Springer Proceedings in Business and Economics

Evangelos Grigoroudis
Michael Doumpos *Editors*

Operational Research in Business and Economics

4th International Symposium and 26th National Conference on Operational Research, Chania, Greece, June 2015

Springer

Springer Proceedings in Business and Economics

More information about this series at http://www.springer.com/series/11960

Evangelos Grigoroudis • Michael Doumpos
Editors

Operational Research in Business and Economics

4th International Symposium and 26th National Conference on Operational Research, Chania, Greece, June 2015

Springer

Editors
Evangelos Grigoroudis
Decision Support Systems Laboratory
Technical University of Crete
Chania, Greece

Michael Doumpos
Financial Engineering Laboratory
Technical University of Crete
Chania, Greece

ISSN 2198-7246 ISSN 2198-7254 (electronic)
Springer Proceedings in Business and Economics
ISBN 978-3-319-33001-3 ISBN 978-3-319-33003-7 (eBook)
DOI 10.1007/978-3-319-33003-7

Library of Congress Control Number: 2016946077

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

The Hellenic Operational Research Society (HELORS) was founded in 1963, with the aim to promote operational research (OR) in Greece and to provide a forum seeking to support the use of analytical methodologies in the Greek public and private sector.

One of the main activities of HELORS is the organization of its annual conference. Since the first national conference in 1977, HELORS has organized 26 National Conferences. Starting from 2012, the annual conference of HELORS is organized together with an international symposium, which provides a forum for exchanging ideas about the theory and practice of modern OR, not only among Greek researchers but also a broader international audience.

The 2015 event (4th International Symposium and 26th National Conference on Operational Research) was held in Chania, Greece, during June 4–6. The scientific program included 85 presentations by Greek and foreign researchers. The covered topics included all recent advances in operational research, including new methodological developments as well as applications and case studies in a wide range of fields, such as energy and the environment, management, logistics and supply chains, finance, transportation, public services, and health care.

This edited volume was prepared on the occasion of the above event. After a review process, 14 papers were selected for this book. The papers cover recent advances on a wide range of areas, adopting an applied perspective that covers the contributions of OR in the broad field of business and economics. The contents of the book can be (roughly) grouped into four main thematic areas.

The first group of papers covers topics related to the management of supply chains, organizational performance, and strategic management. The book starts with the paper by Panayiotou, Stavrou, and Gayialis who present an application of a business process modeling approach to design supply chain processes in the case of a SME manufacturing company. The second paper covers a similar topic. In particular, Nikolaou and Zervas examine the role of environmental information in the decisions of managers/owners to adopt environmental practices into their supply chain management. In the third paper, Karampatsa, Grigoroudis, and Matsatsinis provide a literature overview of retail category management, focusing on methods and approaches for assortment and shelf space planning and other related topics. The next two papers cover issues related to organizational performance. First, Kitsios, Champipi, and Grigoroudis present the use of a multicriteria decision aid approach to develop a model for assessing the likelihood of success of new services in the cultural and creative industries. In the next paper, Mitroulis and Kitsios examine how differentiation and competitive innovation strategy affect organizational performance, both in financial and nonfinancial terms. The last paper of the first thematic area of the book, by Krasadaki and Matsatsinis, presents a decision aiding process for strategic management in the agricultural sector and its pilot implementation in farms operating in the island of Crete, Greece.

The second thematic area includes two papers about financial decision making. The first paper, by Nikolaidis, Doumpos, and Zopounidis, examines the predictive power of analytical models in behavioral credit scoring under population drift conditions due to a deteriorating macroeconomic environment. In the next paper, Karakalidis and Sifaleras present the implementation of a library of portfolio optimization models in the AMPL mathematical programming modeling language.

The third group consists of three papers that focus on the use of optimization approaches (metaheuristics and analytical methods) for production systems and logistics. In the first paper in this group, Boulas, Dounias, and Papadopoulos present the implementation of a genetic programming approach for analyzing serial production lines and extracting useful measurements and line characteristics. The next paper by Rogdakis, Marinaki, Marinakis, and Migdalas presents a new algorithm for the vehicle routing problem together with its application to a real case study. The paper by Baazaoui, Hanafi, and Kamoun deals with a real-world application of cutting mousse blocks proposed by an industrial company, based on a mixed-integer linear programming formulation, which is used to derive an upper bound for this complex optimization problem.

The book closes with three papers about inventory systems and energy systems planning. First Konstantas, Ioannidis, Grigoroudis, and Kouikoglou develop simple models for understanding how the dynamics of quality affect customer satisfaction and profitability in make-to-stock manufacturing systems, focusing on a Markovian, single-stage system facing random demand. In the next paper, Krommyda, Skouri, Konstantaras, and Ganas formulate optimal replenishment policies for an inventory model for seasonal products, taking into account the warehouse capacity and credit period options. The book closes with the paper by Kanellos, Prousalidis, and Tsekouras, who present a dynamic programming approach for optimal demand side management and power generation scheduling in all electric ships, subject to operation, environmental, and travel constraints.

Closing this short preface, we would like to express our sincere gratitude to all participants of the 4th International Symposium and 26th National Conference on Operational Research, who supported the event, and in particular to the authors who contributed with their papers to this volume. We should further thank all those who devoted considerable time to review the submitted papers.

Chania, Greece

Evangelos Grigoroudis
Michael Doumpos

Contents

List of Contributors

Mariem Baazaoui MODILS Laboratory, Faculty of Economics and Management of Sfax, University of Sfax, Sfax, Tunisia

Konstantinos Boulas Management and Decision Engineering Laboratory (MDE-Lab), Department of Financial and Management Engineering, Business School, University of the Aegean, Chios, Greece

Eleni Champipi Cultural Organizations Management, Hellenic Open University, Patra, Greece

Michael Doumpos Financial Engineering Laboratory, Technical University of Crete, Chania, Greece

Georgios Dounias Management and Decision Engineering Laboratory (MDE-Lab), Department of Financial and Management Engineering, Business School, University of the Aegean, Chios, Greece

Sotiris P. Gayialis Sector of Industrial Management and Operational Research, School of Mechanical Engineering, National Technical University of Athens, Zografos, Athens, Greece

Ioannis Ganas Department of Accounting, Technological Educational Institute of Epirus, Preveza, Greece

Evangelos Grigoroudis School of Production Engineering and Management, Technical University of Crete, Chania, Greece

Saïd Hanafi LAMIH Laboratory, University of Valenciennes and Hainaut-Cambrésis, Famars, France

Stratos Ioannidis School of Production Engineering and Management, Technical University of Crete, Chania, Greece

Hichem Kamoun MODILS Laboratory, Faculty of Economics and Management of Sfax, University of Sfax, Sfax, Tunisia

Fotis D. Kanellos School of Production Engineering and Management, Technical University of Crete, Chania, Greece

Alexis Karakalidis Department of Applied Informatics, School of Information Sciences, University of Macedonia, Thessaloniki, Greece

Marina Karampatsa Decision Support Systems Laboratory, Technical University of Crete, Chania, Greece

Fotis Kitsios Department of Applied Informatics, School of Information Sciences, University of Macedonia, Thessaloniki, Greece

Dimitrios Konstantas Financial Engineering Laboratory, Technical University of Crete, Chania, Greece

Ioannis Konstantaras Department of Business Administration, School of Business Administration, University of Macedonia, Thessaloniki, Greece

Vassilis S. Kouikoglou School of Production Engineering and Management, Technical University of Crete, Chania, Greece

Evangelia Krassadaki Decision Support Systems Laboratory, Technical University of Crete, Chania, Greece

Iris-Pandora Krommyda Department of Business Administration of Food and Agricultural Enterprises, University of Patras, Agrinio, Greece

Magdalene Marinaki School of Production Engineering and Management, Technical University of Crete, Chania, Greece

Yannis Marinakis School of Production Engineering and Management, Technical University of Crete, Chania, Greece

Nikolaos F. Matsatsinis Decision Support Systems Laboratory, Technical University of Crete, Chania, Greece

Athanasios Migdalas Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Department of Business Administration, Industrial Logistics, Technology and Social Sciences, Luleå University of Technology, Luleå, Sweden

Dimitrios Mitroulis Department of Applied Informatics, School of Information Sciences, University of Macedonia, Thessaloniki, Greece

Dimitris Nikolaidis Financial Engineering Laboratory, Technical University of Crete, Chania, Greece

Ioannis E. Nikolaou Department of Environmental Engineering, Democritus University of Thrace, Xanthi, Greece

Nikolaos A. Panayiotou Sector of Industrial Management and Operational Research, School of Mechanical Engineering, National Technical University of Athens, Zografos, Athens, Greece

Chrissoleon Papadopoulos Department of Economics, Aristotle University of Thessaloniki, Thessaloniki, Greece

John Prousalidis School of Naval Architecture and Marine Engineering, National Technical University of Athens, Zografos, Athens, Greece

Ioannis Rogdakis School of Production Engineering and Management, Technical University of Crete, Chania, Greece

Angelo Sifaleras Department of Applied Informatics, School of Information Sciences, University of Macedonia, Thessaloniki, Greece

Konstantina Skouri Department of Mathematics, University of Ioannina, Ioannina, Greece

Vasileios P. Stavrou Sector of Industrial Management and Operational Research, School of Mechanical Engineering, National Technical University of Athens, Zografos, Athens, Greece

George J. Tsekouras Department of Electrical Engineering and Computer Science, Hellenic Naval Academy, Piraeus (Hatzikiriakio), Greece

Anastasios Zervas Department of Environmental Engineering, Democritus University of Thrace, Xanthi, Greece

Constantin Zopounidis Financial Engineering Laboratory, Technical University of Crete, Chania, Greece

Audencia Business School, Nantes, France

The Application of a Business Process Modeling Architecture in the Supply Chain of a Manufacturing Company: A Case Study

Nikolaos A. Panayiotou, Vasileios P. Stavrou, and Sotiris P. Gayialis

Abstract Business process modeling is aimed at the design and documentation of business processes. Business process models are used to analyze processes, to reduce their complexity, to evaluate their performance and finally to assist business process improvement. In this light, a number of modeling architectures, methods and tools have been developed in order to assist scientists and practitioners to model and manage business processes. In addition, supply chain management importance is increasingly being recognized as it integrates and synchronizes business processes across the extended supply chains.

This paper deals with the application of a specific business process modeling architecture in order to design supply chain processes in the case of a SME manufacturing company. The modeling architecture has been developed in the context of the "Odysseus" research project, which deals with the management of demand variability in modern supply chains. The architecture covers different supply chain views such as processes and activities, organization, information systems, risk management and decision making. These views are covered by the modeling architecture using nine selected and interconnected ARIS methods. The architecture is applied in a Greek SME company producing electrical equipment. The production process of the equipment consists of in-house as well as sub-contracted phases performed by Greek and European manufacturers. The coordination of the related supply chain processes is performed by the company under discussion. Due to the extended degree of collaboration, the need for accurate planning, coordination and controlling in the supply chain is highly increased, making business process modeling an ideal enabling approach.

Keywords Business processes modeling • Architecture • ARIS methods • Case study • Manufacturing company • Subcontractors

N.A. Panayiotou • V.P. Stavrou (✉) • S.P. Gayialis
Sector of Industrial Management and Operational Research, School of Mechanical Engineering, National Technical University of Athens, Zografos, Athens, Greece
e-mail: vstavru@yahoo.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_1

1 Introduction

Business process modeling allows the description of both core and supportive business processes. Core business processes are considered processes which provide significant added value to the business operation. Such added value could be achieved by cost reduction and/or performance enhancement (Jiménez et al. 2009). Business process modeling is used mostly for business process improvement (BPI) and business process reengineering (BPR). Business process modeling, information technologies and operational research are connected in the context of BPR, in order to achieve operations analysis and improvement (Emiris et al. 2001). In BPI and BPR projects, a series of methodological steps are followed, which are similar to those defined by operational research methodologies. Operational research (OR) is a discipline that deals with the application of advanced analytical methods to help the stakeholders make better decisions (Lancaster and Management 2013). Main topics of operational research approaches may include scheduling, production planning, forecasting, supply chain design, facilities allocation, and so on (INFORMS.org 2015). Operational research methodological steps can be defined as follows (Taha 2007):

- Recognition of the problem
- Formulation of the problem
- Model construction
- Solution finding
- Process definition
- Solution implementation
- Repeat and refine

Business process is defined as a specific order of events and activities across time and place, with a clear beginning, an end, and clearly defined inputs and outputs. Business processes form the structure through which the organization does what is necessary in order to produce value for its customers (Capgemini 2004). A process model is considered a representation of one or more processes and their associations that an enterprise performs. Process modeling represents a mechanism for describing the current or future state of a business process (Capgemini 2004). Business process modeling always has been at the core of both organizational redesign and informational systems development (Giaglis 2001). Similarly to mathematical modeling in OR applications, business process models are used as a prerequisite step for process analysis and problem solving.

The detailed analysis and optimization of these processes is leading to high levels of competitiveness in relation to other competitive companies. It should be noticed that the use of individual processes has led to a series of problems, providing users with only a part of the necessary information. For that reason, integrated processes with multi views architecture are preferred. Integrated architectures help enterprises to construct diagrams in a wide range of different views. In addition, integrated architectures, allows the decomposition of processes into different views, in order to reduce complexity (Roser and Bauer 2005). In this effort, a number of different architectures have been developed, such as ARIS (Davis 2001, 2008; Scheer 2000) and Zachman Framework (Zachman 1987). These architectures include integrated methods, offering views from different perspectives in order to reduce complexity, visualizing processes in an optimal way. Process modeling is also used in order to increase business knowledge (Bandara et al. 2005; Recker et al. 2009) with the help of graphical integrated tools (Curtis et al. 1992; Davenport 2005).

Business process modeling can assist the design, analysis and improvement of contemporary supply chains or other business entities which are characterized by complexity (Gayialis et al. 2015). Supply chain is the network of organizations which are involved, through upstream and downstream linkages, in the different processes that produce value during product production and delivery to the ultimate consumer (Christopher 1992; Mentzer et al. 2001). Supply chain management is a topic which has been adequately addressed in the literature (Cooper and Ellram 1993; La Londe and Masters 1994; Lambert et al. 1998). In the last decades, supply chains have been expanded significantly beyond the boundaries of the enterprise, involving a broad range of organizations. The importance of documented business processes, through business processes modeling, rises significantly in order to achieve interoperability and agility.

The extended supply chain is depicted in Fig. 1. The supply chain is divided in two main groups of process and partnering members: supply and demand, while the production is the process that coordinates all other phases.

The most suitable way for visualization of business processes and different partners of an extended supply chain is through an integrated architecture. In this paper a modeling architecture is applied in the supply chain processes of a Greek manufacturing company. A business process modeling architecture is a term occurred in the literature (Telecken et al. 2004; Morrison et al. 2011; Lampathaki et al. 2013; Xu et al. 2005). The term business process modeling architecture refers to the structure of the modeling components (methods and tools) and their correlation. Business process modeling architectures is an integrated enterprise architecture as it integrates various methods in order to cover different enterprise views. Especially Panayiotou and Tatsiopoulos (Panayiotou and Tatsiopoulos 2013)

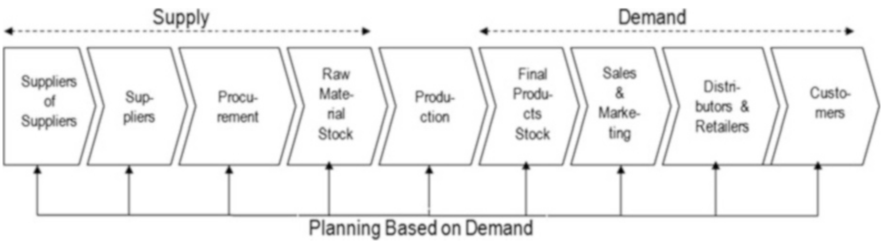


Fig. 1 Extended supply chain

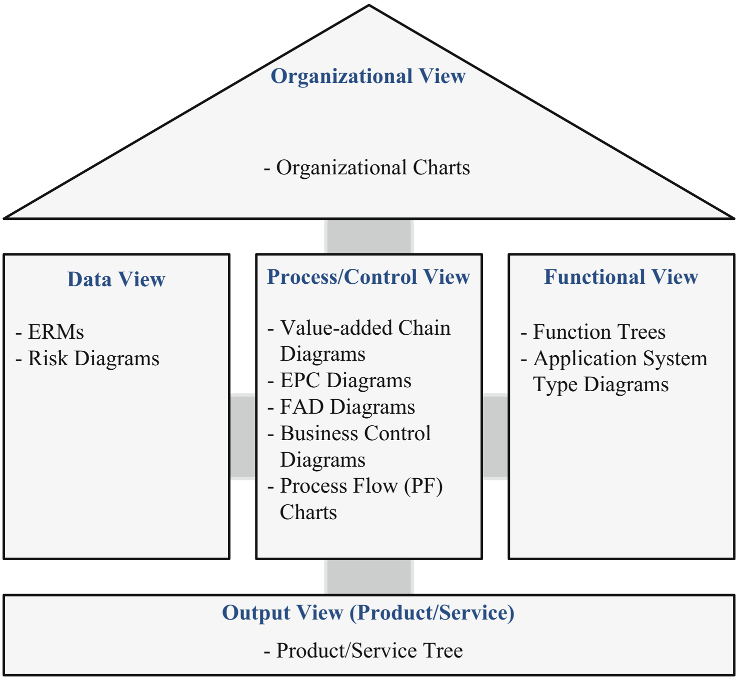


Fig. 2 Methods of the integrated architecture within different views of "ARIS House"

describe the construction of an ARIS-based business process modeling architecture, while Gayalis et al. (2013) refers to business process modeling architecture in order to construct a reference model and Grundy et al. (1998) make extensive reference on process modeling architectures for software process modeling. The architecture which will be used in the paper has been developed in "Odysseus" research project, in the context of Thales research program. The aim of the project was to develop a reference model which manages in a holistic way the variability in contemporary global supply chain networks (Ponis et al. 2015). The developed reference model was based on a customized modeling architecture which integrates 11 different ARIS modeling methods (Fig. 2), using ARIS platform toolset.

In this paper, the application of the aforementioned modeling architecture gives the opportunity to present modeling efforts of supply chain processes in real-life organizations. The "Odysseus" modeling architecture was validated in the supply chain processes of an electronic equipment manufacturer (case study) in order to demonstrate its effectiveness in supply chain design and improvement.

2 The Adopted Integrated Architecture

As discussed earlier, the adopted architecture was developed in the context of a research project. The project team specified a series of criteria, in order to define the architecture requirements and needs. Based on these criteria, ARIS integrated architecture was selected as the basis for the adaptation of a set of modeling methods. These methods cover the different views of contemporary demand-driven supply chains and were used for the construction of a supply chain reference model. These criteria are summarized as shown in Table 1.

The selection and adaptation of the integrated architecture was based on the aforementioned set of criteria, intending to assist the design of effective supply chain processes.

The ability of the architecture for modeling different views of the supply chain includes:

- Organizational view: depicts the organizational structure and the roles of the partners of the supply chain.
- Information view: represents IT systems or enterprise software applications and their interfaces.
- Decision view: emphasizes on the decisional activities, their alternatives flows and the parameters for decision making.
- Risk and control view: depicts the various risks within the supply chain processes and the application of controls for risk management.
- Algorithms view: represents the algorithms that can assist decision making in order to achieve improved business performance (for example in demand forecasting).
- Process view: models business processes in an integrated way, combining objects from all the other modeling views.

Table 1 Architecture's selection criteria and requirements

Criterion	Requirements
Representation and integration of selected views	Organizational view
	Information view
	Decision view
	Risk and control view
	Algorithms view
	Process view
Models reusability	Ease of use
	Software tool support
Coverage of different types of processes	Internal processes
	External processes
	Collaborative processes
Reference models management	Process variability management
	Abstraction level management
	Variants management

Models reusability was provided by the ease of use criterion and the existence of a supportive software tool criterion. Coverage of different types of processes (internal, external and collaboration processes) was also a discrete criterion. The criteria for reference models management had to do with the management of the variants of business process models as well as the ability to model different versions of the processes and the different abstraction of a reference model (generic, general or particular models).

The above criteria were covered by ARIS architecture. A set of eleven modeling methods were selected and adapted in order to cover all the different supply chain views. These methods were included in the different perspectives (views) of the "ARIS House" (Scheer 2000).

Figure 2 summarizes the methods used in the integrated architecture, being categorized in five different perspectives of an organization, according to "ARIS House". A short presentation of these methods is following.

Organizational View Using an *Organizational Chart* the hierarchical structure of business divisions and departments is defined. It can also be used for the identification of business partners that cooperate with the company in specific processes.

Process/Control View The *Value-added Chain Diagrams (VCDs)* defines the core business processes and their connections in the value chain. These processes are further analyzed in detail with the use of *Event-driven Process Chain Diagrams (EPCs)*, describing the steps of every process, including a set of activities, triggers and events. EPC is the most important method in the architecture as it brings together the static resources of the company (systems, organization, data, etc.) and structures them in order to deliver a sequence of tasks or activities, adding business value. Even more detailed information concerning process activities can be included in *Function Allocation Diagrams (FADs)* which define the relationships between an activity and its resources, in addition with the data it transforms. A FAD also connects an activity with various other modeling objects from different modeling views (like risks, software applications or *documents*). *Business Control Diagrams* provide controls and solutions in order to reduce and mitigate the risks which are defined in the *Risks Diagrams*. In Business Control Diagrams every risk is related with controls and solutions. Finally, the *Program Flow Chart* describes the information flow within any computer program, and the order of the various steps which are executed, which is useful for the development of business process automation.

Data View IT applications list and hierarchy are modeled in the *Application System Type Diagrams* of the functional view. The most important data used and exchanged between the IT systems are modeled through the *Entity Relationship Management Diagrams (ERMs)*. In addition, *Risk Diagrams* are used in order to model process risks, including group of risks and risks categories. Risks are defined for various activities of the business processes, especially for decisional activities. These risks can be further connected with specific activities in FADs.

Functional View Business processes are decomposed to sub-processes and activities. This hierarchical decomposition is presented with the use of *Function Trees*. Business processes are also supported by IT applications which are modeled in the *Application System Type Diagrams* defining a library of IT systems used by the company. The IT applications are structured, providing a hierarchical classification.

Output View (Product/Service) Every company produces products or services. A hierarchy of the products and sub-products and the evolution of the product and services, from resources (inputs) to outputs along the way, are presented thorough *Product/Service Tree Diagram*. Every product or service could also be assigned to activities through EPCs or FADs.

The aforementioned methods of the modeling architecture are interconnected each other. For example, through the Value Added Chain Diagram a context diagram of the business processes and their connection could be modeled. For every business process an EPC is developed. Next, each activity of EPC (with its inputs and outputs) is analyzed separately in FAD diagrams. In addition, risks (which are presented in risk diagram) are also analyzed in FADs, and the way that the enterprise can deal with them is shown in Business Control Diagrams. The interoperability between the methods is depicted in Fig. 3.

The detailed recording of integrated processes could also lead to a strategic alignment of the involved companies supply chain, especially in this case which consists of many companies (Ciborra 1997). Strategic alignment was originally defined as concerning the inherently dynamic fit between external and internal domains, such as the product/market, strategy, administrative structures, business

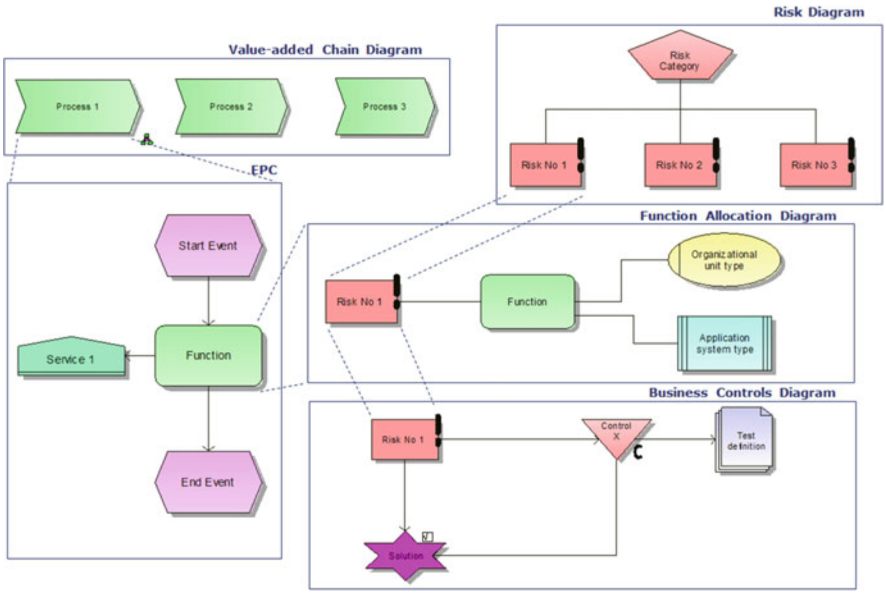


Fig. 3 Integrated ARIS methods of the architecture

processes and IT. Through the whole process and information systems the company can converse with their subcontractors ensuring timely delivery of raw materials, products etc. Likewise, through strategic alignment the organization's structure and its resources can be linked with its strategy and business environment enabling decision makers to collect meaningful data based on their current processes.

3 Case Study: Application of the Integrated Architecture in a Manufacturing Company

3.1 Case Study Company Description

In this paper, the previously presented architecture is implemented in a Greek SME company producing electrical equipment. It is a manufacturing company (which employs around 40 employees) producing switches, sockets, fans, bells, buttons and other electrical equipment, having a long presence (about 40 years) in the Greek market. The company aims to ensure the correct production process of its products and the continuous improvement in order to meet the increasing needs of its customers. The case study focuses on the production process of switches and sockets (which is similar in both cases).

The case study was selected because SMEs exhibit special interest, forming the backbone of the Greek economy and employing the majority of its employees. This company is a characteristic case study in order to implement the abovementioned architecture and its interconnected ARIS methods in a specific company. It is also a good case in order to investigate the behavior of this integrated modeling architecture in order to design or improve supply chain processes in the case of a SME production company. The company under discussion is also used as a pilot case in order to test the architecture's applicability in real world cases.

