodes (the critical nodes) in order to optimize or restrict a given network degradation metric. The problem can be defined either by upper-bounding

F. Barbosa (✉) · A. de Sousa
Instituto de Telecomunicações, Universidade de Aveiro, 3810-193 Aveiro, Portugal
e-mail: fabiobarbosa@ua.pt

A. de Sousa
e-mail: asou@ua.pt

A. Agra
CIDMA, Dept. Matemática, Universidade Aveiro, 3810-193 Aveiro, Portugal
e-mail: aagra@ua.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_2

the number of critical nodes and maximizing the degradation metric, or by lower-bounding the degradation metric and minimizing the number of critical nodes.

One most used network degradation metric is the pairwise connectivity defined as the number of node pairs that remain connected when the critical node set is removed, as in [1]. This work considers a given number c and define CND as the identification of c critical nodes minimizing the pairwise connectivity showing that the problem is NP-hard. It proposes a compact Integer Linear Programming (ILP) model which is not able to solve realistic sized instances, and a multi-start local search heuristic as an alternative for large problem instances. The work in [2] addresses two CND variants. The first variant is the same as defined in [1]. In the second variant, for a given integer L, the aim is to identify a minimum set of critical nodes, so that the largest connected component in the remaining graph contains no more than L nodes. For both variants, the authors propose alternative more compact ILP models, together with reformulations and valid inequalities. In [3], an ILP model with a non-polynomial number of constraints is proposed to the minimum pairwise connectivity CND version and a branch-and-cut method is described exploiting the fact that the linear relaxation of the model can be solved in polynomial time. In [4], a weighted version of the pairwise connectivity is used as the network degradation metric, i.e., a weight is associated to each node pair and the aim is to minimize the total weight of the connected node pairs. This work proposes ILP models which are more efficiently solved by standard solvers than the ones proposed in [1, 2] and presents computational results showing that realistic sized networks up to 75 nodes and 99 edges can be solved within seconds.

In [5], the CND problem is dealt with considering a distance-based connectivity metric, i.e., to take into consideration not only the node pairs that become disconnected but also the shortest path distance penalties between node pairs that remain connected. For a given graph with associated node costs and a given cost budget, this work considers the identification of a set of nodes within the budget whose removal maximally degrades the connectivity metric. It proposes a general ILP model that can be adapted to different distance-based metrics by proper parameter definition. In [6], the critical elements can be either nodes or links. In this work, the aim is to identify a minimum cardinality critical set of elements, referred to as a $\beta-$disruptor, whose removal results in a given pairwise connectivity target ($0 \leq \beta < 1$ denotes the connectivity fraction target). More recently, [7] assumes a budget constraint considering associated link and node costs, and extends the previous work in [6] to the case where the $\beta-$disruptor can be a mix of links and nodes. In both works, approximation methods are proposed to solve the different problem variants.

CND problems have been considered in different contexts (social networks, power grids, military networks, biology, and so on). Recently, CND problems are gaining special attention in the vulnerability evaluation of telecommunication networks to large-scale disasters. Disaster based failures can seriously disrupt any telecommunication network due to either natural, technological or malicious human causes [8] and a key component when dealing with these issues is the vulnerability evaluation of current networks against such failures [9]. In the particular case of malicious human attacks, node shutdowns, although harder to realize, are the most rewarding in the attackers perspective. The solutions provided by CND for a given number c are a worst-case scenario for simultaneous failures of up to c nodes and, when comparing different network topologies, the higher the CND value is, the more robust the network becomes to such failures.

In here, we study the CND problem in the context of transparent optical networks. In such networks, data is converted at the source into light, routed to the destination through an all-optical path, named *lightpath*, and converted back to electronic domain at the destination. To work properly, the routing path from source to destination of a ligthpath must be bounded by a *transparent reach* value which is imposed by the optical degradation suffered by the lightpath both on fibre links and on intermediate optical nodes. The optical degradation suffered by a lightpath while traversing an intermediate node is usually modelled by a given fibre length value δ, i.e. by considering it equivalent to the degradation incurred due to the transmission over a given fibre of length δ.

If some network nodes are considered critical due to some reason, then, the optical network design must take into consideration this fact. An example is [10] where the network design approach proposed in [11] is adapted to the design of a transparent optical network minimizing the impact of the simultaneous failure of a given set of critical nodes. In that work, the critical node set is given while here the aim is to determine the set of critical nodes of a given transparent optical network.

The CND variant addressed here considers, as in [4], a given weight associated to each node pair and a given positive integer c defining the number of critical nodes. Nevertheless, differently from all previous works, in this problem variant, a node pair is considered connected only if the surviving network provides it with a shortest path not higher than a given positive value T representing the transparent reach of the optical network. Moreover, the length of a path depends both on the length of its links and on its number of intermediate nodes. To describe the problem in a compact way, we would need an arc-based ILP model which requires for each pair of nodes many additional arc variables and flow conservation constraints to define the associated path. Instead, we define the problem with a path-based formulation, as in [3], and we propose an exact algorithm based on row generation to solve it. Finally, as in other works for other CND problem variants [2, 6, 7], we compare the optimal solutions of the exact method with node centrality based heuristics showing that such approaches provide solutions which are far from optimal.

The paper is organized as follows. Section 2 describes the path-based ILP model defining the CND problem in the context of transparent optical networks. Section 3 describes the row generation based approach used to solve the problem. Section 4 describes the node centrality based heuristics used in the computational results. The computational results are presented and discussed in Sect. 5. Finally, Sect. 6 presents the main conclusions of the work.

2 Path-Based ILP Model

Consider a transparent optical network represented by a graph $G = (N, E)$ where $N = \{1, \ldots, n\}$ is the set of network nodes and $E \subseteq \{(i, j) \in N \times N : i < j\}$ is the set of fibre links. For each link $(i, j) \in E$, parameter l_{ij} represents its length. For each pair of nodes (i, j), with $i \in N, j \in N, i < j$, parameter w_{ij} represents the connectivity weight of the node pair. The transparent reach of the network is denoted by parameter $T > 0$ and the fibre length equivalent to the degradation suffered by a lightpath while traversing an intermediate node is denoted by parameter $\delta > 0$. We assume that $l_{ij} \leq T$ for all fibre links; otherwise, such link is worthless and can be removed.

The set of all paths in G between $i \in N$ and $j \in N, i < j$, with length not greater than T, is denote by P_{ij}. This set is defined only for non adjacent nodes, i.e., for $(i, j) \notin E$. For each path $p \in P_{ij}$ the following binary parameters are defined: parameter β_k^p indicates whether node k (which can be an end node) is in path p or not, and parameter α_{kt}^p indicates whether link $(k, t), k < t$ is in path p or not. So, P_{ij} is composed by all paths p such that $\sum_{k=1}^{n-1} \sum_{t=k+1}^{n} \alpha_{kt}^p l_{kt} + \delta\big(\sum_{k=1}^{n} \beta_k^p - 2\big) \leq T$. Although δ can be incorporated in the link length and, therefore, the use of parameters β_k^p could be omitted, we opted to include them in order to ease the reading.

Parameter $c \in \mathbb{N}$ represents the number of critical nodes. For each node $i \in N$, we consider a binary variable v_i indicating whether i is a critical node or not. For each node pair (i, j), with $i, j \in N : i < j$, the binary variable u_{ij} is 1 if nodes i and j are connected through a path satisfying the transparent reach T, and 0 otherwise.

A path formulation for the CND problem is given by the following ILP model.

$$\min z := \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij}\, u_{ij} \tag{1}$$

$$s.t. \sum_{i=1}^{n} v_i \leq c \tag{2}$$

$$u_{ij} + v_i + v_j \geq 1, (i, j) \in E, \tag{3}$$

$$u_{ij} + \sum_{k=1}^{n} \beta_k^p v_k \geq 1, (i, j) \notin E,\ p \in P_{ij}, \tag{4}$$

$$v_i \in \{0, 1\}, \qquad i \in N, \tag{5}$$

$$u_{ij} \in \{0, 1\}, \qquad i, j \in N : i < j. \tag{6}$$

The objective (1) is to minimize the total weighted connectivity in the surviving graph, i.e. the sum of the weights of the node pairs that remain connected after the critical nodes are removed. Constraint (2) ensures that at most c nodes are selected as critical nodes (in any optimal solution, c nodes are selected). Constraints (3) guarantee that a pair of adjacent nodes is connected if none of the two nodes is a critical node. Constraints (4) are the generalization of constraints (3) for the node pairs that are not adjacent in G: the node pair (i, j), with $i < j$, is connected if there is a path $p \in P_{ij}$ such that none of its nodes is a critical node. Finally, constraints (5)–(6) are the variable domain constraints.

Notice that constraints (6) can be replaced by $u_{ij} \geq 0$. Since variables v_i are binary, constrains (3)–(4) impose $u_{ij} \geq 1$ if (i, j) is a connected node pair, and will be redundant in presence of constraints $u_{ij} \geq 0$, otherwise. As the objective function is a minimization function, then $u_{ij} = 1$ if (i, j) is a connected pair and $u_{ij} = 0$ otherwise. The resulting mixed integer linear problem (MILP) will be considered henceforward.

3 A Row Generation Approach

The path formulation presented in the previous section includes the family of constraints (4) whose number increases exponentially with the input data. The exact number of constraints depends on the graph topology, the length of the links and on the parameters T and δ. However, the MILP can become too large for relative small size instances. Here we propose an exact algorithm, based on row generation, where inequalities (4) are initially ignored and the relaxed MILP problem is solved. Then, the separation problem associated with inequalities (4) is solved. If a violated inequality is found, it is added to the model and the MILP is solved again. The process is repeated until no violated inequality is found. The exact algorithm is described in Algorithm 1.

The separation problem associated with constraints (4) is solved in the following way. First, we compute the subgraph that results from G when the critical nodes and the corresponding incident edges are removed and we add δ to the length of all non-removed edges. Then, we determine the shortest paths between all pairs of nodes in this subgraph with the new lengths. Finally, each shortest path whose length is not higher than $T + \delta$ is used to generate a new inequality (4) that is added to the model (by adding δ to each edge length and since, for each path, the number of intermediate nodes is equal to the number of edges minus one, the shortest path value with the new lengths is equal to the length value with the original lengths plus δ).

4 Node Centrality Based Heuristics

Heuristic methods based on node centrality measures are commonly used in the literature to quickly compute sets of critical nodes. Algorithm 2 presents a general heuristic framework for using these measures. In each iteration a node is selected according to the chosen node centrality measure (step 3) and removed from the graph (step 4). The heuristic finishes when c nodes are selected.

Algorithm 1 Exact algorithm for the CND problem.

Solve MILP model without constraints (4) and let (u^*, v^*) be the optimal solution
repeat
  Set NCuts $\leftarrow 0$ and $C \leftarrow \{i \in N : v_i^* = 1\}$
  Update the subgraph graph $G^C = (N \setminus C, E^C)$ where $E^C = \{(i, j) \in E : i, j \notin C\}$
  for all node pair $(i, j) \notin E^C$ with $i < j$ **do**
    Run Dijkstra algorithm (adding δ to the length of each edge) to find the shortest path $p_{ij} \in P_{ij}$ and its length d_{ij}
    if $d_{ij} \leq T + \delta$ and $u_{ij}^* + \sum_{k=1}^{n} \beta_k^{p_{ij}} v_k^* = 0$ **then**
      Add constraint (4) corresponding to path p_{ij}
      NCuts $\leftarrow$ NCuts $+1$
    end if
  end for
  if NCuts > 0 **then**
    Solve MILP model with the added constraints. Update (u^*, v^*)
  end if
until Ncuts = 0

Algorithm 2 Iterative heuristic approach based on node centrality.

Set $C \leftarrow \emptyset$ and $G' \leftarrow (N, E)$
for all $k = 1$ to c **do**
  Using the selected node centrality measure, select the central node i of graph G'
  Remove from graph G' node i and all edges incident to node i
  Set $C \leftarrow C \cup \{i\}$
end for

We consider three node centrality measures to select the central nodes in graph G':

- Node degree centrality. The selected node is the one with highest degree in graph G'.
- Node closeness centrality. The closeness of node i is defined as the sum of the inverse of the distances between i and each of the remaining nodes: $c(i) = \sum_{j \in N \setminus \{i\}} \frac{1}{\mathrm{d}_{ij}(G')}$, where $\mathrm{d}_{ij}(G')$ is the shortest path length between nodes i and j in G'. The node with highest closeness is selected.
- Node betweenness centrality. For graph G', the betweenness of node i is the number of shortest paths between all nodes in G', with length not greater than T, that include node i as an intermediate node. The node with highest betweenness is selected.

Again, the shortest path lengths computed for the Closeness and Betweenness centralities consider the length δ associated to each intermediate node and are computed in the same way as described in the previous section for the separation problem.

5 Computational Results

Here we report the computational experiments carried out to test the proposed exact solution approach for the CND problem and to compare it with the centrality based heuristics. Additionally, some insight on the solutions for the CND problem is given.

All computations were performed using the optimization software *Gurobi Optimizer* version 7.5.1, with programming language *Julia* version 0.6.0, running on a PC with a Intel Core i5, 1.7 GHz (up to 2.4 GHz) and 6 GB RAM.

The test instances are based on the Germany50 network topology, a telecommunication backbone network with 50 nodes and 88 edges [12]. The transparent reach depends on the Optical Transport Units installed. Current values go up to 2500 km. Hence, for the transparent reach parameter T we consider values in $\{1417, 1500, 1600, 1800, 2000, 2500\}$, where value 1417 is the maximum length among the shortest paths between all node pairs of Germany50 (a smaller value does not allow the network to be optically transparent between all node pairs). In all cases, we have considered $\delta = 60$ km.

Table 1 presents the results obtained for $c \in \{2, 3, 4, 5, 6\}$ and considering the scenario where each pair of nodes has an unitary connectivity weight i.e. $w_{ij} = 1$ for all $i, j \in N$ with $i < j$. In addition to the number of critical nodes c and the transparent reach T given in the first two columns, column UB provides the trivial upper bound when all pairs of remaining nodes are connected (i.e., the critical nodes do not turn the surviving network into more than one component), which is given by $\frac{(n-c)\times(n-c-1)}{2}$. The next three columns show the objective function value (in this case of unitary weights it coincides with the total number of connected node pairs after the removal of the critical nodes) of the feasible solution obtained with the heuristic based on the corresponding centrality measure. Hence, a feasible solution with value equal to UB means that all the remaining nodes are connected after the critical nodes have been removed. The last four columns are obtained with the exact approach. Column *CND* gives the optimal value, column *Iterations* gives the number of times the relaxed MILP was solved running Algorithm 1, column *Time* gives the total elapsed running time in seconds, and column *Cuts* gives the total number of constraints (4) added to the model in order to reach the optimal solution.

Although the node centrality measures are commonly used in the literature to quickly compute critical node sets, it is possible to conclude from the results that these sets are not minimizing the global connectivity of the graph. Nevertheless, the total running time of all heuristics based on the node centrality take less than half a second (not presented in the table), while the exact approach, in some cases, take almost one minute.

Figure 1 presents a graphical scaled representation of the network and the optimal critical node sets for each $c \in \{2, 3, 4, 5, 6\}$. Figure 2 gives a similar representation of the critical node sets selected using the node centrality based heuristics for $c = 6$. These figures illustrate the reason behind the difference between the CND objective values and the number of connected node pairs obtained using node centrality based methods. On one hand, if the critical node set is optimally selected (i.e minimizing

Table 1 Computational results for unitary weights

c	T (km)	UB	Degree	Closeness	Between	CND	Iterations	Time (s)	Cuts
2	1417	1128	1126	1127	1126	1026	17	36	3645
	1500		1128	1128	1128	1036	18	46	3689
	1600		1128	1128	1128	1036	18	49	3756
	1800		1128	1128	1128	1036	18	57	3816
	2000		1128	1128	1128	1036	18	54	3826
	2500		1128	1128	1128	1036	18	54	3826
3	1417	1081	1076	1076	1060	711	5	3	2557
	1500		1081	1081	1071	711	5	3	2603
	1600		1081	1081	1081	711	5	3	2593
	1800		1081	1081	1081	711	5	3	2620
	2000		1081	1081	1081	711	5	3	2620
	2500		1081	1081	1081	711	5	3	2620
4	1417	1035	1025	880	834	640	11	20	2892
	1500		1032	907	869	640	11	25	2969
	1600		1034	929	908	640	10	19	3031
	1800		1035	976	950	640	10	21	3157
	2000		1035	1010	989	640	10	22	3273
	2500		1035	1035	1035	640	10	22	3359
5	1417	990	980	809	682	496	7	9	3092
	1500		987	836	711	496	7	10	3196
	1600		989	867	743	496	7	11	3296
	1800		990	924	804	496	7	12	3429
	2000		990	959	862	496	7	14	3518
	2500		990	990	990	496	7	14	3600
6	1417	946	933	653	486	415	12	36	3319
	1500		940	678	487	415	12	38	3407
	1600		944	708	487	415	12	37	3492
	1800		946	766	487	415	12	35	3574
	2000		946	825	487	415	12	37	3607
	2500		946	946	487	415	12	36	3615

the global connectivity), the graph resulting from the removal of the critical nodes is disconnected into several components. On the other hand, node centrality based methods do not aim to disconnect the graph but only to select the most influential nodes (using node centrality criteria), which results is less disconnected graphs. When the resulting graph is not disconnected, the total number of connected node pairs increases for larger transparent reach values up to a point such that its value becomes equal to the upper bound.

Regarding the impact of the transparent reach value T in the results, Table 1 shows that this parameter has little impact on the CND optimal value for this network. This

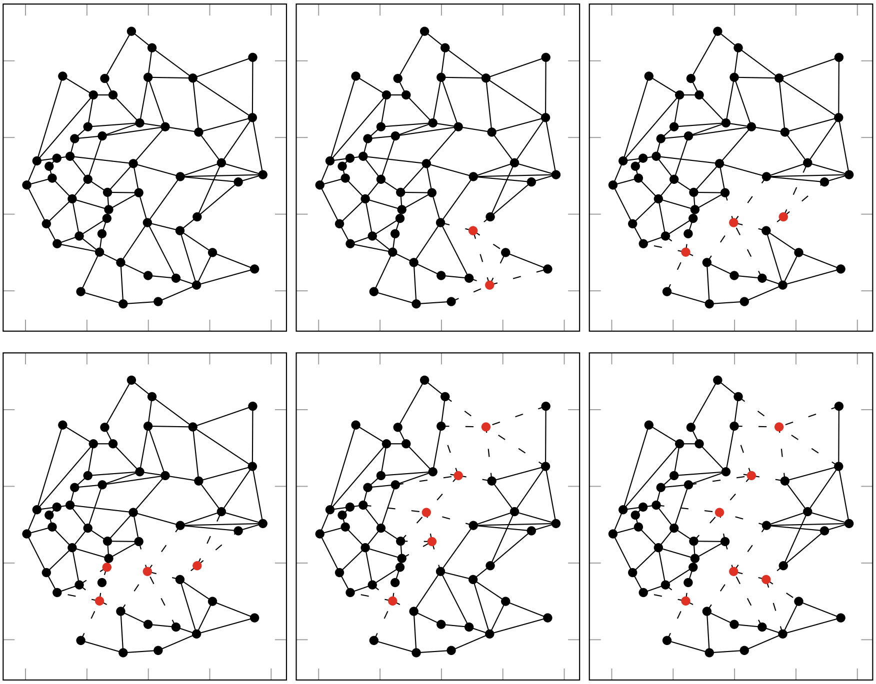


Fig. 1 Germany50 on top left, and the network resulting from the removal of the optimal critical node set for each size $c \in \{2, 3, 4, 5, 6\}$ (for any $T \geq 1500$)

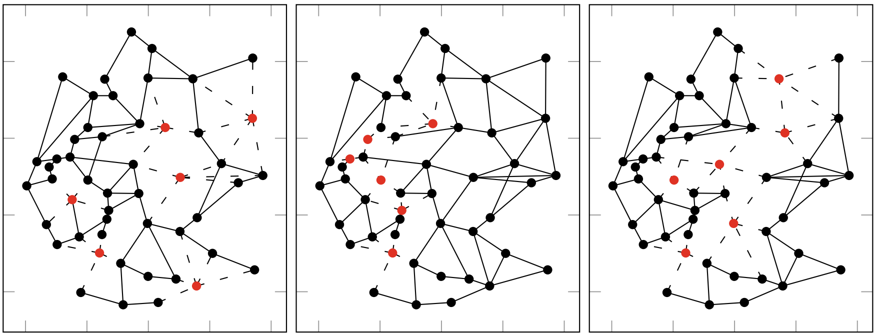


Fig. 2 The network resulting from the removal of node sets computed using centrality methods: Degree, Closeness and Betweenness, respectively (for $c = 6$)

can be explained by observing that when the critical nodes are removed from graph G, the resulting graph becomes disconnected and each resulting component fully satisfies connectivity for any $T \geq 1417$ km. However, this behaviour is not observed when node centrality based methods are used. In several instances the total number of connected node pairs increases for larger transparent reach values T to a point that the upper bound of the problem is reached. In this case, the graph resulting from the removal of the selected critical nodes is totally connected, as can be seen in Fig. 2 for the Degree and Closeness cases.

In order to test the effect of having different connectivity weights between different pairs of nodes, next we consider the case where the nodes corresponding to the five largest German cities (in terms of population) have higher impact than all other 45 nodes. First, we assign a node weight of 4 to the nodes corresponding to the five larger cities and a node weight of 1 to all other nodes. Then, the connectivity weight w_{ij} between node i and node j is given by the multiplication of the weights of the two nodes.

In Table 2, we present the results obtained with the exact approach for the CND problem with the weights computed as explained above. For these weight values an upper bound (column UB) is obtained when the critical nodes are c nodes that do not correspond to the largest cities and the resulting subgraph is fully connected. In these cases, the upper bound is given by the number of node pairs with two largest cities multiplied by 4^2 plus the number of node pairs with one largest city multiplied by 4 plus the number of node pairs with no largest cities multiplied by 1, i.e. UB $:= 4^2 \times \frac{5\times 4}{2} + 4 \times 5 \times (45 - c) + \frac{(45-c)\times(45-c-1)}{2}$. The remaining columns have the same meaning as the corresponding ones in Table 1. The last two columns were added to compare these cases against the unitary weights cases. Column *Con. Pairs* gives the total number of connected node pairs after the removal of the critical nodes of these cases. Column *Prev. CND* gives the (previous) CND optimal value for the unitary weights cases.

Concerning the performance of the exact algorithm proposed in Sect. 3, by comparing the number of iterations, running time and number of added cuts between Tables 1 and 2, one can observe that the results are nearly identically. That means the weights do not have a great impact on the performance of CND method presented in Algorithm 1. Concerning the number of connected node pairs of the CND solutions, as expected, the number of connected node pairs considering different weights is higher than in the previous cases. This is because now the optimal set of critical nodes is a mixture between selecting nodes representing the largest cities and nodes that disconnect more the network.

Table 3, presents the ratio between the optimal values and the corresponding theoretical upper bounds. These results show that the more realistic scenario with different weights show that the network under consideration is less resilient to multiple node failures than the simplest scenario of considering equal importance to all node pairs. Moreover, as expected for both cases, the percentage of non-critical node pairs that remain connected decreases for larger values of critical nodes c.

Table 2 Computational results for different weights

c	T (km)	UB	CND	Iterations	Time (s)	Cuts	Con. Pairs	Prev. CND
2	1417	1923	1577	15	24	2755	1127	1026
	1500		1577	14	21	2798	1127	1036
	1600		1578	14	23	2830	1128	1036
	1800		1578	14	23	2862	1128	1036
	2000		1578	14	23	2868	1128	1036
	2500		1578	14	22	2868	1128	1036
3	1417	1861	1224	6	6	2647	711	711
	1500		1224	6	6	2672	711	711
	1600		1224	6	6	2688	711	711
	1800		1224	6	6	2698	711	711
	2000		1224	6	6	2702	711	711
	2500		1224	6	6	2702	711	711
4	1417	1800	1044	8	11	3291	675	640
	1500		1044	8	12	3368	675	640
	1600		1044	8	12	3432	675	640
	1800		1044	8	12	3511	675	640
	2000		1044	8	12	3525	675	640
	2500		1044	8	12	3526	675	640
5	1417	1740	850	12	35	3902	526	496
	1500		850	11	31	4052	526	496
	1600		850	11	29	4184	526	496
	1800		850	11	32	4374	526	496
	2000		850	11	32	4507	526	496
	2500		850	11	33	4697	526	496
6	1417	1681	653	10	26	3823	446	415
	1500		653	10	27	3981	446	415
	1600		653	10	29	4125	446	415
	1800		653	10	29	4179	446	415
	2000		653	10	30	4278	446	415
	2500		653	10	31	4377	446	415

Table 3 Ratio (%) between CND optimal value and the upper bound ($T \geq 1600$)

c	2	3	4	5	6
Unitary	91.8	65.8	61.8	50.1	43.9
Weighted	82.1	65.8	58.0	48.9	38.9

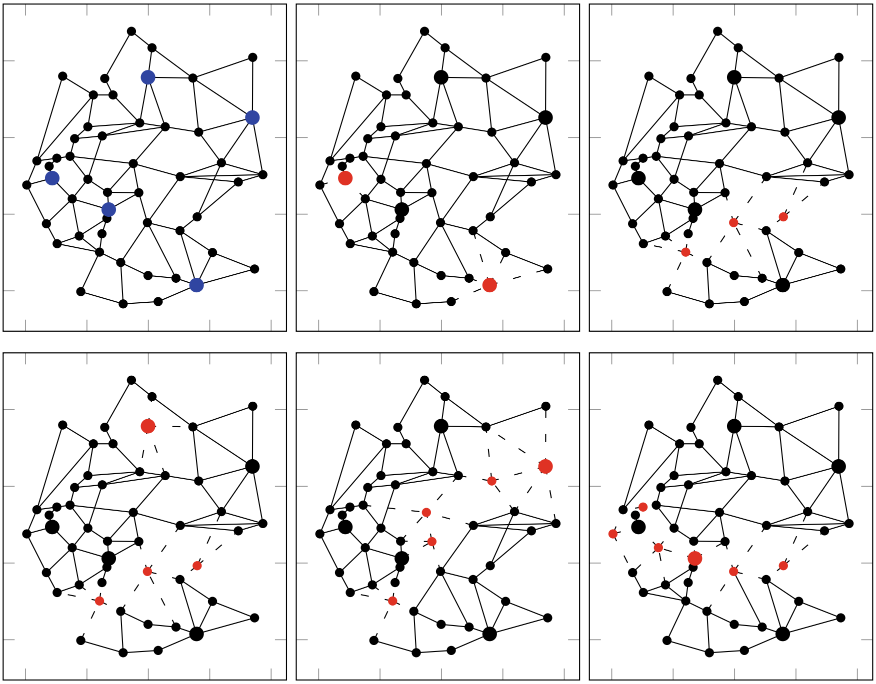


Fig. 3 Germany50 with the five main cities highlighted on top left, and the network obtained with the removal of the optimal node set for $c \in \{2, 3, 4, 5, 6\}$ ($T \geq 1600$)

Figure 3 depicts the network where the five nodes corresponding to the largest cities are highlighted in blue, and the optimal critical node sets obtained for the weighted values and for the different values of c.

Comparing Fig. 3 with Fig. 1, one can observe that, with exception of $c = 3$, the optimal critical node set changes when different weights are considered. For example, with just two critical nodes, instead of disconnecting the graph (as in Fig. 1), the optimal solution is obtained by selecting two nodes corresponding to largest cities. In the last scenario with 6 critical nodes, the set of critical nodes changes considerably from the unitary weights to the different weights case. With unitary weights, the CND solution splits the network into three components each one with a large number of nodes. With different weights, the optimal solution is also obtained by splitting the graph into three components but the components are not balanced in terms of number of nodes (there is one component with only two nodes). Instead, one node corresponding to a largest city is selected as a critical node and the other four nodes representing largest cities are split among the 3 components.

6 Conclusions

In this work, we have addressed the Critical Node Detection (CND) problem in the context of a real transparent optical backbone network, a problem which is gaining a special interest in the vulnerability evaluation of networks. A path-based ILP model was proposed. Although path-based ILP formulations are not as efficient as the compact models for the traditional CND problem, such compact formulations do not allow to include directly the connectivity constraints based on bounded path lengths, as imposed by transparent optical networks. Based on the path formulation, an exact approach, based on row generation, was described allowing to compute the optimal set of critical nodes for the Germany50 network topology. The computational results also showed that the heuristics derived from the commonly used node centrality measures to quickly identify critical nodes, are not able, in general, to identify the optimal critical node set. Moreover, the results have shown that the tested backbone network has not a topology resilient to multiple node failures. In fact, with a simultaneous failure of only 10% of the network nodes ($c = 5$), it is possible to reduce the global connectivity of the network in about 50%. On top of that, the computational results show that in the more realistic scenario where node pairs have different weights, the simultaneous failure of the critical nodes is able to reduce the network connectivity even more than in the unitary weights case.

For a given network topology, the CND solution provides a worst-case measure of the network vulnerability to multiple node failures. As future research, we aim to develop efficient methods, both deterministic and stochastic, to upgrade the current network topology aiming to improve its CND value turning it more robust to multiple node failures.

Acknowledgements This article is based upon work from COST Action CA15127 ("Resilient communication services protecting end-user applications from disaster-based failures—RECODIS"), supported by COST (European Cooperation in Science and Technology), and from project CENTRO-01-0145-FEDER-029312 (ResNeD) supported by FEDER Funds and National Funds through FCT (Fundação para a Ciência e a Tecnologia), Portugal. First author was supported by FCT through Ph.D. grant SFRH/BD/132650/2017. Second author was supported by FCT through CIDMA within project UID/MAT/04106/2013.

References

1. Arulselvan, A., Commander, C.W., Elefteriadou, L., Pardalos, P.M.: Detecting critical nodes in sparse graphs. C&OR **36**, 2193–2200 (2009)
2. Veremyev, A., Boginski, V., Pasiliao, E.: Exact identification of critical nodes in sparse networks via new compact formulations. Optim. Lett. **8**, 1245–1259 (2014)
3. Di Summa, M., Grosso, A., Locatelli, M.: Branch and cut algorithms for detecting critical nodes in undirected graphs. Comput. Optim. Appl. **53**(3), 649–680 (2012)
4. Santos, D., de Sousa, A., Monteiro, P.: Compact models for critical node detection in telecommunication networks. Electron. Notes Discret. Math. **64**, 325–334 (2018)

5. Veremyev, A., Prokopyev, O., Pasiliao, E.: Critical nodes for distance-based connectivity and related problems in graphs. Networks **66**(3), 170–195 (2015)
6. Dinh, T., Xuan, Y., Thai, M., Pardalos, P., Znati, T.: On new approaches of assessing network vulnerability: hardness and approximation. IEEE/ACM Trans. Netw. **20**(2), 609–619 (2012)
7. Dinh, T., Thai, M.T.: Network under joint node and link attacks: vulnerability assessment methods and analysis. IEEE/ACM Trans. Netw. **23**(3), 1001–1011 (2015)
8. Rak, J., et al.: RECODIS: resilient communication services protecting end-user applications from disaster-based failures. In: Proceeding of ICTON, paper We.D1.4. (2016)
9. Gomes, T., et al.: A survey of strategies for communication networks to protect against large-scale natural disasters. In: Proceeding of RNDM, 2016, pp. 11–22 (2016)
10. Barbosa, F., de Sousa, A., Agra, A.: The design of transparent optical networks minimizing the impact of critical nodes. Electron. Notes Discret. Math. **64**, 165–174 (2018)
11. Agra, A., de Sousa, A., Doostmohammadi, M.: The minimum cost design of transparent optical networks combining grooming, routing, and wavelength assignment. IEEE/ACM Tran. Netw. **24**(6), 3702–3713 (2016)
12. Orlowski, S., Wessaly, R., Pioro, M., Tomaszewski, A.: SNDlib 1.0 survivable network design library. Networks **55**(3), 276–286 (2010)

A Benders Decomposition Algorithm for the Berth Allocation Problem

Flávia Barbosa, José Fernando Oliveira, Maria Antónia Carravilla and Eduardo Ferian Curcio

Abstract In this paper we present a Benders decomposition approach for the Berth Allocation Problem (BAP). Benders decomposition is a cutting plane method that has been widely used for solving large-scale mixed integer linear optimization problems. On the other hand, the Berth Allocation Problem is a NP-hard and large-scale problem that has been gaining relevance both from the practical and scientific points of view. In this work we address the discrete and dynamic version of the problem, and develop a new decomposition approach and apply it to a reformulation of the BAP based on the Heterogeneous Vehicle Routing Problem with Time Windows (HVRPTW) model. In a discrete and dynamic BAP each berth can moor one vessel at a time, and the vessels are not all available to moor at the beginning of the planning horizon (there is an availability time window). Computational tests are run to compare the proposed Benders Decomposition with a state-of-the-art commercial solver.

Keywords Berth allocation · Mixed integer linear problem · Benders decomposition

F. Barbosa (✉) · E. F. Curcio
INESC TEC, Porto, Portugal
e-mail: fbarbosa@inesctec.pt

E. F. Curcio
e-mail: eduardo.curcio@fe.up.pt

J. F. Oliveira · M. A. Carravilla
INESC TEC, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
e-mail: jfo@fe.up.pt

M. A. Carravilla
e-mail: mac@fe.up.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_3

1 Introduction

Benders Decomposition is a solution method used for solving large-scale mixed integer linear programming problems. It can be described as a divide-and-conquer strategy: in each iteration, new constraints are added to the problem, making it progress towards a solution. The variables of the original problem are divided into two subsets. A first-stage master problem is solved over the first set of variables. Once these variables are fixed, the values for the second set of variables are determined in a second-stage subproblem. The resulting subproblem is a continuous linear program and the standard duality theory can be used to develop cuts.

According to [15], for more than five decades the Benders Decomposition algorithm has been used to tackle problems in many diverse fields. Computational approaches based on Benders Decomposition to the constrained minimum break problem are proposed in [14]. In [10] a Benders Decomposition approach for the generalized formulation of the integrated aircraft routing and crew scheduling was implemented. A Benders-like decomposition approach was proposed in [2] for solving a capacitated facility location problem with two decision makers. Exact solution algorithms based on Benders decomposition are presented in [8] for the traveling salesman problem with risk constraints. This paper develops a Benders Decomposition approach for the Berth Allocation Problem (BAP).

The BAP has gained prominence in Operations Research, because the demand for vessel transportation has increased over the last years. In [6] it is pointed out that in container transport, vessels arrival schedule should be adopted in advance. In [18] the container terminal operations are divided into quayside and landside operations. Quayside operations include the allocation of berths to arriving vessels, problem known as the Berth Allocation Problem (BAP), the assignment of available cranes to vessels, known as the Quay Crane Assignment Problem (QCAP) as well as the scheduling of cranes, which is known as the Quay Crane Scheduling Problem (QCSP).

Given a set of vessels and a set of berths with availability time windows, the BAP main goal is to determine in which berth each vessel must be moored, the precedence relationship between the vessels assigned to the same berth and the time each vessel starts being serviced. Each vessel has an arrival time, an expected departure time from the port and berth dependent processing times, and the objective is to minimize the weighted sum of vessel service times, i.e. waiting time until service starts plus processing time. Throughout the literature, the BAP has already been modeled in different ways. In [13] the BAP is modeled as a sequencing problem for a port located in the eastern coast of India. In [7] the BAP is represented in a time-space diagram: the horizontal axis represents the time units and the vertical axis represents berth sections. Analogously the vessel is represented as a rectangle whose length is the processing time and whose height is the vessel size. In [1] the BAP is viewed as a heterogeneous vehicle routing problem with time windows (HVRPTW), in which berths corresponds to vehicles and there is a single depot.

This paper is organized as follows. Section 2 gives a description of the Benders Decomposition algorithm and its enhancements. Section 3 proposes a formulation of the Berth Allocation Problem to which the decomposition will be applied. Section 4 details the Benders Decomposition algorithm applied to the BAP. Section 5 reports the results.

2 Benders Decomposition

Benders Decomposition is a cutting plane method which reduces the search region by adding linear constraints, preserving the original feasible region.

Suppose a mixed integer linear problem of the form:

$$\min c^T x + f^T y \tag{1}$$
$$\text{s.t. } Ax + By \geq b \tag{2}$$
$$y \in Y \tag{3}$$
$$x \geq 0 \tag{4}$$

If (1)–(4) is an easier optimization problem in x when y is fixed, y are refered as "complicating variables" in [5].

With y fixed to a feasible integer configuration $\bar{y}$, the resulting model to be solved is given by:

$$\min c^t x \tag{5}$$
$$\text{s.t. } Ax \geq b - B\bar{y} \tag{6}$$
$$x \geq 0 \tag{7}$$

with the associate dual problem:

$$\max \ (b - B\bar{y})^T u \tag{8}$$
$$\text{s.t. } A^T u \leq c \tag{9}$$
$$u \geq 0 \tag{10}$$

Defining z as the objective function of (5)–(7) and $\bar{u}$ as the variable values of the dual problem (8)–(10), the valid inequality

$$z \geq (b - By)^T \bar{u} \tag{11}$$

is a *Benders optimality cut*. In each iteration of the Benders algorithm, a master problem is solved:

$$\min z \tag{12}$$
$$\text{s.t. } z \geq (b - By)^T \, \bar{u} \tag{13}$$
$$z \in \Re \tag{14}$$
$$y \in Y \tag{15}$$

whose solution $\bar{y}$ is the master problem solution and will be used to define the following subproblem (5)–(7).

If the subproblem (primal problem) is infeasible for a fixed $\bar{y}$, the dual formulation is unbounded, according to [17]. In this case, it is necessary to add a feasibility cut. Let $\bar{\alpha}$ be the extreme ray of the dual formulation. The *Benders feasibility cut*

$$\bar{\alpha}^T \, (b - By) \leq 0 \tag{16}$$

is formulated and added to the master problem in order to eliminate the infeasible solution.

It's noteworthy to mention that the master problem gives a lower bound (LB) and the subproblem gives an upper bound (UB) for the original problem. The procedure iterates until $UB - LB < \varepsilon$.

Some different enhancement strategies may be proposed to improve and accelerate the convergence of the Benders Decomposition method, most of them taking into account the special characteristics of each problem. The two most important ones are presented below.

2.1 Combinatorial Benders Cut

The Benders Decomposition can also be used as an alternative to the "big-M" approach, where large positive coefficients are introduced to activate/deactivate the conditional constraints. Suppose a constraint of the form:

$$a_i^T x \geq b_i - (1 - y_{j(i)}) * M \tag{17}$$
$$y_{j(i)} \in \{0, 1\} \;\; \forall j \; (i) \tag{18}$$
$$x \geq 0 \tag{19}$$

The binary variables force some feasibility properties:

$$y_{j(i)} = 1 \Rightarrow a_i^T x \geq b_i \; \forall i \in I \tag{20}$$

In [3] an automatic problem reformulation for mixed integer linear problems involving logical implications modeled through big-M coefficient was proposed and computationally analyzed.

Due to the presence of the big-M coefficients, the linear relaxation of the mixed integer linear problem model is poor and the resulting Benders cuts are weak and still depend on the big-M values. Therefore, the classical Benders approach can be viewed as a tool to speed-up the solution of the LP relaxation. The aim of this approach is to remove the model dependency on the big-M coefficients.

The master problem is solved to integrality. If this problem turns out to be infeasible, then the original problem also is. Otherwise, let y^* be an optimal solution. If the subproblem is infeasible for this solution, a Minimal Infeasible Subsystem $C \subset I$ is sought, i.e., any inclusion-minimal set of row-indices of system (20) such that the linear subsystem:

$$a_i^T x \geq b_i \tag{21}$$

has no feasible solution x.

At least one binary variable $y_{j(i)}$ has to be changed in order to break the infeasibility and these implication constraints are modeled through the following Combinatorial Benders' (CB) cuts:

$$\sum_{j \in C: y^*_{j(i)}=0} y_j + \sum_{j \in C: y^*_{j(i)}=1} \left(1 - y_j\right) \geq 1 \tag{22}$$

CB cuts of this type are generated in correspondence to a given infeasible solution y^*, and added to the master problem.

2.2 *Optimality Cut Disaggregation*

If the Benders subproblem can be separated into independent subproblems, disaggregated cuts can be obtained in order to accelerate convergence of the Benders Decomposition algorithm. The subproblems are solved in parallel and multiple cuts formed by the dual optimal solutions are added to the Benders master problem simultaneously.

Each subproblem k generates an optimality cut, analogous to (11). According to [16], these cuts include the same information as the primal Benders cut and restrict the solution space of the master problem in a more accurate-exact way.

3 Berth Allocation Problem Formulation

There are several models the BAP. This paper closely follows [1]. The BAP is considered discrete and dynamic, and viewed as a heterogeneous vehicle routing problem with time windows (HVRPTW), in which berths correspond to vehicles and there is a single depot. Let N be the set of nodes and A the set of arcs. By using the HVRPTW model, the problem is defined on a graph $G = (V, A)$ where the set $V = N \cup \{o, d\}$ contains a vertex for each vessel and vertices o and d represent, respectively, the origin and destination nodes for any route in the graph.

The set of arcs is a subset of $V \times V$. Let N be the set of vessels and M the set of berths. Each vessel $i \in N$ has an arrival time a_i, an expected departure time from the port b_i, processing times p_i^k, dependent on the assigned berth $k \in M$, and a relative importance v_i, which represents a service priority and is modeled as a weight in the objective function. For the origin and destination vertices, the berths can be available at different times inducing a time window $\left[s^k, e^k\right]$. Each binary variable l_{ij}^k, $k \in M$, $(i, j) \in A$, takes the value one if vessel j immediately succeeds vessel i at berth k and is zero otherwise. Each continuous variable x_i^k, $i \in V$, $k \in M$, gives the time when vessel i starts being serviced at berth k (if vessel i does not use berth k, $x_i^k = a_i$). The variables x_o^k and x_d^k define respectively the start and end time of activities at berth $k \in M$. Defining $M_{ij}^k = \max\left\{b_i + p_i^k - a_j, 0\right\}$, the following model can be written:

$$\min \sum_{i \in N} \sum_{k \in M} v_i \left(x_i^k + p_i^k \sum_{j \in N \cup \{d\}} l_{ij}^k \right) \tag{23}$$

$$\text{s.t.} \sum_{k \in M} \sum_{j \in N \cup \{d\}} l_{ij}^k = 1 \ \forall i \in N \tag{24}$$

$$\sum_{j \in N \cup \{d\}} l_{oj}^k = 1 \ \forall k \in M \tag{25}$$

$$\sum_{i \in N \cup \{o\}} l_{id}^k = 1 \ \forall k \in M \tag{26}$$

$$\sum_{j \in N \cup \{d\}} l_{ij}^k = \sum_{j \in N \cup \{o\}} l_{ji}^k \ \forall k \in M, i \in N \tag{27}$$

$$x_i^k + p_i^k - x_j^k \leq \left(1 - l_{ij}^k\right) M_{ij}^k \ \forall k \in M, (i, j) \in A \tag{28}$$

$$x_i^k \geq a_i \ \forall k \in M, i \in N \tag{29}$$

$$x_i^k + p_i^k \sum_{j \in N \cup \{d\}} l_{ij}^k \leq b_i \ \forall k \in M, i \in N \tag{30}$$

$$s^k \leq x_o^k \ \forall k \in M \tag{31}$$

$$x_d^k \leq e^k \ \forall k \in M \tag{32}$$

$$l_{ij}^k \in (0, 1) \ \forall k \in M, (i, j) \in A \tag{33}$$

$$x_i^k \geq 0 \ \forall k \in M, i \in V. \tag{34}$$

The objective function (23) minimizes the weighted sum of the vessel service times. Constraints (24) state that each vessel must be assigned to exactly one berth k, while constraints (25) and (26) guarantee respectively that for each berth k the degree of the origin and destination nodes is one. Constraints (27) ensure the flow conservation for the remaining vertices. Constraints (28) guarantee consistency between berthing time and mooring sequence on each berth. Constraints (29) and (30) enforce the time window requirements for each vessel. Constraints (31) and (32) enforce the berth availability time windows. Finally, constraints (33) and (34) define the domains of the decision variables.

Given the intractability of the problem, as discussed in [1] the basic formulation is improved. First, a class of valid inequalities is derived to improve the lower bound of the x_i^k variables:

$$s^k l_{oj}^k + \sum_{i \in N} \left(\max\left\{a_i, s^k\right\} + p_i^k\right) l_{ij}^k \leq x_j^k \; \forall k \in M, j \in N \tag{35}$$

Then variables l_{ij}^k are fixed to guarantee that, if vessel j arrives before vessel i and has a shorter processing time, then vessel j will not follow vessel i at berth k.

Besides that, the concept of berth types instead of individual berths was created as some berths are identical (in the sense of processing times). The parameter β_k, $k \in M$, was introduced to represent the number of berths of type k in the problem instance. Constraints (25) and (26) and the domain of variable l_{od}^k were also adequately modified.

4 Decomposition Approach for the Berth Allocation Problem

According to [11], the computational complexity of the BAP lies in the dynamic arrival process of the vessels. If the vessels have a release date and they are not allowed to berth before the expected arrival time, the problem is NP-hard. On the other hand, if the arrival time does not impose a restriction on timing for mooring, the problem reduces to an assignment problem, solvable in polynomial time. Considering the former case, the following decomposition is suggested. The master problem is an assignment problem and contains only the binary variables l_{ij}^k. The master problem is solved to optimality and the solution is sent to the subproblem. If the subproblem is infeasible, a feasibility cut is generated. If the subproblem is feasible, an optimality check is performed based on the Fundamental Theorem of Duality, as in the traditional Benders Decomposition.

Master problem:

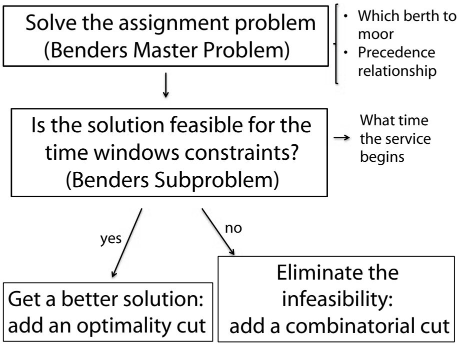
$$\min \sum_{i \in N} \sum_{k \in M} \left(p_i^k \sum_{j \in N \cup \{d\}} l_{ij}^k \right) \tag{36}$$

$$\text{s.t. } (24)\text{–}(27),\ (33) \tag{37}$$

Subproblem:

$$\min \sum_{i \in N} \sum_{k \in M} \left(x_i^k\right) \tag{38}$$

$$\text{s.t. (28)–(32), (35), (34)} \tag{39}$$



4.1 Master Problem Improvement

In order to try to reduce the number of infeasible subproblems, we incorporate in the master problem some information about the subproblem constraints. The following constraints are added to the master problem: if the processing time of all vessels being serviced at berth k is considered, it needs to respect the time window of the berth:

$$\max\left\{s^k, \min_{i \in N}\{a_i\}\right\} + \sum_{i \in N} \sum_{j \in N \cup \{d\}} p_i^k l_{ij}^k \leq \min\left\{e^k, \max_{i \in N}\{b_i\}\right\} \; \forall k \in M \tag{40}$$

4.2 Multi-cut Approach

When infeasibility origins on constraints (28), it means that the solution has a subtour. And this does not depend on the berth to cause infeasibility: if a given sequence of vessels is a subtour at berth k, the same sequence is a subtour at any other berth. Therefore, a subtour elimination constraint for all berths may be added:

$$\sum_{i,j \in C: l_{ij}^k = 1} \left(1 - l_{ij}^k\right) \geq 1 \; \forall k \in M \tag{41}$$

These constraints are the “The Subtour Formulation” proposed by [12] for the Traveling Salesman Problem. If the subproblem infeasibility is in constraints (28), the subtour elimination constraints (41) are incorporated in the master problem.

4.3 Subproblem Disaggregation

After the master problem is solved, the list of vessels allocated to each berth is known, being the only subproblem task to decide the time each vessel must be berthed. Therefore, it can be separated in $|M|$ disconnected subproblems, each subproblem generating an optimality cut for each berth $k \in M$.

5 Computational Results

The implemented Benders Decomposition algorithm follows the strategy presented in [19]. It uses an incumbent solution in the branch-and-bound search tree to be passed to the sub-problem for Benders cut generation and the master problem is solved to optimality only once. Callbacks are used to intervene in the branch-and-bound tree search process and add the benders cuts generated to the master problem as lazy constraints. This method has the advantage that, by using a single search tree, a node is never revisited and a truly superior solution is never overlooked.

All the procedures described in Sect. 4 were implemented in C++ on an Intel Xeon Core processor model E5-W2687 3.10 GHz with 128 GB RAM and IBM Ilog CPLEX 12.6. CPLEX and the decomposition approach were given an one hour CPU time limit. The set of instances I2 from [4] were used in the computational experiments, which are presented in Table 1. There are 5 problem sizes, corresponding to a given number of vessels and berths: 25 vessels and 5 berths (25 $\times$ 05), 25 vessels and 7 berths (25 $\times$ 07), 25 vessels and 10 berths (25 $\times$ 10), 35 vessels and 7 berths (35 $\times$ 07) and 35 vessels and 10 berths (35 $\times$ 10). Ten instances were tested for each problem size and the results regarding the monolithic solution of the model (two first columns) and the Benders decomposition application (three following columns) are presented. For the monolithic solutions the objective function value and the corresponding GAP are presented. For the Benders decomposition the values of the upper and lower bounds are provided (the solution is optimal if they are equal), together with the value of GAP. We recall that all these results concern solutions obtained within the one hour time limit.

Only in 5 instances, out of 50, Benders Decomposition was able to outperform the monolithic model solved by CPLEX. However, interesting enough, these are among the largest instances tested: 35 vessels and 7 berths and 35 vessels and 10 berths. It is noteworthy that for the BAP model (23)–(34), the Benders subproblem may present multiple optimal solutions, because the objective function is a linear combination of the constraints. It generates weak optimality cuts, and for this reason

Table 1 Comparison between CPLEX and Benders Decomposition

Problem size (vessels × berths)	Instance	Monolithic solution		Benders decomposition		
		GAP	Objective	Master (LB)	Sub (UB)	GAP
25 × 05	1	0.000498967	6559	6308	6666	0.053705371
	2	0.0288021	7882	7404	8610	0.140069686
	3	0.0331372	6914	6447	7327	0.120103726
	4	0.00947696	8843	8597	9376	0.083084471
	5	0.0120393	7598	7235	8234	0.121326208
	6	0.0342558	7444	6856	7908	0.133029843
	7	0.000959669	7751	7463	8534	0.125498008
	8	Optimal	7789	7601	8554	0.111409867
	9	Optimal	8556	8318	9239	0.099686113
	10	0.0335009	8579	8032	9055	0.112976256
25 × 07	1	optimal	10088	9932	10763	0.077208957
	2	Optimal	12086	11922	12635	0.05643055
	3	Optimal	9807	9572	10559	0.093474761
	4	0.00123054	9984	9799	11050	0.11321267
	5	Optimal	10763	10577	11683	0.094667466
	6	0.00808516	12434	12137	13378	0.09276424
	7	0.0157357	13218	12854	13972	0.080017177
	8	0.0299663	10478	10458	11618	0.099845068
	9	0.0215913	10884	9982	11189	0.107873805
	10	0.022655	12580	12072	13268	0.090141694
25 × 10	1	0.0113352	11998	11727	12418	0.055645031
	2	Optimal	11693	11526	12246	0.058794708
	3	0.00907693	13661	13391	14023	0.045068816
	4	0.0170101	16696	16223	17096	0.051064577
	5	0.00689249	11897	11623	12790	0.091243159
	6	0.00323535	14120	13941	14575	0.043499142
	7	0.000388922	14913	14785	15528	0.047849047
	8	0.00131053	14498	14289	14866	0.0388134
	9	0.00250406	14776	14599	15339	0.048243041
	10	0.00594059	15150	14896	15689	0.050544968
35 × 07	1	0.0662803	15012	13754	15298	0.100928226
	2	0.0823235	17577	15834	17760	0.108445946
	3	0.0969267	15651	13878	15932	0.128922922
	4	0.0582261	16247	15028	16950	0.11339233
	5	0.13841	18538	15692	17474	0.101980085
	6	0.17487	17277	13968	15725	0.111732909
	7	0.0658902	17706	16215	17956	0.096959234
	8	0.0744419	17067	15509	17642	0.120904659
	9	0.0675465	18039	16560	18526	0.106121127
	10	0.0359825	16700	15846	17868	0.113163197

(continued)

Table 1 (continued)

Problem size (vessels × berths)	Instance	Monolithic solution		Benders decomposition		
		GAP	Objective	Master (LB)	Sub (UB)	GAP
35 × 10	1	0.0576736	21308	19831	21496	0.077456271
	2	0.0790474	19332	17585	18874	0.068295009
	3	0.0198889	19810	19190	20915	0.082476691
	4	0.0645441	21843	20226	21890	0.076016446
	5	0.0408344	20108	19070	20365	0.063589492
	6	0.0919004	19717	17645	19570	0.098364844
	7	0.0554788	18106	16812	18894	0.110193712
	8	0.0347748	19957	18931	20651	0.083288945
	9	0.126682	15408	13144	14967	0.121801296
	10	0.0454113	19973	18779	20456	0.081980837

the decomposition progresses slowly. Moreover, the dual solutions are degenerate and the Pareto optimal cuts (see [9]: if the primal subproblem is degenerate it is possible to select the dual solution that is the closest to the interior of the master problem polyhedron to produce stronger cuts) can not be used to improve Benders Decomposition performance.

6 Conclusion

A model for the Berth Allocation Problem (BAP) as a Heterogeneous Vehicle Routing Problem with Time Windows was presented in this paper, and a Benders Decomposition approach was proposed for the BAP. Benders Decomposition is a cutting plane method where in each iteration new constraints (cuts) are added to the problem. In this paper, several cuts were proposed, implemented and tested.

The computational tests performed indicated that Benders Decomposition may be an interesting approach to solve the BAP. Although being competitive with monolithic model resolution with CPLEX, in general Benders Decomposition does not outperform CPLEX. However, the exception lies on some of the largest instances, indicating that for the most difficult instances Benders Decomposition may have a better performance.

Many real-world systems state change frequently due to unforeseen events. Most of the computational time running scheduling systems is spent in rescheduling, caused by changes in customer priorities, unexpected equipment maintenance, etc. This results in a requirement for frequent re-optimization. For example, when a crane in an automated container terminal malfunctions, a new equipment schedule for the entire port facility must be available within five to ten minutes, otherwise the handling of vessels will be delayed. When such unexpected problems come up, the terminal operator must be ready to change the service system; developing a tool that re-optimizes the system and quickly find a solution to the problem help improve the dynamics of the terminals and consequently their revenue. The results provided by Benders Decomposition in the Berth Allocation Problem open the possibility of incorporating this algorithm in a decision support system to re-optimize solutions whenever unforeseen events occur. Indeed, fixing variables and optimizing the others is inherent to Benders Decompositions algorithms.

Acknowledgements The authors acknowledge the Project "TEC4Growth—Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/NORTE-01-0145-FEDER-000020" is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

Appendix

Supplementary material: Computational results.

References

1. Buhrkal, K., Zuglian, S., Ropke, S., Larsen, J., Lusby, R.: Models for the discrete berth allocation problem: a computacional comparison. Transportation Research Part E: Logistics and Transportation Review **47**(4), 461–473 (2011)
2. Caramia, M., Mari, R.: A decomposition approach to solve a bilevel capacitated facility location problem with equity constraints. Optimization Letters **10**(5), 997–1019 (2016)
3. Codato, G., Fischetti, M.: Combinatorial benders' cuts for mixed integer linear programming. Operations Research **54**(4), 756–766 (2006)
4. Cordeau, J.F., Laporte, G., Legato, P., Moccia, L.: Models and tabu search heuristics for the berth allocation problem. Transportation Science **39**(4), 526–538 (2005)
5. Geoffrion, A.M.: Generalized benders decomposition. Journal of Optimization Theory and Applications **10**(4), 237–260 (1972)
6. Grubišić, N., Vilke, S., Barić, M.: A Contribution to Berth Allocation Problem Solution with Draft Restrictions. Pomorski zbornik **49**(1), 127–142 (2015)
7. Guan, Y., Cheung, R.K.: The berth allocation problem: models and solution methods. Or Spectrum **26**(1), 75–92 (2004)
8. Huang, Z., Zheng, Q.P.: Decomposition-based exact algorithms for risk-constrained traveling salesman problems with discrete random arc costs. Optimization Letters **9**(8), 1553–1568 (2015)
9. Magnanti, T.L., Wong, R.T.: Accelerating benders decomposition: Algorithmic enhancement and model selection criteria. Operations Research **29**(3), 464–484 (1981)
10. Mercier, A., Cordeau, J.F., Soumis, F.: A computational study of Benders decomposition for the integrated aircraft routing and crew scheduling problem. Computers & Operations Research **32**(6), 1451–1476 (2005)
11. Monaco, M.F., Sammarra, M.: The berth allocation problem: A strong formulation solved by a lagrangean approach. Transportation Science **41**(2), 265–280 (2007)

12. Pataki, G.: Teaching integer programming formulations using the traveling salesman problem. SIAM REV **45**, 116–123 (2003)
13. Pratap, S., Nayak, A., Cheikhrouhou, N., Tiwari, M.K.: Decision support system for discrete robust berth allocation. IFAC-PapersOnLine **48**(3), 875–880 (2015)
14. Rasmussen, R.V., Trick, M.A.: A Benders approach for the constrained minimum break problem. European Journal of Operational Research **177**(1), 198–213 (2007)
15. Rahmaniani, R., Crainic, T.G., Gendreau, M., Rei, W.: The Benders decomposition algorithm: A literature review. European Journal of Operational Research **259**(3), 801–817 (2017)
16. Tang, L., Jiang, W., Saharidis, G.K.: An improved benders decomposition algorithm for the logistics facility location problem with capacity expansions. Annals of Operations Research **210**(1), 165–190 (2013)
17. Taskın, Z.C.: *Benders decomposition.* Wiley Encyclopedia of Operations Research and Management Science. John Wiley & Sons, Malden (MA) (2010)
18. Theodorou, E., Diabat, A.: A joint quay crane assignment and scheduling problem: formulation, solution algorithm and computational results. Optimization Letters **9**(4), 799–817 (2015)
19. Vatsa, A.K., Jayaswal, S.: A new formulation and benders decomposition for the multi-period maximal covering facility location problem with server uncertainty. European Journal of Operational Research **251**(2), 404–418 (2016)

Capacitated Vehicle Routing Problem with Heterogeneous Fixed Proprietary Fleet and Outsourcing Delivery—A Clustering-Based Approach

Ricardo Bertoluci, António G. Ramos, Manuel Lopes and João Bastos

Abstract This paper describes a solution method that was created with the objective of obtaining a more efficient finished goods distribution process for a food industry company. The finished goods distribution process involves the use of the companys own fleet to serve a specific group of customers, and the use of outsourcing transportation services that can make direct and transshipment customer deliveries. The complexity of the problem is due to the need to decide which customers should be served by each of the outsourcing transportation services, direct or transshipment, and to find cost efficient solutions for the multiple vehicle routing problems created. First, an original clustering method consisting of a logical division of the customer orders using a delivery ratio based on the transportation unit cost, distance and order weight, is used to define customer clusters by service type. Then, an exact method based on a mixed integer programming model, is used to obtain optimal vehicle routing solutions, for each cluster created. The solution method for the company real instances, proved able to reach the initial proposed objectives and obtain promising results that suggest an average reduction of 34% for the operational costs, when compared to the current distribution model of the company.

Keywords Outsourcing · Heterogeneous fleet · Clustering · Vehicle routing problem

R. Bertoluci
School of Engineering, Polytechnic of Porto, Porto, Portugal
e-mail: 1120698@isep.ipp.pt

A. G. Ramos (✉)
INESC TEC and CIDEM, School of Engineering, Polytechnic of Porto, Porto, Portugal
e-mail: agr@isep.ipp.pt

M. Lopes
CIDEM, School of Engineering, Polytechnic of Porto, Porto, Portugal
e-mail: mpl@isep.ipp.pt

J. Bastos
INESC TEC and School of Engineering, Polytechnic of Porto, Porto, Portugal
e-mail: jab@isep.ipp.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_4

1 Introduction

This work is based on a distribution problem of a Portuguese company dedicated to the charcuterie business with production facilities located in the northwest of Portugal. The majority of the customers are located in Portugal and range from small shops to large food retailers.

The current distribution model is characterised by a private owned fleet that supplies a restricted group of customers, and an outsourced delivery service for the remaining clients. The customers with deliveries from the private fleet are restricted in number and geographical location. This constraint is related with the size of the private fleet, since it is composed by three vehicles, with the proximity of the customers to the production facility and the average size of the customers orders. This last condition is also related with the need for a more personalised service required by these customers, since they usually do not have unloading infrastructures. The routing of the private fleet does not follow any specific method, and is defined by the *common sense* of company drivers.

The customer orders served by the outsourcing delivery service are currently sent through the depot of the outsourcing company that act as a transshipment platform. The outsourcing company has a cost structure based on a unit weight cost for each handling of the goods, i.e, there is a cost when the goods are picked up at the factory and an additional one when they are picked up on the outsourced depot to be delivered to the customers. This presents an inefficiency since a large number of customers is located between the factory and the outsourcing depot. Currently the company has no control over the outsourcing distribution process.

The goal of the company is to create a new distribution model where the routing and scheduling of vehicles is not based on the intuition of the company drivers or the willingness of the outsourcing company, but where the company decides if a customer is served directly from the factory or through the outsourced depot. Therefore, the model should be able to determine all necessary routing and scheduling for the private fleet, which orders and routes for direct outsource delivery, and which orders delivered through the outsourcing depot.

As a key process in the distribution model, the routing and scheduling of vehicles plays an important role in its success. This problem, known in literature as the vehicle routing problem (VRP), since it was introduced by [2], has been thoroughly studied in literature, as have its numerous variants that are a result of the incorporation of a growing number of real world constraints to the more elementary VRP [1, 3, 6].

According to [4], the VRP problem consists in designing the routes of vehicles with known capacity, that must deliver to a set of customers the required order, from a single depot so as to minimise a given objective such as total distance travelled or route costs. The routes are designed in such a way that each location is visited only once by one vehicle and the capacity of each vehicle is not exceeded.

The variants of VRP include multiple depots, multiple trips, multiple vehicle types, loading constraints, time constraints or drivers constraints, just to mention a few [3]. The problem addressed in this work has several of these additional constraints, namely: a heterogeneous fleet; fixed dimensions; vehicle capacity; vehicle weight limits; limited vehicle distance range and ownership of the fleet. The problem is referred as the Capacitated Vehicle Routing Problem (CVRP) with Heterogeneous Fixed proprietary Fleet and Outsourcing Delivery. In this work, we propose a clustering based algorithm, that begins by clustering the customers and then determines the routes and scheduling of vehicles for each cluster using a mixed integer linear programming model (MILP).

In Sect. 2 the proposed algorithm is presented, followed in Sect. 3 by the presentation and discussion of the computational results obtained for the problem instances, using the solution approach proposed. In Sect. 4 the final conclusions are presented.

2 The Vehicle Routing Problem Clustering-Based Algorithm

We begin this section with an overview of the proposed solution process. This is followed by the presentation of the clustering approach. This section ends with the definition of the mixed integer linear programming model used.

2.1 Overview

The proposed algorithm for the clustering-based approach to the CVRP with heterogeneous fixed proprietary fleet and outsourcing services, is composed by two clustering procedures and a MILP model.

The algorithm starts by building a cluster composed by all the customers served by the proprietary fleet. The MILP model is then run for this cluster. Next, a clustering procedure is performed with the remaining customers. For each of the obtained clusters the MILP is then run. The operational costs of the obtained routes is then compared against the operational costs of using the outsourced depot. With the garnered results it can be established for each customer's order the corresponding delivery mode, i.e., by proprietary fleet, direct outsourcing delivery or through the outsourcing depot.

The pseudo code of the algorithm is described in Algorithm 1.

Algorithm 1: Vehicle routing problem clustering based algorithm

```
Input: Let P be a set of customers to be served by the privately owned fleet
Let O be a set of customers to be served by the outsourced fleet
Output: Routes with delivery mode;
Clustering (Set of Customers)
    groups customers in clusters;
    return Set C of clusters c_i (i = 1, .., m);
MILP (c_i)
    determines the routes for a given Cluster;
    return Set of routes;
begin
    MILP(P);
    C ← Clustering (O);
    for each of the clusters in C do
        R ← MILP(c_i);
        for each of the generated routes in R do
            if direct_shipping_cost ≤ 3PL_depot_cost then
            Route_with_direct_shipping;
            else Route_with_3PL_depot_shipping;
        end
    end
end
```

2.2 *Customer Clustering for Outsourced Delivery*

The main goal of clustering the customers is to transform the decision for outsourcing services, i.e., determine which deliveries must be made directly to the customer and those that must be made via the outsourcing depot, into several VRP problems that translate the structure of the business.

The clustering approach for customers with orders delivered by outsourcing should evidence the geographical regions where the deliveries should be made directly to the customer and those where they should be done via a third party warehouse.

Considering that the cost of direct transport has a fixed cost per used vehicle and a variable cost charged by distance unit, i.e., depends on the distance travelled to make the delivery, it is expected that for large orders the direct transport is more advantageous relative to the cost via a third party warehouse, which is charged by weight unit.

Since the objective will always be to minimise operational costs, a comparison of the unit cost of the two types of service is performed. This unit cost ratio UCR is required to evaluate under which conditions the use of direct transport or third party warehouse transport is more advantageous.

The evaluation is done using the delivery ratio dw_{ij} determined by (1) where the index i represents an origin node and the index j represents a destination node of an arc, d_{ij} the distance between the nodes and w_j the weight of the order demand at j.

$$dw_{ij} = \frac{d_{ij}}{w_j} \tag{1}$$

Additionally it was determined that customers closely located, i.e., within maximum distance range Max^{Range} should be in the same cluster regardless of the order weight.

The clustering procedure begins by determining city groups, i.e., grouping customers that are located in the same city. Then, determines for each city group the sum of the orders of the grouped customers. Next, all pairwise delivery ratios including city groups and origin-depot are calculated.

The clusters are then initialised according to the following criteria:

- if there are delivery ratios starting at the origin-depot d_{0j} less than equal to the UCR, initialise the cluster with the customer j that has the lower ratio.
- in case the previous condition is not met, if the distance between two city groups is less or equal to Max^{Range} initialise the cluster with the two closer city groups.
- otherwise, add all city groups to a cluster.

Considering the city groups of the cluster as the origin, assign to the cluster all the city groups with a delivery ratio less than equal to the UCR and distance less or equal to Max^{Range}, and then repeat it for all the newly assigned city groups until no more city groups can be added. If there are city groups not assigned to a cluster, a new one should be initiated.

The clustering procedure is illustrated in Fig. 1. Consider a set of points representing locations where 0 represents the origin-depot and numbers 1 to 9 city groups. The city group chosen to initialise the cluster is city group 1, the only one with a delivery ratio smaller than UCR from the origin-depot.

In Fig. 1a, it is considered that city group 1 has city groups 2 and 3 within range, which are added to the cluster. Since from 2 there is no city group in range, no city

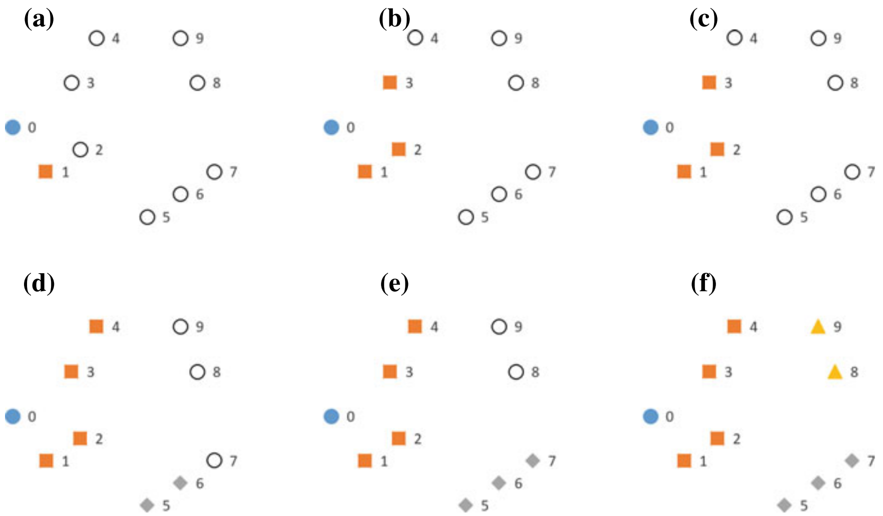


Fig. 1 Example of the city clustering

groups are added. With origin in 3, city group 4 is in range, being added to the cluster. Since from 4 no other city group meets the requirements, the cluster is closed. In Fig. 1d, city groups 5 and 6 initialise a new cluster to which 7 is added. Since no other city group can be added to the cluster, it is closed. The remaining city groups, 8 and 9 considered to have a $d_{89} \geq UCR$, are grouped in a new cluster.

2.3 Mixed Integer Linear Programming Model

Given the previous algorithm description, the MILP model addresses a VRP problem as follows.

Let $G = (N, A)$ be a directed graph where $N = \{0, 1, \ldots, n + 1\}$ is a set of nodes, and $A = \{(i, j) : i, j \in N, i \neq j\}$ is the set of direct arcs. The node 0 denotes the origin-depot, i.e., the start of the route, node $n + 1$ denotes the destination-depot, i.e, the end of the route and the set $C = N \cap \{0, n + 1\}$ is composed by the remaining nodes that represent the n customers. For each arc $(i, j) \in A$, d_{ij} is the distance between the nodes (i, j). Each customer i has a non-negative known demand q_i with a given weight w_i. The fleet is composed by several vehicles, with $V = \{1, ..., K\}$. Each $k \in V$ has a capacity Q_k, a weight limit W_k, a fixed cost per use F_k, and a travel cost per distance unit denoted by c_k. Let $Dist^{Max}$ be the maximum distance allowed for a route assigned to a vehicle of the privately owned fleet. Let M be a large number. The objective is to determine the set of routes that minimise the sum of travel costs such that every route starts and ends at a depot and is associated to a vehicle, each customer belongs to exactly one route and the sum of all the demands of any route does not exceed the capacity of the assigned vehicle and its weight limit.

The following decision variables are used in the formulation:

$x_{ij}^k = \begin{cases} 1, \text{if vehicle } k \in V \text{ traverses arc } (i, j) \in A \\ 0, \text{ otherwise} \end{cases}$

l_i^k = distance travelled by vehicle k to visit location i

$$\min \sum_{k \in V} \sum_{i \in C \cup \{0\}} \sum_{j \in C \cup \{n+1\} | i \neq j} c_k d_{ij} x_{ij}^k + \sum_{k \in V} \sum_{j \in C} F_k x_{0j}^k \tag{2}$$

$$\text{subject to:} \sum_{k \in V} \sum_{i \in C \cup \{0\} | i \neq j} x_{ij}^k = 1, \qquad \forall j \in C \tag{3}$$

$$M \sum_{j \in C} x_{0j}^k \geq \sum_{i \in C \cup \{0\}} \sum_{j \in C \cup \{n+1\} | i \neq j} x_{ij}^k, \qquad \forall k \in V \tag{4}$$

$$\sum_{j \in C} x_{0j}^k \leq 1, \qquad \forall k \in V \tag{5}$$

$$\sum_{i \in C \cup \{0\} | i \neq j} x_{ij}^k = \sum_{i \in C \cup \{n+1\} | i \neq j} x_{ji}^k, \qquad \forall j \in C, \forall k \in V \tag{6}$$

$$\sum_{i \in C \cup \{0\}} \sum_{j \in C | i \neq j} q_j x_{ij}^k \leq Q_k, \qquad k \in V \tag{7}$$

$$\sum_{i \in C \cup \{0\}} \sum_{j \in C | i \neq j} w_j x_{ij}^k \leq W_k, \qquad \forall k \in V \tag{8}$$

$$\sum_{i \in C \cup \{0\}} \sum_{j \in C \cup \{n+1\} | i \neq j} d_{ij} x_{ij}^k \leq Dist^{Max}, \qquad \forall k \in V \tag{9}$$

$$(l_i^k + d_{ij}) - M(1 - x_{ij}^k) \leq l_j^k, \qquad \forall (i, j) \in A, \forall k \in V \tag{10}$$

$$l_i^k \geq 0 \qquad \forall i \in N, \forall k \in V \tag{11}$$

$$x_{ij}^k \in \{0, 1\}, \qquad \forall (i, j) \in A, \forall k \in V \tag{12}$$

The objective function (2) intends to minimise the total operational cost, i.e., the sum of the travel costs and the fixed costs of requested direct delivery routes. Constraints (3) ensure that all customers present on the model must be visited exactly once by one vehicle. Constraints (4) ensures that each time a vehicle is to supply a customer, it must exit from the origin-depot, i.e, the production plant. Constraints (5) ensures that a vehicle can not leave the origin-depot towards more than one customer. Constraints (6) referred to as the conservation of flow constraints, ensure the continuity of a vehicle's flow, thus ensuring that when a vehicle reaches a certain node j, it has to leave towards another node. Constraints (7) and (8) guarantee the feasibility of the loads. Constraints (9) ensure that the maximum distance to be travelled by each vehicle on a route shall not exceed $Dist^{Max}$. However, these constraints only apply to the private fleet since for the outsourced fleet the distance limit is unnecessary.

Constraints (10) are sub-tour elimination constraints modelled with an extension of the Miller-Tucker-Zemlin (MTZ) inequalities for the TSP [5]. Usually the MTZ constraints are applied to VRP problems with time windows delivery constraints using the service and travel time instead of using the distance, as in this case. Therefore, this constraint basically translates a distance counter to reach customer j. Constraints (11) and (12) define the domain of the decision variables.

In spite of the different approaches for the private fleet and the outsourced fleet, the MILP used for determining the routing solutions will be the same for the two available fleets. As previously mentioned, the outsourced fleet constraints (9) will not be included.

3 Computational Results

This section presents the results of the computational experiments run to evaluate the performance of the proposed algorithm.

The clustering algorithm was implemented using Microsoft Excel 2016 and VBA and the MILP model was implemented in IBM ILOG Cplex 16.3 solver and run on an computer with a quadcore Intel(R) Core(TM) i5-6200U CPU @ 2.30 GHz with 6 GigaBytes of RAM running the Windows 10 Home 64 bit operating system.

3.1 Test Instances

The problem tests used to evaluate the effectiveness of the algorithm were obtained from the company, from 10 operational days during April 2017. A summary of the instances is presented in Table 1. The first column indicates the number of the instance, the second and third columns indicate the number of customers and the total weight of the orders to be delivered by the private owned fleet. Columns 4, 5 and 6 indicate the number of customers with orders to be delivered by the outsourced fleet, the number of pallets in the orders and the total weight of the orders.

Table 2 presents for each of the available vehicles the ownership, the capacity, the weight limit, the unit distance cost and the fixed cost per direct delivery route.

Table 3 indicates the parameters used in the MILP model and the city clustering procedure. These values were provided by the company.

Table 1 Instances summary

Instance	Proprietary customers		Outsourcing customers		
	n	Weight [kg]	n	#Pallets	Weight [kg]
1	15	5504.8	25	65	22914.9
2	15	5868.6	29	49	12204.4
3	21	4623.3	32	106	35377.9
4	12	1773.0	31	88	29118.1
5	10	3528.5	33	94	32021.9
6	20	4550.0	28	71	28338.8
7	8	3925.7	25	75	27855.3
8	11	2945.2	38	112	41426.5
9	14	4332.5	25	57	19159.4
10	18	6080.0	24	54	21547.7

Table 2 Vehicle data

Vehicle	Ownership	Capacity [# pallets]	Weight limit [kg]	Unit travel cost [€/Km]	Fixed cost [€/route]
1	Private	16	6500	1.24	–
2	Private	7	1000	1.14	–
3	Private	7	1000	1.14	–
4	Outsourcing	33	22000	0.90	150

Table 3 Instances parameters used in the MILP and clustering

Parameters	Values
Unit cost [€/kg]	0.09
UCR	0.1%
$Dist^{Max}$ [km]	225
Max^{Range} [km]	100

3.2 *Performance of the Algorithm*

In Table 4 the computational results are presented for the private fleet delivery. The table is divided in to two sections. The first section (Current solution) presents the current solution of the company and the second section (MILP solution), presents the results for the MILP model run for the customers with deliveries via private fleet. In both sections, column (n) represents the number of customers, column (Weight [kg]) the weight of the orders delivered, column (k) the number of vehicles used, column ([€]) the transport costs and column ([€/kg]) the resultant unit weight cost. Additionally, in the MILP solution section, column (Time [s]) represents the processing time in seconds. As it can be observed in the results, there is an increase in the number of customers served and cargo weight transported. Nevertheless, when regarding cost variation, both in absolute and percentage value, there is a decrease in these costs. It was therefore possible to deliver to more customers, transport more weight and reduce the operational cost.

The computational results for the outsourced deliveries are presented in Tables 5 and 6. Table 5 presents the results for the MILP model run without clustering, the

Table 4 Private fleet delivery

Inst.	Current solution					MILP solution					
	n	Weight		Cost		n	Weight		Cost		Time [s]
		[kg]	k	[€]	[€/kg]		[kg]	k	[€]	[€/kg]	
1	13	5077	2	559	0.110	15	5505	2	496	0.090	9.13
2	10	5584	3	394	0.071	15	5869	2	309	0.053	493.03
3	17	3689	2	330	0.089	21	4623	1	268	0.058	3.83
4	8	721	1	216	0.299	12	1773	2	343	0.193	19.98
5	9	3289	2	521	0.158	10	3528	2	452	0.128	0.34
6	16	3351	2	344	0.103	20	4550	2	434	0.095	766.39
7	7	3153	2	391	0.124	8	3926	2	386	0.101	0.33
8	9	2249	2	460	0.204	11	2945	2	473	0.161	2.08
9	11	3521	2	430	0.122	14	4332	2	502	0.116	4.72
10	16	5622	2	592	0.105	18	6080	2	516	0.085	19.72
Total	116	36255		4236	0.117	144	43132		4178	0.097	

Table 5 Outsourcing solutions variants

	MILP							Algorithm 1 solution (without clustering)									Algorithm 1 solution								
			Best integer solution	Best bound	Gap	Time	Total cost	Direct delivery			Depot delivery					Direct delivery			Depot delivery				
									Weight	Total		Weight	Total	Total cost			Weight	Total		Weight	Total	Total cost	
Inst.	*n*	[kg]	[€]	[€]	[%]	[s]	[€/kg]	*n*	[kg]	[€]	*n*	[kg]	[€]	[€]	[€/kg]	*n*	[kg]	[€]	*n*	[kg]	[€]	[€]	[€/kg]
1	25	22915	1578	1464	7.2	3600	0.069	25	22915	1578	0	0	0	1578	0.069	14	14849	775	11	8066	726	1501	0.066
2	29	12204	1447	1447	0.0	82	0.119	5	5484	417	24	6720	605	1021	0.084	2	4660	185	27	7544	679	864	0.071
3	32	35378	2136	1782	16.6	3600	0.060	32	35378	2136	0	0	0	2136	0.060	22	31418	1658	10	3960	356	2015	0.057
4	31	29117	1903	1515	20.4	3600	0.065	13	20703	1035	18	8414	757	1792	0.062	16	25725	1244	15	3393	305	1550	0.053
5	33	32022	2066	1579	23.5	3600	0.065	33	32022	2066	0	0	0	2066	0.065	18	24128	1231	15	7894	710	1941	0.061
6	28	28339	1609	1441	10.4	3600	0.057	10	20165	667	18	8174	736	1403	0.049	18	25154	835	10	3185	287	1122	0.040
7	25	27855	1672	1556	6.9	3600	0.060	25	27855	1672	0	0	0	1672	0.060	12	19164	769	13	8691	782	1551	0.056
8	38	41427	2589	1809	30.1	3600	0.062	38	41427	2589	0	0	0	2589	0.062	20	35177	1643	18	6250	562	2205	0.053
9	25	19160	1372	1266	7.7	3600	0.072	8	9315	485	17	9845	886	1371	0.072	15	13529	803	10	5631	507	1310	0.068
10	24	21548	1203	1203	0.0	25	0.056	4	13453	226	20	8095	729	955	0.044	4	13452	226	20	8096	729	955	0.044
Total	290	269965	17573	15061	123	28907	0.065	193	228717	12871	97	41248	3712	16583	0.061	141	207257	9370	149	62708	5644	15013	0.056

Table 6 Outsourced delivery

Inst.	Current solution				Algorithm 1 solution							
					Direct delivery			Depot delivery				
		Weight	Cost			Weight	Total		Weight	Total	Total cost	
	n	[kg]	[€]	[€/kg]	*n*	[kg]	[€]	*n*	[kg]	[€]	[€]	[€/kg]
1	27	23342	2101	0.09	14	14849	775	11	8066	726	1501	0.066
2	34	12489	1124	0.09	2	4660	185	27	7544	679	864	0.071
3	37	36313	3268	0.09	22	31418	1658	10	3960	356	2015	0.057
4	35	30170	2715	0.09	16	25725	1244	15	3393	305	1550	0.053
5	34	32262	2904	0.09	18	24128	1231	15	7894	710	1941	0.061
6	32	29538	2658	0.09	18	25154	835	10	3185	287	1122	0.040
7	26	28628	2577	0.09	12	19164	769	13	8691	782	1551	0.056
8	40	42123	3791	0.09	20	35177	1643	18	6250	562	2205	0.053
9	28	19971	1797	0.09	15	13529	803	10	5631	507	1310	0.068
10	26	22006	1981	0.09	4	13452	226	20	8096	729	955	0.044
Total	319	276841	24916	0.09	141	207257	9370	149	62708	5644	15013	0.056

Table 7 Overall costs comparisons results

Inst.	Current solution			Algorithm 1 solution			Var [%]
	Private	Outsourced	Total	Private	Outsourced	Total	
1	559	2101	2660	496	1501	1997	−25
2	394	1124	1518	309	864	1173	−23
3	330	3268	3598	268	2015	2282	−37
4	216	2715	2931	343	1550	1893	−35
5	521	2904	3425	452	1941	2393	−30
6	344	2658	3002	434	1122	1555	−48
7	391	2577	2968	386	1551	1937	−35
8	460	3791	4251	473	2205	2678	−37
9	430	1797	2227	502	1310	1812	−19
10	592	1981	2573	516	955	1471	−43
Total	4236	24916	29152	4178	15013	19191	−34

results for Algorithm 1 run without clustering, i.e., the routes of the MILP model solution are classified according to the route cost, as direct or depot delivery, and the results for Algorithm 1. Column (n) represents the number of customers, column (Weight [kg]) the weight of the orders delivered, column ([€]) the transport costs and column ([€/kg]) the applied unit weight cost. In section MILP column (Best Integer) represent the best integer solution determined by the CPLEX solver and column (Best Bound) the lower bound determined by the solver. The percentage difference to the best bound is presented in column (Gap) and the processing time in column (Time). Sections Algorithm 1 solution (without clustering) and Algorithm 1 solution are divided in three subsections, the Direct delivery which presents the results of the orders with Direct delivery, the Depot delivery which presents the results of the deliveries that go through the outsourced depot, and section Total Cost presents the total operational cost of the outsourced deliveries. As it can be observed, the clustering reduces the overall cost by 9.5% when compared to the solution without clustering and a 14,6% when compared with the MILP solution. If compared to the lower bound determined by CPLEX the reduction is 0.3%. These results confirm the efficiency of the proposed clustering approach.

Table 6 is divided in two major sections. The first section (Current solution) presents the current solution of the company and the second section (Algorithm 1 solution), presents the results for the Algorithm 1. Since the company currently uses only the outsource delivery service through the depot, and this service has a cost of 0.09 €/kg, the total cost of transport is the total weight to be transported multiplied by the unit weight cost. It must be noted that the customers that were not able to have their orders delivered by the private fleet, had their orders delivered by the outsourcing service. This means that a comparison based on the total cost ([€]) would be biased. From Table 6 it can be observed that the Total unit weight cost determined in each of the instances decreased when compared to the current solution. On average this decrease was 38%.

Finally, Table 7 allows the comparison of the total transport costs of the current solution of the company, with the solution from the proposed Algorithm. The two sections, Current solution and Algorithm 1 solution present the total cost using the private fleet (Private) the Outsourced services (Outsourced) and the sum of the two (Total). The last column presents the overall variation in percentage between the solutions. Since the overall results were obtained for the same delivered orders, it is possible to make a direct comparison between the current and the new method on a cost base. Therefore, for these test instances the new algorithm achieved an overall reduction of 34% for the total costs, ranging from 19% for instance 9–48% for instance 6.

4 Conclusions

In this paper we proposed a cluster-first routing-second VRP algorithm for a case study of a production company. The developed approach offers very promising results regarding the current method used in the company. The good performance of the proposed algorithm was demonstrated by the results obtained using a set of problem instances generated using real data from the company. The results obtained also highlight the benefits of the cluster-first routing-second approach. In future works we aim to extend the solution method using a matheuristic approach.

Acknowledgements We acknowledge the financial support of CIDEM, R&D unit funded by the FCT—Portuguese Foundation for the Development of Science and Technology, Ministry of Science, Technology and Higher Education, under the Project "UID/EMS/0615/2016". This research was also financed by the ERDF—European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation—COMPETE 2020 Programme within project "POCI-01-0145-FEDER-006961", and by National Funds through the Portuguese funding agency, FCT—Fundação para a Ciência e a Tecnologia as part of project "UID/EEA/50014/2013".

References

1. Braekers, K., Ramaekers, K., Van Nieuwenhuyse, I.: The vehicle routing problem: state of the art classification and review. Comput. Ind. Eng. **99**, 300–313 (2016)
2. Dantzig, G.B., Ramser, J.H.: The truck dispatching problem. Manag. Sci. **6**(1), 80–91 (1959)
3. Lahyani, R., Khemakhem, M., Semet, F.: Rich vehicle routing problems: from a taxonomy to a definition. Eur. J. Oper. Res. **241**(1), 1–14 (2015)
4. Laporte, G., Gendreau, M., Potvin, J.-Y., Semet, F.: Classical and modern heuristics for the vehicle routing problem. Int. Trans. Oper. Res. **7**(4–5), 285–300 (2000)
5. Miller, C.E., Tucker, A.W., Zemlin, R.A.: Integer programming formulation of traveling salesman problems. J. ACM (JACM) **7**(4), 326–329 (1960)
6. Montoya-Torres, J.R., Franco, J.L., Isaza, S.N., Jiménez, H.F., Herazo-Padilla, N.: A literature review on the vehicle routing problem with multiple depots. Comput. Ind. Eng. **79**, 115–129 (2015)

Modeling Supply Chain Network: A Need to Incorporate Financial Considerations

Alexandra Borges, Dalila B. M. M. Fontes and José Fernando Gonçalves

Abstract In the past few years, important supply chain decisions have captured managerial interest. One of these decisions is the design of the supply chain network incorporating financial considerations, based on the idea that establishment and operating costs have a direct effect on the company's financial performance. However, works on supply chain network design (SCND) incorporating financial decisions are scarce. In this work, we address a SCND problem in which operational and investment decisions are made in order to maximize the company value, measured by the Economic Value Added, while respecting the usual operational constraints, as well as financial ratios and constraints. This work extends current research by considering debt repayments and new capital entries as decision variables, improving on the calculation of some financial values, as well as introducing infrastructure dynamics; which together lead to greater value creation.

Keywords Supply chain network design problem (SCND) · Financial considerations · Value maximization

1 Introduction

One of the most significant changes in the paradigm of modern business management is the entering in a new era in which firm performance and competitive advantage are linked to supply chain (SC) performance. Traditionally, SC and logistics studies have

A. Borges
Faculdade de Economia da Universidade do Porto and Instituto Politécnico de Viana do Castelo, Avenida do Atlântico, n644, 4900-348 Viana do Castelo, Portugal
e-mail: aborges@estg.ipvc.pt

D. B. M. M. Fontes (✉) · J. F. Gonçalves
Faculdade de Economia da Universidade do Porto and LIAAD/INESC TEC, Rua Dr. Roberto Frias, 4200-464 Porto, Portugal
e-mail: fontes@fep.up.pt

J. F. Gonçalves
e-mail: jfgoncal@fep.up.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_5

focused on cost minimization; however, more recently, such studies have widened their scope to ensure alignment with the company strategy enhancing profitability and thus, company value [10].

SCND determines the supply infrastructure as well as the key operations (production, distribution, and storage) to be performed at each location in the resulting network in order to respond to customer orders in an economically efficient manner [16]. Henceforth, SCND involves decision making at both the strategic and tactical levels. The former addresses decisions regarding the number, locations, and sizes of the facilities (plants, warehouses, and distribution centers); while the latter addresses the ones regarding production (product mix and produced quantities at each plant), inventory (which products and quantities are held at each facility), and distribution (product flows). These decisions are interrelated since tactical decisions are limited by the network made available by strategic-level decisions and therefore, should be addressed simultaneously; nevertheless, most approaches deal with these two decision levels separately.

In this work we propose a deterministic multiple-period, multi-echelon, and multi-product mixed integer linear programming model (MILP) that simultaneously addresses the strategic and tactical decisions. Moreover, our model is adaptive in nature and is capable of dealing with market dynamics, up to a certain extent, since it allows to change the infrastructure during the planning horizon, rather than just setting it at the beginning.

Many authors, see, e.g., [6, 8, 13, 14, 16], consider financial factors to have a strong impact on the configuration of SCs and point out the importance of transfer pricing, corporate income taxes, currency exchange rates, tariffs, duties, among others. Furthermore, sustainability and growth of the SC rely heavily on financial operations, since they ensure the feasibility of production and distribution operations. Thus, the MILP model proposed maximizes the company value rather than minimizing cost. The company value is measured by the EVA (Economic Value Added), which deducts the cost of all capital invested during the planning horizon to the net operating profits after taxes generated during the same period.

The remaining sections of this paper are organized as follows. Section 2 provides a brief literature review on SC models with financial considerations. Section 3 describes the problem and provides the MILP model developed, as well as the notation used. Section 4 uses the case study from [11] and reports and compares the results obtained with those reported in [11]. Finally, Sect. 5 draws some conclusions and points out future research directions.

2 Supply Chain Models With Financial Issues: A Review

As previously mentioned, several authors advocate the use of financial analysis in order to make decisions regarding the SCND, since they, typically, involve large investments. For example, Shapiro [19] suggests that strategic planning should also include the analysis of corporate financial decisions regarding the acquisition of capital needed to expand an existing SC or to acquire companies with complementary product lines. However, optimization models considering both corporate financial planning and SC planning are scarce [14, 18].

As it can be seen in the focused review of multi-stage SC modeling [1], most works consider just the tactical decisions, i.e., production, inventories, and flows, while minimizing operational costs or maximizing revenue. The authors identified as research needs the discussion of the adequacy or appropriateness of existing SC performance measures and the inclusion of other decision variables that although highly impact the performance of the SC are, usually, ignored. A more recent review is that of [14], in which the authors identify characteristics that a facility location model should have to adequately address SC planning needs, as well as several research directions that still require intensive research. In addition, the authors also noticed that most research is cost-oriented and almost none is dedicated to financial objectives. Furthermore, Presutti and Mawhinney [4] stated that 70% or more of manufacturing companies expenditures are related to SC activities, which highlights the potential impact of an effectively managed SC in contributing to the overall improvement of the financial performance.

More recently, works contemplating financial issues have been reported in the literature; however most researchers consider the financial issues only as financial factors (e.g., return on investment, taxes, exchange rates, transfer prices, etc.) to be evaluated or at most restricted; but very few consider them as decision variables (in many cases even when optimizing some sort of financial measure, see e.g., [7]). Most of the works considering financial decisions have done so in the context of specific chemical SCs and belongs to a group of authors with common work (see, e.g., [9, 22] and the references therein), many extending previous works. For example, [9] maximizes the corporate value, extending the work in [5] that extended the work in [2] to multi-product, multi-echelon SCs, which maximized the net revenues obtained from cash transactions over the entire planning horizon; [21] minimizes the opportunity costs of annualized capital investment and cash/material inventory minus the benefit to stockholders by managing the cash flows associated with production activities (raw materials, manufacturing, inventory, and sales), which was then extended in [22] by considering multinational SC activities and investigating the influence of exchange rates and taxes on operational decisions; [17] models the purchasing quantity with respect to capital constraints and payment delays and studies the impact of capital on the operational costs, the return on capital investment, and the financial costs and [15] determines how much to order and when to pay the suppliers by considering maximum cash and loan amounts; [11] determines a SC network under uncertainty by maximizing shareholder value and in [12] the authors incorporate sale and lease-back; and [18] considers the design of a closed-loop SC for a single product such that the change in equity is maximized while satisfying a set of budgetary constraints (balances of cash, debt, securities, payment delays) and also making decisions on current and fixed assets and liabilities.

In here and following on the work in [11], we propose a deterministic MILP model to determine which facilities (warehouses and distribution centers) to open and when, as well as the quantities to be produced in each plant, to be stored in each facility (plants, warehouses, and distribution centers), to be transported between facilities and between distribution centers and customers in order to satisfy customer demands; while maximizing the EVA. Thus, we also consider decisions on bank loans, bank repayments, and new capital entries; however these must respect the long term debt to equity ratio, which provides an indication on how much debt a company is using to finance its assets, and cash coverage ratio that measures the firms capacity to meet interest payments. Our model extends the aforementioned work by (i) using an improved version of the financial considerations, since we consider the net debt in the financial statement analysis, which balances loans and repayments; (ii) improving some financial calculations, namely: depreciation, by keeping track of assets age, and real cash value, rather than assuming it as a fixed percentage of profit; (iii) providing the firm with the means to create an accounts payable policy, rather than assuming that all payments are made in cash; (iv) establishing new plants; and (v) allowing for a dynamic infrastructure, since facilities can be established at every time period over the planning horizon, rather than just at its beginning. Nevertheless, our model is deterministic while that of [11] considers uncertainty through scenarios analysis. However, note that to handle the scenarios, the authors create copies of the production variables, inventory variables, and flow variables, which are then used in independent copies of the constraints using such variables and the objective function is the weighted average of the net sales calculated using the aforementioned variables. Moreover, the possibility that different scenarios may lead to different network structures is not considered. Regarding the case study used, all values are constant across scenarios, except for demand.

3 Problem Definition and Formulation

The purpose of our supply chain network design problem (SCND) is to determine the entire manufacturing and distribution network for a company. In particular, our model considers the design of a multi-product, multi-echelon supply chain network, in a dynamic environment, allowing for decision making at every period of the planning horizon. Our problem integrates operational and corporate financial decisions and constraints, thus the mathematical formulation requires variables and parameters of both topics. After providing a detailed explanation of the operational and the financial aspects involved, we introduce the notation used and then the mathematical programming model developed.

Operational considerations:
Customers are divided into zones and the demand of each customer zone, in each period, is known. To satisfy customer demand, the firm can decide on which facilities to open in each time period. There are three types of facilities, namely: plants, warehouses, and distribution centers. Once a plant is established it can produce any product. Production quantities, in each period, are subject to maximum (capacity) and minimum (economic viability) production limits, as well as to resources

availability. Products may remain at the plant, as inventory, or be dispatched to one or several warehouses, if available. Inventory levels at the plants have upper limits for each product in each time period. Warehouses receive the products from one or more plants and then distribute them among the existing distribution centers, which in turn supply customers. Note that no back-order is allowed. Distribution centers receive the product or products from one or more warehouses and may supply one or more customers. Total inventory level, in each period, at both warehouses and distribution centers have upper and lower limits. In addition, safety stocks are imposed in each period for each product at both warehouses and distribution centers. Product transportation between the different types of facilities also has capacity limits. Finally, any decision regarding opening a facility is taken at the beginning of the period and it becomes available immediately.

Financial considerations:
The firm's operation leads to costs and revenue. The revenue is obtained from selling the products at known prices per product, customer zone and time period. Both fixed and variable costs are incurred with production and transportation. In addition, variable costs are incurred with inventory held at each facility (plants, warehouses, and distribution centers). Moreover, there are also fixed costs with establishing facilities. Finally, financial conditions are imposed/defined regarding cost of capital, long term borrowing interests, tax rate, depreciation rates, and financial ratios.

Therefore, in order to solve the firm's problem, one has to determine in each time period which facilities to open, the production amounts for each product in each plant, the material flows between facilities, and the inventory levels at each facility, as well as bank loans and repayments and new capital entries in order to satisfy customer demand while maximizing the company value, using Economic Value Added [20].

Notation

Sets and Indices:

- E: set of production resources, indexed by e;
- I: set of products, indexed by i;
- J_l: set of locations for facility type l (1-plant, 2-warehouse, and 3-distribution center, 4-customer), indexed by j, k, and m;
- $\mathscr{T}$: Set of T planning periods comprising the planning horizon, indexed by s and t;

Parameters:

- O_{ijt}: demand of product $i \in I$ by customer zone $j \in J_4$ in period $t \in \mathscr{T}$;
- R_{je}: availability of resource $e \in E$ in plant $j \in J_1$;
- ρ_{ije}: marginal needs by product $i \in I$ of resource $e \in E$ at plant $j \in J_1$;
- P_{ij}^{max}: maximum production capacity of product $i \in I$ in plant $j \in J_1$;
- P_{ij}^{min}: minimum production of product $i \in I$ in plant $j \in J_1$;
- I_{ijt}^{max}: maximum storage capacity for product $i \in I$ in plant $j \in J_1$ in period $t \in \mathscr{T}$;

SS_{ijt}: safety stock of product $i \in I$ at facility $j \in J_l$, with $l = 2, 3$ at the end of period $t \in \mathscr{T}$;
S_j^{max}: maximum storage capacity at facility $j \in J_l$, with $l = 2, 3$;
S_j^{min}: minimum storage capacity at facility $j \in J_l$, with $l = 2, 3$;
Q_{jk}: maximum transportation quantity from facility $j \in J_l$ to facility $k \in J_{l+1}$, with $l = 1, 2, 3$;
SP_{ijt}: selling price of product $i \in I$ in customer zone $j \in J_4$ in period $t \in \mathscr{T}$;
C_{jt}: cost of establishing a facility at location $j \in J_l$, $l = 1, 2, 3$, in period $t \in \mathscr{T}$;
FPC_{ijt}: fixed production cost of product $i \in I$ at plant $j \in J_1$ in period $t \in \mathscr{T}$;
VPC_{ijt}: variable production cost of product $i \in I$ at plant $j \in J_1$ in period $t \in \mathscr{T}$;
FTC_{ijkt}: fixed transportation cost of product $i \in I$ from facility $j \in J_l$ to facility $k \in J_{l+1}$, with $l = 1, 2, 3$, in period $t \in \mathscr{T}$;
VTC_{ijkt}: variable transportation cost of product $i \in I$ from facility $j \in J_l$ to facility $k \in J_{l+1}$, with $l = 1, 2, 3$, in period $t \in \mathscr{T}$;
IC_{ijt}: inventory cost per unit of product $i \in I$ at facility $j \in J_l$, $l = 1, 2, 3$, in period $t \in \mathscr{T}$;
r_t: cost of capital rate at the end of period $t \in \mathscr{T}$;
TR_t: tax rate at the end of period $t \in \mathscr{T}$;
IR_t: long-term interest rate at the end of period $t \in \mathscr{T}$;
DPR_{st}: depreciation rate at the end of period $t = s, \ldots, T$ for facilities established in period $s \in \mathscr{T}$;
CR_t: lower limit on cash ratio at the end of period $t \in \mathscr{T}$;
CCR_t: lower limit on cash coverage ratio at the end of period $t \in \mathscr{T}$;
CUR_t: lower limit on current ratio at the end of period $t \in \mathscr{T}$;
ROA_t: lower limit on return on assets ratio at the end of period $t \in \mathscr{T}$;
ROE_t: lower limit on return on equity ratio at the end of period $t \in \mathscr{T}$;
ATR_t: lower limit on assets turnover ratio at the end of period $t \in \mathscr{T}$;
PMR_t: lower limit on profit margin ratio at the end of period $t \in \mathscr{T}$;
QR_t: lower limit on quick ratio at the end of period $t \in \mathscr{T}$;
$LTDR_t$: upper limit on long-term debt ratio at the end of period $t \in \mathscr{T}$;
CP_t: upper limit on new capital entries at the end of period $t \in \mathscr{T}$;
α_t: outstanding revenues coefficient at the end of period $t \in \mathscr{T}$;
μ_t: outstanding payables coefficient at the end of period $t \in \mathscr{T}$;

Decision and Auxiliary Variables:

p_{ijt}: quantity of product $i \in I$ produced in plant $j \in J_1$ in period $t \in \mathscr{T}$;
x_{ijkt}: quantity of product $i \in I$ transported from facility $j \in J_l$ to facility $k \in J_{l+1}$, with $l = 1, 2, 3$, in period $t \in \mathscr{T}$;
q_{ijt}: inventory of product $i \in I$ held in facility $j \in J_l$, with $l = 1, 2, 3$, in period $t \in \mathscr{T}$;
y_{jt}: binary variable taking the value 1 if facility $j \in J_l$, with $l = 1, 2, 3$, is opened in period $t \in \mathscr{T}$ and 0 otherwise;

w_{jst}: binary variable taking the value 1 if facility $j \in J_l$, with $l = 1, 2, 3$, was opened in period $s \in \mathcal{T}$ and is still open in period $t \in \{s, \ldots, T\}$ and 0 otherwise;
u_{ijt}: binary variable taking the value 1 if product $i \in I$ is produced at plant $j \in J_1$ in period $t \in \mathcal{T}$ and 0 otherwise;
z_{jkt}: binary variable taking the value 1 if there is transportation from facility $j \in J_l$ to facility $k \in J_{l+1}$, with $l = 1, 2, 3$, in period $t \in \mathcal{T}$;
b_t: amount borrowed in period $t \in \mathcal{T}$;
rp_t: bank repayment in period $t \in \mathcal{T}$;
ncp_t: new capital entry in period $t \in \mathcal{T}$;

Mathematical Programming Model

The objective is to maximize the firm's value, which is calculated by the summation over the planning horizon of the net operating profit after taxes net of the cost of the invested capital as given in Expression (1).

$$\text{Maximize} \sum_{t \in \mathcal{T}} \left(NOPAT_t - r_t CI_t \right). \tag{1}$$

In each period $t \in \mathcal{T}$, the $NOPAT_t$, which is given by Eq. (2), is obtained by deducting the period costs of sales CS_t (production, transportation, inventory holding, and inventory changing costs), the period depreciation costs (depreciation of the operational facilities), and the period long-term liabilities, which are given by the firm's long term debt (LTD_t) from the revenue obtained from selling the products.

$$NOPAT_t = \left(\sum_{i \in I} \sum_{j \in J_4} SP_{ijt} O_{ijt} - \left(CS_t + \sum_{l=1}^{3} \sum_{j \in J_l} \sum_{s=1}^{t} DPR_{st} C_{js} w_{jst} + IR_t LTD_t \right) \right) (1 - TR_t),\ t \in \mathcal{T}, \tag{2}$$

where $CS_t = PC_t + TC_t + IC_t - (IV_t - IV_{t-1})$ (see Eqs. (3)–(6) and $LTD_t = LTD_{t-1} + b_t - rp_t$.

$$PC_t = \sum_{i \in I} \sum_{j \in J_1} (FPC_{ijt} u_{ijt} + VPC_{ijt} p_{ijt}),\ t \in \mathcal{T}, \tag{3}$$

$$TC_t = \sum_{l=1}^{3} \sum_{j \in J_l} \sum_{k \in J_{l+1}} (FTC_{jkt} z_{jkt} + \sum_{i \in I} VTC_{ijkt} x_{ijkt}),\ t \in \mathcal{T}, \tag{4}$$

$$IC_t = \sum_{l=1}^{3} \sum_{j \in J_l} \sum_{i \in I} IC_{ijt} \frac{q_{ijt} + q_{ijt-1}}{2},\ t \in \mathcal{T}, \tag{5}$$

$$IV_t - IV_{t-1} = \sum_{i \in I} \left(\frac{\sum_{j \in J_1} VPC_{ijt}}{|J_1|} \sum_{l=1}^{3} \sum_{j \in J_l} (q_{ijt} - q_{ijt-1}) \right),\ t \in \mathcal{T}. \tag{6}$$

The cost of the invested capital, referred to in (1), refers to the opportunity cost of making a specific investment and is obtained by multiplying the invested capital by the cost of capital rate (a rate of return that could have been earned by putting the same money into a different investment with equal risk). The capital invested, which is given by Eq. (7), is the amount of money used to fund the project, which is given by the residual claim or interest of the investors in assets and the long-term debt. The former being composed of the previous period equity, the net operating profit after taxes of the current period (assumed to stay in the company), and the new capital entries of the current period.

$$CI_t = E_t + LTD_t, \text{ where } E_t = E_{t-1} + NOPAT_t + ncp_t, \text{ with } t \in \mathscr{T}. \tag{7}$$

The constraints that must be satisfied can be divided into two groups, namely the operational constraints and the financial constraints. The former include demand satisfaction (8), production limits (9)–(10), resources availability (11), inventory limits (12)–(15), inventory balance (16)–(17), transportation limits (18), not allowing a facility to be opened more than once (19), and logical constraints imposing opening facilities in order to have open facilities (20) and the existence of open facilities for sending (21) and receiving (22) products, and also forcing production status variables to be 1 whenever needed (23).

$$\sum_{j\in J_3} x_{ijkt} = O_{ikt}, \qquad \forall i \in I, k \in J_4, t \in \mathscr{T}, \tag{8}$$

$$p_{ijt} \geq P_{ij}^{min} \sum_{s=1}^{t} y_{js}, \qquad \forall i \in I, j \in J_1, t \in \mathscr{T}, \tag{9}$$

$$p_{ijt} \leq P_{ij}^{max} \sum_{s=1}^{t} y_{js}, \qquad \forall i \in I, j \in J_1, t \in \mathscr{T}, \tag{10}$$

$$\sum_{i\in I} \rho_{ije} p_{ijt} \leq R_e, \qquad \forall j \in J_1, e \in E, t \in \mathscr{T}, \tag{11}$$

$$q_{ijt} \leq I_{ijt}^{max} \sum_{s=1}^{t} y_{js}, \qquad \forall i \in I, j \in J_1, t \in \mathscr{T}, \tag{12}$$

$$\sum_{i\in I} q_{ijt} \geq S_j^{min} \sum_{s=1}^{t} y_{js}, \qquad \forall l = 2, 3, j \in J_l, t \in \mathscr{T}, \tag{13}$$

$$\sum_{i\in I} q_{ijt} \leq S_j^{max} \sum_{s=1}^{t} y_{js}, \qquad \forall j \in J_l, t \in \mathscr{T}, \tag{14}$$

$$q_{ijt} \geq SS_{ij} \sum_{s=1}^{t} y_{js}, \qquad \forall i \in I, l = 2, 3, j \in J_l, t \in \mathscr{T}, \tag{15}$$

$$q_{ijt-1} + p_{ijt} - \sum_{k\in J_2} x_{ijkt} - q_{ijt} = 0, \qquad \forall i \in I, j \in J_1, t \in \mathscr{T}, \tag{16}$$

$$q_{ijt-1} + \sum_{m \in J_{l-1}} x_{imjt} - \sum_{k \in J_{l+1}} x_{ijkt} - q_{ijt} = 0, \;\; \forall i \in I, l = 2, 3, j \in J_l, t \in \mathscr{T}, \tag{17}$$

$$\sum_{i \in I} x_{ijkt} \leq Q_{jk} z_{jkt}, \qquad \forall l = 1, 2, 3, j \in J_l, k \in J_{l+1}, t \in \mathscr{T}, \tag{18}$$

$$\sum_{t=1}^{T} y_{jt} \leq 1, \qquad \forall l = 1, 2, 3, j \in J_l, \tag{19}$$

$$w_{jst} = y_{js}, \qquad \forall l = 1, 2, 3, j \in J_l, s, t \in \mathscr{T}, t \geq s \tag{20}$$

$$\sum_{i \in I} \sum_{k \in J_{l+1}} x_{ijkt} \leq M \sum_{s=1}^{t} y_{js} \quad \forall l = 1, 2, 3, j \in J_l, t \in \mathscr{T}, \tag{21}$$

$$\sum_{i \in I} \sum_{j \in J_l} x_{ijkt} \leq M \sum_{s=1}^{t} y_{ks} \qquad \forall l = 1, 2, k \in J_{l+1}, t \in \mathscr{T}, \tag{22}$$

$$p_{ijt} \leq M u_{ijt}, \qquad \forall i \in I, j \in J_1, t \in \mathscr{T}. \tag{23}$$

The financial constraints impose financial ratios, which are a useful section of financial statements and provide standardized measures of a firm's performance, efficiency, leverage, and liquidity. In addition, new capital entries are limited and repayments must prevent an ever increasing debt.

This work uses the categories defined by [3] and sets minimum/maximum threshold values for performance ratios (return on equity (24) and return on assets (25)), efficiency ratios (profit margin (26) and asset turnover (27)), liquidity ratios (current (28), quick (29), and cash (30) ratios), and leverage ratios (long-term debt (31) and cash coverage (32)).

$$\frac{NOPAT_t}{E_t} \geq ROE_t, \qquad \forall t \in \mathscr{T}, \tag{24}$$

$$\frac{NOPAT_t}{NFA_t + CA_t} \geq ROA_t, \qquad \forall t \in \mathscr{T}, \tag{25}$$

$$\frac{NOPAT_t}{\sum_{i \in I} \sum_{j \in J_4} SP_{ijt} O_{ijt}} \geq PMR_t, \qquad \forall t \in \mathscr{T}, \tag{26}$$

$$\frac{\sum_{i \in I} \sum_{j \in J_4} SP_{ijt} O_{ijt}}{NFA_t + CA_t} \geq ATR_t, \qquad \forall t \in \mathscr{T}, \tag{27}$$

$$\frac{CA_t}{STD_t} \geq CUR_t, \qquad \forall t \in \mathscr{T}, \tag{28}$$

$$\frac{C_t + \alpha_t \sum_{i \in I} \sum_{j \in J_4} SP_{ijt} O_{ijt}}{STD_t} \geq QR_t, \quad \forall t \in \mathscr{T}, \tag{29}$$

$$\frac{C_t}{STD_t} \geq CR_t, \qquad \forall t \in \mathscr{T}, \tag{30}$$

$$\frac{LTD_t}{E_t + LTD_t} \leq LTDR_t, \qquad \forall t \in \mathscr{T}, \tag{31}$$

$$\frac{EBIT_t + DPR_t}{IR_t LTD_t} \geq CCR_t, \qquad \forall t \in \mathscr{T}, \tag{32}$$

where NFA_t, CA_t, C_t, and STD_t stand for net fixed assets, current assets, cash, and short-term debt in period $t \in \mathscr{T}$, respectively, and are given by (33)–(36).

$$NFA_t = NFA_{t-1} + \sum_{l=1}^{3} \sum_{j \in J_l} C_{jt} y_{jt} - DPR_t,\ t \in \mathscr{T}, \tag{33}$$

$$CA_t = C_t + \alpha_t \sum_{l=1}^{3} \sum_{j \in J_l} C_{jt} y_{jt} + IV_t, t \in \mathscr{T}, \tag{34}$$

$$\begin{aligned} C_t = {} & C_{t-1} + \alpha_{t-1} \sum_{i \in I} \sum_{j \in J_4} SP_{ijt} O_{ijt} + (1 - \alpha_t) \sum_{i \in I} \sum_{j \in J_4} SP_{ijt} O_{ijt} + ncp_t + b_t \\ & - \mu_{t-1}(PC_{t-1} + TC_{t-1} + IC_{t-1}) - (1 - \mu_t)(PC_t + TC_t + IC_t) \\ & - TR_{t-1}(EBIT_{t-1} - IR_{t-1} LTD_{t-1}) - \sum_{l=1}^{3} \sum_{j \in J_l} C_{jt} y_{jt} - IR_t LTD_t - rp_t, t \in \mathscr{T}, \end{aligned} \tag{35}$$

$$STD_t = \mu_t (PC_t + TC_t + IC_t) + (EBIT_t - IR_t LTD_t) TR_t, t \in \mathscr{T}, \tag{36}$$

where the earnings before interest and taxes in each period $EBIT_t$ is given by

$$EBIT_t = \sum_{i \in I} \sum_{j \in J_4} SP_{ijt} O_{ijt} - CS_t - \sum_{l=1}^{3} \sum_{j \in J_l} \sum_{s=1}^{t} DPR_{st} C_{js} w_{jst}, t \in \mathscr{T}. \tag{37}$$

Investment funds come from limited new capital entries and/or bank loans (38) and new capital entries and repayments are limited as stated by constraints (39) and (40), respectively.

$$\textstyle\sum_{l=1}^{3} \sum_{j \in J_l} C_{jt} y_{jt} = ncp_t + b_t, \qquad \forall t \in \mathscr{T}, \tag{38}$$

$$ncp_t \leq CP_t, \qquad \forall t \in \mathscr{T}, \tag{39}$$

$$rp_t \geq IR_t LTD_t. \qquad \forall t \in \mathscr{T}, \tag{40}$$

Finally, constraints (41)–(43) define the nature of the variables.

$$b_t, rp_t, ncp_t \geq 0, \qquad \forall i \in t \in \mathscr{T}, \tag{41}$$

$$p_{ijt}, q_{ijt}, x_{ijkt} \geq 0, \qquad \forall i \in I, l = 1, 2, 3 j \in J_l, k \in J_{l+1}, t \in \mathscr{T}, \tag{42}$$

$$y_{jt}, w_{sjt}, u_{ijt}, z_{jkt} \in \{0, 1\}, \qquad \forall i \in I, l = 1, 2, 3, j \in J_l, k \in J_{l+1}, s, t \in \mathscr{T}, s \leq t. \tag{43}$$

4 Case Study

The model proposed in this work is tested on the case study used by [11].

Company Alpha already exists and currently has three plants, each being able to produce six of the seven products within its production capacity limits (lower production limits are assumed to be zero) and having nonzero initial inventories. Regarding warehouses and distribution centers, currently, the company holds none and has identified four potential locations for the former and six for the latter. All warehouses and all distribution centers have a minimum and a maximum handling capacity for each time period, as well as safety stock requirements. Product flows between facilities and between distribution centers and customers have upper limits. Selling prices and demands for each of the seven products and each of the eight customer zones are known. The initial nonzero balance of the company is also taken into account. This specific case study considers a 4-years planning horizon.

The proposed model was solved using Gurobi 7.0 solver incorporated in Visual Studio 15.0. The model for the problem instance described was solved optimally in 3.02 CPU seconds.

The total value created amounts to 2,326,120 monetary units.

The optimal network structure consists of the three existing plants (P1, P2 and P3) and three warehouses (W1, W2 and W3) and four distribution centers (DC1, DC2, DC3 and DC5) that are opened in the first year, see Fig. 1. The networks for time periods 2 and 3 and for time period 4 are given in Figs. 2 and 3, respectively, which are provided in the Appendix. The network structure remains the same over the 4-years planning horizon as no other opening decisions are made; however, plant 2 is only used in the first year. This, indicates that closing decisions should be considered.

Regarding the flows between facilities, apart from differences in flows values, there are some changes in the used flows, but not many. These changes mainly occur from year 1 to year 2, since most flows are kept for the remaining periods and with the same value. Nevertheless, a decreasing trend regarding the number of used flows was observed, which is mostly likely due the fact that in [11] the initial inventories for each product at each plant are assumed to be the maximum capacity.

Tables 1, 2 and 3 provide the transportation flows between facilities and between facilities and customers for both our model and that of [11]. These flows are the total value transported regardless of the product type over the 4-years planning horizon.

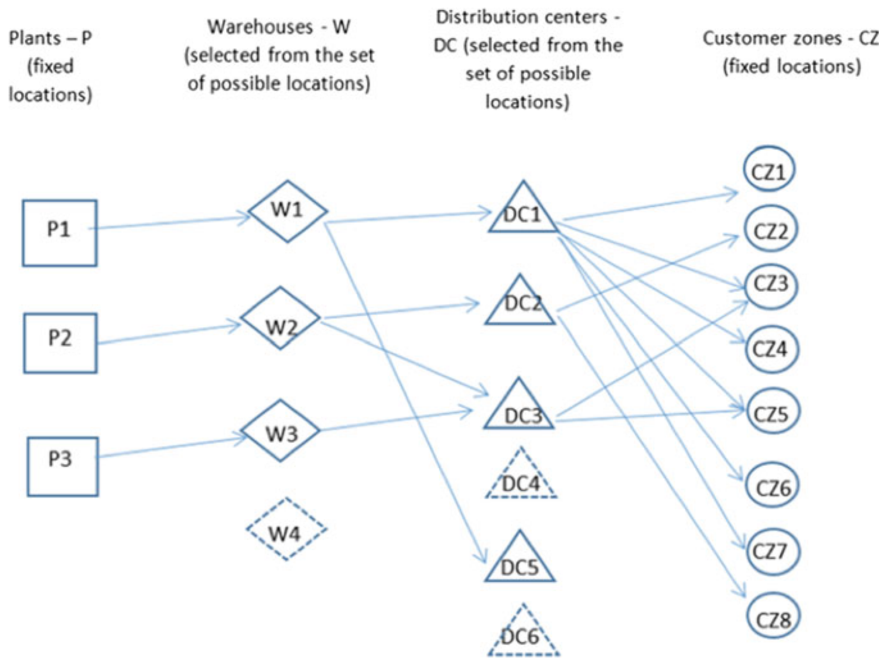


Fig. 1 Network structure and used product flows for the first time period

Table 1 Total aggregated flows transported from the plants to the warehouses. (a) obtained by our model. (b) reported by [11]

(a)

	W1	W2	W3	W4
PL1	7541			
PL2		2174		
PL3			2759	

(b)

	W1	W2	W3	W4
PL1	1685	971	1678	1784
PL2	481	1038	523	1383
PL3	419	744	945	1019

Table 2 Total aggregated flows transported from the warehouses to the distribution centers. (a) obtained by our model. (b) reported by [11]

(a)

	DC1	DC2	DC3	DC4	DC5	DC6
W1	6502				942	
W2		1758	411			
W3			2713			
W4						

(b)

	DC1	DC2	DC3	DC4	DC5	DC6
W1		1211			1349	
W2		876			1818	
W3		1819			1263	
W4	1576	893			1608	

The results of our model show that it is possible to create more value for the company if the correct financial decisions are taken, since the value of the company we found is 32.40% larger than that of the one found in [11]. Recall that in comparison to the work of [11], we consider that both bank repayments and new capital entries

Table 3 Total aggregated flows transported from the distribution centers to the customer zones. (a) obtained by our model. (b) reported by [11]

(a)

	CZ1	CZ2	CZ3	CZ4	CZ5	CZ6	CZ7	CZ8
DC1	1351		100	2019	109	1416	1444	
DC2		1517						194
DC3			1531		201			1350
DC4								
DC5					929			
DC6								

(b)

	CZ1	CZ2	CZ3	CZ4	CZ5	CZ6	CZ7	CZ8
DC1								1544
DC2				2018	1239		1443	
DC3								
DC4								
DC5	1351	1516	1630			1416		
DC6								

are also decisions to be made; although limits to their values are imposed. Our model allows for a balance between loans, repayments, and new capital entries in order to decrease debt and keep a good financial performance; while imposing an upper limit on new capital entries to finance investments because of the known large costs of this funding option and a lower limit on bank repayments to prevent an ever increasing debt. Moreover, the company is allowed to create an accounts payable policy. Furthermore, we also improve on some financial calculations such as depreciation, since the lifetime of each asset is known and the exact depreciation values, if any, in each period can be calculated; real value of the cash is used (rather than an assumed percentage of profit); financial ratios, in accordance with accounting rules, incorporate the net value of fixed assets (not their total value).

5 Conclusions and Future Work

This work is particularly relevant as it extends the scope of financial issues considered in SCND in a dynamic environment.

This paper proposes a mathematical model to address the problem of designing a supply chain network integrating strategical and tactical decisions with financial decisions and considerations, namely: number, locations, and sizes of the facilities (plants, warehouses, and distribution centers); production (product mix and produced quantities at each plant), inventory (which products and quantities are held at each facility), distribution (product flows); and investment decisions (loans and bank repayments). Regarding the constrains, in addition to the usual operational constrains, we also consider minimum and/or maximum threshold values for performance ratios, efficiency ratios, liquidity ratios, and leverage ratios. The objective function of the model is the maximization of the company's value.

The work proposed here extends that of [11] since it uses a formulation in a dynamic environment that approaches reality, allowing to make decisions for every time period. In addition, it considers accounting rules and uses less assumptions: considers decisions on repayments and new capital entries, improves depreciation and cash calculations, and allows for creating an accounts payable policy.

This work is limited in several ways and some extensions are already been considered. Currently, we are already working on incorporating the ability of closing and/or selling existing facilities, in addition to being able to open new plants whenever necessary. Another extension to address in the very near future is the incorporation of stochastic elements.

Acknowledgements We acknowledge the support of FEDER/COMPETE2020/NORTE2020/ POCI/ PIDDAC/MCTES/FCT funds through grants PTDC/EEIAUT2933/2014 and 02/SAICT/ 2017-31821/ POCI-01-0145-FEDER-031821.

Appendix

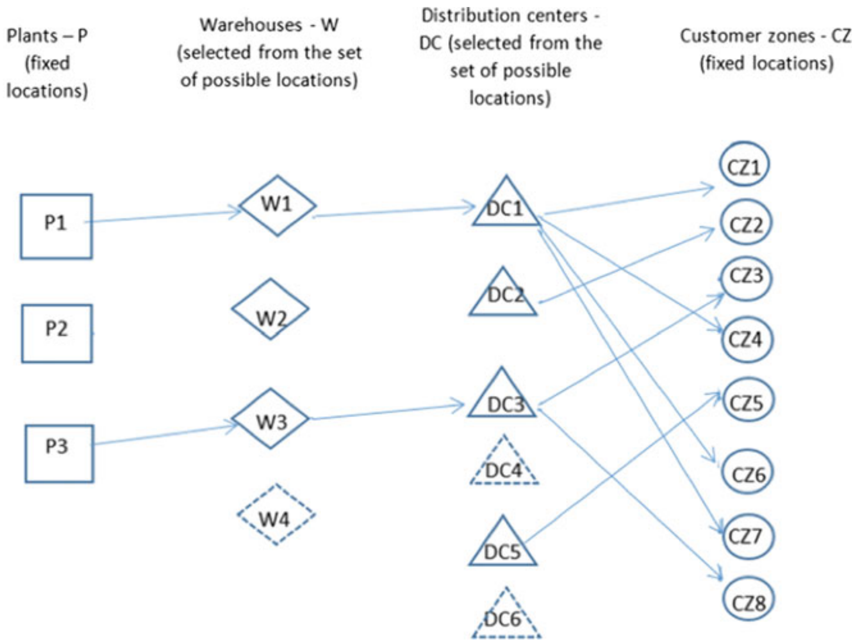


Fig. 2 Network structure and used product flows for the second and third time period

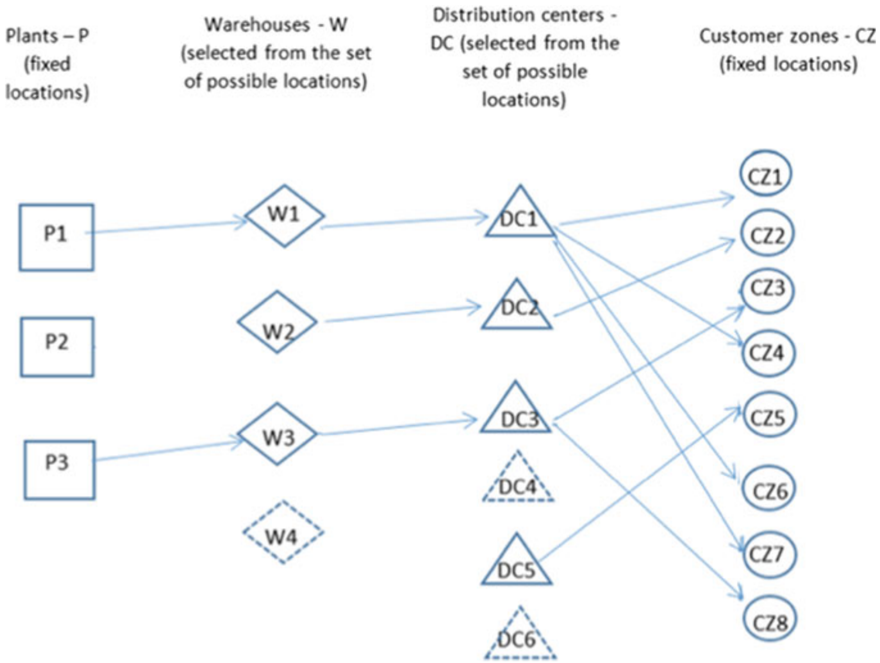


Fig. 3 Network structure and used product flows for the fourth time period

References

1. Badell, M., Nougues, J., Puigjaner, L.: Integrated on line production and financial scheduling with intelligent autonomous agent based information system. Comput. Chem. Eng. **22**, S271–S277 (1998)
2. Badell, M., Romero, J., Huertas, R., Puigjaner, L.: Planning, scheduling and budgeting value-added chains. Comput. Chem. Eng. **28**(1–2), 45–61 (2004)
3. Brealey, R.A., Myers, S.C., Allen, F.: Principles of Corporate Finance, 10th edn. (2011)
4. Elgazzar, S.H., Tipi, N.S., Hubbard, N.J., Leach, D.Z.: Linking supply chain processes performance to a company financial strategic objectives. Eur. J. Oper. Res. **223**(1), 276–289 (2012)
5. Guillén, G., Badell, M., Espuna, A., Puigjaner, L.: Simultaneous optimization of process operations and financial decisions to enhance the integrated planning/scheduling of chemical supply chains. Comput. Chem. Eng. **30**(3), 421–436 (2006)
6. Hammami, R., Frein, Y., Hadj-Alouane, A.B.: Supply chain design in the delocalization context: relevant features and new modeling tendencies. Int. J. Prod. Econ. **113**(2), 641–656 (2008)
7. Hammami, R., Frein, Y., Hadj-Alouane, A.B.: A strategic-tactical model for the supply chain design in the delocalization context: mathematical formulation and a case study. Int. J. Prod. Econ. **122**(1), 351–365 (2009)
8. Klibi, W., Martel, A., Guitouni, A.: The design of robust value-creating supply chain networks: a critical review. Eur. J. Oper. Res. **203**(2), 283–293 (2010)
9. Laínez, J.M., Puigjaner, L., Reklaitis, G.V.: Financial and financial engineering considerations in supply chain and product development pipeline management. Comput. Chem. Eng. **33**(12), 1999–2011 (2009)

10. Lambert, D.M., Cooper, M.C., Pagh, J.D.: Supply chain management: implementation issues and research opportunities. Int. J. Logist. Manag. **9**(2), 1–20 (1998)
11. Longinidis, P., Georgiadis, M.C.: Integration of financial statement analysis in the optimal design of supply chain networks under demand uncertainty. Int. J. Prod. Econ. **129**(2), 262–276 (2011)
12. Longinidis, P., Georgiadis, M.C.: Integration of sale and leaseback in the optimal design of supply chain networks. Omega **47**, 73–89 (2014)
13. Meixell, M.J., Gargeya, V.B.: Global supply chain design: a literature review and critique. Transp. Res. Part E: Logist. Transp. Rev. **41**(6), 531–550 (2005)
14. Melo, M.T., Nickel, S., Saldanha-Da-Gama, F.: Facility location and supply chain management-a review. Eur. J. Oper. Res. **196**(2), 401–412 (2009)
15. Moussawi-Haidar, L., Jaber, M.Y.: A joint model for cash and inventory management for a retailer under delay in payments. Comput. Ind. Eng. **66**(4), 758–767 (2013)
16. Papageorgiou, L.G.: Supply chain optimisation for the process industries: advances and opportunities. Comput. Chem. Eng. **33**(12), 1931–1938 (2009)
17. Protopappa-Sieke, M., Seifert, R.W.: Interrelating operational and financial performance measurements in inventory control. Eur. J. Oper. Res. **204**(3), 439–448 (2010)
18. Ramezani, M., Kimiagari, A.M., Karimi, B.: Closed-loop supply chain network design: a financial approach. Appl. Math. Model. **38**(15–16), 4099–4119 (2014)
19. Shapiro, J.F.: Challenges of strategic supply chain planning and modeling. Comput. Chem. Eng. **28**(6–7), 855–861 (2004)
20. Stern, J.M.: One way to build value in your firm, a la executive compensation. Financ. Exec. **6**(6), 51–55 (1990)
21. Yi, G., Reklaitis, G.V.: Optimal design of batch-storage network with financial transactions and cash flows. AIChE J. **50**(11), 2849–2865 (2004)
22. Yi, G., Reklaitis, G.V.: Optimal design of batch-storage network considering exchange rates and taxes. AIChE J. **53**(5), 1211–1231 (2007)

Performance Evaluation of European Power Systems

Mário Couto and Ana Camanho

Abstract Electric power systems are facing significant challenges regarding their organization and structure. Energy infrastructures are crucial to ensure a transition to low-carbon societies, contributing to sustainable development. This paper uses Data Envelopment Analysis to compare the performance of the power systems in 16 European countries using data available to the public. Three perspectives were considered, focusing on technical aspects affecting quality of service, network costs and environmental impact. It is proposed a new formulation of the DEA model that estimates a composite indicator (CI) aggregating individual indicators which should be minimized. The benchmarking results can give insights to electric operators, regulators and decision-makers on the strengths and weakness of national power systems and disclose the potential for performance improvements. Based on the outcomes from the CI model, Austria, Croatia, Denmark, Germany, Greece, Ireland, Italy and Netherlands are identified as the benchmarks for the power systems in the Europe. The discussion of the results is intended to raise public awareness on the performance of the European power systems and contribute to the definition of public policies for the promotion of continuous improvement.

Keywords Electric power systems · Data envelopment analysis · Composite indicators · International benchmarking

1 Overview

Electric infrastructures have a strategic position in today's society due to their impact on the environment and industrial development. Therefore, the electric sector is crucial to ensure the transition to low-carbon societies. This requires a growing share

M. Couto (✉)
Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
e-mail: up200503817@fe.up.pt

A. Camanho
INESC TEC, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
e-mail: acamanho@fe.up.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_6

of renewable sources, improved energy efficiency and the promotion of the electrification of the transport sector [1]. The electric sector has been experimenting liberalization and restructuring processes over the three last decades. Historically, these processes started in Chile and Argentina in 1980, spread to England and Wales in 1990, and then to other European countries, Australia, New Zealand, Latin American, and the United States [2]. In the European context, liberalization plays an important role in the setting of the European common energy market. The creation of this common market has three targets: reduction of carbon intensity, security of supply, and establishment of fair and affordable prices for the consumers [1]. Initially, the goal of the liberalization process was the introduction of competition in the sector, through the creation of competitive wholesale markets organized in terms of pool mechanisms to establish the relationships between generator entities and distributors or large consumers [2]. The conditions of access to the networks were a key feature to develop electricity markets. As consequence, the need to turn these access conditions transparent and fair to all users led to unbundling the traditional vertically integrated companies. This allowed the separation of generation assets from transmission network activities and from the distribution sector. In the second step of the liberalization process, distribution network activities were decoupled from retailing activities. This lead to the creation of purely commercialization companies, without the ownership of network assets and not responsible for the operation, maintenance or expansion of the networks. From this point of view, the frameworks at the transmission and distribution levels are very similar. In both cases, there are network activities usually provided as a natural monopoly, at a national or regional level. Furthermore, there are several users in terms of generation companies, retailing entities and consumers that must pay tariffs for the use of the network assets. Besides the organizational changes, the liberalization and restructuring processes also had consequences at the regulatory level. The new structure of the electric value chain clearly created the need to regulate transmission and distribution network operators, as they are not subjected to competition. According to [3], one of the consequences of the unbundling of the traditional vertically integrated companies was the need to pass new tariff codes, establishing a tariff system that sets the rules to determine the tariffs for each activity. The existence of new players in the electric sector requires the end user tariffs to reflect the costs all along the value chain, from generation to demand (additive tariffs). At the European level, the Regulatory Agencies adopt different strategies to set the tariffs for the use of the networks. The role of Regulatory Agencies is to guarantee economic viability conditions for network companies, the promotion of economic and technical efficiency, the increase of quality of supply levels and a more stable and less volatile environment [4]. The assessment of performance of regulated operators requires the use of models that can account for different regulatory objectives. Benchmarking has been adopted in several countries to compare the performance of different operators. This paper describes a benchmarking analysis of the electric systems of European countries, which can complement the comparative analysis usually conducted at the utilities level. Unlike most previous studies, which have focused either on the performance of generation assets or the performance of grid operators to support the regulatory role for network businesses, we position our study at a country level. We aim to provide a new perspective concerning the alignment of national power systems with the paradigms of sustainable development, according to Elkington (1997) framework [5], named Triple Bottom Line (TBL). This framework states that economic, environmental and social aspects must be simultaneously taken into account when assessing the ability to meet the needs of the present without compromising the needs of future generations [6, 7]. The purpose of our study is to benchmark the performance of power systems of European countries concerning technical and economic aspects of network operation and environmental aspects of electricity sourcing. In the scope of this research, the analysis of power systems was performed considering the quality of supply, network costs and the share of renewables in energy generation. Given the degree of heterogeneity in terms of operating conditions and legislation settings in EU countries, these results are not intended to be used for regulatory purposes. Instead, this study is a formative evaluation that can allow a comparison of the status of the power systems in different countries in terms of the three sustainability pillars. This analysis can highlight the issues that should be addressed to facilitate the creation of a common electric market, and contribute to the design of policies to reduce asymmetries among European countries. Alternative approaches that could be used for regulatory purposes or target setting should take into account the heterogeneity of countries. This would require procedures such as a-prior clustering of countries, the use of conditional efficiency measures or second stage analysis to correct the performance estimates for the effect of non-discretionary or exogenous factors affecting the operation of power systems in different countries. This type of assessments is out of the scope of the current paper. The paper is organized as follow: Sect. 2 provides a literature review about application of DEA in electric sector, sector Sect. 3 describes the context setting to perform the research, sector Sect. 4 presents the model formulation, finally, in the Sect. 5 the results are presented and discussed.