3.2 Case Study Research Approach

Following Yin's case study research methodology (Yin 2003) and according to similar work in business process modeling and case study research (Panayiotou et al. 2015), a number of methodological steps were followed in order to test the applicability of the architecture in the abovementioned SME company and to answer the question of "how the application of the business process modeling architecture in the supply chain of a manufacturing company can assist the improvement of its operation". These steps (described in detail below) include:

- Collection of case data in a structured way, using multiple sources of evidence (information concentration),

- Application of the business process modeling architecture, in the case company (AS-IS modeling, identification of improvement areas, and prioritization of improvement initiatives).
- Evaluation of the results (TO-BE modeling).

As a concept, the case study is considered a suitable practice for a different range of applications. In order to come up with new conclusions regarding the way this architecture suites in SMEs producing electrical equipment and have interconnected supply chain, and highlighting the importance of SMEs in the Greek market, this case study will be used as a pilot for future case studies, investigating the way supply chain processes in the manufacturing SMEs look like (the unit of analysis will be the company). The whole research lasted 2 months in close cooperation with the company by a research team of three people (one researcher, one senior researcher and one research supervisor).

Four conditions related to design quality according to Yin (2003) are used for the evaluation of the research presented in this paper. Construct validity is achieved by ensuring multiple source of evidence and establish a chain of evidence. Internal validity outlines the quality of the research design by establishing casual relationships between objects (pattern matching, explanation building, logic models etc.). External validity is achieved by using theory in single case studies and exploring the way this study's findings can be generalized to other SMEs of various industries. Finally, reliability is achieved by building a case study protocol and develop a case study database (so data collection procedures, business process modeling approach etc.) can be repeated with similar results.

The construct validity of the case study data was strengthened in several ways. For example, several people were interviewed. The research team consisted of three researchers (Two researchers were involved in every interview). The approach for data collection described increased also the reliability of the case study. The data collected during interviews were discussed in a daily session, where processes were clarified. Business models were gradually created and they were reviewed, in order to formulate the final version. A possible replication of the study was also ensured by creating a database with notes, documents, and narratives collected during the research (strengthening the validity of the study). Under these circumstances, the aforementioned process modeling architecture and methods can be reproduced. The case study leads to an accurate observation of the reality. The main evidences are both the company documents and the interviews held. The research team tried to crosscheck information collected using different data sources.

During the analysis, a set of explanations and results were emerged, both regarding the companies involved and the models developed. A number of business process models were also developed. Regarding external validity, the modeling architecture was developed in the context of the "Odysseus" research project. The framework can be generalized in similar SMEs companies, while further testing in other SMEs could better validate the research results, improving the external validity. The research team ensured also the transparency of the research, clarifying a careful research procedure (case study protocol) in order to ensure the series of results extracted. The research protocol validated the research way.

3.3 General Description of the Study

The supply chain processes of the company face many challenges in a highly competitive environment. A critical success factor is the cooperation with its suppliers and subcontractors. The production process depends on the timely procurement of raw materials from domestic and international suppliers, as well as the implementation of certain process stages from subcontractors. In the AS IS phase of the company, the processes that require cooperation with suppliers and subcontractors are managed mostly empirically and are based on the experience of few key employees.

Although the company posses an ERP system in order to support its processes, it has not fully utilized its functionality and has not used it for on-line cooperation with business partners. The retirement of the Production Planning Manager, an old and very experienced employee, responsible for production planning and sub-contractors management created serious problems in the production process, leading, to bad coordination with subcontractors, uneven production schedules and increased production lead times. The owner of the company realized that a systematic approach in business process design and control was needed in order to guarantee the continuity of business process execution and stable operations.

This main aim of the paper consists on the application of a business process modeling architecture in order to design supply chain processes in the case of a Greek SME production company producing electrical equipment.

The adoption of the architecture took place in order to design supply chain processes in the manufacturing SME, aiming to redesign them based on the process modeling architecture, including the following methodological steps:

- Information concentration
- AS-IS modeling
- Identification of improvement areas
- Prioritization of improvement initiatives
- TO-BE modeling

These methodological steps try to cover many phases of BPM lifecycle. BPM lifecycle establishes a sustainable process management capability, empowering organizations to embrace and manage process changes successfully (Com Cloud Study). Because it incorporates human resources, technology, culture, roles and responsibilities, data content, applications and infrastructure the approach enables fully informed decision-making right across an organization. Information concentration is a part of design part of BPM lifecycle while modeling belongs to modeling phase. Identification and prioritization of improvement areas as well as TO BE modeling are part of reengineering phase, as they include retrieving process performance from modeling phase, identifying potential problems and applying enhancements in the forthcoming design of the process.

The study started with the information collection, in order to identify the existing corporate strategy which is important for the determination of the existing processes. Next, project team conducted the AS-IS process modeling. In the case study under discussion, a series of modeling methods were applied. The re-designed processes enforce the accurate and efficient production of the company's products. Methodological steps are described in the next paragraphs.

3.4 Application of the Business Process Modeling Architecture

The adoption of the previously presented architecture was decided for the analysis of the existing business processes taking into account all the aspects covered. The architecture was utilized by the case company in order to document knowledge in a systematic manner and to help future reengineering efforts.

The coordinator of the analysis of business processes was the new Production Planning Manager and the development of the necessary diagrams was based on interviews with key employees of the company as well as on existing documentation, like ISO processes, organizational charts, informal written processes and IT manuals. At first, the *Organizational Chart* and the *Application System Type Diagram* were developed, documenting the existing resources. The seven supply chain processes that were in the scope of the analysis were depicted in a *Value-Added Chain Diagram* (presented in Fig. 4). The diagram was used in order to represent the processes in the highest level. These specific processes are the primary processes of an enterprise according to Porter (1985).

The most challenging part of the analysis was the recognition of activities and their representation in *EPC Diagrams*. An exemplary *EPC Diagram* is depicted in Fig. 5, concerning the production planning process. A more analytical representation of the resources used by each activity was carried out in *Function Allocation Diagrams (FADs)*, where needed in order to include all different business objects associated with every activity (risks, applications, documents, organizational roles). The completion of EPCs and FADs permitted the identification of existing risks at an activity-level, which had never been done in a formal manner in the past.

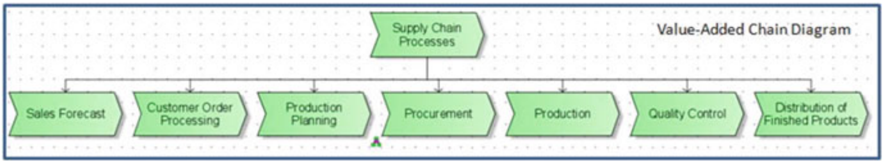


Fig. 4 Value-added chain diagram

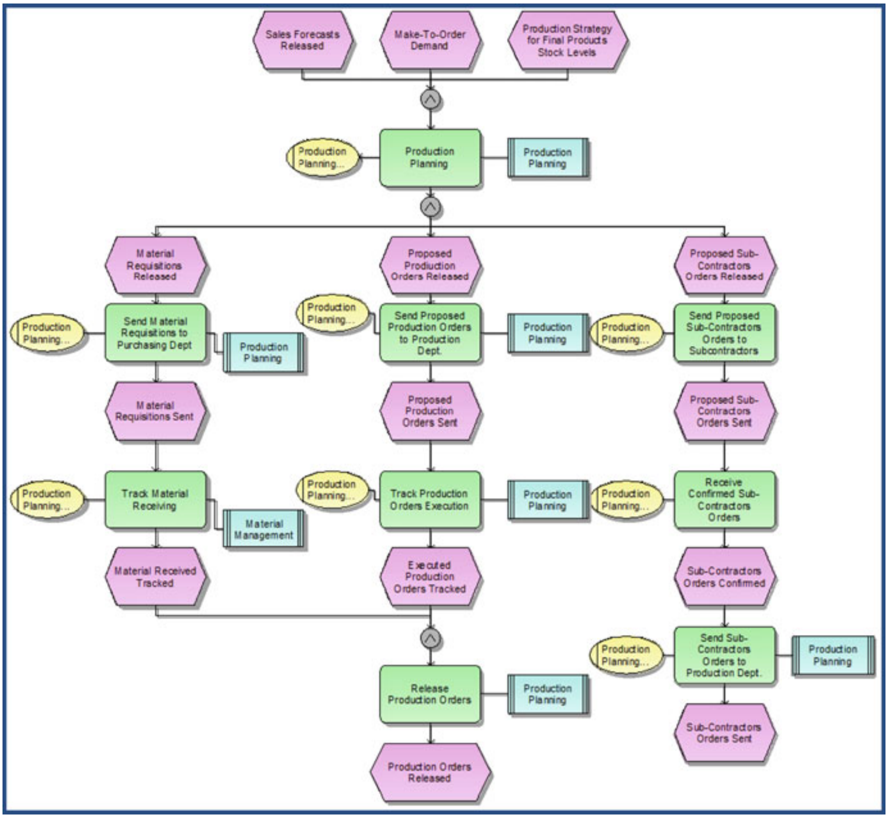


Fig. 5 EPC diagram

The identification of specific risks enabled the pursuit of controls that could minimize the risks. Figure 6 represents a *Risk Diagram* concerning the production planning process, the identified risks related in an activity of the production planning process, represented in a *Function Allocation Diagram (FAD)* and the corresponding risk controls in a *Business Controls Diagram*. In the example shown, one of the recognized risks in the production planning activity is the inability to satisfy the demand. This is an operational risk and it can be controlled through the use of alternative forecasting methods and the calculation of Mean Absolute Deviation (MAD) in order to monitor the deviation of the forecast compared with the actual demand for each final product. A solution could be the development of forecasting algorithms for the final products and the calculation of MAD. The algorithms needed to be developed can be represented in a *Program Flow Chart (PF)*.

All ARIS modeling methods of the architecture (as presented in Fig. 2) were applied, except Product/Service Tree, as the case study emphasized on process flow rather than products.

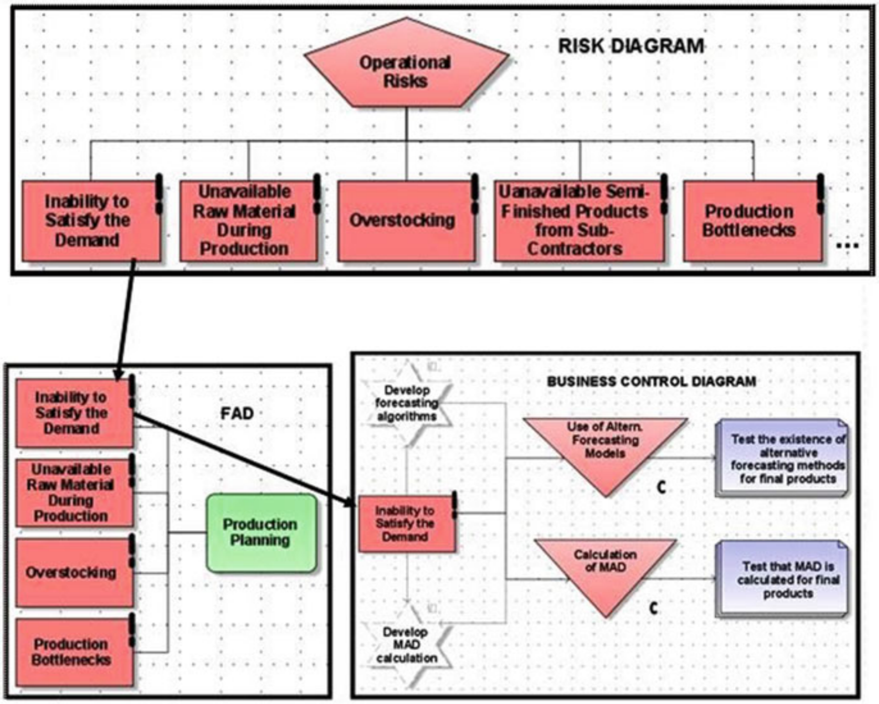


Fig. 6 Integration of risk diagram, fad and business control diagram

3.5 Results of Architecture's Application

Following the modeling architecture, a set of business process models were developed. The company was able for the first time to document and evaluate its supply chain processes. A quick win from the implementation of the models was the fact that different process aspects were documented in a formal way, facilitating business continuity and management. Different stakeholders (information systems, "horizontal" processes, internal audit) had an integrated point of view for business processes. However, the most important benefit was the introduction of the culture of continuous improvement in the company.

The business process modeling highlighted the existence of important weaknesses in the company. This was the trigger for process improvement actions. The identification of improvement areas included: problematic forecasting generation, inaccurate material requisitions, problematic procurement orders status monitoring, problematic coordination with sub-contractors, inefficient planning process for the management of emergent make-to-order customer orders, problematic production execution.

List business process improvements also developed and prioritized according to the implementation easiness and the expected benefit. Some of these improvement initiatives included: formal agreement and renegotiation on the suppliers' lead times, master data maintenance, introduction of new processes for improved inventory control, introduction of new forecasting method.

Improved processes of the supply chain were designed and the TO-BE business models were documented, using the same ARIS methods as the AS-IS models. The adoption of the presented architecture in the case study made the recording and visualization of the supply chain processes an easy and simple procedure which added value to the company.

4 Conclusions

This paper presents the application of a business process modeling architecture in the supply chain of a manufacturing Greek SME company. The implemented architecture is based on eleven ARIS methods divided in the five perspectives (views) of the ARIS House. All methods of the architecture are interconnected and integrated.

The adoption of such architectures and tools in specific companies or sectors can help them deal with the problem they face. In the business management area, business process documentation was established with integrated process views regarding different stakeholders. The application of business process modeling architecture in the case company guided to the identification of specific improvement areas such as the problematic forecasting generation or the problematic coordination with sub-contractors and eventually to specific actions for business process improvement.

This business process modeling architecture can assist a company's operation through the whole supply chain. A primary role in implementing this architecture is the reduction of resistance in change through transparency and common understanding. This will lead to good supply chain collaboration. During this effort, a number of challenges needed to be addressed. These challenges are summarized below:

- Existing knowledge available only in few employees' minds.
- Difficulty in identifying important complicated business issues.
- Undocumented real business processes.
- Neglected views in company's business thinking.
- Existing company culture.

For all these reasons, building architectures which assist in business processes modeling and monitoring should be encouraged. The presented architecture should be tested in additional case studies (companies operating in different sectors) in order to extend the validation of its effectiveness. The enhancement of the architecture towards business process execution would increase its offered benefits.

Acknowledgments The research efforts described in this paper are part of the research project "A Holistic Approach for Managing Variability in Contemporary Global Supply Chain Networks" in research action: "Thales—Support of the interdisciplinary and/or inter-institutional research and innovation" which is implemented under the Operational Programme: Education and Lifelong Learning, NSRF 2007–2013 and is co-funded by European Union (European Social Fund) and Greek Government.

References

Bandara W, Gable G, Rosemann M (2005) Factors and measures of business process modelling: model building through a multiple case study. Eur J Inf Syst 14(4):347–360

Capgemini (2004) Business process modeling defined. URL: http://www.cdc.gov/std/program/bpmm/overview.pdf. Last Retrieved 25 Sept 2015

Christopher ML (1992) Logistics and supply chain management. Pitman Publishing, London

Ciborra CU (1997) De profundis? Deconstructing the concept of strategic alignment. Scand J Inf Syst 9(1):67–82

Com Cloud Study (2015) Discuss six sigma, BPM life cycle and total quality management. https://sites.google.com/site/comcloudhelper/home/ipcc/business-process-management-it/discuss-sixsigma-bpm-life-cycle-and-total-quality-management-tqm. Retrieved 20 Dec 2015

Cooper M, Ellram L (1993) Characteristics of supply chain management and the implication for purchasing and logistics strategy. Int J Logist Manag 4(2):13–24

Curtis B, Kellner K, Over J (1992) Process modeling. Commun ACM 35(9):75–90

Davenport T (2005) The coming commoditization of processes. Harv Bus Rev 83(6):100–108

Davis R (2001) The event-driven process chain, business process modeling with ARIS: a practical guide. Springer Science & Business Media, London, pp 111–139

Davis R (2008) ARIS, design platform, advanced process modeling and administration. Springer Science & Business Media, London

Emiris DM, Koulouriotis DE, Matsatsinis NF (2001) Modeling of business processes and functions of an industrial unit for ERP system application. Oper Res: An Int J 1(2):181–195

Gayialis S, Ponis S, Tatsiopoulos I, Panayiotou N, Stamatiou D-R (2013) A knowledge-based reference model to support demand management in contemporary supply chains. 14th European conference on knowledge management, Kaunas University of Technology, pp 236–244

Gayialis SP, Ponis ST, Panayiotou NA, Tatsiopoulos IP (2015) Managing Demand in supply chain: the business process modeling approach. 4th international symposium and 26th national conference on operational research, Chania, Greece, pp 73–79

Giaglis G (2001) A taxonomy of business process modeling and information systems modeling techniques. Int J Flex Manuf Syst 13(2):209–228

Grundy JC, Apperley MD, Hosking JG, Mugridge WB (1998) A decentralized architecture for software process modeling and enactment. IEEE Internet Comput 5:53–62

INFORMS.org (2015) What is operations research? URL: https://www.informs.org/About-INFORMS/What-is-Operations-Research. Last Retrieved 25 Sept 2015

Jiménez M, Piattini M, Vizcaíno A (2009) Challenges and improvements in distributed software development: a systematic review. Adv Softw Eng (3). doi:10.1155/2009/710971

La Londe BJ, Masters J (1994) Emerging logistics strategies: blueprints for the next century. Int J Phys Distrib Logist Manag 24(7):35–47

Lambert K, Douglas M, James R, Ellram L (1998) Fundamentals of logistics management. Chapter 14. Irwin/McGraw-Hill, Boston, MA

Lampathaki F, Koussouris S, Psarras J (2013) Business process modeling, business process re-engineering. Decision Support Systems Laboratory NTUA. URL: http://academics.epu.ntua.gr/LinkClick.aspx?fileticket=nNPcBN7bjX0%3D&tabid=930&mid=2218. Last Retrieved 25 Nov 2015

Lancaster University, Management School (2013) What is operational research? URL: http://www.lancaster.ac.uk/lums/study/masters/programmes/msc-operational-research-management-science/what-is-operational-research/. Last Retrieved 25 Sept 2015

Mentzer J, DeWitt W, Keebler J, Min S, Nix N, Smith C, Zacharia Z (2001) Defining supply chain management. J Bus Logist 22(2):1–25

Morrison ED, Ghose AK, Dam HK, Hinge KG, Hoesch-Klohe K (2011) Service-oriented computing of the series lecture notes in computer science strategic alignment of business processes. ICSOC 2011 Workshops 221: 9–21

Panayiotou NA, Tatsiopoulos IP (2013) Supporting the design of a management accounting system of a company operating in the gas industry with business process modeling. In: Advances in production management systems, competitive manufacturing for innovative products and services. Springer, Berlin, pp 686–692

Panayiotou NA, Gayialis SP, Evangelopoulos NP, Katimertzoglou PK (2015) A business process modeling-enabled requirements engineering framework for ERP implementation. Bus Process Manag J 21(3):628–664

Ponis S, Gayialis S, Tatsiopoulos I, Panayiotou N, Stamatiou D, Ntalla A (2015) An application of AHP in the development process of a supply chain reference model focusing on demand variability. Oper Res Int J 15(3):337–357

Porter ME (1985) The competitive advantage: creating and sustaining superior performance. Free Press, New York. Republished 1998

Recker J, Rosemann M, Indulska M, Green P (2009) Business process modeling—a comparative analysis. J Assoc Inf Syst 10(4):335–363

Roser S, Bauer B (2005) A categorization of collaborative business process modeling techniques. Seventh IEEE International Conference on E-Commerce Technology Workshops, pp 43–51

Scheer A (2000) ARIS—business process modeling, 3rd edn. Springer, Berlin

Taha H (2007) Operations research: an introduction, 8th edn. Pearson Prentice Hall, Upper Saddle River, NJ

Telecken TL, Valdeni de Lima J, Montgomery Barroso F (2004) Process modeling architectures with namespace and XML technology, CLEI- Conferencia latinoamericana de informatica, 2004, Arequipa, Peru, pp 54–65

Xu H, Li Z, Bo-sheng Z (2005) Enterprise process modeling methodology and environment supporting project management. J Shanghai Univ 9(6):493–500

Yin R (2003) Case study research: design and methods. Sage Publications, London

Zachman J (1987) A framework for information systems architecture. IBM Syst J 26(3):276–292

How Environmental Knowledge of Managers Plays a Critical Role in Implementing Green Supply Chain Management

Ioannis E. Nikolaou and Anastasios Zervas

Abstract This paper aims to examine how environmental information affects the decisions of managers/owners to incorporate environmental concerns into their supply chain management. A research framework is constructed through a literature review. This is based on certain research hypotheses which are particularly related to the knowledge of managers about environmental impacts of Small and Medium Sized Enterprises (SMEs), and of the barriers and opportunities faced by SMEs when implementing certain practices to green their supply chain management. The findings show that knowledge and environmental information plays a critical role in managers' decisions to adopt environmental practices across supply chain management, while the economic crisis seems to negatively affect their intention to implement any environmental practice.

Keywords Green supply chain • Barriers • Weaknesses • Environmental practices • Corporate environmental management • EMSs

1 Introduction

Many assumptions have been made to explain why managers are willing or unwilling to adopt environmental management practices in order to make the traditional ways of production and operation environmentally friendly. The adoption of such practices seems to offer important benefits to SMEs such as cost reduction, improved reputation, enhanced market share, easy access to financial markets, attractiveness to the investor community, and increased satisfaction of staff and consumers (Zhu and Sarkis 2004). The theoretical foundation of firms' initiatives to incorporate green concerns into their supply chain management is classified in two general theories: (a) a reactive-driven approach is a response of firms to legislation pressure (Walker et al. 2008), and (b) a proactive-driven

I.E. Nikolaou (✉) • A. Zervas
Department of Environmental Engineering, Democritus University of Thrace, Xanthi, Greece
e-mail: inikol@env.duth.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_2

approach which can be explained as a strategy of firms either to avoid potential risks or exploit new opportunities (Carter and Dresner 2001).

Other authors explain these strategies as the influence of environmental policy on managers' decisions to adopt green supply chain management (G-SCM) (Lee 2008). Following classical classification of environmental policy tools, Tsireme et al. (2012) notice that not only "*command and control*" and "*market-based*" instruments explain the intention of managers of SMEs to adopt G-SCM but also value-driven parameters play a critical role in these decisions. Lee (2008) provides indirect parameters which affect managers decisions such as suppliers with an environmentally friendly profile, government involvement and green supply readiness in order to adopt green strategies in the overall supply chain.

However, SMEs should overcome various weaknesses in order to adopt G-SCM (Carter and Dresner 2001; Lee 2008). Some essential weaknesses are the lack of financial resources and environmental skilled employees as well as environmental knowledge of firms make managers reluctant to adopt G-SCM (Mathiyazhagan et al. 2013).

This paper sheds light on how information regarding environmental problems, barriers and weaknesses influence managers of SMEs to invest in G-SCM. A conceptual model has been developed which is based on certain research hypotheses. These have been tested in a number of Greek SMEs which are located in the industrial area of Thessaloniki, Greece.

The rest of the paper consists of four sections. The first section includes the theoretical background and hypotheses of the research. The second section presents the methodology of this paper and the third section describes the results and discussion. Finally, the last section includes the conclusion.

2 Theoretical Underpinning: Hypotheses Development

G-SCM has increased in popularity in corporate environmental management. The current literature covers many topics regarding forward and backward supply chain management. Srivastava (2007) classifies green supply chain management in three categories; importance of G-SCM, green design and green operations. The first category includes academic work which focuses on the relationships of corporate performance and green supply chain strategies. The second category focuses on examining green design tools (e.g., Life Cycle Analysis, Environmentally Conscious Design) which are useful and suitable for greening products and operations. The last category includes green operations such as green manufacturing, reverse logistic and waste management. Sarkis et al. (2011) present nine theories which are utilized in the general literature of G-SCM and explain the reasons for which firms adopt G-SCM strategies on a voluntary basis. These theories emphasize differentiations in a wide range of issues such as environmental, social and economic.

The environmentally driven theories of G-SCM are based on ecological modernization where firms seek to protect natural resources by implementing new innovative achievements. The social theories of G-SCM focus on the stakeholder theory where firms address environmental protection goals by taking into consideration the needs of different stakeholder groups (e.g., customers, suppliers, financial sectors). Institutional theories and social networks are also categorized in social theories where incentives for firms to adopt G-SCM strategies are associated with institutions' policies and existence and strength of social networks. Economic theories constitute the majority of theories and focus on explaining the incentives for firms to adopt G-SCM practices through decreasing transaction costs (transaction cost theory), increasing the resources of firms (resource – based view of firms) and elimination of asymmetric information (information theories). Seuring and Muller (2008a) highlighted that the economic dimension appears in all G-SCM literature. The environmental dimension appears in 73 % of the G-SCM literature examined and the remainder concentrates on sustainable development (both environmental and social dimensions).

Darnall et al. (2008) classify the benefits for firms of adopting G-SCM in two categories: internal and external. The internal incentives are associated with developing new control and inventory systems for materials management, better and improved communication between different departments, and effective employee training regarding environmental and innovation issues. The external incentives are related to the aim of firms to align their daily operations with a range of social norms and values (legitimating of their operations). Seuring and Müller (2008b) identified that managers and owners consider two reasons as significant in order to adopt G-SCM such as the pressure of regulation and pressure from NGO. Linton et al. (2007) recognize three basic reasons which influence the decision making of firms in the adoption of G-SCM: to comply with environmental regulations, to achieve public interest requirements and to exploit new competitive opportunities. Tsireme et al. (2012) consider the reasons for corporate G-SCM adoption as a result of command and control instruments (e.g., environmental legislation), economic instruments (e.g., market-based instruments) and voluntary instruments (e.g., ISO 14011, EMAS).

However, a new approach in the field of corporate environmental management indicates that human resources play a critical role in the adoption of environmental management practices (Daily and Huang 2001; Muduli et al. 2013). Daily and Huang (2001) maintain that management commitment, environmental training programs for employees and rewards systems for staff may facilitate firms to adopt environmental management practices. Muduli et al. (2013) examine similar factors for the correct implementation of G-SCM practices by employees. They identify that human resources are a very important factor for implementing a business environmental strategy. In particular, senior management support is considered very important in order to implement environmental management strategies within the supply chain. The implementation of voluntary environmental management practices creates different types of organizational culture inside an organization such as employees who put emphasis on diverse facets of sustainability such as staff development, environmental protection and economic values (Linnenluecke and Griffith 2010).

Green et al. (1996) highlight that managers which are informed on environmental issues are more likely to adopt G-SCM practices. This implies that managers and staff who are better informed concerning their responsibilities for the natural environment might shift to G-SCM more easily (Darnall et al. 2008). Environmental awareness of managers/owners of SMEs regarding environmental impacts for which their firms are responsible might be a potential driver that leads them to adopt G-SCM (Gadenne et al. 2009).

A rational hypothesis is:

H1: The higher the environmental awareness of managers/owners of SMEs, the easier their decision is to invest in G-SCM.

SMEs have a wide range of alternative practices to G-SCM such as: green purchasing, green suppliers and green supply chain practices. Zhu and Sarkis (2004) classify green supply chain management practices in four categories: internal environmental management, external G-SCM practices, investment recovery and eco-design. The first category includes top management commitment, environmental management systems (e.g., ISO 14001) and environmental auditing systems. The second category encompasses practices regarding suppliers' environmental strategies and purchasing issues as well as cooperation with customers for making environmentally friendly products and packaging design. The third category refers to design products by introducing the protection of the natural environment and the fourth category includes investment recovery practices such as products and materials recovery at the end of product life cycle. Diabat and Govindan (2011) expand the list with strategies with reverse logistics and reducing energy consumption throughout the supply chain. Nikolaou et al. (2013) provide guidelines to incorporate sustainability and corporate social responsibility principles into reverse logistics. Their methodology is based on the Global Reporting Initiative mainly for transparency reasons.

Similarly, many authors focus on green purchasing techniques and the selection of green suppliers (Min and Galle 1997; Zsidisin and Siferd 2001). Min and Galle (1997) classify green purchasing methods in two categories: source reduction (e.g., recycling, reuse, remanufacturing) and waste minimization (e.g., biodegrading, nontoxic incineration). They also identified that the most important reason for adopting a technique for greening their purchases is legal requirements and avoiding potential liabilities for the disposal of hazardous materials. Additionally, many academics have proposed mathematical techniques for assisting firms in selecting suppliers with environmentally friendly performance (Noci 1997; Bai and Sarkis 2010). These techniques are based on multicriteria (e.g., AHP, ANP) and mathematical methodologies (e.g., uzzy) in order to provide standard steps in procedures of selecting green suppliers (Handfield et al. 2002; Awasthi et al. 2010).

The adoption of such practices in turn will entail benefits for SMEs. Similarly, these practices will be adopted more easily by managers/owners who are informed of/about the benefits of environmental management practices. Pujari et al. (2003) identify a significant statistical relationship between corporate environmental performance and top management support and supplier involvement to green supply chain management. Chiou et al. (2011) recognize that greening the suppliers is positively associated with green innovation, corporate environmental performance and corporate competitive advantage. These results are based on a study of 124 companies in Taiwan. Carter et al. (2000) support that managers' choices in purchasing practices might play a critical role in corporate environmental performance. For example, recycled packaging contributes to eliminating the cost of packaging and transportation costs due to reducing packaging weight.

A rational hypothesis is:

H2: The managers who know the benefits of green supply chain management are more likely to implement these practices.

However, some factors hinder SMEs to adopt environmental strategies. For instance, the limited financial resources and the lack of time to take on additional initiatives (e.g., environmental) are two indicative barriers for SMEs (Cote et al. 2008). Similarly, Govindan et al. (2014) mention that SMEs have fewer opportunities to produce practical environmental information, have smaller amount of resources (e.g., technology, funds, time, environmentally educated employees) and personal incentives (e.g., personal views of managers/owner regarding environmental performance). Walker et al. (2008) state that there are barriers which hinder firms in the adoption of G-SCM practices such as the local character of a green supply chain project, a lack of resources, a lack of essential information, a small number of suppliers, and relevant legislation.

Additional another range of internal barriers could play a negative role in adopting G-SMC such as institutional norms and values. Bowen et al. (2001) introduce the concept of 'environmental illiteracy' of employees who work in purchasing departments as a very important barrier for firms in adopting environmental management practices. The majority of managers consider green supply chain management as additional costs. Moreover, the lack of mechanisms and tools facilitating the collection of environmental data and employees with relevant experience are very important internal barriers. It is obviously that managers who are informed about the barriers might be reluctant to adopt G-SCM.

A rational hypothesis is:

H3: Managers informed about the barriers of G-SCM are more likely to be reluctant to implement such practices.

3 Methodology

This section describes the methodology employed. A questionnaire research methodology was used. The methodology includes four sub-sections. The first illustrates the research framework which outlines the rational connections of hypotheses development. The second section includes criteria used to select the sample. The third section analyzes the main parts of the questionnaire and the final section describes basic variables which are statistically tested.

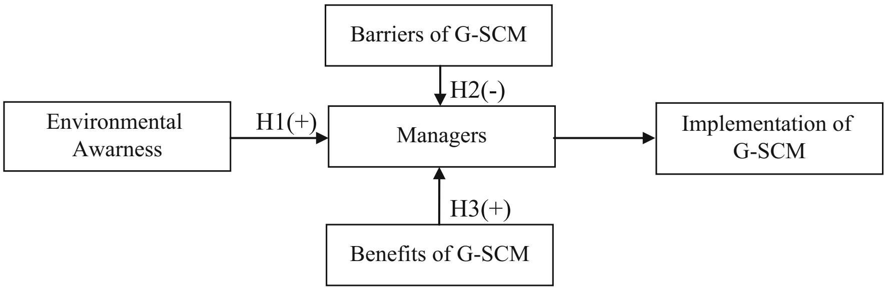


Fig. 1 Research framework

3.1 Research Framework

The research framework assists in developing an investigation tool of identifying the relationships of rational hypotheses developed in the previous section. The research strategy of this paper is based on Fig. 1. This implies that managers with environmental knowledge are more likely to implement G-SCM practices. This is a rational and expected result since managers are informed regarding G-SCM practices and similar success stories of other firms. Similarly, the managers who are well informed about the benefits of G-SCM might be willing to implement these practices. This shows that managers are informed regarding new innovations, reducing operational costs, improving reputation and gaining competitive advantage. Finally, managers who are informed about the barriers SMEs face when implementing G-SCM practices are more likely to be reluctant to adopt such practices.

3.2 Sample Selection

The data used in this research is from a questionnaire survey conducted among managers in the Greek manufacturing industry. The research focuses mainly on general managers/owners since the small size of sampled firms restricts their capability for greater specialization and several distinct departments (e.g., logistics, accounting, human resources). These companies were selected by using three criteria: the type of firm (SMEs), the geographical proximity (the industrial area of Thessaloniki, Greece) and previous experience in supply chain management practices (implementation of any type of supply chain management). This study focuses on 250 SMEs from different manufacturing sectors with less than 10 employees. Managers provided answer only for their firms. 185 questionnaires were sent via e-mail and 65 questionnaires via face-to-face meetings. The response rate is 32 % or 80 questionnaires. Research was conducted from 2014 to 2015. Table 1 shows the geographical allocation of SMEs in the area of Thessaloniki.

Table 1 Geographical distribution of the sample

Area	Frequency	%
Eastern Thessaloniki	42	52.5
Western Thessaloniki	23	28.75
Central Thessaloniki	15	18.75
Total	80	100

Table 2 Sampled SMEs per sectors

Industry	Number of firms	Average of employees
Food industry	23	8
Painting manufacturing	21	8.5
Electronic manufacturing	15	10
Chemical industry	9	10
Wine industry	7	7.5
Steel industry	5	9.5

In particular, the research was focused on the industrial area of Thessaloniki (Sindos) where approximately 700 firms operate. As aforementioned, only 250 firms (36 %) operate in manufacturing sector, the rest of the sample operates in service sectors.

Table 2 indicates the distribution of firms per size and industry using the number of employees. There are six main industrial sectors which range from the food industry to the steel industry. The majority of firms belongs to the food industry and paint manufacturing. The electronic and chemical industries employ the majority of employees (average 10 employees).

3.3 *Questionnaire Structure*

The data used in this study is arisen from questionnaire responses of Greek managers. The final version of questionnaire is tested prior to the data collection by three experts in green supply chain management and two managers. After final corrections as result of experts feedback, the questionnaire is classified in four parts and includes 18 questions in total. Each question is measured in a 5-point Likert scale: 1 not at all, 2 not very much, 3 quite a lot, 4 much, and 5 very much. The first part of questionnaire intends to draw data regarding the knowledge of managers/ owners about general environmental problems and the responsibilities of SMEs to the natural environment. The second part includes questions to draw information regarding the viewpoints of managers/owners on environmental practices which are available for SMEs in the context of supply chain management. The third part of the questionnaire includes questions concerning the barriers to the adoption/implementation of environmental management practices across the supply chain which SMEs encounter. The effects of the economic crisis on the decision of managers to implement environmental management practices are also investigated. The last part includes a number of potential benefits for SMEs when adopting G-SCM.

3.4 Variable Description: Statistical Test

Data was analyzed using SPSS software. In particular, some Spearman rank correlations were run. The basic variables which are used to answer the hypotheses developed in the previous section are presented in Table 3. The variables ENVPRO, FIRMRESP, and ENVSYT seek information regarding the level of knowledge of managers/owners about general environmental issues and the responsibilities of SMEs to them. Additionally, the variables ENVPRAC, ENVSUP and ENVTRANS are associated with environmental management practices in the context of supply chain. Finally, variables ENVBAR and ENVIBEN provide information regarding the potential barriers and benefits of firms in order to implement environmental green supply chain management.

It is worth noting that the Table 3 includes 35 variables which are which are codified questions. The initial (18) questions provide many alternative options for respondents. For example, a question regarding the knowledge of environmental management practices (one question) includes 5 alternative options for respondents such as wastewater, water consumption, paper consumption, plastic consumption, energy consumption, air pollution.

4 Results and Discussion

This section analyzes the findings of this study. The analysis has been made according to the three hypotheses which were developed in the second section.

4.1 How Environmental Awareness of Managers/Owners of SMEs Affect Their Propensity to Invest in Green Supply Chain Management Practices

Table 4 shows the level of knowledge of respondents regarding general environmental problems and the environmental responsibilities of SMEs. Respondents consider climate change to be the most significant problem and also consider natural resource depletion to be an important responsibility of SMEs. The second important option of respondents is regarding water pollution (ENVPRO1). This is a rational finding since the majority of managers run firms in water-intensive sectors and are highly-dependent on water and wastewater legislation. However,

Table 3 Variable description

Variable	Description	Average	St deviation
ENVPRO1	Water pollution	0.33	0.122
ENVPRO2	Climate change	0.83	0.343
ENVPRO3	Natural resource depletion	0.15	0.313
ENVPRO4	Energy pressures	2.35	1.319
FIRMRESP1	Wastewater	3.28	1.133
FIRMRESP2	Water consumption	2.41	1.144
FIRMRESP3	Paper consumption	3.42	1.282
FIRMRESP4	Plastic consumption	3.24	1.490
FIRMRESP5	Energy consumption	3.11	1.335
FIRMRESP6	Air pollution	3.93	1.028
ENVSYST1	Environmental Management Systems (ISO 14001)	3.20	1.300
ENVSYST2	Eco-label	3.14	1.243
ENVPRAC1	Paper recycle	4.34	0.912
ENVPRAC2	Renewable Energy use	4.14	0.898
ENVPRAC3	Cooperation with NGOs	4.38	0.780
ENVSUP1	Recyclable raw materials	4.48	0.633
ENVSUP2	Recyclable packaging	3.22	1.145
ENVSUP3	Reuse packaging	3.26	1.255
ENVTRANS1	Clean technology	2.19	1.160
ENVTRANS2	Vehicles management	4.44	0.820
ENVTRANS3	Common orders with other firms	2.55	1.175
ENVTRANS4	Less use of packaging	2.12	1.458
ENVBAR1	Lack of environmental legislation	4.15	1.175
ENVBAR2	Less environmental knowhow	0.86	0.348
ENVBAR3	Low environmental awareness of suppliers	0.22	0.310
ENVBAR4	Low environmental awareness of staff	2.35	1.422
ENVBAR5	Lack of financial resources	3.22	1.122
ENVBAR6	Lack of subsidies	2.44	1.130
ENVBAR7	Effects of economic crisis	3.43	1.275
ENVBENEF1	Reduction of penalties	2.28	1.485
ENVBENEF2	Lending incentives	1.84	1.342
ENVBENEF3	Improved infrastructure	1.16	2.321
ENVBENEF4	Information campaigns	3.25	1.421
ENVBENEF5	Tax reduction	2.20	1.115
ENVBENEF6	Government subsidies	2.45	1.108

respondents seem to have little knowledge regarding water consumption (FIRMESP2) and paper consumption (FIRMESP3). This implies that respondents are aware of wastewater problems but not issues regarding water consumption. This is a rational finding since water scarcity issues are not reported in the area where the firms examined are located.

Table 4 Knowledge of managers/owners about environmental problems

	Not at all (1)	Not very much (2)	Quite a lot (3)	Much (4)	Very much (5)	Ranking[a] (4) + (5)
ENVPRO1	6	5	5	51	33	2
ENVPRO2	4	5	1	40	50	1
ENVPRO3	1	8	24	37	28	5
ENVPRO4	2	9	32	41	15	7
FIRMRESP1	1	15	18	48	18	4
FIRMRESP2	1	6	20	49	24	3
FIRMRESP3	23	32	24	12	7	10
FIRMRESP4	12	35	28	17	6	9
FIRMRESP5	2	13	22	37	24	6
FIRMRESP6	5	20	40	21	12	8

[a]This column is the sum of the fourth column (much) and fifth column (very much)

Table 5 shows Spearman correlations. In particular, a high positive correlation seems to be achieved between the knowledge of managers/owners regarding environmental issues and SMEs' environmental responsibility with their decision to implement G-SCM. For example, those managers/owners who were informed about water pollution, are more likely to adopt environmental management practices such as ISO 14001 ($r_s = 0.396$, $p < 0.01$), Eco-label ($r_s = 0.671$, $p < 0.01$) and paper recycling ($r_s = 0.228$, $p < 0.01$). This is rational since the majority of firms examined operate in sectors where severe legal requirements regarding wastewater treatment are in place. Additionally, those managers/owners who consider the discharge of wastewater a significant responsibility of their firms, are willing to implement ISO 14001 ($r_s = 0.423$, $p < 0.01$) and clean technology ($r_s = 0.340$, $p < 0.01$). This shows that managers would face potential water problems through holistic environmental management practices which could offer win-win solutions such economic and environmental factors (Porter and van der Linde 1995).

4.2 How the Knowledge About Benefits of Green Supply Chain Management Practices Are Positively Related to Green Supply Chain Management Practices

Table 6 shows the level of knowledge of respondents regarding the benefits gained by SMEs as a result of/when incorporating environmental concerns into the supply chain management. Respondents consider the lending incentives (ENVBENEF2) a very important benefit as well as the reduction of penalties as a result of non-compliance (ENVBENEF1) with environmental regulations. Respondents want to avoid potential penalties which are associated with environmental issues such as water pollution. It is worth noting that government subsidies are ranked third in the list. It seems that respondents prefer to avoid potential litigation risks,

Table 5 First hypothesis correlations

	ENVSYST1	ENVSYST2	ENVPRAC1	ENVPRAC2	ENVPRAC3	ENVSUP1	ENVSUP2	ENTRNS1
ENVPRO1		0.396**	0.671**	0.228*				
ENVPRO2	0.396**		0.474**	0.331**				
ENVPRO3	0.671**				–0.334**	0.285**		
ENVPRO4	0.228*	0.331**					0.423**	
FIRMRESP1	0.423**							0.340**
FIRMRESP3					0.552**	0.652**		
FIRMRESP4					0.285*	0.431**	0.242*	
FIRMRESP6					0.294**			

*Correlation is significant at the 0.05 level (2-tailed)
**Correlation is significant at the 0.01 level (2-tailed)

Table 6 Knowledge of managers/owners for benefits of green supply chain management

	Not at all (**1**)	Not very much (**2**)	Quite a lot (**3**)	Much (**4**)	Very much (**5**)	Ranking (**4**) + (**5**)[a]
ENVBENEF1	5	2	21	57	15	2
ENVBENEF2	1	0	7	46	44	1
ENVBENEF3	0	7	29	49	15	4
ENVBENEF4	0	18	33	27	22	6
ENVBENEF5	0	10	33	39	18	5
ENVBENEF6	0	1	33	52	13	3

[a]This column is the sum of the fourth column (much) and fifth column (very much)

Table 7 Second hypothesis correlations

	ENVSYST1	ENVSYST2	ENVPRAC1	ENVPRAC2	ENTRNS2
ENVBENEF1	0.330**	0.393**	0.278*		
ENVBENEF2	0.224*	0.240*			
ENVBENEF3	0.373**	0.598**			
ENVBENEF4			0.256*	0.256*	0.423**
ENVBENEF5		0.257*			
ENVBENEF6	−0.242*		−0.242*		

*Correlation is significant at the 0.05 level (2-tailed)
**Correlation is significant at the 0.01 level (2-tailed)

while preferring to cooperate with the private rather than public sector in order to raise financial resources to invest in environmental management practices. The avoidance of public authorities could be explained through cultural, sociological and structural characteristics. The worst rating in the list is for improved infrastructure (ENVBENEF3) and governmental information campaigns (ENVBENEF2). They are consistent in their response by demonstrating reluctance for the public sector. Social capital and trust of a number of public institutions are very limited at times of financial crises.

Table 7 shows Spearman correlations for the second hypothesis. There is a positive correlation between the knowledge of managers/owners about the benefits of implementing G-SCM. In particular, managers/owners who consider lending incentives important are positively associated with practices IS0 14001 ($r_s = 0.330$, $p < 0.01$), Eco-Label ($r_s = 0.393$, $p < 0.01$) and Renewable energy ($r_s = 0.278$, $p < 0.01$). It is obvious that the managers who have greater trust in the private sector for raising financial resources they show a preference to invest in a holistic environmental strategy in order to exploit new opportunities. Those managers who consider governments subsidies to be important are negatively associated with ISO 14001 ($r_s = -0.242$, $p < 0.01$) and Renewable energy ($r_s = -0.242$, $p < 0.01$). This shows that managers who are after state money (without risk) have presented a reactive view for environmental problems. They seem to be reluctant regarding voluntary strategies such as ISO 14001 and renewable energy. This could be explained by firms only addressing legal requirements for environmental protection.

4.3 How the Knowledge of Managers/Owners About the Barriers of Implementing Green Supply Chain Management Practices Make Them Reluctant to Adopt These Practices

Table 8 shows the level of knowledge of respondents regarding the barriers faced by SMEs in their effort to implement G-SCM. The most significant barrier is the lack of awareness of staff regarding green practices (ENVBAR4). The respondents classify the lack of environmental legislation (ENVBAR1) in second place. This implies that middle and low level managers have a low level of knowledge regarding environmental legislation. This is confirmed from the majority of the literature regarding SMEs and environmental management (Brammer et al. 2012). Hillary (2004) points out that a significant factor that affects the degree of compliance of SMEs with environmental legislation is the absence of a central source to inform them. Similarly, managers classify low environmental know-how in third place. This is confirmed by a number of studies of the corporate environmental management field where the success of environmental management practices is linked to the environmental experience of managers and employees (Boiral 2002). It is worth noting that the economic crisis is of little importance for SMEs (ENVBAR7). This implies that managers consider it an important strategy to adopt environmental management practices. They believe new opportunities could be exploited and they could gain sustainable competitive advantage for other firms. They also state that their exports could be strengthened mainly to countries with high average salaries, disposable income and GDP.

Another significant finding is the low classification of a lack of financial resources as a barrier to SMEs. This finding is not confirmed by other studies. The majority of studies show that managers of SMEs maintain that one important factor which hinders their effort to green their firm and supply chain is a lack of financial resources and opportunities to identify them. However, the respondents support that financial opportunities for environmental strategies do exist.

Table 8 Knowledge of managers/owners about barriers to the implementation of green supply chain management practices

	Not at all **(1)**	Not very much **(2)**	Quite a lot **(3)**	Much **(4)**	Very much **(5)**	Ranking **(4) + (5)**[a]
ENVBAR1	4	7	22	46	20	2
ENVBAR2	0	9	28	46	16	3
ENVBAR3	1	12	28	48	10	5
ENVBAR4	1	5	9	37	46	1
ENVBAR5	6	13	33	37	10	6
ENVBAR6	2	6	30	44	16	4
ENVBAR7	4	16	22	2	12	7

[a]This column is the sum of the fourth column (much) and fifth column (very much)

Table 9 Third hypothesis correlations

	ENVSYST2	ENVPRAC1	ENVPRAC2	ENVPRAC3	ENVSUP1	ENVSUP2	ENTRNS1
ENVBAR1	0.233**						
ENVBAR2		0.272*	0.381**	0.231*			
ENVBAR3	0.302**		0.259*				
ENVBAR4			0.302**		0.230*		0.249*
ENVBAR5		0.334**	0.249**	0.285*			
ENVBAR6		0.229*		0.301**			
ENVBAR7						−0.229*	

*Correlation is significant at the 0.05 level (2-tailed)
**Correlation is significant at the 0.01 level (2-tailed)

Table 9 shows Spearman correlations. There is a positive correlation between barriers and environmental management practices of G-SCM (e.g., Eco-label; $r_s = 0.233, p < 0.01$). The explanation is that the negative influence of these barriers is very limited in managers' opinion. The Economic crisis affects negatively the decision of managers to adopt environmental management practices ($r_s = -0.229$, $p < 0.01$). This means that their decision to adopt green supply chain strategies move in different direction in relation to economic crisis. Therefore, the recovery of the economy from financial crisis will drive managers to adopt green supply chain practices.

5 Conclusions

This paper aims to identify how the knowledge of managers/owners regarding environmental responsibility of firms, barriers and benefits of implementing environmental management practices in the overall supply chain management affect their final decision to implement these strategies. The contribution of this paper is to connect previous literature with the knowledge of managers which is a very important factor in the adoption of green supply chain management practices. The commitment of top management for implementing environmental management practices is a very significant reason for the success of corporate environmental policy (Daily and Huang 2001; Linnenluecke and Griffiths 2010).

The previous literature mainly focuses on knowledge of G-SCM practices and did not investigate the benefits and barriers to adopting these practices. The findings show that several barriers and benefits exist which seem to play a critical role in managers/owners decisions lending incentives being one the main driving factors. Similarly, many barriers are identified that affect managers/owners decisions to adopt green supply chain management, such as the effect of the economic crisis and the lack of financial resources. In particular, the first hypothesis has been supported. Knowledge regarding environmental topics and G-SCM practices seems to be very important. Managers consider significant to have knowledge regarding water pollution, climate change, natural depletion and energy pressures as well as water consumption practice, wastewater practice and plastic consumption practice.

In addition, the second hypothesis has been supported. The majority of managers are positive about investing in G-SCM practices in the case where they are well-informed on the benefits for SMEs of adopting such practices. An important contribution of these findings is the higher emphasis of managers on the private financial sector (e.g., private loans) rather than the public financial sector (e.g., subsidies). This indicates that managers understand environmental strategies as strategic tools which offer new perspectives and opportunities and thus tend adopt proactively G-SCM practices. Finally, the third hypothesis is partially supported. Although the majority of the parameters have been supported such as lack of environmental experience of staff, lack of environmental legislation and less know-how, little significance has been given to the effects of the economic crisis and lack of financial resources. In particular, a remarkable finding shows that managers put low emphasis on the economic crisis as a barrier to adopting G-SCM practices.

As with every research study, there is a limitation which should be mentioned: the small sample of only 80 SMEs. Although the response rate was 32 % and the majority of present research of SMEs utilize small samples (Lee 2008; Holt and Ghobadian 2009) due to lack of response to questionnaires (e.g., lack of time, professional privilege), the merit of this research is the contribution to empirical knowledge regarding new theoretical hypotheses and not to make general theory. Another limitation of this study which associated with the previous one (the small sample of firms) is that the inability to identify significant models with more than two variables (pair wise comparisons) to improve deeper understand of the existing correlations and support strongly our findings. This research lays the foundations for developing a number of future research studies on G-SCM and SMEs. In particular, studies should be conducted to identify how the level of knowledge could affect the decision of managers to implement G-SCM practices.