2 Literature Review

In what concerns the electric grids, the DEA technique has been widely applied for utilities regulation. Indeed, benchmarking analyses have often been used for the evaluation of electricity transmission and distribution companies [8]. Several countries use incentive regulation for electric utilities, which requires the use of performance measurement techniques to assess relative efficiency and compute efficiency factors for tariff adjustment once that the competition is simulated through regulation. Benchmarking regulation aims to make regional monopolies compete with a reference (an efficient model company) such that competition is achieved by comparison with peers [9]. Reviewing the literature on this topic, the work of [10] provided guidelines for the regulation of power transmission and distribution in China. The efficiency of companies was estimated using DEA, and the results were complemented with the identification of one or more yardstick companies through intensive numerical experiments. The shadow prices corresponding to each input and output variable were also determined. In the first phase of the study, the companies were clustered in groups according to their scale size and benefits received, in accordance with the abundance of the power transmission and distribution companies in the area they operate within China. This study focused primarily on the costs, electricity charges and other technology and economy related variables. Then, the data from different groups was classified and a relative overall efficiency measure was calculated using DEA. Simab et al. [11] combined the use of DEA with fuzzy c-means clustering (FCM) to obtain the parameters of a reward and penalty scheme for performance-based regulation. The FCM algorithm contributed to identify similar distribution companies and gather them into different clusters, while the DEA was used to set a quality target for each electric distribution company. In some countries in Latin-American, Iran, Italy Portugal and Norway, DEA is used as a benchmarking method for the regulation of the distribution sector, with favorable results widely acknowledged [12–21]. Concerning the electric distribution sector, in [20] the authors used DEA to estimate the potential cost savings from horizontal mergers among distribution companies in Norway. They compared the results of post-merger performance to identify plausible pre-merger motivations. It is also made a comparison between regulated revenues and the regulated efficiency improvement requirements before and after the mergers. From the survey of the literature, an important conclusion is the lack of consensus on the choice of input and output variables to be included in the benchmarking models. A possible explanation is the variation in data availability in different settings, as well as the diverse objectives of the studies. In [18], the authors discuss the best choice of input-output variables to measure technical efficiency in a cost model and in a cost-and-quality model for the largest Italian distribution utility. Their results show that quality has a significant effect to identify the efficiency scores in different geographic zones. Although the scope of this work is not the regulation of operators, the conclusions from previous works are important to define the indicators to be used in our approach. Additionally, they provide the contextual framing for the analysis of power systems with a triple bottom line perspective, including economic and social aspects regarding network operation and environmental aspects concerning the energy sourcing.

3 Context Setting

The performance evaluation of the European power systems was conducted observing the three perspectives of the triple bottom line (economic, environment, social). This analysis aims to explore the alignment of power systems with the sustainable development paradigm and their contribution to environmental and social welfare. Transposing the triple bottom line concept for the analysis conducted in this research, the economic pillar is evaluated using the network costs, the social impact is assessed by the continuity of supply (SAIDI and SAIFI) provided to the customers and by the losses level in the transmission process. Finally, the environmental impact is evaluated using the amount of electricity generated from renewable sources. Table 1

Table 1 Indicators used in the DEA model

Dimension	Indicators
Social dimension: quality of supply	SAIDI (minutes)
	SAIFI (interruptions per year)
	Losses (in % of consumption)
Economic dimension: cost performance	Cost per kilometer (k €/km)
Environmental dimension: environmental performance	Share of non-renewable energy (%)

presents the indicators used for the performance evaluation of the power systems of European countries. All data used is available in the public domain.

A critical issue that often hinders this type of analysis is the unavailability of data for some years, as well as the dispersion of information in multiple sources/databases. Our analysis used the year 2014 as a reference, as this was the most recent year with reliable data available for most countries. The only exception were the indicators for quality of supply, which were collected for the year 2013. The data collected for the performance assessment of 16 European countries are reported in Table 2. The data sources are also reported.

Normally, the quality of supply is evaluated in three dimensions: continuity of supply, power quality and commercial quality. The continuity of supply is related to the number and duration of interruptions. In this research, we focused the analysis on the indicators more commonly used by electric utilities to represent the continuity of supply [26]. This topic has been consistently in the agenda of the regulators, and most European countries have standards for the levels of System Average Interruption Duration Index (SAIDI) and the System Average Interruption Frequency Index (SAIFI), which characterize the quality of service provided to the customers. These are the indicators used in our analysis, both measured on an annual basis using the formulas shown in (1) and (2).

$$SAIDI = \frac{\sum U_i N_i}{N_T} \tag{1}$$

$$SAIFI = \frac{\sum \lambda_i N_i}{N_T} \tag{2}$$

The notation used is as follows: N_i is the number of customers for the network point i, U_i is the annual outage time for the network point i (sum of all annual time interruptions), λ_i is the failure rate at the network point i and N_T is the number of customers served by the electric grid. None of these indicators is weighted according to the consumption. They are influenced by the level of investment in the network (lines reinforcement, automated devices, network redundancy) and maintenance actions. The values collected for SAIDI and SAIFI are for unplanned interruptions, excluding the exceptional events like extreme weather conditions. These indicators are directly provided by CEER (Council of European Energy Regulators) in [22]. Losses in the

Table 2 Indicators used in the DEA model

Source	CEER - Council of European Energy Regulators [22]			ENTSOE - European Network of Transmission Systems Operators for Electricity [23]	DataBank The World Bank Data [24]	EUROSTAT-European Statistics [25]		PORDATA - Portugal - Data Base [36]		
Country	SAIDI (minutes)	SAIFI (interruptions per year)	Total circuit length (1000 km)	Annual consumption (GWh)	Losses (GWh)	% Network costs	Share of non-renewable energy (%)	Energy Price for householders (€/kWh)	Network costs (k€/km)	Losses (% consumption)
Austria	33.96	0.820	257.351	69294	3284	29.65	29.9	0.183	14.6	4.74
Croatia	176.10	1.940	143.422	16407	1764	31.82	54.7	0.207	7.53	10.75
Czech Republic	98.01	1.690	246.178	62000	3847	53.54	86.1	0.218	29.45	6.20
Denmark	11.25	0.320	171.026	33349	1973	27.96	51.5	0.286	15.6	5.92
France	68.10	0.870	1437.85	465051	34545	35.19	81.7	0.144	16.43	7.43
Germany	15.32	0.470	1797.756	529369	24159	22.56	71.9	0.286	19.00	4.56
Greece	96.00	1.600	252.757	49258	4149	15.08	78.1	0.213	6.26	8.42
Hungary	67.21	1.040	167.475	39521	3631	37.39	92.7	0.211	18.63	9.19
Ireland	82.00	1.178	176.552	26188	2028	27.56	77.1	0.216	8.84	7.74
Italy	42.27	1.63	1315.245	308428	19541	17.35	66.6	0.244	9.92	6.34
Poland	254.90	3.020	766.329	146909	10250	38.30	87.6	0.245	17.99	6.98
Portugal	88.70	1.750	232.153	48797	5209	26.91	47.9	0.280	15.81	10.67
Spain	58.20	1.420	763.398	258131	26393	21.10	62.2	0.242	17.26	10.22
Sweden	70.77	0.970	559.608	135533	7334	38.50	36.8	0.149	13.86	5.41
The Netherlands	23.00	0.300	264.426	110941	4394	30.06	90.0	0.168	21.21	3.96
United Kingdom	54.71	0.589	789.315	330636	27440	20.90	82.2	0.168	14.70	8.30
Average	77.53	1.225	583.803	164363	11246	29.62	68.6	0.216	15.44	7.30
Standard Deviation	60.00	0.690	506.569	159036	10774	9.35	18.9	0.044	5.52	2.10
Median	67.66	1.109	260.889	90118	4802	28.80	74.5	0.215	15.71	7.20

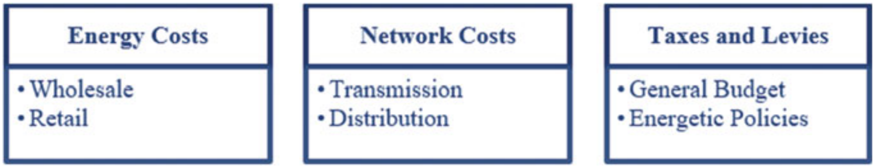


Fig. 1 Elements of the end-consumer price

electric grids are a significant part of the overall losses in the electric power system. They can be defined as the amount of electric energy that is needed by an electric system to supply the demand. A reduction of network losses would make an important contribution to increase energy efficiency in electricity supply. Nevertheless, there is a minimum level from which it is no longer possible to reduce the losses. The electric losses are usually grouped in two main components: technical and non-technical. The first one corresponds the conversion of the electric energy to heat and noise during the transmission process. The second component is related to the energy delivered and consumed, but it is not recorded by a meter due to a several reasons, such as illegal consumption of electricity and errors in metering, billing and data processing. According to the last report on losses of the CEER [24], there is a lack of harmonized definitions and rules to quantify the different components of losses across Europe, especially in case of the non-technical component. Thus, in this paper losses are treated as a global value, computed in percentage of the consumption (Losses/Annual Energy Consumption), without distinguishing between the technical and non-technical components. As the dimensions of the European electric systems are very different, it is not appropriated to consider the losses in absolute values, since larger countries will have larger values of absolute losses. The energy bills paid by consumers have three components: energy, network and taxes/levies [27], as shown in Fig. 1. The energy costs are subdivided in wholesale and retail prices. The wholesale price corresponds to the burdens supported by the companies to deliver energy, including fuel purchase, fuel shipping and processing, costs of constructing, operating and decommissioning of power stations. On the other hand, the retail price is relative to the sale of energy to the final consumer. The network costs correspond to the costs incurred by transmission and distribution companies to support maintenance and expansion of the grids. Lastly, the taxes and levies cover the general taxation or specific levies defined by governments to support energy and climate policies.

In the scope of this paper, the evaluation of economic performance of the European electric systems was based on the network component of costs. The indicator used was the network cost per kilometer. This data was obtained from EUROSTAT, using the information about the different components of the electric price for households. The circuits lengths for the different voltage levels and the electric consumption were also considered to obtain the cost per kilometer. The network cost for each country was computed using the expression shown in (3).

$$\text{Cost (per km)} = \frac{p_E \times E_C \times \%\text{network costs}}{L_{total}} \quad (3)$$

The notation is as follows: p_E is the final electric price for households (€/kWh, excluding the power tariff) weighted by the PPS (Purchasing Power Standard), E_C is the electric consumption of each European country and L_{total} corresponds to the total length of the electric circuits encompassing all voltage levels in the value chain (Extra High Voltage, High Voltage, Medium Voltage and Low Voltage). %Network Costs represents the proportion of network costs in the total price for the end consumer. It should be noticed that p_E, E_C, L_{total} were obtained in public database presented in Table 2.

We also collected from EUROSTAT the share of electricity from non-renewable sources in gross electricity consumption.

4 Methodology

DEA is a non-parametric method that derives a single summary score of relative efficiency for each decision-making unit (DMU) under assessment when compared to a set of homogenous units. DEA estimates a relative efficiency score by the extent to which the DMU (firm or country) under assessment matches or falls short of the best performance levels observed in peers. The peers with best performance constitute the anchors used for the estimation of a piece-wise linear frontier enveloping the data of the sample under analysis. There are a number of considerations involved in the construction of a DEA model, including the most appropriate returns to scale assumption and the appropriate direction to be used in the measurement of the distance to the frontier. Regarding the returns to scale assumption, constant returns to scale (CRS) assume that changes in the inputs result in equiproportional changes in the outputs, irrespectively of the scale size of the unit under assessment. This means that maximum productivity levels is assumed to be potentially achievable for all dimensions of the operation. When the indicators underlying the evaluation of performance are defined as percentages or ratios of outputs to inputs (which implicitly assume a fair comparison is possible irrespectively of the original scale size of the operation or dimension of the countries under assessment), the DEA model used should assume CRS. Thus, the model used in our study was built under a CRS assumption. With respect to the orientation of the projection towards the frontier, an input oriented model aims to minimise the amount of inputs consumed, and an output orientation aims to maximise the quantity of the outputs produced. In the performance assessment reported in this paper, we depart from the traditional DEA paradigm considering the transformation of inputs (resources) into outputs (service delivered) and consider the paradigm underlying the construction of composite indicators (CIs). In a CI we only have performance indicators to be aggregated, so we can assume that all units are comparable. Thus, following [28–31] who first suggested this type of approach, we can have a unitary indicator underlying the evaluation of every DMU, interpreted as a "helmsman" attempting to steer the DMU towards the maximization of performance. This paper develops a new formulation for the composite indicator model, as shown in (4). It differs from the formulation proposed by [32], known as benefit of the doubt, because the indicators considered in our assessment should all be minimized, as lower values of the indicators correspond to better performance. In our case, we have to consider a dummy output equal to one and several inputs to be aggregated in the evaluation of the performance level of each country.

$$\text{Max} = u \tag{4}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} v_i x_{ij_0} = 1 \tag{5}$$

$$u - \sum_{i=1}^{m} v_i x_{ij} \leq 0,\ j = 1, \ldots, n \tag{6}$$

$$v_i \geq \varepsilon, i = 1, \ldots, m \tag{7}$$

Model (4)–(7) is computed separately for each country and the subscript j_0 refers to the country whose relative performance is under evaluation. x_{ij} is the value of the indicator i $(i = 1, \ldots, m)$ observed for country j. u and v_i are the decision variables of the linear programming model. v_i corresponds to the weight assigned to indicator i $(i = 1, \ldots, m)$ for the evaluation of performance, and u is the weight assigned to the dummy output variable. It also corresponds to the optimal value of the composite indicator score for the country under evaluation, as this is the only component of the objective function. Model (4) aims to find the values for the weights u and v_i that enable the estimation of the performance of country j_0 in the best possible light. The performance (efficiency) measure for the country j_0 is maximised (corresponding to a composite indicator of performance resulting from the aggregation of the performance indicators specified), subject to the constraint that the efficiency measures must be less than or equal to one for all countries in the sample when evaluated with similar weights (second constraint). The first constraint corresponds to the normalization of the denominator of the efficiency measure (ratio of the weighted dummy output over the weighted aggregation of inputs) for the country under assessment (j_0). If using the optimal weights for country j_0 no other country reaches a score of aggregate performance higher than the value assigned to country j_0, it implies that the country defines the best-practice frontier. In this case the objective function of (4) returns a score equal to one. Otherwise, country j_0 is considered inefficient, meaning that it is located below the frontier.

5 Results and Discussion

The performance of the European power systems was evaluated using model (4). The results of the composite indicator model are shown in Table 3. The average composite indicator score is 89.9%, representing the average spread of performance

Table 3 Results of the DEA overall assessment

Country	Overall CI score (%)
Austria	100.00
Croatia	100.00
Czech Republic	70.08
Denmark	100.00
France	81.64
Germany	100.00
Greece	100.00
Hungary	69.34
Ireland	100.00
Italy	100.00
Poland	73.70
Portugal	79.31
Spain	73.10
Sweden	97.40
The Netherlands	100.00
United Kingdom	93.68

of European Countries in relation to a best-practice frontier, which is assigned a reference performance value of 100%.

Analyzing the results from Table 3, the countries with best performance when they can freely choose the weighting system that shows them in the best possible light are Austria, Croatia, Denmark, Germany, Greece, Ireland, Italy and Netherlands. The other countries (Czech Republic, France, Hungry, Poland, Portugal, Spain, Sweden and UK) have score for simultaneous improvements in quality of supply, network losses, energy sourcing and network costs. Note that the flexibility in the choice of weights means that the indicators may be assigned quite different weights across the 5 performance dimensions considered in the DEA model (including zero weights). Thus, the countries located in the best-practice frontier may occupy a range of different positions along the frontier, according to their profile in the dimensions considered. For example, Greece is efficient mainly because of the cost per km indicator, despite the modest performance in other indicators such as SAIDI and share of renewables. It is thus important that administrative authorities interpret the CI scores with caution, exploring the root causes of good performance signaled by the results obtained. For example, the countries with a lower cost due to judicious investment must be distinguished from situations of lack of investment. The trade-offs between quality and cost should also be taken into account, as a country may adopt an ambitious investment plan in the electric grid to achieve remarkable improvements in the quality of supply, such that the low performance in one dimension is justified by improvements in other dimensions. The potential for improvement in the different indicators considered in the analysis is summarized in Table 4. These results

Table 4 Outcomes for the DEA models

Description	SAIDI		SAIFI		Losses		Cost per kilometer		Share of non-renewables	
	Target	Gain (%)	Target	Gain (%)	Target	Gain (%)	Target	Gain (%)	Target	Gain (%)
Austria	33.96	0	0.82	0	0.05	0	14.6	0	29.85	0
Croatia	176.1	0	1.94	0	0.11	0	7.53	0	54.71	0
Czech Republic	28.4	−71.02	0.56	−67.08	0.04	−29.92	17.95	−39.05	60.35	−29.92
Denmark	11.25	0	0.32	0	0.06	0	15.6	0	51.51	0
France	37.82	−44.47	0.71	−18.36	0.06	−18.36	13.41	−18.36	52.14	−36.17
Germany	15.32	0	0.47	0	0.05	0	19	0	71.86	0
Greece	96	0	1.6	0	0.08	0	6.26	0	78.08	0
Hungary	41.29	−38.56	0.72	−30.66	0.06	−30.66	12.92	−30.66	57.58	−37.9
Ireland	82	0	1.18	0	0.08	0	8.84	0	77.1	0
Italy	42.27	0	1.63	0	0.06	0	9.92	0	66.58	0
Poland	36.34	−85.74	1.05	−65.16	0.05	−26.3	13.26	−26.3	40.38	−53.9
Portugal	70.35	−20.69	1.12	−35.96	0.06	−41.02	12.54	−20.69	38	−20.69
Spain	42.54	−26.9	1.04	−26.9	0.06	−44.74	12.62	−26.9	45.49	−26.9
Sweden	44.31	−37.39	0.93	−3.71	0.05	−2.6	13.5	−2.6	35.82	−2.6
The Netherlands	23	0	0.3	0	0.04	0	21.21	0	90.02	0
United Kingdom	30.39	−44.44	0.55	−6.32	0.06	−23.05	13.77	−6.32	58.43	−28.88

show the dimensions that require more attention in order to align practices with the best levels observed in other countries. For example, Portugal has a potential for improvement of at least 21% in all indicators considered, with the largest scope for improvement being in SAIFI (36%) and losses (41%). The results reported in Table 4 should be evaluated with caution since the required reductions may not be easy to achieve. Cross-country comparisons always involve some degree on heterogeneity in terms of operating conditions and regulatory settings, such that it may be difficult to transfer good-practices observed elsewhere to a different context. Nevertheless, some general conclusions can be drawn in terms of priority areas for improvement.

In order to keep losses at reasonably low levels, national regulators have established incentive mechanisms that deliver rewards (or penalties) for network operators whose losses are below (or above) a pre-set target level [33]. These mechanisms aim to encourage distribution system operators to adopt policies leading to the optimization of losses through network reconfiguration, operational procedures and investment decisions. In general, losses are proportional to the amount of energy delivered and to the distance between generation and consumption, and inversely related to the voltage level of the network. Consequently, any measures or actions focused on reducing or smoothing the demand for energy, motivating distributed generation, and upgrading the voltage level of the network, will have a positive impact on losses. However, such actions require investments. Given that the final tariffs are built based on an additive principle, additional investments to enhance efficiency can lead to an increase in the final price. Although such investment may be targeted for improvements of quality of service and network efficiency, the regulator has a key role to control the level of the network investment to avoid its rise to unreasonable levels, with undesirable effects on prices. As the energy sector has a capital-intensive nature, the policies must be carefully considered and decisions must be made with a medium/long term horizon, whereby remarkable improvements are not achievable in a short term. Quality of supply also requires considerable investments to support improvements. It is significantly affected by the percentage of underground cables of the distribution networks. An improvement of the quality of supply could require better maintenance policies, an increase in the automation level to support a faster service restoration, or a careful evaluation of the possibility of replacement of overhead lines by underground lines in critical areas. A country with the majority of the electric grid comprised by overhead lines has more difficulties to ensure higher levels of service. The dispersion of the consumption can influence the quality of supply. Typically, the rural zones have less redundancy comparing with the urban zones, since it is more difficult to obtain payback for the investments. Thus, when a line has a fault, the impact on the quality of supply can be more serious than in the urban areas. For this reason, some countries analyze the quality of supply indicators per zones. In further research, it would be interesting to decouple the indicators per zones (rural, urban, suburban) for each country. Furthermore, nowadays there is a tendency to focus on asset management, aiming to optimize the assets lifecycle and implement predictive maintenance. This type of maintenance allows smarter decisions about when and where interventions should be done to reduce overall maintenance costs. The development of Smart-Grids is also a pathway to have access to large amounts of data, which can lead to better performance of electric grids in terms of quality of supply and losses. Therefore, in the coming years, data analytics is expected to play an important role to improve electric grids management, moving from an investment-oriented perspective to a more operational perspective. Next, we conducted individual assessments focused on cost performance, quality of service and environmental performance. This is intended to provide detailed insights of individual performance dimensions. The cost performance assessment considered the indicator network cost per kilometer. The quality of supply considered a composite indicator of SAIDI, SAIFI and Losses, computed using model (4)–(7) to aggregate these three indicators using weights determined by optimization. The environmental indicator considered the indicator share of energy from nonrenewable sources. The results of these analysis are reported in Table 5. The results obtained for the individual performance dimensions are also pictorially illustrated in Figs. 2 and 3. These figures plot the results of the overall performance assessment (x axis) against the results of the performance assessments focused on individual dimensions (cost performance in shown on the left of Fig. 2, quality of supply on the right of Fig. 2, and environmental performance on Fig. 3. The dashed horizontal lines in the graphs separate the first and last quartiles (percentile 25% and percentile 75% for the individual performance dimensions).

Concerning cost performance, Greece presents the best performance, followed by Croatia, Ireland and Italy. As previously explained, this assessment considers exclusively network cost per km, and these countries have a significant percentage of electric circuits comprised by overhead lines [22]. The price to install and maintain

Table 5 Results of the performance assessment of specific dimensions

Country	Cost performance (%)	Quality of supply (%)	Environmental performance (%)	Overall CI score (%)
Austria	42.88	85.11	100	100
Croatia	83.13	37.04	54.56	100
Czech Republic	21.26	64.52	34.66	70.08
Denmark	40.13	100	57.95	100
France	38.10	54.05	36.55	81.64
Germany	32.95	100	41.54	100
Greece	100	47.62	38.23	100
Hungary	33.60	43.48	32.19	69.34
Ireland	70.81	51.95	38.72	100
Italy	63.10	63.49	44.83	100
Poland	34.80	57.14	34.08	73.70
Portugal	39.60	37.38	62.29	79.31
Spain	36.27	39.31	47.97	73.10
Sweden	45.17	74.07	81.16	97.40
The Netherlands	29.51	100	33.16	100
United Kingdom	42.59	50.85	36.33	93.68

overhead lines is smaller than in underground circuits. In contrast, despite the higher prices, underground lines provided a better quality of supply since they are not exposed to external agents (weather, trees, birds) like the aerial lines. In the report about energy prices [27], the European Commission verified that for households the network cost ranged from 0.022 and 0.097 €/kWh. Due to the non-standardized data information on energy costs is not easy to find explanations about the differences in network costs. Such differences can be assigned to a wide differentiation of national network tariffs regulation, cost allocation practices and physical differences in the network infrastructure. It would be interesting to compare the percentages of network costs used to remunerate the investments made in the electric grid and the part to remunerate the operational costs. However, the available data does not allow such comparison. Comparing the graphics from Fig. 2, one can conclude that the countries with a larger undergrounded component (Austria, Denmark, Germany, Netherlands) are in the top performance quartile for the quality of supply, while for the economic performance they just appear in the second and third quartile. An opposite case occurs for Italy, Croatia, Ireland and Greece, as they are in the top quartile for cost performance, but they are located in the second and third quartiles regarding quality of supply. A lower level of the cost per kilometer should be balanced with a suitable performance in terms of quality of supply. For instance, Greece presents a good performance for the cost but, it is in the second quartile for the quality of performance. Regarding environmental performance (Fig. 3), Austria, Portugal, Denmark and Sweden are the countries located in the top percentile and can be

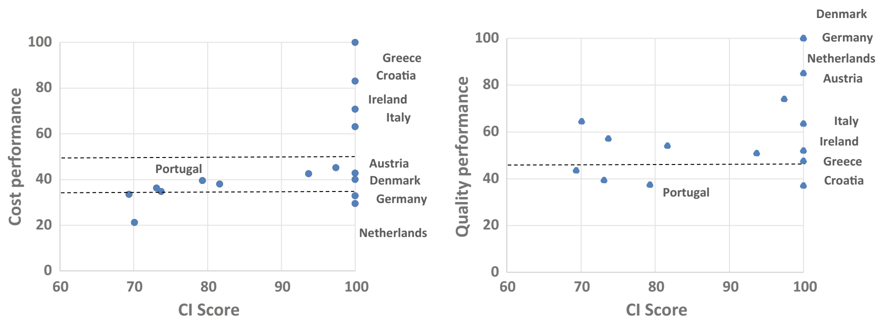
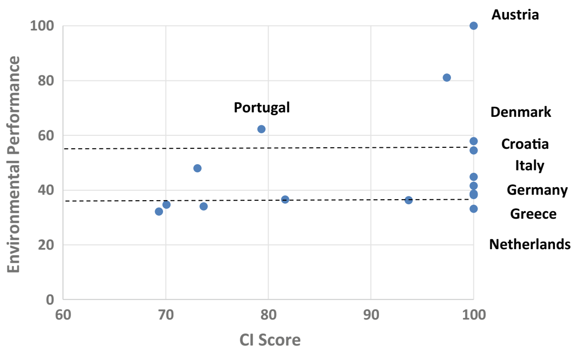


Fig. 2 Cost performance versus overall CI score (left) and quality performance versus overall CI score (right)

Fig. 3 Environmental performance versus overall composite indicator score



considered the best performers. Austria and Sweden have important water resources to produce electricity through hydro power plants. An overview of the energy system in Austria [34] shows that about 61% of the generation comes from hydro stations, 8% from wind and 2% from solar energy. Sweden has a similar situation to Austria, with 40% and 10% of renewable energy coming from hydro and wind [35], respectively.

The Portuguese and Danish cases are different. Since the end of 80s, Denmark was one of the leaders in terms renewables sources, installing wind parks onshore and offshore. About 43% of the generation comes from wind parks and 18% uses biofuels sources and waste stations. Due to the latest oil prices crisis in 2007 and 2008, Portugal launched an ambitious program to increase the usage of endogenous resources, mainly wind. As consequence, over the last decade Portugal became one of the main European countries where a major part of the generated electricity comes from renewables. Hydro power plants and wind parks represent about 50% of the generation from sources without polluting emissions. The inefficient countries in terms of environmental performance need to increase the share of renewable energy in their generation portfolio. Obviously, to act on this vector, the availability of the energy resource in the territory is crucial. For instance, the exposition to the solar radiation or the existence of windy days to produce energy with acceptable costs are key factors to ensure the generation through renewables sources. Meanwhile, the adoption of measures to improve the energy efficiency is also important for ensuring a better usage of the resources. Besides, the increasing share of renewables brings new challenges for the operation of the electric grid. The combination of several sources of renewable energy (solar, wind, hydro) needs storage capacity to store energy in the periods in which the demand is smaller. For instance, during the sunlight hours, the demand is smaller, and the surplus of solar generation should be stored to be used during the peak hours of the load diagram. An increasing share of renewables on the generation portfolio should not conduct to negative effects like increase of energy costs due to incentives to support renewable energy or to contract ancillary services to handle with the variability and uncertainty of the renewable sources.

6 Conclusions

In this paper, a single model is presented to assess the performance of European Power Systems of 16 European countries. Traditionally, the DEA technique has been applied to compare different companies at national or region level for regulatory purposes. In this work, similar models and principles to those applied in regulatory settings were adopted to compare the performance of electric power systems. The data available to perform this benchmarking analysis is scarce and spread across different sources, making it difficult to collect the information required for a cross-country comparison. Our study used the most recent year with data available for the majority of the indicators (year 2014), and complemented this analysis with data on quality of supply for the year 2013. We started in an overall analysis of the performance of the electric system, using a composite indicator aggregating cost, quality and environmental aspects. This was followed by a detailed analysis of the different dimensions of performance, focusing on the three sustainability pillars of the Triple Bottom Line approach [5, 6]. From the results of the DEA model, one concludes that the performance of a country can vary in the three dimensions. Austria, Croatia, Denmark, Germany, Greece, Ireland, Italy and Netherlands are the best practices according to the model presented in this paper. From the results, losses and quality of service indicators are the main vectors to be addressed in some countries like Czech Republic, France, Hungry, Poland, Portugal, Spain and United Kingdom. Therefore, this type of benchmarking analysis is crucial to provide information on the vectors that each country should act to improve its efficiency, taking into account its performance in the other dimensions. Regulatory Agencies and decisions makers can use this methodology as tool to prioritize actions leading to enhanced energy systems. An important highlight from this work is the need to standardize the practices in the European states regarding the computation of losses, quality supply indicators, and estimation of the energy prices. Data on these topics should be organized in public repositories, to ensure the transparency of the information and the availability of data for benchmarking purposes at a European level. Further developments of this research can consider a wider range of indicators to construct new models with a more comprehensive view of the electric systems performance. For instance, in next upgrades of this approach, other components of the final tariffs can be incorporated to analyze the cost performance.

References

1. European Commission, Market Observatory for Energy, Quarterly Report on Electricity Markets (2017)
2. Charun, R., Morande, F., Raineri, B.R.: (De)regulation and competition: the electric industry in Chile, 1st edn. in English. ILADES, Santiago, Chile (1997)
3. Saraiva, J.T.: Estimation of the congestion cost of the Portuguese national transmission network evolution from 1998 to 2008. In: 2012 9th International Conference on the European Energy Market, pp. 1–7, Florence (2012)
4. Saraiva, J.T., da Silva, J.P., Ponce de Leo, M.T.: Evaluation of the marginal based remuneration - a case study using the Portuguese transmission network. In: 2001 IEEE Power Tech Proceedings, pp. 6, Porto (2001)
5. Elkington, J.: Cannibals with Forks. The Triple Bottom Line of 21st Century Business, 1st edn. Capstone Publishing Limited, Oxford (1997)
6. Elkington, J.: Towards the sustainable corporation: win-win-win business strategies for sustainable development. Calif. Manag. Rev. **36**(2), 90–101 (1994)
7. UNUnited Nations, Report of the World Commission on Environment and Development: Our Common Future. Technical Report 1, United Nations, New York (1987)
8. Farsi, M., Filippini, M.: Regulation and measuring cost-efficiency with panel data models: application to electricity distribution utilities. Rev. Ind. Organ. **25**, 119 (2004)
9. Rudnick, H., Donoso, J.A.: Integration of price cap and yardstick competition schemes in electrical distribution regulation. IEEE Trans. Power Syst. **15**, 1428–1433 (2000)
10. Chengcheng, L., Yanling, W., Ju, G.: Research on the regulation of power transmission and distribution based on DEA, In: 2008 IEEE International Conference on Automation and Logistics, pp. 141–146, Qingdao (2008)
11. Simab, M., Alvehag, K., Soder, L., Haghifam, M.R.: Designing reward and penalty scheme in performance based regulation for electric distribution companies. IET Gener. Transm. Distrib. **6**(9), 893–901 (2012)
12. Cheng, X., BjØrndal, E., BjØrndal, M.: Cost efficiency analysis based on the DEA and StoNED models: case of Norwegian electricity distribution companies. In: 11th International Conference on the European Energy Market (EEM14), pp. 1–6, Krakow (2014)
13. Tascheret, C., Rattá, G., Andreoni, A.M.: Methodology to determine the optimal electricity distribution tariff using benchmarking techniques. In: 13th International Conference on the European Energy Market (EEM), pp. 1–5, Porto (2016)
14. Simab, M., Haghifam, M.R.: DEA efficiency for the benchmarking of Iranian electric distribution utilities. In: 20th International Conference and Exhibition on Electricity Distribution - Part 1, CIRED 2009, pp. 1–4, Prague (2009)
15. Roman, J., Gomez, T., Mofioz, A., Peco, J.: Regulation of distribution network business. IEEE Trans. Power Deliv. **14**(2), 662–669 (1999)
16. Recordon, E., Rudnick, H.: Distribution access pricing: application of OFTEL rule to a yardstick competition scheme. IEEE Trans. Power Syst. **17**(4), 1001–1007 (2002)
17. Avalos-Gonzalez, J.A., Rico-Melgoza, J.J., Madrigal, M., Madrigal, M.: Total quality management indicators and DEA for benchmarking the Mexican electrical industry. In: 2006 IEEE International Engineering Management Conference, pp. 388–392, Bahia (2006)
18. Cambini, C., Fumagalli, E., Croce, A.: Output-based incentive regulation: benchmarking with quality of supply in electricity distribution. In: 2012 9th International Conference on the European Energy Market, pp. 1–8, Florence (2012)

19. Santos, S.P., Amado, C.A., Rosado, J.R.: Formative evaluation of electricity distribution utilities using data envelopment analysis. J. Oper. Res. Soc. **62**(7), 1298–1319 (2011)
20. Agrell, P.J., Bogetoft, P., Grammeltvedt, T.E.: The efficiency of the regulation for horizontal mergers among electricity distribution operators in Norway. In: 12th International Conference on the European Energy Market (EEM), pp. 1–5, Lisbon (2015)
21. Agrell, P.J., Bogetoft, P., Tind, J.: DEA and dynamic yardstick competition in scandinavian electricity distribution. J. Product. Anal. **23**, 173201 (2005)
22. CEER, Benchmarking Report 5.2 on the Continuity of Electricity Supply, February 2015. www.ceer.eu
23. ENTSOE, Power Statistics. www.entsoe.eu/data/statistics/
24. Electric power transmission and distribution losses (% output) (2017). https://data.worldbank.org/
25. European Commission, EUROSTAT, Electricity and heat statistics 2014 (2014). http://ec.europa.eu/eurostat
26. IEEE Guide for Electric Power Distribution Reliability Indices - Redline. IEEE Std 1366-2012 (Revision of IEEE Std 1366-2003) - Redline, pp. 1–92 (2012)
27. European Commission, Energy Prices and costs in Europe, Brussels (2014)
28. Koopmans, T.: Analysis of production as an efficient combination of activities. In: Activity Analysis of Production and Allocation, pp. 33–97. Wiley, New York (1951)
29. Cook, W.D., Kress, M.: A data envelopment model for aggregating prefence rankings. Manag. Sci. **36**(11), 13021310 (1990)
30. Lovell, C.A.K.: Measuring the macroeconomic performance of the taiwanese economy. Int. J. Prod. Econ. **39**(1–2), 165–178 (1995)
31. Lovell, C.A.K., Pastor, J.T., Turner, J.A.: Measuring macroeconomic performance in the OECD: a comparison of european and noneuropean countries. Eur. J. Oper. Res. **87**(3), 507–518 (1995)
32. Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T.: An introduction to benefitt of the doubt composite indicators. Soc. Indic. Res. **82**(1), 111–145 (2007)
33. CEER- Council of European Energy Regulators, CEER Report on Power Losses (2017)
34. IEA-International Energy Agency, Austria-Energy System Overview (2017). https://www.iea.org/media/countries/Austria.pdf
35. IEA-International Energy Agency, Sweden-Energy System Overview (2017). https://www.iea.org/media/countries/Sweden.pdf
36. PORDATA, Electricity prices for households and industrial users (2017). www.pordata.pt

The Demand for Healthcare Services and Resources: Patterns, Trends and Challenges in Healthcare Delivery

Sofia Cruz-Gomes, Mário Amorim-Lopes and Bernardo Almada-Lobo

Abstract Together with the significant improvement in health and longevity came a number of health and economic concerns related to the demand for healthcare services and resources: changes in the patterns of health and illness, increasing amount and complexity of healthcare services demanded, rising health expenditures and uncertainty about whether there will be enough human, physical and financial resources to deliver the healthcare services needed. This paper aims to draw attention to the importance of planning the demand for healthcare in the aforementioned context, to create awareness of the need for a comprehensive study on the demand for healthcare services and resources and to propose an integrated approach for planning them, to inform managers and policy-makers on what can be the main challenges on assuring healthcare delivery in the future.

Keywords Healthcare · Demand · Integrated framework · Planning

1 Introduction

The improved living conditions and the advances in medical science that occurred in the past decades led to significant improvements in health and in longevity, which is now mainly due to the declining mortality among older people [1]. However, together with this achievement came a number of changes that are raising global health and economic concerns. Healthcare systems around the world are now facing several challenges related to the demand for healthcare services and resources. Changes in

S. Cruz-Gomes (✉) · B. Almada-Lobo
INESC TEC, Faculty of Engineering, University of Porto, Porto, Portugal
e-mail: scruzgomes@fe.up.pt

B. Almada-Lobo
e-mail: almada.lobo@fe.up.pt

M. Amorim-Lopes
INESC TEC, Faculty of Engineering, University of Porto and Católica Porto Business School, Porto, Portugal
e-mail: mario.lopes@fe.up.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_7

patterns of health and illness among individuals, increasing amount and complexity of healthcare services demanded by the populations, rising health expenditures over time and uncertainty about whether there will be enough resources to deliver the demanded services are some of the main current matters calling the attention of health researchers, managers and governments.

Health is one of the highest priorities for people all around the world and promotion and protection of health are crucial to both human welfare and sustained socioeconomic development [2]. Thus, a better understanding of these topics—allowing an accurate planning that can help to ensure that health systems can continue to provide quality and timely healthcare services and to contribute to healthy populations in a sustainable way—is of major worth.

This paper aims to highlight the importance of planning the demand for healthcare in the challenging current health and economic environment, to create awareness of the need for a comprehensive study on the demand for healthcare services and resources and to propose an integrated approach for planning them, to inform both managers and policy-makers on what can be the main challenges on assuring healthcare delivery in the future. The remainder of the paper is organized as follows. The next section presents and describes the problem under consideration. In Sect. 3, the scientific literature on the different problem dimensions is briefly revised and some research opportunities on assessing and planning for the future needs of healthcare services and resources are pointed out. Section 4 proposes an integrated approach to address the identified gaps and the main contributions that may result from such work. The last section is devoted to conclusions and final remarks.

2 The Problem

Overall, the problem can be summarized as the need to ensure that the necessary healthcare services will be provided to patients who need them. However, ensuring that future demand will be met depends on planning the different dimensions of the problem: (1) the healthcare services that will be demanded by the population; (2) the health human resources (HHR) that will be required to deliver these services; (3) the financial resources that will have to be available for all the inputs that are needed to produce them; and (4) the quality and timely delivery of healthcare to all who need them. Thus, for a better understanding of the problem of planning the demand and delivery of healthcare services, we decompose it in four different blocks, as represented in Fig. 1. Each block represents one of the aforementioned dimensions, which are logically and sequentially connected as the figure shows.

The first block represents the sub-problem of planning the demand for healthcare services. There are many factors driving this demand, which have been hardly studied by health and economic researchers, both theoretically and empirically: socio-economic, epidemiological and health-related, demographic and individual characteristics and behaviors. Nowadays, in a context of rising co-existence of multiple health conditions, it is recognized that epidemiological factors are extremely

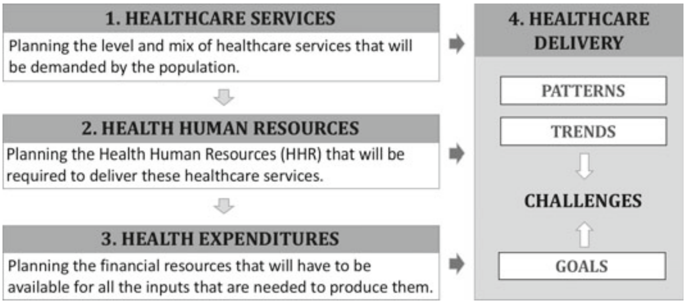


Fig. 1 Problem framing

relevant on driving the decision of seek for care. Efforts should be made to understand the patterns of the demand according to these factors, specially focusing on the impact of the evolution of multiple diseases co-existence over the time and across people from different contexts and with different individual characteristics. Understanding these complex and dynamic interrelations is of most interest to produce accurate forecasts on the demand for healthcare evolution, which are crucial to several purposes, such as planning for the capacity needed, inform on regional disparities and serve as basis for the sub-problem represented on block 2, where knowing the outputs to be produced is imperative to determine the inputs that may be needed.

The second block concerns to the HHR demand planning sub-problem. Here, the main issue is to translate the healthcare services demanded in HHR requirements. There are many interesting insights arising from the study of this topic. In addition to the more obviously useful information on the future demand for HHR, so that an adequate planning avoids future imbalances and their consequences, a deeper knowledge on the effect of the productivity in the delivery of healthcare over time and on several possible allocations of the existent HHR between different healthcare services are some of the useful information that can be used by health managers on pursuing their goals. Additionally, since the human resources bill is usually the largest single item in the budget for health, the requirements of HHR have a preponderant influence in health expenditures, to which the following block is dedicated.

In the third block, the sub-problem of forecasting health expenditures is represented. Ultimately, the healthcare services delivery is dependent on the availability of the financial resources needed to provide all the necessary inputs to produce these services—physical and human. The expenditures depend on a range of factors, whose relevance should be assessed and integrated in forecast models. Planning for the future health expenditures is critical at government level, as they are ultimately responsible for ensuring the continuity and sustainability of healthcare delivery.

Finally, the fourth block represents the sub-problem of dealing with the challenges in healthcare delivery. From the three previous blocks, insights on what can be these main challenges arise from the analysis of the main trends in health and in healthcare, and from the study on the demand for healthcare, on the subsequent HHR requirements and on the financial sustainability of healthcare. In a context of growing incidence of chronic diseases and existence of multiple morbidities, patients have changing needs and the volume and complexity of healthcare is rising, as well as health expenditures. A better understanding of the main trends and challenges should culminate in the study and evaluation of different modes of delivering care, ensuring that access, quality and sustainability of healthcare delivery are safe.

3 Literature Review

This section is divided in three distinct sub-sections, each one of them directly related to one of the sub-problems 1–3 previously described.

3.1 Healthcare Services

In his seminal work, Grossman [3] presents a health production model establishing the relation between the demand for health and for healthcare, arguing that the demand for healthcare is a derived demand, as it derives from the demand for 'good health'. By recognizing health as a capital good and that individuals have an initial stock of health that depreciates with time and that can be enlarged by investment, he synthesized that when purchasing healthcare services, consumers are not demanding these services per se: rather, they want to produce 'good health' (the commodity) using healthcare services (the market goods) and their own time as inputs. This insight revolutionized the way economists studied health related topics: it turns the attention to health determinants, marking the begining of health economics as a sub-discipline by replacing the healthcare economics that prevailed until then, where the demand for healthcare services was studied as being a regular demand for a final consumption good [4].

Since then, the demand for healthcare is recognized as a multidimensional and complex demand, and studies analyzing the determinants of healthcare demand started to include both types of factors: the ones that are known to drive the demand for consumption goods, such as the income, the prices and the individual preferences, and others related to the demand for the investment good: health. It is now recognized that the demand for healthcare ultimately depends on the seek for health and that addressing epidemiological factors and its evolution is of most importance for studying the problem of planning the demand for healthcare services.

In the last decades, the improved socio-economic conditions and the advances in medical science led to significant improvements in health and longevity [1], responsible for a significant demographic transformation and a subsequent drastic change in the epidemiological profiles and in the patterns of health and illness. Chronic diseases are now the world's leading causes of death, and their burden is increasing rapidly [5, 6]. As the prevalence of chronic conditions continues to rise [7], several attempts to understand the main causes and consequences of these epidemiological changes have been made, which are reflected in a vast scientific literature on the topic. It is now known the prevalence of the most common chronic diseases and the main factors on which it depends. It has also been showed that increases in the prevalence of chronic diseases rise the demand for healthcare and generate new challenges on ensuring the delivery and financing of healthcare services [8]. Furthermore, it is known that chronic conditions tend to cluster, as people with one chronic condition are more likely to also have others [9, 10]. This fact drove the researchers' attention to the study of the co-existence of multiple morbidities, which got a significant attention in last years.

Due to the efforts that have been made on getting a better understanding of the main determinants of the co-existence of multiple morbidities, we now know that age and the co-occurrence of multiple morbidities are significant and positively related [11, 12], which helps to explain the recent rise of multiple morbidities. Studies also show that prevalence of multiple morbidities is higher in women than in men [13, 14], maybe because women are generally more affected by non-fatal conditions [15], and that socio-economic conditions are inversely associated with multiple morbidities [16]. In this context, interesting studies on the effect of socio-economic conditions were undertaken considering a space component. Overall, they conclude that multiple chronic conditions occur earlier and more frequently in deprived areas, a conclusion that can be very helpful to plan for the resources needed in each location, according to their different healthcare needs. Multiple morbidity is known to be associated with a decline in health outcomes and a subsequent increase in the utilization of healthcare services and resources [17], as people with multiple conditions tend to have a higher vulnerability to diseases and less resistance to health threats [13]. In fact, it has been shown that patients with multiple chronic diseases demand for a greater array of healthcare services and professionals [18] and that they are more likely to seek care (and do it more often) than patients without multiple chronic diseases. This fact was observed for all different types of healthcare services: primary care, inpatient care, ambulatory care and emergency care [19].

Although both the prevalence of multiple morbidity and the impact of different health conditions on healthcare demand have been widely studied, the complex interaction between multiple co-existent diseases, demographic and socioeconomic factors and its impact on the demand for healthcare services is less understood [20]. As an higher prevalence of multiple conditions increases the demand for healthcare services and heightens its complexity [9], a deeper analysis of the interacting influences that lead to the complex pattern in the use of health services, addressing the evolution on morbidity patterns and planning for the needs of the different types of healthcare services (by location, service and specialty), are of most importance for clinicians, researchers, managers and policy-makers, to improve healthcare delivery.

The methods used to understand and predict the epidemiological evolution evolved from simple state-transition and statistical methodologies, that analyze and forecast the incidence and prevalence of specific diseases, to more complex epidemiological models, representing the epidemiology of disease occurrence and its association with a range of related factors and processes [21]. Due to the complex nature of the interaction between epidemiology and several individual, social and demand-related factors, computational models and simulations, such as system dynamics, discrete-event simulation, network analysis, and agent-based modeling, are now seen as central research tools in epidemiology [22].

Within these approaches, agent-based modeling is recognized as a promising approach to model the complex interactions and processes related to health conditions. Although this type of models has been applied for single chronic diseases [23], it is still underused among researchers and a broader use of agent-based modeling to provide insights on population health and consequent demand for healthcare services and resources is missing [24].

3.2 Health Human Resources

Since human resources account is usually the largest single item in the healthcare budget, and that no health system can deliver healthcare services without them, health human resources are widely recognized as the most important input of the health system [25]. Although health workforce planning is not new, the attention and resources allocated to workforce planning have increased in recent years [26], as imbalances in the healthcare workforce are becoming a major concern in both developed and developing countries [27].

HHR planning comprises the study and analysis of four key elements: supply, demand, a potential gap, and possible solutions to solve imbalances [28, 29]. Studies focused on the supply-side aim to forecast the future HHR, analyzing factors that influence the movement of professionals into, through and out of the healthcare workforce, as well as their motivation and organization [30]. On the other hand, studies approaching the demand-side aim to predict the HHR that will be needed, or demanded, in the future. These studies focus on the evolution of factors driving the demand for healthcare services and on the estimation of HHR required to efficiently deliver these services. Perhaps due to the relative simplicity of the data required to address the supply-side of the problem, considerable attention has been placed on supply approaches, with fewer endeavors on the demand-side [31].

On the demand-side, four main approaches to project future health worker requirements can be identified in the literature, which differ in the way the required healthcare services are identified: health needs, services utilization/demand, service targets and workforce-to-population-ratio [26, 32–35]. From these approaches, only the latter directly estimate the healthcare workforce requirements; the other three require the estimation of HHR requirements as a step succeeding the estimation of future healthcare services [33].

Several methods of converting services into workforce requirements can be found in the literature. The most commonly used consists of assessing the time required to complete tasks, measured by direct observation (time-motion studies, activity sampling techniques and patient flow analysis) or by expert's opinion [36]. This method is not very demanding in terms of data and it is the only one involving the healthcare providers. However, defining the necessary tasks can be difficult and time-consuming, and there is the risk of over-estimating the HHR requirements [35]. Another method, based on productivity patterns, consists in applying labor productivity measures to the expected needs for healthcare services. This approach can be based on the maintenance of productivity levels or on productivity benchmarks to be achieved [37, 38].

Involving theoretical economic fundamentals in HHR planning has also become popular. Economics, which is ultimately about studying the allocation of limited resources to unlimited necessities, can bring useful insights to the problem, namely on the opportunity cost of human resources allocation between alternative productive processes [39]. In this context, the estimation of production functions relating healthcare inputs (HHR and capital in health facilities) to healthcare outputs (delivered services) has become a widely used method in HHR planning [40]. Although the economic concepts and assumptions underlying the use of production functions are suitable to the problem [41], the complexity of handling production functions with multiple outputs contributed to a rising interest in the creation of indexes combining outputs [42], and in nonparametric approaches for the estimation of production frontiers, such as Data Envelopment Analysis [43, 44]. An inverse production function can also be used to model the production of healthcare services [45]. This function, commonly known as Input Requirements Function (IRF), is another way of overcoming the limitation of handling multiple outputs [46], keeping the key economical concepts of the production functions. Using IRF to understand the demand for labour has been a very popular approach in several sectors, including banking, insurance, manufacturing, electricity, railways and agriculture [47, 48]. However, despite the strengths and adequacy of using IRF and besides its popularity in other fields, the use of IRF in the healthcare field is almost inexistent.

3.3 Health Expenditures

Within the literature on health expenditures (HE) two different types of studies can be found: the ones that want to provide a better understanding on the determinants of health expenditure and the ones that intend to forecast health expenditures [49]. The first type of studies urged from the interest on understanding the main factors causing the rapid increase of HE in many developed countries by 1960 [50]. More recently, the concerns on the financial sustainability of the upward trend in national health expenditures in recent decades boosted the development of forecasting models to project HE evolution and inform on its sustainability [51]. Overall, most of the health expenditure literature still only tries to understand past drivers of HE, rather than project health expenditures into the future.

The models to forecast health expenditures can be classified according to the aggregation level of the unit of analysis. Hollenbeck [52] distinguished between disaggregate models, where a micro unit such as individuals or households is considered, and aggregate models, which are dedicated to the analysis of a macro unit, such as a cohort of individuals or a whole country or region. Later, Astofi et al.

[53] refined this classification identifying three different classes of forecast models: micro, component-based, and macro. Although the choice between the available models must always be driven by the policy questions that need to be answered, some authors believe that the evolution in both the data availability and the computing power will lead the forecasts of HE to complicated micro-level models [53]. However, Zhao [49] showed that when the main goal is to accurately forecast aggregate health expenditure, these models require considerably more data and effort, and projections may be worse than the ones obtained with macro-level models.

In the past decades, several macro studies have tried to address the determinants of HE. The drivers of HE have been grouped according to their nature in several categories: macroeconomic, organizational, technological, demographic, lifestyle and environmental. More succinctly, some authors [54] distinguished the factors influencing the future trajectory of HE as demand factors, such as aging, health status and income; and supply factors, such as technological progress, productivity and health prices.

Within the wide range of HE determinants, income was the first to be related to health expenditures [55] and it is still considered the most important factor in explaining differences in the level and growth of HE [56]. In his seminal work, Newhouse [57] found the GDP to be statistically significant on explaining HE, a fact that is nowadays generally accepted, after several other authors confirmed the existence of a positive correlation between variations in national income and variations in HE. These findings show that national HE are highly related to a country's state of economic development, which according to Wagner's law [58] states that government spending increases as the national income increases, mainly due to an increase in demand for public services. Population aging is also considered as a major determinant of HE [59] and another of the most studied factors. The literature on this topic, however, is not unanimous. While some authors argue that HE largely depend on the size and structure of the population and tend to rise due to an increase in life expectancy and consequent decline of the health status of the population, expansion of morbidity and burden of healthcare [60], others give preference to the argument that aging, per se, is not that relevant: HE are considerably higher in the period preceding death and increases on the life expectancy only postpone the expenditures, rather than raise them [54]. These two opposite theoretical standpoints result in different factors considered in empirical analysis: while the first leads to the study of demographic indicators, such as the share of elderly population or the life expectancy, the latter includes pure health-related variables translating the needs for care, such as health-status or morbidity [61].

The supply factors have also been pointed as a relevant driver of HE. The inclusion of factors such as the number of physicians or the number of beds can be explained by the direct dependency between HE and physical and human resources' levels of a health system. The number of physicians is the most considered explanatory variable in this class for three main reasons: first, because healthcare is a labor-intensive industry, where the human resources are more representative of health expenditures than the physical resources [62]; second, the number of physicians has been the indicator selected to capture the supply-induced-demand effect, as it is known that supply and demand for healthcare are not independent from each other and that asymmetric information exists between physicians and patients [63]; and third, because from a higher access to healthcare usually comes a higher utilization, specially in the presence of unmet needs [64]. Overall, these studies found a positive and statistically significant relation between the number of doctors and the aggregate HE. Also from the supply-side, technological progress has also been cited as a major driver of HE. The pioneer work of Newhouse [56], considering the relevance of technological change on health expenditures, has been extended by several other authors [65–67] studying the impact of technological advances on the evolution of the HE. Many proxies for changes in health technology have been used, such as surgical procedures performed, life expectancy [65], pharmaceutical spending [68] or R&D expenditures [66]. In most of the time-series studies, a time trend [69] or a time effect [67] are frequently used as proxies for the technological progress. Overall, these studies confirmed that technological progress is, in fact, a major determinant of health expenditures.

Several other factors can be found in the literature. The vast list of health expenditures determinants studied so far includes many other less studied drivers and variables with lower demonstrated relevance, such as health prices, insurance coverages, total or young population, education expenditures, life-style factors, utilization indicators and indicators related to the health system itself, as the share of public health expenditure or the out-of-pocket payments. However, and despite the number of studies reviewing the determinants of HE and trying to anticipate its evolution, the factors driving HE remain only imperfectly understood and empirical explanatory forecasts applying time-series approaches are almost inexistent.

4 Research Opportunities and the Proposed Approach

This section presents some of the gaps identified in the literature and proposes research directions both to assess and plan the future needs for healthcare services and resources and to provide insights on the patterns and trends of healthcare delivery, helping the several health stakeholders facing the challenges on healthcare delivery. Despite the undeniable interest of the field and the growing number of studies devoted to such issues, the problem is not closed, as there are no established preferred methodologies to approach it.

4.1 Healthcare Services

Despite the rising interest on understanding the complexity inherent to the co-existence of multiple health conditions and its impact on the demand for healthcare, there is clearly space to contribute to the knowledge on this topic. First, the analysis of multiple co-existent conditions is usually made considering only the chronic conditions, and a better understanding on the interaction between chronic and non-chronic disease is also of major interest. Second, most of the studies intended to provide knowledge for a specific disease (or group of diseases) and its association with other conditions that are previously known to be related, which can hide less obvious relations between conditions. Third, studies dedicated to the analysis of co-existent diseases and their evolution over time consider a simple count of conditions or, more succinctly, if the individual has or not two or more chronic diseases at the same time. Forth, studies analyzing longitudinal changes in morbidity over time and through the life course are limited, as well as analysis considering cohorts effects on the co-existence of health conditions. Finally, most of the research is only focused on the analysis of historical data and does not go further in assessing for the future healthcare services demanded to contribute to a more informed planning.

The literature reviewed points to several topics worthy of investigation, where it stands out the lack of a wider approach, capable of dealing with the complexity of the problem as a whole.

An agent-based simulation model could be developed to detect the complex patterns on the evolution of multiple conditions through their association with the co-occurrence of other current or previous chronic and non-chronic conditions, as well as of other relevant factors, such as age, gender, and residence location. These patterns may then be used to simulate how illness and morbidity will evolve and the subsequent services utilization, using projections for some relevant variables (e.g., demographic) and considering specific 'what if' scenarios (e.g., change in the prevalence of a specific condition) and the results on the expected healthcare utilization may allow to infer whether the current physical capacity would be enough to provide the expected volume of healthcare services. Furthermore, as this simulation model may include both a space and a time dimension, it would also allow for a further assessment of eventual regional asymmetries, and to describe how illnesses evolve over time, age and cohorts.

4.2 Health Human Resources

Ensuring the delivery of healthcare services is also crucially dependent on planning for the HHR needed to properly deliver these services. The literature reviewed on this topic shows there is still room to contribute to the extension and diversity of the available methods to forecast the HHR demand. In fact, while there is a growing agreement that planning the demand for HHR should be based on healthcare needs, less accordance exists on the best way to translate needs in HHR requirements [28].

A new approach to empirically quantify the relation between healthcare services and workforce requirements, modeling the relation between services and human resources, may result in an interesting contribution to the challenging task of translating healthcare services to HHR requirements if conjoining four main aspects: (1) analyze healthcare services by medical specialty, which have the advantage of capturing some relevant specificities that otherwise pass unnoticed; (2) analyze labor productivity evolution and its impact on the healthcare services delivered over time; (3) consider a specification that assumes that labor can be planned and sized according to the needs; and (4) analyze HHR requirements for different types of healthcare services. This approach may rend in an efficient tool for the estimation of the HHR required and it may also contribute to a deeper knowledge of the healthcare delivery process by revealing possible options for HHR allocation and opportunity costs of labor. Furthermore, by accounting for the technological progress and for the productivity of health professionals, it may allow the analysis of the labor productivity evolution and its impact on the healthcare services delivered over time, which is critical to plan for the future needs of HHR.

To empirically quantify the relation between healthcare services and the HHR requirements in a multi-output context, a Labor Requirements Function may be proposed, relating the number of physicians with a set of specialty specific workload and capital variables. This methodology is based on the assumption that health decision makers do not control the demand for healthcare services, but they can size and adjust the level of HHR in response to a given expected demand. By using period fixed effects, this methodology allows to infer on the impact of labor productivity and technological progress on healthcare delivery. Furthermore, elasticities of mean labor use, marginal rates of technical substitutions between healthcare services and returns to scale could be analyzed to provide further insights on opportunity costs of labor and different possibilities for allocate the existing HHR. These insights could further be used in simulation models to study the effect of specific 'what-if' scenarios and infer on the impact of utilization or demographic changes on the demand for HHR.

4.3 Health Expenditures

The rising health expenditures experienced over the last decades have urged researchers to review their determinants and to try to anticipate the respective evolution. However, the factors driving health expenditures remain only imperfectly understood and empirical explanatory forecasts applying time-series approaches are almost inexistent.

A contribution to the field must compile both the existent knowledge on health expenditure drivers and the main insights on the adequacy of the models and approaches in a macro-level forecast of HE to inform governments and managers on their expected evolution and on the sustainability of healthcare delivery. An interesting contribution may rely on an explanatory forecast of HE associating the following main characteristics: (1) Macro-level analysis of HE at the single-country level to provide an aggregate view and identify national specificities and trends; (2) Multi-variable approach, considering the most relevant drivers of HE, both on the demand-side, represented by socio-economic and demographic variables; and on the supply-side, including the technology and the number of physicians; (3) Time series techniques to detect non-stationarity and long-run relationships between the variables; (4) Multi-equation time-series model estimation, exploring and describing both the long-run relationships and the short-run dynamics between the variables; (5)Forecast HE using the estimated model.

To infer on how much will the health expenditures reach in the short- and long-run future if no actions are taken, the use of a Vector Error Correction (VEC) model may be pursued: a multi equation time series econometric model relating HE and some other relevant macro indicators. This model would allow to account for both non stationarity of the data and possible existence of cointegration. Through this model, the main factors driving the changes in health expenditures, as well as both the long-run relations and the short run dynamics between health expenditures and its drivers may be investigated. Furthermore, it would be possible to use the estimated model to forecast health expenditures for the future, based on the projections available for the main determinants that are found to impact health expenditures.

4.4 Integrated Approach

To approach the research opportunities identified, we propose an integrated framework to the problem of planning the demand and delivery of healthcare services. This four-steps framework, represented in Fig. 2, aims to help ensuring that healthcare services will be provided to whom may need them. Each of the four steps is directly related to one of the four dimensions identified in the problem: (1) the healthcare

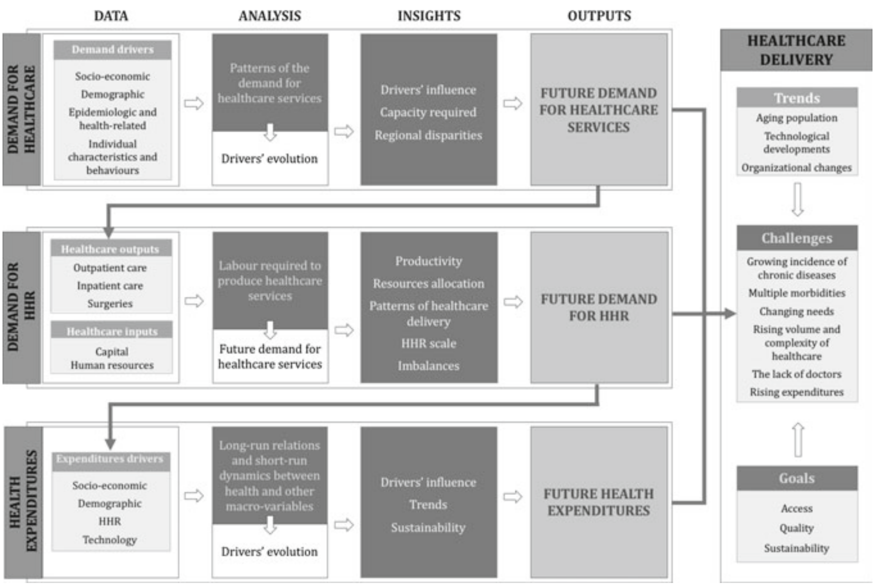


Fig. 2 Integrated approach

services that will be demanded; (2) the health human resources that will be required to deliver these services; (3) the financial resources that will have to be available for all the inputs that are needed to produce them; and (4) the quality and timely delivery of healthcare. For each of the three first steps, the data required, the core analysis suggested and the main insights and outputs resulting from the analysis are described in the figure. Furthermore, the connection between outputs of one step and data required to perform the step following are also represented: Step 2 should use the level and mix of healthcare services that are expected to be demanded in the future to estimate the HHR requirements, and Step 3 should consider these HHR requirements as a driver to forecast HE. The fourth step consists on the joint analysis and evaluation of the insights and outputs resulting from the three previous steps, on the main trends and patterns of healthcare delivery. This information should be used to infer about what can be the main challenges on the future of healthcare delivery, and to provide knowledge and tools to deal with them, enabling alternative modes of healthcare delivery and ensuring the continuity and sustainability of healthcare delivery.

5 Conclusion and Future Work

This paper draws attention to the current health and economic concerns related to the future demand and delivery of healthcare: changes in the patterns of health and illness, increasing amount and complexity of healthcare services demanded, rising health expenditures and uncertainty about whether there will be enough human resources to deliver the healthcare services. These issues have been calling the attention of health researchers, managers and governments and highlighting the importance of planning the demand for healthcare services and resources.

As we attempted to show, and despite the undeniable interest of the field and the growing number of studies devoted to these issues, the problem is not closed. By pointing out to some research opportunities we hope to stimulate future research focused both on providing a better understanding of the factors driving the demand for healthcare services and resources and on developing models to accurately forecast their expected evolution for the future: the healthcare services that will be demanded, the HHR that will be required to deliver these services and the financial resources that will be needed for all the inputs required to produce them.

Furthermore, we plan to apply the proposed four-step framework, making use of econometric and operational research methods, to assess and plan the future needs for healthcare services and resources. As so, we hope to contribute for this undeniably interesting field of research and to make a fruitful contribution to society by providing useful insights on the patterns and trends of healthcare delivery, which may help on the challenging task of planning for the future demand and delivery of healthcare services.

Acknowledgements This work is financed by the ERDF—European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation—COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT—Fundação para a Ciência e a Tecnologia within project POCI-01-0145-FEDER-016738.