References

Awasthi A, Chauhan SS, Goyal SK (2010) A fuzzy multicriteria approach for evaluating environmental performance of suppliers. Int J Prod Econ 126(2):370–378

Bai C, Sarkis J (2010) Integrating sustainability into supplier selection with grey system and rough set methodologies. Int J Prod Econ 124(1):252–264

Boiral O (2002) Tacit knowledge and environmental management. Long Range Plann 35(3): 291–317

Bowen FE, Cousins PD, Lamming RC, Faruct AC (2001) The role of supply management capabilities in green supply chain. Prod Oper Manag 10(2):174–189

Brammer S, Hoejmose S, Marchant K (2012) Environmental management in SMEs in the UK: practices, pressures and perceived benefits. Bus Strat Environ 21(7):423–434

Carter CR, Dresner M (2001) Purchasing's role in environmental management: cross functional development of grounded theory. Supply Chain Manag 37(3):12–26

Carter CR, Kale R, Grimm CM (2000) Environmental purchasing and firm performance: an empirical investigation. Transport Res E Logt 36(3):219–228
Chiou T-Y, Chan HK, Lettice F, Chung SH (2011) The influence of greening the suppliers and green innovation on environmental performance and competitive advantage in Taiwan. Transport Res E Logt 47(6):822–836
Cote RP, Lopez J, Marche S, Perron GM, Wright R (2008) Influence practices and opprotunities for enviornmetnal supply chain management in Nova Scotia SMEs. J Clean Prod 16(15):1561–1570
Daily BF, Huang S-C (2001) Achieving sustainability through attention to human resource factors in environmental management. Int J Oper Prod Manag 21(12):1539–1552
Darnall N, Jolley GJ, Handfield R (2008) Environmental management systems and green supply chain management: complements for sustainability? Bus Strat Environ 18:30–45
Diabat A, Govindan K (2011) An analysis of the drivers affecting the implementation of green supply chain management. Resour Conserv Recycl 55:659–667
Gadenne DL, Kennedy J, McKeiver C (2009) An empirical study of environmental awareness and practices in SMEs. J Bus Ethics 84:45–63
Govindan K, Kaliyan M, Kannan D, Haq AN (2014) Building supply chain system capabilities in the age of global complexity: emerging barriers analysis for green supply chain management implementation in Indian industries using analytic hierarchy process. Int J Prod Econ B 147: 555–568
Green K, Morton B, New S (1996) Purchasing and environmental management: interactions, policies and opportunities. Bus Strat Environ 5:188–197
Handfield R, Walton SV, Sroufe R, Melnyk SA (2002) Applying environmental criteria to supplier assessment: a study in the application of the analytical hierarchy process. Eur J Oper Res 141(1):70–87
Hillary R (2004) Environmental management systems and the smaller enterprise. J Clean Prod 12: 561–569
Holt D, Ghobadian A (2009) An empirical study of green supply chain management practices amongst UK manufacturers. J Manuf Tech Manag 20(7):933–956
Lee S-Y (2008) Drivers for the participation of small and medium-sized suppliers in green supply chain initiatives. Int J Supply Chain Manag 13(3):185–198
Linnenluecke M, Griffith A (2010) Corporate sustainability and organizational culture. J World Bus 45:357–366
Linton JD, Klassen R, Jayaraman V (2007) Sustainable supply chains: an introduction. J Oper Manag 25:1075–1082
Mathiyazhagan K, Govindan K, Noorulhaq A, Geng Y (2013) An ISM approach for the barrier analysis in implementing green supply chain management. J Clean Prod 47:283–297
Min H, Galle WP (1997) Green purchasing strategies: trends and Implications. Int J Purch Mater Manag: 1–17
Muduli K, Govindan K, Barve A, Kannan D, Geng Y (2013) Role of behavioural factors in green supply chain management implementation in Indian mining industries. Resour Conserv Recycl 76:50–60
Nikolaou IE, Evangelinos KI, Allan S (2013) A reverse logistics social responsibility evaluation framework based on the triple bottom line approach. J Clean Prod 56:173–184
Noci G (1997) Designing 'green' vendor rating systems for the assessment of a supplier's environmental performance. Eur J Purch Supply Manag 3(2):103–114
Porter ME, van der Linde C (1995) Toward a new conception of the environment-competitiveness relationship. J Econ Perspect 9(4):97–118
Pujari D, Wright G, Peattie K (2003) Green and competitive: Influences on environmental new product development performance. Int J Bus Res 56(8):657–671
Sarkis J, Zhu Q, Lai K-H (2011) An organizational theoretic review of green supply chain management literature. Int J Prod Econ 130:1–15
Seuring S, Muller M (2008) From a literature review to a conceptual framework for sustainable supply chain management. J Clean Prod 16:1699–1710

Seuring S, Müller M (2008) Core issues in sustainable supply chain management—a Delphi study. Bus Strat Environ 17:455–466

Srivastava SK (2007) Green supply-chain management: a state-of the-art literature review. Int J Manag Rev 9(1):53–80

Tsireme A, Nikolaou EI, Georgantzis N, Tsagarakis KP (2012) The influence of environmental policy on the decisions of managers to adopt G-SCM practices. Clean Technol Environ 14(14): 953–964

Walker H, Di Sisto L, McBain D (2008) Drivers and barriers to environmental supply chain management practices: lessons from the public and private sectors. J Purch Supply Manag 14(1):69–85

Zhu Q, Sarkis J (2004) Relationships between operational practices and performance among early adopters of green supply chain management practices in Chinese manufacturing enterprises. J Oper Manag 22:265–289

Zsidisin GA, Siferd SP (2001) Environmental purchasing: a framework for theory development. Eur J Purch Supply Manag 7(1):61–73

Retail Category Management: A Review on Assortment and Shelf-Space Planning Models

Marina Karampatsa, Evangelos Grigoroudis, and Nikolaos F. Matsatsinis

Abstract Retail Category Management addresses a series of questions and demands decisions for category managers on critical issues such as product assortment and shelf-space planning. Product assortment planning involves listing decisions based on consumer behavior and substitution effects. Shelf space allocation involves facing and replenishment decisions based on space elasticity effects and constraints of limited shelf space and restocking capacity. The complexity of these questions has grown significantly in recent years due to product proliferation and various consumer choice effects in the retail environment. It is an increasingly difficult task for category managers to find an effective assortment due to consumer preferences instability and the extremely large number of possible assortments. This chapter presents an updated review on scientific models that deal with assortment and shelf space planning and other related topics, such as consumer response to stock-outs and consumer perceptions of assortment variety. One of the main objectives of this literature review is to show that shelf space allocation models do not clearly and comprehensively address assortment selection, neglect substitution effects between products, and ignore the stochastic nature of demand. Assortment planning models on the other hand mostly ignore shelf space constraints and neglect space depend demand.

Keywords Retailing • Assortment planning • Shelf-space planning • Product substitution • Variety perceptions

1 Introduction

For retailers, the building block for any kind of management decision is the *category*. A category is defined as a distinct manageable group of products that consumers perceive to be related and/or substitutable (FMI 1995). This group of SKUs can sometimes be segmented into *subcategories* or *classes*. Category

M. Karampatsa • E. Grigoroudis (✉) • N.F. Matsatsinis
Decision Support Systems Laboratory, Technical University of Crete, Chania, Greece
e-mail: mkarabatsa@isc.tuc.gr; vangelis@ergasya.tuc.gr; nikos@ergasya.tuc.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_3

management in retail (hereinafter referred to as Retail Category Management—RCM) is a process of managing categories as Strategic Business Units, producing enhanced business results by focusing on delivering consumer value. It was developed as a strategy for retailers to successfully compete in each retail category for the shopper's loyalty and dollars (Blattberg 1995) and is designed to help retailers provide of the right mix of products, at the right price, with the right promotions, at the right time, and at the right place (Gruen and Shah 2000).

In this study, we focus on the literature related to RCM and especially on scientific models that deal with assortment and shelf space planning. RCM addresses a series of questions and demands decisions for category managers on critical issues such as *which* products to include in an assortment and *how to allocate* these products to shelves (issues related to assortment and shelf space planning respectively).

Several mutually reinforcing trends in retail environment have made these decisions two of the most critical marketing and operational decisions in the industry nowadays. One such trend is product proliferation and the fact that competition for shelf space is at an all-time high (Ball 2004; Hübner 2011; Murray et al. 2010). Since the 1990s, there has been a significant proliferation in new product items in retail stores as both manufacturers and retailers saw it as a strategic way for increasing market shares (Drèze et al. 1994). Retailers have also introduced new categories into their stores (e.g. fresh produce, organic products) in order to satisfy diverse consumer needs. The average number of different items in overall store assortments had increased by 20 % between 1970 and 1980, by 75 % between 1980 and 1990 (Greenhouse 2005) and by 30 % between 2000 and 2009 (Hübner 2011). These new products and categories are putting a tremendous demand on the available shelf space which is practically fixed and every day becomes scarcer for existing stores.

At the same time, retailers have been experiencing a steady increase in their operational costs as a consequence of carrying the aforementioned large assortments of products and a pressure to improve their operational efficiency because of the increased levels of competition in the industry. Krafft and Mantrala (2010) in their book "Retailing in 21st century" indicate that competiveness of both U.S. retail and global marketplace is escalating. Reyes and Frazier (2007) suggest that while gross margins average about 28 % of sales, net profits after taxes are only 1 % for the grocery retail industry in U.S., causing retailers to continually search for operational efficiency and even greater customer satisfaction.

How to control such spiraling costs from increased levels of product assortments and competition in an industry already marked by one of the thinnest profit margins? Not surprisingly, German retailers and consumer goods producers recently rated "optimization of product portfolio and category management" the most important task for achieving performance goals according to a survey of McKinsey and Company (Hübner 2011). Mantrala et al. (2009) in their seminal work "Why is assortment planning so difficult for retailers?" refer that assortment decisions have an enormous impact in the short and long term profitability of a retailer.

Retailers in their attempt to offer a balance among *variety* (number of categories), *depth* (number of Stock Keeping Units—SKUs within a category) and *service level* (the number of individual items of a particular SKU) are constrained by the amount of money they can invest in inventory and by their physical space. Therefore, offering more variety may limit the depth within categories and the service level, or both. By making appropriate trade-offs with respect to variety, depth, and service levels, retailers hope to satisfy customers' needs by providing the right merchandise in the right store at the right time. If they fail, customers defect, causing losses in both current and future sales.

Furthermore, a reality faced by retailers in their pursuit of operational efficiency, is that the majority of shoppers take their purchase decision in-store (Drèze et al. 1994). This led many retailers to divert a proportion of their marketing budget from out-of-store media advertising to in-store marketing (Chandon et al. 2009; Ge et al. 2009). Consequently, it is becoming extremely important for retailers how to optimally organize their product assortments to generate more profit from their existing and limited shelf space (Greenhouse 2005; Hübner 2011; Lim et al. 2004; Murray et al. 2010).

The goal of this study is to identify and structure these planning problems and constraints and their effects in RCM. It provides a summary of empirical studies, consumer choice and demand estimation theories, and the latest advances in scientific modeling in the area of RCM.

The remainder is organized as follows. The second section sets the scope of RCM and other related definitions. The third section review the published empirical studies on topics related to retail category management, such as customer driven substitution, consumer perceptions of assortment variety, and the sales—space relationship between unit sales and self-space. The forth section review the most popular approaches for demand estimating, while the fifth section identify research progress in assortment and shelf space planning. The concluding section provides an outlook on further research areas.

2 Definitions and Scope of Retail Category Management

In the retail environment, products are grouped together into categories to reflect customers' needs based on how the product is used, consumed or purchased. RCM aims to provide the shoppers and consumers with what they want, where they want it, and when they want it. Consistent with the above analysis, RCM can be characterized as a multi-perspective same-time decision problem (Hübner 2011), where a series of questions have to be solved jointly: what to list (assortment planning), how to put the products into the shelves (shelf planning), how often to replenish (inventory planning) and how to prize (price planning).

Table 1 Retail category management decision aspects

Decision aspects	Decision area	Effects and constraints
Assortment planning	Assortment size, listing (delisting) of products	• Substitution demand • Budget and space constraints
Shelf space planning	Shelf space allocation, number of facings	• Space-elastic demand • Space constraints, restocking capacity
Inventory planning	Restocking and replenishment frequency	• Space-elastic demand • Customer reactions to stock-out • Operational restocking constraints and costs
Price planning	Pricing of products	• Price-elastic demand • Retailer's price image, category role, budget constraints

According to Table 1, these planning questions are defined as:

- ***Assortment planning*** involves decisions about (de-) listing products. According to Kök et al. (2009), a retailer's assortment is defined by the set of products carried in each store at each point in time. Consistent with the aforementioned definition, the goal of assortment planning is to specify an assortment that maximizes sales subject to various constraints, such as limited budget for purchasing products, limited shelf space for displaying products, and other constraints such as the desire to have at least two vendors for each type of product. One main feature that needs to be reflected in assortment optimization is the integration of consumers' substitution behavior; their willingness, in other words, to accept a substitute when their favorite product is not available. Traditionally, assortment planning belongs to the field of strategic planning and has a long-term planning horizon.
- ***Shelf space planning*** refers to the problem of allocating the available shelf space among the selected items to be included in the product assortment (Hariga et al. 2007). This is done by assigning facing quantities to individual products under the constraints of limited shelf sizes and restocking capacity (Hübner and Kuhn 2012). The goal of shelf space planning is to improve the return of inventory investment, enhance consumer satisfaction by reducing out-of-stock occurrences and increase sales and profit margins (Yang 2001). An underlying assumption in shelf space planning models is that the sales of a product depend on space allocated to it as well as the space allocated to its competitive products (own- and cross-space elastic demand). Therefore, many researchers have estimated the demand using space elastic functions. These functions are mathematical expressions where the coefficients are called space elasticities. Space elasticity is defined as the sensitivity of the customer to the inventory (or number of facings) displayed in terms of quantity bought (Hariga et al. 2007). Shelf space planning belongs mainly to operative planning and has a mid- to short-term horizon.
- ***Inventory planning*** deals with the restocking decisions about the timing and the size of the orders to be placed for each of the products included in the

assortment. Other relevant planning questions are related to shelf replenishment processes, such as refilling quantities and refilling cycles. Proper inventory planning strives to ensure that markdowns due to overstock and lost sales due to stock-outs are minimized. Inventory planning has a similar to shelf space planning horizon and it also belongs to operative planning.

- ***Price planning*** sets prices for each of the products included in the assortment relatively to other product prices and competition. The goal for a retailer when optimizing pricing in a category of products is to drive sales, volume, and profit while maintaining his price image and the category's role. An underlying assumption in price management is the price elasticity of demand. Price elasticity measures the responsiveness of demand to changes in price for a particular product as well as for its competitive products (own- and cross-price elastic demand). Price planning belongs either to tactical or operative planning depending on retailer and the way a company selects to execute their strategic planning. Consequently, price planning usually has a mid- to short-term horizon.

3 Related Literature and Empirical Studies

In this section, we briefly review the literature and the published empirical studies on topics related to retail category management, such as customer driven substitution, consumer perceptions of assortment variety, and the sales—space relationship between unit sales and self-space.

3.1 Customer Driven Substitution

Consumers sometimes cannot find their favorite product in a store and settle for another similar product instead. This is called substitution according to Kök and Fisher (2007). Demand substitution arises in two basic forms: Stock-out- and assortment-based substitution (Kök et al. 2009):

- **Stock-out-based substitution** is the switch to an available variant by a customer when his favorite product is carried in the store, but is stocked-out at the time of his shopping trip.
- **Assortment-based substitution** is the switch to an available variant by a customer when his favorite product is not carried in the store's assortment.

When customers cannot find their favorite product in a store they visit, because it is either temporarily stocked out or not carried at all, they tend to behave in a variety of ways:

- **Substitute**: buy one of the available items in the category of his/her interest,
- **No purchase**: decides to leave the store and forgo the purchase

Table 2 Empirical studies on consumer substitution behavior

Authors	No purchase (%)	Delay (%)	Store switch (%)	Substitution: Same brand (%)	Substitution: Other brand (%)
Walter and Grabner (1975)	–	3	14	19	64
Emmelhainz et al. (1991)	–	13	14	41	32
Campo et al. (2000) (Cereals)	4	49	3	44	
Campo et al. (2000) (Margarine)	2	30	2	66	
Zinn and Liu (2001)	–	15	23	62	
Gruen et al. (2002)	9	15	31	19	26
Sloot et al. (2005)	3	23	19	18	36
ECR Europe (2003)	9	17	21	16	37
van Woensel et al. (2007)	–	6	10	84	
Verbeke et al. (2013)	–	23.4	20.1	56.6	

- **Delay**: decides to come back later for his/her favorite product,
- **Store switch**: decides to leave the store and search for his/her favorite product at another shop.

Table 2 summarizes the findings of empirical studies on consumer substitution behavior. For example, Sloot et al. (2005) examine how consumers behave when their preferred bread product is out-of-stock, based on 3800 customer interviews performed in three stores of a large Dutch grocery retail chain. Their key observation is that even though customers have a high willingness to substitute (up to 54 %), still 19 % of they may (in case of a stock-out) decide to buy their bread elsewhere, which is still a large amount of lost sales. Gruen et al. (2002) surveyed more than 71,000 customers worldwide and found that 26 % of customers encountering a stock-out situation substitute brand, 19 % substitute another product of the same brand, 31 % buy the same product at another store, and 15 % choose to delay or cancel their purchase.

According to ECR Europe (2003), when facing an out-of-stock, European consumers are more likely to switch brand than retailers (37 % and 16 % respectively), with the rage of observed behaviors varying widely. Another 9 % of consumers will decide not to purchase at all, which means that 9 % of the intended purchases facing a stock-out are definitely lost. According to the same study, the industry forfeits approximately 4 billion euros in sales yearly, because of this 9 % of consumers that decide to opt not to buy at all. For an industry suffering from low margins, this is unacceptable.

The previously mentioned studies examine the consumer response to stock-outs, albeit they do not clearly exclude assortment-based substitution. One of the few studies that investigate the consumer response to out-of-stocks as opposed to permanent assortment reductions (PAR) is the paper of Campo et al. (2000). They report that although the retailer losses in case of a PAR may be larger than those in case of an OOS, there are also significant similarities in consumer reactions in the two cases and OOS reactions for an item can be indicative of PAR responses for that item.

The above mentioned papers showed that between 44 and 84 % of demand can be substituted. The substitution potential, according to other studies, depends on product, situation and consumer characteristics. Fitzsimons (2000) deduces that substitution behavior is driven by the personal commitment to the non-stocked item and the difficulties of switching initial choice. Ge et al. (2009) conclude that the presence of sold-out products may influence the consumers' purchases in two possible ways. First, sold-out products may create a sense of urgency for consumers to expedite their purchases, which the authors call an immediacy effect. Second, sold-out products may enhance the perceived attractiveness of products similar to the sold-out products, which the authors refer to as an informational cascades effect. Nevertheless, all the above studies verify that substitution behavior is inherent in the retail environment and needs to be reflected in category planning optimization.

3.2 Consumer Perceptions of Assortment Variety

This section is dedicated to empirical studies that focus on the way consumers perceive and evaluate product assortments. Consumer assortment evaluations have been shown to be one of the top three criteria, along with location and price, in determining retail patronage (Broniarczyk and Hoyer 2010). In the 1980s and early 1990s, retailers assumed that larger product assortments better met consumer needs and this led to the expansion of the average number of different items included in overall store assortments (see chapter "The Application of a Business Process Modeling Architecture in the Supply Chain of a Manufacturing Company: A Case Study"). However, several empirical studies put this assortment proliferation into question.

As part of its 1993 report, the Food Marketing Institute conducted a field study in which it reduced SKUs in six product categories (cereal, pet food, salad dressing, spaghetti sauce, toilet tissue, and toothpaste) at 24 test stores. Superfluous SKUs were eliminated and shelf space was held constant, with more space now allocated to items with higher market share. The results showed no significant impact of SKU reduction on sales, which suggested that efficient assortment might result in cost savings without loss of sales. However, it was difficult to draw definite conclusions as the magnitude of SKU reduction in these categories was ambiguous (Broniarczyk and Hoyer 2010). In another series of in-store studies, Drèze et al. (1994) examined a 10 % reduction of low-volume SKUs in eight test categories over a 4-month period. Their shelf-reset procedure was similar to that in the Food Marketing Institute study, with shelf space held constant and freed-up shelf space reallocated to high-volume SKUs. Across eight categories, they found that sales increased by 4 % for 30 test stores compared with 30 control stores. These results began to indicate the possibility that smaller assortments might even have a positive effect on sales for retailers.

Recent empirical studies entirely challenge the common assumption that having more choices is necessarily more motivating for consumers than having fewer. Consumer choice models often assume that customers are perfectly knowledgeable about their preferences and the product alternatives which are offered to them. Therefore, consumers are always better off when they choose from a broader assortment of products. This implicit assumption that more choice benefits consumers is consistent with the classical economic and psychological theories that larger assortments should always be beneficial to consumers because they provide a potentially better probability that consumers will find their ideal product and offer flexibility for variety seekers (Chernev 2003; Iyengar and Lepper 2000). On the contrary, recent findings have suggested that consumers might perceive large assortments negatively and that increasing the choice set may confuse them, create frustration or a sense of being overwhelming, increasing the probability of delaying their choice or not choosing at all (Dhar 1997; Iyengar and Lepper 2000).

To complicate even more the retailer's challenge when choosing an assortment, an empirical study conducted by Broniarczyk et al. (1998) showed that the actual variety of an assortment may not match the perceived variety a consumer experiences. This perception can be influenced by:

- the space devoted to a category (Broniarczyk and Hoyer 2010),
- the presence or absence of a favorite item (Amine and Cadenat 2003; Broniarczyk et al. 1998),
- the similarity of items (Hoch et al. 1999),
- or the arrangement of the assortment and how it is organized (Morales et al. 2005; Simonson 1999).

3.3 The Sales: Space Relationship

In their pursuit of operational efficiency, a key reality faced be retailers is the fact that the majority of consumers decides about their final purchases in the store (Chandon et al. 2009; Drèze et al. 1994; Ge et al. 2009). A series of experimental studies conclude that product location and facing area can influence consumers' attention and purchase behavior. For example, Chandon et al. (2009), using an eye-tracking experiment, found that the number of facings has a strong impact on consumer attention. Going from 4 to 8 facings increased the probability of noticing the item by 28 % and the probability of reexamining it by 40 %. But adding additionally four facings only added an extra 7 % to noting and extra 19 % to reexamination. They identify facing variation as the most significant effect among in-store factors, even stronger than vertical and horizontal positioning and pricing.

Other field experiments tried to quantify the relationship between demand and space. Even though, all agreed that a significant relationship exists, they couldn't agree in their estimates of space elasticities (most vary between 0.15 and 0.8). The diversity in empirical estimates of space elasticities is mainly due to the different methodology and data used in each study (Desmet and Renaudin 1998).

Among many studies using store experimentation, the most interesting are the study of Curhan (1972) which, taking a large sample, found an average value of 0.212 for space elasticity, and the large scale work of Drèze et al. which concluded that the number of facings allocated to a product is one of the least important success factors, whereas the position of the product on the shelf seems to be far more important for determining sales. Drèze et al. (1994) concluded that supermarket retailers should expect 4–5 % profit gains in category sales from better product positioning and space allocation. Both draw the conclusion that most products receive an over-allocation of shelf space. In another study (Hansen and Heinsbroek 1979), the authors surveyed about 20 experiments in order to evaluate the space elasticity of demand and they found a low average space elasticity of 0.15. Finally, in their research, Corstjens and Doyle (1981) found an even lower mean value of 0.086.

4 Consumer Choice Modeling and Demand Estimation

Consumer choice models constitute the fundamental platform for category planning. In this chapter, we provide a review of those models that may be classified as: utility based models, exogenous demand models, and space-elastic demand models. In assortment planning, the most popular approaches for demand estimating are *multinomial logit models* and *exogenous models*. In shelf space planning, on the other hand, the demand rate is formulated as a function of the shelf space allocated to products *(space-elastic models)*. In Sect. 4.4, we also briefly discuss the estimation of the demand models mentioned in the preceding sections, depending on the type of data available.

Before proceeding, we will define the notation for category planning in a single category or subcategory at a single store. This notation is common throughout this chapter.

N The set of products in a category or subcategory $N = 1, 2, 3, \ldots n$
S The subset of products carried in the assortment $S \subset N$
r_i Selling price of product i
c_i Purchasing cost of product i
λ Mean number of customers visiting the store per period

4.1 Utility Based Models

Utility based choice models assume that consumers are rational utility maximizers and associate a utility Ui with each product $i \in N$. In addition, for the consumers that choose not to purchase any product there is a no-purchase option denoted $i = 0$, with an associated utility U_0. According to utility based models, each customer when offered an assortment chooses the option with the highest utility in $S_0 = S \cup \{0\}$.

4.1.1 Multinomial Logit Model

The Multinomial logit (MNL) model is a discrete choice model that is commonly used in marketing and economics literature, and has recently been applied in assortment and inventory planning as well. MNL models use consumer preferences in order to compute the probability of each product, carried in the assortment, to be chosen by each customer visiting the store. The model has the appeal that even though it is stochastic in nature, it manages to capture decision variables in its formulation.

As mentioned before, each customer chooses the variant i with the highest utility U_i among the set of the available options, $i \in S_0$. The utility U_i can be decomposed into two parts, the deterministic component u_i and a random component ξ_i.

$$U_i = u_i + \xi_i$$

The ξ_i are assumed to be identical and independent Gumbel random variables with mean zero and scale parameter μ. A higher μ implies a higher degree of heterogeneity among the customers. This suggests that though the different customers have the same expected utility for a given product, the realized utility will be different. This may be due to the heterogeneity of preferences across customers or unobservable factors in the utility of the product.

The Gumbel distribution is closed under maximization. Hence, the probability that a customer chooses product $i \in S_0$ is

$$p_i(S) = \frac{e^{u_i/\mu}}{\sum_{k \epsilon S_0} e^{u_k/\mu}}$$

This closed form expression makes the MNL model an ideal method for modeling customer choice in analytical studies. Starting with the seminal work of Guadagni and Little (1983), an enormous number of marketing researchers found that MNL model is very useful in estimating demand for a group of products. In Sect. 4.4, we will briefly discuss the parameter estimation of MNL model, depending on the data available.

The most important criticism of MNL model originates in its Independence Irrelevant Alternatives (IIA) property. This property can be stated as follows: *The ratio of the probabilities of any two alternatives is independent of the choice set*. That is, for any choice sets C_1 and C_2 such that $C_1 \subseteq C_n$ and $C_2 \subseteq C_n$, and for any alternatives i and j in both C_1 and C_2, we have:

$$\frac{p(i|C_1)}{p(j|C_1)} = \frac{p(i|C_2)}{p(j|C_2)}$$

The IIA property of MNL models is a limitation for some practical applications. For example, IIA property does not hold when there are groups of alternatives in the choice set which are more similar to each other than with the rest of products. Consider an assortment consisting of two products from different brands. If brand loyalty is high, adding a new product can cannibalize the sales of the product with the same brand more than the sales of the alternative product. Unfortunately, IIA property does not capture this significant feature of consumer choice.

Various schemes have been proposed for dealing with the IIA property. The Nested Logit (NL) model introduced by Ben-Akiva and Lerman (1985) is one. Usually in NL models a two-stage nested process is used. Subsets of alternatives which are similar are grouped in hierarchies or nests, e.g. the choice set is partitioned into nests of SKUs from the same brand. The consumer chooses with a certain probability one of the nests, from which he then chooses a variant from that particular nest. In the NL model the IIA property no longer holds, when two alternatives do not belong to the same nest. However, different nest structures will produce different results and that leads to the main disadvantage of the NL models. Their use requires the knowledge of key attributes and their hierarchy for consumers and makes estimation problems more difficult (Kök et al. 2009).

4.1.2 Locational Demand Model

Lancaster (1966, 1975) proposed an extension of MNL model, the locational model of consumer choice behavior which is also utility-based. In this model, products are perceived as bundles of characteristics or attributes and individual preferences are defined on these characteristics rather than on the products themselves. We assume that each characteristic or attribute is quantifiable. We define the preference spectrum or the attribute space to be the space of all possible combinations of levels of attributes, where each point corresponds to a potential product location in the category. Each individual is characterized by an ideal point in the attributes space, which corresponds to his most preferred combination of attributes.

Suppose that there are m characteristics of a product. Let z_j denote the location of variant j in R^m. Consider a consumer whose ideal product is defined by $y \in R^m$. The utility of variant j for the consumer is

$$U_j = k - r_j - g(y, z_j)$$

where k is a positive constant, r_j is the price, and $g: R^m \to R$ is a distance function, representing the disutility associated with the distance from the consumer's ideal point. The consumer chooses the variant that gives him the maximum utility.

There is one major difference between the locational choice model and the MNL model. In the locational choice model, IIA property does not hold and substitution between products is localized to products that are close to each other in the attribute space. Hence, the firm can control the rate of substitution between products by selecting their locations to be far apart or close to each other.

4.2 Exogenous Demand Model

Exogenous demand (ED) models are the most common demand model in the literature of inventory management. ED model specify the demand directly for each product and what a consumer does when his preferred product is not available. Consumers choose from an item set $N = \{1\ 2, \ldots, i, \ldots, j, \ldots n\}$. If the preferred item j is not available, either because it is out-of-stock or because it is not included assortment at all, an individual consumer might accept another item i as a substitute according to a defined substitution probability μ_{ji}. The parameter μ_{ji} makes it possible to differentiate between items with varying substitution rates.

The substitution probability matrix for multiple items can be written as:

$$\mu_{ij} = \begin{bmatrix} 0 & \mu_{12} & \cdots \\ \mu_{21} & 0 & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}$$

In ED models the substitution probability depends on the type of substitution considered. For example, sometimes the substitution probability is equally distributed among the competing products, while in other cases the probability either is possible among neighboring products only or is considered to be a weighted ratio of the demand of the competing products.

4.3 Space Elastic Demand Models

Profit optimization models in shelf space planning formulate the demand rate as a function of the shelf space allocated to products. According to (Kök et al. 2009), this view seems especially relevant for fast moving products whose demand is sufficiently high that a significant amount of inventory is carried on the shelf.

Space models cluster around the following two streams: Own-space elastic demand models and Own- and cross-space elastic demand models.

4.3.1 Own-Space Elastic Demand Models

The demand function under this model is written as a function of the space allocated to an item. For example, if the item i's demand is denoted by d_i, the following applies:

$$d_i = a_i(k_i b_i)^{\beta_i}$$

The base demand is a_i, k_i, is the number of facings allocated to item i, b_i is the item's breadth, β_i the space elasticity expressed as a power function. In Sect. 3.3, we briefly discuss the empirical estimates of space elasticities and the fact that most of them vary between 0.15 and 0.8.

4.3.2 Own- and Cross-Space Elastic Demand Models

Some shelf space models deal with the effect of cross-space elasticity as it was proposed in the influential work of Corstjens and Doyle (1981). They formulate a demand function that additionally includes cross-space elasticity γ_{ji} as follows:

$$d_i = a_i(k_i b_i)^{\beta_i} \prod_{\substack{j=1 \\ i \neq 1}}^{I} (k_i b_i)^{\gamma_{ji}}$$

where $i = 1, 2, \ldots, I$ and the set of facings of product i is $k_k = 1, 2, \ldots, K_i$. The cross-space elasticity γ_{ji} quantifies the effects of neighboring items j on the sales of the item i.

Discussion on cross-space effects is ambiguous in the pertinent literature according to Hübner and Kuhn (2011a). On the one hand, Kök et al. (2009) argue that there is no empirical evidence that product level demand can be modeled with cross- space elasticity and on the other hand Zufryden (1986) describes that consideration of cross-elasticity at an individual level would be impossible in practice due to the overwhelming number of cross-elasticity terms that would need to be estimated.

4.4 Demand Estimation

This section is referring to the estimation of demand according to MNL and ED models. The estimation method depends on the data available.

4.4.1 Demand Estimation with MNL Model

Two different types of data are usually used in estimating the parameters of MNL model:

- Panel data

  Panel data present histories of purchases for a sample of households. The seminal work of Guadagni and Little (1983) was the first successful attempt to estimate the parameters of MNL model using panel data. After them, an enormous number of papers followed trying to understand the impact of marketing mix variables on demand of products (Guadagni and Little 2008). Consider the purchase decision of a household visiting a store at time t. The deterministic component of the utility u_{it} is specified as a linear function of m independent variables including constants specific to each product, price, promotion, brand loyalty of the household to the brand of product i (taken to be the exponentially weighted average of past purchases of the brand, treated as 0–1 variables) and finally size loyalty (which is analogous to brand loyalty).

  Let $x_{it} = (x_{it1}, x_{it2}, \ldots, x_{itm})$ denote the vector of these attributes for the household's shopping trip at time t, S_{0t} denote the assortment at time t including the no-purchase option, and $\beta = (\beta_1, \beta_2, \ldots, \beta_m)$ denote the vector of common coefficients.

$$u_{it} = \beta x_{it},\ i = 1, 2, \ldots, n$$

The outcome y_{it} of the choice of a household in time t is given by the panel data and is

$$y_{it} = \begin{cases} 1,\ \textit{if product i was chossen in time t} \\ 0,\ \textit{otherwise} \end{cases}$$

The β are unknown constants to be determined by calibration. Calibration is done by maximum likelihood, using the probability formulation of MNL and the above formulation of u_{it} to calculate the likelihood function. In order to obtain the maximum likelihood estimates (MLE) for the β, the log of the likelihood function is used by multiplying the probability of the observed choices across all t:

$$L(\beta) = \sum_t \sum_i y_{it} \left(\beta^T x_{it} - \ln \sum_{k \in S_{0t}} e^{\beta^T x_{kt}} \right)$$

The log-likelihood function is concave (McFadden 1974), therefore any nonlinear optimization technique can be used to find MLE estimate of β.

Fader and Hardie (1996) extended the work of Guadagni and Little and suggested an attribute based approach to estimate demand from panel data. Consumer utility for a product is expressed in terms of its attributes, dropping the product-specific dummy variables. The attribute-level utilities are used to determine choice probabilities (based on an MNL model) and the likelihood of observing the given transaction data. They use maximum likelihood to estimate the parameters of the model. Estimation is parsimonious as they only need to estimate attribute-level utilities, and their model has the added advantage of being able to estimate demand of new products.

- Sales transaction data
  Consider a demand process where consumer arrivals follow a Poisson process with rate λ and consumers select a product according to MNL model. If customer arrivals, purchases and no-purchase outcomes are completely observed, then one can estimate the demand parameters using maximum likelihood methods. However, in practice only purchases are observed, and hence it is not possible to distinguish between a period with no arrivals, and a period with arrivals but no purchases.
  Talluri and van Ryzin (2004) use sales transaction data (records of purchase time and product choice for each customer) to estimate demand in the context of airline revenue management. Sales transaction data is an incomplete data set because only the arrivals of customer that made a purchase are recorded. To overcome this problem Talluri and van Ryzin (2004) use the Expectation-Maximization (EM) algorithm to correct for the missing data. This method starts with arbitrary initial estimates of the demand parameters and uses Bayes rule to estimate the missing data. These estimates are now used to compute the conditional expected value of the likelihood (the expectation step), and the resulting expected log-likelihood function is maximized to generate new estimates for the demand parameters (the maximization step). This procedure is repeated until it converges.

4.4.2 Demand Estimation with Exogenous Demand Model

Here we review the methodology proposed by Kök and Fisher (2007) to estimate assortment based substitution rates using sales data. Kök and Fisher use an exogenous demand model to study a joint assortment selection and inventory planning problem in the presence of shelf-space constraints. They consider assortment-based substitution and assume that substitution demand accrues to available products in proportion to their original market share, as in the MNL model. They provide a process for estimating demand and substitution rates and apply their method to data from a large Dutch grocery retailer.

The data set they use includes the number of customers visiting each store each day, sales for each product each store each day and values of variables like price and promotion and other variables that influence demand such as weather, holidays. They use a three stage hierarchical model for demand estimation: (1) *purchase incidence*: whether or not to buy form a subcategory, (2) c*hoice*: which particular product in the subcategory to buy, (3) q*uantity*: how many units to buy. The demand for product i is

$$D_i = K\pi p_i q_i$$

where K is the number of customers visiting the store the given day, π is the probability of purchase incidence, p_i is the choice probability and q_i is the average quantity of units that a customer buys given the purchase incidence and choice of product i. The product choice p_i is modeled with MNL framework. It is computed as p_{iht}, where subscript h denote store index and t denote time index (day of store visit), from sales data as the ratio of number that bought product i to number of the customers that bought any product from the subcategory at store h on day t. It is computed as the probability of purchase history π_{ht} from sales data as the ratio of customers that bought any product form the given subcategory to the number of customers visited the store h on day t. Finally, they compute q_{iht} form sales data as the number of units of product i divided by the number of customers who bought product i at store h on day t. More details can be found in the paper.

Kök and Fisher also proposed a methodology to estimate substitution rates using sales data as well. Their methodology can be briefly explained as follows. They suppose that there are two types of stores: those that carry full assortments N and those that carry smaller assortments $S \subset N$ with 100 % service rate (i.e. no stock-out-based substitution). They observe D_i for product $i \in S$ from sales data and they also observe d_i for product $i \in N$ from sales data from a store that carries full assortment. They conclude that substitution exists if $\sum_{i \in S} D_i > \sum_{i \in S} d_i$ Finally, they find the substitution rate that minimize the total error between the theoretical and the observed sales across all available data from multiple stores and different time periods. More details of the methodology can be found in the paper.

The proposed methodology by Kök and Fisher has many advantages. First of all, the substitution rate estimated by the methodology for assortment based substitution can be also used for stock-out-based substitution. According to Campo et al. (2004), there are significant similarities in consumer reactions to a permanent assortment reduction and to stock-outs. Another advantage is that it enables us to estimate the demand rates of products in a store including those that have never been carried in that particular store.

5 Scientific Models for Retail Category Planning

This section is dedicated to scientific models that have been introduced in the field of retail category planning and, in particularly, to scientific models for assortment and shelf space planning. As we demonstrate in the following sections of the chapter, shelf space models have a deterministic nature, whereas assortment models are predominately stochastic. An extensive presentation of the related papers can be found in Appendix, sorted by the year of publication.

5.1 Scientific Models for Assortment Planning

In this section, we focus on the retailer's problem of choosing, from a set of N potential SKUs in a retail category, K SKUs to be carried at each store so as to maximize revenue or profit. We review scientific models and algorithms for demand forecasting and assortment optimization and demonstrate their use in practical applications. Assortment optimization research has been based on both stylized models intended to provide insight into structural properties of optimal assortments and decision support models intended to guide a manager planning retail assortments.

The main body of literature on assortment models focuses on assortment decisions at a single store. Most of the papers on assortment optimization include both the selection of items to stock (assortment selection) and deciding the inventory levels for each item in the assortment (inventory optimization). Table 3 summarizes the major assortment planning studies and structures them according to key characteristics. Specifically, the studies are grouped by the demand estimation model (Multinomial Logit, Exogenous Demand or Locational Choice model), the type of substitution (static or dynamic) and the solution method that has been used in each model, along with the average number of items used in the test cases. Moreover, the table column entitled "Model considerations" contains specific references to the type of solutions provided by each model, such as assortment, inventory and price optimization. Finally, the table in column "Model enhancements" exhibits other directions that each study has contributed to, such as consumer search cost and supplier selection.

In Sects. 5.1.1–5.1.3, we review papers categorized according to the demand model that they are based on. In Sect. 5.1.4, we review papers that deal with assortment planning in decentralized supply chains. Finally, Sect. 5.1.5 discusses dynamic assortment planning models in which the retailer updates his demand estimates every period and updates his assortment throughout the season.

5.1.1 Assortment Planning with MNL

Van Ryzin and Mahajan (1999), in their influential work, study an assortment planning approach with assortment-based substitution, where the demand derives from stochastic choice processes in which individual purchase decisions are made according to a Multinomial Logit (MNL) random utility model. Their objective is utility maximization in the setting of the Newsboy problem. They assume identical costs and prices for all items. Mahajan and van Ryzin (2001) study the same problem under dynamic substitution (assortment-based and stock-out-based substitution). By comparing the results of a stochastic gradient algorithm with two newsvendor heuristics, they conclude that the retailer should stock more of the more popular variants and less of the less popular variants than a traditional newsvendor analysis suggests.

Table 3 Overview in assortment planning models

	Authors (by year)	Demand model	Substi-tution	Model considerations			Model enhancements	Solution	
				Assortment opt/tion	Inventory decisions	Pricing opt/tion		Method	Experimental data[a]
1.				x	x		Identical prices and costs	Specialized heuristic	8
2.	Smith and Agrawal (2000)	ED	Static	x	x		Limited space	Lagrange relaxation, one-dimensional search	5
3.	Chong et al. (2001)	MNL	Static	x			Product similarities, limited space	Local improvement heuristic	47
4.	Mahajan and van Ryzin (2001)	MNL	Dyn.	x	x			Stochastic gradient algorithm	10
5.	Rajaram and Tang (2001)	ED	Dyn.	x			Order quantities	Service rate heuristic	7
6.	Agrawal and Smith (2003)	ED	Dyn.	x			Probability of purchase sets of items	0–1 mixed integer linear problem	5
7.	Cachon et al. (2005)	MNL	Static	x			Consumer search cost	Specialized heuristic	8
8.	Singh et al. (2005)	MNL	Dyn.	x	x		Different supply chain structures		
9.	Gaur and Honhon (2006)	LC	Static	x	x		Location	Specialized heuristic	5
10.	Caro and Gallien (2007)	DL	–	x				Greedy and Index policy heuristics	720
11.	Kök and Fisher (2007)	ED	Dyn.	x			Limited space	Iterative heuristic	29
12.	Li (2007)	MNL	Dyn.	x			Store traffic	Specialized heuristic	5
13.	Maddah and Bish (2007)	MNL	Static	x	x	x		Specialized heuristic	3

14.	Shah and Avittathur (2007)	ED	Dyn.	x	x			Specialized heuristic	3
15.	Hopp and Xu (2008)	MNL	Dyn.	x		x	Competition	Fluid network heuristic	3
16.	Aydin and Hausman (2009)	MNL	Static	x	x		Slotting fee, identical prices and costs		
17.	Smith (2009)	MNL	Static	x			Retailer choice, brand coverage and display constraints	Empiricalanalysis over the optimization	117
18.	Yücel et al. (2009)	ED	Static	x			Supplier selection, limited space	MIP model	10
19.	Honhon et al. (2010)	LC	Dyn.	x	x			Dynamic programming heuristic	5
20.	Miller et al. (2010)	MNL	Static	x			Attractiveness of competing retailers' assortments	Mixed Integer linear programming	117
21.	Sinha et al. (2012)	MNL	Dyn.	x		x	Perceptual maps	Genetic Algorithm	3000
22.	Rooderkerk et al. (2013)	ED	Static	x	x (EOQ)	x	Similarity effects	Neighborhood search heuristic	61
23.	Sauré and Assaf (2013)	DL		x			Limited space	No optimization takes place	

[a]Number of products in a product category
MNL Multinomial Logit, *ED* Exogenous demand, *LC* Locational demand, *DL* Demand learning

There was also many attempts to extend the van Ryzin and Mahajan model. For example, Cachon et al. (2005) study the van Ryzin and Mahajan model in the presence of consumer search. Maddah and Bish (2007) try to extend it by considering the pricing decisions as well. Li (2007), on the other hand, assumes that the demand for each product is proportional to a random store traffic volume and extends van Ryzin and Mahajan to allow for unequal unit prices and costs across variants.

There is also an interesting category of models that try to deal with the Independence of Irrelevant Alternatives (IIA) property. As it was indicated in the Sect. 4.1.1 paragraph, the major criticism of the MNL model stems from this property. For example, Chong et al. (2001) extend the classical Guadagni and Little (1983) model using a nested MNL model. They propose three new brand-width measures that capture the similarities and the differences among products within and across brands. Another example is the work of Sinha et al. (2012). They developed a category management tool that it was successfully implemented in a wine company (Beringer California Collection). Their model uses a mixed MNL model in order to estimate what they call as transferable and non-transferable demand (substitutable and non-transferable demand) of each product and use a Genetic algorithm to optimize the assortment.

Finally, we would like to refer to the works of Smith (2009) and Miller et al. (2010) who try to incorporate into their models the heterogeneous customer preferences using the notion of “consideration sets”. Smith (2009) develops an assortment optimization model that includes heterogeneous consumers′ store and product choices in the MNL framework. He indicates that under two extreme cases (the retailer is either a monopolist or a marginal player in the market) the model can be (approximately) solved via a simple ranking procedure. He also conducts an empirical analysis over the optimization of an assortment of DVD players, and argues that a substantial increase in expected profit may be achieved when customers are segmented according to consideration sets and the retailer uses this information correctly. Miller et al. (2010) consider the retailer’s assortment selection problem with heterogeneous customers and test the impact of different consumer choice models on the optimal assortment. They develop a sequential choice model in which customers first form Consideration Sets and then make product choices based on the MNL model.

5.1.2 Assortment Planning with Exogenous Demand Models

In this subsection, we review related assortment planning models that use exogenous demand (ED). One of first papers that introduce the use of ED models in assortment planning was the work of Smith and Agrawal (2000). They develop a probabilistic demand model that capture the effects of substitution and a methodology for selecting item inventory levels so as to maximize the total category profit. They use the binomial distribution to model the demand of each item, whereas the supply process is modeled as a Newsboy model.

Rajaram and Tang (2001) and Shah and Avittathur (2007) also use the basic Newsboy model into their formulation. Shah and Avittathur (2007) study multi-item inventories and propose a heuristic for assortment decision and stocking levels. Rajaram and Tanganalyze the impact of product substitution on retail merchandizing with a service-rate heuristic. They evaluate the impact of substitution on order quantities and expected profits and show that substitution reduces shortages and overstocking costs.

Kök and Fisher (2007), in their influential work, employ an exogenous stochastic model of substitution and develop an estimation methodology for the substitution rates using sales data from different stores with varying assortments. They propose an iterative optimization heuristic for the assortment planning and inventory problem with out-of-assortment (OOA) and out-of-stock (OOS) in the presence of constraints on shelf space, maximum inventory levels and order lead times. Shelves are divided into facings, but do not account for space elasticity.

Yücel et al. (2009) model an assortment and inventory problem under consumer-driven demand substitution that takes into account other issues such as supplier selection, shelf space constraints and poor quality procurement. They conclude that neglecting consumer substitution, excluding supplier selection or ignoring space limitations all have significant impact on the efficiency of retail assortments.

Finally, Rooderkerk et al. (2013) note that a product is a stronger substitute for similar products than it is for dissimilar ones. They call this phenomenon "the similarity effect" and argue that discrete choice models that do not account for similarity effects may lead to suboptimal assortment decisions. Attempting to address this challenge, they develop a methodology that consists of an attribute-based model of store-level SKU sales and utilize a very large neighborhood search heuristics to provide solutions to the resulting optimization problem.

5.1.3 Assortment Planning Under Locational Choice Demand Models

In this subsection we review the work of Gaur and Honhon (2006) who studied the assortment planning model under the locational choice demand. The authors consider both static and dynamic substitution. Under static substitution, they show that the distance between products in the optimal assortment is large enough so that there is no substitution between them. Under dynamic substitution, they derive lower and upper bounds on the optimal expected profit and propose two heuristics based on these bounds. In a numerical study using 3150 problem instances, they find that the average optimality gaps of the heuristics are 1.44 and 1.24 %. The optimality gap of each heuristic decreases as mean demand increases, as consumer preferences become more homogeneous, as the profit margin increases, and as the degree of substitutability of products increases.

5.1.4 Assortment Planning in Decentralized Supply Chains

Singh et al. (2005) and Aydin and Hausman (2009) are the only cases in the literature that incorporate supply chain considerations into assortment planning. In the first work, the authors present an analytical model that facilitates comparison of assortment and stocking decisions in alternative supply chain structures, and try to describe the critical influence of the structure of a supply chain on the depth of the optimal assortment chosen by the decision-maker in each supply chain. They investigate a traditional channel where the retailer stocks and owns the inventory and makes the assortment and stocking decisions and a drop-shipping channel where the wholesaler stocks and owns the inventory for multiple retailers and makes both the decisions. They demonstrate that the structure of the optimal assortment in both channels is given by a single threshold policy. A key result is that over a wide range of problem parameters, it is cost-efficient for the wholesaler in the drop-shipping channel to offer a larger assortment to her customers than the retailer in the traditional channel. Moreover, the assortment offered by the wholesaler in the drop-shipping channel increases with the number of retailers in the channel and her relative margin.

In the latter work (Aydin and Hausman 2009), the authors study a single-retailer, single-manufacturer supply chain, where the retailer decides what assortment to offer to end customers. They find that the retailer chooses a narrower assortment than the supply chain optimal assortment since her profit margins are lower than that of the centralized (vertically integrated) supply chain. The manufacturer can induce coordination by paying the retailer a per-product fee, resembling the slotting fees in the grocery industry, while making both parties more profitable.

5.1.5 Dynamic Assortment Planning with Demand Learning

Even though, retailers implicitly know that they should modify or adapt their assortments over time in response to environmental trends, all of the assortment planning papers reviewed in the previous sections consider static assortment planning problems and do not consider revising or changing assortment selection as time elapses.

One of the first attempts into this area was made by Caro and Gallien (2007). Caro and Gallien study a discrete-time finite-horizon problem using a multi-armed bandit formulation and Bayesian learning. At each time period, the retailer must decide the subset of products to offer based on historical sales data. This problem relates to the classical exploration versus exploitation trade-off. The firm must decide whether to optimize revenues based on the current information (exploitation), or try to learn more about the demand of products not in the assortment with the hope of identifying popular products (exploration). The authors propose a simple index policy based on a relaxation of the dynamic program. In most of this inventory-related research, however, pricing policies and their impact on revenues and demand learning are not investigated.

Sauré and Assaf (2013) also study a dynamic assortment planning problem in which an MNL demand model drives consumer choice with a capacity constraint. They present an assortment planning algorithm that simultaneously learns the underlying MNL parameters and optimizes profit, by balancing the tradeoff between exploration and exploitation.

5.2 *Scientific Models for Shelf Space Planning*

Managing shelf space is critical for retailers in order to attract customers and optimize profit. The decisions relating to the products to be stocked among a large number of competing products and the amount of shelf space to allocate to those products is a question central to retailing. As shelf space is a scarce and fixed resource, and the number of potentially available products continually increases, retailers have a high incentive to make these decisions correctly.

Table 4 summarizes the major shelf space studies and structures them according to key characteristics. Specifically, the studies are grouped by the type of demand model (space elastic and/or price elastic demand) that has been used in each model, as well as other demand effects such as marketing and substitution effects. Moreover, the table contains specific references to the type of costs taking into account by each model (e.g. listing, ordering, and replenishment costs). Finally, the studies are grouped by the solution method that has been used in each model, along with the average number of items used in the test cases.

Shelf space models cluster around the following three streams: shelf space allocation models with own space-elastic demand (Sect. 5.2.1), shelf space allocation models with own space- and cross-space elastic demand (Sect. 5.2.2) and integrated shelf space allocation and inventory planning models (Sect. 5.2.3).

5.2.1 Shelf Space Allocation Models with Own Space-Elastic Demand

For the first stream of space allocation models, an influential study was published by Hansen and Heinsbroek (1979). They formulate the total category profit as a maximization problem with the number of facings as a decision variable. The authors also incorporate into their model cost and demand effects, but do not reflect cross-product relations. Further developments in this stream are published by the following researchers. Zufryden (1986) suggests the use of dynamic programming to solve a problem of shelf-space allocation with space elasticity and marketing variables. His formulation allows the consideration of general objective-function specifications and provides integer solutions.