References

1. World Health Organization: World Report on Ageing And Health. World Health Report, pp. 1–260 (2015)
2. World Health Organization. 2010 'Health Systems Financing - The path to universal coverage'. World Health Report, pp. 1–128 (2015)
3. Grossman, M.: On the concept of health capital and the demand for health. Polit. Econ. **80**(2), 223–255 (1972)
4. Linddren B.: Foreword to the 2017 Edition. In: Grossman, M. (ed.) The Demand for Health, pp. xi–xxvi. Columbia University Press (1972)
5. Alwan, A., et al.: Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. Lancet **376**(9755), 1861–1868 (2017)
6. Beaglehole, R., et al.: Priority actions for the non-communicable disease crisis. Lancet **377**(9775), 1438–1447 (2011)
7. Ward, B.W., Schiller, J.S., Goodman, R.A.: Multiple chronic conditions among US adults: a 2012 update. Prev. Chronic Dis. **17**, 11–E62 (2014)
8. Barcel, A., Aedo, C., Rajpathak, S., Robles, S.: The cost of diabetes in Latin America and the Caribbean. Bull. World Health Organ. **81**(1), 19–27 (2003)
9. Fortin, M., Bravo, G., Hudon, C., Vanasse, A., Lapointe, L.: Prevalence of multimorbidity among adults seen in family practice. Ann. Fam. Med. **3**(3), 223–228 (2005)
10. Wolff, J.L., Starfield, B., Anderson, G.: Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. Arch. Intern. Med. **162**(20), 2269 (2002)
11. Banerjee, S.: Multimorbidity: older adults need health care that can count past one. Lancet **385**(9968), 587–589 (2015)
12. Canizares, M., Hogg-Johnson, S., Gignac, M.A.M., Glazier, R.H., Badley, E.M.: Increasing trajectories of multimorbidity over time: birth cohort differences and the role of changes in obesity and income. J. Gerontol. Ser. B **45**, 387–416 (2017)
13. Glynn, L.G., et al.: The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. Fam. Pract. **28**(5), 516–523 (2011)
14. Smith, S.M., Ferede, A., O'Dowd, T.: Multimorbidity in younger deprived patients: an exploratory study of research and service implications in general practice. BMC Fam. Pract. **9**, 6 (2008)
15. Marengoni, A., et al.: Aging with multimorbidity: a systematic review of the literature. Ageing Res. Rev. **10**(4), 430–439 (2011)
16. Barnett, K., Mercer, S.W., Norbury, M., Watt, G., Wyke, S., Guthrie, B.: Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet **380**(9836), 37–43 (2017)
17. Caughey, G.E., Vitry, A.I., Gilbert, A.L., Roughead, E.E.: Prevalence of comorbidity of chronic diseases in Australia. BMC Public Health **8**(1), 221 (2008)
18. Vogeli, C., et al.: Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J. Gen. Intern. Med. **22**(3), 391–395 (2007)
19. Boyd, M., Fortin, C.M.: Future of multimorbidity research: how should understanding of multimorbidity inform health system design? Public Health Rev. **32**(2), 451–474 (2010)
20. Roos, N.P., Carrire, K., Friesen, D.: Factors influencing the frequency of visits by hypertensive patients to primary care physicians in Winnipeg. CMAJ **159**(7), 777–783 (1998)

21. Garner, M.G., Hamilton, S.A.: Principles of epidemiologic modelling. Rev. Sci. Tech. Off. Int. Epiz. **30**(2), 407–416
22. Habtemariam, T., Tameru, B., Nganwa, D., Beyene, G., Ayanwale, L., Robnett, V.: Epidemiologic modeling of HIV/AIDS: use of computational models to study the population dynamics of the disease to assess effective intervention strategies for Decision-making. Adv. Syst. Sci. Appl. **8**(1), 35–39
23. Nianogo, R.A., Arah, O.A.: Agent-based modeling of non-communicable diseases: a systematic review. Am. J. Public Health **105**(3), 20–31 (2015)
24. Li, Y., Lawley, M.A., Siscovick, D.S., Zhang, D., Pagn, J.A.: Agent-based modeling of chronic diseases: a narrative review and future research directions. Prev. Chronic Dis. **13** (2016)
25. World Health Organization: Health systems: improving performance. World Heal. Rep. **78**(1), 1–215 (2000)
26. Ono, T., Lafortune, G., Schoenstein, M.: Health workforce planning in OECD countries: a review of 26 projection models from 18 countries. OECD Heal. Work Pap. **62** (2013)
27. Zurn, P., Dal Poz, M.R., Stilwell, B., Adams, O.: Imbalance in the health workforce. Hum. Resour. Health **2**(13), 1–12 (2004)
28. Murphy, G.T., Birch, S., Mackenzie, A.: Needs-based health human resources planning: the challenge of linking needs to provider requirements (2007)
29. Al-Sawai, A., Al-Shishtawy, M.M.: Health workforce planning: an overview and suggested approach in Oman. Sultan Qaboos Univ. Med. J. **15**, 27–33 (2015)
30. World Health Organization: Models and tools for health workforce planning and projections. World Heal. Rep. **3**, 1–19 (2010)
31. O'Brien-Pallas, L., Baumann, A., Donner, G., Murphy, G.T., Lochhaas-Gerlach, J., Luba, M.: Forecasting models for human resources in health care. J. Adv. Nurs. **33**(1), 120–129 (2001)
32. Amorim-Lopes, M., Almeida, A.S., Almada-Lobo, B.: Handling healthcare workforce planning with care: where do we stand? Hum. Resour. Health **13**(1), 38 (2015)
33. Bärnighausen, T., Bloom, D.E.: Changing research perspectives on the global health workforce. NBER Work Pap. Ser. (2009)
34. Roberfroid, D., Leonard, C., Stordeur, S.: Physician supply forecast: better than peering in a crystal ball? Hum. Resour. Health **7**, 10 (2009)
35. Dreesch, N., et al.: An approach to estimating human resource requirements to achieve the millennium development goals. Health Policy Plan **20**(5), 267–276 (2005)
36. Lipscomb, J., Parmigiani, G., Hasselblad, V.: Combining expert judgment by hierarchical modeling: an application to physician staffing. Manag. Sci. **44**(2), 149–161 (1998)
37. Dall, T., West, T., Chakrabarti, R., Iacobucci, W.: The complexities of physician supply and demand: projections from 2013 to 2025 final report association of American Medical Colleges. Assoc. Am. Med. Coll. 1–68 (2015)
38. Hall, T.L.: Chile health manpower study: methods and problems. Int. J. Heal. Serv. **1**(2), 166–184 (1971)
39. Scott, R.D., Solomon, S.L., McGowan, J.E.: Applying economic principles to health care. Emerg. Infect. Dis. **7**(2), 282–285 (2001)
40. Pourmohammadi, K., Hatam, N., Bastani, P., Lotfi, F.: Estimating production function: a tool for hospital resource management. Shiraz E Med. J. **15**(4), 1–7 (2014)
41. Santías, F.R., Cadarso-Suárez, C., Rodrguez-Álvarez, M.X.: Estimating hospital production functions through flexible regression models. Math. Comput. Model **54**(7), 1760–1764 (2011)
42. Castelli, A., Street, A., Verzulli, R., Ward, P.: Examining variations in hospital productivity in the English NHS. Eur. J. Heal. Econ. **16**(3), 243–254 (2015)
43. Newhouse, J.P.: Frontier estimation: how useful a tool for health economics? J. Health Econ. **13**(3), 317–322 (1994)
44. Hollingsworth, B., Dawson, P.J., Maniadakis, N.: Efficiency measurement of health care: a review of non-parametric methods and applications. Health Care Manag. Sci. **2**(3), 161–172 (1999)
45. Diewert, W.E.: Funcional forms for revenue and factor requirements functions. Int. Econ. Rev. **15**(1) (1974)

46. Lipscomb, J., Kilpatrick, K.E., Lee, K.L., Pieper, K.S.: Determining VA physician requirements through empirically based models. Health Serv. Res. **29**(6), 697–717 (1995)
47. Heshmati, A.: Labour demand and efficiency in Swedish savings banks. Appl. Financ. Econ. **11**(4), 423–433 (2001). Kumbhakar, S.C.: Labour-use efficiency in Swedish social insurance offices. J. Appl. Econ. **10**, 33–47 (1995)
48. Kumbhakar, S.C., Zhang, R.: Labor-use efficiency and employment elasticity in Chinese manufacturing. Econ. e Polit. Ind. **40**(1), 5–24 (2013)
49. Zhao, J.: Forecasting health expenditures: methods and applications to international databases. CHEPA Work Pap. Ser. (2015)
50. Getzen, T.E.: Measuring and Forecasting Global Health Expenditures. Chapter Glob. Heal. Econ. Public Policy (2014)
51. Marino, A., James, C., Morgan, D., Lorenzoni, L.: Future trends in health care expenditure: a modelling framework for cross-country forecasts. OECD Heal. Work Pap. **5** (2017)
52. Hollenbeck, K.: A review of retirement income policy models. Upjohn Inst. Work Pap. **95**(38) (1995)
53. Astolfi, R., Lorenzoni, L., Oderkirk, J.: Informing policy makers about future health spending: a comparative analysis of forecasting methods in OECD countries. Health Policy **107**(1), 1–10 (2012)
54. La Maisonneuve, C.D., Martins, J.O.: Public spending on health and long-term care: a new set of projections. OECD Econ. Policy Pap. **6**(6), 1–39 (2013)
55. Newhouse, J.P.: Medical-care expenditure: a cross-national survey. J. Hum. Resour. **12**(1), 115–125 (1977)
56. Newhouse, J.P.: Medical care costs: jow much welfare loss? J. Econ. Perspect. **6**(3), 3–21 (1992)
57. Newhouse, J.P.: Forecasting demand for medical care for the purpose of planning health services (1974)
58. Musgrave, R.A., Jacob, A., Stuart, C., Barone, E.: Classics in the theory of public finance (1967)
59. Przywara, B.: Projecting future health care expenditure at European level: drivers, methodology and main results. Econ. Pap. **717**, 1–83 (2010)
60. Fogel, R.W.: Forecasting the cost of U.S. health care in 2040. J. Policy Model **31**(4), 482–488 (2009)
61. Schulz, E.: The influence of supply and demand factors on aggregate health care expenditure with a specific focus on age composition. ENEPRI Res. Rep. **16**, 45 (2005)
62. Mohammed, B.: Determinants for demand for health care services in Mekelle City. Ethiop. Health Reforms **22**(5), 102–110 (2013)
63. Gerdtham, U.G., Lothgren, M.: On stationarity and cointegration of international health expenditure and GDP. J. Health Econ. **19**(4), 461–475 (2000)
64. San, E., Mar, O., Peiming, W.: The effects of an ageing population and other factors on aggregate health care expenditure in Singapore, ICEB-15 (2005)
65. Dreger, C., Reimers, H.E.: Health care expenditures in OECD countries: a panel unit root and cointegration analysis. IZA Discuss Pap. Ser. **1469** (2005)
66. Okunade, A., Murthy, V.: Technology as a 'major drivers' of health care costs: a cointegration analysis of the newhouse conjecture. J. Health Econ. **21**, 147–159 (2002)
67. Di Matteo, L.: The macro determinants of health expenditure in the United States and Canada: assessing the impact of income, age distribution and time. Health Policy **71**(1), 23–42 (2005)
68. Atella, V., Bhattacharya, J., Carbonari, L.: Pharmaceutical price controls and minimum efficacy regulation: evidence from the United States and Italy. Health Serv. Res. **47**(1), 293–308 (2012)
69. Blomqvist, A.G., Carter, R.A.L.: Is health care really a luxury? J. Health Econ. **16**(2), 207–229 (1997)

Planning the Delivery of Home-Based Long-Term Care: A Mathematical Programming-Based Tool to Support Routes' Planning

Daniel Espadinha and Teresa Cardoso-Grilo

Abstract The adequate planning of home-based long-term care (HBLTC) is essential in the current European setting where long-term care (LTC) demand is increasing rapidly, and where home-based care represents a potential cost-saving alternative from traditional inpatient care. Particularly, this planning should involve proper route planning to ensure visits of health professionals to patients' homes. Nevertheless, literature in the specific area of HBLTC planning is still scarce. Accordingly, this paper proposes a tool based on a mathematical programming model—the LTC^{routes}—for supporting the daily planning of routes to visit LTC patients' homes in National Health Service-based countries. The model allows exploring the impact of considering different objectives relevant in this sector, including the minimization of costs and the maximization of service level. Patients' preferences, traffic conditions and budget constraints are also considered in the proposed model. To illustrate the applicability of the model, a case study based on the National Network of LTC in Portugal is analyzed.

Keywords Health care planning · Long-term care · Home-based care · Route planning · Mathematical programming

1 Introduction

Home Health Care (HHC) delivery is an important component of health care systems, having the potential to reduce costs of health care provision and free capacity in overcrowded acute care settings [14]. Nevertheless, this potential is found not only for general acute care services, but also in the Long-Term Care (LTC) sector.

D. Espadinha
Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal

T. Cardoso-Grilo (✉)
Centre for Management Studies of Instituto Superior Técnico (CEG-IST), Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal
e-mail: teresa.sofia.grilo@iscte-iul.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_8

LTC can be provided in both formal and informal settings [11]. In fact, although informal delivery of LTC by families traditionally represents a cornerstone of LTC delivery, the development and strengthening of formal Home-Based LTC (hereafter-called HBLTC) by health professionals currently represents a policy priority across European countries. This is not only due to the expected growing demand for LTC, but also to the decreasing reliance on informal care provided by families, due to the increase in female employment and migration of the younger members to urban centers [9, 19, 22].

Within this context, and considering the current pressures to reduce health care spending across European countries, such as is the case of Portugal, there is clearly the need to plan the adequate provision of HBLTC. In fact, when considering European countries where health care services are provided within a National Health Service (NHS), such as it is the case of Portugal [17], home-based care services poise an appealing cost shift and potential cost-saving alternative from traditional inpatient care.

This adequate planning implies, among other issues, the planning of routes for HBLTC provision. This planning is typically performed manually, resulting in high organizational efforts and potentially non-optimal routes—and this may result in higher costs incurred in the care delivery process, as well as on poor service levels.

According to [9], a review on HHC routing and scheduling models revealed that mathematical programming models represent a key method when the aim is to support the planning of routes, with the most frequent approach being extensions of the Vehicle Routing Problem (VRP). Still, when considering the particular case of route planning of HBLTC services, no related-study was found.

Within this context, this paper develops a general tool based on a mathematical programming model that can be used to support the planning of routes to patients' home in NHS-based countries, so as to respond to the growing demand of HBLTC—this model will be hereafter called LTC^{routes}. The particular case of home health care provided within the scope of the National Network of LTC (*Rede Nacional de Cuidados Continuados Integrados*, RNCCI) in Portugal is used as case study to illustrate the usefulness of the proposed approach. The LTC^{routes} aims at informing the practitioners on the optimal routes to visit their patients' homes on a daily basis. The model allows exploring the impact of considering different objectives relevant in this sector, such as the minimization of costs (measured in different ways, including travelling time and number of vehicles required to perform the routes) and the maximization of service level (measured by the percentage of patients visited). Patients' preferences, traffic conditions and budget constraints are also imposed to explore the impact on planning decisions.

This paper contributes to the literature in the area by:

i. Proposing a tool that can be used to support route planning decisions of HBLTC delivery on a daily basis;
ii. Exploring the impact of accounting for different objectives relevant in this sector, such as the minimization of the costs (measured in different ways) and the maximization of service level;

iii. Proposing a planning model that allows accounting for patients' preferences and traffic conditions;
iv. Proposing a generic approach that can be used to plan HHC delivery in general, and that can be applied to other regions of Portugal or other countries with a NHS.

This paper is organized in six sections. Section 2 presents a literature review on key studies in the area, followed by some background information in Sect. 3. Section 4 presents the mathematical details of the proposed model. Section 5 presents the case study and the results obtained for different objectives. Finally, Sect. 6 presents the conclusions and possible lines of further research.

2 Literature Review

A vast literature exists on HHC logistics management. Gutirrez and Vidal [10] provide a framework presenting three different dimensions that should be considered when planning HHC delivery: (i) the planning horizon, (ii) management decisions, and (iii) services processes. Regarding the planning horizon, and depending on whether one is dealing with strategic, tactical or operational decisions, the time horizon may be long (years), medium (months or weeks) or short (days), respectively. Concerning the second dimension, different groups of management decisions may need to be accounted for: network design, transportation management, staff management and inventory management. Finally, the third dimension describes the services processes at the medical, patient and support services levels.

Within this setting, this paper is focused on a specific challenge related to transportation management, namely, the HBLTC routing problem. Accordingly, a tool is proposed to support the planning of HBLTC routes on an operational level, thus providing information on how to plan the routes to patients' homes on a daily basis. To the authors knowledge, no study exists proposing planning models to solve the HBLTC routing problem, but several studies exist dealing with routing problems in the health care sector in general. For this reason, the focus of this review will be on studies proposing methods to solve health care routing problems, since these methods have potential to be applied for the particular case of the LTC sector.

Home Health Care as a Vehicle Routing Problem

The framework presented by [10] show how HHC planners face complex and challenging planning problems on different decision levels, including routing decisions. And according to [9], mathematical programming models represent the most widely used approach to deal with routing problems in HHC settings, with VRP playing a key role in this area, particularly, Home Health Care Vehicle Routing Problems (HHCVRP). HHCVRP formulations differ in several dimensions. Particularly, one can find (i) deterministic or stochastic models, and (ii) mono- or multi-objective models:

i. Since not all data is known in advance and uncertainty in demand is a reality faced by many HBLTC providers, stochastic models play a central role when planning in real-world situations (see, for instance, [5] for a comprehensive review on stochastic planning in general; and [18] for stochastic HHC planning). Still, most literature in the area is based on deterministic models that ignore the impact of uncertainty on planning decisions [9];
ii. Regarding the number of objectives, many existing studies opt to consider only one single objective (see, for instance, [1, 3]). Nevertheless, multiple and conflicting objectives may need to be pursued—e.g., [6] present one of the few multi-objective approaches in the area, by proposing a bi-objective model focused on the minimization of total costs and minimization of client inconvenience. And when multiple objectives are accounted for, different approaches may be followed [8]: (i) One can identify compromise solutions, the so-called Pareto optimal or non-dominated solutions; (ii) Alternatively, one can identify the best non-dominated solution through the aggregation of this multiplicity of objectives into an overall objective function using a set of weights.

This review is thus focused on deterministic models, since the aim of this paper is the proposal of a tool for supporting daily route planning of HBLTC services while ignoring the impact of uncertainty. Also, acknowledging the relevance of accounting for multiple objectives, this paper explores the impact of different objectives but through the use of a mono-objective model, i.e., different versions of the objective function are considered depending on the planning circumstances—in fact, a wide variety of objectives may need to be accounted for, as described below. And concerning the constraints that should be considered when the aim is to ensure an adequate planning of HHC routes, a set of key constraints may be worth to be considered, as also described in detail below.

Key Objectives in HHCVRP

A key objective found in VRP studies is usually the minimization of travelling costs (either in monetary terms or in distances travelled) (see, for instance, [20]). However, in HHC settings, rather than minimizing traveling distances, most studies propose the minimization of the travel time spent by caregivers traveling to patients' homes, or, alternatively, the travel cost associated with moving between patients' homes, since working times are often considered as the main cost factor [9]. According to [9] other objectives include (i) maximizing the preferences of patients (for a specific caregiver or visit time window), (ii) minimizing the number of nurses needed, (iii) maximizing the number of tasks performed, and (iv) maximizing fairness, i.e., balancing the workload amongst the staff. Still, although relevant in contexts of budgetary cuts in

Table 1 Key features analysed in home health care routing problems, and in the LTC sector in particular

References	Multiple objectives		Preferences	Traffic conditions	LTC
	Travel time/cost	Service level			
[3]	✓				
[20]	✓				
[6]	✓		✓		
LTC^{routes}	✓	✓	✓	✓	✓

health where it is not possible to fully satisfy health care demand [2], no study was found considering service level-related objectives.[1]

Table 1 summarize key studies and features that are relevant to be considered when developing planning models to deal with the HHC routing problem. It can be concluded that no study exists considering service level-related objectives (accounting for the reality that it is not always possible to fully satisfy health care demand) together with cost objectives, which is essential for countries facing budgetary cuts in health (such as happens in several European countries, including in Portugal). Also, traffic conditions are not typically considered in existing studies, and no study considers the specificities of the LTC sector. Furthermore, patients' preferences are especially relevant in HBLTC, representing a key aspect that should be considered in planning models in this sector. Accordingly, it can be concluded that there is space to develop research devoted to the development of planning models based on mathematical programming so as to support the route planning of HBLTC, and so the LTC^{routes} fills this gap in the literature. Up to the authors' knowledge, the model proposed by Braekers et al. [6] is the one more closely related to the model proposed in this study (the LTC^{routes} model), but it still does not account for service level-related objectives and traffic conditions, and it is also applied to an health care context other than LTC.

3 Problem Statement

The present paper proposes the LTC^{routes} model to support the planning of routes to patients' home within the HBLTC sector, with background information being presented in this section.

[1]Within the context of this study, service level related-objectives represent objectives related to the maximization of the number of patients that can be served within a limited budget—note that this concept differs from the one associated with patient inconvenience and preference focused in [13] or [6].

Planning Home-Based Long-Term Care (HBLTC) Delivery

Being focused on the planning of routes to ensure HBLTC delivery in NHS-based countries, this paper considers the Portuguese National Network of LTC (RNCCI) as reference. This network was created in 2006 so as to ensure the formal provision of LTC and ensures the delivery of institutional care (IC), ambulatory care (AC) and home-based care (HBC) [16, 19]. For the purpose of this paper, only the HBC component of the network (i.e., HBLTC) is considered for planning purposes. HBLTC comprises the delivery of both social and health care services by a multidisciplinary team of professionals, including physicians, nurses and physiotherapists, and its adequate planning involves making decisions on the optimal routes to be performed by these professionals so as to visit patients' homes on a daily basis.

Planning Objectives

Depending on the planning context, different objectives may need to be pursued. Particularly, policy statements and literature in the area suggest that a key policy objective in any NHS-based country include ensuring universal coverage of LTC demand [2, 9, 19]. Still, when considering the current European context of severe budget cuts, there are very high pressures to place cost considerations as a key priority when planning health care delivery. Since a key cost component associated with HBC delivery are travelling costs, minimizing these costs appears as a key policy objective in this area—and in particular, minimizing these travelling costs may imply minimizing different cost components, such as travelling time and the number of vehicles required to visit patients. Also, since it may not be possible to fully satisfy HBLTC demand, maximizing service level is also a key objective in the area.

Key Constraints

Depending on the planning context, some particular constraints may also need to be considered when planning the daily routes for practitioners in the field of HBLTC. Particularly, the preferences of patients for a specific time period for their visit and the impact of traffic conditions on different periods of the day are key constraints to be considered in this context, and may have a significant impact in planning decisions (as confirmed in Sect. 2).

4 Mathematical Formulation of the Model

The notation used for the model formulation is presented below, together with the mathematical formulation of the objectives and key constraints of the model.

4.1 Notation

Indices and Sets

$v \in V$	Vehicles
$j, i \in J = Jo \cup Jm$	Nodes representing primary health care centers (PHCC) ($jo, io \in Jo \subseteq J$) and municipalities ($jm, im \in Jm \subseteq J$), with municipalities representing patients' homes
$w \in W$	Periods of the day
$p \in P$	Patient type

Parameters

M_v	Maximum daily work time for each vehicle v, in minutes
N	Number of vehicles available
K_w	Duration, in hours, of each period of the day w
D_p	Duration, in minutes, of the visit to a patient belonging to type p
$N_{p,j}$	Total demand of patients belonging to type p on node j
N_{jm}	Total number of municipality nodes
$T_{i,j}$	Travel time, in minutes, from node i to node j
$TRAFFIC_w$	Impact of traffic on each period w of the day
$PREF_{jm,p,w}$	Preference of a patient belonging to type p in municipality jm to be visited during the period of the day w

Variables

$x_{i,j,v,w}$	Binary variable that equals 1 if vehicle v travels between nodes i and j during the period of the day w; and zero otherwise
k_v	Binary variable that equals 1 if vehicle v is used; and zero otherwise
y_v	Maximum time, in minutes, that a vehicle v can be used per day
$s_{i,j,v,p,w}$	Number of patients belonging to type p served in node j by vehicle v, with this vehicle traveling from node i, during the period of the day w
u_{jm}	Variable to be used for eliminating subtours involving each municipality jm, as suggested by [15]
z_1	Time traveled minimization variable
z_2	Vehicle minimization variable
z_3	Service level maximization variable.

4.2 Objetive Functions

Depending on the planning circumstances, the objectives to be considered may differ.

As noted before, minimizing costs play a key role in this sector, and such cost can be measured in monetary terms, distance or time traveled. In this study, a key objective to be considered is the minimization of traveling costs, with costs being measured in two different ways:

i. Minimizing costs in terms of time traveled (Eq. 1)—this will have as consequence an increase in the number of patients served per vehicle.

$$Min\ z_1 = \sum_{i \in I} \sum_{\substack{j \in J \\ j \neq i}} \sum_{v \in V} \sum_{w \in W} T_{i,j} \times TRAFFIC_w \times x_{i,j,v,w} \tag{1}$$

ii. Minimizing costs in monetary terms, by minimizing the number of vehicles required to visit patients' home (Eq. 2; with the number of vehicles being used as a proxy for investment costs)—this is important from a management and investment planning point-of-view, since it gives information on the resources needed to satisfy HBLTC demand.

$$Min\ z_2 = \sum_{v \in V} k_v \tag{2}$$

Furthermore, and still within the context of budgetary cuts in health, it may be relevant to maximize the demand to be satisfied with the available resources. This objective will be hereafter mentioned as service level-related objective (Eq. 3).

$$Max\ z_3 = \sum_{i \in I} \sum_{j \in J} \sum_{v \in V} \sum_{p \in P} \sum_{w \in W} s_{i,j,v,p,w}. \tag{3}$$

4.3 Constraints

A set of constraints is considered in the model, and are described below.

Routes-Related Constraints

Equation 4 ensures that a vehicle that enters a node j (either a municipality or a PHCC) must also leave from it, ensuring the continuity of the route.

$$\sum_{\substack{i \in I \\ i \neq j}} x_{i,j,v,w} - \sum_{\substack{i \in I \\ i \neq j}} x_{j,i,v,w} = 0 \quad \forall j \in J, v \in V, w \in W \tag{4}$$

Equation 5 eliminates the routes between the same node.

$$x_{j,j,v,w} = 0 \quad \forall j \in J, v \in V, w \in W \tag{5}$$

Equation 6 prevents subtours between municipalities, as suggested by [15].

$$u_{jm} - u_{im} + N_{jm} \times x_{jm,im,v,w} \leq N_{jm} - 1 \quad \forall jm \in Jm, im \in Im, v \in V, w \in W, jm \neq im \tag{6}$$

Equation 7 prevents vehicles from traveling between PHCC. And this because after leaving a PHCC a vehicle must visit at least one of the municipalities.

$$x_{jo,io,v,w} = 0 \quad \forall jo \in Jo, io \in Io, v \in V, w \in W, jo \neq io \tag{7}$$

Equation 8 ensures that each vehicle departs from only one PHCC, i.e., each vehicle has a stationary PHCC.

$$\sum_{jo \in Jo} \sum_{j \in J} \sum_{w \in W} x_{jo,j,v,w} \leq 1 \quad \forall v \in V \tag{8}$$

Service Level-Related Constraints

Equation 9 simply imposes that the satisfied demand can never surpass the total demand.

$$\sum_{i \in I} \sum_{v \in V} \sum_{w \in W} s_{i,j,v,p,w} \leq N_{p,j} \quad \forall j \in J, p \in P \tag{9}$$

Nevertheless, if the aim is to impose full demand satisfaction, Eq. 9 is transformed in Eq. (10).

$$\sum_{i \in I} \sum_{v \in V} \sum_{w \in W} s_{i,j,v,p,w} = N_{p,j} \quad \forall j \in J, p \in P \tag{10}$$

Equation 11 ensures that patients are served only if the respective municipality is visited by a vehicle.

$$s_{i,j,v,p,w} \leq N_{p,j} \times x_{i,j,v,w} \quad \forall i \in I, j \in J, p \in P, v \in V, w \in W \tag{11}$$

Patients' Preference-Related Constraints

Equation 12 guarantees that the preference of the patients to be visited at a specific period of the day is respected.

$$\sum_{\substack{i \in I \\ i \neq j}} \sum_{v \in V} s_{i,jm,v,p,w} \geq PREF_{jm,p,w} \quad \forall jm \in Jm, p \in P, w \in W \tag{12}$$

Vehicle-Related Constraints

Equation 13 defines the maximum visit time per vehicle, and ensures that each vehicle has time enough to travel between nodes, to serve their patients and to return to the PHCC within the regulated daily work time.

$$y_v = M_v - \sum_{i\in I}\sum_{\substack{j\in J\\ j\neq i}}\sum_{w\in W} T_{i,j} \times TRAFFIC_w \times x_{i,j,v,w} \quad \forall v \in V \tag{13}$$

Equation 14 defines the maximum visit time per vehicle.

$$\sum_{i\in I}\sum_{\substack{j\in J\\ j\neq i}}\sum_{p\in P}\sum_{w\in W} s_{i,j,v,p,w} \times D_p \leq y_v \quad \forall v \in V \tag{14}$$

Equation 15 ensures that each vehicle cannot be used for a number of hours higher than the regulated daily work time—and this include the time used for travelling (first term) and the time devoted for care provision (second term). Similarly, Eq. 16 ensures that the time available per period of the day w cannot be exceeded.

$$\sum_{i\in I}\sum_{\substack{j\in J\\ j\neq i}}\sum_{w\in W} T_{i,j} \times TRAFFIC_w \times x_{i,j,v,w} + \sum_{i\in I}\sum_{\substack{j\in J\\ j\neq i}}\sum_{p\in P}\sum_{w\in W} s_{i,j,v,p,w} \times D_p \leq M_v \quad \forall v \in V \tag{15}$$

$$\sum_{i\in I}\sum_{\substack{j\in J\\ j\neq i}} T_{i,j} \times TRAFFIC_w \times x_{i,j,v,w} + \sum_{i\in I}\sum_{\substack{j\in J\\ j\neq i}}\sum_{p\in P} s_{i,j,v,p,w} \times D_p \leq K_w \quad \forall v \in V, w \in W \tag{16}$$

Equation 17 defines maximum number of vehicles that can be used.

$$\sum_{v\in V} k_v \leq N \tag{17}$$

According to Eq. 18, if no visit is performed using a given vehicle v, that vehicle is not used.

$$k_v \leq \sum_{i\in I}\sum_{\substack{j\in J\\ j\neq i}}\sum_{w\in W} x_{i,j,v,w} \quad \forall v \in V \tag{18}$$

And Eq. 19 imposes the use of each vehicle, even if it is required for only one visit.

$$k_v \geq x_{i,j,v,w} \quad \forall i \in I, j \in J, v \in V, w \in W \tag{19}$$

Non-negativity and Binary Variables

Non-negativity conditions are given by Eqs. (20)–(21).

$$y_v \geq 0 \quad \forall v \in V \tag{20}$$

$$s_{i,j,v,p,w} \geq 0 \quad \forall i \in I, j \in J, p \in P, v \in V, w \in W \tag{21}$$

Binary variables are given by Eqs. (22)–(23).

$$x_{i,j,v,w} \in \{0, 1\} \quad \forall i \in I, j \in J, v \in V, w \in W \tag{22}$$

$$k_v \in \{0, 1\} \quad \forall v \in V \tag{23}$$

5 Case Study

In this section the LTC^{routes} model is applied to real data from a Portuguese region to illustrate how it can be used to support the planning of routes within the HBLTC sector. Specifically, the model is applied in the context of *Lisboa Ocidental*, in *Lisboa e Vale do Tejo*, one of the regions covered by the National Network of LTC (RNCCI) in Portugal.

5.1 Dataset and Assumptions Used

Within *Lisboa e Vale do Tejo*, the model was applied in the specific context of *Lisboa Ocidental*, comprised by 5 municipalities (*Ajuda*; *Alcntara*; *Santa Maria de Belm*; *Santo Condestvel*; *São Francisco Xavier*). And three PHCC exist for providing HBLTC within the scope of the RNCCI in this area (*Ajuda* PHCC; *Alcntara* PHCC; *Santo Condestvel* PHCC).

The main dataset used for applying the model includes:

i. The total demand of LTC for both home-based care and inpatient care per municipality in *Lisboa e Vale do Tejo* [7];
ii. Travel times between nodes, calculated using *Google Maps*.

Also, several assumptions were made for this application:

i. Health units are opened from 8:00 a.m. to 8:00 p.m., 7 days per week, as it is the case in the RNCCI in Portugal;
ii. HBLTC patients are visited once per week, allowing for the daily plan proposed in the model;
iii. The average duration of a visit per patient is 25 min;
iv. Considering patients requiring inpatient care, but receiving home-based care as a substitute service:
    a. These patients need to be visited every day—these patients require care on a daily basis in institutional settings, that's why substituting inpatient care by home-based care imply daily visits;
    b. The medium duration of each visit for these patients is 50 min (assuming twice the time required for those needing home-based care);
v. The work day is divided in 4 periods, morning (8:00–11:00), lunch period (11:00–14:00), afternoon (14:00–17:00), and evening (17:00–20:00)—the morning and

evening periods are the ones with more traffic in Lisbon, thus justifying the need to distinguish between these 4 periods;

vi. Each team/vehicle can only work 8 h/day (480 min/day).

Also, for illustrative purposes, the preference of patients is considered as follows: 20% prefer to be served in the morning; 40% prefer the lunch period; 10% prefer the afternoon; and 30% prefer the evening. And it is considered that preferences are respected for half of the patients. Also, traffic conditions characterizing the different periods of the day were taken into account when determining the time needed to perform each route.

5.2 *Scenarios Under Study*

Figure 1 summarizes the different scenarios under analysis in this paper.

The first scenario (Scenario 1) refers to a situation where one wants to plan the daily routes of HBLTC without any budget constraint, i.e., the question to be answered is "*How can we plan daily routes so as to fully satisfy daily demand, and this while using the minimum possible time?*". In this case the objective would be to minimize the travel time (Eq. 1), while accounting for traffic conditions and patients' preferences. This scenario is chosen for analysis since minimizing travel time is a common objective found in the literature. Two versions of this scenario are explored:

i. Scenario 1A: One wants to plan daily routes while ensuring the full satisfaction of home-based care demand (Eq. 10);
ii. Scenario 1B: An extra amount of home-based care demand is considered, namely, 10% of the total inpatient demand should be satisfied through home-based care. Transferring institutionalized patients to home-based care settings is considered for two reasons. Not only patients prefer to stay at home close to their relatives,

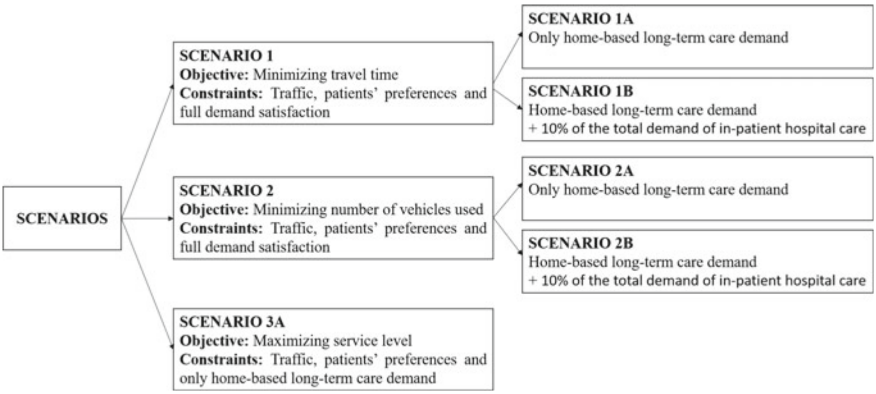


Fig. 1 Scenarios' under study

but also the Portuguese state can save money with this transference because home-based care is a cheaper alternative when compared to institutional care (one should be aware that home-base care can be provided as a substitute service of inpatient care).

On the other hand, Scenario 2 is analyzed to explore the amount of resources needed to fully satisfy HBLTC. Particularly, one wants to explore how many vehicles are needed to fully satisfy the demand in each municipality. The objective would be the minimization of vehicles used (Eq. 2). The same two versions explored under Scenario 1 are explored under Scenario 2 (Scenario 2A and Scenario 2B). One should be aware that the number of vehicles used can be used as a proxy for costs, and so by minimizing the number of vehicles we are minimizing costs, which is also a key objective to be considered in this sector (particularly, in NHS-based countries with severe health care budget constraints). Finally, Scenario 3A considers the current context of budgetary cuts in health where it is not possible to fully satisfy health care demand, and thus intends to answer the question "*How much of the total demand is it possible to be satisfied using the available budget?*". In this case the objective would be to maximize the number of served patients (Eq. 3), while accounting for the number of vehicles available, traffic and preferences constraints. This is a key scenario for any NHS-based country with budget constraints that limit the number of patients to be served, such as happens in Portugal.

5.3 Results

Results obtained under each scenario are described below. These results were obtained with the General Algebraic Modeling System (GAMS) 23.7 using CPLEX 12.0 on a Two Intel Xeon X5680, 3.33 GHz computer with 12 GB RAM.

Scenario 1

According to Scenario 1A, the minimum daily travel time needed to fully satisfy the current demand for home-based LTC is 100,4 min. Under this scenario, 12 routes should be performed so as to fully satisfy this demand—3 distinct routes, each traveled by 4 different vehicles. One should be aware that each route needs to be performed by several vehicles, because one single vehicle is not enough to visit all the patients in the area until the end of the day.

According to Scenario 1B, so as to absorb 10% of the inpatient demand through HBLTC, 14 daily routes need to be planned (6 distinct routes, each travelled by different vehicles), increasing the minimum daily travel time to 131,9 min.

Scenario 2

To fully satisfy the current demand for HBLTC in *Lisboa Ocidental*, and according to Scenario 2A, all the PHCC need to operate with 6 vehicles (with these vehicles being used to perform 6 distinct routes). Accordingly, this number of vehicles is far less than the number of vehicles obtained under Scenario 1. This lower number of vehicles is thus possible by allowing for longer (and, consequently, more expensive) routes.

On another hand, according to Scenario 2B, the added 10% of inpatient demand results in a higher number of vehicles needed—in particular, 12 different vehicles are needed to fully satisfy the imposed demand.

Scenario 3A

Under Scenario 3A, and considering that only 5 vehicles were to be available for the 3 PHCC, 28% of the HBLTC demand would remain unsatisfied, meaning that those patients would need to be referenced to other ACES of the RNCCI and/or wait for a vacancy in the network.

Table 2 summarizes the results obtained under each of the five scenarios under study.

5.4 Computational Results

Key computational results obtained when running the model under each scenario are presented in the Table 3.

Table 2 Key results obtained under the scenarios presented in Fig. 1

Scenarios	Total travel time (minutes)	No. vehicles needed	% Satisfied HBLTC demand	No. different daily routes planned
Scenario 1A	100.4	12	100	3
Scenario 1B	131.9	14	100	6
Scenario 2A	196.4	6	100	6
Scenario 2B	259.0	12	100	12
Scenario 3A	257.7	5	72	5

Table 3 Key computations results per scenario

Scenarios	Time (minutes)	Gap	Iterations	Equations	Integer variables	Variables
Scenario 1A	4.17	0	926 952	27 352	6 750	23 106
Scenario 1B	16.67	0	1 881 892	27 352	6 750	23 106
Scenario 2A	16.67	0	1 524 901	27 352	6 750	23 106
Scenario 2B	2.44	0	126 083	27 352	6 750	23 106
Scenario 3A	0.3	0	91 957	4 627	1 125	3 856

5.5 Discussion of Results

From Table 2 it is clear the importance of analyzing conflicting objectives. From the results obtained one can see that:

i. Scenario 1 (where one aims at minimizing travel time) results in a more organized and standard daily route planning;
ii. Under Scenario 2 (in which the aim is to minimize the number of vehicles needed), it is clear that longer and more expensive routes may be required when the aim is to have the lowest investment in vehicles;
iii. In Scenario 3 one can see the impact that a lack of resources (in this case, vehicles) can have on the percentage of satisfied demand.

Under Scenarios 1A and 1B, it is possible to see that with an already optimized route plan (result of minimizing the travel time), the satisfaction of extra demand can be obtained without a costly investment (see how only 2 more vehicles allow for the satisfaction of 10% of the total inpatient demand).

Regarding Scenario 2A, it shows an alternative evaluation of the problem by finding the minimum number of vehicles to fully satisfy the HBLTC demand, and it is possible to see that with half the resources it is possible to fully satisfy the existing demand at the cost of some longer routes. It sheds light on a quick investment oriented solution to solve the current situation of long waiting times for patients to start being treated. However, Scenario 2B shows that, to achieve the same goal of adding 10% of the inpatient demand, the added investment is way higher compared to Scenario 1.

On the other hand, the awareness that there are budgetary cuts in health that make it impossible to fully satisfy health care demand is key to understand the importance of studying scenarios of resource scarcity and comprehend the impact that unitary investments can have in the overall performance of the network. Scenario 3A thus represents a key scenario to be explored under these circumstances.

On a final note, it adds value to the whole model the fact that it considers, even at a simple level, the traffic conditions in different periods of the day and the patients' preferences for their visit time.

6 Conclusions and Further Work

Long-term care demand is growing at a fast pace and many NHS-based countries like Portugal are not prepared to answer back. On the other hand, home-based health care in general, and home-based long-term care (HBLTC) in particular, has been appointed by some studies as an important component of health care systems, having the potential to reduce costs of health care provision and free capacity in overcrowded acute care settings.

Within this setting, there is clearly the need to plan the adequate provision of HBLTC, with this planning being especially relevant in National Health Service

(NHS)-based countries facing high pressures to reduce and control health care spending (such as happens in Portugal). This adequate planning implies, among other issues, the planning of routes for HBLTC provision, with mathematical programming models representing the most widely used approach for that purpose.

This study thus proposes the development of the LTC^{routes} model. This is an optimization model based on mathematical programming developed to support the planning of routes to patients' home within the HBLTC sector in any NHS-based country. This model aims at informing health practitioners on how they can plan their routes and on the required investments in order to meet HBLTC demand. Multiple conflicting objectives that are relevant in this sector are considered, such as the minimization of costs (in terms of travelling time and number of vehicles required to perform the routes) and the maximization of service level (in terms of the percentage of patients visited). Patients' preferences, traffic conditions and budget constraints are also imposed to explore the impact on planning decisions.

The key contributes of this paper are as follows:

i. It proposes a generic tool that can be used to support route planning decisions of HBLTC delivery on a daily basis;
ii. It explores the impact of accounting for different objectives relevant in this sector, such as the minimization of costs and the maximization of service level;
iii. It proposes a planning model that allows accounting for patients' preferences and traffic conditions as constraints;
iv. It proposes a generic approach that can be used to plan HHC delivery in general, and that can be applied to other regions of Portugal or other countries with a NHS.

To demonstrate the usefulness of the proposed model, it was applied to a Portuguese case study in *Lisboa Ocidental* performed, and the particular case of home health care provided within the scope of the National Network of LTC (RNCCI) was considered. The main results confirm the importance of analyzing conflicting objectives, since different routes are obtained as result when different planning objectives are imposed. The minimization of travel times reflects a more organized and standard daily route planning, whereas minimizing the vehicles needed to fully satisfy the existing demand is key when planning future investment. Also, the maximization of the service level, measured in terms of satisfied demand, shows the impact of the lack of investment on the number of patients that cannot be served by the RNCCI.

In terms of further research, different lines are worth to be pursued. First, as a simplistic and general model, it is certainly possible to take this base and deepen the constraints considered (such as lunch breaks, nurse preferences, team skills, team composition, overtime, weekly planning). Another research direction can be the development of a stochastic model so as to explore the impact of uncertainty in route planning decisions. Moreover, given the importance noted in analyzing multiple conflicting objectives, a multi-objective model that allow exploring the joint impact of multiple objectives, such as minimizing costs together with maximizing preferences, should also be pursued. Particularly, the augmented ε-constraint method may be used for that purpose. Also, and in addition to the policy objectives mentioned in this study, other key objectives may worth to be considered for planning purposes. In particular, quality of care may need to be considered through the maximization of health and wellbeing benefits—and within such circumstances, QALYs may be used as a proxy for health benefits, while the ICECAP instrument may be used to estimate wellbeing benefits. Developing an easy-to-use tool based on the proposed mathematical programming model should also be pursued, thus supporting the route planning on a daily basis with real professionals. Finally, an additional line of research would involve exploring the use of alternative means of transport and its impact on the routes and on costs.

References

1. Akjiratikarl, C., Yenradee, P., Drake, P.R.: PSO-based algorithm for home care worker scheduling in the UK. Comput. Ind. Eng. **53**(4), 559–583 (2007)
2. Baker, M.: Making Sense of the NHS White Papers, 2nd edn. Radcliffe Medical Press, Oxon (2000)
3. Bachouch, R.B., Guinet, A., Hajri-Gabouj, S.: A decision-making tool for home health care nurses' planning. Supply Chain Forum Int. J. **12**(1), 14–20 (2011)
4. Bertels, S., Fahle, T.: A hybrid setup for a hybrid scenario: combining heuristics for the home health care problem. Comput. Oper. Res. **33**(10), 2866–2890 (2006)
5. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer, New York (1997)
6. Braekers, K., Hartl, R.F., Parragh, S.N., Tricoire, F.: A bi-objective home care scheduling problem: analyzing the trade-off between costs and client inconvenience. Eur. J. Oper. Res. **248**(2), 428–443 (2016)
7. Cardoso, T., Oliveira, M., Barbosa-Póvoa, A., Nickel, S.: Modeling the demand for long-term care services under uncertain information. Health Care Manag. Sci. **15**(4), 385–412 (2012)
8. Cohon, J.L.: Multiobjective Programming and Planning. Academic Press Inc., London (1978)
9. Fikar, C., Hirsch, P.: Home health care routing and scheduling: a review. Comput. Oper. Res. **77**, 86–95 (2017)
10. Gutirrez, E.V., Vidal, C.J.: Home health care logistics management: Framework and research perspectives. Int. J. Ind. Eng. Manag. **4**(3), 173–182 (2013)
11. Knapp, M., Somani, A.: Health financing: long term care, organization and financing. In: Carrin, G., Buse, K., Heggenhougen, K., Quah, S.R. (eds.) Health Systems Policy, Finance, and Organization, pp. 250–9. AcademicPress, San Diego (2009)
12. Lanzarone, E., Matta, A.: Robust nurse-to-patient assignment in home care services to minimize overtimes under continuity of care. Oper. Res. Health Care **3**(2), 48–58 (2014)
13. Maya Duque, P.A., Castro, M., Srensen, K., Goos, P.: Home care service planning. The case of Landelijke Thuiszorg. Eur. J. Oper. Res. **243**(1), 292–301 (2015)
14. Milburn, A.B.: Operations research applications in home healthcare. In: Hall, R. (eds) Handbook of Healthcare System Scheduling. International Series in Operations Research & Management Science, vol. 168. Springer, Boston (2012)
15. Miller, C.E., Tucker, A.W., Zemlin, R.A.: Integer programming formulation of traveling salesman problems. J. Assoc. Comput. Mach. **7**, 326–329 (1960)
16. Ministry of Health: Decreto-lei n 101/2006: Cria a Rede Nacional de Cuidados Continuados Integrados [Create the National Network of Long-Term Care], Dirio da Repblica: I Srie-A, no 109 de 6 de Junho (2006)
17. Ministry of Health: Lei n 48/90: Lei de bases da saúde [Fundamental principles of health], Diário da República: I Série-A, n 195 de 24 de Agosto (1990)
18. Rodriguez, C., Garaix, T., Xie, X., Augusto, V.: Staff dimensioning in homecare services with uncertain demands. Int. J. Prod. **53**(24), 7396–410 (2015)

19. Simões, J., Augusto, G.F., Fronteira, I., Hernández-Quevedo, C.: Portugal: health system review. Health Syst. Transition **19**(2), 1–184 (2017)
20. Trautsamwieser, A., Hirsch, P.: Optimization of daily scheduling for home health care services. J. Appl. Oper. Res. **3**(3), 124–136 (2011)
21. Vidal, C.J., Gutirrez, E., Gutierrez, V.: Home health care logistics management problems: a critical review of models and methods. Revista Facultad Ingeniería Univ. Antioq. **68**, 160–175 (2013)
22. World Health Organization: Home-based long-term care: report of a WHO study group. World Health Organization, Geneva (2000)

Selection of a Strategic Plan Using an Integrated AHP-Goal Programming Approach

Dalila B. M. M. Fontes, Teresa Pereira and Márcia Oliveira

Abstract This work proposes a multi-criteria decision making model to assist in the choice of a strategic plan for a world-class company. The Balanced Scorecard (BSC) is a support tool of Beyond Budgeting that translates a company's vision and strategy into a coherent set of performance measures. However, it does not provide help in choosing a strategic plan. The selection of a strategic plan involves multiple goals and objectives that are often conflicting and incommensurable. This paper proposes an integrated Analytic Hierarchy Process-Goal Programming (AHP-GP) approach to select such a plan. This approach comprises two stages. In the first stage, the AHP is used to evaluate the relative importance of the initiatives with respect to financial indicators/KPIs; while in the second stage a GP model incorporating the AHP priority scores is developed. The GP model selects a set of initiatives that maximizes the earnings before interest and taxes (EBIT) and minimizes the Capital Employed (CE). The proposed method was evaluated through a case study.

Keywords Decision-making · AHP · Goal programming · AHP-GP · Balanced scorecard

D. B. M. M. Fontes (✉)
Faculdade de Economia da Universidade do Porto, and LIAAD/INESC TEC,
Rua Dr. Roberto Frias, Porto 4200-464, Portugal
e-mail: fontes@fep.up.pt

T. Pereira
Institute of Engineering of Porto of Polytechnic Institute of Porto, CIDEM, ISEP,
Rua Dr. Antnio Bernardino de Almeida 431, Porto 4200-072, Portugal
e-mail: mtp@isep.ipp.pt

M. Oliveira
Skim Technologies, London, UK
e-mail: marcia@skim.it

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_9

1 Introduction

In volatile, dynamic, and complex business environments, organizations can differentiate themselves through strategic planning. Management control systems play a crucial role in the implementation of a strategy and in the performance of an organization. Beyond Budgeting (BB) is an innovative management control system, which has the advantage of adaptability by fostering continuous planning rather than an adherence to a restrictive fixed annual budget. One of the support tools of BB is the Balanced Scorecard (BSC), which is a conceptual framework for translating the strategic objectives of an organization into a set of operational attributes.

The selection of a strategic plan (i.e., a plan of initiatives) involves multiple goals and objectives. Although these objectives can be complementary, very often they are conflicting and incommensurable. Therefore, the selection of a strategic plan which optimizes all objectives and goals simultaneously is a difficult problem. Since there is no strategic plan selection in the performance of the BSC, such a decision can be improved using a multi-objective optimization (MOO) method, such as the Analytic Hierarchy Process-Goal Programming (AHP-GP) approach proposed here. Similar approaches have been proposed in several application areas, see, e.g., [2, 11, 12, 14, 16, 18, 19]. It has been shown to be an efficient and effective tool for modeling and analyzing problems that involve multiple and conflicting objectives and thus, finding solutions involving trade-offs.

The work reported was conducted at the department of Planning and Performance management of Nors Group. In 2013, the Nors Group replaced the traditional budgeting model with the Beyond Budgeting model. Changing the control management model has increased the responsibility of all companies in the group, which led to the need of defining a strategic plan, i.e., a plan of initiatives. This study aimed to develop a model that integrates the financial vision of the organisation's strategy with a strategic plan, taking into account the restrictions on each objective and initiative, and also the relative importance of goals.

The remainder of this paper is organized as follows. Section 2 presents the theoretical background and provides a general review of multicriteria decision analysis methods, with emphasis on the Analytic Hierarchy Process (AHP), on Goal Programming (GP), and on the AHP-GP approach. Section 3 introduces the proposed methodology to select a strategic plan and offers a brief explanation of some issues that have relevance to the real-world application, including organizational structure, planning, and application of the Beyond Budgeting model, and details the implementation of the proposed method to the selection of a strategic plan in the Nors Group. Section 4 highlights the contributions of this work, as well as the results obtained, and points out future research directions.

2 AHP-GP Methodology

The AHP method [17] is widely used in practice due to its relative simplicity and flexibility. AHP allows to formalize the structure of a decision problem, to check the level of inconsistency, and to efficiently process both qualitative and quantitative data without involving any cumbersome mathematics. Despite these advantages, decision analysts emphasize some drawbacks of the use of AHP for multi-criteria decision making. Bana e Costa and Vansnick [5] have shown that the eigenvalue method, which is used to derive the AHP priority vector, can violate a condition of order preservation that is fundamental in decision aiding; Goodwin and Wright [8] point out as the biggest flaws of AHP (i) the large number of comparisons that may be required, (ii) the possibility of rank reversal, and (iii) the discrepancy of the nine-point scale. However, practice shows that the AHP is working well in various integrated approaches, one of the most popular being with GP. GP is an optimization technique that offers organized ways of considering more than one objective at a time. The basic idea is to establish a numerical goal for each objective and then find a feasible solution (i.e., a solution that satisfies all constraints) such that the deviations from the goal values set to the objectives are minimized. The AHP can be used to elicit a set of preferential weights of criteria that are then incorporated into the GP model.

The first use of the AHP-GP approach appears to be that of [7] in the context of military planning. Since then, it has been applied to a variety of other fields.

Kruger and Hattingh [13] use an AHP-GP approach to solve the problem of allocating internal auditing time among competing projects. AHP was used to deal with qualitative risk assessments, providing a risk score for each audit project. Then a weighted GP model was developed to determine an optimal allocation of internal auditing time using the preceding risk evaluation scores, as well as quantitative information provided by the decision maker (DM) regarding the projects.

In [3] an AHP-GP approach is used to define the best strategies for the maintenance of critical centrifugal pumps in an oil refinery. The AHP evaluates the maintenance policy priority scores and then the best is chosen using a Lexicographic GP model. Arunraj and Maiti [1] proposes an AHP-GP approach to a similar problem for a benzene extraction unit of a chemical plant, in which the policy selection is based on equipment risk and maintenance costs. This work improves on that of [3] as the risk contribution of different equipment is computed and supplied to the experts for rating maintenance policies. More recently, [16] proposes a combined AHP-GP model to determine a maintenance strategy for the most critical electrical equipment in a big scale hydroelectric power plant, preceded by a combined AHP-TOPSIS method for the identification of the most critical equipment. The authors report an improvement of the equipment downtime of about 77%. In this work, the authors also provide a brief literature review regarding the use of optimization models in maintenance policy, following the classification proposed in [6], where a comprehensive review can be found.

Barbosa and Gomes [2] propose an AHP-GP approach to assess and improve efficient and sustainability in the chemical industry. The AHP was used to identify and evaluate the variables of interest. Then, a GP model was used to determine the decisions to be made in order to achieve the goals previously set.

Ho [9] proposes to use the AHP-GP approach for selecting the best set of warehouses taking into account both qualitative (e.g., customer satisfaction) and quantitative factors (e.g., total cost, total delivery days, effectiveness of capacity utilization for the warehouse). The AHP is first used to determine the relative importance of alternative warehouses and, then, the GP model is formulated to select the best set of warehouses considering the limited available resources. In a more recent work, [11] addresses the facility location problem considering several objectives. The decisions involve the choice of tenants for a large commercial area in a shopping mall to be divided into rental units and then rented. The study uses a combined AHP-GP approach. In the combined approach the AHP is used to prioritize the goals. The GP model is then solved considering these priorities.

Approaches combining AHP and multi-choice GP (MCGP) have also been proposed in education. For example, [14] addresses the problem of selecting an appropriate set of information systems tools, considering the time limitations of teachers and students, for e-learning adoption in university courses. The AHP is used to obtain the weights for the evaluation factors based on different course purposes and then GP uses these weights as goal coefficients in the objective function to choose a subset of tools that provides the least weighted under-achievement of students and teachers satisfaction. Another study using a similar approach is that of [12], which aims at determining an efficient course plan following the Bologna process. The proposed approach was applied in an industrial engineering department. Again the AHP is used to determine the weights of the criteria used to evaluate the courses, which are then used to a MCGP model that considers multiple aspiration levels to obtain an efficient solution.

An AHP-GP approach for asset allocation is reported in [18]. The model proposed considers several levels of market conditions (e.g., recession, trough, recovery, expansion, peak) and of the investor's risk profile (e.g., conservative, moderately conservative, moderate, moderately aggressive, and aggressive). The AHP method is used to analyze the percentage of investment to be allocated to each asset class (e.g., bonds, stocks, liquid assets) by an investor. Then, the scores of asset classes are integrated into a two-step optimization model. The first step consists in determining the combination of funds that will produce maximum returns for a given level of risk. However, this model does not take into account the DM's ideal goal about the proportion of stocks, bonds, and liquid assets. Hence, the next step is the goal optimization, that minimizes deviations with respect to the ideal ratios of stocks, bonds, and liquid assets obtained by the AHP.

Recently, some other interesting applications have been reported. For example, [19] proposes a fuzzy AHP-GP approach to the Shelter location problem. Fuzzy AHP is used since potential sites are evaluated in terms of qualitative criteria and with limited information. The sites suitability indices, given by the priority vectors from fuzzy AHP, are used as input to the GP model, which determines a subset of sites that minimize the deviations form anticipated goals, while satisfying a set of constraints (budget, thresholds for amenities, etc.). Wichapa and Khokhajaikiat

Fig. 1 Problem framework, Adapted from Nors Group [15]

[20] address the problem of determining the location of infectious waste disposal centers. Potential locations are, usually, defined by the legislation and then evaluated by experts that define relevant factors. These evaluations are used within the fuzzy AHP in order to determine priority weights and then fed to the GP model that, in addition to the goal constraints, also has to satisfy demand, capacity, and equipment constraints.

As seen in the aforementioned articles, the combined AHP-GP approach has been used on a wide scope of decision problem areas. Nevertheless, it has not yet been used to aid in the selection of a strategic plan.

3 Proposed Methodology for the Selection of a Strategic Plan

3.1 Problem Definition

This study aims to develop a model which helps to select a strategic plan optimizing several objectives simultaneously.

Organizations usually need to set global objectives that reflect their strategy. These objectives have to be operationalized by various financial indicators and Key Performance Indicators (KPIs) since, usually, the global objectives cannot be directly measured. The set of financial indicators and KPIs represent the financial view of the strategy and show the direction the company wants to take. To accomplish an organization's strategy it is essential to convert its financial view into a set of initiatives (or strategic plan), whose implementation ensure that the aforementioned indicators are achieved. An overview of these steps is provided in Fig. 1.

Each initiative has, at the same time, a positive (e.g., increasing profitability) and a negative (e.g., increasing costs) contribution to the objectives. Therefore, it is not apparent which of the initiatives should be implemented in order to achieve the expected best results. In addition, the organization may have a preference regarding which initiatives to implement and also on initiative's preferences. Thus, it is important to take into account both the established objectives and the organization's preferences regarding the initiatives in order to select a strategic plan.

The BSC is an appropriate tool for designing operational strategies, but it does not help in the choice of a strategic plan under several objectives, that may be conflicting and incommensurable [10]. Thus, this work proposes an AHP-Goal Programming integrated approach to address this problem.

3.2 *Integrated AHP-GP Approach*

The proposed approach comprises two stages, namely: the application of the AHP to determine the relative importance (i.e., scores) of initiatives, according to the DM's preferences; and the solution of a GP model incorporating the final AHP output and other management information in order to select a strategic plan, i.e., the set of initiatives to be implemented, that achieves the established objectives. Considering that the GP approach combined with the AHP is (i) relatively simple and clear for application in real-world projects; (ii) useful in decision problems where the multiple goals are conflicting, and (iii) widely used for different types of problems, this method was chosen for the study and it is discussed in more detail in this section.

3.2.1 AHP

Regarding the AHP hierarchy tree, our decision problem comprises three levels: (i) overall goal, (ii) criteria, and (iii) alternatives, as illustrated in Fig. 2.

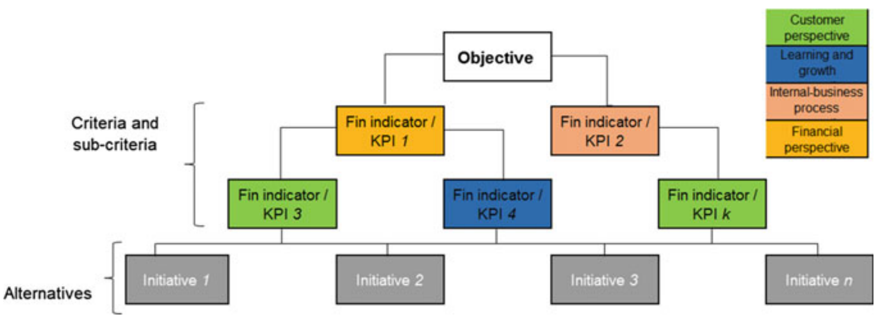


Fig. 2 Hierarchical structure of the decision problem

At the top of the hierarchy tree is the financial indicator that shows the ambition of the company and which is operationalized by various financial indicators/KPIs, since the global objective is directly immeasurable. The financial indicators represent the financial vision of the strategy (e.g., Sales, Working Capital Needs), while the KPIs describe them. Therefore, the levels below consist of financial indicators and KPIs (criteria and sub-criteria), with each color representing one of the four perspectives of the BSC, namely, customer perspective, learning and growth, internal-business process, and financial perspective. The last level consists of the set of the initiatives (i.e., the alternatives), proposed by the DM to achieve the global objective, that are to be evaluated with respect to the defined criteria.

The AHP outcomes are: (i) the global w_{g_j} and local w_{l_j} weights for each financial indicator/KPI j; (ii) the local weight $p_{l_{ij}}$ for each initiative i with respect to each financial indicator/KPI j; and (iii) the global weight w_i for each initiative i.

The final result of the AHP analysis is the normalized vector of the relative importance scores of initiatives with respect to financial indicators/KPIs. Thus, the overall score of initiatives takes into account not only the initiatives' performance, but also the financial indicators/KPIs' weights. These scores are then incorporated into the GP model, so that it can take into account the DM's preferences.

3.2.2 The GP Model

The next step is to build a GP model. This multicriteria decision aid model enables the DM to specify desirable goals and constraints for each objective, in order to find a solution that satisfies these constraints and tries to meet the goals as close as possible. The lexicographic or pre-emptive GP approach [4] is adopted in this study, since it allows the different goals to be ranked according to different priority levels that reflect the target allocated to them by the DM. The aim of the GP model is to help the DM find a solution that best satisfies the target set for each objective.

Before providing the lexicographic GP model, let us first introduce the notation used. Let $J = J' \cup J''$ be the set of financial indicators/KPIs and $J'$and J'' the subsets of the financial indicators/KPIs for which upper u_j and lower l_j bounds are imposed, respectively, and $G \subset J$ the set of m financial indicators/KPIs for which a goal (target value) has been established. Also, let k be the number of priorities specified and consider G divided into k subsets $G = G_1 \cup G_2 \cup ... \cup G_k$, where G_l for $l = 1, \ldots, k \leq m$ is the subset of financial indicators/KPIs associated with priority l. Further, let d_g^+ and d_g^- denote the deviation variables associated with over-achievement and under-achievement of the goal value established for the financial indicator/KPI $g \in G \cap J'$ and $g \in G \cap J''$, respectively. Consider also I the set of n initiatives, a_{ij} the expected contribution of initiative $i \in I$ to financial indicator/KPI $j \in J$, w_i the weight of initiative $i \in I$, which is obtained from the AHP stage, and d^- the deviation variable associated with (the under-achievement of) the execution of all initiatives. Finally, the decision variables are represented as binary structural variables x_i, which take the value 1 if initiative $i \in I$ is executed and 0 otherwise.

To formulate a generic lexicographic goal programme algebraically, a function $P_l(\cdot)$ is defined for each priority level $l = 1, \ldots, k$ and it involves the subset of deviation variables associated with the goals of the corresponding priority level, that is, deviation variables for the financial indicators/KPIs in G_l. The objective function is defined as the sequential minimization of these functions, see expression (1). The exact function $P_l(\cdot)$ depends on the nature of the problem to be formulated as a goal programme and is, typically, given by a weighted sum of the deviations involved. In the lexicographic GP model the objective function consists in the minimization of the unwanted deviations from the aspiration levels, considering one priority at the time. Deviations of a higher priority level are regarded as infinitely more important than that of deviations of a lower priority level. This leads to a series of sequential optimizations, each of which with reduced feasible region as the values of the deviations of higher priority level must be maintained.

$$Lex \min F = \left[P_1(d_g^+, d_g^- : g \in G_1), P_2(d_g^+, d_g^- : g \in G_2), \ldots, P_k(d_g^+, d_g^- : g \in G_m) \right]. \tag{1}$$

The GP model has two types of constraints: goal constraints and system or functional constraints. The goal constraints are written as in expressions (2)–(4):

$$\sum_{i \in I} a_{ij} x_i - d_g^+ \leq u_g, \forall g \in G \cap J', \tag{2}$$

$$\sum_{i \in I} a_{ij} x_i + d_g^- \geq l_g, \forall g \in G \cap J'', \tag{3}$$

$$\sum_{i \in I} w_i x_i + d^- = 1. \tag{4}$$

Constraints (2) are used for goals where it is only necessary to minimize the deviation associated with the over-achievement of a goal; while constraints (3) are used for the goals where it is only necessary to minimize the deviation associated with the under-achievement of a goal. If a goal has both over and under-achievement deviations to be minimized, then it will be associated with both constraints (2) and (3). Constraint (4) is associated with the execution of all initiatives, which is desirable.

The system or functional constraints are limitations to the scope of decisions and thus, define the feasible region as any possible strategic plan must satisfy such constraints. Thus, they do not have deviation variables. The system constraints are defined in Eqs. (5) and (6) and ensure that the values of all financial indicators/KPIs are within their bounds.

$$\sum_{i \in I} a_{ij} x_i \leq u_j, \forall j \in J' \setminus G, \tag{5}$$

$$\sum_{i \in I} a_{ij} x_i \geq l_j, \forall j \in J'' \setminus G. \tag{6}$$

Finally, constraints (7) and (8) define the nature of the variables: assignment variables x_i are binary, while deviation variables are continuous and non-negative.

$$x_i \in \{0, 1\}, \forall i \in I, \tag{7}$$

$$d_g^+, d_g^-, d^-, \geq 0, \forall g \in G. \tag{8}$$

3.3 Model Implementation

The proposed methodology has been empirically evaluated using the data of one of the companies that integrates the Nors Group, which is the leader in the Portuguese after-market for the distribution of parts for heavy vehicles, buses, and workshop equipment.

The study is divided into the 4 steps outlined in Fig. 3:

1. Collection of data and information regarding the global objectives, quantitative targets, and strategic initiatives;
2. Computation of criteria and initiatives' weights and evaluation of initiatives;
3. Development and solution of a lexicographic GP model;
4. Analysis of the results.

In accordance with the financial vision of the company's strategy, i.e., increase of the Residual Revenue, two main goals were defined: to increase EBIT and to decrease Capital Employed. As it can be seen in Fig. 4, to increase the EBIT it is necessary to increase Sales without decreasing the percentage of Gross Margin (GM %). However, increasing Sales leads to an increase of Stock and Clients. Regarding the decrease of the Capital Employed, there are two important conditions that need to be satisfied in order to achieve this reduction: the decrease of Stock and the decrease

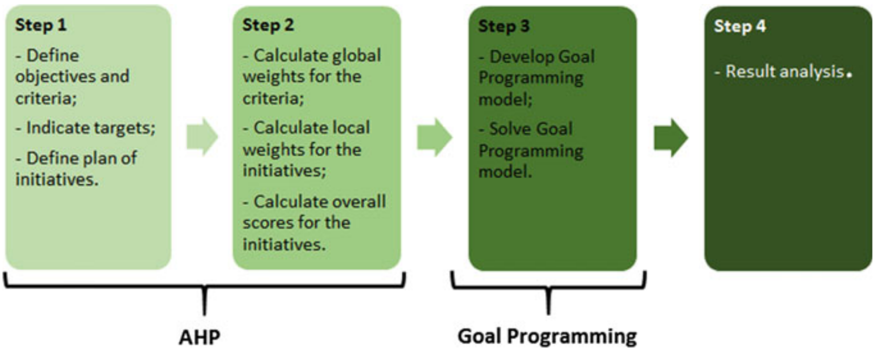


Fig. 3 Steps of the proposed methodology

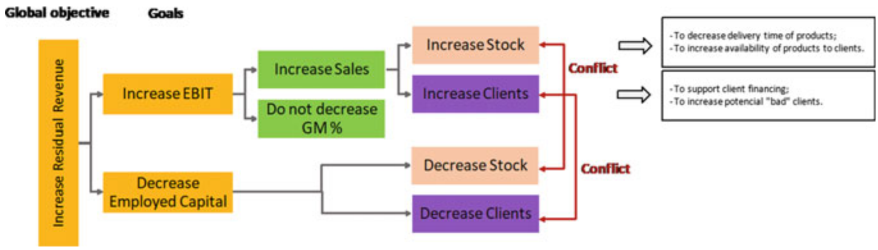


Fig. 4 Objective and goal definition

of Clients. Thus, it can be concluded that these two goals are conflicting as we cannot achieve one without impairment of the other.