An interesting paper by Yang and Chen (1999) assumes linear profit as long as the number of facings for each product is between some lower and upper bounds. They formulate their shelf space allocation problem with vertical and horizontal space allocation effects and, although there are no cross-effects, this is the first

Table 4 Overview of shelf space planning models

		Demand model					Solution	
	Authors (by year)	Space elasticity	Price elasticity	Other demand effects	Cost	Facings decisions	Method	Experimental data[a]
1	Hansen and Heinsbroek (1979)	Osp			Replenishment, Out-of-stock constraints	Total-length	Lagrange, Specialized heuristic	6443
2	Corstjens and Doyle (1981)	Osp, CSp			Procurement, Inventory, OOS	Number	Geometrical programming	5
3	Zufryden (1986)	OSp		Marketing effects	Procurement, Inventory, OOS	Total-length	Dynamic programming	40
4	Bultez et al. (1989)	Osp, CSp		Promotion, visibility	Restocking	Number	Specialized heuristic	4
5	Borin et al. (1994)	CSp		Assortment		Number	Simulated annealing	18
6	Urban (1998)	OSp, CSp		Assortment	Inventory, Order quantity	Number	GRG, GA	54
7	Yang (2001)	Osp				Number	Knapsack algorithm	10
8	Lim et al. (2004)	OSp, CSp		Product associations	Procurement, Inventory	Number	Tabu search	100
9	Hwang et al. (2005)	OSp		Positioning	Procurement, Inventory, Order quantity	Number	Gradient search heuristic, GA	4
10	Hariga et al. (2007)	Osp, CSp		Displayed quantity, Assortment	Inventory, Ordering	Number	MINLP model (LINGO)	4
11	Reyes and Frazier (2007)	OSp	OPr			Number	Non-linear Integer Goal programming	4
12	Abbott and Palekar (2008)	Osp, CSp		Shelf depletion	Restocking	Number	Upper bound heuristic	4
13	Bai et al. (2008)	OSp		Positioning	Product surplus, Ordering, Inventory	Number	Greedy algorithms, Meta-heuristic	5
14	Gajjar and Adil (2008)	OSp		Positioning		Number	Specialized heuristic	200

15	Ramaseshan et al. (2008)	OSp, CSp			Procurement, Shelf space, Ordering	Number	GRG	14
16	Hansen et al. (2010)	OSp, CSp		Positioning		Number	Meta-heuristics, simulation	100
17	Murray et al. (2010)	OSp, CSp	OPr, CPr	Positioning		Number[b]	MINLP model (BONMIN)	100
18	Gajjar and Adil 2011	OSp				Number	Local search heuristic	200
19	Hübner and Kuhn (2011a)	OSp, CSp		Substitution	Listing, Inventory, Replenishment	Number	MIP model (CPLEX)	250
20	Hübner and Kuhn (2011b)	OSp, CSp		Substitution	Listing, Overstocking, Undersupply	Number	MIP model (CPLEX)	25
21	Irion et al. (2012)	OSp, CSp	OPr		Inventory, Replenishment	Number	MIP model (LINGO)	6
22	Tsao et al. (2014)	OSp, CSp		Promotion	Inventory, Replenishment	Number	MIP model (CPLEX)	6

[a]Number of products in an product category
[b]Number of facings in different orientations
OSp Own space elasticity, *CSp* Cross space elasticity, *OPr* Own price elasticity, *CPr* Cross price elasticity

shelf-space model to include location effects. Yang (2001) proposes a knapsack heuristic for this model.

His heuristic allocates shelf space with respect to a descending order of sales profit for each item per display length, but finds an optimal solution only for simplified versions. Hwang et al. (2005) assume that the display level significantly influences the sales of products and propose an integrated mathematical model, which combines the shelf space allocation model and inventory-control model with the objective of maximizing the retailer's profit. Due to the complexity of the integrated model, Hwang et al. proposed a gradient search heuristic and a genetic algorithm to resolve the model.

Two non-linear models are provided by Bai and Kendal (2005) and Bai et al. (2008), where demand is a function of the amount of displayed inventory and the space elasticity of the product. Gajjar and Adil (2008) try to propose an alternate linear model for this non-linear model by Bai and Kendal (2005), using piecewise linearization. They obtained the tight upper bounds using linear programming relaxation in their linear model. Gajjar and Adil (2011) develop a local search heuristic which obtained better results compared to greedy heuristics, simulated annealing heuristics and heuristics proposed by Bai et al. (2008). Another non-linear problem was proposed by Hansen et al. (2010). They present a retail shelf-space decision model that incorporates a nonlinear profit function, vertical and horizontal effects, and product cross-elasticity. They investigate metaheuristics for decision models with facing-dependent demand and vertical and horizontal location effects.

Reyes and Frazier (2007) propose a shelf-space allocation model that considers the tradeoff between profitability and customer service level under the nonlinear demand of price and shelf space allocation. Finally, Murray et al. (2010) jointly optimize a retailer's decisions for product prices, display facing areas, display orientations, and shelf-space locations in a product category. They solve the resulting optimization problem with an MINLP-solve.

5.2.2 Shelf Space Allocation Models with Own Space- and Cross-Space Elastic Demand

In their work, Corstjens and Doyle (1981) formulate a maximization problem based on individual and cross-product demand. They suggest a method for allocating shelf space to categories and include cross-space elasticity in the demand calculation. Their signomial geometric programming method optimizes shelf space for categories, but Borin et al. (1994) show that the reported solutions violate the constraints in seven out of ten cases. In addition, the multiplicative model predicts zero demand for a given category if the space of any other category is set to zero. Bultez et al. (1989) apply Corstjens' model at brand level, assuming identical cross-elasticity within product groups.

Borin et al. (1994) extend the demand function of Corstjens and Doyle (1981) by allowing simultaneous decisions about assortment selections and shelf space allocations. They describe unmodified, modified, acquired and stock-out demand. Modified demand arouses from in store attractiveness of products. Acquired demand captures assortment decisions. Stock-out demand arises when total demand exceeds shelf inventory and consumers switch to other items. They assume constant market size and distribute the volume of delisted items according to the market shares of remaining items. The resulting model optimizes return on inventory and is solved using simulated annealing heuristic procedures.

Irion et al. (2011) further extend the Corstjens and Doyle's model. They consider purchasing costs, interest rates and listing costs. The model chooses integer facings to optimize profit under shelf space and facing constraints. They propose a piece-wise linearization technique for approximating the complicated nonlinear model of shelf space allocation problem that relaxed the non-convex optimization problem into a linear Mixed Integer Program (MIP). Proposed MIP not only generated near optimal solutions, but also provided a posteriori error bound to evaluate the quality of the solution.

Finally, Lim et al. (2004) build on Yang and Chen (1999) model by optimizing with meta-heuristics. They present two extensions to this model. First, they consider product groupings, where cross-product affinity is modeled as a linear combination of the Yang and Chen (1999) profit function and an additional profit term. Second, they consider a general nonlinear profit function.

5.2.3 Integrated Shelf Space Allocation and Inventory Planning Models

Although some existing space allocation models (e.g. Borin et al. 1994) use return on inventory as the objective and take stock-outs into consideration, they do not explicitly include the conventional inventory control decisions as variables. Urban (1998) provides the first enhancement with available inventory and replenishment systems by integrated the inventory control model with the product assortment and space allocation models. In addition, a greedy search and a genetic algorithm were developed to resolve the integrated model. Inventory control integration with space allocation is proposed by Hwang et al. (2005). They optimize order quantity, facings and shelf positioning using a generalized reduced gradient algorithm and a genetic algorithm, but exclude assortment decisions.

Abbott and Palekar (2008) develop retail replenishment models considering a linear function of sale rate to effective shelf space. They determine the optimal replenishment cycles for products given the costs of restocking and sales effects of inventory-elastic demand. They present exact and approximate solutions for single- and multi-product cases by formulating an economic order quantity problem with time-varying. The model does not optimize assortment and facing because it requires an initial space assignment as input.

Hariga et al. (2007) develop an integrated model that consider product assortment, self-space allocation, and inventory replenishment problems. The decision variables are display locations, order quantities and the number of facings in each display area. The non-linear problem could be solved exactly for a four-item case, but requires a heuristic for a larger, practical case. They also omit integer facing values.

Ramaseshan et al. (2008) determine a heuristic for shelf space allocation, product assortment, and inventory quantities. Their model uses the generalized reduced gradient algorithm to generate an approximate solution for up to 14 items.

Hübner and Kuhn (2011a, b) extend traditional shelf space models by integrating out-of-assortment substitution effects and replenishment constraints. The proposed models take into account facing-dependent restocking constraints and manage to reflect the actual retailers' replenishment policy. A transformation of the MINLP into a specialized knapsack problem is proposed and the model is capable of dealing with large, realistic category sizes with up to 250 items using CPLEX. Furthermore, the models reflect basic supply levels ensuring appropriate service levels with limited replenishment capacity.

Tsao et al. (2014), motivated by the significant influence of trade allowance on the efficiency of shelf- space allocation, develop a model for optimizing category-level shelf-space management. A multi-player Retailer Stackelberg game is introduced to model the interaction between retailer and manufacturers. With this framework, a retailer maximizes profit by taking the manufacturers' trade allowance response into account, which provides a realistic approach of simultaneously determining the shelf-space and trade promotion decisions under the consideration of interactions among products. A piecewise linearization method is employed to reformulate the Mixed Integer Nonlinear Programming (MINLP) problem into a linear MIP problem for optimal solutions.

6 Conclusions

This paper presented an updated review on topics related to RCM, such as scientific models that deal with assortment and shelf space planning. The complexity of these planning tasks has grown tremendously in recent years due to product proliferation and various consumer choice effects in the retail environment. It is an increasingly difficult task for category managers to find an effective assortment due to consumer preferences instability and the exponential number of possible assortments. We showed that, despite longstanding recognition of its importance, no dominant methodology for RCM exists and scientific models address only some of the factors that make assortment and shelf space planning so challenging.

Shelf space allocation models do not clearly and comprehensively address assortment selection, neglect substitution effects between products and ignore the stochastic nature of demand. The generally proposed method of handling assortment decisions for non-listed items is to assume no demand (no assortment-based substitution). Assortment planning models, on the other hand, mostly ignore shelf space constraints and neglect space depend demand. Even though, the studies have contributed to various directions, such as supplier selection and purchasing sets of items, only three studies include limited space constraints. Moreover, none of the studies analyzed take into account the space elasticity effects of a higher number of facings.

In summary, academic models should endeavor to provide more valuable insights and directions to practitioners for category planning. The industry could benefit greatly from scientific models that provide insights into which factors prevail and which have less importance. For example, more empirical work is needed in understanding the impact of different variables in store choice and purchasing behavior. Other demand effects based on empirical findings (e.g. variety perceptions or other marketing variables) must be incorporated in the future models. Finally, the problems of assortment selection, space allocation, pricing and replenishment are interrelated. Consequently, in the future the focus must be on integrated models that reflect these interrelated issues.

In the future, we strongly believe that retailing will continue to evolve as a wonderful new application domain in Operational Research. Particularly in RCM, it seems to us that academic research could make a tremendous difference in retailing by contributing in this area, much as it has done in other areas like finance, marketing and strategy (Kök et al. 2009). According to Fisher (2009) "retailing is an industry in transition from decision making almost exclusively based on art to a well-balanced blend of art and science". Retailing will always be a mix of art and science, with the relative proportions depending on the nature of the product category. For example, there is a greater tendency for apparel to depend on art than there is for other categories like food of hardware. Over the last decade, the amount of science in this mix has been steadily increasing. Many successful analytic software companies have been emerged, some started with the help of academics and based on their research. Retailers care most about efficiency and profitability. So, even though they are by nature suspicious of scientific implementations, any use of science that can demonstrate a significant improvement in results will get a serious look from them.

References

Abbott H, Palekar US (2008) Retail replenishment models with display-space elastic demand. Eur J Oper Res 186(2):586–607

Agrawal N, Smith SA (2003) Optimal retail assortments for substitutable items purchased in sets. Naval Res Logist 50(7):793–822

Amine A, Cadenat S (2003) Efficient retailer assortment: a consumer choice evaluation perspective. Int J Retail Distrib Manag 31(10):486–497

Aydin G, Hausman WH (2009) The role of slotting fees in the coordination of assortment decisions. Prod Oper Manag 18(6):635–652

Bai R, Kendal G (2005) An investigation of automated planograms using a simulated annealing based hyper-heuristic. In: Metaheuristics: progress as real problem solvers. Springer, New York, NY, pp 87–108

Bai R, Burke EK, Kendal G (2008) Heuristic, meta-heuristic and hyper-heuristic approaches for fresh produce inventory control and shelf space allocation. J Oper Res Soc 59:1387–1397

Ball D (2004) Consumer-goods firms duel for shelf space. Wall Str. J, October 22, pp. B2

Ben-Akiva ME, Lerman SR (1985) Discrete choice analysis: theory and application to travel demand. MIT Press, Cambridge, MA
Blattberg RC (1995) Category management, guides 1–5. Food Marketing Institute, Washington, DC
Borin N, Farris PW, Freeland JR (1994) A model for determining retail product category assortment and shelf space allocation. Decis Sci 25(3):359–384
Broniarczyk SM, Hoyer WD (2010) Retail assortment: more ≠ better. In: Retailing in the 21st century. Springer, Berlin, pp 271–284
Broniarczyk SM, Hoyer WD, McAlister L (1998) Consumers' perceptions of the assortment offered in a grocery category: the impact of item reduction. J Mark Res 35(2):166–176
Bultez A, Gijsbrechts E, Naert P, Abeele PV (1989) Asymmetric cannibalism in retail assortments. J Retail 65(2):153
Cachon G, Terwiesch C, Xu Y (2005) Retail assortment planning in the presence of consumer search. Manuf Serv Oper Manag 7(4):330–346
Campo K, Gijsbrechts E, Nisol P (2000) Towards understanding consumer response to stock-outs. J Retail 76(2):219–242
Campo K, Gijsbrechts E, Nisol P (2004) Dynamics in consumer response to product unavailability: do stock-out reactions signal response to permanent assortment reductions? J Bus Res 57(8):834–843
Caro F, Gallien J (2007) Dynamic assortment with demand learning for seasonal consumer goods. Manag Sci 53(2):276–292
Chandon P, Hutchinson JW, Bradlow ET, Young SH (2009) Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. J Mark 73(6):1–17
Chernev A (2003) When more is less and less is more: the role of ideal point availability and assortment in consumer choice. J Consum Res 30(2):170–183
Chong J-K, Ho T-H, Tang CS (2001) A modeling framework for category assortment planning. Manuf Serv Oper Manag 3(3):191–210
Corstjens M, Doyle P (1981) A model for optimizing retail space allocations. Manag Sci 27(7): 822–833
Curhan RC (1972) The relationship between shelf space and unit sales in supermarkets. J Mark Res 9(4):406–412
Desmet P, Renaudin V (1998) Estimation of product category sales responsiveness to allocated shelf space. Int J Res Mark 15(5):443–457
Dhar R (1997) Consumer preference for a no-choice option. J Consum Res 24(2):215–231
Drèze X, Hoch SJ, Purk ME (1994) Shelf management and space elasticity. J Retail 70(4):301–326
ECR Europe (2003) ECR-optimal shelf availability: increasing shopper satisfaction at the moment of truth. ECR Europe, Helsinki
Emmelhainz MA, Stock JR, Emmelhainz LW (1991) Consumer responses to retail stock-outs. J Retail 67(2):138–147
Fader PS, Hardie BGS (1996) Modeling consumer choice among SKUs. J Mark Res 33(4): 442–452
Fisher ML (2009) Or FORUM-rocket science retailing: the 2006 Philip McCord Morse lecture. Oper Res 57(3):527–540
Fitzsimons GJ (2000) Consumer response to stockouts. J Consum Res 27(2):249–266
FMI (1995) Efficient Consumer Response Report. Food Marketing Institute. Washington DC
Gajjar HK, Adil GK (2008) A piecewise linearization for retail shelf space allocation problem and a local search heuristic. Ann Oper Res 179(1):149–167
Gajjar HK, Adil GK (2011) Heuristics for retail shelf space allocation problem with linear profit function. Int J Retail Distrib Manag 39(2):144–155
Gaur V, Honhon D (2006) Assortment planning and inventory decisions under a locational choice model. Manag Sci 52(10):1528–1543

Ge X, Messinger PR, Li J (2009) Influence of soldout products on consumer choice. J Retail 85(3):274–287
Greenhouse S (2005) How Costco became the Anti-Wal-Mart. New York Times
Gruen TW, Shah RH (2000) Determinants and outcomes of plan objectivity and implementation in category management relationships. J Retail 76(4):483–510
Gruen TW, Corsten DS, Bharadwaj S (2002) Retail out-of-stocks: a worldwide examination of extent, causes and consumer responses. Grocery Manufacturers of America, Washington, DC
Guadagni PM, Little JDC (1983) A logit model of brand choice calibrated on scanner data. Mark Sci 2(3):203–238
Guadagni PM, Little JDC (2008) Commentary—a logit model of brand choice calibrated on scanner data: a 25th anniversary perspective. Mark Sci 27(1):26–28
Hansen P, Heinsbroek H (1979) Product selection and space allocation in supermarkets. Eur J Oper Res 3(6):474–484. doi:10.1016/0377-2217(79)90030-4
Hansen JM, Raut S, Swami S (2010) Retail shelf allocation: a comparative analysis of heuristic and meta-heuristic approaches. J Retail 86(1):94–105
Hariga MA, Al-Ahmari A, Mohamed A-RA (2007) A joint optimisation model for inventory replenishment, product assortment, shelf space and display area allocation decisions. Eur J Oper Res 181(1):239–251
Hoch SJ, Bradlow ET, Wansink B (1999) The variety of an assortment. Mark Sci 18(4):527–546
Honhon D, Gaur V, Seshadri S (2010) Assortment planning and inventory decisions under stockout-based substitution. Oper Res 58(5):1364–1379
Hopp W, Xu X (2008) A static approximation for dynamic demand substitution with applications in a competitive market. Oper Res 56(3):630–645
Hübner AH (2011) Retail category management: decision support systems for assortment, shelf space, inventory and price planning. Springer, Berlin
Hübner AH, Kuhn H (2011a) Retail shelf space management model with space-elastic demand and consumer-driven substitution effects. Technical Report, KU Eichstatt-Ingolstadt, Ingolstadt
Hübner AH, Kuhn H (2011b) Shelf and inventory management with space-elastic demand. In Hu B, Morasch K, Pickl S, Siegle M (eds) Operations Research Proceedings 2010. Springer, Berlin, pp 405–410
Hübner AH, Kuhn H (2012) Retail category management: state-of-the-art review of quantitative research and software applications in assortment and shelf space management. Omega 40(2): 199–209
Hwang H, Choi B, Lee M-J (2005) A model for shelf space allocation and inventory control considering location and inventory level effects on demand. Int J Prod Econ 97(2):185–195
Irion J, Lu J-C, Al-Khayyal F, Tsao Y-C (2011) A hierarchical decomposition approach to retail shelf space management and assortment decisions. J Oper Res Soc 62:1861–1870
Irion J, Lu J-C, Al-Khayyal F, Tsao Y-C (2012) A piecewise linearization framework for retail shelf space management models. Eur J Oper Res 222(1):122–136
Iyengar SS, Lepper MR (2000) When choice is demotivating: can one desire too much of a good thing? J Pers Soc Psychol 79(6):995–1006
Kök GA, Fisher ML (2007) Demand estimation and assortment optimization under substitution: methodology and application. Oper Res 55(6):1001–1021
Kök GA, Fisher ML, Vaidyanathan R (2009) Assortment planning: review of literature and industry practice. In: Retail supply chain management. Springer, New York, NY, pp 99–153
Krafft M, Mantrala MK (eds) (2010) Retailing in the 21st century. Springer, Berlin
Lancaster K (1966) A new approach to consumer theory. J Polit Econ 74:132–157
Lancaster K (1975) Socially optimal product differentiation. Am Econ Rev 65:567–585
Li Z (2007) A single-period assortment optimization model. Prod Oper Manag 16(3):369–380
Lim A, Rodrigues B, Zhang X (2004) Metaheuristics with local search techniques for retail shelf-space optimization. Manag Sci 50(1):117–131
Maddah B, Bish EK (2007) Joint pricing, assortment, and inventory decisions for a retailer's product line. Nav Res Logist 54(3):315–330

Mahajan S, van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. Oper Res 49(3):334–351

Mantrala MK, Levy M, Kahn BE, Fox EJ, Gaidarev P, Dankworth B, Shah D (2009) Why is assortment planning so difficult for retailers? A framework and research agenda. J Retail 85(1):71–83

McFadden D (1974) Conditional logit analysis of qualitative choice behaviour. In: Frontiers of econometrics. Academic Press, New York, NY, pp 105–142

Miller CM, Smith SA, McIntyre SH, Achabal DD (2010) Optimizing and evaluating retail assortments for infrequently purchased products. J Retail 86(2):159–171

Morales A, Kahn BE, McAlister L, Broniarczyk SM (2005) Perceptions of assortment variety: the effects of congruency between consumers' internal and retailers' external organization. J Retail 81(2):159–169

Murray CC, Talukdar D, Gosavi A (2010) Joint optimization of product price, display orientation and shelf-space allocation in retail category management. J Retail 86(2):125–136

Rajaram K, Tang C (2001) The impact of product substitution on retail merchandising. Eur J Oper Res 135(3):582–601

Ramaseshan B, Achuthan NR, Collinson R (2008) Decision support tool for retail shelf space optimazation. Int J Inf Technol Decis Mak 07(03):547–565

Reyes PM, Frazier GV (2007) Goal programming model for grocery shelf space allocation. Eur J Oper Res 181(2):634–644

Rooderkerk RP, van Heerde HJ, Bijmolt THA (2013) Optimizing retail assortments. Mark Sci 32(5):699–715

Sauré D, Assaf Z (2013) Optimal dynamic assortment planning with demand learning. Manuf Serv Oper Manag 15(3):387–404

Shah J, Avittathur B (2007) The retailer multi-item inventory problem with demand cannibalization and substitution. Int J Prod Econ 106(1):104–114

Simonson I (1999) The effect of product assortment on buyer preferences44The paper has benefited from the comments of Ran Kivetz. J Retail 75(3):347–370

Singh P, Groenevelt H, Rudi N (2005) Product variety and supply chain structures. Working Paper, University of Rochester, Rochester

Sinha A, Sahgal A, Mathur S (2012) Category Optimizer™: a dynamic-assortment, new-product-introduction, mix-optimization, and demand-planning system. Mark Sci 32(2):221–228

Sloot LM, Verhoef PC, Franses PH (2005) The impact of brand equity and the hedonic level of products on consumer stock-out reactions. J Retail 81(1):15–34

Smith SA (2009) Optimizing retail assortments for diverse customer preferences. In: Retail supply chain management, vol 122. Springer, New York, NY, pp 183–205

Smith SA, Agrawal N (2000) Management of multi-item retail inventory systems with demand substitution. Oper Res 48:50–64

Talluri K, van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. Manag Sci 50(1):15–33

Tsao Y-C, Lu J-C, An N, Al-Khayyal F, Lu RW, Han G (2014) Retailer shelf-space management with trade allowance: a Stackelberg game between retailer and manufacturers. Int J Prod Econ 148:133–144

Urban TL (1998) An inventory-theoretic approach to product assortment and shelf-space allocation. J Retail 74(1):15–35

Van Ryzin G, Mahajan S (1999) On the relationship between inventory costs and variety benefits in retail assortments. Manag Sci 45(11):1496–1509

van Woensel T, van Donselaar K, Broekmeulen R, Fransoo J (2007) Consumer responses to shelf out-of-stocks of perishable products. Int J Phys Distrib Logist Manag 37(9):704–718

Verbeke W, Farris PW, Thurik R (2013) Consumer response to the preferred brand out-of-stock situation. Eur J Mark 32(11/12):1008–1028

Walter CK, Grabner JR (1975) Stockout cost models: empirical tests in a retail situation. J Mark 39(3):56–60

Yang M-H (2001) An efficient algorithm to allocate shelf space. Eur J Oper Res 131(1):107–118
Yang MH, Chen WC (1999) Study on shelf space allocation and management. Int J Prod Econ 60: 309–317
Yücel E, Karaesmen F, Salman FS, Türkay M (2009) Optimizing product assortment under customer-driven demand substitution. Eur J Oper Res 199(3):759–768
Zinn W, Liu PC (2001) Consumer response to retail stockouts. J Bus Logist 22(1):49–71
Zufryden FS (1986) A dynamic programming approach for product selection and supermarket shelf-space allocation. J Oper Res Soc 37(4):413–422

Cultural and Creative Industries Innovation Strategies for New Service Development Using MCDA

Fotis Kitsios, Eleni Champipi, and Evangelos Grigoroudis

Abstract The continuous development of new services is a prerequisite for the development and prosperity of Museums in today's competitive environment. So far, however, limited research effort has been devoted on the development process of innovative services within cultural and creative industries and particularly Museums. Given the crucial importance of cultural organisations for the dynamic cultural and creative industries, this research work aims at studying the development of new innovative services in the Museums of Athens and their ability and capacity to adopt innovation management (i.e., development, organisational, functional and technological aspects of management) as the basic strategy for diversification. To achieve this task, a systematic desk-top research took place in order to identify and analyse models, practices and processes available on the design and development of new services. Based on the results of this work, a survey was designed aiming to identify the current status with regards to innovative services development and innovation management within Athenian Museums. The fieldwork was carried out in 62 museums in Athens and investigated the development process of 184 different services. The analysis was based on data collected via an in depth structured interview with questionnaires from museum directors knowledgeable about new service development in their organisation. A multicriteria methodology was used to examine the potential of a predictive model for successful new service development projects in the cultural industry.

F. Kitsios (✉)
Department of Applied Informatics, School of Information Sciences, University of Macedonia, Thessaloniki, Greece
e-mail: kitsios@uom.gr

E. Champipi
Cultural Organizations Management, Hellenic Open University, Patras, Greece
e-mail: lenahabipi@yahoo.gr

E. Grigoroudis
Decision Support Systems Laboratory, Technical University of Crete, Chania, Greece
e-mail: vangelis@ergasya.tuc.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_4

Keywords New service development • Innovation management • Strategy • Cultural industry • Multicriteria decision analysis

1 Introduction

The cultural heritage is understood as a double entity of financial and cultural capital. It constitutes a tradable commodity, which can be sold to large public audience groups. As cultural capital, it is connected with both the preservation of old items and their messages, and also with their contribution to education and civilisation. As financial capital, cultural heritage contributes to the financial growth, becomes a central part of local communities and is incorporated in lifestyles that serve a new order of financial things. It also takes part into the development of the tourist industry and specific parts of it, such as cultural tourism. The cultural heritage is used to create directly "saleable" tourist products and to offer new jobs.