The global objective Residual Revenue (level 1), its division into more specific objectives (EBIT and Capital Employed CE—level 2), and subsequential operationalization by various financial indicators and KPIs (levels 3–6) is graphically depicted in Fig. 5. Finally, the last level consists of the set of initiatives that are to be assessed in terms of the financial indicators/KPIs specified in the previous levels.

The DM has then evaluated the relative importance of each criterion by performing pairwise comparisons and providing a numerical evaluation using Saaty's nine-point scale. This was carried out by comparing each criterion with the one that is immediately above in the AHP hierarchy tree, according to the hierarchy tree (see Fig. 5). Each manager belonging to the expert group gave his/her own rating and then, whenever necessary, the ratings were discussed and adjusted in order to reach a consensus. These ratings were then used to obtain the local weights. The global weight of each criterion was computed by multiplying its local weigh by the global weight of the criterion on the level immediately above, i.e., parent criterion. Both the local and the global weights of the criteria (see Figs. 6 and 7) were computed using the demo version of the software "MakeItRational".

The initiatives have been evaluated by the DM. In order to facilitate the assessment, the DM was asked to obtain the expected contribution of each initiative to each financial indicator/KPI. This numerical information served as a basis for the DMs assessment and corresponds to the expected values (a_{ij}) used in the GP model (see Table 1). The corresponding local weights $p_{l_{ij}}$ are given in Table 2 and have been obtained by applying the eigenvalue method to the pairwise comparison matrices of the initiatives. As it can be seen, none of the initiatives has the highest local weight (in green) in all criteria, nevertheless initiative 2 is more promising since it achieves the best value in 9 out of the 12 criteria considered.

The overall relative scores w_i are obtained as in Eq. (9) and reported in Fig. 8. It should be noticed that, the w_i values have been calculated using all 14 criteria (see Fig. 5, which depicts the decision hierarchy of the problem); however, Table 2 reports the local weights of the initiatives only for 12 of the 14 initiatives. All data regarding

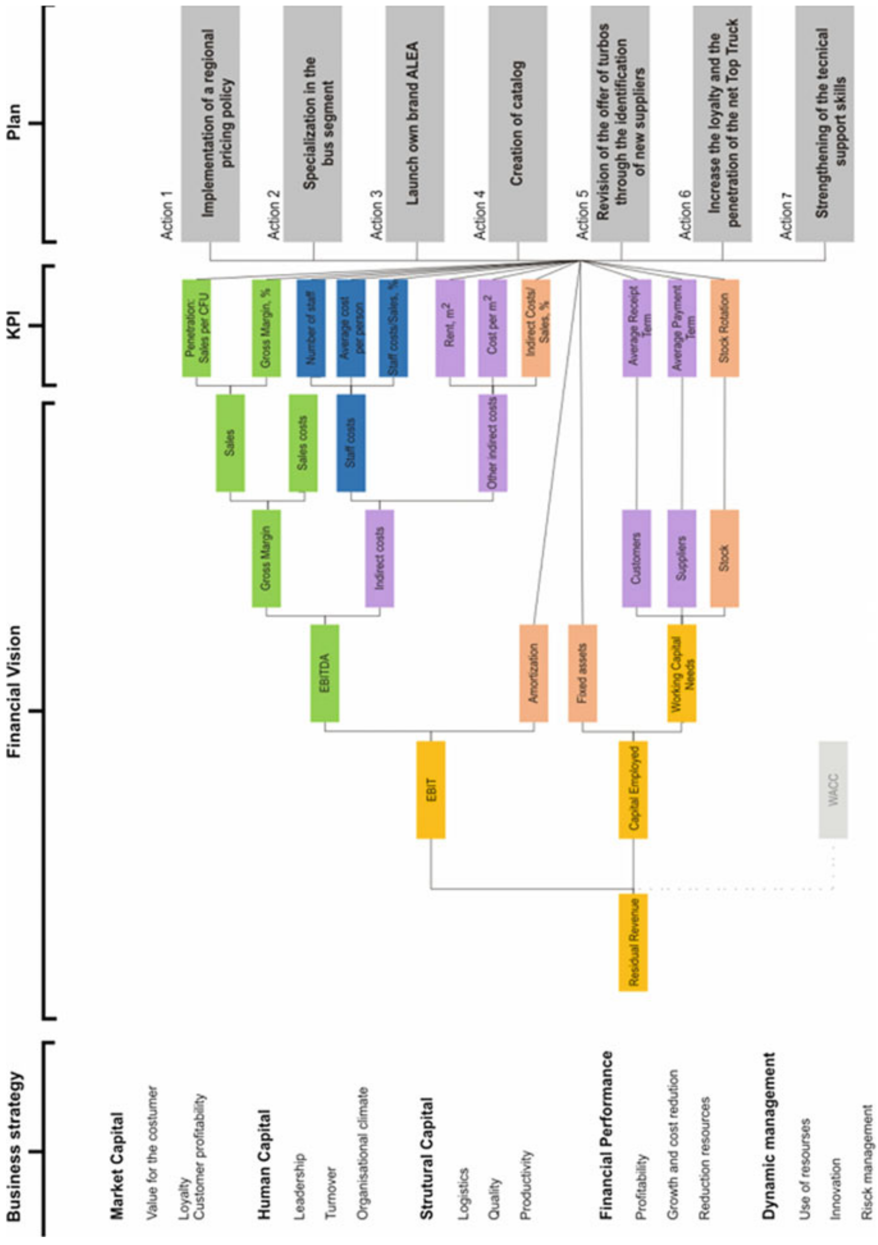


Fig. 5 The decision hierarchy of the decision problem at hand

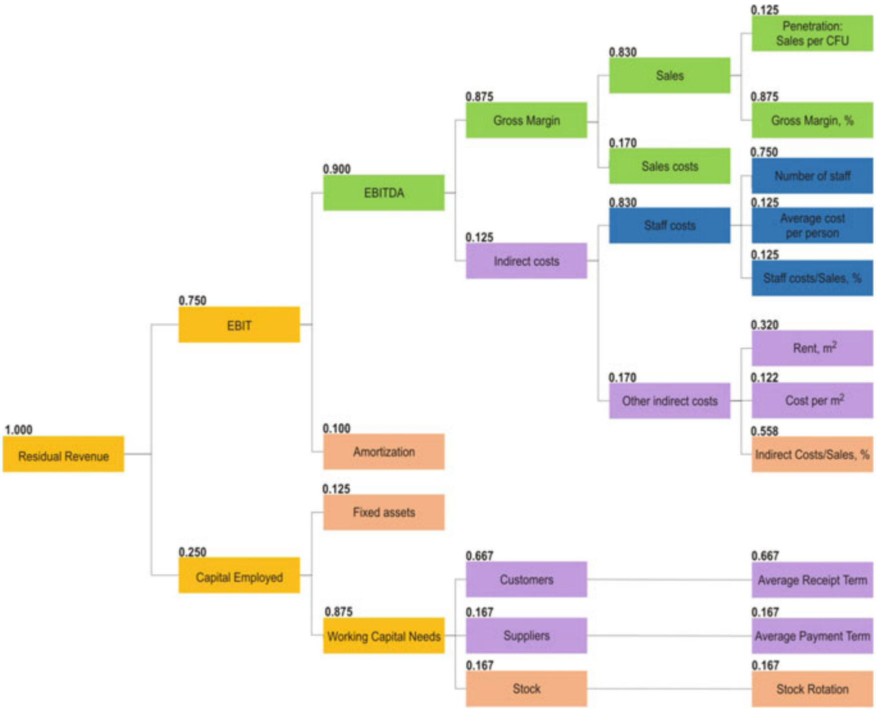


Fig. 6 Criterion local weights (w_{l_j})

Amortization and Fixed assets, except for criterion weights, has been considered confidential and thus, cannot be reported here.

$$w_i = \sum_{j \in J} w_{g_j} p_{l_{ij}}, \quad \text{for all } i = 1, 2, \ldots, n. \tag{9}$$

The next step of the study is to build the GP model. The objective function is the minimization of the undesirable deviations from the goals set to EBIT, CE, and the number of executed initiatives (see Eq. (10)). The goal constraints (11)–(13) set the goals for EBIT, CE, and number of executed initiatives, respectively. These inequalities state that the objective is to achieve the desired value; however allowing smaller values. Since the difference between the "goal" value and the achieved value is minimized, the"goal" is not satisfied only whenever impossible do so. Constraints (14) and (15) state the nature of the decision variables and the remaining constraints, i.e., the system constrains, are given, in matrix format, in Table 3.

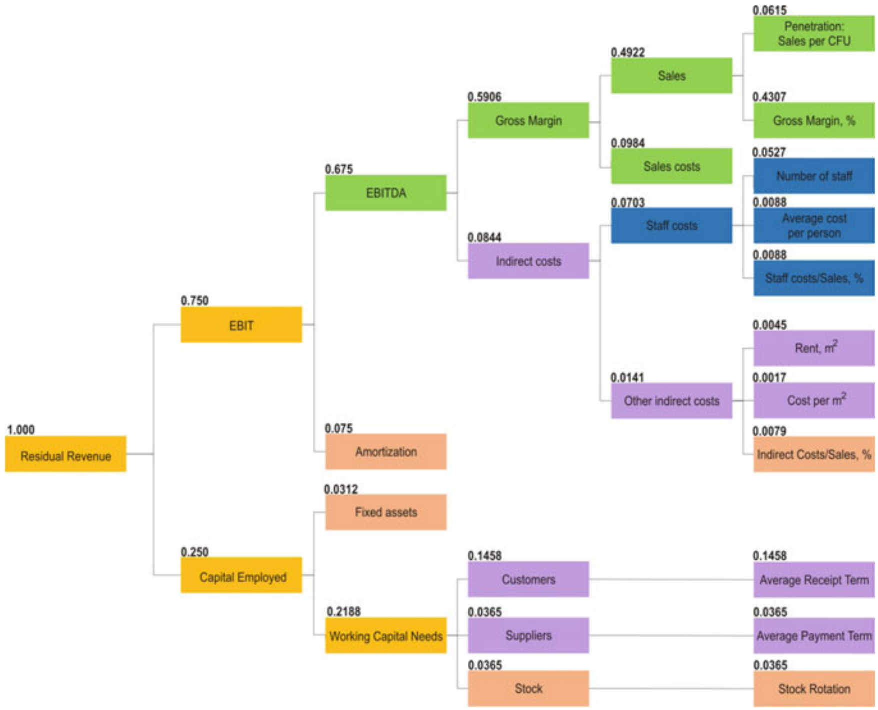


Fig. 7 Criterion global weights (w_{g_j})

$$Lex \min F = \{d_1^-, d_2^+, d_3^-\} \tag{10}$$

subject to:

$$198x_1 + 239.39x_2 + 90.51x_3 - 40x_4 + 102x_5 + 75x_6 - 30x_7 + d_1^- \geq 283.08, \tag{11}$$

$$520x_2 - 300x_4 - d_2^+ \leq 500.09, \tag{12}$$

$$0.22x_1 + 0.19x_2 + 0.21x_3 + 0.06x_4 + 0.12x_5 + 0.12x_6 + 0.07x_7 + d_3^- = 1, \tag{13}$$

$$x_i \in \{0, 1\}, \forall i \in I, \tag{14}$$

$$d_1^-, d_2^+, d_3^- \geq 0. \tag{15}$$

To solve the above model together with the system constrains (given in Table 3), we iteratively solve linear programming (LP) models as follows:

1. First solve the LP model resulting from minimizing the first priority objective function d_1^- subject to all original constraints, that is, find a solution such that the EBIT under-achievement is as small as possible. Let EBIT^1 represent such value.

Table 1 Performance matrix (a_{ij})

Criteria (j)	Initiatives (i)						
	1	2	3	4	5	6	7
Sales costs	−78.000	810.614	−140.510	0.000	198.000	225.000	0.000
Penetration: Sales per CFU	27.516	45.869	25.476	25.476	30.574	30.574	25.476
Gross Margin. %	0.671%	0.171%	0.580%	0.000%	0.063%	−0.047%	0.000%
Number of staff	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Average costs per person	0.000	0.000		0.000	0.000	0.000	0.029
Staff costs/Sales, %	−0.044 %	−0.426 %	0.000 %	0.000 %	−0.110 %	−0.110 %	0.124 %
Rent, m2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cost per m2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Indirect cost/Sales, %	−0.110 %	−0.466 %	0.206 %	0.165 %	−0.274 %	−0.274 %	0.124%
Average receipt term	−0.617	−0.162	0.000	0.000	−1.530	−1.530	0.000
Average payment term	0.235	2.351	0.273	−0.120	−0.590	−0.669	0.000
Stock rotation in days	0.462	3.929	0.835	−6.353	−1.154	−1.310	0.000

Table 2 Local weights of the initiatives ($p_{l_{ij}}$)

	Criteria (j)											
Initiatives (i)	Sales per CFU	GM, %	Sales Costs	Staff N°	Av. costs per/p.	Staff costs/ Sales, %	Rent, m2	Costs per m2, %	Ind.Costs/ Sales, %	ART, days	APT, days	Stock rotation, days
	min in ↑15%	min in ↑1,5%	max in ↑7%	max in ↑1%	min in ↓8%	→ 9%	→ 11321	→ 22.50%	max in ↑ 1%	min in ↓3%	min in ↑3%	max in ↑7%
1	0.1090	0.3650	0.0810	0.1250	0.1540	0.0750	0.1430	0.1430	0.0950	0.1330	0.1610	0.0330
2	0.3670	0.1060	0.3930	0.1250	0.1540	0.3310	0.1430	0.1430	0.3510	0.1330	0.4810	0.4460
3	0.0407	0.3650	0.0810	0.1250	0.1540	0.0750	0.1430	0.1430	0.0540	0.0670	0.1480	0.0740
4	0.0407	0.0410	0.0360	0.1250	0.1540	0.0750	0.1430	0.1430	0.0540	0.0670	0.0580	0.0740
5	0.2010	0.0410	0.1870	0.1250	0.1540	0.2070	0.1430	0.1430	0.1750	0.2670	0.0290	0.1700
6	0.2010	0.0410	0.1870	0.1250	0.1540	0.1940	0.1430	0.1430	0.1750	0.2670	0.0290	0.1700
7	0.0407	0.0410	0.0360	0.2500	0.0770	0.0440	0.1430	0.1430	0.0950	0.0670	0.0950	0.0330

↑ increase ↓ decrease → keep on

2. Then, solve the LP model resulting from the minimization of the second priority objective function d_2^+ subject to all original constraints and the additional constraint

$$198x_1 + 239.39x_2 + 90.51x_3 - 40x_4 + 102x_5 + 75x_6 - 30x_7 \geq \mathrm{EBIT}^1.$$

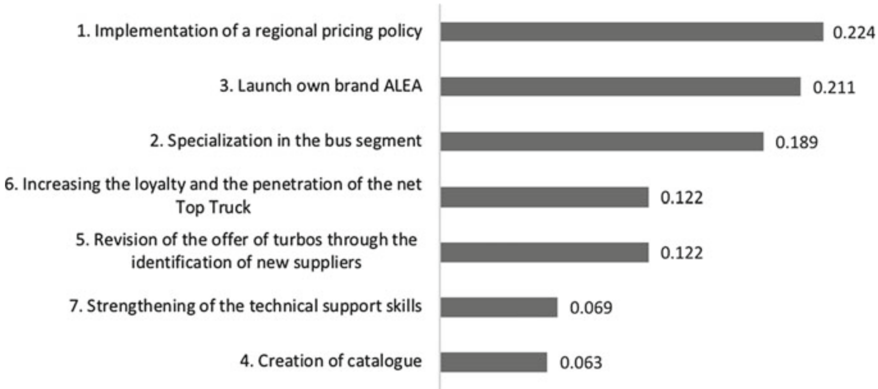


Fig. 8 Overall weights of initiatives (w_i)

Table 3 System constraints

Type of constraints	Financial indicator / KPI, j	Expected contribution of initiative i to financial indicator / KPI j, a_{ij}							Upper / lower bound for financial indicator / KPI j, t_j	
System constraints	EBITDA	198.000	239.386	90.510	-40.000	102.000	75.000	-30.000	≥	205.082
	Working Capital Needs	0.000	520.000	0.000	-300.000	0.000	0.000	0.000	≤	718.095
	Gross Margin	198.000	389.386	140.510	0.000	102.000	75.000	0.000	≥	659.498
	Indirect costs	0.000	150.000	50.000	40.000	0.000	0.000	30.000	≤	454.416
	Customers	0.000	300.000	0.000	0.000	0.000	0.000	0.000	≤	407.288
	Suppliers	0.000	200.000	0.000	0.000	0.000	0.000	0.000	≤	375.333
	Stock	0.000	420.000	0.000	-300.000	0.000	0.000	0.000	≤	686.140
	Sales	120.000	1,200.000	0.000	0.000	300.000	300.000	0.000	≥	1,905.475
	Sales costs	-78.000	810.614	-140.510	0.000	198.000	225.000	0.000	≤	1,245.977
	Staff costs	0.000	0.000	0.000	0.000	0.000	0.000	30.000	≤	214.861
	Other indirect costs	0.000	150.000	50.000	40.000	0.000	0.000	0.000	≤	239.555
	Penetration: Sales per UCF	27.516	45.869	25.476	25.476	30.574	30.574	25.476	≥	57.857
	Gross Margin, %	0.007	0.002	0.006	0.000	0.001	0.000	0.000	≥	0.004
	Number of staff	0.000	0.000	0.000	0.000	0.000	0.000	1.000	≤	1.000
	Average costs per person	0.000	0.000		0.000	0.000	0.000	0.029	≤	2.340
	Staff costs/Sales, %	0.000	-0.004	0.000	0.000	-0.001	-0.001	0.001	≤	0.002
	Indirect costs/Sales, %	-0.001	-0.005	0.002	0.002	-0.003	-0.003	0.001	≤	0.001
	Average Receipt Term	-0.617	-0.162	0.000	0.000	-1.530	-1.530	0.000	≤	-3.433
	Average Payment Term	0.235	2.351	0.273	-0.120	-0.590	-0.669	0.000	≤	2.075
	Stock Rotation in days	0.462	3.929	0.835	-6.353	-1.154	-1.310	0.000	≤	6.698

An optimal solution to this LP model minimizes the over-achievement of the CE ensuring, however, that the best possible value of EBIT is maintained. Let CE^2 denote the best CE obtained under such conditions.

3. Finally, the third priority objective function is minimized subject to all original constraints and to

$$198x_1 + 239.39x_2 + 90.51x_3 - 40x_4 + 102x_5 + 75x_6 - 30x_7 \geq \text{ EBIT}^1,$$
$$520x_2 - 300x_4 \leq \text{ CE}^2.$$

An optimal solution to this LP model is an optimal solution to our original problem. The one obtained is reported in Table 4.

Table 4 Optimal solution of the combined AHP-GP approach

Deviations	Initiatives	Goal Priorities		
		(P1) EBIT ↑	(P2) Cap. Emp ↓	(P3) Initiatives
$d_1^- = 0$ $d_2^+ = 0$ $d_3^- = 0.211$	1	198	0	1 X 0.224
	2	239	520	1 X 0.189
	3	0	0	0 X 0.211
	4	-40	-300	1 X 0.063
	5	102	0	1 X 0.122
	6	75	0	1 X 0.122
	7	-30	0	1 X 0.069
Total		544	220	0,789

4 Conclusions

In this work, we propose an AHP-GP approach to address the problem of selecting a strategic plan. The proposed model integrates the financial vision of the organization's strategy with a strategic plan. To the best of our knowledge, this is the first research work addressing such a problem using an AHP-GP integrated approach. The AHP is used to obtain relative importance weights for the initiatives, and the corresponding results are integrated within a lexicographic Goal Programming model that minimizes the deviations of the goals established for the financial indicators/KPIs found relevant.

The presented case study in the Nors Group provided the motivation for the research and served to illustrate how the suggested framework can be applied in practice. The case study solution is a strategic plan that consists of six initiatives. These initiatives were evaluated and then selected from an initial set of seven initiatives. The company's expert group found the final result satisfactory and consistent with their expectations.

Acknowledgements We acknowledge the financial support of NORTE 2020 - North Portugal Regional Operational Programme, ERDF European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme, and by National Funds through the Portuguese funding agency, FCT - Fundao para a Ciłncia e a Tecnologia within projects NORTE-01- 0145-FEDER-000020, POCI-01-0145-FEDER-031821, and PTDC/EEI-AUT/2933/2014 POCI-01-0145-FEDER-016858.

References

1. Arunraj, N., Maiti, J.: Risk-based maintenance policy selection using AHP and goal programming. Saf. Sci. **48**(2), 238–247 (2010)
2. Barbosa, L.C., Gomes, L.F.A.M.: Assessment of efficiency and sustainability in a chemical industry using goal programming and AHP. Proc. Comput. Sci. **55**, 165–174 (2015)

3. Bertolini, M., Bevilacqua, M.: A combined goal programming AHP approach to maintenance selection problem. Reliab. Eng. Syst. Saf. **91**(7), 839–848 (2006)
4. Charnes, A., Cooper, W.W.: Management Models and Industrial Applications of Linear Programming. Wiley (1961)
5. Bana e Costa, C.A., Vansnick, J.C.: A critical analysis of the eigenvalue method used to derive priorities in AHP. European J. Oper. Res. **187**(3), 1422–1428 (2008)
6. Ding, S.H., Kamaruddin, S.: Maintenance policy optimization literature review and directions. Int. J. Adv. Manuf. Technol. **76**(5–8), 1263–1283 (2015)
7. Gass, S.I.: A process for determining priorities and weights for large-scale linear goal programmes. J. Oper. Res. Soc. **37**(8), 779–785 (1986)
8. Goodwin, P., Wright, G.: Decision Analysis for Management Judgment, 3rd edn. Wiley, Chichester (2004)
9. Ho, W.: Combining analytic hierarchy process and goal programming for logistics distribution network design. In: 2007 IEEE International Conference on Systems, Man and Cybernetics, ISIC, pp. 714–719. IEEE (2007)
10. Hope, J., Fraser, R.: Beyond Budgeting: How Managers Can Break Free from the Annual Performance Trap. Harvard Business Press (2003)
11. Jamshidi, H.: An empirical application of goal programming and AHP in real estate assessment. J. Manag. Res. **9**(3), 1–11 (2017)
12. Kırış, Ş.: AHP and multichoice goal programming integration for course planning. Int. Trans. Oper. Res. **21**(5), 819–833 (2014)
13. Kruger, H., Hattingh, J.: A combined AHP-GP model to allocate internal auditing time to projects. ORiON **22**(1), 59–76 (2006)
14. Lin, T.C., Ho, H.P., Chang, C.T.: Evaluation model for applying an e-learning system in a course: an analytic hierarchy processmulti-choice goal programming approach. J. Educ. Comput. Res. **50**(1), 135–157 (2014)
15. Nors Group: Group performance book manual mydarwin. Technical Report, Federal Energy Regulatory Commission (2013)
16. Özcan, E.C., Ünlüsoy, S., Eren, T.: A combined goal programming-AHP approach supported with TOPSIS for maintenance strategy selection in hydroelectric power plants. Renew. Sustain. Energy Rev. **78**, 1410–1423 (2017)
17. Saaty, T.L.: How to make a decision: the analytic hierarchy process. Eur. J. Oper. Res. **48**(1), 9–26 (1990)
18. Sedzro, K., Marouane, A., Assogbavi, T.: Analytical hierarchy process and goal programming approach for asset allocation. J. Math. Financ. **2**(01), 96 (2012)
19. Trivedi, A., Singh, A.: A hybrid multi-objective decision model for emergency shelter location-relocation projects using fuzzy analytic hierarchy process and goal programming approach. Int. J. Proj. Manag. **35**(5), 827–840 (2017)
20. Wichapa, N., Khokhajaikiat, P.: Solving multi-objective facility location problem using the fuzzy analytical hierarchy process and goal programming: a case study on infectious waste disposal centers. Oper. Res. Perspect. **4**, 39–48 (2017)

Improving Inventory Management in an Automotive Supply Chain: A Multi-objective Optimization Approach Using a Genetic Algorithm

João N. C. Gonçalves, M. Sameiro Carvalho and Lino Costa

Abstract Inventory management represents a cornerstone inherent to any supply chain, regardless of industry type. Nevertheless, uncertainty phenomena related to demand and supply can induce overstock or even inventory stock-outs occurrences which, in turn, jeopardize one of the major principles of supply chain management: deliver the right product at the right place, at the right time and to the right cost. This situation may also be aggravated in automotive supply chains, due to their complexity in terms of entities involved. This research paper explores a multi-objective optimization model and applies it to a real industrial company, to address an inventory management problem. Moreover, a genetic algorithm is used to determine solutions corresponding to the order size and to a safety factor system. The obtained results are compared to the current strategy adopted by the company. At this point, the advantages and the drawbacks of the model implementation are assessed. Based on a set of logistic performance indicators, it is showed that the adoption of a smaller order size is potentially beneficial to the overall levels of inventory and to the value of inventory on–hand, without compromising the service level. Assertively, the proposed model reveals to be an useful tool to practitioners involved in automotive electronic supply chains.

Keywords Inventory management · Multi-objective optimization · Automotive supply chain management

J. N. C. Gonçalves (✉) · M. S. Carvalho · L. Costa
Algoritmi R&D Center, University of Minho, Braga, Portugal
e-mail: jncostagoncalves@gmail.com

M. S. Carvalho
e-mail: sameiro@dps.uminho.pt

L. Costa
e-mail: lac@dps.uminho.pt

J. N. C. Gonçalves · M. S. Carvalho · L. Costa
Department of Production and Systems, University of Minho, Braga, Portugal

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_10

1 Introduction

Within modern companies, the concept of Supply Chain Management (SCM) plays a key role in promoting their success, achieving their objectives and, above all, guaranteeing customer satisfaction. Without loss of generality, supply chain (SC) is commonly understood as a system composed by entities who develop symbiotic relationships by promoting both information and material exchanges from suppliers to the end-consumers. This concept was driven by the studies of Oliver and Webber in [1] (as cited in [2]), who have introduced the first-ever definition of SCM in 1982. Currently, SCM continues to be of the utmost importance in company performance at a time when firms are becoming increasingly more competitive. In fact, to put it in a nutshell, some of the primary goals of SCM relate to enhanced demand planning and forecast; lead time shrinkage; integration, coordination and information exchange between supply chain members; logistic service quality and overall minimization of logistic costs, namely purchasing, production, transportation and inventory costs (see [3] and the references cited therein). Let us focus now on inventory management. At this point, both the intense market competitiveness and the need to meet costumer requirements have triggered, amongst other features, uncertainty factors for companies, particularly related to the increase, reduction, cancellation or even forward-backward movements of customer orders.

Hence, in view of the cross-cutting impact that uncertainty phenomena have within SCM processes, several issues derived therefrom may lead to SC disruptions. Particularly, in what concerns the global automotive electronic sector, which entails large and complex supply chains, uncertainty factors are easily magnified and potentially lead to serious struggles in adapting production needs to customer demand. In this context, when demand variability levels are high, demand amplification effects might occur—leading to a vicious cycle characterized by further demand forecast inefficiencies and misaligned production plans. Consequently, demand amplification effects translate into serious supply risks, by increasing the volatility of supplier orders. Somewhere in this process, it is the logical corollary of these issues that SC will eventually suffer from overstock or even inventory stock–outs that, in turn, jeopardize one of the major principles of SCM: deliver the right product at the right place, at the right time and to the right cost. In this respect, inventory management is considered to be an important topic to contribute for the logistics effectiveness of supply chains [4], especially in what concerns the mitigation of supply chain disruptions [5].

The key role of inventory management is, in fact, enhanced in [4], in which the authors referred that ineffective inventory policies can entail significant and widespread damages to companies and supply chains. Moreover, by warning that there is no ready–made strategy to be followed when inventory approaches are adopted (see [6]), these authors stressed that inventory potentially can assist in mitigating supply chain disruptions. Besides, effective inventory policies can also avoid economic downturns (see, e.g., [7]).

In the light of the foregoing considerations, this paper focuses on the inventory reality of a real and major automotive electronic company that is affected by demand and supply lead time variability, which increase even more the levels of uncertainty. Furthermore, over time, this company has been facing an increasing and permanent pressure to meet customer's orders. To reach it, the managers tend to carry the lower possible level of inventory without losing sight of the SCM objectives mentioned above. This principle is positively correlated with the concept of *Just-In-Time* [8]. Therein too lies the complex task of establishing optimal inventory levels which not only provide the lowest stock costs, but also avoid both out-of-stocks and service level breakdowns. Thus, owing to the inherent complexity of automotive industries, strongly related to demand and supply variability, arises the necessity of modeling mechanisms capable to assist the decision–making regarding inventory control.

All in all, the problem that is intended to be tackled in this paper is that of inventory management, particularly related to determine how often and how much to order a certain product of a major automotive electronic company. Hence, based on the continuous–review control mechanism, a multi-objective optimization problem (MOOP) studied in [9] (updated from [10]) is considered in order to concomitantly minimize the expected annual costs for setup and holding inventory under lost sales, as well as the expected annual frequency of stock–out occurrences. For that, the Non-dominated Sorting Genetic Algorithm II (NSGA–II) [11] is used to provide the Pareto set and front solutions for the proposed problem. In addition, this work contributes to the current scientific literature by comparing two distinct logistics scenarios (pre–model implementation and post-model implementation) under real numerical data from a major automotive electronic company—demonstrating the industrial applicability of the proposed model.

The remainder of this paper is organized as follows. In Sect. 2, some relevant scientific literature on multi-objective optimization applied to inventory management is provided. Section 3 introduces and characterizes the MOOP, as well as some fundamental results associated to Pareto optimality concepts. Then, the solution procedure that forms the basis to the application of the optimization approach to the company's context is described in Sect. 4. Lastly, Sect. 5 depicts the numerical results and discussion, followed by concluding remarks and future lines of research in Sect. 6.

2 Related Work

The application of multi-objective optimization to inventory management has merited the attention of researchers over time. Agrell, in his seminal work [12], presented a model for inventory control aiming to aid the decision–making process in what concerns production and operations management. Concretely, the proposed model intended to minimize total costs, stock–out units and frequency of stock–out occurrences. It should be noted that, at that time, Agrell already stated the importance of concomitantly computing the order size and the safety factor using multi-objective optimization procedures. From a related perspective, Tsai and Chen [13] proposed a multi-objective model to compute the order quantity as well as the reorder point, which turns out to be a key factor in safety stock decisions. The ultimate goal of that model is threefold, namely concerned to the minimization of total inventory costs, average inventory levels and frequency of inventory shortages. Additionally, the authors designed new methodologies related to ranking and selection procedures in order to try to select the best candidate solution. This selection can turn out to be a challenging problem for decision–makers, due to the difficulty of ascertaining the trade–offs and repercussions associated with them. A threefold objective, similar to the one stated in [12], was also studied in [14], in which inventory management strategies are discussed by taking advantage of particle swarm optimization and genetic algorithms, under a multi-objective model related to hybrid backorder inventory. Alongside this study, other interesting research works concerning hybrid approaches to tackle inventory problems were also conducted (see, e.g., [15] and the references cited therein). Still considering the three objectives stated by Agrell [12], Tsou [10] published a multi-objective particle swarm optimization (MOPSO) technique to compute the set of efficient solutions of the reorder point and order size. Then, a ranking method procedure called TOPSIS is applied to provide a compromise solution which is congruent with the decision–maker preference.

In major companies, a common and effective approach to inventory management is the establishment of vendor–managed–inventory (VMI) policies. However, VMI environments require different assessment approaches when compared to traditional inventory policies. In this context, Liao et al. [16] proposed a multi-objective model in which, amongst other objectives, intends to infer the levels of safety stock to attain a certain service level.

Given the high variability and uncertain phenomena present within modern supply chains, especially in automotive ones, order crossover might occur, potentially entailing stock–outs or even lost sales. Aiming to tackle this issue, Srivastav and Agrawal [17] proposed a multi-objective mixture inventory system with multi-objective clustering selection (MOCS) algorithm. The approach used reveals to be effective in reducing inventory costs and increasing service levels. Regarding the model structure, it has the same objectives as in [10, 12] but they are modeled in a different way. The same authors proposed in [18] a multi-objective mixture inventory with similarities to the work published by Tsou in [10]. However, one of the novelties in [18] is the development of linear regression expressions to estimate relevant logistics indicators such as the safety factor or even optimal cost and lot size. From another perspective, a deterministic multi-objective inventory model for electronic components was proposed in [19] using a vector evaluated genetic algorithm. This model is suitable for single–items with deterioration features. Nonetheless, apart from the fact that the evolutionary approach used does not provide the Pareto front in a direct way, lead time is considered to be zero—which both constitute a considerable disadvantage that does not match the reality of electronic parts supply chain.

By noting also that inventory management is deeply linked with the logistics quality of suppliers, a multi-objective evolutionary algorithm was presented by Türk et al. in [20] to supplier selection and inventory planning. For that, the authors used three evolutionary algorithms (NSGA–II, SPEA2 and IBEA) and tuning approaches, concluding that the proposed approach does not compromise the trade–off between risk and cost by providing a set of solutions that fulfill it. A different approach is given in [21], in which fuzzy stochastic programming is applied to multi-objective modeling, under single–item context, in order to compute minimum stock levels and order quantities. Interestingly, this study was applied to a military context and ensured superior performance characteristics in extreme scenarios related to uncertain phenomena, when compared to the well–known fixed order quantity (r, Q) and service level (s, S) inventory models.

3 Multi-objective Optimization Problem

This section intends to introduce and describe the MOOP that forms the basis for all the mathematical analyses hereinafter presented. The MOOP analyzed herein is symbiotically related to a well known control mechanism called continuous–review (r, Q) (see, e.g., [22]), in which an order of size Q is placed to the supplier whenever stock levels fall below a threshold level r—called reorder point. In terms of model assumptions, let us consider the existence of a single product and that an order is received in full, after a pre–specified lead time L, and added to the existing stock. Additionally, the demand during lead time (DDLT) is considered to be normally distributed, with mean μ_{DDLT} and standard deviation σ_{DDLT}.

In the first instance, the MOOP considered was studied in [10], presenting three objectives, i.e.

$$\underset{Q,k}{\text{Minimize}} \quad C(Q,k) = ADQ^{-1} + hc\left(\frac{Q}{2} + k\sigma_{DDLT}\right) \tag{1}$$

$$\underset{Q,k}{\text{Minimize}} \quad N(Q,k) = DQ^{-1}\int_r^{\infty} \varphi(x)\,dx \tag{2}$$

$$\underset{Q,k}{\text{Minimize}} \quad S(Q,k) = DQ^{-1}\int_r^{\infty} (x-r)\varphi(x)\,dx \tag{3}$$

$$\textit{subject to} \quad 0 < Q \leqslant D, \tag{4}$$

$$0 \leqslant k \leqslant D\sigma_{DDLT}^{-1}\,, \tag{5}$$

Equation (1) refers to the minimization of the expected annual total costs $(C(Q, k))$. On the other hand, Eqs. (2) and (3) concern the minimization of the expected annual frequency of stock–out occurrences $(N(Q, k))$ and the expected annual number of items disrupted $(S(Q, k))$, respectively. Regarding the problem constraints, Eqs. (4) and (5) ensure that the order size (Q) should be positive and less or equal than the average annual demand (D), and that safety stock must be a non-negative quantity that would never exceed the average annual demand, respectively. See Table 1 for the description of the variables and parameters used in the above MOOP.

Table 1 Summary of the notation used in the MOOP proposed in [10]

Variable/Parameter	Description
C	Expected annual total costs
N	Expected annual frequency of stock–out occurrences
S	Expected annual number of items disrupted
φ	Probability distribution function of the demand within lead time
A	Ordering cost
D	Average annual demand
Q	Order size
L	Lead time (in days)
c	Unit item cost
h	Inventory carrying rate
k	Safety stock factor
r	Reorder point
μ_{DDLT}	Average demand during lead time
σ_{DDLT}	Standard deviation of demand during lead time

However, in a subsequent research study published in [9], the authors stated that the three objectives mentioned above are not conflicting among each other. Concretely, by noticing that the objectives traduced in Eqs. (2) and (3) are redundant, only differing in the associated unit of measure, these authors proposed a bi-objective optimization problem aiming to overcome the redundancy under lost sales associated with the above MOOP with three objectives. Thus, by dropping out Eq. (3) and rearranging Eq. (1), the resultant non-redundant bi-objective model is described as follows.

$$\underset{Q,k}{\text{Minimize}} \quad C(Q,k) = ADQ^{-1} + hc\left[\frac{Q}{2} + k\sigma_{DDLT} + DQ^{-1}\sigma_{DDLT}\Big(\varphi(k) - k(1-\Phi(k))\Big)\right] \tag{6}$$

$$\underset{Q,k}{\text{Minimize}} \quad N(Q,k) = DQ^{-1}\int_k^{\infty} \varphi(x)\,dx \tag{7}$$

$$\textit{subject to} \quad \sqrt{2AD(hc)^{-1}} \leqslant Q \leqslant D, \tag{8}$$

$$0 \leqslant k \leqslant D\sigma_{DDLT}^{-1}, \tag{9}$$

where Φ represents the cumulative distribution function of DDLT. At this point, following the work of [9], Eq. (6) traduces the minimization of the expected annual cost for setup and holding inventory under lost sales, and Eq. (7) pertains to minimize the expected annual frequency of stock–out occurrences. Concerning the inequality constraints, the only difference lies in Eq. (8), which imposes that the order size may not be less than the economic order quantity or greater than the average annual demand. It should be emphasized that the two objectives measured in Eqs. (6) and (7) are often perceived as the chosen trade–off in inventory management, notwithstanding the model does not consider the stock–out costs (see [9] and the references cited therein).

For the sake of relevance, before moving on to the application of the proposed multi-objective optimization model to a real industrial context in Sect. 4, let us provide some introductory remarks on Pareto optimality concepts that will be used in further sections.

Let $f(x)\colon \mathbb{R}^n \rightarrow \mathbb{R}^m$, defined as $f(x_1, \ldots, x_n) = (f_1, \ldots, f_m)$ with $m > 1$, be the function to be minimized over the n–dimensional space $\mathbb{R}^n$. Let us also consider $g(x)$ as the vector of $p \geqslant 1$ inequality constraints, defining the feasible region Ω. Thus, a minimization MOOP with m objectives and p constraints can be characterized as follows:

$$\text{Minimize} \quad f(x) = [f_1(x), f_2(x), \ldots, f_m(x)]^T \tag{10}$$

$$\textit{subject to} \quad x \in \Omega \tag{11}$$

$$\Omega = \{x : g_j(x) \leqslant 0\,, \quad j = 1, 2, \ldots, p\}\,. \tag{12}$$

Since the objectives conflict among each other, there is no single optimal solution for a MOOP but a set of efficient or non-dominated solutions known as *Pareto Optimal Set*. This lead us to the concept of Pareto dominance. Under this concept, for a minimization MOOP and two arbitrary n–dimensional decision variable vectors x and y, x is said to dominate y $(x \prec y)$ iff $f_i(x) \leqslant f_i(y)$, $\forall i \in \{1, \ldots, m\}$ and $\exists i \in \{1, \ldots, m\}$ such that $f_i(x) < f_i(y)$. Therefore, x and y are said to be incomparable if neither x nor y dominates the other and they are not equal.

On the basis of these Pareto dominance concepts, a solution x is Pareto optimal if it is not dominated by any other solution on the feasible space. Pareto optimal solutions are incomparable to each other and represent different trade–offs in terms of the objectives. Thus, the main aim of a multi-objective optimization algorithm lies in finding an approximation to the Pareto optimal set. The set of the images by f of all points in the Pareto optimal set is called *Pareto front*. For more details on this subject, the reader is referred to [23].

4 Application of the MOOP to a Real Industrial Case Study

In order to test the industrial applicability of the MOOP previously described, real data related to a product of a major automotive electronic company is used. All in all, this section aims to introduce the solution procedure that will form the basis for the numerical comparisons between the results derived from the model application to company's reality and the ongoing approach used by it.

As alluded to earlier, the numerical analyses further presented have been based on a single product of the concerned company (hereinafter referred as *Product 1*). In this context, real logistics data relating to year 2017 was collected and compiled. At

Table 2 Summary of the parameters for the multi-objective optimization model application

Product ID	A	c	D	h	σ_{DDLT}	μ_{DDLT}	Q
1	25	5.16	48.28	0.5	43	449	2000

this point, for $L = 12$, statistical work showed that demand during lead time approximately follows a Gaussian distribution $N(449, 43)$. The list of model parameters for the appropriate application of the optimization model is summarized in Table 2. Regarding the parameter A, it was not easy to properly assess and estimate all the costs related to create and process an order to a certain supplier (e.g., costs of processing purchase requisitions or even the labor costs involved when the products are received), as well as in respect of transportation. Hence, all the parameter values are real, with the exception of A that, due to the difficulty of measuring it within an industrial context, was estimated based on the real features of *Product 1*.

It should be emphasized that despite of many scientific literature assume that demand distribution is known, (statistically speaking) long–term demand is normally forecasted using historical data. In this respect, the adoption of inefficient forecasting models by companies can entail serious logistics issues, mainly related to overstock or, eventually, stock–out occurrences and service level breakdowns. This fact motivates the need to conceive a new methodology to control inventory levels over time.

Aiming to simulate the performance of the MOO model on the inventory management of *Product 1*, two scenarios related to pre and post–implementation of the model are designed. For both simulation scenarios, the assessment and analysis of the obtained results are based on the following key performance indicators measured on a yearly basis:

- Inventory levels
- Inventory on–hand costs
- Number of placed orders
- Turnover rate
- Coverage rate

The reason for the selection of these indicators relates to both data availability and reliability provided by the organization. Regarding the numerical implementation, the search for optimal solutions (in the Pareto sense) within complex search spaces requires algorithms that avoid some drawbacks of traditional approaches (e.g., weighted method and ε–constraint method), namely the necessity of more than one iteration to obtain an approximation to only one Pareto optimal solution or even the insensitivity in what concerns the shape of the Pareto front [23]. Thus, this paper takes advantage of an evolutionary approach for multi-objective optimization called *Elitist Non–dominated Sorting Genetic Algorithm* (NSGA–II). This algorithm is considered to be a computationally efficient approach that, in each iteration (also called generation), implements *elitism*. In other words, when dealing with two points, the one characterized by the lower rank, in terms of non–domination, is chosen. However, it might occur that two points have the same rank. In these cases, crowding distance comes into play to decide which one has the higher distance (see [11] for more details about this algorithm).

The computation procedure of the MOO model using NSGA–II was performed in the programming language `R` [24], by taking advantage of the function `nsga2` to provide the set of non–dominated solutions. In this process, the number of generations is set to be 100, under a population size of 140.

5 Numerical Results and Discussion

This section reports the numerical results derived after the model application to the company context, using the parameters available on Table 2. Concretely, a comparison of the results obtained using the MOO model and the current approach adopted by the company is here provided.

Figure 1 portrays the sets of all non–dominated solutions obtained. For the Pareto front plot, it is recalled that the x–axis (f_1) relates to the expected annual cost for

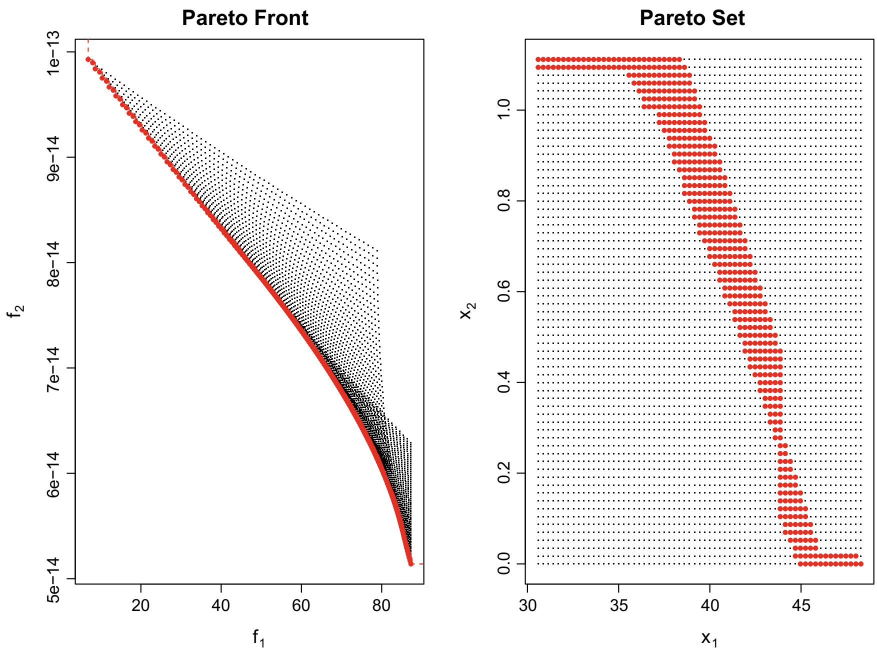


Fig. 1 Pareto front for the trade–off between the objectives C (f_1) and N (f_2) and Pareto optimal set for Q (x_1) and k (x_2)

setup and holding inventory under lost sales and that the y–axis refers to the expected frequency of stock–out occurrences (f_2).

Firstly, note that the Pareto front suggests that there is no linear or *quasi*–linear dependency between the two objectives characterized by f_1 and f_2. Moreover, it can be inferred that higher levels of expected annual costs for setup and holding inventory (f_1) are required to obtain lower expected annual frequencies of stock–out occurrences (f_2). Naturally, lower levels of expected annual costs are associated to a higher number of expected annual frequencies of stock–outs. From this, one can conclude that a sharp reduction in the values of f_1 potentially leads to higher chances of stock–out occurrences and, consequently, production line stoppages and further losses in production and sales volumes. Conversely, Pareto optimal set suggests that higher safety factor levels (x_2) are related to lower values of Q (x_1). This fact is natural since smaller order sizes are related to higher numbers of placed orders to suppliers which, in turn, suggests high levels of safety factor to prevent stock–out occurrences.

Hence, in accordance with the foregoing and based on the output results from the application of NSGA–II, two scenarios are now compared. Note that due to space limit, the full list of Pareto solutions is not listed here. Nevertheless, Table 3 presents 15 Pareto solutions from 140 found by NSGA–II.

The above analyses together with the efficient solutions obtained were discussed with the managers of the company and the adoption of $Q = 37$ was considered. The company managers chose this value for Q since it reveals a good compromise solution to balance the minimization trade–off between the two proposed objectives. At this point, top managers did not express preferences of one objective over another. This solution is highlighted in bold type on Table 3. Between the solutions 7 and 10, the managers have opted to the latter due to the smaller stock–out frequency. It is worth pointing out that a technique to rank the obtained non–dominated solutions (e.g., TOPSIS) was not applied since for this bi-objective model Pareto set and front are self–explanatory and already provide sufficient information to meet the business requirements of the company.

On this basis, a simulation process was undertaken to assess the logistics repercussions of adopting $Q = 37$ instead of $Q = 2000$ on the company inventory records. These scenarios aim to simulate the benefits of the model implementation in the company, by confronting the results derived from its application with the current approaches. The summary of the comparison between them is presented in Table 4.

Regarding the comparisons in what concerns the recorded levels of inventory, both in terms of units as well as in value, Fig. 2 allows to infer that, if the company had used the proposed model as a supporting tool for the decision–making, both the inventory levels in units (Fig. 2a) and the recorded value of inventory on–hand (Fig. 2b) would be significantly lower. In fact, when using the proposed model, the new values for stock in units and value are almost three times lower than the ones recorded with the current methodology used by the company.

Table 3 List of 15 (out of 140) non–dominated solutions obtained from NSGA–II

# Instance	$\approx Q$	k	$C(Q,k)$	$N(Q,k)$
1	48	5.188913×10^{-6}	87.287491	5.140280×10^{-14}
2	34	1.109468	25.412899	9.011910×10^{-14}
3	35	1.105876	33.446853	8.627815×10^{-14}
4	32	1.110666	13.819280	9.580706×10^{-14}
5	34	1.072007	28.447386	8.890523×10^{-14}
6	42	0.4765033	75.292520	6.417178×10^{-14}
7	37	1.029903	43.486233	8.169080×10^{-14}
8	34	1.107318	26.438539	8.963706×10^{-14}
9	43	0.5554719	76.171123	6.366719×10^{-14}
10	**37**	**1.018627**	**47.151751**	$\mathbf{7.991840 \times 10^{-14}}$
11	33	1.104654	21.743920	9.194887×10^{-14}
12	47	4.880267×10^{-3}	85.973450	5.334644×10^{-14}
13	33	1.106130	22.330156	9.164834×10^{-14}
14	36	1.110429	35.573982	8.525269×10^{-14}
15	40	0.9413747	59.026523	7.406437×10^{-14}

Table 4 Overall performance indicators for *Product 1* with and without MOO over 2017

Time period	Inventory levels		Inventory on–hand (in €)		Number of placed orders	
	With MOO	Without MOO	With MOO	Without MOO	With MOO	Without MOO
January	65 199	104 636	300 483	482 236	2	2
February	62 161	171 181	286 481	776 410	3	0
March	41 527	229 566	191 385	961 629	3	2
April	28 788	224 284	132 675	939 502	2	0
May	39 570	209 345	182 366	876 924	3	0
June	44 433	165 985	204 778	695 292	3	0
July	52 147	142 095	240 330	595 217	3	1
August	56 281	173 166	259 382	725 370	3	2
September	73 471	176 588	338 606	739 706	2	1
October	71 863	81 824	331 195	342 751	3	1
November	84 517	107 223	389 513	449 145	3	2
December	79 068	127 692	364 401	534 887	1	1
Total	**699 025**	**1 913 585**	**3 221 597**	**8 119 069**	**31**	**12**

With regard to the number of placed orders to suppliers, Fig. 3 suggests that, by using multi-objective optimization, the number of placed orders required to fulfill demand is higher than without using it. However, the necessity of placing a higher number of orders to suppliers can be overtaken by noticing that the total ordering

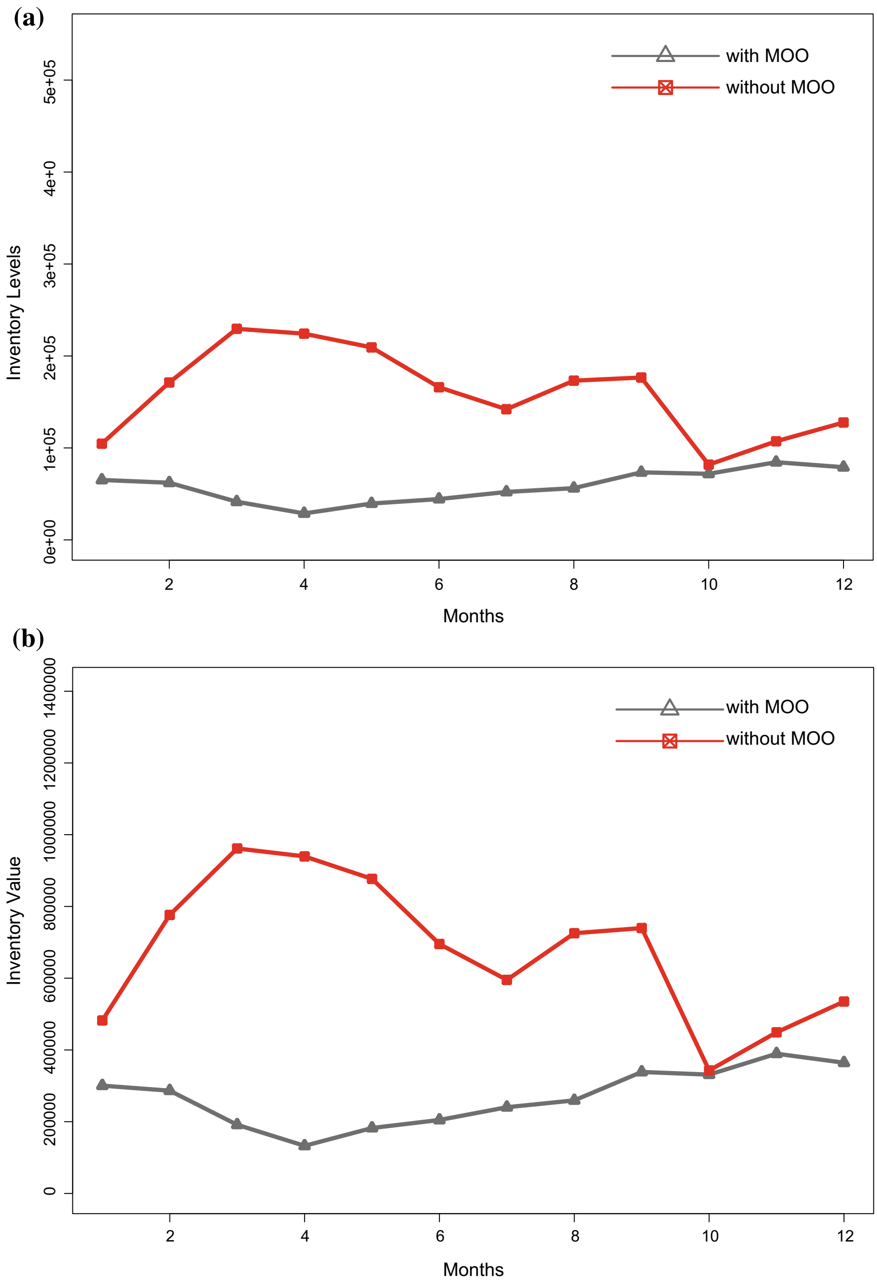


Fig. 2 Inventory levels and inventory value for *Product 1* with and without MOO over 2017

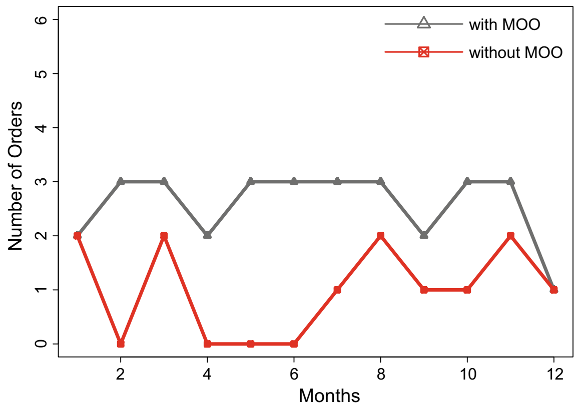


Fig. 3 Number of placed orders to suppliers for *Product 1* with and without MOO over 2017

Table 5 Turnover and coverage rates for *Product 1* with and without MOO in 2017

Time period	Scenarios	
	With MOO	Without MOO
Turnover rate	8.03	2.93
Coverage rate	0.12	0.34

costs derived by the MOO implementation appear to be derisory in comparison to the inventory levels before and after the model implementation. Indeed, this statement is underpinned by the fact that the inventory levels produced with the multi-objective model are smaller. From another perspective, the adoption of small order sizes Q can contribute to minimize crossover order events (see [17]).

All in all, the current strategy adopted by the company produces significantly higher values of stock costs on–hand, confirming the view that the application of the MOO model is beneficial for the company.

On the other hand, also note that turnover rate is higher when the model is applied (see Table 5). This means that attaining high levels of stock profitability is easier with the adoption of the proposed model in detriment of the current approach adopted by the company. Nonetheless, from higher levels of profitability comes the greater possibility of stock–out occurrences since the levels of stock are smaller—as one could expect, higher values of turnover translate into smaller rates of coverage.

6 Conclusions and Future Work

In this research work, a multi-objective optimization model is applied, in an exploratory way, to a major automotive supply chain, purposing to improve inventory management by, concomitantly, determining how often and how much to order a given product and minimizing both annual costs and the expected frequency of stock–out occurrences. For that, several logistic performance indicators are used to assess the benefits of the model implementation. Moreover, by taking advantage of evolutionary computation, NSGA–II is used to provide the set of non–dominant solutions for the proposed multi-objective optimization problem. In this regard, a Pareto trade–off curve between the expected annual cost for setup and holding inventory under lost sales and the expected annual frequency of stock–out occurrences is generated, providing some guidelines to industrial practitioners.

As main results, with the adoption of smaller order sizes Q, it stands out a remarkable reduction in both inventory stock levels (63%) and inventory on–hand (60%), when compared to the current strategy adopted by the company. At this point, the turnover rate resultant from the model application increases significantly. However, the order size Q required to achieve these targets imposes an increase in the number of placed orders, notwithstanding the overall benefits inherent to its adoption. Besides, the adoption of smaller order sizes is not always accepted by associated suppliers, since they often impose minimum order quantities to the companies.

Regarding the caveats related to the implementation of the proposed model in the concerned company, it is underlined not only the difficulty inherent to the collection and compilation of reliable data, but also to the estimation process of certain model parameters. These obstructions hampered the simulation processes and their ability to produce other kind of analyses and comparisons.

As further research work, it is intended to improve the complexity of the model, as well as test and compare the model dynamics for products with different characteristics using different evolutionary computation strategies.

Acknowledgements The authors would like to acknowledge the comments and suggestions from the two reviewers, which improved the quality of the paper. This work has been supported by ALGORITMI R&D Center, under COMPETE: POCI-01-0145-FEDER-007043 and FCT–Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013.

References

1. Oliver, R.K., Webber, M.D.: Supply–chain management: logistics catches up with strategy. In: Christopher, M. (ed.) Outlook, Booz, Allen and Hamilton Inc. Reprinted 1992. Chapman Hall, London (1982)
2. Barbosa-Póvoa, A.P., da Silva, C., Carvalho, A.: Opportunities and challenges in sustainable supply chain: an operations research perspective. Eur. J. Oper. Res. **268**(2), 399–431 (2018)
3. Vlahakis, G., Apostolou, D., Kopanaki, E.: Enabling situation awareness with supply chain event management. Expert Syst. Appl. **93**, 86–103 (2018)
4. Boone, C.A., Craighead, C.W., Hanna, J.B., Nair, A.: Implementation of a system approach for enhanced supply chain continuity and resiliency: a longitudinal study. J. Bus. Logist. **34**(3), 222–235 (2013)
5. Sarkar, S., Kumar, S.: A behavioral experiment on inventory management with supply chain disruption. Int. J. Prod. Econ. **169**, 169–178 (2015)
6. Lee, H.L.: Aligning supply chain strategies with product uncertainties. Calif. Manag. Rev. **44**(3), 105–119 (2002)
7. Singh, S., McAllister, C.D., Rinks, D., Jiang, X.: Implication of risk adjusted discount rates on cycle stock and safety stock in a multi-period inventory model. Int. J. Prod. Econ. **123**(1), 187–195 (2010)

8. Sugimori, Y., Kusunoki, K., Cho, F., Uchikawa, S.: Toyota production system and kanban system materialization of just-in-time and respect-for-human system. Int. J. Prod. Res. **15**(6), 553–564 (1977)
9. Tsou, C.-S., Wu, B.-H., Lee, Y.-H.: Bi–objective inventory management through evolutionary multi–objective optimization. In: 2010 International Conference on Economics, Business and Management, IPEDR, vol. 2 (2011)
10. Tsou, C.-S.: Multi-objective inventory planning using MOPSO and TOPSIS. Expert Syst. Appl. **35**(1–2), 136–142 (2008)
11. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)
12. Agrell, P.J.: A multicriteria framework for inventory control. Int. J. Prod. Econ. **41**(1–3), 59–70 (1995)
13. Tsai, S.C., Chen, S.T.: A simulation-based multi-objective optimization framework: a case study on inventory management. Omega **70**, 148–159 (2017)
14. Srivastav, A., Agrawal, S.: Multi-objective optimization of hybrid backorder inventory model. Expert Syst. Appl. **51**, 76–84 (2016)
15. Mahnam, M., Yadollahpour, M.R., Famil-Dardashti, V., Hejazi, S.R.: Supply chain modeling in uncertain environment with bi-objective approach. Comput. Ind. Eng. **56**(4), 1535–1544 (2009)
16. Liao, S.-H., Hsieh, C.-L., Lai, P.-J.: An evolutionary approach for multi-objective optimization of the integrated location-inventory distribution network problem in vendor-managed inventory. Expert Syst. Appl. **38**(6), 6768–6776 (2011)
17. Srivastav, A., Agrawal, S.: Multi-objective optimization of mixture inventory system experiencing order crossover. Ann. Oper. Res. 1–18 (2018)
18. Srivastav, A., Agrawal, S.: Multi-objective optimization of a mixture inventory system using a MOPSO-TOPSIS hybrid approach. Trans. Inst. Meas. Control **39**(4), 555–566 (2017)
19. Yadav, A., Mishra, R., Kumar, S., Yadav, S.: Multi objective optimization for electronic component inventory model & deteriorating items with two-warehouse using genetic algorithm. Int. J. Comput. Technol. Appl. **9**(2), 881–892 (2016)
20. Türk, S., Özcan, E., John, R.: Multi-objective optimisation in inventory planning with supplier selection. Expert Syst. Appl. **78**, 51–63 (2017)
21. Bean, W.L., Joubert, J.W., Luhandjula, M.: Inventory management under uncertainty: a military application. Comput. Ind. Eng. **96**, 96–107 (2016)
22. Silver, E.A., Pyke, D.F., Peterson, R.: Inventory Management and Production Planning and Scheduling, vol. 3. Wiley, New York (1998)
23. Miettinen, K.: Nonlinear Multiobjective Optimization. International Series in Operations Research and Management Science. Kluwer Academic Publishers, Dordrecht (1998)
24. Team, R.C.: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2017)

Consistent Consolidation Strategies in Grocery Retail Distribution

Sara Martins, Pedro Amorim and Bernardo Almada-Lobo

Abstract In the food retail sector, maintaining the food quality across the supply chain is of vital importance. The quality of the products is dependent on its storage and transportation conditions and this peculiarity increases the supply chain complexity relatively to other types of retailers. Actually, in this industry there are three types of food supply chains: frozen, chilled and ambient. Moreover, food retailers run different store formats, of different sizes, assortments and sales volume. In this study we research the trade-off between consolidating a range of products in order to perform direct deliveries to the stores versus performing separate delivery routes for products with different transportation requirements. A new consistency dimension is proposed regarding the periodicity that a consolidation strategy is implemented. The aim of this paper is to define a consolidation strategy for the delivery mode planning that allows to smooth the complexity of grocery retail operations. A three-step approach is proposed to tackle a real size problem in a case-study with a major Portuguese grocery retailer. By changing the consolidation strategy with a complete consistent plan the company could reach annual savings of around 4%.

Keywords Grocery retail · Consolidation · Distribution

1 Introduction

Since the 90's the retail sector has been evolving, with retailers controlling and managing the flow of products and adopting different distribution strategies to stay competitive. This paper focuses on the grocery retail sector, which has peculiarities

S. Martins (✉) · P. Amorim · B. Almada-Lobo
INESC TEC and Faculty of Engineering, University of Porto,
Rua Dr. Roberto Frias, s/n, 4600-001 Porto, Portugal
e-mail: sara.martins@fe.up.pt

P. Amorim
e-mail: pamorim@fe.up.pt

B. Almada-Lobo
e-mail: almada.lobo@fe.up.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_11

that increase the complexity of supply chain planning due to the management of products with distinct temperature requirements, such as frozen, chilled and ambient products. As 1/3 of the global food produced is lost or wasted [5], maintaining the food quality across the supply chain is of vital importance. Food quality degradation depends on environmental conditions of storage and transportation facilities [9]. In order to guarantee the products quality, grocery retailers have to be manage properly the required temperature range when distributing orders from the DCs. The same reasoning holds for the transportation of products. From hereafter, products with the same temperature requirements are considered as a group that is managed in the same way.

To be closer to customers and promote a convenient shopping experience, new stores are open every year. Nowadays, grocery retailers manage a range of stores from hypermarkets to convenience stores that hold distinct characteristics such as, different levels of net sales and product assortments. Hence, a mix of orders of different sizes and products are required to be distributed.

In such context, with multiple products flowing from distinct facilities to stores with heterogeneous needs, the delivery mode planning is a complex and challenging distribution decision. This planning decision is at the core of master planning [6] and can be analyzed in two levels: strategic and tactical/operational [10]. This work tackles the tactical/operational level of the delivery mode planning that decides how the products should flow inside the retailers distribution network. At this planning level, a focus on the distribution process is required in order to define the consolidation and transportation rules between the distribution centers (DCs) and the stores [10]. Replenishment and shipment consolidation are within the most important issues in today's supply chain planning [4].

In order to smooth the grocery retailing operations, i.e. reduce the disruption that consolidation synchronization can induce, a certain level of consistency in the strategies implemented needs to be accomplished. [7] present a survey on consistency in vehicle routing problems. The authors discuss consistency measured in 3 dimensions: arrival time (customer deliveries around the same time of the day), person-oriented (deliveries performed by the same driver and driver visiting the same area), and delivery consistency (defined interval between deliveries or fixed order sizes). In this work it is proposed a fourth dimension related with consolidation consistency, focusing on the supply side. Instead of defining every day the consolidations that will have to be performed and synchronized, the consolidations are repeated with a certain periodicity.

Overall, this paper studies the consolidation strategies grocery retailers can use to enhance their distribution process. Moreover, it proposes for the first time the creation of consistent plans for the consolidation strategies. The research addresses the particular case of grocery distribution using multi-compartment vehicles (MCVs) with longitudinal divisions, which impose some restrictions on the operations. This type of vehicles have been neglected in the literature despite their great utilization in practice. Analysis are performed based on a case-study from an European grocery retailer.

The remainder of this paper organized as follows. In Sect. 2 the possible consolidation strategies and the consistency levels proposed are discussed, followed by the problem description in Sect. 3. Section 4 presents the three-step approach proposed and Sect. 5 analyses the results achieved. Finally, Sect. 6 provides conclusions and further research directions.

2 Consistent Consolidation Strategies

In order to guarantee the quality of the products, grocery retailers have to manage properly the required temperature range when distributing orders from the DCs. The retailer's network can be composed of different DCs that operate as storage or cross-docking facilities, receiving the products from the suppliers and afterwards, sending them to the stores. Usually, each DC operates independently and is responsible for the set of allocated products [8]. Moreover, some retailers integrate production centers of fresh products in their network in order to have more control in the quality and availability of the products.

Orders are placed by each store for each product to a specific DC. In order to achieve full-truck load (FTL) different orders can be consolidated. This consolidation can be made by combining orders from different products to the same store (product consolidation) and/or orders from different stores (spatial consolidation) [3]. However, different characteristics and requirements have to be taken into consideration regarding the consolidation strategy. For example, in products consolidation, products that are not allocated to the same DC require further procedures, such as cross-docking or milk-run. Furthermore, the joint delivery of products makes it easier to achieve an FTL for one store, but it could lead to unloading problems at the store site, if the receiving area is not prepared to receive a delivery with a large order size. Spatial consolidation is defined based on a routing plan, with the goal of minimizing total transportation costs. Overall, distinct consolidation strategies have different advantages and disadvantages that are summarized in Table 1.

A particular feature of product consolidation in grocery retailing is the requirement for the use of multi-compartment vehicles (MCVs). Some works on grocery distribution have already analyzed the benefits of using MCVs against single-compartment vehicles [2, 12]. However, these works were focused on the application of MCVs, which allow for transverse divisions (here called flexible MCVs), making it possible to jointly distribute products to a route of stores. Nevertheless, other types of MCVs can be used, as those with longitudinal divisions (here called simple MCVs) that are less expensive than the transverse ones, but impose more restrictions to the distribution [10]. Hence, although there is a trend for grocery retailers to use flexible MCVs, due to their costs, many retailers still prefer to use the simple MCVs.

The consolidation of products from distinct temperatures imposes additional challenges to the synchronization of DCs and transportation operations, increasing the risk of delays [11]. When products are allocated in the same location but in different DC zones, a vehicle performs a milk-run to each zone and all orders have to be

Table 1 Advantages and disadvantages of distinct consolidation strategies

Consolidation	Advantages	Disadvantages
Spatial	• Risk of delays is lower • MCVs are not required	• Difficult to achieve FTL for one store • Longer routes might be required to achieve FTL • Stores receive the different products separately
Product	• Stores receive the different products together • Earlier supplies to the stores • Easier to achieve FTL for one store	• Store's unloading docks can be a bottleneck • MCVs are required • Additional handling of the products might be required (in case of cross-docking)

prepared to avoid idle time. If cross-docking is necessary to consolidate products from distant facilities, an additional space in the DC is required to perform the operation, which should be quick.

3 Problem Description

The aim of this work is to define consolidation strategies to be followed with a given consistent periodicity. Retailers can decide upon the level of consolidation consistency they desire to use and best suits their operation. Four levels of consolidation consistency are proposed: full consistency, monthly consistency, week day consistency or absent. The full consistency imposes the performance of the same type of consolidation every day, which means the same strategy can be maintained despite annual seasonality, simplifying the operational processes that are recurrently repeated. In the monthly consistency, the consolidation strategy changes every month, and increases the operational complexity. The consistency based on the week day also involves some operational complexity but, as the operational activities are already commonly planned per week days due to the delivery patterns, the complexity is smaller than for the monthly consistency. Finally, the absent consistency is when the retailer decides every day the consolidations that will perform.

Simple MCVs are considered in this problem, which only allow to consolidate products from different temperatures if a direct delivery (single drop route) is performed. Therefore, a spatial and product consolidation for the same trip is not allowed. While spatial consolidation is frequently used in global distribution, product consolidation needs to be carefully analyzed, specially in grocery distribution, due to the operation requirements. In this problem setting, the aim is to study if it is preferable to perform spatial consolidation for each product separately or perform a product consolidation for each individual store. In the first case, delivery routes of stores are defined for each product, and in the latter direct deliveries to each store with a group of products is established. Hence, the overall problem has to consider the vehicle routing problem. Moreover, cross-docking of products from distant DCs is also analyzed. Due to quality control, it is assumed that there is a main DC where the cross-docking operation is performed. This operation has a warehouse cost associated to the additional material handling required. It is considered that the DCs have limited capacity to perform the operations.

The distribution is performed using an heterogeneous fleet of vehicles, distinct in terms of capacity. Some stores have restrictions on the type of vehicle that visits them, e.g. vehicles with trailers are forbidden in the middle of villages and therefore the stores cannot be visited by them. The delivery days of each product to each store are defined by the delivery pattern and stores have pre-defined time-windows to be fulfilled.

To summarize, this problem can be characterized as a heterogeneous fleet (HF), site dependent (SD), time-window (TW), consolidation consistent (C), periodic (P), vehicle routing problem (VRP) with cross-docking (CD).

4 Methodology

Since this problem is an extension of the vehicle routing problem, which is known to be NP-hard, with a set of additional operational constraints, in general, it is not possible to solve the problem exactly. A three step approach is proposed in this paper, dividing the problem in three sub-problems that are solved sequentially. The first two steps are used to define a store profile, i.e., identify the best stores for product consolidation that satisfy the operational requirements. In the third step, the overall solution cost is evaluated. Figure 1 gives an overview of the methodology. A further description is provided in the next subsections.

1. **Consolidation Strategy**
   In the first step, the stores having promising conditions to achieve near FTLs due to product consolidation are identified. A mixed-integer programming model is proposed at this stage. The stores that do not fulfill the requirements for consolidation are put aside for posterior routing (S1, in Fig. 1).
2. **Distribution Centers Capacity**
   In the second step, the DC's capacity to perform the consolidations required for each of the promising stores (provided by step 1) is studied. Some stores are put aside at the beginning of this stage due to operational constraints found in a pre-processing phase (S2, in Fig. 1). As the DC might not have enough capacity to perform all consolidations, an heuristic approach is designed to select the consolidation strategies that should be implemented (S4, in Fig. 1). The stores not selected are put aside for posterior routing (S3, in Fig. 1).
3. **Solution Evaluation**
   Finally, the overall solution is evaluated based on three cost components, namely the transportation cost of supplying directly the stores with product consolidation,

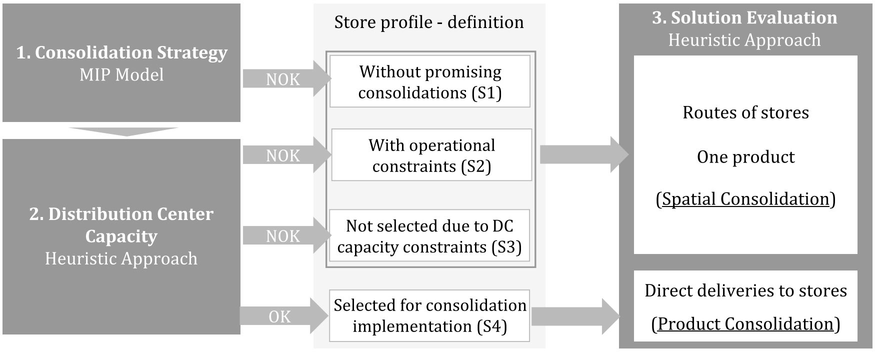


Fig. 1 Three-step approach scheme proposed for the consistent consolidation problem of grocery retail distribution (Note: NOK means not ok)

the transportation and warehouse cost of moving the products to the consolidation DC, and the transportation cost of the stores putted aside for separate routing of products. The latter is quantified by means of a meta-heuristic approach proposed in the literature for rich vehicle routing problems.

4.1 Consolidation Model

As mentioned before, a mixed-integer programming model is proposed to identify for which stores it is possible to achieve near FTL by consolidating different products. The objective is to maximize the number of direct deliveries to the store, taking into consideration the operational characteristics. This sub-problem can be defined as follows.

A retailer distributes a set of products $p \in P$ to a set of stores S, according to their delivery pattern. Therefore, the stores do not receive the full set of products every day. Let D be the set of days in the planning period and Q_{ds}^{p} the quantity of product p to be delivered to store s at day d, with $Q_{ds}^{p} = 0$ if the day d does not belong to the delivery pattern of product p for store s.

Direct deliveries to a store are only acceptable if a minimum utilization (U) of the vehicles can be accomplished, otherwise separate route of stores for each product could be more beneficial. The consolidation of products to a store might result in an overall quantity slightly bigger than a vehicle capacity. In practice, when the difference is very small, the surplus is left at the DC for posterior supply (back-orders). However, these situations need to be controlled. To avoid an excess of back-order situations, the number of times it can happen to a store has to be bounded by an upper limit that depends on the product (L^{p}). The binary variable B_{ds}^{p} is used as an auxiliary variable to determine if store s has surplus of product p at day d or not.

Depending on the level of consistency to be implemented, let T be the set of periods to be consistent, e.g. in case of a week day consistency T represents the days of the week, and in a monthly consistency it represents the months. Further, let the subset $D_t \subset D$ be the days that belong to period t.

In order to simplify the model, let parameter Q^p_{ds} represent the proportion of the vehicle capacity used by the quantity of product p to be delivered to store s at day d. Since we are considering a site-dependent heterogeneous fleet of vehicles, the proportion of an order for store s is determined having as base the maximum capacity of a vehicle eligible to visit store s. For instance, an order of 11 pallets to be delivered to a store that can be visited by a 33 pallet vehicle as an associated quantity of $Q = 0.333(3)$.

The problem comprises two decision variables. The binary decision variable X^p_{ds} is set to 1 if product p is consolidated for store s at day d, 0 otherwise. This decision variable dictates the products to be consolidated, i.e. determines for each store which products should be consolidated and delivered directly on each day. For example, if store 1 on day 1 should consolidate products 1 and 3 and deliver product 2 in a single product route, the correspondent variables will be set as $X^1_{11} = X^3_{11} = 1$ and $X^2_{11} = 0$. To control the direct deliveries performed, the integer decision variable F_{ds} represents the number of FTLs for store s at day d. The consolidation model can be formulated as follows:

$$max \sum_{d \in D} \sum_{s \in S} F_{ds} \tag{1}$$

$$\sum_{p \in P} (X^p_{ds} \cdot Q^p_{ds}) \geq F_{ds} \cdot U \qquad \forall s \in S, d \in D \tag{2}$$

$$\sum_{p \in P} (X^p_{ds} \cdot Q^p_{ds}) \leq F_{ds} \cdot U + 1 \qquad \forall s \in S, d \in D \tag{3}$$

$$X^p_{ds} \geq X^f_{ds} \qquad \forall p, f \in P : p \neq s \in S, d \in D \tag{4}$$

$$B^p_{ds} \geq X^p_{ds} \cdot Q^p_{ds} - F_{ds} \qquad \forall p \in P, s \in S, d \in D \tag{5}$$

$$\sum_{d \in D} B^p_{ds} \leq L^p \qquad \forall p \in P, s \in S \tag{6}$$

$$X^p_{ds} = X^p_{ys} \qquad \forall p \in P, s \in S, t \in T, d, y \in D_t : d \neq y \tag{7}$$

$$X^p_{ds} \in \{0, 1\}, F_{ds} \in Z^+_0 \qquad \forall p \in P, s \in S, d \in D \tag{8}$$

The objective function aims to maximize the number of FTLs and therefore direct deliveries to the stores. Constraint (2) imposes a minimum utilization of the vehicles for the consolidation strategy to be eligible for the store. Constraints (3) impose a limit on the maximum number of FTLs that can be sent to a store. Preferred consolidations, i.e. a group of products f can only be consolidated if product p is consolidated, are enforced by constraint (4). The control of backorder situations is guaranteed by constraints (5) and (6). The first defines the days when backorders happen and the latter ensures that the upper limit is not exceeded. Constraint (7) secures that the consolidation strategy is implemented according to the level of consistency imposed. Finally, constraint (8) defines the variables domains.

4.2 Distribution Centers Capacity

After the first step, a group of stores with promising consolidations for direct deliveries are identified. However, the DC where the consolidations are performed might not have enough capacity to perform them all, as it still needs to be operational for the preparations of the remaining stores that will be supplied in a route. In order to analyze the DC capacity and select the stores for which to perform the consolidations, a heuristic procedure is proposed. The heuristic starts by defining a ranking of the stores, which is afterwards used to prioritize the allocation of stores to the DC capacity, without exceeding it, and therefore to select the stores to be supplied with a product consolidation. The overall scheme of the proposed heuristic is presented in Algorithm 1, followed by a detail description.

Algorithm 1 DC capacity check

C_{stores} : set of stores with promising consolidations (given by step 1)
$Selected_{stores} = \emptyset$
for each store s in C_{stores} **do**
    calculate priority rank R_s ▷ step 2.1
    calculate earliest (E_s) and latest (L_s) time the store can be loaded ▷ step 2.2
end for
Sort C_{stores} by priority rank
repeat ▷ step 2.3
    select store s with the highest rank in C_{stores}
    try to allocate to DC the loading of store s to the earliest time between $[E_s, L_s]$
    if allocation is successful **then**
        insert store s into $Selected_{stores}$
    end if
    remove store s from C_{stores}
until $C_{stores} = \emptyset$
return $Selected_{stores}$

Step 2.1 The stores are ranked in order to define a priority allocation to the DC. The ranking is defined based on five dimensions: distance between the store and the DC; dispersion of other stores around the store in analysis (this measure is calculated based on the number of stores in a 50 km radio); number of FTLs to be sent to the store; average vehicle utilization; number of back-order days. This ranking aims to give consolidation priority to stores that are far from the DC, in isolated areas, with a high number of FTLs achieved with high utilization and small number of back-order situations.

Step 2.2 For each store it is calculated the interval of time when the vehicles should be loaded considering the operational restrictions of each of the products to be consolidated. Distinct products have different processing times, specially if they are prepared in a production center or delivered from suppliers through cross-docking. The stores have delivery time-windows that need to be respected. Additionally, in a pre-processing phase it needs to be confirmed with the stores if they have capacity to receive all products consolidated at once.

Step 2.3 The stores are allocated to the DC according to their ranking priority while satisfying the interval of time they should be loaded. The allocations are only successful if the DC capacity holds. At the end, the $Selected_{stores}$ will be supplied with a product consolidation direct delivery.

4.3 Solution Evaluation

Note that at the third step, it is already defined a profile for each store as presented in Fig. 1. Therefore, there is a group of stores that were selected to be directly supplied using a given consistent consolidation plan, being the remaining stores supplied in routes separated by product. Overall, the solution cost has three components: (i) the transportation cost of supplying directly the stores with a consolidated load; (ii) the transportation and warehouse cost of moving the products to the consolidation DC; and (iii) the transportation cost of the stores to be supplied in product separate routes.

The transportation cost of supplying directly the stores with a consolidated load is easily calculated since the consolidation model provided the number of FTLs to be sent to each of the selected stores through the planning period. The number of FTLs just needs to be multiplied by the distance between the store and DC and summed up. The transportation and warehouse cost of moving the products to the consolidation DC is defined by calculating the number of vehicles that need to be sent for DC with products for consolidation (and the respective travel cost), and the number of additional pallets that were processed in the DC due to this operation.

The transportation cost of the stores to be supplied in product separate routes is calculated recurring to an Adaptive Large Neighborhood Search (ALNS) algorithm proposed by [1] for rich vehicle routing problems. This approach was used as it considers a heterogeneous fleet site dependent vehicle routing problem with multiple time windows, similar to our problem setting for separate products. The main idea of the ALNS algorithm is to destroy and repair parts of the solution, using different operators in an adaptive manner, in order to find better ones. Overall, the solution approach solves the routing problem for each product flow considering the stores not selected for consolidation. Note that, if the stores selected do not consolidate the full set of products, they will also be considered for routing regarding the products not consolidated.

5 Computational Experiments

The methodology proposed in this work was applied to a case-study from an European grocery retailer, which is briefly described below. Section 5.2 describes preliminary tests performed to analyze the influence of different consolidation strategies and Sect. 5.3 present the final results of the consolidation strategy defined.

5.1 *Case-Study*

The retailer that motivated this work manages a set of DCs (operating as storage and cross-docking) and production centers located in Europe. Both cross-docking facilities and production centers have processing times that need to be taken into consideration when planning the distribution. The company manages seven distinct product flows, each with a distinct origin (within or not the same location), which can be transported with three different temperature ranges. Preferred consolidation of products are defined due to temperature ranges and operational reasons. For instance, fresh products are considered the main type of product in the chilled temperature and therefore, other chilled products, such as meat, can only be consolidated with other products if the fresh products are also consolidated. Each product has a predetermined delivery pattern, i.e. the days of replenishment are defined, as well as the delivery time-windows. In case of product consolidation, the product with the earliest delivery time dictates the overall delivery time-window. The distribution is performed seven days a week, using simple MCVs. Since the retailer had two days of the week where the number and quantity of supplies were very different comparing with the remaining days, along the planning period, these days were removed from the consolidation analysis.

5.2 *Consolidation Analysis*

A preliminary test was performed only using the consolidation strategy model (Sect. 4.1) to analyze the evolution of the number of promising stores for distinct problem settings. It was analyzed the distinct consolidation strategies (full consistency, monthly consistency, week day consistency or absent), the vehicle minimum utilization (80, 85 and 90%) and the permission or not of back-orders within limits. Figure 2 presents the percentage of stores with promising consolidations to achieve FTL for the different problem settings.

The results show that the minimum utilization of the vehicles imposed for FTL seem to have an impact on the number of stores for consolidation, reaching deviations of around 10% of stores between a 85 and 90% of minimum utilization. The

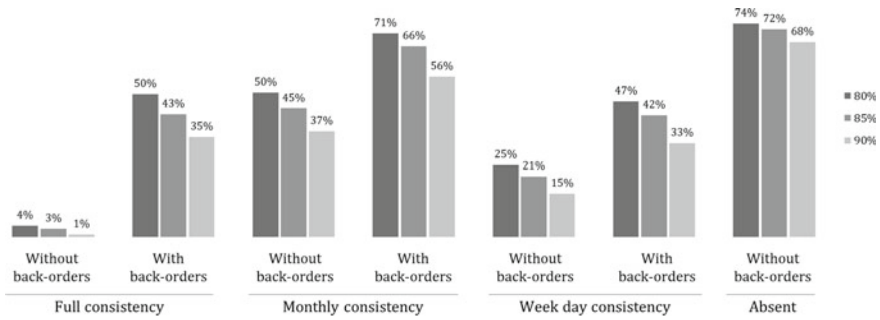


Fig. 2 Percentage of stores with promising consolidations to achieve FTL for different problem settings

permission of back-order situations also has a high influence on the number of stores for consolidation.

The absence of a consistent delivery plan, not allowing back-orders, would have around 68 and 74% and of stores with promising consolidations for FTL. With the permission of back-orders, the monthly consistency achieves a similar percentage of stores. These percentage reduces in the cases for week day consistency and full consistency, which are similar between each other. However, it is worthy of note that the week day consistency allows to increase the number of FTL supplies to the stores by 25%, which could result in an overall reduction of the distribution costs. The higher FTL supplies of the week day consistency (over the monthly consistency) are triggered by additional product consolidations leveraged by the delivery patterns.

5.3 *Final Results*

Since this analysis shows that with a full consistency plan a reasonable number of stores are identified for promising consolidations, and it is clearly the strategy with less disruption to the operation, the complete methodology was applied for this case. The stores with promising consolidations for direct deliveries, defined by the consolidation model were used in the distribution center capacity step. Half of the stores were allocated and selected for consolidation implementation. Note that it was assumed that the full load of products for each FTL is performed at once (conservative scenario), which reduces the DC capacity flexibility. If the distinct products were loaded at different times it could lead to a more flexible management of the DC capacity. At the end, the new distribution planning, with the selected stores supplied directly with a consolidated load, resulted in an increase of the direct supplies costs, but consequently in a reduction of separate product routes. Overall, the final results showed that the retail company could reach annual savings around 4%.

6 Conclusions

This paper analyzes the consolidation strategies a grocery retailer can implement in order to enhance their distribution. The idea of consistent consolidation strategies is proposed in order to smooth retailer operations. Moreover, it focuses on grocery distribution with simple MCVs, which are used in practice but neglected in the literature to the best of our knowledge. A three-step approach is developed to tackle a real case problem, which would be very complex to solve exactly. The problem incorporates the operational complexity of the activity and is divided in three sub-problems: consolidation strategy, DC capacity and routing. The methodology proposed is easy to use for practical problems and therefore to help companies optimizing their delivery mode planning.

Further research can be developed on this topic. The heuristic procedure of step two is very conservative and could be improved, as well as the implementation of a loop with the third step in order to better select the stores for consolidation. Other methodologies could also be developed to tackle the complete problem, such as branch-and-price.

Acknowledgements "Project "TEC4Growth—Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/NORTE-01-0145-FEDER-000020" is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF)."

References

1. Amorim, P., Parragh, S.N., Sperandio, F., Almada-Lobo, B.: A rich vehicle routing problem dealing with perishable food: a case study. Top **22**(2), 489–508 (2014)
2. Derigs, U., Gottlieb, J., Kalkoff, J., Piesche, M., Rothlauf, F., Vogel, U.: Vehicle routing with compartments: applications, modelling and heuristics. OR Spectr. **33**, 885–914 (2011)
3. Gümüş, M., Bookbinder, J.H.: Cross-docking and its implications in location-distribution systems. J. Bus. Logist. **25**(2), 199–228 (2004)
4. Gürbüz, M.Ç., Moinzadeh, K., Zhou, Y.-P.: Coordinated replenishment strategies in inventory/distribution systems. Manag. Sci. **53**(2), 293–307 (2007)
5. Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R., Meybeck, A.: Global Food Losses and Food Waste. Food and Agriculture Organization of the United Nations, Rome (2011)
6. Hübner, A.H., Kuhn, H., Sternbeck, M.G.: Demand and supply chain planning in grocery retail: an operations planning framework. Int. J. Retail. Distrib. Manag. **41**(7), 512–530 (2013)
7. Kovacs, A.A., Golden, B.L., Hartl, R.F., Parragh, S.N.: Vehicle routing problems in which consistency considerations are important: a survey. Networks **64**(3), 192–213 (2014)
8. Kuhn, H., Sternbeck, M.G.: Integrative retail logistics: an exploratory study. Oper. Manag. Res. **6**(1–2), 2–18 (2013)
9. Labuza, T.P., et al.: Shelf-Life Dating of Foods. Food & Nutrition Press, Inc (1982)
10. Martins, S., Amorim, P., Almada-Lobo, B.: Delivery mode planning for distribution to brick-and-mortar retail stores: discussion and literature review. Flex. Serv. Manuf. J. 1–28 (2017)

11. Min, H.: A personal-computer assisted decision support system for private versus common carrier selection. Transp. Res. Part E: Logist. Transp. Rev. **34**(3), 229–241 (1998)
12. Ostermeier, M., Martins, S., Amorim, P., Hübner, A.: Loading constraints for a multi-compartment vehicle routing problem in grocery distribution (working paper) (2016)

Dynamic Approaches to Solve the Smart Waste Collection Routing Problem

Carolina Soares de Morais, Tânia Rodrigues Pereira Ramos and Ana Paula Barbosa-Póvoa

Abstract A Dynamic Inventory Routing Problem model embedded into a rolling horizon solution approach is developed, along this paper, to solve the Smart Waste Collection Routing Problem. This allows the definition of dynamic waste collection routes that explore the use of real-time information on the bins fill-level, over a medium-term horizon. Opposite to a published short-term approach, based on the solution of the Vehicle Routing Problem with Profits that maximize daily profits, the present approach leads to better results translated into higher operational profits. This evidence is shown through the comparison of the solution of both the short-term and the medium-term approaches in a set of small instances where different active rolling horizon intervals are tested. A large instance obtained from a real waste collection system case study is also studied, and the results confirm the conclusions obtained when solving smaller instances.

Keywords Inventory routing problem · Vehicle routing problem with profits · Dynamic routes · Sensors · Waste collection

1 Introduction

The Smart Waste Collection Routing Problem introduced by Ramos et al. [15] considers the use of real-time information on the bins fill-level (transmitted by volumetric sensors located inside the bins) to define smart collection routes that

---

C. S. de Morais (✉) · T. R. P. Ramos · A. P. Barbosa-Póvoa
Centre for Management Studies, Instituto Superior Técnico (CEG-IST), Universidade de Lisboa, Avenida Rovisco Pais, 1049-001 Lisbon, Portugal
e-mail: carolina.morais@tecnico.ulisboa.pt

T. R. P. Ramos
e-mail: tania.p.ramos@tecnico.ulisboa.pt; tania.ramos@iscte-iul.pt

A. P. Barbosa-Póvoa
e-mail: apovoa@tecnico.ulisboa.pt

T. R. P. Ramos
Business Research Unit, ISCTE (BRU-ISCTE), Instituto Universitário de Lisboa, Avenida das Forças Armadas, 1649-026 Lisbon, Portugal

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_12

maximize operations profit (difference between revenues obtained from the recyclable waste collected and the transportation costs of collecting that waste). The problem was modeled as a Vehicle Routing Problem with Profits (VRPP) that selects the waste bins to be visited and the optimal visiting sequence for each day so as to maximize the amount of waste collected while reducing the total travelled distance and satisfying the vehicles capacity and bins capacity. As stated by Archetti et al. [5], the key characteristic of the class of VRPPs is that the set of customers to serve is not given; therefore, two different decisions have to be taken: which customers to serve; and how to sequence the visits in each route; for that, a profit is associated with each customer that makes such a customer more or less attractive.

However, when defining dynamic routes, one important aspect to have in mind is that the access to real-time information may lead to short-term solutions as the data used by the model is constantly updated and a daily period is often assumed. In this context, a longer period approach should be targeted exploring medium-term solutions where events in a close future should also be accounted for (e.g. weekly). For example, if a VRPP model is solved every day, targeting the maximization of a daily profit, the final solution may result in the collection of a certain number of bins that, if collected latter in the week could guarantee higher amount of waste being collected at a minimum cost. To overcome such limitation, Inventory Routing Problem (IRP) models should be explored as these allow profit maximization for a given time horizon. As stated by Coelho et al. [7], the IRP integrates inventory management, vehicle routing, and delivery-scheduling decisions, and, in this context, the supplier has to make three simultaneous decisions: when to serve a given customer; how much to deliver to this customer when it is served; and how to combine customers into vehicle routes.

However, the classical IRP models are not able to deal properly with the available real-time information since a static planning horizon is traditionally considered using the available real-time information at the first period of time and considering estimates for the days ahead. In this way, dynamic approaches coupled with IRP models should be developed to allow a continuous data updating available every day as transmitted by the sensors. Thus, the dynamic approaches should explore the new available information on the bins fill-level, changing, based on this, the planning in order to prevent collection routes from occurring unnecessarily or defining new necessary collection routes to ensure that bins do not overflow.

In this sense, this work proposes a dynamic IRP approach to solve the Smart Waste Collection Routing Problem and compares this new approach to the VRPP model proposed by Ramos et al. [15]. The comparison is done through the solution of instances derived from real data provided by a Portuguese waste collection company. Additionally, a real instance from the same company is solved.

This paper is structured as follows: in the next section, a literature review on dynamic solution approaches for the waste collection is presented; then, the solution methods proposed in this paper are discussed; next, the solution approaches are applied to small instances and, in sequence, to a large instance from a case study; finally, the conclusions about this work are presented.

2 Literature Review

In the existent literature, waste collection problems have been mostly modeled as Vehicle Routing Problems (VRP) that consider a predefined set of bins to be collected (see, for example, the review of Ghiani et al. [9]). A collection situation where the definition of the set of containers to visit is to be optimized has been addressed by Aras et al. [4], Aksen et al. [1] and Anghinolfi et al. [3]. In these works, sites are selected to be visited considering their profit, but no real-time information on the bins' fill-level was used. This is, however, a current challenge as the availability of data is nowadays a reality. In this setting, it becomes important to develop optimization approaches that consider the choice of bins to collect based on real-time information. Few works have, however, looked to this problem. Mes [13], Anagnostopoulos et al. [2] and Gutierrez et al. [10] explored the selection of waste bins to be emptied based only on its weight (and not on the profit that may be attained when collecting them). Furthermore, the selection step was not integrated with the routing step. More recently, Ramos et al. [15] explored such opportunity and integrated the selection of waste bins and the routing steps in a single mathematical model. A VRPP model was developed whose objective is to maximize the daily profit. Bins were selected to be visited every day based not only on its weight but also on its location.

Different from the described short-term approaches (e.g., daily), IRP models deal with a given time horizon considering both routing and inventory issues (Baita et al. [6]). In this sense, these models explore longer-term approaches, maximizing, for example, the profit over a given planning horizon. The work by Aksen et al. [1] explored this problem and sites were selected to be visited considering the maximization of the associated profit for a given time horizon. However, difficulties exist when trying to couple IRP models with real-time information as the dynamism required to deal with every day-updatable information, received by the sensors, is not straightforward when using IRP models (IRP models typically consider only the real-time information on the bins' fill-level for the first day of the planning period; for the days ahead, the model deals with estimates calculated using the expected daily accumulation rate). In this context, dynamic IRP approaches are required, but these are not easily found in the literature, when applied to waste collection. Mes et al. [14] considered a reverse inventory routing problem for collecting waste from underground containers for which the information about current fill-levels is available anytime. The authors consider the waste deposits as stochastic and developed a parametrized heuristic for solving the short-term IRP for waste collection dynamically, taking into account the long-term impact, and using simulation for tuning the heuristic parameters. Most of the Dynamic IRP in the literature deal with an IRP in which demands can vary across the time horizon and planning must be made at the beginning of each of several periods; for each variation, a new IRP is solved. This is the case of Coelho et al. [8] that describe and compare four solution policies for the Dynamic Stochastic IRP (DSIRP), divided into reactive and proactive policies, all implemented in a rolling horizon fashion. The reactive policies are characterized by triggering a replenishment order to bring the inventory position up, whenever the inventory reaches a reorder point; while the proactive policies solve the problem as a deterministic IRP.

Concluding, it is possible to state that there is still space to explore dynamic IRP approaches to solve the Smart Waste Collection Routing Problem, coupling real-time information with the profit maximization for the entire planning horizon, in an updatable approach.

3 Solution Methods to Solve the Smart Waste Collection Routing Problem

Two different dynamic approaches to solve the Smart Waste Collection Routing Problem are studied in this work: (1) a VRPP approach (short-term approach), proposed in the work of Ramos et al. [15]; and (2) a new dynamic IRP approach (medium-term approach).

3.1 Short-Term Approach

In this first approach, dynamic routes are defined through a VRPP model, targeting daily profit maximization. A heuristic method decides the best days in which the collection operation should be performed; for the selected days, the VRPP model is run to define the collection routes for that day. The heuristic rule chooses to collect bins as late as possible, i.e., when at least one bin is about to overflow. The routes to be performed at day t are decided in the morning of day t, after receiving the real-time information on the bins fill-level (see Fig. 1).

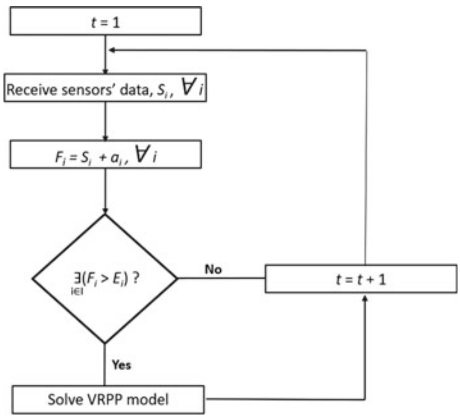


Fig. 1 Short-term approach

An expected daily accumulation rate a_i is added to the amount of waste S_i transmitted by the sensors, generating F_i, a parameter that represents an estimate of the amount of waste in bin i at the beginning of the next day (collection routes are performed at the end of the day). At day t, if there are no bins closely to overflow, the next iteration is set to be carried out on the next day ($t = t + 1$); otherwise, if there is at least one bin closely to overflow, i.e., if there is at least one bin, say i for which F_i is greater than its capacity E_i, the VRPP model is run and routes are defined. For solving this problem, the VRPP model proposed by Ramos et al. [15] is applied considering the desired service level equal to 100% (no bins can overflow).

3.2 Medium-Term Approach

This approach uses an IRP model to maximize profit over the entire planning period (T). The model defines the best days and the best routes to operate, in a given period of time, generating a routing plan that can then be updated with new available real-time information. To allow this update, a rolling horizon solution approach was developed to solve the IRP model in each day within the planning horizon (instead of just solving the IRP model at the first day of the planning horizon, as usual). This allows the use of the data transmitted daily by the sensors. The IRP model is then solved dinamically.

In this approach, the IRP model is solved every day t, in the morning, after receiving the sensors information on the bins' fill-level. When solving the dynamic IRP approach model, the planning period is divided in two intervals: an active interval (AI), that includes the days for which the original binary variables ($x_{ijt}, y_{it} \in \{0, 1\}, \forall i, j \in I, t \in \{1, ..., AI\}, (i \neq j)$) are considered, plus a relaxed interval (RI), which includes the remaining days and where the binary variables are relaxed, assuming positive values ($x_{ijt}, y_{it} \geq 0, \forall i, j \in I, t \in \{AI + 1, ..., T\}, (i \neq j)$). The active interval is then constituted by a current period, which is being solved at the moment, and an auxiliary period, which helps the model to consider future events. As stated by Meira et al. [12], the auxiliary period function is to guide the model, helping to anticipate future problems. The resulting plan for the current period in the first iteration is then fixed for the next iteration, which will be carried out at the next day $t+1$ (see Fig. 2). Successively, for all the next iterations, the resulting plan for the previous period is considered as fixed information.

The length of the active interval (AI) has impact on the Objective Function Value (OFV) and also on the computational time. As stated by Marquant et al. [11], the interval length needs to be carefully chosen in order to reduce the total solving time without affecting the quality of the results. If the interval length is too small, the optimal operating strategy found will be less efficient than the full horizon problem, where all periods are known from the outset. In this sense, by varying AI, different configurations for the application of the rolling horizon approach can be defined (see Fig. 3). In the next section, different active interval lengths will be tested.

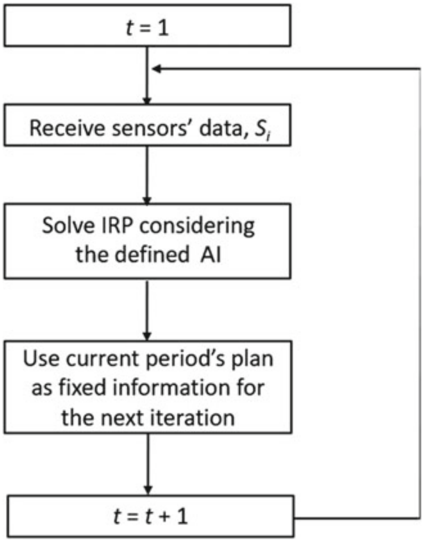


Fig. 2 Medium-term approach

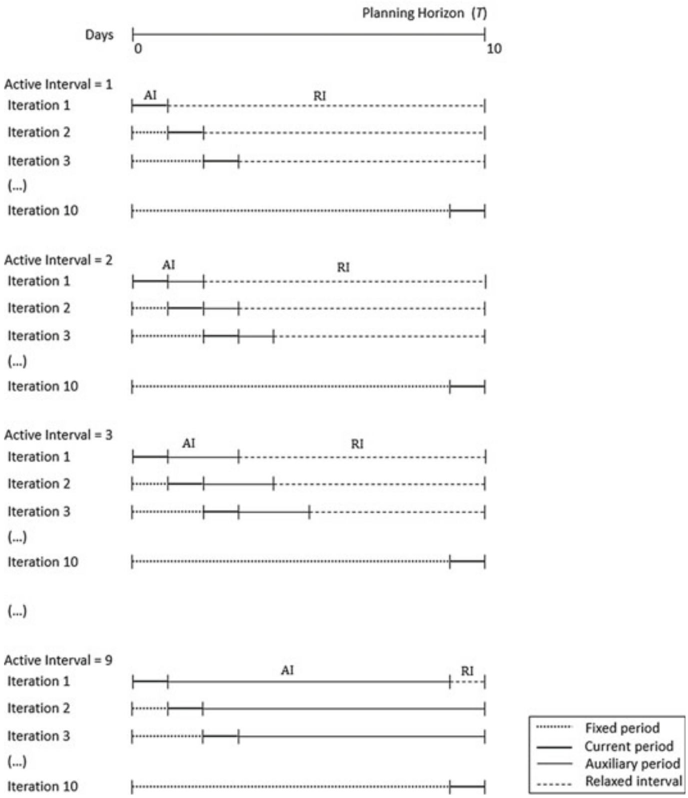


Fig. 3 Different configurations for the rolling horizon approach

The IRP model embedded in the medium-term approach is based on the Selective and Periodic IRP model proposed by Aksen et al. [1] and is as follows.

Sets
$I = \{1, 2, ..., n + 1\}$: set of n waste bins and the depot $n+1$
$T = \{1, 2, ..., t\}$ set of t days in the planning horizon

Parameters
C: travelling cost per distance unit (in €)
R: selling price per kg of a recyclable material (in €)
ω: penalty for the use of the vehicles (in €)
Q: vehicle capacity (in kg)
B: waste density (in kg/m^3)
d_{ij}: distance between node i and node j
S_{i0}: amount of waste in kg at bin i at the first day of the planning horizon (calculated using the information given by the sensor (in m^3) and the material density (in kg/m^3)
a_{it}: expected daily accumulation rate of bin i at day t (in kg)
E_i: capacity of bin i (in kg)

Decision variables
x_{ijt}: binary variable indicating if edge (i, j) is visited at day t, $(i, j \in I)(t \in T)$
y_{it}: binary variable indicating if waste bin i is visited at day t, $(i \in I)(t \in T)$
f_{ijt}: positive variable representing the flow between node i and node j at day t, $(i, j \in I)(t \in T)$
w_{it}: positive variable representing the amount of waste collected at waste bin i at day t, $(i \in I)(t \in T)$
u_{it}: positive variable representing the amount of waste at waste bin i at the end of day t, $(i \in I)(t \in T)$
k: integer variable on the number of vehicles to use $(k \in K)$

Model

$$\max \mathrm{P} = R \sum_{i \in In\{n+1\}} \sum_{t \in T} w_{it} - (c \sum_{i \in In\{n+1\}} \sum_{j \in In\{n+1\}} \sum_{t \in T} x_{ijt} d_{ij} + k\omega) \tag{1}$$

s.t.

$$\sum_{j \in I,(j \neq i)} f_{ijt} - \sum_{j \in I,(j \neq i)} f_{jit} = w_{it}, \forall i \in In\{n+1\}, t \in T \tag{2}$$

$$f_{ijt} \leq (Qk - a_{jt} x_{ijt}, \forall i, j \in I, t \in T, (i \neq j) \tag{3}$$

$$f_{ijt} \leq Qk - w_{jt}, \forall i \in I, j \in In\{n+1\}, t \in T, (i \neq j) \tag{4}$$

$$f_{ijt} \geq w_{it} - BigM(1 - x_{ijt}), \forall i \in In\{n+1\}, j \in I, t \in T, (i \neq j) \tag{5}$$

$$\sum_{j \in I,(j \neq i)} x_{jit} = y_{it}, \forall i \in In\{n+1\}, t \in T \tag{6}$$

$$\sum_{j \in I,(j \neq i)} x_{ijt} = y_{it}, \forall i \in In\{n+1\}, t \in T \tag{7}$$

$$\sum_{i \in In\{n+1\}} x_{in+1t} = \sum_{i \in In\{n+1\}} x_{n+1it} \tag{8}$$

$$w_{it} \leq BigMy_{it}, \forall i \in In\{n+1\}, t \in T \tag{9}$$

$$u_{it} \leq BigM(1 - y_{it}), \forall i \in In\{n+1\}, t \in T \tag{10}$$

$$u_{i0} = S_{i0}, \forall i \in In\{n+1\} \tag{11}$$

$$u_{it} = u_{it-1} + a_{it} - w_{it}, \forall i \in In\{n+1\}, t \in T \tag{12}$$

$$u_{n+1t} = \sum_{i \in In\{n+1\}} w_{it}, \forall t \in T \tag{13}$$

$$u_{it} \leq E_i B - a_{it+1}, \forall i \in In\{n+1\}, t \in T \tag{14}$$

$$x_{ijt}, y_{it} \in \{0, 1\}, \forall i, j \in I, t \in T, (i \neq j) \tag{15}$$

$$f_{ijt}, w_{it}, u_{it} \in \mathbb{R}^+, \forall i, j \in I, t \in T, (i \neq j) \tag{16}$$

$$k \in \mathbb{N} \tag{17}$$

The objective function (1) considers the maximization of profit (P), defined as the difference between the revenues from selling the waste collected and the transportation cost. Constraint (2) represents the flow balance at each waste bin i. Constraints (3) and (4) provide upper bouns on the flow variables f_{ijt} by considering the vehicle capacity and the waste amount collected from node j when a vehicle travels from i to j at day t. Lower bounds on the flow variables are ensured by constraint (5). Variables x_{ijt} and y_{it} are linked in constraints (6) and (7), which ensure that the incoming/outgoing degree of waste bin i must be equal to 1, if waste bin i is visited at day t; and equal to 0, otherwise. The depot's mass balance is provided in constraint (8). Constraint (9) ensures that the collection amount at waste bin i at day t must be zero unless it is visited at that day. Constraint (10) guarantees that the inventory at waste bin i must be zero at the end of day t, if it is visited on that day. Constraint (11) ensures that the initial inventory at waste bin i is equal to the information sent by the sensors on the waste bin' fill-level. Constraint (12) updates the final inventory at waste bin i at day t. Constraints (13) and (14) guarantee that no waste bin i can overflow. The domain of the variables is given by constraint (15), (16) and (17).

4 Application to Small Instances Derived from Real Data

In this section, the short-term and medium-term approaches are tested and compared using small instances derived from real data provided by a Portuguese waste collection company. Moreover, different active interval lengths (AI) for the medium-term approach are explored and compared with the static IRP (IRP without rolling horizon). The instances consider data from 7 up to 13 containers, during a 10-days time horizon (T), and only one vehicle. These small instances were selected so that when applying the static IRP model, the optimality is reached within the established time limit (4600 s).

Both VRPP and IRP mathematical models were implemented using GAMS 24.6.1 and solved with CPLEX Optimizer 12.6.3 on an Intel Xeon CPU X5680 @ 3.33 GHz.

The obtained results are translated into Table 1. These involve the Objective Function Value (OFV), in euros; the computational time (CPU), in seconds; and the time interval between two consecutive routes (N), in days, resulting from the application of the VRPP model, static IRP model (IRP without rolling horizon) and dynamic IRP approach (IRP with rolling horizon). For the rolling horizon approach, the active interval was tested with $AI = \{1, ..., (T - 1)\}$. To illustrate the results, Table 1 shows the results for $AI = 1$ day, $AI = 5$ days and $AI = 9$ days. All instances were solved to optimality. For instance #1, solving the VRPP model for each one of the 10-days results in a routing plan with two routes, one route at day 1 and another route at day 6, hence $N = 5$ days, and the total profit is 11.2€. Solving the same instance with the IRP model, the routing plan also includes two routes, but at different days, day 1 and day 5, hence $N = 4$ days. Moreover, this routing plan leads to a higher total profit of 12.2€.

When solving the dynamic IRP approach, a negative profit is obtained when $AI = 1$ day, a similar profit to the VRPP solution is obtained with $AI = 5$ days, and the optimal profit is reached with $AI = 9$ days.

Figure 4 shows the evolution of the OFV, in euros, when applying the dynamic IRP approach with different active intervals. We also compare the length of the active interval AI with the time interval between consecutive routes (N) of the VRPP solution (short-term approach). It can be seen from this figure that when the length of AI is at least equal to N+2, the optimal solution is always reached. For example, instances #1, #2 and #3 reach the optimal solution when applying the dynamic IRP approach with $AI = 7$, which is equal to N+2 (in those cases, $N = 5$).

Table 2 compares the OFV and the computational time resulting from the application of the VRPP model, the static IRP model, and the dynamic IRP approach. When comparing the static IRP with the VRPP, it can be seen that the static IRP model provides better OFV for all test instances, with an average improvement of 11%, reaching a maximum difference of 34% in instance #6. This is justified by the fact that IRP models consider a profit maximization for the entire horizon instead of a daily profit maximization as the VRPP approach does. For the IRP solutions, a higher number of bins are visited in the last route of the planning horizon, consequently these are found fuller than if collected before.

Table 1 Results for the application of the VRPP model, the static IRP model and the dynamic IRP approach

Instance	Number of bins	VRPP			Static IRP			Dynamic IRP (AI = 1)			Dynamic IRP (AI = 5)			Dynamic IRP (AI = 9)		
		OFV(€)	CPU(s)	N	OFV(€)	CPU(s)	N	OFV(€)	CPU(s)	N	OFV(€)	CPU(s)	N	OFV(€)	CPU(s)	N
#1	7	11.2	0.5	5	12.2	11.5	4	−27.7	3.4	3	11.0	4.5	4	12.2	23.6	4
#2	10	13.7	0.6	5	17.2	38.4	4	−34.0	3.4	3	16.0	9.1	4	17.2	123.1	4
#3	13	33.3	0.6	5	42.6	1,454.9	4	2.7	4.6	3	41.3	26.4	5	42.6	329.2	4
#4	12	60.0	4.3	6	60.1	1,696.3	6	−268.7	6.8	1	13.0	23.0	4	60.1	1,077.5	6
#5	12	50.93	1.5	6	51.4	1,929.2	6	−139.7	7.8	2	51.4	63.2	6	51.4	2,180.2	6
#6	12	30.5	1.3	7	40.7	198.0	9	−225.4	6.6	1	21.1	31.5	6	40.7	30.3	8
#7	12	34.7	3.2	6	39.4	2,568.8	7	−120.1	6.1	1	30.7	1,353.7	5	39.4	1,676.6	7
#8	12	55.8	1.9	9	56.1	1,593.7	9	−260.4	9.2	1	36.6	93.2	6	56.1	392.5	9
#9	12	46.0	1.7	8	46.0	4,503.8	8	−201.2	7.1	1	24.9	337.3	6	46.0	461.4	8
#10	12	34.4	2.2	8	34.7	1,853.8	8	−126.5	12.6	2	0.6	348.1	7	34.7	700.4	8

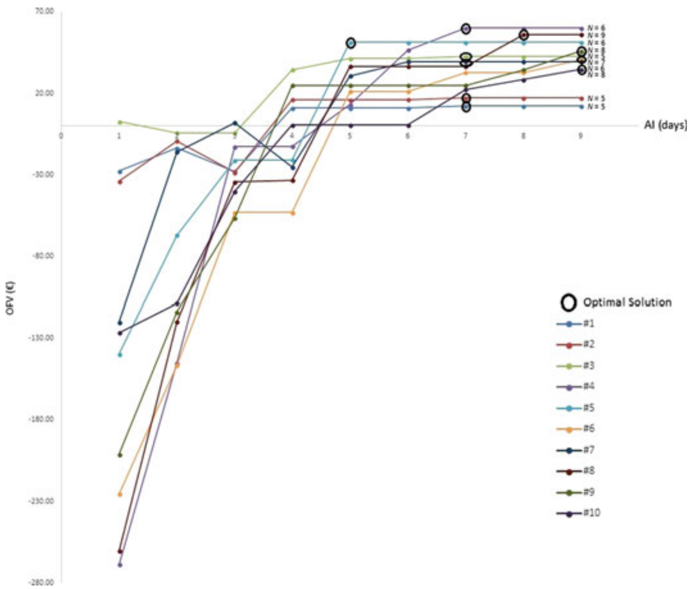


Fig. 4 OFV resulting from the application of the dynamic IRP approach for all instances

Table 2 Comparison between the application of the VRPP model, the static IRP model and the dynamic IRP approach

Instance	Number of bins	Static IRP versus VRPP		Dynamic IRP (AI = 1) versus Static IRP		Dynamic IRP (AI = 9) versus Static IRP	
		$OFV(\%)^{a}$	$CPU(\%)^{a}$	$OFV(\%)^{b}$	$CPU(\%)^{b}$	$OFV(\%)^{b}$	$CPU(\%)^{b}$
#1	7	8.8	2,164.3	−328.4	−70.2	0.0	105.6
#2	10	25.5	6,009.7	−297.7	−88.6	0.0	220.7
#3	13	27.8	242,791.2	−93.6	−99.7	0.0	−77.4
#4	12	0.1	39,165.4	−547.3	−99.6	0.0	−36.5
#5	12	0.8	132,492.0	−372.0	−99.6	0.0	13.0
#6	12	33.7	14,586.1	−653.2	−96.7	0.0	−84.7
#7	12	13.6	79,999.0	−404.8	−99.8	0.0	−34.7
#8	12	0.6	83,079.1	−563.9	−99.4	0.0	−75.4
#9	12	0.1	269,910.7	−537.0	−99.8	0.0	−89.8
#10	12	0.8	84,899.2	−464.9	−99.3	0.0	−62.2
Average	–	11.2	95,509.7	−426.3	−95.3	0.0	−12.1

[a] $OFV\ (\%) = (OFV\ Static\ IRP - OFV\ VRPP)/OFV\ VRPP$
[a] $CPU\ (\%) = (CPU\ Static\ IRP - CPU\ VRPP)/CPU\ VRPP$
[b] $OFV\ (\%) = (OFV\ Dynamic\ IRP - OFV\ Static\ IRP)/OFV\ Static\ IRP$
[b] $CPU\ (\%) = (CPU\ Dynamic\ IRP - CPU\ Static\ IRP)/CPU\ Static\ IRP$

However, the static IRP model takes longer CPU times when compared to the VRPP model (e.g., for instance #3, the static IRP CPU time is more than 2400 times longer than the VRPP).

Additionally, when comparing the application of the static IRP model *versus* the developed dynamic IRP approach, where a rolling horizon solution approach is explored (Static IRP vs. Dynamic IRP, in Table 2), the results show that the obtained OFV for the dynamic IRP is closer to the optimal OFV, given by the static IRP model, for active intervals close to the defined planning horizon. For example, considering $AI = 9$ days in the dynamic IRP approach, all test instances present an OFV equal to the optimal OFV of the static IRP. Although the computational time increases with the length of the active interval, for 7 out of the 10 instances, the total computational time required in the dynamic IRP approach considering an active interval equal to 9 days is lower than the time required by the static IRP. For the other 3 instances (instances #1, #2 and #5), it can be observed in Fig. 4 that the optimal OFV is obtained when solving the dynamic IRP approach with active intervals smaller than 9 days (7 days for instances #1 and #2 and 5 days for instance #5). Consequently, in these cases, the rolling horizon approach requires lower computational times than when solving the static IRP model (11.4s vs. 11.5s for instance #1; 29.4s vs. 38.4s for instance #2; 63.2s vs. 1929.2s for instance #5).

Furthermore, when using longest active interval lengths (AI), longest interval between routes (N), in days, are obtained. This is because the solution approach in these cases accounts for a longer planning horizon being in this way able to choose the best day to perform a route. Accordingly, smaller active interval lengths lead to worse results, resulting in some cases in negative OFV with shorter intervals between routes. This happens because, when the approach cannot "look-ahead" enough, routes are defined to collect few bins resulting in a higher number of routes and consequently more costs.

Concluding, the shorter the active interval is, the blinder is the dynamic IRP approach. Also, for larger size instances, cases where, depending on the computer and on the solver, the static IRP model can be impossible to solve, the dynamic IRP approach is proved to be a solution. Based on this, a question arises: what would be the best active interval (AI) to consider for the rolling horizon applied inside the dynamic IRP approach? From the analysis of Table 1 and Fig. 4 a tentative rule is proposed. This states that the time interval between routes (N) plus two days appears generically as a good option for the definition of AI. This rule was observed for all instances studied, meaning that the optimal solution was obtained at least for an active interval equal to $N+2$ days (e.g. $AI = N+2$). This will be later on used in the real case-study.

5 Application to a Large Instance Derived from a Real-Case Study

Company X is responsible for the recyclable waste collection at 14 municipalities in Portugal. It performs three different types of collection: undifferentiated, selective of recyclable materials (glass, paper/cardboard and plastic/metal), and organic. For the collection of recyclable material paper/cardboard, the company performs 26 different predefined routes periodically. In this work, route number 3 was selected to be analyzed as it is representative of the global operation. For this, a time horizon of 15 days was established ($T = 15$ days), which allows a comparative analysis between the current situation (blind collection) and the installation of sensors to get real-time information on the bins fill-level.

Between 9th January and 23rd January, i.e., within 15 days, all 68 bins in this route were visited two times (two routes, each one performed by one vehicle); a total of 270 km was travelled and a total of more than 4000 kg of waste was collected. As there was no real-time information on the bins fill-levels, almost 78% of the collected bins had fill-levels less or equal to 50% (106 bins out of a total of 136 visited bins), leading to a low average waste collection ratio of 14.8 kg/km. Table 3 presents the Key Performance Indicator (KPI) values for the current situation, for the time horizon under analysis.

To simulate a future scenario, where sensors are installed inside the waste bins and the actual bins fill-level is known by the operations manager in the beginning of each day, we apply the short-term approach (Scenario 1) and the medium-term approach (Scenario 2) to the company's data.

5.1 *Scenario 1: Short-Term Approach*

In this scenario, the results show the definition of 3 routes respectively in days 1, 7 and 13 (see Table 4), meaning that in this solution $N = 6$ days. Despite defining one more route than the current situation, better results are obtained: the amount of

Table 3 Current situation

KPI	Day 1	Day 13	Total
Profit (€)	51.0	60.1	111.1
Weight (kg)	1,966.4	2,062.3	4,028.7
Distance (km)	135.8	135.8	271.6
Attended bins	68	68	136
Ratio (kg/km)	14.5	15.2	14.8
Vehicles usage rate (%)	49.2	51.5	50.3

Table 4 Short-term approach

KPI	Day 1	Day 7	Day 13	Total
Profit (€)	114.9	24.2	23.4	162.5
Weight (kg)	2,113.5	1,221.3	1,203.8	4,538.6
Distance (km)	85.9	91.8	91.0	268.7
Attended bins	55	61	61	177
Ratio (kg/km)	24.6	13.3	13.2	16.9
Vehicles usage rate (%)	52.8	30.5	22.5	35.3
Computational time (s)	1,714.4	4,600.0	4,600.0	10,914.4

collected waste is increased (4,539 vs. 4,029 kg) while travelling smaller distances (268 vs. 271 km). In this case, the model chooses not to visit empty bins, what results in an increase of the ratio kg per km (16.9 vs. 14.8). For each run, the computational time was limited to 4600 s. Total computational time is the sum of all runs.

5.2 Scenario 2: Medium-Term Approach

In this scenario, the proposed dynamic IRP approach is applied considering $T = 15$ days and $AI = 8$ days, which is justified by the fact that the interval between two consecutive routes in the short-term approach is 6 days ($N = 6$ days), and the analysis done in Sect. 4 shown that $N+2$ is a good option for the definition of AI. In this case, 8 days is a large enough interval to consider the occurrence of the next route, and is an operational interval big enough to provide good solutions in reasonable computational times. Results are shown in Table 5.

This second approach provides better results than the first one. In this case, 4 routes are defined (at days 1, 7, 9 and 15). Regarding the total profit, a larger total profit is obtained when comparing with the short-term approach (185 € against

Table 5 Medium-term approach

KPI	Day 1	Day 7	Day 9	Day 15	Total
Profit (€)	45.7	7.2	−31.9	163.8	184.8
Weight (kg)	1,260.0	814.5	224.1	2,745.3	5,043.9
Distance (km)	74.0	70.2	53.2	97.0	294.4
Attended bins	21	15	5	65	106
Ratio (kg/km)	17.0	11.6	4.2	28.3	17.1
Vehicles usage rate (%)	31.5	20.4	5.6	68.6	31.5
Computational time (s)	4,600.0	4,600.0	4,600.0	947.7	14,747.7

163 €, representing an improvement of 12%). Also, the total amount of waste collected is higher than the amount collected in the first approach (5044 vs. 4539 kg), which corresponds to a smaller number of bins collected, 106 against 177. This shows that the bins collected are fuller (e.g. collected latter in the time horizon). A main disadvantage is that the obtained good results are associated with higher computational times.

6 Conclusions

In the present work, a new dynamic IRP solution approach was proposed to solve the Smart Waste Collection Routing Problem. Such approach, which is a combination of an IRP model with a rolling horizon, proved to be an adequate solution method to solve large real problems, which can be not solvable through exact models. Additionally, the dynamic IRP also presents the advantage of turning the IRP model into a dynamic approach allowing a continuous data updating, fundamental characteristic when exploring the continuous availability of real-time information on the bins' fill-level.

The developed approach proved to result in better operational KPIs (profit, waste collected (kg/km)) than a VRPP approach. However, higher computational times were obtained. This limitation should be further explored in the future.

Acknowledgements This work was supported by the financial support of "Fundação para a Ciência e Tecnologia" (FCT - Portugal), through the research project MIT-EXPL/SUS/0132/2017.

References

1. Aksen, D., Kaya, O., Salman, F.S., Akca, Y.: Selective and periodic inventory routing problem for waste vegetable oil collection. Optim. Lett. **6**, 1063–1080 (2012)
2. Anagnostopoulos, D., Kolomvatsos, K., Anagnostopoulos, C., Zaslavsky, A.: Assessing dynamic models for high-priority waste collection in smart cities. J. Syst. Softw. **110**, 178–192 (2015)
3. Anghinolfi, D., Paolucci, M., Robba, M., Taramasso, A.: A dynamic optimization model for solid waste recycling. Waste Manag. **33**, 287–296 (2013)
4. Aras, N., Aksen, D., Tekin, M.T.: Selective multi-depot vehicle routing problem with pricing. Transp. Res. Part C Emerg. Technol. **19**, 866–884 (2011)
5. Archetti, C., Speranza, M.G., Vigo, D.: Vehicle routing problems with profits. In: Toth, P., Vigo, D. (eds.) Vehicle Routing: Problems, Methods, and Application, 2nd edn. SIAM, Philadelphia (2014)
6. Baita, F., Ukovic, W., Pesenti, R., Favaretto, D.: Dynamic routing-and-inventory problems: a review. Transp. Res. Part A Police Pract. **32**, 585–598 (1998)
7. Coelho, L.C., Courdeau, J.F., Laporte, G.: Thirty years of inventory routing. Transp. Sci. **48**(1), 1–19 (2013)
8. Coelho, L.C., Courdeau, J.F., Laporte, G.: Heuristics for dynamic and stochastic inventory-routing. Comput. Oper. Res. **52**, 55–67 (2014)

9. Ghiani, G., Lagana, D., Manni, E., Musmanno, R., Vigo, D.: Operations research in solid waste management: a survey of strategic and tactical issues. Comput. Oper. Res. **44**, 22–32 (2014)
10. Gutierrez, J.M., Jensen, M., Henius, M., Riaz, T.: Smart waste collection system based on location intelligence. Procedia Comput. Sci. **61**, 120–127 (2015)
11. Marquant, J.F., Evins, R., Carmeliet, J.: Reducing computation time with a rolling horizon approach applied to a MILP formulation on multiple urban energy hub system. Procedia Comput. Sci. **51**, 2137–2146 (2015)
12. Meira, W.H.T., Magatao, L., Relvas, S., Barbosa-Povoa, A.P., Neves, F., Arruda, L.V.R.: A matheuristic decomposition approach for the scheduling of a single-source and multiple destinations pipeline system. Eur. J. Oper. Res. (to appear) (2018)
13. Mes, M.: Using simulation to assess the opportunities of dynamic waste collection. In: Use Cases of Discrete Event Simulation, vol. 34, pp. 1564–1576. Springer (2012)
14. Mes, M., Schutten, M., Rivera, A.P.: Inventory routing for dynamic waste collection. Waste Manag. **34**, 1564–1576 (2014)
15. Ramos, T.R.P., Morais, C.S., Barbosa-Povoa, A.P.: The smart waste collection routing problem: alternative operational management approaches. Expert. Syst. Appl. **103**, 146–158 (2018)

Supply Chain Resilience: An Optimisation Model to Identify the Relative Importance of SC Disturbances

João Pires Ribeiro and Ana Barbosa-Póvoa

Abstract Supply Chains (SC) have been facing a vast set of events that can endanger their operations and produce permanent damage, which are unknown until its occurrence. This led to an increased awareness of SC Resilience to deal with such SC uncertainty. This work uses a Mixed Integer Linear Programming model to study the relative importance of the different types of events that threaten SC operations exploring the use of two indicators Expected NPV (ENPV) and Expected Service Level (ECSL). From this work, it can be said that upstream disturbances are of greater importance and because of that should be more efficiently managed, generating value and ultimately a competitive advantage to companies that deploy resilience concerns to their operations.

Keywords Supply chain · Resilience · Optimisation · Disturbances

1 Introduction

Supply Chains (SC) are an essential system of enterprises and have been receiving increased importance with globalisation. However, globalisation has also exposed such systems to high levels of uncertainty and risk. In this setting, SC Resilience (SCR) should be a main SC concern, as this allows such systems to be able to react to unforeseen events [9]. As proposed by Ribeiro and Barbosa-Póvoa [7]

> A resilient supply chain should be able to prepare, respond and recover from disturbances and afterwards maintain a positive steady state operation at an acceptable cost and time.

J. P. Ribeiro (✉) · A. Barbosa-Póvoa
CEG-IST, Instituto Superior Técnico, Universidade de Lisboa,
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
e-mail: pires.ribeiro@tecnico.pt

A. Barbosa-Póvoa
e-mail: apovoa@tecnico.ulisboa.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_13

Such definition appears as a coherent and comprehensive definition of SCR that should be followed when managing SC. The risk of occurring an event causing significant damages to the SC is today a reality, proven by recent events as stated in the SC literature [3, 4, 10].

In this context, the importance of studying SCR is today a requisite. However, SCR is still far from being thoroughly studied, and there is still the need to understand SCR main characteristics comprehensively.

Most published SCR models have been presenting and studying the resilient concerns by modelling the SC response against a set of disruptions. However, little attention has been given to which kind of disruptions companies should devote efforts and so improve its resilience capabilities. Here the importance of the different types of SC disruptions is the motivation for the study on the particular impacts of such events so as to understand how companies should design their SC to better respond to disruptive events, thus improving the operations resilience. For that, a mixed integer linear programming (MILP) model for the design and planning of SC was developed considering uncertainty in demand. A case study of a European closed loop SC is solved to provide useful information regarding the relationship between different SC configurations and different types of disruptions (Production facilities, Supply and Transportation), aspect not comprehensive addressed in the literature, but very important to support the supply chain managers when managing scarce resources in the design of resilient supply chain.

This paper is structured as follows. In Sect. 2 a brief contextualization on SC disturbances is presented. In Sects. 3 and 4 the problem is defined and the case study is presented. Section 5 presents the results of the application of the model on the case study. Finally, Sect. 6 presents the discussion and conclusions.

2 Supply Chain Disturbances

The primary relevance brought by this work is to aid in SC decision making regarding which events should efforts be directed. This is studied by exploring the framework proposed by [7], constituted by four pillars (Performance Level, Speed, Focus Event, Adaptive Framing) and where the Focus Event is the central element, given its importance in SCR. Deciding on what are the type of events that should be primarily considered in SCR is a topic of discussion [3, 5–7].

The categorisation of such events is essential and should embrace the SCM challenges and provide the most useful channels to face such uncertainties. The work presented by [8] proposes five categories of disturbances (Supply, Transportation, Production Facilities, Communications, Human Resources) as the summarisation of all the events that can cause harm to SC and should be dealt as SCR concerns. However, one can consider that the impact of the latter two types of failures ultimately leads to failures in the first three categories and thus we can focus on the first three types of disruptions (Supply, Transportation and Production Facilities).

Based on this, these disturbances are the events to be studied along this paper, making use of the model developed.

3 SC Model Characterisation

The primary goal of this paper is to examine the extent of the impact caused by several types of disturbances in SC operation, while considering different SC design strategies, so as to identify where should a supply chain manager invest primarily the available resources when designing resilient supply chains. For that an optimisation model for the design and planning of SC is applied, for model details see [2]. The model is a multi-product and multi-period Mixed Integer Linear Programming model where uncertainty in demand is considered. This allows the study of SC Resilience under different SC structures: forward, reverse or closed-loop supply chain.

Following the conclusions on the supply chain resilience metrics of Cardoso et al. [2] where flow complexity is the most adequate supply chain resilience metric two objectives functions are studied: the Expected Net Present Value (ENPV), considering the probability of each scenario occurrence (Pb_s) (Expression 1); and the flow complexity (Expression 2). Additionally, and in order to ensure the best financial result while generating comparable results when considering the maximisation of SC flow complexity, a two step approach is followed: First the SC network configuration in retrieved, considering the model with an objective function maximising SC flow complexity, for all time periods. Second, the SC network configuration found previously is imposed to the objective function where ENPV is maximised (Expression 1). This procedure ensures two sets of results. A first one obtained when only financial concerns are considered and a second one, which results from the application of the two step approach, where flow complexity is maximised.

$$Max \sum_{s} pb_s \times NPV_s \tag{1}$$

$$Max \sum_{t} FC_t = \sum_{t} \sum_{a} \sum_{b} (ForwardFlows_{abt} + ReverseFlows_{abt}) \tag{2}$$

The SC flow complexity (FC) was adopted as it is identified in the literature as a natural mean to provide resilience to SC [1, 2, 6], since the increase in forward and reverse flows provide more flexibility and redundancy to the SC, therefore is seen as an appropriate method to test SC resilient behaviour.

The model, apart from the objective functions involves a diverse set of constraints related to: entities capacities; material balances; demand and supply needs. Providing this model, it is possible to assess the impact of disruptions analysing the SC resilience through the obtained expected service level while considering investments in technology and facilities as well as flow and inventory allocation.

4 Case Study

The case study of a European SC is here analysed, based on that presented by [2], with the values associated scaled down due to confidentiality reasons. The original SC is distributed throughout Europe, with one production facility and one warehouse in Germany and the remaining entities (4 raw materials suppliers, 3 possible final products suppliers, considered as outsourcing, and 18 cities acting as markets) broadly distributed in the continent. The company has the possibility of expanding the current SC by installing new plants in Spain and Italy, with each one of the plants with a set of specific suppliers and new warehouses in the UK, France, Italy and Spain. With the introduction of new facilities the company is also considering the expansion or the upgrade of the existent facilities.

Uncertainty in demand is introduced using a scenario tree, based on the assumption that demand is known in the beginning and in the consecutive branches a probability is assumed to each of the three possibilities (Optimistic, Realistic and Pessimistic), Fig. 1.

Within the case study, different SC structures can be designed: Case A—a forward supply chain; Case B—similar to A, but now integrating also reverse flows between consecutive echelons; Case C—similar to B, but plants and markets can directly exchange products, thus bypassing the warehouses; Case D—similar to B, but with the possibility of transhipment at plants, disassembling centres and warehouses; Case E—the most general case, encompassing all the previous ones, forming a closed-loop supply chain where plants send products directly to markets and can also receive directly from markets the end-of-life products. Transhipment is allowed at plants, warehouses and disassembling centres.

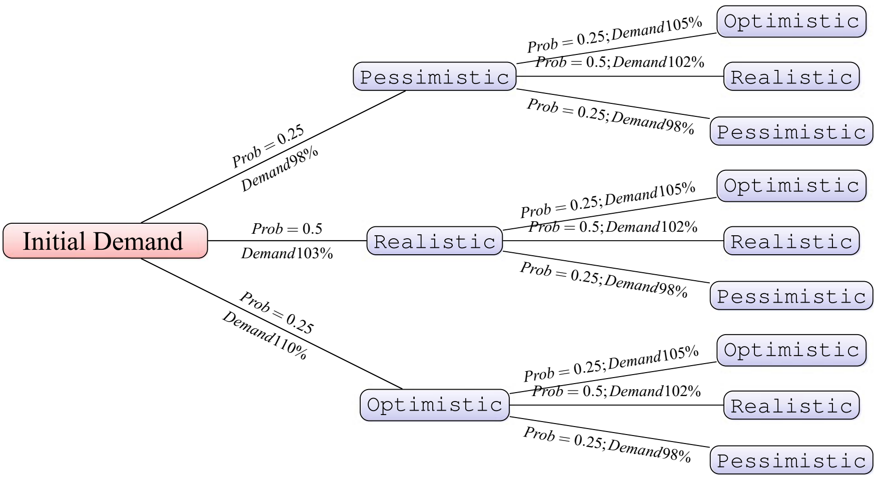


Fig. 1 Demand uncertainty scenario tree

For each SC network configuration case, from Case A to Case E, four operational conditions are studied. One reference case, representing the steady state operation and where the SC network is defined. The defined SC are then faced with three examples of different SC disturbances:

1. **Disturbance 1—Production facilities**: A production disturbance is studied by assuming the occurrence of an event causing all production to stop in the most important plant for one time period.
2. **Disturbance 2—Supply**: A failure in the supply of raw materials is considered. This is implemented by stopping all production of most important suppliers for one time period.
3. **Disturbance 3—Transportation**: A transportation disturbance is introduced by unexpectedly stopping the biggest existent flow between plants and warehouses.