Museums, taking part into the modern cultural industry, have been reformed both internally, with changes and reclassifications, and externally, in shape and type. The changes have been made in public organisations, as well as in the private sector. Museums are considered as non-profitable organisations, where social targets prevail (education, maintenance, research, etc.). Nevertheless, they have similarities with organisations whose target is profit, since they follow commercial and financial targets by offering visitors an alternative activity of "leisure time industry".

The commercialisation of museums, especially the public ones, is an issue that is under criticism from various scientific and also social groups. It is, however, required from them to be economically viable as private businesses, to adopt marketing techniques, and to develop an active approach towards the public, both to their sponsors and their competitors, if they want to "survive" in a period of time that the public and private resources are constantly disappearing.

The constant development of new services is a prerequisite for the growth and the welfare of Museums in the modern competitive business environment. For this reason, it is considered as a complex and difficult work. Until today, the research work which has been done relating to the development process of new services in the museums is limited.

Taking into consideration the crucial importance of cultural organisations as a dynamic cultural industry, the purpose of this research is to study the development process of new services in Athens Museums and their competence in the adoption of management of innovation (productive, developmental, organisational, operational, technological) as a basic strategy of their differentiation.

In this framework, some scholars (Camarero and Garrido 2008, 2009, 2012; Bakhshi and Throsby 2010; Camison et al. 2009; Barczak et al. 2006; Castañer and Cavotta 2010; Castañer 2014) attempted to investigate the financial performance through the development of new services at Museums (financial results, revenues from funding and sponsorships), the impact of organisational size on novelties and their business management, on the technology and creating value, as well as on the social performance that a cultural mission achieves. This way they tried to encourage cultural institutes to be directed to the market, services and innovation. The present study, however, is the first empirical research of the success factors in New Service Development (NSD), studying specifically the cultural and creative industries, as well as the first one that analyses the innovation process in NSD using multicriteria methodologies.

The rest of the paper is organised as follows: Section 2 reviews studies of success factors, which are critical in service innovation, and discusses the relevant empirical research. Section 3 describes the data and methods used in the analysis, whereas Sect. 4 presents the obtained results. Finally, Sect. 5 concludes the paper and proposes some future research directions.

2 Literature Review

Services are a key pillar of the economy of every country today. NSD and service innovation are key factors for profitability and competitiveness of businesses today. But several studies indicate the limited research that has been conducted in this field. These researches are related mainly in financial services and in recent years researchers focused on tourism, due to its continuous increasing importance. The academic community and businesses have focused more on these research fields in recent years.

Studies researchers started from 1980. After investigating service characteristics which distinguish them from products, researchers turned their attention to identifying the stages necessary to develop successful services and define the success factors of new and innovative services.

Empirical research on new product development goes back to the 1970s, initially focusing on factors that influence the success of new tangible products. However, there are significant differences between products and services that distinguish New Product Development (NPD) from NSD: lack of tangibility, simultaneous production and purchase, variability and corruptibility (Kotler 2000; Reicheld and Sasser 1990). Thus, in later research efforts, the success factors for service development were approached by separate and specialised studies that used methodologies varying from direct matched pairs to multivariate tools and techniques.

On the early stages, research on NSD discriminated successful from unsuccessful services (Myers and Marquis 1969; Rothwell et al. 1974), as these two categories were approached separately from each other. Later researches introduced control groups that allowed the discrimination between success and failure, as well as the simultaneous comparison between multiple successful and unsuccessful services (Cooper and Kleinschmidt 1987; De Brentani 1986; Edgett and Parkinson 1994; Maidique and Zirger 1984). The comparison between success and failure in services was also the subject of studies that were performed during the 90’s and

00's. These particular studies tried to identify the factors that lead to success in NSD (Flikkema 2008; Van Riel et al. 2004; De Jong and Vermeulen 2003; Buisson et al. 1997; Martin and Horne 1995). The majority of these studies use standard statistical methods and multivariate analysis. However, there are very limited studies that analyse data from cultural and creative industries aiming at a predictive model and even fewer ones that specialise and focus on Greek industry.

The comparison method has also been used with good results in a number of studies on new services. For example, Edgett and Parkinson (1994) used it to compare new service development in British building societies that were registered and maintained active membership status in the Building Societies Association. Edgett and Parkinson (1994) used it to compare new service development activities in UK banks and building societies. In a study of new commercial service companies Cooper and de Brentani (1991) compared successful and unsuccessful services in a way similar to the methodology previously used by Cooper (Edgett and Parkinson 1994; Ernst 2002).

The success of the comparative methodology for tangible new product studies earlier and for new services later, indicated that this approach would be suitable and reliable for this study. A useful framework has been provided by these studies for similar work in a service setting, as each attempt has identified new product development characteristics that effectively discriminate between successful and unsuccessful new products (Edgett and Parkinson 1994).

Ottenbacher et al. (2006) applied multivariate statistics techniques (i.e., discriminant analysis) and identified that seven factors play a distinctive role in the outcome of high contact NSD in hospitality services. Also, in another survey he determined nine factors that promote successful service innovations (Ottenbacher and Gnoth 2005). In this context, Angelopoulos et al. (2010) suggest a conceptual model in order to investigate CSF in NSD projects in electronic government. They connected previous NSD studies with the new digital economy services. The authors justify and build on the existing literature that advocates the use of CSF as a tool to study the implementation of these projects (Angelopoulos et al. 2010).

A later study in the hotel sector used a multicriteria methodology (i.e., UTADIS method) to examine and compare the results with other popular classification methods, in order to develop a predictive model for NSD success (Kitsios et al. 2009). The same determinant factors emerged as critical, but the study offered an improved accuracy in the prediction of the success in future new services, which was also measured with a Likert-type scale. The criteria, which seemed to have a lower contribution to the success of NSD, were further examined, in order to investigate whether an intensive performance in these criteria follows a predictable behaviour for the NSD result (Kitsios et al. 2013, 2015).

The differentiation strategy approach for Greek museums in order to handle the demanding business environment is very difficult and must be effective. The challenge for the museum executives is very significant and nobody wants to be excluded from such a holistic development that comes upon. Thus, it is important to examine if it is possible for someone to investigate the development of successful new services in museums. It is also important to examine if there are any factors that are able to discriminate the successful from the unsuccessful new services.

3 Methodology

3.1 Data and Variables

Based on the aforementioned framework, this study extends the use of Multicriteria Decision Analysis (MCDA) in predicting the success of new services in the Greek cultural and creative industries. The main purpose of the study is to give insights on NSD in the museum industry, focusing on the city of Athens.

Examining the descriptive variables that appear in literature and are tested by trial and error in various studies of development of new services (Cooper 2001; De Brentani 2001; Cooper and De Brentani 1991; Harrington and Ottenbacher 2010; Kitsios et al. 2009; Voss et al. 2006; Camarero and Garrido 2009; Kitsios 2005; 2006), a set of variables is identified that are sufficient to describe the phenomenon of the development of new services in Museums.

Collecting determinant factors of success in NSD defined in Greek and international literature, a 126-factor questionnaire was formulated and used in interviews conducted in 70 different cultural organisations of Athens. There was a positive response from 62 organisations, 40 out of which were public and 22 private. The total number of new services, developed and completed by the interviewed organisations during the last five years was 184, including successes and failures. Data were collected by direct in depth interviews with the museum managers.

To identify the determinants of success or failure for a new service, 126 variables were developed and tested in nine categories (company profile, new services generally, description of new service offered, idea generation sources, activities for new service development, organisation, resource allocation, market potentiality, market synergy). First, respondents were asked to select and refer to one new service. Then, they were asked to indicate the level of quality of performance, with the way each of the 126 variables reflected the events that occurred during the NSD process. Success and failure was defined by each respondent in terms of their own company's interpretation of whether or not the new service met their success criteria. The variables were measured using a 5-point Likert-type point scale anchored at each end with "percentage of 0 %—not done" and "100 %—completely done". This approach produced a more reliable rating than continuous scales (Churchill 1987). Moreover, a five-point Liker-type scale was adopted for measuring the dependent variable that refers to the new service success or failure. This is one of the most important advantages of the study, since the majority of previous researches consider the new service success as a dual dependent variable, ie success or failure (Kitsios et al. 2009). The adopted larger measurement scale is more consistent with practice, while it may significantly improve the accuracy of the prediction model.

Out of the initial set of 126 variables, 33 were finally selected for the development of reliable success/failure prediction models. These were selected with correlation analysis of the variables in the same category; those with the lower correlation were preferred.

Given the classification of the projects as successful or failures, the objective of the analysis was to explore the development of a reliable success identification/prediction model, which aggregates all the relevant information as described by the selected variables. MCDA methods, often encountered in decision-making situations, are well-suited for such kind of data. Thus, in this study, a MCDA classification methodology, namely the UTADIS method, was employed to implement a disaggregation approach. The next section describes briefly the employed multicriteria methodology.

3.2 Multicriteria Methodology

As mentioned earlier in Sect. 3.1, the UTADIS method, which implements a disaggregation approach (Zopounidis and Doumpos 1999; Doumpos and Zopounidis 2002) has been selected for the present study.

The UTADIS method leads to the development of an additive value function that is used to estimate the expected outcome of each NSD project. The developed additive function has the following general form:

$$V(\mathbf{x}_i) = \sum_{j=1}^{n} w_j v_j\left(x_{ij}\right) \in [0,\ 1] \tag{1}$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ is the description of project i on the set of n evaluation criteria (i.e., the 25 selected variables), w_j is the trade-off constant of criterion j and $v_j(\cdot)$ is the marginal value function of criterion j. The trade-off constants are non-negative, sum up to 1 and are often interpreted as proxies for the relative important of the criteria in the mode. On the other hand, the marginal value functions provide a mechanism for decomposing the aggregate result (global value) in terms of individual assessment at the criterion level. Both the global value $V(\mathbf{x}_i)$ and the marginal values are normalized in [0, 1], with higher values associated with higher likelihood of success.

To avoid the estimation of both the criteria weights and the marginal value functions, the transformation $u_j\left(x_{ij}\right) = w_j v_j\left(x_{ij}\right)$ can be used. This way, the additive value function (1) is simplified as follows:

$$V(\mathbf{x}_i) = \sum_{j=1}^{n} u_j(x_{ij}) \tag{2}$$

The assignment of a project i into one of the k predefined classes is determined by comparing its global value $V(\mathbf{x}_i)$ to the $k-1$ thresholds $0 < t_1 < t_2 < \ldots < t_{k-1} < 1$ that distinguish the classes. Thus, a project i is assigned to group ℓ if $t_\ell \leq V(\mathbf{x}_i) < t_{\ell-1}$.

The estimation process for the additive value function and the cut-off thresholds use a set of data to fit the model (reference set of m projects) using linear programming techniques. The objective of the method is to develop an optimal additive model that minimises the classification error for the projects in the reference set. This is achieved through the solution of the following optimisation problem:

$$\begin{aligned}
&\min \sum_{i=1}^{m} \left(\sigma_i^+ + \sigma_i^-\right) \\
&\text{s.t.}: \sum_{j=1}^{n} u_j(x_{ij}) + \sigma_i^+ - t_\ell \geq \delta && \text{for each project } i \text{ from class } \ell = 1, \ldots, k-1 \\
&\sum_{j=1}^{n} u_j(x_{ij}) - \sigma_i^- - t_{\ell-1} \leq -\delta && \text{for each project } i \text{ from class } \ell = 2, \ldots, k \\
&u_j(x_{ij}) - u_j(x_{i'j}) \geq 0 && \text{for all } j = 1, \ldots, n \text{ and projects } i, i' \text{ with } x_{ij} \geq x_{i'j} \\
&\sum_{j=1}^{n} u_j(x_{j*}) = 0, \ \sum_{j=1}^{n} u_j\left(x_j^*\right) = 1 && j = 1, \ldots, n \\
&t_\ell - t_{\ell+1} \geq \varepsilon && \ell = 1, \ldots, k-2 \\
&\sigma_i^+, \sigma_i^-, t_\ell \geq 0 && i = 1, \ldots, m, \ \ell = 1, \ldots, k-1
\end{aligned} \tag{3}$$

where σ_i^+ (σ_i^-) is the classification error for project i from group ℓ with respect to the lower (upper) threshold of the group, x_{j*} and x_j^* denote the anti-ideal and ideal performances on criterion j, and ε, $\delta \geq 0$ are user defined constants. With a piecewise modelling of the marginal value functions, the above optimization formulation can be expressed in linear programming form. A detailed description the model and its properties can be found in the works of Zopounidis and Doumpos (1999) and Doumpos and Zopounidis (2002).

4 Results

One of the most important results of the UTADIS method refers to the estimation of the weights of the examined factors. These weights are shown in Table 1 and represent the overall relative importance or the contribution of a variable in the new service success/failure.

Table 1 Estimated weights for the examined variables

Criteria/variables	Weights (%)	Notation
The objectives were expressed as contribution in the final income and/or profit of the company	2.68	x_1
“Good” projects which did not fit with the strategy of the organisation were dismissed at times	3.03	x_2
A systematic effort was pursued by the company for the capture and collection of new ideas	2.62	x_3
The idea was derived from insistence and not from passive search	3.76	x_4
Group decision based on informal discussion—not using specific techniques (criteria lists, ranking forms, etc.)	2.58	x_5
Individual decision taken informally (not using specific techniques)	2.63	x_6
Existed access in secondary published data	3.07	x_7
The preliminary assessment of the market and technical needs was well supported by written evidence (documented)	13.08	x_8
Customer opinion of the new product was obtained very early in the development process	2.74	x_9
Official go/kill decision making process was developed after conducting “business” and economic analyses	2.63	x_{10}
A superficial analysis was made: not formal, approximately guesses and estimates	2.56	x_{11}
A multi-disciplinary project team was organised with members coming from different departments of the company in order to support the new service development project	2.60	x_{12}
The cross-functional team was responsible for the result	2.55	x_{13}
Public opinion assessment was designed	2.57	x_{14}
The new service development period was shortened	2.71	x_{15}
The concept of service underwent numerous revisions throughout the development process	2.57	x_{16}
During the various stages of the development process, a series of “go / kill” decisions took place	2.61	x_{17}
Enough time was invested in testing the new service to ensure that all technologies were working appropriately	2.67	x_{18}
The organisation was certain for planning new service technically—there were no errors or technical failures	2.65	x_{19}
Training frontline staff for the conception of new ideas	2.78	x_{20}
Developed and shared comprehensive manual with clear instructions for use and distribution of the new service	2.60	x_{21}
The promotion plan was tested in selected audience segment only	2.55	x_{22}
The promotion plan was tested in selected geographical point	2.81	x_{23}
The costs were reviewed	2.57	x_{24}
Promotion by commercial flyers, web, exhibitions, press conference and advertisements, but not special promotion or training for the sales force	2.60	x_{25}
Very limited effort was put: nothing really exceptional for the promotion	2.56	x_{26}
A full-scale launch occurred with an identifiable set of marketing activities specific to this product	2.56	x_{27}

(continued)

Table 1 (continued)

Criteria/variables	Weights (%)	Notation
There was a high level of awareness within the company where this new product was being developed	2.69	x_{28}
The various employees who participated in the development of this service were appropriately qualified for their mission	2.52	x_{29}
Sufficient resources—people, time and money—were committed so that operational or process activities could be undertaken effectively and on time	3.03	x_{30}
We knew well the size of the potential market for our product	3.03	x_{31}
The product fitted well with the current image of our company in the marketplace	2.71	x_{32}
By the time we commercialised our product, we understood our potential customers' needs and desires for their product	2.70	x_{33}

Table 2 Criteria account more than 45 % of the total trade-offs

Criteria/variables	Weights (%)	N
The preliminary assessment of the market and technical needs was well supported by written evidence (documented)	13.08	x_8
The idea was derived from insistence and not from passive search	3.76	x_4
Existed access in secondary published data	3.07	x_7
"Good" projects which did not fit with the strategy of the organisation were dismissed at times	3.03	x_2
Sufficient resources—people, time and money—were committed so that operational or process activities could be undertaken effectively and on time	3.03	x_{30}
We knew well the size of the potential market for our product	3.03	x_{31}
The promotion plan was tested in selected geographical point	2.81	x_{23}
Training frontline staff for the conception of new ideas	2.78	x_{20}
Customer opinion of the new product was obtained very early in the development process	2.74	x_9
The product fitted well with the current image of our company in the marketplace	2.71	x_{32}
The new service development period was shortened	2.71	x_{15}
By the time we commercialised our product, we understood our potential customers' needs and desires for their product	2.70	x_{33}

Overall, twelve criteria accounted more than 45 % of the total trade-offs. These criteria are summarised in Table 2. As it can be concluded, criterion x_8 is very significant according to the UTADIS method. The preliminary assessment of the market and technical needs was well supported with written evidence (documented) emerged as the most decisive and critical factor in NSD success, having a high weight in the UTADIS model and a high significance in the final success ratio of new services. 13.08 % of the final service success is guaranteed by forecasting the service performance, rather than beginning the procedure and organising it step by step, as the project goes by.

In addition, the following criteria are significant according to the UTADIS method:

- x_4: The idea was derived from insistence and not from passive search,
- x_7: Existed access in secondary published data,
- x_2: "Good" projects which did not fit with the strategy of the organisation were dismissed at times,
- x_{30}: Sufficient resources—people, time and money—were committed so that operational or process activities could be undertaken effectively and on time, and
- x_{31}: We knew well the size of the potential market for our product

The weights of all these criteria are similar and add 15.92 % in the final success ratio.

Finally, other important criteria include the following:

- x_{23}: The promotion plan was tested in selected geographical point,
- x_{20}: Training frontline staff for the conception of new ideas,
- x_9: Customer opinion of the new product was obtained very early in the development process,
- x_{32}: The product fitted well with the current image of our company in the marketplace,
- x_{15}: The new service development period was shortened, and
- x_{33}: By the time we commercialised our product, we understood our potential customers' needs and desires for their product emerged also as critical factors.

It is very important to mention that the multicriteria methodology classified the results with an extremely high accuracy, reaching an 82.06 % of correct overall classification in all categories. This high level accuracy indicates that the multicriteria methodology performs quite well for the classification of all the new service projects. Detailed classification results are given in Table 3.

Finally, the aggregation of the estimated criteria weights into the main factor categories justifies the aforementioned findings. In particular, as shown in Fig. 1, the preliminary assessment (e.g., resources and time investment, market target focus) and the idea generation are the most important dimensions, having an

Table 3 Detailed classification results (%)

		Predicted category				
		1	2	3	4	5
Actual category	1	0.00	3.80	0.00	0.00	0.00
	2	0.00	58.15	0.00	0.00	0.00
	3	0.00	0.00	1.63	0.00	0.00
	4	0.00	0.00	4.35	22.28	0.00
	5	0.00	0.00	2.17	7.61	0.00

Fig. 1 Estimated importance for the main factor categories

aggregated weight of 13.1 % and 6.4 %, respectively. Strategy focus is also an important category of variables affecting the success/failure of new services (weight 5.7 %), while market synergy, sources training, market assessment and sales testing, systems planning and control, new service development planning, idea preliminary assessment, organization, business analysis, process, market launch and project team creation seem to play a less important role but with a weight value close 5 % each.

It would be also very useful to analyse how the studied criteria factors contribute in the overall success of new services according to their performance level. This can be examined by the estimated marginal value functions resulted by the UTADIS method. The value functions of the six most important criteria are presented in Fig. 2.

The curve shape of the value functions indicates the demanding level of the examined criteria. In particular, a convex form shows a high demanding level, i.e., the contribution of this criterion to the new service success/failure is not high, unless its requirements are greatly fulfilled. The opposite occurs in the case of a concave form, i.e., the contribution of this criterion to the new service success/failure is high, even if its requirements are not greatly fulfilled. Based on the results of Fig. 2, the following should be noted:

(a) The weight of the most important criterion "A preliminary assessment of technical needs fully supported by written evidence" differs from all the others and its value function has a different form too. As it can be observed, this particular criterion plays a very important role in the new service success, only

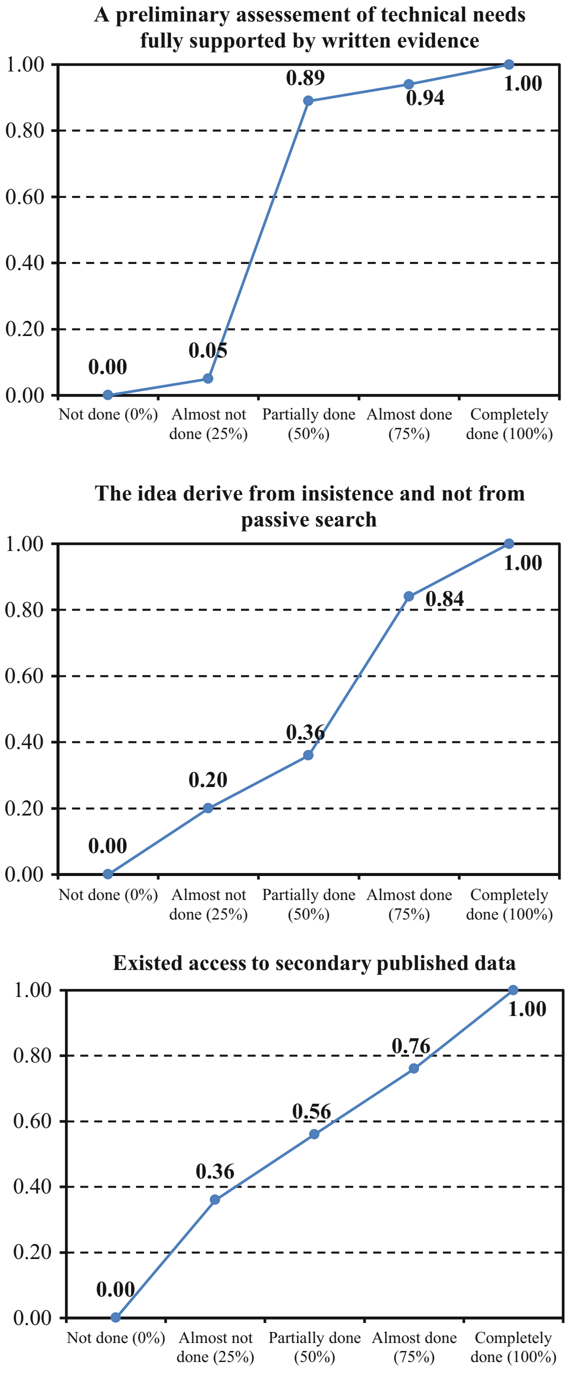


Fig. 2 Value functions (normalized) for the most important criteria

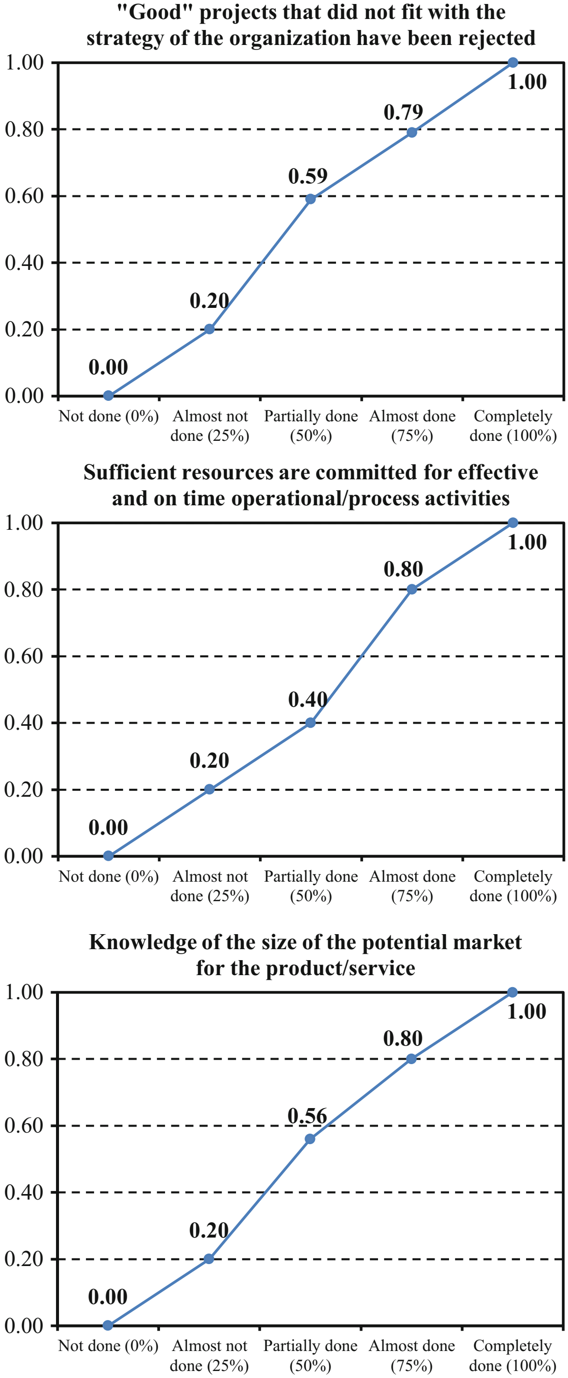


Fig. 2 (continued)

if it is partially, almost or completely fulfilled (i.e., their performance should be greater than 50 %).

(b) The other criteria seem to have the same weights and their demanding levels are similar. In particular, the "The idea derive from insistence and no from passive search", "Existed access in secondary published data", ""Good" projects which did not fit with the strategy of the organization have dismissed at times", "Sufficient resources—people, time and money—are committed so that operational or process activities can be undertaken effectively and on time" and "We knew well the size of the potential market for our product", appear to have a neutral demanding level, i.e., the larger their performance, the grater their impact to the new service success.

5 Conclusions

The main aim of this study is to highlight the NSD procedure in cultural and creative industries especially in Athens museums and to give an insight of how museums managers undertake the necessary activities in order to ensure that the new services are successful and profitable. The predictive model that was eventually developed can accurately distinguish success and failure cases of new services, setting relative weights for each one of the factors that have emerged as critical. Thus, there is a clear and important implication that there is correlation between the development procedure and the eventual success or failure of a new service.