Such disturbances are studied considered the two objective functions, above defined, resulting in four different series of results for each case. The first two series are related with ENPV maximisation, the solid blue line when only the ENPV maximisation is considered, and in the dashed light blue line the ENPV result in the two-step approach considering the flow complexity (FC) maximisation. The remaining two series are the result of the Expected Customer Service Level (ECSL), the percentage of demand met, where the solid red line represents the ECSL result obtained with the ENPV maximising objective function and the dashed line in red represents the result from the two-step approach.

The MILP model is implemented in the GAMS software, and the results are presented in Sect. 5.

5 Results

5.1 Disturbance 1

As mentioned, in Disturbance 1, a production failure is considered. From Fig. 2 it can be seen that when comparing the SC network with increased complexity (Fig. 2, Plots and), for all configurations studied, against a SC designed for maximum economic return (Fig. 2, Plots and), an improvement in performance is observed both on the of economic return as well as customer service.

Cases A and B present similar results for both objective functions. The difference between such cases is represented by the introduction of reverse flows which are of no use to improve ECSL. With the increased complexity, the ENPV of Case B decreases slightly, when compared with Case A, due to the added cost of implementing more complexity without providing additional economic return or increased responsiveness to meet customer demand.

For cases C and E the economic and service level outcome are very much identical since the infrastructures are very similar. The added flexibility in Case E is of great

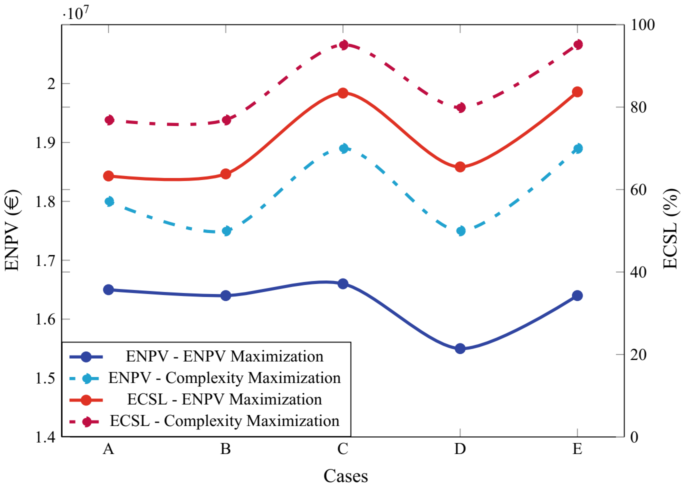


Fig. 2 Results of ENPV and ECSL for Disturbance 1

benefit, since it adds better responsiveness without implying costs if those options are not used by the SC.

Case D presents itself as the worst performing case since its added complexity (allowing transhipment) comes as no use when facing a disruption in the production facility. When maximising flow complexity in the SC, the results for this SC configuration improve. Particularly ENPV results move from being the weaker performer in the simpler strategy to an ENPV in line with Case B when SC flow complexity is maximised. This can be seen as a result of the increased responsiveness of all tiers of the SC that, possibly by increased redundancy better meet customer demands leading to an improvement in the economic return. Is important to mention that this improvement, brought by increased flow complexity, is not a consequence of the different characteristics between Case B and Case D (transhipment). In fact the ECSL is in line with case B allowing to consider the contribution of transhipment to SCR to be small.

5.2 Disturbance 2

Supply-side disturbances are tested in this case, by assuming a failure in suppliers such that production facilities do not receive upstream goods. In this case, Cases C and E (Fig. 3, Plot), that have lower flow complexity, appear with the worst ENPV when suffering the disturbance. However for the same cases when maximising flow complexity there is an increase in economic and service quality performance for all types of SC, Fig. 3, Plots and .

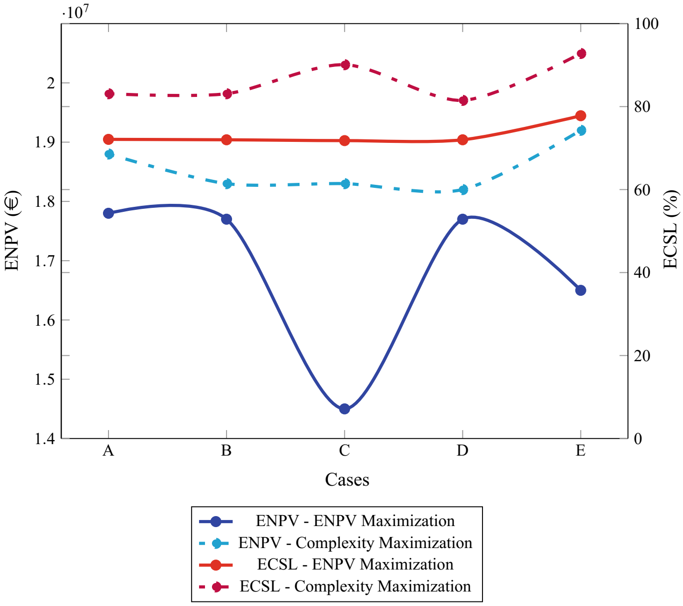


Fig. 3 Results of ENPV and ECSL for Disturbance 2

When designing the SC to maximise profit under steady-state conditions there is the danger of producing severe consequences under a disturbance event, Case C during disturbance 2 is an excellent example of such. There is a noticeable drop in ENPV much because the added complexity (plants and markets being able to communicate directly) does not provide any enhanced solution for the lack of raw materials upstream. With increased redundancy, the SC can increase inventory thus better coping with the changes in operational conditions. Taking advantage of the risk pooling between the different entities, the SC can reduce purchases, developing the responsiveness to meet customer demand while simultaneously improving economic return. On the other hand, Case E provides the SC that can return higher ECSL when facing disturbances. Also, a significant improvement in ENPV between the profit-driven strategy and when designing the SC with increased flow complexity is observed. This added value in ENPV comes from the added flexibility, brought by the increment in possible flows, that combines the possibility to add inventory and to take advantage of the communication between entities allowing for a higher amount of sales, compensating the increased costs of a more complex network.

Cases A, B and D show very similar performance, both in ENPV and ECSL. While the three cases reach the same level of sales, the increased SC responsiveness observed in Cases B and D, with higher costs of investment allows for a decrease in the use of outsourcing, thus improving control, margins and economic return.

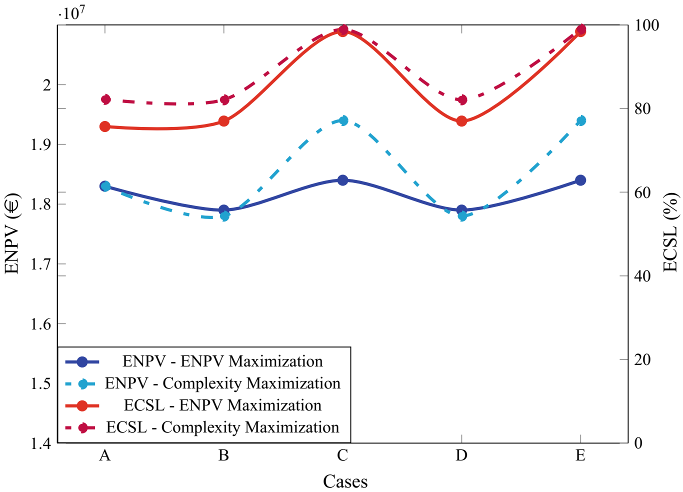


Fig. 4 Results of ENPV and ECSL for Disturbance 3

5.3 Disturbance 3

Transportation failures are a permanent risk for SC managers, delays or even complete failures in the transport of goods can occur. When studying such disturbance and comparing it to the previous events it is observed a weaker impact on SC performance. Higher ENPV and ESCL values close to 100% in both strategies of SC design, Fig. 4.

Cases A, B and D have very similar results between them and with small improvements when a more complex network is deployed, a behaviour caused by the same considerations as in Disturbance 2, Sect. 5.2.

For this type of disturbances, the SC configurations provided by Cases C and D are more resilient. These two cases return a performance very similar between them and can also improve the economic return while increasing flow complexity. The better performance is due to the ability of plants and markets being able to communicate directly, bypassing warehouses. With this relevant characteristic, the operation can avoid the disruptive effects of the transportation failure passing through intermediate entities.

6 Discussion and Conclusions

Accounting for the results previously analysed, in Fig. 5 is represented the mean relative variation between maximising ENPV or ECSL when flow complexity is accounted for (by the two-step approach) and the ENPV without flow complexity consideration (Expression 1), for each type of disturbance considered.

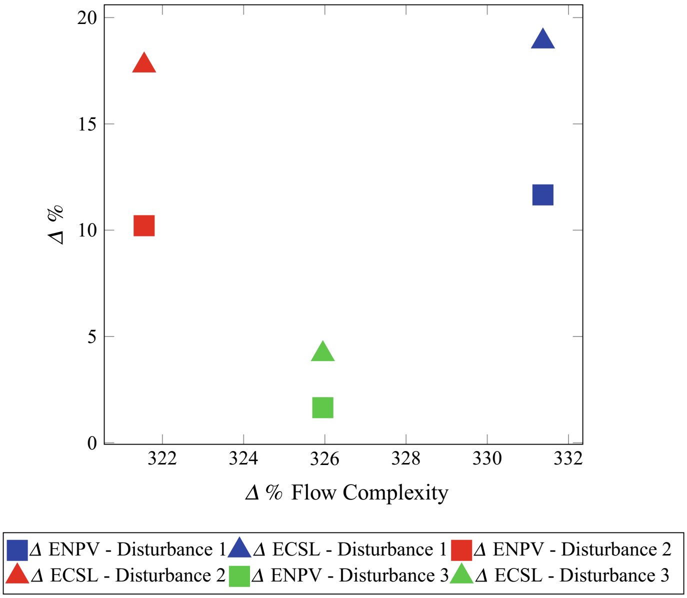


Fig. 5 Delta

When analysing such results it can be observed an improvement in the resilient behaviour brought by the increased SC flow complexity on economic (Squares) and customer service level (Triangles) for the different types of disturbances—a positive variation in %.

In conclusion, it can be said that companies can retrieve beneficial outcomes if considering and investing in a SC with resilient behaviour, by increasing SC flow complexity. The increased observed responsiveness introduces significantly different ways for the SC to face disturbances when compared to the scenario of deciding just based on steady-state conditions.

Additionally, it is seen that the results of ENPV and ECSL are different between the disturbances in study, Fig. 5. For all scenarios, there is a better performance associated with the SC structures that provide a higher SC flow complexity. Also, when disturbances occur, there is a more prominent gain in ECSL than ENPV, since the SC can better respond to the challenges of meeting customer demand.

Transportation failures, represented by Disturbance 3 (Fig. 5 in green), is the scenario where there is a smaller gain from the increased flow complexity, whereas response to Disturbance 1 and 2 (Fig. 5 in blue and red respectively) can be profoundly enhanced if a SC design with resilience concerns is deployed, especially the capability to meet customer demand while improving economic return. This behaviour comes as a result of the impact of the type of disturbance impacting the SC. Production and raw materials supplier are upstream entities that influence all the players downhill since there are fewer ways of compensating such failure in a leaner SC. With an increased amount of possible flows, there is increased flexibility and better responsiveness to changes in steady-state conditions.

From the above analysis, it can be concluded that upstream disturbances are critical and should be the ones where decision-makers should consider investing resources towards SC resilience. Also as managerial insights, companies should invest in contingency plans involving the different players in the SC enhancing communication and collaboration between entities thus advancing through a more resilient SC.

Future work should continue to pursue the challenge of identifying and understanding SCR drivers and the relationship between those. There should be efforts to understand better the disturbances that influence SC performance and can be dealt with more efficiently. The relationship between SC flow complexity and the positive impact on SCR should be the focus of more significant attention.

Acknowledgements The authors acknowledge the project DPI2015-67740-P (MINECO/FEDER).

References

1. Adenso-Diaz, B., Mena, C., García-Carbajal, S., Liechty, M.: The impact of supply network characteristics on reliability. Supply Chain. Manag. Int. J. **17**(3), 263–276 (2012)
2. Cardoso, S.R., Barbosa-Póvoa, A.P., Relvas, S., Novais, A.Q.: Resilience metrics in the assessment of complex supply-chains performance operating under demand uncertainty. Omega **56**, 53–73 (2015)
3. Hohenstein, N.O., Feisel, E., Hartmann, E., Giunipero, L.: Research on the phenomenon of supply chain resilience: a systematic review and paths for further investigation. Int. J. Phys. Distrib. Logist. Manag. **45**(1/2), 90–117 (2015)
4. Kamalahmadi, M., Parast, M.M.: A review of the literature on the principles of enterprise and supply chain resilience: major findings and directions for future research. Int. J. Prod. Econ. **171**, 116–133 (2016)
5. Ponomarov, S.: Antecedents and consequences of supply chain resilience: a dynamic capabilities perspective (2012)
6. Ponomarov, S.Y., Holcomb, M.C.: Understanding the concept of supply chain resilience. Int. J. Logist. Manag. **20**(1), 124–143 (2009)
7. Ribeiro, J.P., Barbosa-Povoa, A.: Supply chain resilience: definitions and quantitative modelling approaches-A literature review. Comput. Ind. Eng. **115**, 109–122 (2018)
8. Rice, J.B., Caniato, F.: Building a secure and resilient supply network. Supply Chain Management Review **7**(5), 22–30 (2003) (Ill)
9. Tang, C.S.: Robust strategies for mitigating supply chain disruptions. Int. J. Logist. Res. Appl. **9**(1), 33–45 (2006)
10. Wang, J., Muddada, R.R., Wang, H., Ding, J., Lin, Y., Liu, C., Zhang, W.: Toward a resilient holistic supply chain network system: concept, review and future direction. IEEE Syst. J. **10**(2), 410–421 (2016)

Supply Chain Purchasing Domain Optimization in a Portuguese Retail Company

Ana Teixeira, Eliana Costa e Silva, Cristina Lopes and João Ferreira Santos

Abstract In this paper we address a case study of the purchasing management section in a Portuguese company in the Retail sector. The purchasing management costs in this company were found to depend largely on the storage mode of the products. Therefore we developed a mathematical model for optimizing the purchasing management costs. In the model we address how to order the products and which storage mode to choose, in order to minimize costs and fulfil the demand. Real instances regarding the monthly demand for one year are tested and the results show that the model can reduce the ratio between operational costs and merchandise costs in almost every instance.

Keywords Supply chain · Retail sector · Mixed integer linear programming · Decision-making

---

A. Teixeira · E. Costa e Silva (✉)
CIICESI, ESTG / P.PORTO – Center for Research and Innovation in Business Sciences and Information Systems, School of Management and Technology / Polytechnic of Porto, Porto, Portugal
e-mail: eos@estg.ipp.pt

A. Teixeira
e-mail: 8130097@estg.ipp.pt

C. Lopes
LEMA,CEOS.PP, ISCAP / P.PORTO – Mathematical Engineering Lab, Centre for Organisational and Social Studies of P.PORTO, Accounting and Business School / Polytechnic of Porto, Porto, Portugal
e-mail: cristinalopes@iscap.ipp.pt

J. F. Santos
Auchan Retail Portugal, Porto, Portugal
e-mail: j-p-santos@iol.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_14

1 Introduction

Globalization and constant changes in the business environment have contributed to the creation of opportunities, enabling organizations to expand their businesses around the world. As a consequence, over the past decade, markets have gone from being independent to integrated in a dynamic environment, requiring companies to develop new and varied products and services, boosting competitiveness [23].

Thus, increasing consumer demand has led to greater uncertainty in demand forecasting and, as a result, efficient and strategically well-managed Supply Chain Management (SCM) has gained increasing importance [20]. In this context, as the SCM issues are developed, the studies refer more frequently to the construction of mathematical models in order to optimize various problems [24] in relation to the various domains of Supply Chain (SC).

The contribution of this paper is a new mathematical model for optimizing one domain of the SC in a Portuguese company in the Retail sector, where the purchasing management costs depend largely on the storage mode of the products. In the model we address how to order the products and which storage mode to choose, in order to minimize costs.

2 Literature Review

2.1 The Evolution of the Concept of Supply Chain Management

In the middle of the XX century, the term of SCM was unknown, having evolved from the concept of logistics and previously of the physical distribution [16]. Since the 90's, the interest for the area began to increase, and these days SCs are considered *"base systems of any organization wishing to operate in a global competitive economy"* [1, p. 15].

Given its importance in the academic and business context, the rapid increase of publications on SCM multiple definitions were proposed and no uniform definition of SCM exist. This led to confusion among researches and professinals [27]. Mentzer et al. [21, p. 18] extensively examined several concepts and defined SCM *"as the systemic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole"*. Similarly, the authors assume SC *"as a set of three or more entities (organizations or individuals) directly involved in the upstream and downstream flows of products, services, finances, and/or information from a source to a customer"* [21, p. 4]. In summary, SCM is seen as a complex network of organizations seeking to ensure the satisfaction of markets [6], with the right products, in the right location, with right quantity and at the right time [2].

According to Seuring [26, p. 1513], *"managing supply chains in a sustainable manner has become an increasing concern for companies of all sizes and across a wide range of industries"*. In fact, the number of publications has increase largely in recent years [28]. Sustainability includes the concept of triple bottom line to SCM, which points out that the activities of companies, besides focusing on improving the economic performance of the company, should also be concerned with social and environmental prevention [5], giving rise to a new concept of Sustainable Supply Chain Management (SSCM).

With regard to the definition of SSCM, several studies extend this theme to other related terms, namely purchase, supply, supply and SC, in that it emphasizes the strategic recognition of these activities in the performance of the company [25, 28]. Likewise, it is important to mention that SC managers can decide on the different sustainability issues as they make decisions about supplier selection, vehicle routing, location decisions, packaging options, which allow, for example, cost savings, better working conditions, shorter delivery times, safer storage and transport, and more [5]. Strategic purchasing and supply are still frequent topics covered in the SSCM literature, but less often addressed. However, in today's hypercompetition era, the purchasing function in many organizations has been seen as a vital strategic component in SCs [12].

2.2 The Purchasing Domain in Supply Chain

In recent years, purchasing has ceased to be considered a purely administrative function to play a very important strategic role in organizations [13]. Some of the reasons for this are the increased levels of outsourcing, the increased use of the internet and information technology, the greater emphasis on SCM, both in academic and business settings, the globalization and the continuous increase of companies' efforts to reduce costs and increase quality [18].

Indeed, planning for purchases is about clarifying issues such as: *In what material and how many units should be purchased?*, *Which suppliers?* and *What kind of cooperation should be privileged?* [9]. Thus, it is important to define, on one hand, outsourcing strategies that stipulate a minimum and maximum number of suppliers by category and quota of supply, and afterwords establish the selection and contracting of suppliers based on issues such as price, trade terms, reliability or lead time, thereby ensuring basic contracts that include quantity discounts and improved delivery efficiency [17].

Therefore, the purpose of the purchasing in SC is more than just the cost of goods purchased. Other factors such as the quality of goods or services should be considered, seeking to meet or exceed the requirements of operations in a timely and economic manner, to negotiate contracts, establish alliances and act as a link between suppliers and various internal departments throughout SC [18].

Due to the fact that companies concentrate on certain competences, they are increasingly dependent on their suppliers. In this context, as purchases is the basis of supplier relationships and seek to establish mutually beneficial relationships with suppliers, the increase of the supplier's response capacity in SC is an additional advantage for the company, since it contributes to effective market competition [4, 13, 18]. For this reason, it can be said that companies with strategic purchases have a more efficient supplier response capacity and, in return, acquire the products more quickly, contributing to the company's advantage and performance [4]. However, although the procurement domain represents entry into SC and is primarily concerned with the provision of resources for the whole SC, it is important to note that its planning must take into account the different domains not only of purchases but also of storage, distribution and sales [10].

Warehouse planning involves making decisions about the size and number of warehouses, the organization of production processes and the capacity of the production system [9]. Warehouse costs must consider transport, fixed, stock and picking costs. In addition, it must be decided whether the products are stored in central or regional warehouses or if the storage is done through (i) direct store delivery; (ii) Cross-Docking (XD), (iii) Picking-By-Line (PBL), and (iv) Picking-By-Store (PBS) [17]. With regard to distribution, the structuring of the distribution system and warehouse locations should be determined as well as the use of different distribution channels and vehicle routes [9], considering the trade-off between infrastructure costs, stock maintenance, transportation and customer service targets [17]. Finally, sales include strategies regarding the type of store and planning of its location, making decisions related to the size of stores, geographic expansion or increase in the density of the SC network [17].

Therefore, given the various decisions to be made for SC planning, operational research, based on quantitative methods, has contributed over the last few years to structuring and modeling problems in the different domains of SC, acting as decision support systems for organizations [1].

2.3 Optimization Models

The deployment of innovative optimization models and new methods has emerged as a support to reach an effective and efficient SC, because in addition to the different decisions to be taken with a view to good planning, the current business world is increasingly uncertain and vulnerable [7]. Thus, although in recent years the number of problem-solving efforts has increased, there are still few studies that apply quantitative models (about 12%), and most are not based on real data [3, 26].

The role of operational research in SC is increasingly linked to applications such as SCM, reverse logistics, location, sustainable manufacturing, among others [8]. In addition, it is present in different industries, especifically in Electric and Electronic industries (19.6%) and Agricultural industry (17.6%); however, Fashion industry, Retail and Grocery (6%) are less studied [24].

With regard to quantitative models, there are a large number of methods that have been applied, with the possibility of combining models, seeking a better resolution of the problem [3, 14, 26]. The adopted methodology must take into account the nature and type of problem. According to Govidan et al. [14], (i) the problems of prices and coordination are usually shaped by approaches to Game Theory; (ii) design and plan problems can use Fuzzy Logic; and (iii) Simulation techniques are present several decision-making processes. Likewise, according to Rajeev et al. [24], the most popular methodologies in SCM include Mixed Integer Linear Programming (MILP), Nonlinear modeling, and Heuristics. With regard to decision-making in the different areas of SC, the review carried out by Mansouri et al. [19], clarifies that Mixed Integer Programming is the most used technique for solving problems between the various interfaces of the SC. However, Govindan et al. [14], indicate that 30.5% of the articles use Linear Programming models and only about 7% deal with Nonlinear Programming.

3 Empirical Study

3.1 The Company and its Challenge

The company under study is recognized worldwide as one of the best food distribution groups, having branded the Portuguese market with some of the most important innovations in the sector. Currently, the group is present in 16 countries and its main vision is to improve the purchasing power and quality of life of the largest number of clients, with responsible, professional and motivated employees. In Portugal, the group has three distinct activities, namely management and construction of shopping centers, banking, and the management of a chain of hypermarkets. The present work focus on the last activity, with emphasis on purchasing.

The purchase manager is responsible for the processes of acquisition and supply of products throughout the SC. His/her mission is to ensure customer satisfaction with products that match their needs, purchased at the best prices, delivered to their respective sales channels and to suppliers offering the best service rates. However, SC has several stakeholders, not only the final consumer, the sales channels, the purchasing direction, the sales direction and the suppliers, but also the transport, supply and logistics platforms. In this context, logistics platforms operate in an outsourcing process and are essential for the entire process. The objective is to supply the sales channels with the lowest operating cost. It incorporates activities such as reception, storage, preparation, dispatch and transfer of information, being present in different parts of the country and grouped according to the typology of the products, that is, frozen, chilled, room temperature, meat products and fresh fish.

Therefore, since purchasing management represents the entrance in the SC and its decisions must be based on the different planning domains, the company proposed a challenge whose main objective is to optimize the decision-making in the purchases for one of its logistics platforms, maintaining an *"end-to-end"* vision of the value chain. For this, different variables present in the complex SC of a hypermarket are considered, namely: how to order the product (pallet or FC[1]); and which storage mode to use (PBL, PBS, XD or in-store delivery). The company is willing to change the way the purchases are made if this reduces the ratio between operational costs and merchandise costs.

3.2 Data Set

For the accomplishment of the empirical study, the company provided a data set, with information on the logistics platform under study, for the period from May 2017 to April 2018, of two suppliers: (i) Supplier I351, that supplies 21 products, with deliveries once a week and (ii) Supplier P940, which provides six products, with deliveries twice a week. For both the rotation objective[2] was of 29 days. These suppliers were selected by the company, since they present different lot dimensions and storage modes, and therefore represent a greater challenge in the decision-making of the purchasing management.

The products can be ordered in FC or in pallet, and the number of units ranges from 6 to 36 in each FC and from 396 to 1980 in each pallet. In addition, when the purchasing manager places the order to the supplier, PBL or XD storage may be used for delivery in FC, and PBS for delivery in pallets. The requests from the platform can only be performed in FC (Fig. 1).

The database provided by the company includes information on the costs incurred in its SC, which are divided in: (i) costs directly related to the price of the merchandise and, (ii) operating costs that guarantee that the product reaches the respective sales channels. In the costs of merchandise, the company included the unit price of the product at the suppliers' location **(Exw)** and the unit value of the pallet **(Pallet)**, which decreases as the number of pallets to be ordered increases. The value of one pallet should also be considered if the order is made in FC. On the other hand, operating costs include rent, picking, transportation and replenishment. The rent of the warehouse **(Rent)** varies according to the size of the lots, i.e., it depends on whether the request to the supplier is made to the pallet or to FC, having a smaller value when ordering the pallet, and is calculated based on the average sale of the item or stock that exists in the warehouse. Picking costs **(Pick)** vary according to the storage mode of the product, i.e., the cost is lower in XD and higher in PBS. Finally, the decision process should also consider the costs related to transportation **(Trans)**

[1]FC = Purchase factor, i.e, number of products per logistic unit.

[2]Estimated time for the sale of the existing stock in the platform based on the average sale of the item.

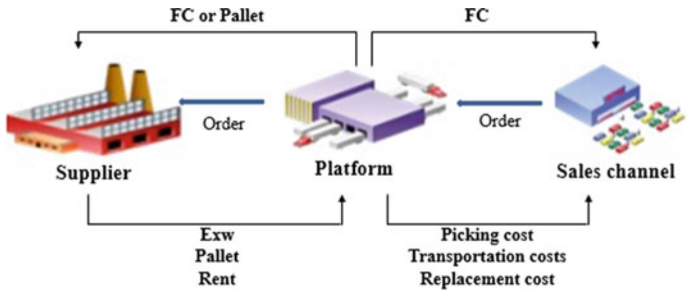


Fig. 1 Cost structure of the company's SC

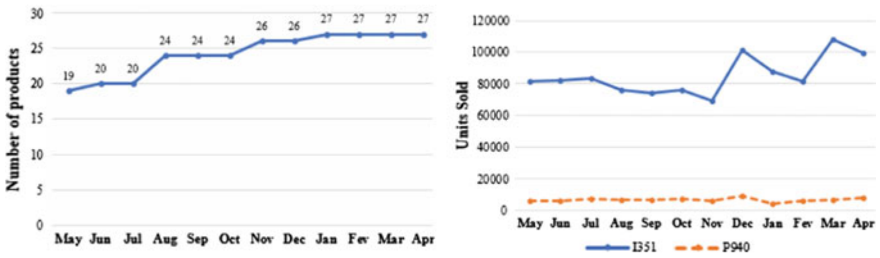


Fig. 2 Quantity of different products (left) and units sold monthly by each supplier (right)

of the merchandise to the respective sales channels and replenishment **(Repl)** of products on the shelves for sale to final consumers. The monthly sales historical information, also provided, will be used as future demand. However the developed model is easily adjusted for different sales forecast, which can be produced with time series techniques.

Figure 2 (left) presents the number of products sold each month. January, February, March and April present a larger variety of products sold, while May, June and July are the months with fewer products are sold. On the other hand, from Fig. 2 (right) we can see that the quantity of products sold by the supplier I351 is much higher than the units sold by the supplier P940, which can be explained by the fact that the first supplies the company in 21 products and the second in only six products. We can also see the number of units from P940 is approximately constant over the months. Supplier I351 shows a greater variation of units sold over the months, e.g., in May, June and July approximately the same quantities are sold and November is the month it sold a smaller quantity of products. December, March and April are the months in which it is sold a greater number of units, higher to 100 000, which can be explained by the festive season of greater consumption that occurs in December and due to the typical promotions of the company in the month of March. The monthly sales suggest that there are months with similar behaviour, while for others this is not true.

In the next section, for testing the quantitative model, we will start by considering instance problems extracted from the data provided by the company. We test the model with the demand for every month, analyse the results of the model concerning three representative months, four randomly chosen products and look at typical solutions per supplier.

4 Model Formulation

For answering the real challenge, the following MILP model has been developed.

Indices and sets are denoted by:

- $i \in Products$—all products;
- $s \in \{XD, PBL, PBS\}$—storage type.

The known parameters are:

- D_i—demand of product i;
- Exw_i —unit cost of product i;
- $Pallet_i^P$—unit cost of pallet for product i;
- $Rent^{s'}$—unit cost of warehouse rent per FC (s'=XD/PBL) or per pallet (s'=PBS);
- $Pick^s$—unit cost of picking for storage mode s;
- $Trans$—unit cost of transportation;
- $Repl$—unit cost of replenishment;
- f_i—number of units per FC;
- p_i—number of units per pallet.
- pfc_i—number of FC per pallet.

Since the company seeks to minimize the SC overall costs, the decision should allow purchasing managers to determine the number of units of each product $i \in Products$ to purchase and their corresponding storage mode $s \in \{XD, PBL, PBS\}$. Therefore the decision variables are:

- $x_i^{s''}$—number of FC of product i with storage $s'' \in \{XD, PBL\}$;
- y_i—number of pallets of product i with storage PBS.

We will need also the following auxiliary variables:

- w_i^k—binary variable that equals 1 if the number of pallets y_i is in the $k-$th price level, for $k \in 1, ..., 6$, and equals 0 otherwise;
- $yAux_i^k$—integer variable that equals y_i if the number of pallets is in the $k-$th price level, for $k \in 1, ..., 6$, and equals 0 otherwise.

The objective function is the minimization of the total costs, which includes the six following components:

$$Cost = \sum_i \left(C_i^{Exw} + C_i^{Pallet} + C_i^{Pick} + C_i^{Trans} + C_i^{Repl} + C_i^{Rent} \right). \tag{1}$$

The first component regards the acquisition costs of the ordered units:

$$C_i^{Exw} = Exw_i\,(f_i \sum_{s''} x_i^{s''} + p_i\, y_i). \tag{2}$$

For "XD" and "PBL", the value of one pallet is a fixed cost that depends only on the product i, while for "PBS", this cost depends on the quantity of pallets to order. The cost of each pallet ($Pallet_i^P$) is a step function defined in six levels according to the number of pallets to purchase:

$$Pallet_i^P = Pallet_i^P(y_i) = \begin{cases} Pallet_i^1, & y_i = 1 \\ Pallet_i^2, & 2 \le y_i \le 3 \\ Pallet_i^3, & 4 \le y_i \le 7 \\ Pallet_i^4, & 8 \le y_i \le 10 \\ Pallet_i^5, & 11 \le y_i \le 15 \\ Pallet_i^6, & 16 \le y_i \end{cases}, \tag{3}$$

where $Pallet_i^1, Pallet_i^2, Pallet_i^3, \ldots, Pallet_i^6$ are constant for each step. Therefore the pallet cost component was modelled considering the above mentioned auxiliary variables w_i^k and $yAux_i^k$ that separate the six levels k of this step function:

$$C_i^{Pallet} = Pallet_i^1\, f_i \sum_{s''} x_i^{s''} + \sum_k Pallet_i^k\, p_i\, yAux_i^k \tag{4}$$

The costs related to picking vary according to the storage mode of the product:

$$C_i^{Pick} = \sum_{s''} Pick^{s''} x_i^{s''} + Pick^{PBS} pfc_i\, y_i, \tag{5}$$

The transportation costs component is given by:

$$C_i^{Trans} = Trans\,(\sum_{s''} x_i^{s''} + pfc_i\, y_i), \tag{6}$$

and the replenishment component in costs is:

$$C_i^{Repl} = Repl(\sum_{s''} x_i^{s''} + pfc_i\, y_i), \tag{7}$$

The rent of the warehouse varies according to the size of the lots, i.e., it depends on whether the request is made in pallets or FC:

$$C_i^{Rent} = Rent^{s''} \sum_{s''} x_i^{s''} + Rent^{PBS}\, y_i. \tag{8}$$

The model will include constrains to ensure customers' demand satisfaction:

$$f_i\,(x_i^{XD} + x_i^{PBL}) + p_i\,y_i \geq D_i \qquad \forall i \tag{9}$$

Also, to separate the storage modes of buying in pallets or in FC, the following constraints guarantee that for XD and PBL the number of units can not exceed the equivalent of one pallet (otherwise we would order in pallets and not in FC):

$$f_i\,x_i^{XD} \leq p_i \qquad \forall i \tag{10}$$

$$f_i\,x_i^{PBL} \leq p_i \qquad \forall i \tag{11}$$

To model the step function of the pallet costs, the following constraints, based on [22], were used.

$$\sum_k w_i^k = 1 \qquad \forall i \tag{12}$$

$$y_i = \sum_k yAux_i^k \qquad \forall i \tag{13}$$

$$0 \leq yAux_i^1 \leq 1w_i^1 \qquad \forall i \tag{14}$$

$$2w_i^2 \leq yAux_i^2 \leq 3w_i^2 \qquad \forall i \tag{15}$$

$$4w_i^3 \leq yAux_i^3 \leq 7w_i^3 \qquad \forall i \tag{16}$$

$$8w_i^4 \leq yAux_i^4 \leq 10w_i^4 \qquad \forall i \tag{17}$$

$$11w_i^5 \leq yAux_i^5 \leq 15w_i^5 \qquad \forall i \tag{18}$$

$$16w_i^6 \leq yAux_i^6 \leq 50w_i^6 \qquad \forall i \tag{19}$$

These constraints separate the levels of the number of pallets and allow to use the different unit costs of the pallets in the objective function.

5 Results

The MILP model presented in the previous section was written using the mathematical programming modelling language AMPL. We used AMPL [11] due to its simplicity and versatility, since it allows interface with several open-source and commercial solvers available in the market. Specifically, the instances for each of the months from May to April were solved using Gurobi [15] solver with AMPL interface.

A total of 12 instances were solved—one per each month. The demand and the unitary costs were provided by the company.[3] For all months, we considered that

[3]Due to confidentiality, all the values presented in this paper were masked.

there were no stocks for any product and therefore the quantity to order is equal to the demand.

The results are presented in Table 1. For all instances, the optimal solution was found. The optimal cost varied from 104, 035.25€ to 165, 409.17€, with a mean value of 126, 980.75€. The problems presented approximately 353 integer and binary decision variables and 323 constraints, mostly inequalities. All problems were easily solved using a minimum of three and a maximum of five simplex iterations.

In Fig. 3 we may observe that the total number of units for all products should be ordered mostly as PBS (in pallets), and less in PBL and XD (in FC). Each ring represents the optimal solution of each month. In these optimal solutions we can see

Table 1 Statistics of the results of the model applied the 12 instances with the monthly demand of 27 products

	Min	Mean	Max	StDev
Objective function (€)	104,035.25	126,980.75	165,409.17	18,451.45
Eliminated constraints	131	136	141	3.07
Eliminated variables	50	52	54	0.85
Number of variables	351	353	355	0.85
Integer variables	163	191.5	235	15.08
Binary variables	107	160.58	190	17.79
Number of constraints	318	323	328	307
Equality constrains	54	54	54	0.00
Inequality constrains	264	269	274	3.07
Simplex iterations	3	3.73	5	0.75

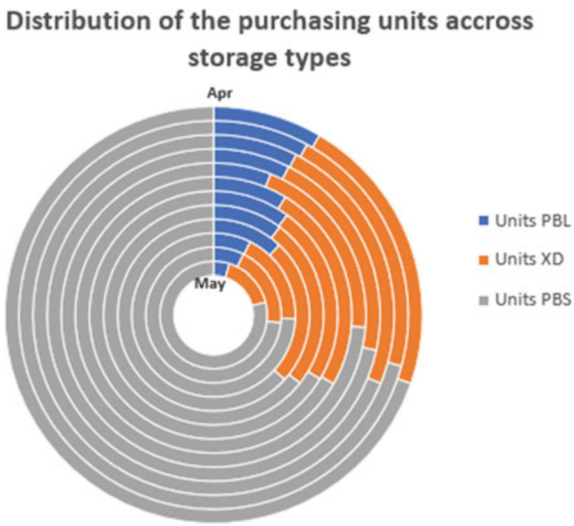


Fig. 3 Distribution of the purchasing orders across the three storage types for the 12 months of the year, in total number of units for all products

a different pattern in each month. In May the larger quantity should be ordered in PBS, as it is the month with overall larger demand. In August, September, October and November, there is a pattern of greater number of units to be ordered in FC with XD and PBL storage modes. The remaining months have a similar pattern of recommended number of units in PBL, XD.

In Fig. 4 the solutions for three different months are depicted. In the pie charts (Fig. 4, top), the total number of units for every product ordered in each month is separated according to the storage type. In the ring charts (Fig. 4, bottom), the same values are also separated by each of the 27 products.

Figure 5 shows the monthly optimal quantity of products and distribution mode for supplier I351 and P940, respectively. For both suppliers, the purchased quantities guarantee that the demand (black line) is fulfilled. The excess units purchased are a small number. We can see that the optimal solution of our model recommends that, for supplier I351, which supplies largest amounts of products, the *PBS* storage mode (purchasing by pallets) is to be used mostly, while *PBL* storage is the less used.

We further analysed for each product the optimal quantities to buy in each storage mode recommended by the proposed model. There were very different solutions regarding different products. Figure 6 shows the distribution of the quantities in the optimal solutions of four different products. For product A, it is always better to buy mostly in pallets (PBS), while for product C it is almost always better to buy in FC, mostly with XD storage. For product B the distribution of the solution varies across

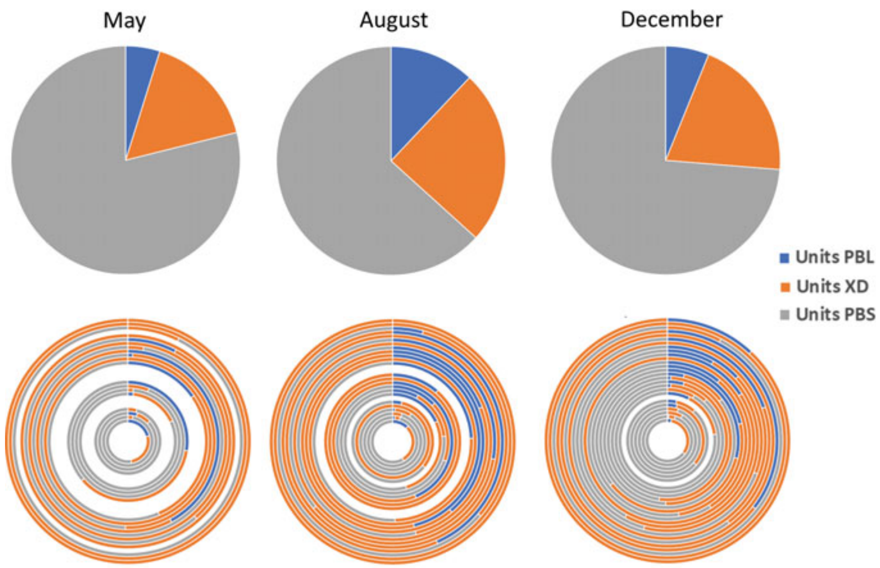


Fig. 4 On top, distribution of purchasing units across storage type during May (left), August (center) and December (right). On the bottom, distribution of purchasing units of each os the 27 products for May (left), August (center) and December (right)

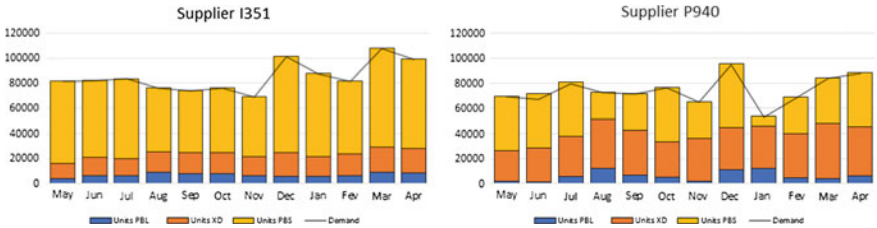


Fig. 5 Monthly evolution of purchasing orders from suppliers I351 and P940

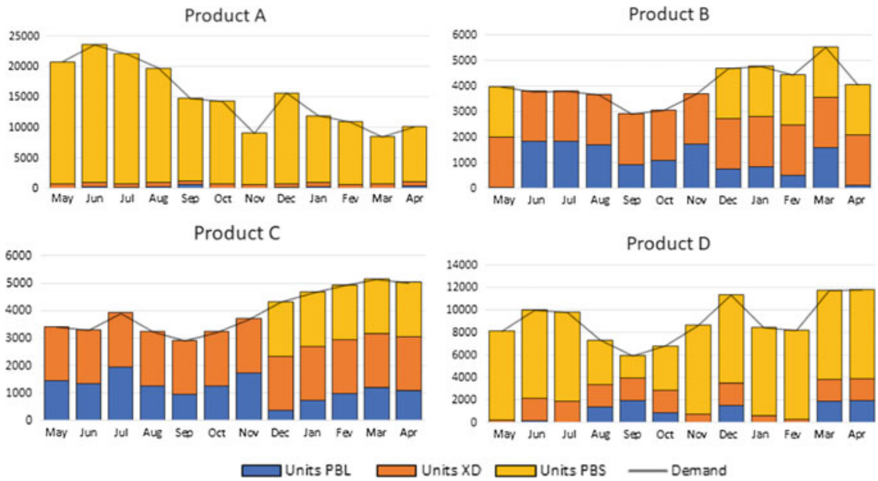


Fig. 6 Monthly evolution of purchasing orders of four different types of products

the months: from June to November the units are divided in PBL and XD, in May and April divided, mainly in XD and PBL, and the remain months are divided in the three systems. Product D also is recommended to be bought divided across the three storage systems, but with a lower number of units to be bought in PBL.

We were interested in understanding how much our model could improve the management costs and how relevant it could be to the company to adopt this solution. We have compared the optimal solutions given by our model with the costs of purchasing every product as the company have done in the past. The comparison

Table 2 Savings in merchandise, operations and total costs, for each of the 12 month instances

Savings (%)	Min (%)	Mean (%)	Max (%)	Std.Dev (%)
Merchandise	5.69	10.00	14.63	2.45
Operations	10.85	16.32	22.09	3.17
Total costs	5.85	10.19	14.85	2.46

is presented in Table 2, with savings being calculated considering as reference the company's current strategy. Looking at the 12 months, we found that our model improves the total costs of the company's SC at least in 5.85% and in maximum of 14.85%. Analysing the costs separately we obtain in the costs of merchandise an average saving of 10% and in operating costs 16.32%.

However, the company is more interested in the ratios between operational costs and merchandise costs per product, than in the overall cost itself. This ratio represents how much is costing the company to move the products along the SC, in terms of a percentage of the merchandise price. Figure 7 shows these ratios computed from our solution, for every product in every month. The ratios are stable across the months, most of them between 1 and 4%, and there is only one product that has a ratio larger than 5% during half of the year.

Therefore, we compared the ratios obtained with the optimal solution given by our model with the costs of purchasing every product as the company have done in the past. The comparison is presented in Table 3. We analysed the products that were purchased over the course of the 12 months, giving a total of 291 ratios and, we found that our model improves the operational costs ratio in almost all situations. In fact,

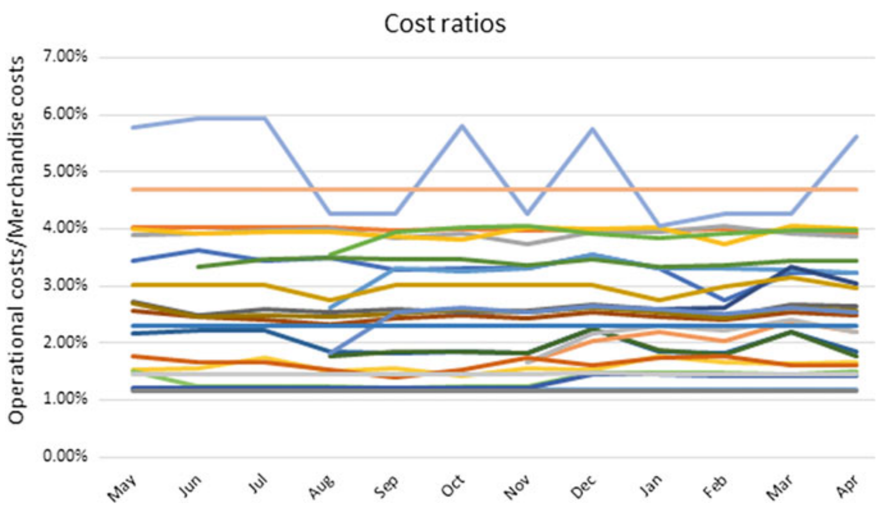


Fig. 7 Ratios between operational costs and merchandise costs in our solution, computed for each product in every month. Each line represents one product

Table 3 Comparison of the ratios between operational costs and merchandise costs for each product in the 12 month instances, if we ordered as the company did, or according to the optimal solution of units with each storage mode given by our model

Our solution improved the ratios?			How much our solution improved			
Total	Yes	No improvement	Min	Mean	Max	Std.Dev
291	289 (99.3%)	2 (0.7%)	0.013%	0.460%	1.756%	0.397%

only in two situations where there was no improvement, i.e., the ratio in our solution is equal to the ratio in the company's current solution. Our solution improved the ratios in at most 1.756%, with a standard deviation of 0.397%.

6 Conclusion

In this work, a case study of the purchasing management section of a company in the Retail sector was studied. A MILP model was developed with the aim of minimizing the costs that influence the decision of purchasing a certain product by pallet or by FC, and what type of storage should be used. The model was tested with real instances from data provided by the company regarding the demand of 27 products from two suppliers, across 12 months. The ratios between operational costs and merchandise costs were computed per product per month. The solutions of the model were compared to the company's current strategy and we found that our model brought improvements in the operational costs ratios for almost every product and month.

References

1. Barbosa-Póvoa, A.: As cadeias de abastecimento e a sustentabilidade. Boletim APDIO **55**, 5–9 (2016)
2. Ballou, R.H.: The evolution and future of logistics and supply chain management. Eur. Bus. Rev. **19**(4), 332–348 (2007)
3. Brandenburg, M., Govindan, K., Sarkis, J., Seuring, S.: Quantitative models for sustainable supply chain management: developments and directions. Eur. J. Oper. Res. **233**(2), 299–312 (2014)
4. Carr, A.S., Smeltzer, L.R.: The relationship of strategic purchasing to supply chain management. Eur. J. Purch. Supply Manag. **5**(1), 43–51 (1999)
5. Carter, C.R., Easton, P.L.: Sustainable supply chain management: evolution and future directions. Int. J. Phys. Distrib. Logist. Manag. **41**(1), 46–62 (2011)
6. Christopher, M.: Logistics & Supply Chain Management. Pearson UK (2016)
7. Choi, T.-M., Govindan, K., Li, X., Li, Y.: Innovative supply chain optimization models with multiple uncertainty factors. Ann. Oper. Res. **257**(1–2), 1–14 (2017)
8. Dias, L.: O papel da investigação operacional. Boletim APDIO **55**, 14–16 (2016)
9. Fleischmann, B., Meyr, H.: Planning hierarchy, modeling and advanced planning systems. Handb. Oper. Res. Manag. Sci. **11**, 455–523 (2003)
10. Fleischmann, B., Meyr, H., and Wagner, M.. Advanced planning. In: Supply Chain Management and Advanced Planning, pp. 81–106. Springer (2008)
11. Fourer, R., Gay, D., Kernighan, B.W.: The AMPL Book. Duxbury Press, Pacific Grove, MA (2002)
12. Giunipero, L.C., Denslow, D., Eltantawy, R.: Purchasing/supply chain management flexibility: moving to an entrepreneurial skill set. Ind. Mark. Manag. **34**(6), 602–613 (2005)
13. González-Benito, J.: Information technology investment and operational performance in purchasing: The mediating role of supply chain management practices and strategic integration of purchasing. Ind. Manag. Data Syst. **107**(2), 201–228 (2007)

14. Govindan, K., Soleimani, H., Kannan, D.: Reverse logistics and closed-loop supply chain: a comprehensive review to explore the future. Eur. J. Oper. Res. **240**(3), 603–626 (2015)
15. Gurobi. Gurobi optimizer 5.0 (2013). http://www.gurobi.com
16. Hou, H., Chaudhry, S., Chen, Y., Hu, M.: Physical distribution, logistics, supply chain management, and the material flow theory: a historical perspective. Inf. Technol. Manag. **18**(2), 107–117 (2017)
17. Hübner, A.H., Kuhn, H., Sternbeck, M.G.: Demand and supply chain planning in grocery retail: an operations planning framework. Int. J. Retail Distrib. Manag. **41**(7), 512–530 (2013)
18. Joyce, W.B.: Accounting, purchasing and supply chain management. Supply Chain Manag. An Int. J. **11**(3), 202–207 (2006)
19. Mansouri, S.A., Gallear, D., Askariazad, M.H.: Decision support for build-to-order supply chain management through multiobjective optimization. Int. J. Prod. Econ. **135**(1), 24–36 (2012)
20. Melo, M.T., Nickel, S., Saldanha-Da-Gama, F.: Facility location and supply chain management-a review. Eur. J. Oper. Res. **196**(2), 401–412 (2009)
21. Mentzer, J.T., DeWitt, W., Keebler, J.S., Min, S., Nix, N.W., Smith, C.D., Zacharia, Z.G.: Defining supply chain management. J. Bus. Logist. **22**(2), 1–25 (2001)
22. MIT OpenCourseWare (2013). http://ocw.mit.edu. *Integer Programming Formulations*. 15.053 Optimization Methods in Management Science
23. Perea, E., Grossmann, I., Ydstie, E., Tahmassebi, T.: Dynamic modeling and classical control theory for supply chain management. Comput. Chem. Eng. **24**(2–7), 1143–1149 (2000)
24. Rajeev, A., Pati, R.K., Padhi, S.S., Govindan, K.: Evolution of sustainability in supply chain management: a literature review. J. Clean. Prod. **162**, 299–314 (2017)
25. Seuring, S., Müller, M.: From a literature review to a conceptual framework for sustainable supply chain management. J. Clean. Prod. **16**(15), 1699–1710 (2008)
26. Seuring, S.: A review of modeling approaches for sustainable supply chain management. Decis. Support Syst. **54**(4), 1513–1520 (2013)
27. Stock, J.R., Boyer, S.L.: Developing a consensus definition of supply chain management: a qualitative study. Int. J. Phys. Distrib. Logist. Manag. **39**(8), 690–711 (2009)
28. Touboulic, A., Walker, H.: Theories in sustainable supply chain management: a structured literature review. Int. J. Phys. Distrib. Logist. Manag. **45**(1/2), 16–42 (2015)

Environmental Performance Assessment of European Countries

Clara B. Vaz and Ângela P. Ferreira

Abstract The European Union (EU) has been promoting an integrated approach to climate protection and energy policy, through a set of key objectives for 2020, 2030 and 2050, linking Europe's green agenda with its need for energy security and competitiveness. This paper aims to evaluate the environmental efficiency of European Countries from 2010 to 2015 towards 2020 targets, through a Data Envelopment Analysis (DEA) model. The DEA model assesses the ability of each country in minimizing current resources while maximizing the gross domestic product (GDP) and minimizing undesirable outputs, such as GhG emissions. The DEA model is based on Directional Distance Function (DDF), imposing weak disposability for the undesirable output (UO). Results obtained show that globally, in the period under analysis, the EU has increased its environmental efficiency which is consistent with the analysis of the indicators of the 2020 climate and energy package.

Keywords Data envelopment analysis · Environmental efficiency · Directional distance function · Europe 20-20-20 targets

1 Introduction

The Europe 2020 strategy, adopted by the European Council in 2007, focuses on smart, sustainable and inclusive growth as a way to overcome the structural weaknesses in Europe's economy. Under the 2020 climate and energy strategies, the EU

C. B. Vaz (✉)
Centre for Management and Industrial Engineering (CEGI/INESC TEC),
Research Centre in Digitalization and Intelligent Robotics (CeDRI),
Instituto Politécnico de Bragança, Campus de Santa Apolónia,
5300-253 Bragança, Portugal
e-mail: clvaz@ipb.pt

Â. P. Ferreira
Research Centre in Digitalization and Intelligent Robotics (CeDRI),
Instituto Politécnico de Bragança, Campus de Santa Apolónia,
5300-253 Bragança, Portugal
e-mail: apf@ipb.pt

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_15

has set three key targets which aim at increasing its energy security, by reducing the dependence on imported energy, creating jobs and help making Europe more competitive. The EU as whole established a (i) 20% cut in greenhouse gas emissions (from 1990 levels), (ii) 20% of energy from renewables and (iii) 20% improvement in energy efficiency, existing an interrelation and mutual support among them.

This study aims at comparing the success of implementing environmental policy-making in Europe concerning those targets. These indicators describe the so-called EU's 20-20-20 climate and energy targets, which are determined for each Member State and yearly reported to European Commission. The European Parliament has continuously supported more ambitious goals. The 2030 climate and energy framework sets at least 40% cuts in GhG emissions (from 1990 levels), 27% share of renewable energy (RE) of EU energy consumption and 27% improvement in energy efficiency. By 2050, the EU should cut GhG emissions to 80% below 1990 levels, as suggested by the low-carbon economy roadmap.

To support the environmental policymaking in Europe, it is necessary to use aggregated indicators of economic and environmental efficiency that take into account simultaneously both energy and non-energy resources in each country. This assures the fair performance comparison among countries, enabling to set goals and to implement effective environmental policies. Following this perspective, this study intends to evaluate the environmental efficiency of European Countries from 2010 to 2015, towards the Europe 2020 targets through a DEA model. The DEA model reflects the ability of each country in minimizing current resources such as capital, labor and energy, maximizing the GDP and minimizing the GhG emissions. The DEA model is based on DDF proposed by [1] and imposing weak disposability (WD) for UOs as proposed by [2, 3] (hereafter CFG model). The contributions of this paper consist of: (i) the analysis of a new dataset, the period 2010–2015 and (ii) the introduction of an additional constraint in CFG model, which can modify the efficiency score of the assessed unit by searching peers whose linear combination reflects more accurately the RE policies observed in the assessed unit.

The paper is organized as follows: after a literature review on environmental and energy efficiency through DEA models in Sect. 2 and Sect. 3 presents the methodology based on DEA technique to assess the environmental efficiency. Section 4 introduces the case study, data description, analysis of the indicators of the 2020 climate and energy package and empirical results derived from the methodology. Finally, the paper's conclusions are provided in Sect. 5.

2 Literature Review

A significant increasing number of articles on DEA applications in energy and environmental efficiency has been developed, particularly after 2000s. Energy efficiency measures the ability to prevent the growth in energy consumption [4]. Since a large portion of GhG emissions comes from energy consumption, the energy efficiency can be considered as one of the most effective ways to reduce the global warming and climate change. The effective use of energy is necessary for both improving corporate competitiveness in business and reducing energy costs for consumers. The benefits from the improved energy efficiency should be incorporated into companies' strategy and energy policy of nations. A review of DEA models related to energy efficiency, published between 2006 and 2015, is presented in [5].

The environmental efficiency is studied in various organizations at the level of companies, industries, regions, and countries to explore how to improve the degree of efficiency on operation and environment [4]. Generally, these studies incorporate inputs, desirable and undesirable outputs in their production factors. Usually, the UOs include man-made emissions of the "kyoto basket" of greenhouse gases, with global warming potential. According to [5] the concept of environmental efficiency indicates how the organizations reduce undesirable production or increase desirable production, particularly against the climate change, rigorous environmental regulation, severe pollution and scarcity of resources. Published papers either focus on the environmental efficiency or eco-efficiency. The eco-efficiency is mostly used in regional and sector level analyses whereas the environmental efficiency is related with country level analysis [5]. This differentiation is supported by the Table 1 that reports several previous studies which assess environmental and energy efficiency at the macroeconomic level through frontier methods.

It is verified that the environmental efficiency is the most explored issue at the country level analysis [6–12, 14, 17, 23–28]. The eco-efficiency is also assessed to country level by [13, 20, 21] as the energy efficiency [15, 16, 18, 19, 22].

Stochastic Frontier Analysis is used by [21] but the DEA is the most frontier method used, mainly due to its non-parametric nature. Several DEA models have been explored for assessment of environmental efficiency such as hyperbolic, radial, slacks-based model, non-radial, range-adjusted model and DDF (c.f. [29]). The main inputs included in these studies are labor, capital and energy while the GDP is always used as output. Almost all studies included UOs which can be CO_2, NOx, SOx, CO or aggregated GhG.

When applying DEA for environmental efficiency assessment, two points are important in selecting models. Firstly, a DEA model needs to include information on all production factors (i.e., inputs, desirable and UOs). Färe et al. [31] were the first to specify an output vector including undesirable and desirable outputs corresponding to weak and strong disposability, respectively. Since then, this concept has dominated most of the environmental production technology [4]. The weak disposability of UOs is used in [6–11, 14]. The strong disposability for the desirable outputs means that lower quantities of outputs can be produced at no cost using the same inputs. The weak disposability for the UOs implies that lower quantities of UOs require either increased quantities of inputs or decreased output production [30]. Secondly, the DEA model should reflect the concerns of the policymakers. Thus, the DDF model is used to assess the environmental performance of EU countries, enabling for increasing outputs while reducing inputs and UOs along a path given by the directional vector. Varying the directional vector, different scenarios might be modelled to represent the concerns of policymakers. For example, [11, 18, 25, 32] define the directional vector equal to the observed variables for each Decision Making Unit (DMU). Other authors [27] used DDF model to decrease only the environmental pressures while maintaining outputs and conventional inputs.

Table 1 Literature review in energy and environmental performance at the country level analysis

Ref[a]	Efficiency	Units	Inputs	Outputs	UO	Model
[6]	Environmental	14 OECD countries, 1990–1995	Capital, labor	Production	CO_2	H, WD-UO
[7]	Environmental	17 OECD countries, 1990	Energy, capital, labor	GDP	CO_2, NOx, SOx	R, CRS; WD-UO
[8]	Environmental	30 OECD countries, 1998–2002	Energy, population	GDP	CO_2	SBM, WD-UO
[9]	Environmental	26 OECD countries, 1995–1997	Labor, energy	GDP	CO_2, NOx, SOx, CO	NR, MI, WD-UO
[10]	Environmental	Eight world regions, 2002	Energy	GDP	CO_2	R, NIRS, VRS, WD-UO
[11]	Environmental	90 countries, 2003–2007	Labor, capital, energy	GDP	CO_2	DDF, WD-UO
[12]	Environmental	70 countries, 1981–2007	Labor, capital, energy	GDP	CO_2	DDF, MI
[13]	Eco-efficiency	22 OECD countries, 1980–2008	GDP	–	CO_2, NOx, SOx	R
[14]	Environmental	63 countries, 1981–2005,	Labor force, capital, energy	GDP	CO_2	DDF, WD-UO
[15]	Energy	EU-15 countries, 1980–2008	RE, Non-RE, Population	GDP	–	R, CRS
[16]	Energy	EU-26 countries, 2009–2013	Capital, labor, energy	GDP	–	R, CRS
[17]	Environmental	95 countries, 1996–2007,	Capital, labor, energy	GDP	CO_2	SBM
[18]	Energy intensity	27 EU members, 2006–2010	Capital, labor, energy	GDP	–	DDF, CRS
[19]	Clean energy use	87 countries, 2004–2010	Capital, energy, labor	GDP, RE	CO_2	SBM, DDF
[20]	Eco-efficiency	OECD countries, 2007	Labor, precipitation, coal	GDP	CO_2	SBM, RAM, UO, ND
[21]	Eco-efficiency	27 EU, 2000–2011	Fossil fuel, RE, capital, labor	GDP	GhG	SF
[22]	Energy	20 OECD countries, 1985–2011	Capital, Labor, RE, N-RE, ND	GDP	CO_2	SBM, UO
[23]	Environmental	31 OECD countries, 2004–2011	Labor, capital, RE	GDP	carbon emission	R, VRS, CRS, MI
[24]	Environmental	26 EU members, 2001–2012,	Labor, capital, N-RE, RE	GDP	GhG	R
[25]	Environmental	28 EU, 1993–2010	–	GDP	CO_2, SO2, NOx	DDF
[26]	Environmental	26 EU, 2001–2012	Labour, capital, energy, RE	GDP	GhG	R, VRS, CRS
[27]	Environmental	28 EU members, 2001–2013	Labour, capital	GDP	NOx, SO2, CO_2	DDF, LI
[28]	Environmental	EU members, 1990–2011	Capital, labor, energy	GDP	CO_2	H

[a]Ref, Reference; H, hyperbolic; R, radial; SBM, slacks-based model; NR, non-radial; RAM, Range-adjusted model; ND, non-discretionary factors; MI, Malmquist index; SF, Stochastic frontier; LI, Luenberger productivity indicator; N-RE, Non-renewable energy

This paper proposes the definition of the environmental production technology by inputs, desirable and UOs, assuming the weak disposability to the last ones, enabling to measure the environmental efficiency in increasing the observed outputs while simultaneously reducing the observed inputs and UOs. This directional vector reflects the environmental policymaking in Europe concerning climate change and energy targets.

3 Methodology

Charnes et al. [33] introduced the DEA as a non-parametric technique to assess the relative efficiency of an homogeneous set of DMUs in producing multiple outputs from multiple inputs. DEA facilitates the construction of a "best practices DMUs" frontier over the existing data. By reference to this frontier, efficiency measures may then be calculated through exploring ratios or distances between observed input and output combinations and frontier input and output combinations.

The notion of a directional technology distance function was introduced in [34, 35], being referred as a shortage function. The DDF for efficiency evaluation is highly popularized by [1], enabling simultaneously to expand outputs and contract inputs according to a directional vector. Consider an environmental production technology T defined by a set of n DMUs j $(j = 1, \ldots, n)$, each consuming m inputs x_{ij} $(x_{1j}, \ldots, x_{mj})$ to produce s desirable outputs y_{rj} $(y_{1j}, \ldots, y_{sj})$ and l UOs b_{kj} $(b_{lj}, \ldots, b_{lj})$. The DDF is a non-oriented measure of efficiency that restricts movements towards the frontier by choosing a priori the directional vector, $g = (g_x, g_y, g_b)$, to be followed. The directional model expands the desirable outputs in the direction g_y and contracts inputs in the direction g_x and UOs in the direction g_b, as denoted by $\overrightarrow{D_o}(x_{ij}, y_{rj}, b_{kj}, g_x, g_y, g_b) = Max\{\beta|(x_{io} - \beta g_x, y_{ro} + \beta g_y, b_{ko} - \beta g_b) \in\ T\ \}$. Hence β is the proportion by which inputs and UOs are contracted and outputs expanded, that can be achieved simultaneously. Note that β is not an efficiency, corresponding to the inefficiency measure. A DMU is efficient when the score of the DDF equals to zero. Scores higher than zero are associated with technical inefficiency, implying that production is not occurring along the frontier. β represents the radial expansion in outputs and radial contraction in inputs and UOs. For example, a value of $\beta = 0.25$ indicates that the DMU could expand desirable outputs by 1/4 and contract inputs and UOs by 1/4.

The directional vector was proposed in [32] as being equal to the observed inputs or outputs, showing the equivalence between the traditional input and output efficiency measures and the DDF model. The application of the DDF for assessing efficiency in the presence of undesirable inputs and/or outputs was introduced in [2, 3]. According to them, for technologies producing both desirable and UOs, the weak disposability should be imposed on the underlying technology, such as those reductions in the UOs (bads) require joint reductions in the desirable outputs (goods). As pointed by Chung et al. [3], 'The fact that goods and bads are jointly produced means that reduction of bads will be "costly": either resources must be diverted to "clean-up" (e.g. scrubbers), production must be cut back, or fines must be paid.'

Thus, imposing weak disposability for the UOs and taking the DDF equal to the observed levels of inputs and outputs [$g = (x_{io}, y_{ro}, b_{ko})$], [3] proposed the CFG model, preserving the linearity of the DEA model, constant returns to scale and the strong disposability of desirable outputs and inputs.

The CFG model is used to assess the environmental efficiency of the countries each consuming inputs to produce desirable outputs and UOs. Strong disposability in outputs implies that desirable outputs can be easily disposed of without cost. Generally, this does not occur with GhG emissions, since the reduction in this UO implies an investment in RE structures and/or energy efficient technologies which is captured by capital in our case study. Then, the weak disposability is imposed on GhG emissions, as referred in Sect. 2.

We propose also that the environmental efficiency of a specific country should be based on countries in which their linear combination in terms of RE share should be at least the share of RE observed in that country. Consider that p_j is the share of RE that is observed in the country j. This methodology requires the inclusion of a new constraint $\sum_{j=1}^{n} \lambda_j\, p_j \geq p_o$, in the CFG model, deriving the model (1). Thus, model (1) enables to assess the environmental efficiency of the countries each consuming inputs to produce desirable outputs and UOs and, in addition, also compares the practices on RE production policy observed in the assessed unit with the RE production policy observed on benchmarks.

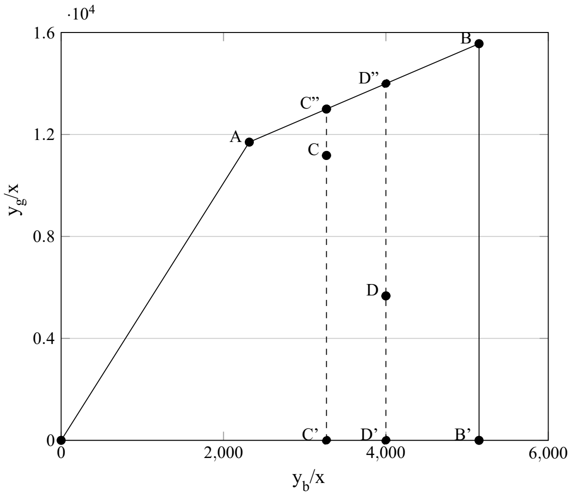
$$
\begin{aligned}
\max \Bigg\{ \beta_o \Big| & \sum_{j=1}^{n} \lambda_j\, x_{ij} \leq x_{io}(1-\beta_o), \quad i = 1, \ldots, m \\
& \sum_{j=1}^{n} \lambda_j\, y_{rj} \geq y_{ro}(1+\beta_o), \quad r = 1, \ldots, s \\
& \sum_{j=1}^{n} \lambda_j\, b_{kj} = b_{ko}(1-\beta_o), \quad k = 1, \ldots, l \\
& \sum_{j=1}^{n} \lambda_j\, p_j \geq p_o, \quad \lambda_j \geq 0, \quad j = 1, \ldots, n \Bigg\}
\end{aligned} \tag{1}
$$

Table 2 Illustrative example and results of models CFG and (1)

DMU	x	y_g	y_b	p_{RE}	Model CFG		Model (1)	
					β_o	Peers	β_o	Peers
A	1	11699.4	2317.4	0.310	0	$\lambda_A = 1$	0	$\lambda_A = 1$
B	1	15560.6	5147.1	0.146	0	$\lambda_B = 1$	0	$\lambda_B = 1$
C	1	11176.5	3267.9	0.092	0.075287	$\lambda_A = 0.61410$, $\lambda_B = 0.31062$	0.075287	$\lambda_A = 0.61410$, $\lambda_B = 0.31062$
D	1	5666.7	4000.0	0.180	0.423592	$\lambda_A = 0.23367$, $\lambda_B = 0.34274$	0.152874	$\lambda_A = 0.34341$, $\lambda_B = 0.50371$

To illustrate the results of environmental efficiency assessed by models CFG and (1), we consider an example with four countries using one input (x : energy) to produce one desirable output (y_g : GDP) and one UO (y_b : GhG emissions), producing a given share of RE (p_{RE}) (see Table 2). Firstly, we assess the efficiency of countries using an input x to produce a desirable output y_g and a UO y_b through the CFG model (see Table 2), without considering p_{RE}. The frontier of the production technology is represented in Fig. 1, including countries A and B which are efficient ($\beta = 0$). C is inefficient, with $\beta = 0.075287$ indicates that it could expand desirable output by 0.075287 and contract input and UO by the same score simultaneously, achieving the targets given by C'' where the input level is (1-0.075287)×1, desirable output level is (1+0.075287)×11176.47 and UO level is (1-0.075287)×3267.94. D is also inefficient and should expand y_g and contract y_b and x by $\beta = 0.423592$, simultaneously. No slacks are observed. Secondly, we assess the environmental efficiency of countries using an input x to produce the UO y_b and a desirable output y_g, given a share of RE (p_{RE}) produced in each country, through the model (1), defining a new production technology (see results in Table 2). A and B keep in the frontier of the production technology while C keeps its inefficiency ($\beta = 0.075287$) and D reduces its inefficiency. C keeps the inefficiency score, as the constraint $\sum_{j=1}^{n} \lambda_j \; p_j \geq p_o$ is verified. That is, $\sum_{j=1}^{n} \lambda_j \; p_j = 0.236$ is higher than its share of RE ($p_{RE} = 0.092$). If C produces a higher share of RE, its inefficiency will reduce until the $\sum_{j=1}^{n} \lambda_j \; p_j \geq p_o$ is satisfied. This happens with D, as its inefficiency is reduced to $\beta = 0.152874$ to become the linear combination of the share of the RE observed on benchmarks ≥ 0.18. In the limit, if the constraint $\sum_{j=1}^{n} \lambda_j \; p_j \geq p_o$ is not satisfied, the DMU becomes efficient. In this work, the environmental efficiency of the European countries is assessed through the models CFG and (1). The case study and the dataset are presented in the next section.

Fig. 1 Illustrative example



4 Case Study

4.1 Data and Variables

The environmental performance assessment is performed for 28 EU Member States, during the 2010–2015 time-frame. Data has been collected from Eurostat regarding Belgium, Bulgaria, Czech Republic (CR), Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Croatia, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Malta, Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Sweden and United Kingdom (UK).

This study uses input and output variables for the DEA analysis based on the rationale of the application, and supported by the information in Table 1. Thus, the environmental production technology is defined by the inputs labor, capital and final energy consumption, considering as desirable output the GDP and as UO the GhG emissions. The GDP by country was collected at market constant prices, in millions of euro (10^6 €). The GDP is measured in chain-linked volumes, reference year 2010. Labor is measured by thousand hours (10^3 h) worked regarding residents and non-residents who work for national producer units. The capital corresponds to the gross fixed capital formation at market prices, in millions of euro (10^6 €), measured in chain-linked volumes, reference year 2010. Final energy consumption by country was collected in 10^6 tonnes of oil equivalent (10^6 TOE). It translates all energy supplied to industry, transport, households, services and agriculture, excluding deliveries to the energy transformation sector and the energy industries themselves. This quantity is relevant for measuring the energy consumption at final place of energy use and for comparing it to the Europe 2020 target. Share of RE in gross final energy consumption (in %) measures how extensive is the use of RE and, by implication, the degree to which renewable sources have substituted fossil and/or nuclear fuels and therefore contributed to the decarbonisation of the EU economy. Aggregated GhG emissions was collected in thousands of tonnes (10^3 tonnes) from the European Environment Agency (EEA). This indicator is also available in terms of total emissions in relation to 1990 emissions (base year 1990). Table 3 summaries the mean and standard deviation (SD) of the above mentioned variables. From the levels of 2010, the mean of capital, labor and GDP increased by 4.6%, 0.2% and 5.7%, respectively, while the mean of final energy consumed and GhG emissions decreased by 6.6% and 9.3%, respectively. The share of RE in gross final energy consumption increased by 25%, on average. Globally, the European countries follow a sustainable trend in terms of climate change and energy policy. Next sections detail the behavior of European countries since 2010, towards Europe 2020 targets.

Table 3 Descriptive statistics

	Capital (10^6 €)		Labor (10^3 h)		GDP (10^6 €)		Energy (10^6 TOE)		Share of RE (%)		GhG (10^3 tonnes)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2010	91 724	136 605	13 329 505	16 455 788	458 141	695 586	41.53	55.47	0.16	0.11	175 340	235 358
2011	93 456	140 861	13 362 316	16 604 252	465 958	711 621	39.62	52.44	0.16	0.11	169 952	227 643
2012	91 175	138 679	13 183 880	16 536 731	463 978	712 101	39.59	53.25	0.18	0.11	167 616	228 598
2013	89 795	136 983	13 090 872	16 464 619	465 174	715 305	39.59	53.98	0.19	0.12	164 244	227 799
2014	92 520	140 474	13 235 801	16 675 349	473 289	727 241	37.96	51.33	0.19	0.12	157 991	216 987
2015	95 907	142 874	13 360 797	16 835 114	484 248	738 769	38.80	52.30	0.20	0.12	158 993	216 622

4.2 Analysis of EU Countries Towards Europe 2020 Targets

As mentioned before, EU sets three targets on the 2020 climate and energy package: reducing its GhG emissions by at least 20%, increasing the share of RE to at least 20% of consumption, and achieving energy savings of 20% or more. At national level, these targets differ according to their financial and economic status, and environmental circumstances affecting the ability to achieve the desired outputs. For instance, on the GhG emissions, there are two policy instruments: the "EU Emissions Trading System" [36], which sets a cap for power stations, industrial plants and also within the aviation sector and the "Effort Sharing Decision", which sets annual GhG emissions for member states in sectors not included in the Emissions Trading System. With regard to the latest, countries' targets vary from 20% cut for the richest countries to a maximum increase of 20% for the poorest countries, though they were still projected to limit emissions. By this way, weaker economies are allowed to increase their emissions to accommodate an economic growth. Regarding the share of renewables in the energy consumption, national targets vary from 10% in Malta to 40% in Sweden, reflecting available resources, different starting points, and ability to further increase it. Member States had also the opportunity to set the energy efficiency targets in terms of either primary or final energy consumption, taking into account factors such as GDP evolution and forecast, development of all sources of renewable and non-renewable energies, and also remaining cost-effective energy-saving potential [37].

Indicators of the 20-20-20 targets for the EU are depicted in Fig. 2, for the time span in analysis (2010–2015), becoming clear that those targets are within reach with a continuous effort in the climate and energy policies. In 2015, GhG emissions had already reached 22.1% lower than 1990 levels and RE provided 16.7% of gross final energy consumption, mainly due to the grid-parity of wind and solar energy. The 2020 final energy consumption target had already been achieved, but it should be noted that some of the reduction may be attributed to the recent economic crisis and years with heating and cooling energy requirements lower than average, which had decreased the consumption.

Figures 3, 4 and 5 detail the indicators for the EU-28 countries, in the time frame under analysis, and set a comparison with their individual targets. From the aggregated GhG emissions (total man-made emissions of the "Kyoto basket") in percentage of the base year 1990 (Fig. 3), it can be seen a generalized trend to decrease the emissions over the years, however in 2015 some countries increased the GhG emissions due to a slight economic recovery. Seven member countries remain above their national reduction target, although all of them show a downward trend, except Ireland. Remaining countries have reached their national targets, but twelve of them, mostly Eastern countries, were allowed to increase their annual emissions in relation to 1990. Belgium, Denmark, Germany, Greece, France, Italy, Finland, Sweden and United Kingdom were able to reduce their emissions below the base value of 1990.

Regarding the share of RE in gross final energy consumption (Fig. 4), all EU countries increased their RE share in the years under analysis and eleven of them

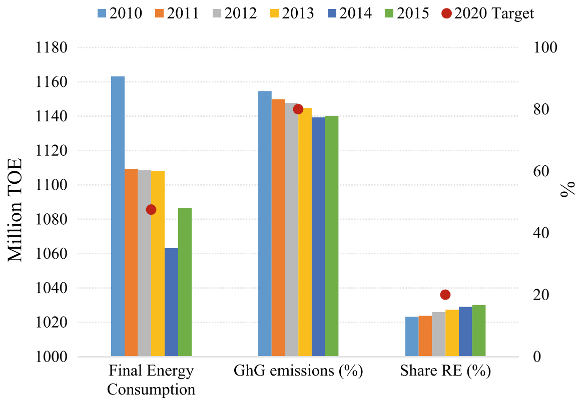


Fig. 2 Indicators of the EU 2020 climate and energy policy

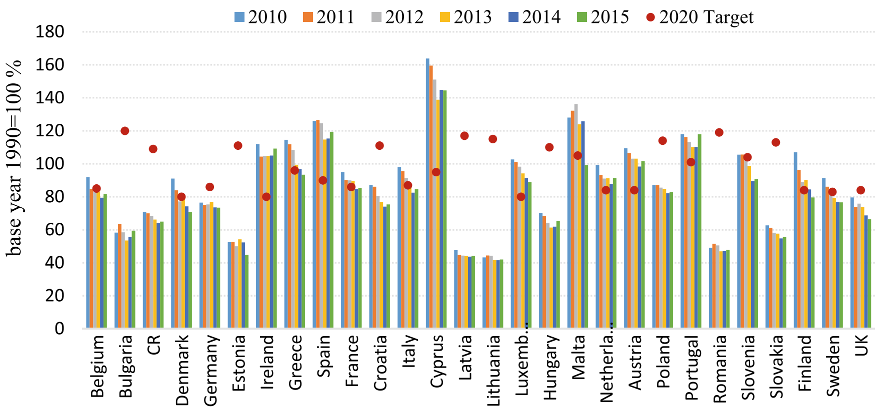


Fig. 3 Aggregated GhG emissions by country and 2020 target (2010–2015)

have already met their 2020 targets, of which Denmark, Estonia, Croatia, Lithuania, Romania, Finland and Sweden have a target higher than 20%. In 2015, the share of RE ranged from 53.9% in Sweden to 5.0% in Luxembourg and Malta. These differences among Member States are mainly due to variations in natural resources, such as the hydropower potential and availability of biomass. From the range of sources, bioenergy is the EU's most important RE source because it contributes to all energy sectors (electricity generation, transport and heating and cooling) [38].

Energy efficiency is assessed in this work through final energy consumption, according to Fig. 5. Most of the countries display an intermittent but overall downward trend. However, much of the decrease may be attributed to financial and economic crisis, rather than to a structural shift in energy savings behaviour. One factor that impacts considerably on the final energy consumption is the heating demand, therefore warmer years, such as 2014, have a significant reduced energy consumption, when compared to the other ones. In 2015, fifteen member states have reached

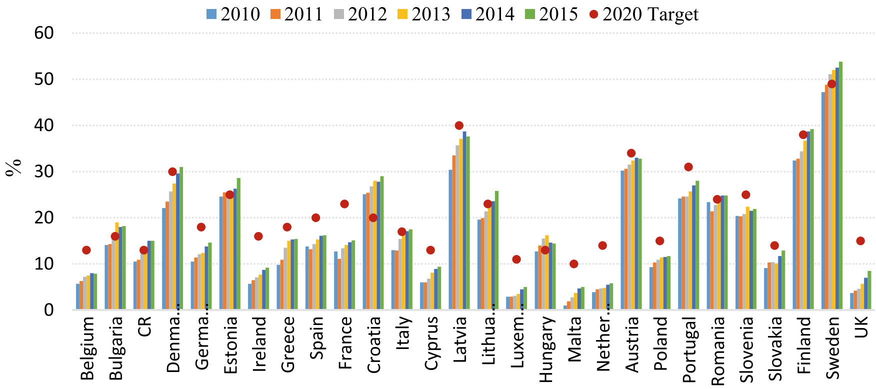


Fig. 4 Share of RE in gross final energy consumption and 2020 target (2010–2015)

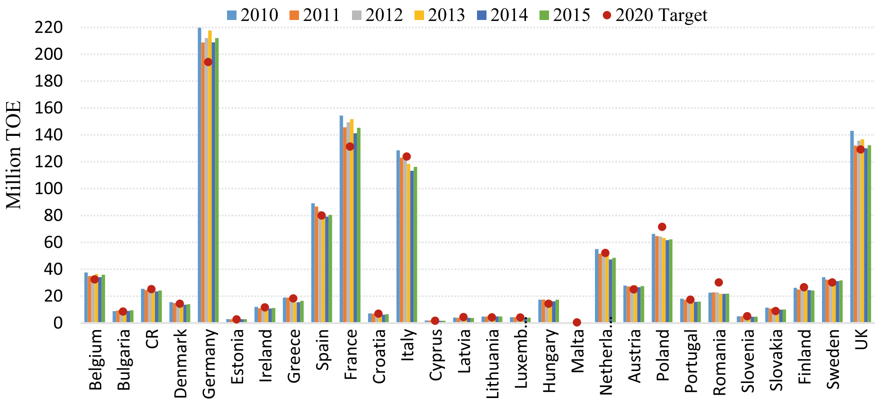


Fig. 5 Final energy consumption by country and 2020 target (2010–2015)

their own targets. From the member states that did not reach their targets, there are three major economies: Germany, France and United Kingdom. The environmental efficiency assessment of EU countries is explored in the next section taking into account both energy and non-energy resources in each country, assuring a fair performance comparison.

4.3 Empirical Results

The environmental efficiency for each country is assessed through the models previously introduced, CFG and (1), deriving the inefficiency scores β^*_{CFG} and β^*_1, respectively, by comparison to the best practices observed until 2015. The RE share is only an additional factor in model (1) to restrict the comparison of the assessed unit

Table 4 Environmental efficiency results through the models CFG and (1)

Model	2010		2011		2012		2013		2014		2015	
	β^*_{CFG}	β^*_1	β^*_{CFG}	β^*_1	β^*_{CFG}	β^*_1	β^*_{CFG}	β^*_1	β^*_{CFG}	β^*_1	β^*_{CFG}	β^*_1
Austria	0.099	0.095	0.102	0.100	0.097	0.096	0.102	0.101	0.088	0.088	0.092	0.091
Belgium	0.109	0.109	0.114	0.114	0.114	0.114	0.109	0.109	0.118	0.118	0.122	0.122
Bulgaria	0.142	0.132	0	0	0.097	0.083	0.157	0.130	0.164	0.141	0.133	0.113
Croatia	0.210	0.172	0.201	0.162	0.190	0.142	0.190	0.134	0.175	0.119	0.179	0.120
Cyprus	0.179	0.166	0.141	0.132	0.079	0.075	0.051	0.036	0	0	0.029	0.009
CR	0.309	0.309	0.306	0.306	0.300	0.300	0.294	0.294	0.293	0.293	0.299	0.299
Denmark	0.004	0.000	0	0	0	0	0.003	0	0	0	0	0
Estonia	0	0	0	0	0.216	0.035	0	0	0	0	0.247	0
Finland	0.127	0.127	0.128	0.125	0.123	0.112	0.108	0.098	0.096	0.081	0.092	0.073
France	0.087	0.087	0.078	0.078	0.077	0.077	0.071	0.071	0.057	0.057	0.056	0.056
Germany	0.073	0.073	0.079	0.079	0.075	0.075	0.069	0.069	0.072	0.072	0.070	0.070
Greece	0.087	0.087	0.084	0.084	0.004	0.004	0	0	0	0	0	0
Hungary	0.211	0.209	0.195	0.192	0.183	0.179	0.194	0.189	0.213	0.209	0.211	0.208
Ireland	0.014	0.014	0	0	0.011	0.011	0.015	0.015	0.029	0.028	0	0
Italy	0.090	0.090	0.078	0.078	0.053	0.053	0.031	0.030	0.017	0.017	0.023	0.023
Latvia	0.209	0	0.233	0.042	0.256	0.056	0.223	0.022	0.213	0	0.199	0
Lithuania	0.182	0.124	0.210	0.174	0.186	0.146	0.180	0.129	0.180	0.125	0.187	0.124
Luxembourg	0	0	0.004	0.004	0.018	0.018	0.011	0.011	0	0	0	0
Malta	0.084	0.084	0.029	0.029	0	0	0.025	0.024	0	0	0.122	0
Netherlands	0.069	0.069	0.078	0.078	0.059	0.059	0.043	0.043	0.041	0.041	0.071	0.071
Poland	0.221	0.221	0.275	0.275	0.262	0.262	0.250	0.250	0.288	0.288	0.286	0.286
Portugal	0.136	0.127	0.095	0.090	0.049	0.044	0.033	0.027	0.036	0.030	0.057	0.051
Romania	0.348	0.345	0.363	0.361	0.363	0.360	0.320	0.317	0.315	0.312	0.320	0.316
Slovakia	0.242	0.239	0.261	0.258	0.217	0.213	0.205	0.202	0.193	0.188	0.237	0.231
Slovenia	0.191	0.143	0.169	0.126	0.148	0.103	0.159	0.101	0.134	0.073	0.120	0.059
Spain	0.145	0.145	0.125	0.125	0.106	0.106	0.098	0.098	0.100	0.100	0.103	0.103
Sweden	0.023	0.023	0.017	0.017	0.008	0.008	0	0	0	0	0	0
UK	0.001	0.001	0	0	0.003	0.003	0.003	0.003	0.001	0.001	0	0
All	Model CFG						Model (1)					
Mean (SD)	0.115 (0.098)						0.097 (0.094)					
No of efficient units	25						31					

to peers. Note that the RE share is not used in CFG model. These results are reported in Table 4. β^*_{CFG} evaluates the capacity of each country in expanding GDP and contracting labor, energy, capital and GhG emissions, simultaneously, by the same proportion. β^*_1 evaluates the same capacity, by imposing to each assessed country the comparison to peers such as their linear combination in terms of RE share should be at least the share of RE observed in the assessed country.

It is observed that since 2010, the mean of environmental efficiency of the economies assessed by the model (1) is 0.097, which means that inputs and GhG emissions can reduce 0.097, on average, and the GDP can increase 0.097, on average (see Table 4). Globally, from 2010, the European countries have increased their environmental efficiency or some of them keep their efficiency status. Four groups of countries are observed taking into account the results from models CFG and (1): (i) countries efficient in both models; (ii) countries that become efficient only in model (1); (iii) inefficient countries that keep efficiency score in both models and (iv) inefficient countries that increase the efficiency in model (1). There are 25 units that are efficient in both models, such as Estonia (in 2010, 2011, 2013, 2014), Luxembourg (in 2010, 2014, 2015), Bulgaria (in 2011), Denmark (2011, 2012, 2014, 2015), Ireland (in 2011, 2015), UK (in 2011, 2015), Malta (in 2012, 2014), Greece (in 2013–2015), Sweden (in 2013–2015) and Cyprus (in 2014). These countries are the benchmarks as they have the best practices in terms of RE production policy and capacity to reduce inputs and GhG emissions and expand GDP, simultaneously. Generally Denmark, Sweden and Luxembourg keep their environmental efficiency status over time. Greece becomes efficient from 2013 and Ireland and UK become efficient in 2015.

Regarding the inefficient units in CFG model, there are only 6 units that become efficient using model (1), such as Denmark (in 2013), Estonia (in 2015), Latvia (in 2010, 2014, 2015) and Malta (2015). These countries should follow the best practices observed using model CFG and keep their RE production policy. Latvia becomes efficient from 2014 while Malta becomes efficient in 2015.

It is observed that 64 units keep inefficiency scores in both models. These countries should emulate the best practices observed on their benchmarks, in terms of energy consumed, GhG emissions and RE production policy. These countries are Belgium, Czech Republic, Germany, Spain, France, Poland, Netherlands during all period and Ireland (in 2010, 2012, 2013), Greece (in 2010–2012), Italy (in 2010–2012), Malta (in 2010, 2011), Finland (in 2010), Sweden (in 2010–2012), UK (in 2010, 2012–2014) and Luxembourg (in 2011–2013). From 2010, it is observed a consistent increasing trend in environmental efficiency of Finland, France, Greece and Sweden. On the other hand, Poland has decreased its environmental efficiency over the time.

There are 73 inefficient units that improve efficiency score in model (1). Thus, β_1^* reflects also the adjustment regarding the RE production policy observed in the assessed unit. This occurs in the case of Croatia, Lithuania, Hungary, Austria, Portugal, Romania, Slovenia and Slovakia during all period and Bulgaria (in 2010, 2012–2015), Estonia (in 2012), Denmark (in 2010), Cyprus (in 2010–2013, 2015), Finland (in 2011–2015), Italy (in 2013–2015), Ireland (in 2014), Latvia (in 2011–2013) and Malta (in 2013). Cyprus has significantly improved their efficiency since 2010.

As by-product of model (1), each inefficient country should follow the best practices of their benchmarks, learning strategies to improve its environmental efficiency. The managerial implications of the environmental efficiency results are explored in detail for Portugal, in 2015, using both models: CFG and (1). In the CFG model the peers are Greece (in 2015, $\lambda = 0.4021$) and UK (in 2015, $\lambda = 0.0527$), deriving a $\beta_{CFG}^* = 0.057$, meaning that the country could expand the GDP and reduce the inputs and the GhG emissions by 0.057. Applying the model (1), the environmental inefficiency decreases to 0.051, to accommodate the new constraint, reflecting an improvement on RE share. In this model, the peers are Cyprus (in 2014, $\lambda = 2.8315$), Greece (in 2015, $\lambda = 0.1539$) and UK (in 2015, $\lambda = 0.0506$), fulfilling the new constraint ($2.8315 \times 0.089 + 0.1539 \times 0.154 + 0.056 \times 0.085 = 0.28$). This outcome entails the reduction of the targets contribution of Greece and UK and the inclusion of an additional benchmark, Cyprus. The λ_j's also incorporate the influence of the inefficiencies due to the simultaneously expansion of GDP and reduction of the inputs and GhG. If there are not observed inefficiencies in terms of the current inputs and outputs (GDP, GhG), the country could become efficient. This occurred with Latvia in 2010, 2014, 2015 and Denmark in 2013. Bear also in mind, that Portugal in 2015 presents slacks regarding final energy consumption (slack = 1.42×10^6 TOE) and labor (slack=2405233x10^3 hours), which result in total reductions to the levels of $(1 - 0.051)16 - 1.42 = 13.76\text{x}10^6$ TOE and $(1 - 0.051)8579294 - 2405233 = 5735380\times10^3$ hours, respectively. In a similar way, Portugal can reduce the level of capital and GhG emissions by (1-0.051) to the levels 26287×10^6 € and 68399×10^3 tonnes, respectively. It also is possible, to increase the level of GDP to 180994×10^6 € , i.e., $172190(1 + 0.051)$.

4.4 EU Countries 2020 Targets and Environmental Efficiency Comparison

From the analyses presented above (Sects. 4.2 and 4.3), it is possible to complement the fulfilment of the EU countries 2020 targets with the environmental efficiency results, which take into account both energy and non-energy resources, in order to establish a relation between them. At the end of 2015, almost all countries have potential to further reduce the GhG emissions and the final energy consumed while increasing GDP except for the benchmarks in that year, i.e., Denmark, Luxembourg, Sweden, Ireland, Greece and UK. Denmark has already met their own targets and Sweden only surpass the final energy consumption target by a small margin. UK and Greece are still below the 2020 target on RE share, but well positioned to achieve the targets regarding GhG emissions and final energy consumption. Luxembourg and Ireland reached only the final energy consumption target, but according to DEA results, they are environmentally efficient. Finland has already achieved their own targets and, according to DEA assessment, it has had a consistent increasing trend in the environmental efficiency. Sweden only lacks the achievement on energy efficiency target (measured by the final energy consumption) and, taking into account the DEA results, it is environmentally efficient. Germany and France have only fulfilled the target concerning the GhG emissions target. The DEA model indicates that there is still potential to reduce GhG emissions and the final energy consumed. Poland presents a decrease in its environmental efficiency and also a bad performance metrics in terms of the EU climate and energy 2020 targets. In fact, the energy sector has been dominated by fossil fuels and only in the past decade, a new energy strategy has been introduced into political agenda to implement an energy mix and the necessary regulations to support renewables [39].

5 Conclusions

This work establishes an analysis of the environmental efficiency of European countries, under the EU 2020 climate and energy strategy. The environmental efficiency is assessed through a DEA analysis, measuring the ability of each country in minimizing current resources, while maximizing the GDP and minimizing GhG emissions, reflecting the environmental policymaking in Europe concerning climate change and energy targets. The DEA model is based on DDF, imposing weak disposability for the UO. A new model (1) is proposed establishing that efficiency comparison for each country is performed only using countries such that linear combination in terms of RE share is at least the share of RE observed in the assessed country, comparing also the practices in terms of RE production policy.

Results obtained show that globally the EU has increased its environmental efficiency which is consistent with the analysis of the indicators of the 2020 climate and energy package, during the 2010–2015 time-frame. Although the environmental efficiency scores vary among the EU member states, countries with higher GDP levels tend to have higher potential in reducing GhG emissions. The study produces a benchmark list (Denmark, Sweden, Luxembourg, Greece, Ireland and UK) in 2015 and also a set of countries which environmental efficiency has increased consistently in the period under analysis: Finland, France, Greece and Sweden. On the other hand, Poland is the only country whose environmental efficiency has been decreasing over time. In fact, Poland performs worse than most of other Eastern European countries and average EU in terms of emission reductions due to its strong dependence on coal. Also, the expansion of renewable energies has not been properly addressed by national incentives and legislation.

The DEA methodology can be used as a support tool in establishing the new targets for each one of the EU country and also in future target setting scenarios. Further analysis could be explored in terms of environmental performance assessment in other countries of the world, establishing their targets in order to guide them to sustainable environmental policies, fulfilling their commitments under the Kyoto Protocol.

References

1. Chambers, R., Chung, Y., Färe, R.: Benefit and distance functions. J. Econ. Theor. **70**, 407–419 (1996)
2. Chung, Y.: Directional distance functions and undesirable outputs. Ph.D. Dissertation, Southern Illinois University (1996)
3. Chung, Y., Färe, R., Grosskopf, S.: Productivity and undesirable outputs: a directional distance function approach. J. Environ. Manag. **51**, 229–240 (1997)
4. Sueyoshi, T., Yuan, Y., Goto, M.: A literature study for DEA applied to energy and environment. Energy Econ. **62**, 104–124 (2017)

5. Mardani, A., Zavadskas, E.K., Streimikiene, D., Ahmad Jusoh, A., Khoshnoudi, M.: A comprehensive review of data envelopment analysis (DEA) approach in energy efficiency. Renew. Sustain. Energy Rev. **70**, 1298–1322 (2017)
6. Zofio, J., Prieto, A.: Environmental efficiency and regulatory standards: the case of CO2 emissions from OECD industries. Resour. Energy Econ. **23**, 63–83 (2001)
7. Färe, R., Grosskopf, S., Hernandez-Sancho, F.: Environmental performance: an index number approach. Resour. Energy Econ. **26**, 343–352 (2004)
8. Zhou, P., Ang, B., Poh, K.: Slacks-based efficiency measures for modeling environmental performance. Ecol. Econ. **60**, 111–118 (2006)
9. Zhou, P., Poh, K., Ang, B.: A non-radial DEA approach to measuring environmental performance. Eur. J. Oper. Res. **178**, 1–9 (2007)
10. Zhou, P., Ang, B., Poh, K.: Measuring environmental performance under different environmental DEA technologies. Energy Econ. **30**, 1–14 (2008)
11. Chiu, C., Liou, J., Wu, P., Fang, C.: Measuring efficiency of decision-making units. Energy Econ. **34**, 1392–1399 (2012)
12. Lin, E., Chen, P., Chen, C.: Measuring green productivity of country: a generalized metafrontier Malmquist productivity index approach. Energy **55**, 340–353 (2013)
13. Camarero, M., Castillo, J., Picazo-Tadeo, A., Tamarit, C.: Eco-efficiency and convergence in OECD countries. Environ. Resour. Econ. **55**, 87–106 (2013)
14. Lin, E., Chen, P., Chen, C.: Measuring the environmental efficiency of countries: a directional distance function metafrontier approach. J. Environ. Manag. **119**, 134–142 (2013)
15. Bampatsou, C., Papadopoulos, S., Zervas, E.: Technical efficiency of economic systems of EU-15 countries based on energy consumption. Energ. Policy **55**, 426–434 (2013)
16. Vaz, C., Ferreira, Â.: Measuring technical efficiency of european countries using DEA. In: Godinho, P., Dias, J. (eds), Assessment Methodologies: Energy, Mobility and Other Real World Application. Imprensa da Universidade de Coimbra, pp. 17–32 (2015)
17. Li, M., Wang, Q.: International environmental efficiency differences and their determinants. Energy **78**, 411–420 (2014)
18. Chang, M.: Energy intensity, target level of energy intensity, and room for improvement in energy intensity: an application to the study of regions in the EU. Energ Policy **67**, 648–655 (2014)
19. Pang, R., Deng, Z., Hu, J.: Clean energy use and total-factor efficiencies: an international comparison. Renew. Sustain. Energy Rev. **52**, 1158–1171 (2015)
20. Rashidi, K., Shabani, A., Saen, R.: Using data envelopment analysis for estimating energy saving and undesirable output abatement: a case study in the Organization for Economic Co-Operation and Development (OECD) countries. J. Clean. Prod. **105**, 241–252 (2015)
21. Robaina, M., Moutinho, V., Macedo, P.: A new frontier approach to model the eco-efficiency in European countries. J. Clean. Prod. **103**, 562–573 (2015)
22. Apergis, N., Aye, G., Barros, C., Gupta, R., Wanke, P.: Energy efficiency of selected OECD countries: a slacks based model with undesirable outputs. Energy Econ. **51**, 45–53 (2015)
23. Woo, C., Chung, Y., Chun, D., Seo, H., Hong, S.: The static and dynamic environmental efficiency of renewable energy: a Malmquist index analysis of OECD countries. Renew. Sustain. Energy Rev. **47**, 367–376 (2015)
24. Madaleno, M., Moutinho, V., Robaina, M.: Economic and Environmental assessment: EU cross-country efficiency ranking analysis. Energy Proced. **106**, 134–154 (2016)
25. Gómez-Calvet, R., Conesa, D., Gómez-Calvet, A., Tortosa-Ausina, E.: On the dynamics of eco-efficiency performance in the European Union. Comput. Oper. Res. **66**, 336–350 (2016)
26. Moutinho, V., Madaleno, M., Robaina, M.: The economic and environmental efficiency assessment in EU cross-country: evidence from DEA and quantile regression approach. Ecol. Indic **78**, 85–97 (2017)
27. Beltrán-Esteve, M., Picazo-Tadeo, A.: Assessing environmental performance in the European Union: Eco-innovation versus catching-up. Energ. Policy **104**, 240–252 (2017)
28. Duman, Y., Kasman, A.: Environmental technical efficiency in EU member and candidate countries: a parametric hyperbolic distance function approach. Energy **147**, 297–307 (2018)

29. Zhou, P., Ang, B., Poh, K.: A survey of data envelopment analysis in energy and environmental studies. Eur. J. Oper. Res. **189**(1), 1–18 (2008)
30. Färe, R., Grosskopf, S., Tyteca, D.: An activity analysis model of the environmental performance of firms - application to fossil-fuel-fired electric utilities. Ecol. Econ. **18**, 161–175 (1996)
31. Färe, R., Grosskopf, S., Lovell, C., Pasurka, C.: Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach. Rev. Econ. Stat. **71**, 90–98 (1989)
32. Färe, R., Grosskopf, S.: Theory and application of directional distance functions. J. Prod. Anal. **13**, 93–103 (2000)
33. Charnes, A., Cooper, W., Rhodes, E.: Measuring efficiency of decision-making units. Eur. J. Oper. Res. **2**(6), 429–444 (1978)
34. Luenberger, D.: Benefit functions and duality. J. Math. Econ. **21**, 461–481 (1992)
35. Luenberger, D.: Microeconomic Theory. McGraw-Hill, New York (1995)
36. Skjaerseth, J., Wettestad, J.: EU Emissions Trading. Routledge, Taylor & Francis Group, London (2013)
37. European Commission. Climate strategies & targets. Accessed Feb 2018
38. Eurostat. Europe 2020 indicators climate change and energy (2017). Accessed Feb 2018
39. Skjaerseth, J.: Implementing EU climate and energy policies in Poland: policy feedback and reform. Environ. Polit. **27**, 498–518 (2018)

A Column Generation-Based Diving Heuristic for Staff Scheduling at an Emergency Medical Service

Hendrik Vermuyten, Joana Namorado Rosa, Inês Marques, Jeroen Beliën and Ana Barbosa-Póvoa

Abstract Staff scheduling involves assigning people to tasks organized in working shifts. It is a complex and time-consuming activity common to several real-world companies while still typically a hand-made task. These problems are usually conditioned by legal and working rules, and by personal preferences. Thus, the challenge is to find schedules that most accurately fit the functionality of the services and equity issues. For this purpose, a column generation-based diving heuristic is proposed to solve a staff scheduling problem at an Emergency Medical Service. The approach is generic and possibly adjusted to several realities and companies. In this context, the heuristic is applied to a real-life problem at Instituto Nacional de Emergência Médica (INEM), obtaining good quality solutions in relatively short running times. The best-found solution is compared with an implemented schedule at INEM, strengthening the practical value of this approach. The ultimate goal is to develop automated tools to support INEM in their staff scheduling activities.

Keywords OR in health services · Emergency medical services · Staff scheduling · Diving heuristic · Column generation

H. Vermuyten (✉) · J. Namorado Rosa · I. Marques · A. Barbosa-Póvoa
Centre for Management Studies, Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
e-mail: hendrik.vermuyten@kuleuven.be

J. Namorado Rosa
e-mail: joana.namorado.rosa@gmail.com

I. Marques
e-mail: ines.marques.p@tecnico.ulisboa.pt

A. Barbosa-Póvoa
e-mail: apovoa@tecnico.ulisboa.pt

H. Vermuyten · J. Beliën
Faculty of Economics and Business, Department of Information Management, Modeling and Simulation, KU Leuven Campus Brussels, Warmoesberg 26, 1000 Brussels, Belgium
e-mail: jeroen.belien@kuleuven.be

M. J. Alves et al. (eds.), *Operational Research*, Springer Proceedings in Mathematics & Statistics 278, https://doi.org/10.1007/978-3-030-10731-4_16

1 Introduction

Staff scheduling is the process of deploying timetables for a set of workers within an organization so as to satisfy demand for various services, while simultaneously ensuring a distinctive level of employee satisfaction. Moreover, staff scheduling also needs to consider legal, organizational and contractual constraints [1]. In Emergency Medical Services (EMSs), staff scheduling is of paramount importance, since shortages in the number of required personnel directly impact the quality of care that patients receive. Furthermore, employee satisfaction cannot be neglected, as undesirable schedules can lead to increased staff turnover and to poorer on-job-performances.

Personnel scheduling problems arise in a wide variety of settings, such as transportation systems, call centers, and health care systems. Different reviews of the literature have been published and propose different classifications for staff scheduling problems [1–3]. In the health care sector, workforce scheduling have been largely dedicated to nurse and physician scheduling. Reviews of models and solution methods for nurse or physician scheduling are also available in the literature [4, 5]. Nevertheless, only few papers deal with staff scheduling in EMSs [6] and are usually dedicated to ambulances by combining the so-called crew scheduling and crew pairing problems. The domain of dispatch centers has not received much attention and is proposed as a potential future research area in [7].

Many solution techniques are applied to solve personnel scheduling problems. A straightforward approach is to model the problem as a standard integer programming (IP) formulation and solve it using a commercial solver [8]. Other researchers use branch-and-price [9], which uses column generation (CG) in each node in the branch-and-bound tree to solve the linear programming (LP) relaxation. Despite the vast improvements in computer hardware and commercial solvers, the high complexity of staff scheduling problems make them difficult to solve to optimality. The easy-to-implement feature of heuristics and the ability to deal with complex constraints or objectives make them well suited to personnel scheduling problems. Genetic algorithms are a popular type of metaheuristic in this field [10]. Recently, MixedIP-based heuristics have been succesfully applied to workforce scheduling [11]. These approaches combine the strengths of both mixed IP and (meta)heuristic approaches. Column generation is also used inside a heuristic framework instead of standard MixedIP approaches [12]. Diving heuristics can be used to obtain integer feasible solutions. Diving algorithms heuristically select branches in the branch-and-price tree using a certain rounding strategy until the first integer solution is found [13].

This paper addresses a staff scheduling problem motivated by the real-life context at Instituto Nacional de Emergência Médica (INEM). The aim is twofold. First, it intends to propose a novel solution approach for the staff scheduling problem at INEM that may be applied to other real-life settings. Second, it aims to enhance INEM productivity and profitability levels, by noticing and proposing solutions for their current scheduling conflicts.

The remainder of this paper is structured as follows. Section 2 introduces the staff scheduling problem and a standard IP model. Next, a column generation-based diving heuristic is proposed in Sect. 3. In Sect. 4, the heuristic is applied to a real-life dataset provided by INEM. Finally, Sect. 5 concludes the paper.

2 Staff Scheduling at an Emergency Medical Service

The EMS is organized in different services and each service is divided in several teams. Each team is responsible for a set of tasks and has a set of staff members with the required skills to perform those tasks. All staff members are assigned to one service and may belong to one or more teams. Preferably workers perform tasks in their own team(s), but it is possible to do tasks from other teams or services to meet the required demand. However, this should be avoided.

This personnel scheduling problem integrates the staffing of the set of services that may share the same workforce. Each service operates 7 days per week and 24 hours per day. Each day of the planning horizon is divided into shifts in which tasks are performed (night, morning and afternoon shifts) and the workload required to each shift (staff level) is known in advance. In this, the problem differs from most problems in the context of staff scheduling at an emergency service, where shift lengths and workload demands are usually not considered as fixed. The reason is that our model schedules only technical personnel for which the demand can be predicted in advance reasonably well.

Legal rules and organizational and contractual issues which need to be satisfied to obtain feasible schedules give rise to the following hard constraints. Legal rules require a minimum resting time between working shifts. Working time regulations set a maximum number of consecutive working days and of consecutive days off for each staff member. Workers must have a minimum number of Sundays off over the planning period. Furthermore, employees are only assigned to tasks for which they have the required skills. For reasons of equity between workers, every staff member needs to work a minimum number of each shift type.

The objective function of the schedules consists of two elements. The primary objective is to ensure functionality of the services. The workload demand should be satisfied as much as possible and is formulated as soft constraints. Thus understaffing and overstaffing are allowed at a penalization cost to be minimized. A second objective for this staff scheduling problem is to increase employee satisfaction and equity among the staff. Workers should have the entire weekend off instead of a single day. The contract hours of each staff member should preferably be met, meaning both overtime and undertime are undesirable. Finally, the number of tasks assigned to staff members belonging to other teams should be minimized.

A standard IP model for this staff scheduling problem is formulated in (1)–(14). The notation is summarized in Table 1. Functionality of the services is modeled by constraints (2)–(8) whereas equity among the staff is formulated by constraints (9)–(12).

$$\min \quad \sum_{d \in \mathrm{D}} \sum_{s \in \mathrm{S}} \sum_{j \in \mathrm{J}} \sum_{t \in \mathrm{T}_j^{\mathrm{J}}} \left(w_j^{RE+} Y_{tds}^{RE+} + w_j^{RE-} Y_{tds}^{RE-} \right) + w^{WO} \sum_{p \in \mathrm{P}} \sum_{w \in \mathrm{W}} \left(Y_{pw}^{WO+} + Y_{pw}^{WO-} \right)$$

$$+ \sum_{p \in \mathrm{P}} \left(w^{H+} Y_p^{H+} + w^{H-} Y_p^{H-} \right) + \sum_{j \in \mathrm{J}} \sum_{g \in \mathrm{G_j}} w_j^G Y_g^G \tag{1}$$

$$\text{s.t.:} \quad \sum_{p \in \mathrm{P}_t^T} x_{ptds} - Y_{tds}^{RE+} + Y_{tds}^{RE-} = R_{tds} \quad \forall t \in \mathrm{T}, d \in \mathrm{D}, s \in \mathrm{S} \tag{2}$$

$$\sum_{t \in \mathrm{T}_p^{\mathrm{P}}} \left(x_{ptd,\mathrm{N}} + x_{ptd,\mathrm{M}} + x_{ptd,\mathrm{A}} \right) \leq 1 \quad \forall p \in \mathrm{P}, d \in \mathrm{D} \tag{3}$$

$$\sum_{t \in \mathrm{T}_p^{\mathrm{P}}} \left(x_{ptd,\mathrm{M}} + x_{ptd,\mathrm{A}} + x_{pt,d+1,\mathrm{N}} \right) \leq 1 \quad \forall p \in \mathrm{P}, d \in \mathrm{D} \setminus \{|\mathrm{D}|\} \tag{4}$$

$$\sum_{t \in \mathrm{T}_p^{\mathrm{P}}} \left(x_{ptd,\mathrm{A}} + x_{pt,d+1,\mathrm{N}} + x_{pt,d+1,\mathrm{M}} \right) \leq 1 \quad \forall p \in \mathrm{P}, d \in \mathrm{D} \setminus \{|\mathrm{D}|\} \tag{5}$$

$$\sum_{t \in \mathrm{T}_p^{\mathrm{P}}} \sum_{r \in \{d, d+1, \ldots, d+\theta^{WD}\}} \sum_{s \in \mathrm{S}} x_{ptrs} \leq \theta^{WD}$$

$$\forall p \in \mathrm{P}, d \in \mathrm{D} \setminus \{|\mathrm{D}|, |\mathrm{D}|-1, \ldots, |\mathrm{D}| - \theta^{WD} + 1\} \tag{6}$$

$$\sum_{t \in \mathrm{T}_p^{\mathrm{P}}} \sum_{r \in \{d, d+1, \ldots, d+\theta^{DO}\}} \sum_{s \in \mathrm{S}} x_{ptrs} \geq 1$$

$$\forall p \in \mathrm{P}, d \in \mathrm{D} \setminus \{|\mathrm{D}|, |\mathrm{D}|-1, \ldots, |\mathrm{D}| - \theta^{DO} + 1\} \tag{7}$$

$$\sum_{t \in \mathrm{T}_p^{\mathrm{P}}} \sum_{d \in \{7-\kappa, 7-\kappa+7, \ldots\}} \sum_{s \in \mathrm{S}} x_{ptds} \leq |W| - \theta^{SO} \quad \forall p \in \mathrm{P} \tag{8}$$

$$\sum_{t \in \mathrm{T}_p^{\mathrm{P}}} \sum_{d \in \mathrm{D}} x_{ptds} \geq \theta_s^{ST} \quad \forall p \in \mathrm{P}, \forall s \in \mathrm{S} \tag{9}$$

$$\sum_{t \in \mathrm{T}_p^{\mathrm{P}}} \sum_{s \in \mathrm{S}} \left(x_{pt,7(w-1)+7-\kappa,s} - x_{pt,7(w-1)+6-\kappa,s} \right) - Y_{pw}^{WO+} + Y_{pw}^{WO-} = 0$$

$$\forall p \in \mathrm{P}, w \in \mathrm{W} \tag{10}$$

$$\sum_{t \in \mathrm{T}_p^{\mathrm{P}}} \sum_{d \in \mathrm{D}} \sum_{s \in \mathrm{S}} L_t x_{ptds} - Y_p^{H+} + Y_p^{H-} = \Theta_p \quad \forall p \in \mathrm{P} \tag{11}$$

$$\sum_{p \in \mathrm{P}_g^G} \sum_{t \in \mathrm{T}_p^{\mathrm{P}} \setminus \mathrm{T}_g^G} \sum_{d \in \mathrm{D}} \sum_{s \in \mathrm{S}} x_{ptds} - Y_g^G = 0 \quad \forall g \in \mathrm{G} \tag{12}$$

$$Y_{tds}^{RE+}, Y_{tds}^{RE-}, Y_{pw}^{WO+}, Y_{pw}^{WO-}, Y_p^{H+}, Y_p^{H-}, Y_g^G \geq 0$$

$$\forall p \in P, t \in T, d \in D, s \in S, w \in W, g \in G \tag{13}$$

$$x_{ptds} \in \{0, 1\} \quad \forall p \in P, t \in \mathrm{T}_p^{\mathrm{P}}, d \in D, s \in S \tag{14}$$

This is a goal programming model where undesired deviations to target values are penalized and minimized as the weighted sum objective function (1). Constraints (2) model the workload demand where understaffing and overstaffing are allowed at a penalty cost minimized in the first term of the objective function. Constraints (3), (4), and (5) force a minimum 2-shifts resting time between consecutive working shifts. Staff members must work a maximum of θ^{WD} consecutive days (6) and have at most θ^{DO} consecutive days off (7). Constraints (8) guarantee at least θ^{SO} Sundays off over the planning period. Constraints (9) set a minimum number of each shift type to be performed by each staff member. The soft requirements contributing for the equity among staff are formulated in (10)–(12), respectively: full weekends off; contract hours; and tasks of other teams performed by members of team g. Deviations to the corresponding target values are penalized, respectively, in the last three terms of the objective function.Regarding the penalization scheme for full weekends off (constraints (10)), the requirements outlined by INEM were followed: a penalty is incurred if a staff member has to work exactly one day in the weekend; and no penalty is incurred if a staff member has to work both weekend days or has both weekend days-off. Thus, as we have a minimization problem, Y_{pw}^{WO+} equals one if staff member p has to work Sunday but not Saturday, while Y_{pw}^{WO-} equals one if staff member p has to work Saturday but not Sunday. Constraints (13) represent the domain for the deviation variables, and constraints (14) model the binary domain for the decision variables.

3 Column Generation-Based Diving Heuristic

Staff scheduling problems are highly complex and commercial solvers are not able to solve these problems effectively when there are a large number of staff using standard IP models. This paper proposes a hybrid approach combining CG and a diving heuristic.

The column generation is first presented in Sect. 3.1 and the diving heuristic is proposed in Sect. 3.2.

3.1 Column Generation

The standard IP model (1)–(14) is reformulated in terms of work patterns (or columns) for each staff member consisting on the tasks to be perfomed during the planning period. The model can be reformulated in (15)–(19) using the additional notation in Table 2.

Table 1 Sets, parameters and variables for the IP model

Sets and indices	
$p \in \mathrm{P}$	Available staff
$t \in \mathrm{T}$	Tasks
$d \in \mathrm{D}$	Days in the planning horizon; the first day of the planning horizon is $d = 1$
$w \in \mathrm{W}$	Full weekends in the planning horizon
$s \in \mathrm{S}$	Shifts, i.e. $\mathrm{S} = \{\mathrm{N, M, A}\}$, (N = Night, M = Morning, A = Afternoon)
$g \in \mathrm{G}$	Teams
$j \in \mathrm{J}$	Services
Subsets	
$\mathrm{P}_t^{\mathrm{T}}$	Staff skilled to perform task t
$\mathrm{T}_p^{\mathrm{P}}$	Tasks that can be performed by staff member p
$\mathrm{P}_g^{\mathrm{G}}$	Staff from team g
$\mathrm{T}_g^{\mathrm{G}}$	Tasks from team g
$\mathrm{T}_j^{\mathrm{J}}$	Tasks from service j
G_j	Teams from service j
Parameters	
κ	First planning day (0 = Monday, 1 = Tuesday, ..., 6 = Sunday)
Θ_p	Contract hours of staff member p, adjusted for holidays
L_t	Duration of task t
R_{tds}	Workload demand (service level) for task t on shift s on day d
θ^{WD}	Maximum number of consecutive working days
θ^{DO}	Maximum number of consecutive days off
θ^{SO}	Minimum number of sundays off
θ_s^{ST}	Minimum number of shift type s scheduled
w_j^{RE+}, w_j^{RE-}	Penalty for overstaffing and understaffing service j, respectively
w^{WO}	Penalty for full weekend off
w^{H+}, w^{H-}	Penalty for excess and shortage hours worked, respectively
w_j^{G}	Penalty for assigning tasks of a team to members of another team in service j
Decision and auxiliary variables	
x_{ptds}	1, if staff member p is assigned to task t on shift s on day d; 0, otherwise
Y_{tds}^{RE+}, Y_{tds}^{RE-}	Overstaffing and understaffing for task t on shift s on day d, respectively
Y_{pw}^{WO+}, Y_{pw}^{WO-}	Penalty variables when staff member p has to work exactly one day in weekend w
	(Sunday and Saturday, respectively)
Y_p^{H+}, Y_p^{H-}	Excess and shortage hours worked for staff member p
Y_g^{G}	The number of tasks of other teams performed by members of team g

Table 2 New sets, parameters and variables

Sets and indices	
$k \in \mathrm{K}_p$	Work patterns for staff member p
Parameters	
a_{pktds}	1, if staff member p is assigned to task t on shift s on day d in work pattern k; 0, otherwise
c_{pk}	Cost of work pattern k for staff member p (based on the violations of the soft constraints)
Decision variables	
z_{pk}	1, if staff member p is assigned to work pattern k; 0, otherwise
a_{tds}	1 if task t is assigned on shift s on day d; 0 otherwise

$$\min \quad \sum_{p\in \mathrm{P}} \sum_{k\in \mathrm{K}_p} \left(c_{pk}\, z_{pk}\right) + \sum_{d\in \mathrm{D}} \sum_{s\in \mathrm{S}} \sum_{j\in \mathrm{J}} \sum_{t\in \mathrm{T}_j^{\mathrm{J}}} \left(w_j^{RE+}\, Y_{tds}^{RE+} + w_j^{RE-}\, Y_{tds}^{RE-}\right) \tag{15}$$

$$\text{s.t.:} \quad \sum_{p\in \mathrm{P}} \sum_{k\in \mathrm{K}_p} a_{pktds}\, z_{pk} - Y_{tds}^{RE+} + Y_{tds}^{RE-} = R_{tds} \quad \forall t \in \mathrm{T}, d \in \mathrm{D}, s \in \mathrm{S} \tag{16}$$

$$\sum_{k\in \mathrm{K}_p} z_{pk} = 1 \quad \forall p \in \mathrm{P} \tag{17}$$

$$z_{pk} \in \{0, 1\} \quad \forall p \in P, k \in \mathrm{K}_p \tag{18}$$

$$Y_{tds}^{RE+}, Y_{tds}^{RE-} \geq 0 \quad \forall t \in T, d \in D, s \in S \tag{19}$$

Objective function (15) minimizes the cost of the work patterns assigned to each staff member measured by the deviations to the target values of the soft requirements contributing for the equity among staff, and the penalty associated with understaffing and overstaffing. Constraints (16) model the workload demand. Constraints (17) enforce the assignment of one work pattern to each staff member. The variables' domain are defined in constraints (18) and (19).

In the CG approach, the LP relaxation of model (15)–(19) (master problem) is initialized with a limited set of work patterns. New promising work patterns (columns) are iteratively added into the master problem. These promising work patterns are generated by the subproblems (pricing problems) solved for each staff member in each iteration of the algorithm. These subproblems consider the constraints related to each staff member individually. The process is repeated until no more promising work patterns (i.e. columns with negative reduced cost) are obtained by solving the subproblems for all the staff. The reduced cost (RC) of a new work pattern k for staff member p is then given by:

$$\mathrm{RC} = c_{pk} - \sum_{t\in \mathrm{T}} \sum_{d\in \mathrm{D}} \sum_{s\in \mathrm{S}} a_{pktds}\, \lambda_{tds} - \mu_p \tag{20}$$

where λ_{tds} and μ_p represent the dual costs associated with constraints (16) and (17), respectively.

The objective function of the pricing problem for each staff member p can then be defined in (21) while the set of constraints consists of Eqs. (3)–(13) with the new defined variables a_{tds} instead of variables x_{ptds}.

$$\begin{aligned}\min \quad & \sum_{t \in \mathrm{T}} \sum_{d \in \mathrm{D}} \sum_{s \in \mathrm{S}} (-\lambda_{tds}\, a_{tds}) + \sum_{w \in \mathrm{W}} \left(w^{WO}\, Y_w^{WO+} + w^{WO}\, Y_w^{WO-} \right) \\ & + w^{H+}\, Y^{H+} + w^{H-}\, Y^{H-} + \sum_{j \in \mathrm{J}} \sum_{g \in \mathrm{G}_j} w_j^G\, Y_g^G \end{aligned} \tag{21}$$

The column generation loop can be implemented in two different ways. Strategy E solves a pricing problem for every staff member in each iteration of the column generation. Thus, in each iteration the number of subproblems solved equals the number of available staff and all the work patterns with negative reduced cost are added to the master problem. Strategy P solves only one pricing problem for a given staff member and a work pattern is added to the master problem in each iteration of the column generation if it has a negative reduced cost.

3.2 Diving Heuristic

The solution obtained by the column generation might fail integrality and thus not represent a feasible staff schedule. A diving heuristic allows to traverse a branch-and-price tree in a depth-first manner until finding a feasible solution, thus speeding up the search for a good integer solution. In the diving heuristic, some integer variables are fixed and the linear program is resolved. The fixing and resolving is iterated until either an integral solution is found or the linear program becomes infeasible.

Two strategies are considered for the branching scheme. Strategy L fixes the largest fractional variable to one while strategy T fixes to one all the fractional variables above a threshold value δ. Each iteration of the heuristic defines the schedule for at least one staff member and the corresponding variables can be removed from the problem. The algorithm ends when a work pattern has been assigned for every staff member.

4 Case Study at INEM

4.1 Case Study

The column generation-based diving heuristic is applied in the context of the staff scheduling problem at INEM. This EMS is organized in two main services: the dispatch center (*Centro de Orientação de Doentes Urgentes*, CODU), and the Emergency Vehicles (EVs). CODU receives the emergency calls, delivers assistance and

Table 3 Penalty values used in the computational tests

Penalty	w^{RE+}_{CODU}	w^{RE+}_{EV}	w^{RE-}_{CODU}	w^{RE-}_{EV}	w^{WO}	w^{H+}	w^{H-}	w^{G}_{CODU}	w^{G}_{EV}
Value	10	10	100	1000	10	1	1	10	20

instructions, and dispatches the convenient EV. EVs move to the scene, the technicians provide assistance and care on scene, and transport the victims to the hospital if required. The scope of the study is the Lisbon region which includes neighborhood areas coordinated by the Lisbon center (Almada, Amadora, Cascais, Elvas, Estremoz, Ponte de Sôr, Sacavém, Seixal, Setúbal, Tomar and Torres Novas). The workforce at INEM consists of *Técnicos de Emergência Pré-Hospitalar* (TEPHs). Each TEPH is assigned to one service, although they can also perform tasks at other service. Within a service, TEPHs may belong to more than one team. The assignment of TEPHs to teams depends on the required skills and on the place of residence.

Work rules at INEM state that staff: cannot work more than 6 consecutive days ($\theta^{WD} = 6$); cannot have more than 5 consecutive days-off ($\theta^{DO} = 5$); must have at least one Sunday off every four weeks ($\theta^{SO} = 1$). For equity concerns, it is required that each staff member works at least 2 shifts of each type ($\theta^{ST}_s = 2, \forall s \in S$). Finally, a standard contract specifies a monthly working time of 140 hours ($\Theta_p = 140, \forall p \in P$). All tasks have a duration of 8 h, except for a single type of tasks, emergency motorcylces, which has a duration of 12 hours. Table 3 shows the penalty values used in the objective function (1). The most important objective is to ensure functionality of the services. This means that understaffing is worse than overstaffing. Moreover, the call center (CODU) can still function if there is some understaffing, while the emergency vehicles (EV) are more sensitive to understaffing. Therefore, $w^{RE-}_{\text{CODU}} = 100$, $w^{RE-}_{\text{EV}} = 1000$, and $w^{RE+}_{\text{CODU}} = w^{RE+}_{\text{EV}} = 10$. The second objective is that every staff member works the amount of hours stipulated in his or her contract (w^{H+} and w^{H-}), that staff members receive full weekends off (w^{WO}), and that tasks are assigned to members of the own team as much as possible (w^{G}_{CODU} and w^{G}_{EV}). Therefore, the values for these weights are smaller. The different weights are measured in different units or dimensions, and thus e.g. w^{H+} and w^{H-} are equal to 1, while w^{WO} is set equal to 10. The INEM dataset includes 289 TEPHs which are organized in 22 teams (5 in CODU and 17 in EVs), and need to be assigned to 61 tasks (10 in CODU and 51 in EVs) within a 4-weeks planning period.

4.2 Results

The standard IP model and the column generation-based diving heuristic were coded in C++14 and compiled with Microsoft Visual Studio 2015 using the callable library of ILOG CPLEX 12.6.2. All tests ran on a PC with an Intel Core i5-5200U CPU of 2.20 GHz and 8 GB of RAM under Windows 10.

Table 4 Results for INEM

Configuration	BFS	$\text{Time}_{\text{total}}$	$\text{COLS}_{\text{total}}$	Obj. root	$\text{Time}_{\text{root}}$	$\text{COLS}_{\text{root}}$	Gap
IP model (1)–(14)	1,901,916	>18,000	–	37,021	159	–	98.05
E/L/2-2	2,137,340	52	578	2,137,340[a]	48	578	98.27
E/L/3-3	–	>18,000	–	1,727,850[a]	57	867	–
E/L/10-1	–	>18,000	–	515,331[a]	224	2890	–
E/T/2-2	2,137,340	35	578	2,137,340[a]	32	578	98.27
E/T/3-3	189,732	488	6075	1,727,850[a]	58	867	80.49
E/T/4-4	164,796	587	7044	1,437,690[a]	72	1156	77.54
E/T/10-1	204,992	434	5082	515,331[a]	224	2890	81.94
P/L/2-2	–	>18,000	–	103,131[a]	58	578	–
P/L/10-1	–	>18,000	–	61,702.8[a]	1765	2890	–
P/T/2-2	92,640	287	1905	103,131[a]	58	578	60.04
P/T/3-3	89,368	1592	4006	86,627.7[a]	130	867	58.57
P/T/4-4	91,434	2432	5042	80,397.6[a]	262	1156	59.51
P/T/10-1	71,112	7852	9181	61,702.8[a]	1765	2890	47.94

[a] Since column generation is stopped prematurely, this is not an actual lower bound

The following notation is used to refer to the different combination of strategies for the column generation-based diving heuristic: A/B/C-D, where A is the column generation loop strategy (E or P), B is the diving strategy (L or T with threshold $\delta = 0.6$), and C-D refer to the stopping criteria used for the column generation (respectively, the number of iterations for the root node, β^{root}, and for each node on the diving, β^{diving}).

The results for the INEM instance are shown in Table 4. In this table, BFS is the objective function value of the best found solution, $\text{Time}_{\text{total}}$ is the total computation time (in seconds), $\text{COLS}_{\text{total}}$ is the total number of columns added, 'Obj. root' is the objective value of the LP relaxation of the root node, $\text{Time}_{\text{root}}$ is the computation time (in seconds) of the root node, $\text{COLS}_{\text{root}}$ is the number of columns added in the root node, and 'Gap' is an upper bound on the optimality gap in percent based on the LP relaxation solved by CPLEX. Since the optimal solution is not known, the optimality gap is calculated relative to the objective value LP relaxation, which is never worse than the objective value of the IP problem. Therefore, this gives an upper bound on the optimality gap. Compared to the standard IP approach, the column generation-based diving heuristic with configuration P/T/10-1 finds a solution almost 27 times better in significantly less time (7852 s).

From the results of Table 4, three observations can be made. First, branching on the largest fractional variable (L) is much slower than branching on all variables (T) with a value above threshold δ. The former branching method is not able to find any solution within the 5 h time limit, irrespective of the column generation scheme and the values for β^{root} and β^{diving}. The only exception is E/L/2-2, but this is a very poor solution. Second, the column generation strategy has an important impact on the solution quality. The strategy where a subproblem is solved for every staff member p (E) is much less efficient. The reason for this is the high level of symmetry in the data as most people have the skills to do most of the tasks. As such, given the same dual vector, for almost every staff member the same work pattern is generated. However, in the master only one or a few of those work patterns are useful to meet the demand for a given set of tasks on the different days. On the other hand, in strategy P the dual variables are always updated after the addition of a single work pattern and thus the subproblems for the other staff generate work patterns with different activities. When the column generation phase is terminated prematurely, this leads to the quality of the solutions found by the former column generation strategy being worse as the available work patterns cannot be combined into a good solution. This can also be seen from the objective value obtained in the root node. Third, increasing β^{root} is more beneficial for the solution quality than increasing β^{diving}.

4.3 Comparison with an Implemented Schedule

The best solution (configuration P/T/10-1) is compared with an implemented schedule at INEM. Some indicators, used for comparison, are summarized in Table 5.

In the INEM schedule, TEPHs are working below their contractual hours. On the other hand, in the best solution a TEPH exceeds his/her working hours by only 3.61% on average. Indeed, in our approach failing to meet workload demands carries a higher penalty than overtime, and therefore overtime is being used to satisfy the workload demand as much as possible. More than one Sundays-off is considered on average, which is in line with hard constraints (8). The entire weekend off is modeled as a soft constraint and therefore this value is lower than in the INEM schedule. The INEM real case gives greater importance to this personnel preference, while the solution case schedule seeks primarily to satisfy the workload demand.

Therefore we may conclude that the column generation-based diving heuristic is an improvement over the time-consuming manual scheduling procedure currently in use at INEM.

Table 5 Comparison between the best solution found and an implemented schedule at INEM

Indicator (per staff member)	INEM schedule			Best solution found		
	CODU	EVs	INEM	CODU	EVs	INEM
Average hours worked (% of contractual hours)	106.9	82.8	88.7	107.8	102.3	103.6
Average number of sundays off	1.69	1.09	1.24	1.18	1.38	1.33
Average number of weekends off	0.96	1.01	0.99	0.67	0.96	0.90

5 Conclusions

This paper addresses a real-life staff scheduling at an EMS. A column generation-based diving heuristic is proposed, decomposing the problem on the staff members and diving to obtain integer solutions. Two column generation strategies and two branching schemes are implemented. The best configuration is the one that solves a single pricing problem for each staff member in each iteration of the column generation, and rounds to one all the variables above a certain threshold as branching scheme for the diving heuristic. The solution approach is tested on the case study of INEM, the Portuguese EMS, and obtains qood quality solutions in reasonable computation time. The best solution found is compared with an implemented schedule at INEM and shows good practical value and potential to be embedded in an expert system to support staff scheduling decisions. Moreover, this personnel problem integrates the scheduling of staff for several services which is a novelty in the literature of workforce scheduling at EMSs.

The performance of the algorithm in datasets with different dimensions (not described in this paper) shows that the solution approach can easily handle more staff and is robust for the level of problem symmetry (measured in terms of the number of staff members that can perform each task) but performs poorly when the planning period is extended. Together with the slow convergence of the column generation this suggests that the pricing problems are the bottleneck of the algorithm. Consequently, alternative solution approaches can be also explored such as constructive heuristics combined with a variable neighborhood search (VNS) method.

Acknowledgements The authors thank the personnel of INEM for the fruitful meetings and discussions and for providing the data for the case study. Hendrik Vermuyten was sponsored by the Research Foundation Flanders (FWO), travel grant no. V418317N.

References

1. Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., De Boeck, L.: Personnel scheduling: a literature review. Eur. J. Oper. Res. **226**(3), 367–385 (2013)
2. Brucker, P., Qu, R., Burke, E.: Personnel scheduling: models and complexity. Eur. J. Oper. Res. **210**(3), 467–473 (2011)
3. Ernst, A., Jiang, H., Krishnamoorthy, M., Owens, B., Sier, D.: An annotated bibliography of personnel scheduling and rostering. Ann. Oper. Res. **127**(1), 21–144 (2004)
4. Burke, E., De Causmaecker, P., Berghe, G., Van Landeghem, H.: The state of the art of nurse rostering. J. Sched. **7**(6), 441–499 (2004)
5. Erhard, M., Schoenfelder, J., Fügener, A., Brunner, J.O.: State of the art in physician scheduling. Eur. J. Oper. Res. **265**(1), 1–18 (2018)
6. Aringhieri, R., Bruni, M., Khodaparasti, S., van Essen, J.: Emergency medical services and beyond: addressing new challenges through a wide literature review. Comput. Oper. Res. **78**, 349–368 (2017)
7. Van den Berg, P.L.J.: Logistics of emergency response vehicles: facility location, routing, and shift scheduling, Ph.D. thesis, Delft University of Technology (2016)

8. Isken, M.W.: An implicit tour scheduling model with applications in healthcare. Ann. Oper. Res. **128**(1), 91–109 (2004)
9. Bard, J.F., Purnomo, H.W.: Preference scheduling for nurses using column generation. Eur. J. Oper. Res. **164**(2), 510–534 (2005)
10. Puente, J., Gómez, A., Fernández, I., Priore, P.: Medical doctor rostering problem in a hospital emergency department by means of genetic algorithms. Comput. Ind. Eng. **56**(4), 1232–1242 (2009)
11. Santos, H.G., Toffolo, T.A., Gomes, R.A., Ribas, S.: Integer programming techniques for the nurse rostering problem. Ann. Oper. Res. **239**(1), 225–251 (2016)
12. Gomes, R.A., Toffolo, T.A., Santos, H.G.: Variable neighborhood search accelerated column generation for the nurse rostering problem. Electron. Notes Discret. Math. **58**, 31–38 (2017)
13. Joncour, C., Michel, S., Sadykov, R., Sverdlov, D., Vanderbeck, F.: Column generation based primal heuristics. Electron. Notes Discret. Math. **36**(C), 695–702 (2010)