The overriding finding of the investigation is that new product success is closely linked to what activities are carried out in the new services process, how well they are executed, and the completeness of the process. That is, people and not solely the nature of the market, the type of technology, or even the synergy or fit between the project and the organisation-doing tasks and, most importantly, people doing them well, contribute strongly to new services success.

Having as main focus the museums services sector in Athens, the overall findings of this study could trigger a comparison with other countries that may be considered as the main competitors of the Greek cultural sector. They could also raise questions of whether the findings can be generalised and used in other sectors of the cultural and creative industries or in the service industries, in general. Finally, a cross-time comparison and application of the new service success model would be an interesting and much needed scientific challenge. Innovation is a necessity for both private and public organisations. Especially in times of crisis, like the current one, the implementation of innovative policies is a key factor for strengthening the position of the cultural organisations. Apparently, two approaches are offered to the cultural organisations to follow: (i) an operational approach based on the market orientation, and (ii) a cultural approach based on the orientation of growth of the achieved services.

References

Angelopoulos S, Kitsios F, Papadopoulos T (2010) "New service development in e-Government: identifying critical success factors. Transform Gov People Process Policy 4(1):95–119

Bakhshi H, Throsby D (2010) Culture of Innovation: an economic analysis of innovation in arts and cultural organisation. NESTA, Londres

Barczak G, Kahn K, Moss R (2006) An exploratory investigation of NPD practices in non-profit organizations. J Prod Innov Manag 23:512–527

Buisson D, Garrett T, Souder WE (1997) Success through customer—driven new product development: A comparison of US and New Zealand small entrepreneurial high technology firms. J Prod Innov Manag 14(6):459–472

Camarero C, Garrido MJ (2008) The role of innovation in the relation between market orientation and performance in cultural organization. Eur J Innov Manag 11(3):413–34

Camarero C, Garrido MJ (2009) Assessing the impact of organizational learning and innovation on performance in cultural organizations. Int J Nonprofit Volunt Sect Mark 25(3):15–32

Camarero C, Garrido MJ (2012) Fostering innovation in cultural contexts: Market orientation service orientation, and innovations in museums. J Serv Res 15:39–58

Camison C, Lapiedra R, Segarra M, Boronat M (2009) A meta-analysis of innovation and organizational size. Organ Stud 25(3):331–61

Castañer X (2014) Cultural innovation by cultural organizations. In: Ginsburgh VA and Throsby D (eds) Handbook of the economics of art and culture, vol 2. Elsevier, pp 264–274

Castañer X and Cavotta V (2010) Creativity and innovation in cultural organizations. Paper Presented at the 16th International Conference on Cultural Economics, Copenhagen.

Churchill G (1987) Marketing research: methodological foundations. Holt, Rinehart and Winston, New York

Cooper RG (2001) Winning at new products: accelerating the process from idea to launch, 3rd edn. Perseus Books, Reading, MA

Cooper RG, De Brentani U (1991) New industrial financial services: What distinguishes winners. J Prod Innov Manag 8(2):75–90

Cooper RG, Kleinschmidt EJ (1987) New products: What separates winners from losers. J Prod Innov Manag 4(3):169–184

De Brentani U (1986) Do firms need a custom-designed new product screening model? J Prod Innov Manag 3:108–19

De Brentani U (1991) Success factors in developing new business services". Eur J Mark 25(2): 33–59

De Brentani U (2001) Innovative versus incremental new business services: Different keys for achieving success. J Prod Innov Manag 18:169–187

De Jong J, Vermeulen P (2003) Organizing successful new service development: A literature review. Manag Decis 41(9):844–858

Doumpos M, Zopounidis C (2002) Multicriteria decision aid classification methods. Kluwer Academic Publishers, Dordrecht

Edgett S, Parkinson S (1994) The development of new financial services: Identifying determinants of success and failure. Int J Serv Ind Manag 5(4):24–38

Ernst H (2002) Success factors of new product development: a review of the empirical literature. Int J Manage Rev 4(1):1–40

Flikkema MJ (2008) Service development and new service performance. PhD Thesis, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Harrington R, Ottenbacher M (2010) Strategic management: An analysis of its representation and focus in recent hospitality research. Int J Contemp Hosp Manag 23(4):439–462

Kitsios F (2005) Innovation management in new service development. PhD thesis, Technical University of Crete, Chania, Greece

Kitsios F (2006) Services marketing in the hospitality economy: an exploratory study. In: Paper presented at the 98th EAAE Seminar, Marketing Dynamics within the Global Trading System: New Perspectives, June-July 2006, Chania, Greece
Kitsios F, Doumpos M, Grigoroudis E, Zopounidis C (2009) Evaluation of new service development strategies using multicriteria analysis: Predicting the success of innovative hospitality services. Oper Res 9(1):17–33
Kitsios F, Moschidis O, Livanis E (2013) Service innovation strategies in Greek hotel sector: an exploratory study using the statistical method of multidimensional analysis. Int J Data Anal Tech Strateg 5(1):49–62
Kitsios F, Grigoroudis E, Giannikopoulos K, Doumpos M, Zopounidis C (2015) Strategic decision making using multicriteria analysis: new service development in Greek hotels. Int J Data Anal Tech Strateg 7:187–202
Kotler P (2000) Marketing management. The millennium edition. Prentice-Hall International Editions, New Jersey
Maidique MO, Zirger BJ (1984) A study of success and failure in product innovation: The case of the U.S. electronics industry. IEEE Trans Eng Manag 31:192–203
Martin CR, Horne DA (1995) Level of success inputs for service innovations in the same firm. Int J Serv Ind Manag 6(4):40–56
Myers S, Marquis D (1969) Successful industrial innovations, NSF 69-17. National Science Foundation, Washington, DC
Ottenbacher M, Gnoth J (2005) How to develop successful hospitality innovation. Cornell Hotel Restaur Admin Q 46(2):205–222
Ottenbacher M, Gnoth J, Jones P (2006) Identifying determinants of success in development of new high-contact services". Int J Serv Ind Manag 17(4):344–363
Reicheld FF, Sasser WE (1990) Zero defects: quality comes to services. Harv Bus Rev (September–October): 105–113
Rothwell R, Freeman C, Horlsey A, Jervis VTP, Roberston AB, Townsend J (1974) SAPPHO updated—project SAPPHO phase II. Res Policy 3:258–291
Van Riel ACR, Lemmink J, Ouwersloot H (2004) High-technology service innovation success: a decision-making perspective. J Prod Innov Manag 21:348–359
Voss G, Montoya-Weiss M, Voss Z (2006) Aligning innovation with market characteristics in the non-profit professional theatre industry. J Mark Res 43(2):296–302
Zopounidis C, Doumpos M (1999) A multicriteria decision aid methodology for sorting decision problems: The case of financial distress. Comput Econ 14(3):197–218

Fostering a Competitive Differentiation Strategy for Sustainable Organizational Performance

Dimitrios Mitroulis and Fotis Kitsios

Abstract Market orientation has always been a critical factor of creating useful business knowledge. As a result, this kind of information is crucial for creating dynamic innovation capabilities. Knowing and affiliating that to their strategy, firms could be able to differentiate easily from their competitors. Having taken into consideration the above framework, they are led to better organizational performance. The only way to achieve such a project is to be always up-to-date and proactive. Added value, competitive advantage, customers' satisfaction are some of the missions that a business should be able to accomplish. The purpose of this research is to present how SMEs' could adjust their strategy depending on their customer and competitors' orientation, innovation capabilities so that organizational performance could be achieved. Differentiation is a weapon which is difficult to use successfully. Most of the researches have shown the importance of innovativeness and performance. However, this research is going to show how differentiation and competitive innovation strategy could affect the organizational performance, not only financially but also non-financially.

Keywords Market orientation • Innovation management • Differentiation strategy • Competitive advantage • Organizational performance

1 Introduction

Being competitive is a crucial capability of today's organizations. Successful innovation management appears to be a catalyst in their growth. How could a firm achieve competitive advantage without being competitive? Market orientation is the answer. According to Avlonitis and Gounaris (1999), customer satisfaction is very important for the firms' actions and strategic moves. Moreover, market orientation is always useful due to offering information that could help firms be

D. Mitroulis (✉) • F. Kitsios
Department of Applied Informatics, School of Information Sciences, University of Macedonia, Thessaloniki, Greece
e-mail: tm1133@uom.edu.gr; kitsios@uom.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_5

competitive (Salavou et al. 2004). The latest researches about market orientation argue that customer satisfaction and proactive market orientation are, in fact, the keys to sustainable competitive advantage. Customers always seek for added value. Firm's mission is to make it a long-term dynamic capability. In order to survive, organizations should be able to innovate. This could be achieved on by creating dynamic capabilities, on which innovation process could be based (Yang 2011; Jiménez-Jiménez et al. 2008). Business capabilities have a significantly sensitive connection with the generation of market oriented knowledge. This could lead firms to differentiate from their competitors and achieve superior competitive advantage. It is very important to quickly overcome as many competition obstacles as possible, so that they get a high percentage of loyalty and market share. However, it is of great importance to disseminate the generated knowledge among the inside units and employees of the organization, generating a constantly reactive system of adapting in every change of its environment.

Firms try to wield their best efforts for achieving profits that make possible their survival, the covering of their expenses and the maximization of their values (Kumar et al. 2011; Jiménez-Jiménez and Sanz-Valle 2011; Salavou et al. 2004; Erdil et al. 2004; Wong 2013). Hence, it is vital that managers identify and understand strategic orientations, for instance market orientation, to make firms likely to achieve competitive advantage that leads to superior organizational performance. Innovation enhances the organizational capability to face the uncertainty and turbulence. Firms could seek new opportunities and exploit the existing ones more efficiently (Gunday et al. 2011; Salavou et al. 2004; Han et al. 1998; Avlonitis and Gounaris 1999; Jiménez-Jiménez et al. 2008; Yang 2011; Tidd 2001; Cepeda and Vera 2007; Avlonitis et al. 2001; Jiménez-Jiménez and Sanz-Valle 2011; Kumar et al. 2011). Moreover, innovation is also comprised the key factor of creating and sustaining of competitive advantages, which as a result leads to the expansion of the organizational performance. Being innovative is related to flexibility. Flexibility facilitates firms to find easier to adapt to their environment, enabling them to leverage opportunities better than their competitors. In order to succeed within a competitive environment, firms ought to be more innovative. Firms must be up-to-date of the changes that continuously appear in the market. This involves staying oriented to their customers, adopting a market orientation strategy (Laforet 2009). The ultimate target of developing a market orientation strategy deals with enhancing organizations with innovativeness and performance.

Market orientation is major for organizations to be competitive in the global market. Responding to the constantly changing customers needs and service, firms have to make sure that they provide sufficient services to their customers. Market orientation is supposed to be an organizational culture that assist firms achieve sustainable competitive advantage via creating superior customer value (Narver and Slater 1990; Han et al. 1998). Customer needs change rapidly. Therefore, market orientation requires a clear understanding of present and future demand dynamics of targeted customers. Strategic orientations such as customer and competitor orientation, are really important for competitive advantages (Sørensen 2009; Zhou et al. 2005; Slater and Narver 1994; Jiménez-Jiménez et al. 2008). Linking market orientation to organizational performance, there is a well-based relationship in the literature, due to showing market orientation as a crucial success factor for organizational performance (Tsiotsou and Vlachopoulou 2011). As a result, the critical success of any company lies within the its ability to serve its customers. These leads firms to the adoption of more market-based strategies (e.g. market orientation) to improve its performance (Li and Zhou 2010; Hooley et al. 2003). Even though a significant amount of studies on market orientation, competitive advantage and organizational performance could be found in the marketing literature, few researches have been investigating the relationship among customer orientation, innovation differentiation and organizational performance. It is vital managers to realize and measure the impact of customer orientation on organizational performance via innovation management. Additionally, a business creating useful dynamic capabilities for the improvement of the organizational performance could achieve sustainability and competitive advantage. The importance of "Market orientation-Innovation management-Organizational performance" chain and its advantages on implementing a differentiation strategy could be crucial for today's management.

The major purpose of this study is to explore the relationship between market orientation, innovation management and organizational performance. The research questions of this study are: What is the importance of utilizing market oriented knowledge through innovation management to achieve superior competitive advantage that leads to greater organizational performance? In the rest of the discussion, this paper reviews the theoretical background and theories following a certain methodology. This is followed by a detailed questionnaire created according to the discussed literature. The questionnaire and the scales that are developed, could lead to a successful future research. Finally, the paper presents the discussions on the findings and academic and managerial implications.

2 Literature Review

Following the framework presented in Fig. 1, this study points out the problem of improving the organizational performance via utilizing market orientation and fostering innovation management as a process of creating a sustainable competitive advantage and generally, sustainable performance. Additionally, this research is based on Webster and Watson (2002) proposed research methodology. In order to create a complete literature review, which is concept-centric, there should be an appropriate organizing framework. These requirements are satisfied by presenting the main concept of the research, setting some helpful definitions and moreover organizing a systematic search among the relevant published papers. Analyzing and categorizing the results of the literature research, we try to increase the knowledge of the subject, leading to the evaluation of the results. Finally, the results are explained. In this section, a review of the found germane literature is resented.

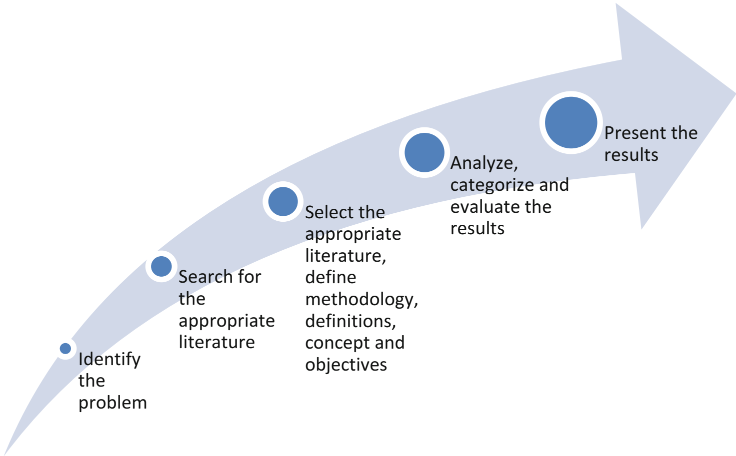


Fig. 1 Framework of the research methodology

2.1 Identifying and Choosing the Relevant Publications

The published work concerning the concept of this study was found by research among a variety of articles in “Google scholar”, “Scopus”, “Science Direct” and “Web of Science”. In order to find out the appropriate and most relevant published work, key-words and some of their combinations have been used: “organizational performance”, “innovation management”, “market orientation”. Gathering all the related published researches, we reached a number of 2,003 publications. However, it was not possible to use all of them due to not being in the same direction with this research. In this manner, the most interesting and relevant published works have been chosen. The abstracts of all the chosen publications were read and categorized according to their concepts. Some of them were rejected due to not fulfilling some certain criteria. Moreover, some were chosen from the references of some articles, as they were really helpful. On the other hand, a large number of publications were rejected due to not being available as a full text. The remaining articles were read and analyzed. The final selection of the articles that have been used in the literature review was done following some criteria:

- The article should be in English.
- The article should be in a peer-viewed journal or conference proceedings or book.
- The article should cover business topics.

Applying the above mentioned filters, the titles and abstracts of the publications were scanned for the above mentioned key words and categorized according to their relevance. Unfortunately, full texts could not be accessed for all of the selected papers, reducing the sample of the used references. In line with review of the publications, the introductions, the discussions and conclusions of all remaining

Fig. 2 Article selection process (Bondarouk and Friebe 2014)

articles were examined, leading to the exclusion of more articles. The citations of the remaining articles were then reviewed, identifying another a number of potentially relevant researches, some of which could not be accessible. All over again, introductions, discussions, and conclusions were examined and irrelevant articles were excluded as they did not cover issues related to the appropriate concept. The rest publications were reviewed based on the full text leading to some more exclusions. The rest were analyzed and categorized according to their concept. However, only the proper papers were qualified to be used in the literature review (Fig. 2).

The final number of the used references is 120, all of which were accepted by the above criteria. These papers were studied, categorized and analyzed, giving the result written below. The researches were categorized according to the central variables of this research study. Each reference focuses either on the variables or the relationship between them. Specifically, the major elements that are discussed in this research are the following:

- Market orientation
- Innovation management
- Competitive differentiation strategy
- Organizational performance

2.2 Results of Research

See Tables 1, 2 and 3.

2.3 Theoretical Background

2.3.1 Market Orientation

Firms try to become more market-oriented and customer-driven in order to create and offer superior customer added value, anticipating greater customer satisfaction and loyalty in the longer term (Hult 2011). Market orientation and its importance in the superior customer value and organizational performance has inspired a rich literature in marketing research studies (Kohli and Jaworski 1990; Narver and Slater 1990; Kohli et al. 1993; Slater and Narver 1994; Jaworski and Kohli 1996; Avlonitis and Gounaris 1997, 1999; Han et al. 1998; Cavera et al. 2001; Tomášková 2009; Hult 2011). It seems to be one of the hottest research methods founded on marketing conceptions and marketing management. In the early 90s of the twentieth century Kohli and Jaworski and Narver and Slater re constructed the idea of market orientation. It is supposed to be a method to redound a more effective management of organizations. This idea was osculated by many researchers among these years. These studies implicated definitions of market orientation, force of market orientation on business performance, implementation of market orientation into management of organizations and methods for measuring market orientation. Additionally, it is well related to organizational innovation and innovation management (Han et al. 1998). A wide range of methods have been stated in the previous years, but most of these methods include the knowledge of Kohli and Jaworski (MARKOR) and Narver and Slater (MKTOR).

Narver and Slater (1990) occupied customers orientation, competitors orientation and interfunctional co-ordination, as behavioral component, in addition to decisions components such as endurance and profitability. It relates two stakeholder orientation, referring to customers and competitors (Hooley et al. 2003; Kumar and Subramanian 2000; Dawes 2000; Farrelly and Quester (2003); Pumphrey 2004; Liu 1995). Measurement of market orientation includes marketing activities, the development of new products, marketing plan, segmentation, management activities, marketing research. Deng and Dart (1999) additionally introduced performance orientation in market orientation research. Gima (1995) implemented market performance in market orientation based on collecting market information, developing a market-oriented strategy and responding to customers, related to customers orientation and competitors orientation. Moreover, organizational prosperity was

Table 1 Market orientation authors and components

Authors	Components of market orientation
Kohli and Jaworski 1990;Vazquez et al. 2002; Farrelly and Quester 2003; Han et al. 1998; Kohli et al. 1993; Jaworski and Kohli 1996	• Gaining information, • Information dissemination, • Information response
Narver and Slater 1990; Harris and Ogbonna 2001; Kumar and Subramanian 2000; Slater and Narver 1994; Mitroulis and Kitsios 2014	• Customer orientation, • Competitor orientation, • Interfunctional co-ordination
Liu 1995; Avlonitis and Gounaris 1997	• Interfunctional co-ordination, • Customer orientation
Deng and Dart (1999)	• Customer orientation • Competitor orientation • Interfunctional co-ordination • Profit orientation
Gima 1995	• Gaining of information • Strategy developing • Implementation of market orientation (All customer oriented)
Fritz 1996; Deshpande and Farley 1998; Hajjat 2002	• Customer orientation
Lado et al. 1998	• Gaining information and analysis: (Customer, Distributor Competitor, Environment) • Interfunctional co-ordination and strategy: (Customer, Competitor, Environment)
Farrell and Oczkowski 2002; Dawes 2000; Farrell 2002; Varela and Rio 2003	• Gaining information (Customer, competitor, interfunctional co-operation) • Information dissemination (Customer, competitor, interfunctional co-operation) • Response on the information (Customer, competitor, interfunctional co-operation)
Avlonitis and Gounaris 1999	• Interfunctional co-ordination • Competitor orientation
Cadogan et al. 1999; Gray et al. 1998	• Gaining information • Information dissemination • Information response • Profit orientation • Orientation on some components of external environment
Akimova 2000	• Competitor advantage orientation • Response to hostility environment
Harrison-Walker 2001; Hooley et al. 2003	• Competitor orientation (Gaining information, information dissemination, interpretation and response) • Customer orientation (Gaining information, information dissemination, interpretation and response) • Business performance orientation

(continued)

Table 1 (continued)

Authors	Components of market orientation
Helfert et al. 2002	• Customer orientation • Profit orientation • Interfunctional co-ordination • Gaining information
Pulendran et al. 2003	• General aspect • Politic aspect • Interaction aspect • Business performance orientation
Bigne et al. 2004; Tomášková 2009; Hult 2011; Sørensen 2009; Zhou et al. 2005; Tsiotsou and Vlachopoulou 2011; Panigyrakis and Theodoridis 2007; Narver et al. 2004; Li and Zhou 2010; Salavou et al. 2004; Zolfagharian 2010; Liao et al. 2011; Raju et al. 2011; Haugland, et al. 2007; Megicks and Warnaby 2008; Ramaseshan et al. 2002; Wood et al. 2000; Lonial et al. 2008; Pumphrey 2004	• Competitor orientation (Gaining information, information dissemination, interpretation and utilization) • End-customer orientation (Gaining information, information dissemination, interpretation and utilization) • Distributor orientation • Orientation on some components of external environment • Interfunctional coordination

introduced connected to customer orientation (Fritz 1996; Deshpande and Farley 1998; Hajjat 2002). Deshpande and Farley (1998) focused on customers' no expressed wishes, identifying firms as a service to customers and management in the sense of long-term customers, in the concept of customer orientation. Lado et al. (1998) developed a different perception of market orientation as a degree of using information about stakeholders (i.e., end customers, distributors, competitors and environment), implementations on strategic activities, interfunctional coordination and response to gaining information about them. Gray et al. (1998) implemented response and profit as parts of market orientation components.

The process of defending market share, success of new products, sells improvement, customer loyalty, return of investments and performance related market orientation, costs, market turbulence, competitors' ability, technology turbulence, customers' ability, market growth, size of the market, entrance barriers and suppliers' ability were also discussed as concepts of market orientation (Oczkowski and Farrell 1998; Cadogan et al. 1999). Avlonitis and Gounaris (1999) dealt with the definition of marketing orientation supposing that market orientation is the same as marketing orientation. They analyzed interfunctional coordination including the approach of risk management, centralization, formalization and moderation of competition (Avlonitis and Gounaris 1997). Akimova (2000) assessed market orientation from the view of gaining competitive advantages, responding to unfriendly environment and performance (Akimova 2000; Helfert et al. 2002). Dawes (2000) highlights the analysis of customer, analysis of customer response, competitors' orientation and information dissemination. Harrison-Walker (2001)

Table 2 Innovation management authors and components

Authors	Components of innovation management
Gunday et al. 2011; Jiménez-Jiménez et al. 2008; Salavou et al. 2004; Yang 2011; Damanpour 1996; Damanpour and Evan 1984; Laforet 2009; Damanpour and Schneider 2006; Leal-Rodríguez et al. 2014; Nasution et al. 2011; Kitsios et al. 2009; Naidoo 2010; Damanpour et al. 1989; Keskin 2006; Low et al. 2007; Matear et al. 2002; Dibrell et al. 2011; Sondergaard 2005; Aldas-Manzano et al. 2005; Mavondo et al. 2005; Tajeddini et al. 2006; Kim and Pennings 2009; Sandvik and Sandvik 2003; Roberts 1999; Thornhill 2006; Brown and Eisenhard 1995; Kitsios and Sindakis 2014; Caves and Ghemawat 1992; Mitroulis and Kitsios 2014; Damanpour & Gopalakrishnan 1998; Wolfe 1994; Baker and Sinkula 2002; Balkin et al. 2000; Lyon and Ferrier 2002; Weerawardenaa and O'Cassb 2004	• Organizational innovation (Administrative innovation, product/service innovation, process innovation, etc.)
Jiménez-Jiménez et al. 2008; Salavou et al. 2004; Yang 2011; Damanpour 1996; Damanpour and Schneider 2006; Leal-Rodríguez et al. 2014; Matear et al. 2002; Dibrell et al. 2011; Aldas-Manzano et al. 2005; Mavondo et al. 2005; Tajeddini et al. 2006;; Sandvik and Sandvik 2003; Mitroulis and Kitsios 2014; Weerawardenaa and O'Cassb 2004	• Acquisition, dissemination and use of new knowledge (Utilizing market-oriented inflows)
Gunday et al. 2011; Salavou et al. 2004; Han et al. 1998; Avlonitis and Gounaris 1999; Jiménez-Jiménez et al. 2008; Yang 2011; Tidd 2001; Cepeda and Vera 2007; Avlonitis et al. 2001; Jiménez-Jiménez and Sanz-Valle 2011; Kumar et al. 2011; Damanpour and Schneider 2006; Leal-Rodríguez et al. 2014; Mavondo et al. 2005; Tajeddini et al. 2006; Kim and Pennings 2009; Sandvik and Sandvik 2003	• Capabilities (Management of the resources creating unique abilities for the organization)
Porter 1980; Jiménez-Jiménez and Sanz-Valle 2011; Kumar et al. 2011; Tajeddini et al. 2006; Mitroulis and Kitsios 2014; Kim and Pennings 2009	• Differentiation (Offering superior, different and unique product/service to the customers)

include competition orientation, customer orientation and business performance stating it as a process of gaining information, information dissemination, interpretation the information and utilization the information (Farrelly and Quester 2003; Varela and Rio 2003; Vazquez et al. 2002). Pulendran et al. (2003) showed market orientation as dependent on marketing plans such as general perspective, rational perspective, political perspective and interactional perspective and interfunctional

Table 3 Organizational performance authors and components

Authors	Components of organizational performance
Ferraresi et al. 2012; Hassim et al. 2011; Teece 2007; Erdil et al. 2004; Herath and Mahmood 2013; Al-alak and Tarabieh 2011; Kotler 2010; Gunday et al. 2011; Salavou et al. 2004; Mitroulis and Kitsios 2014; Han et al. 1998; Jiménez-Jiménez et al. 2008; Li and Zhou 2010; Zhou et al. 2009; Sindakis and Kitsios 2014; Cano et al. 2004; Wong 2013; Damanpour and Schneider 2006; Matear et al. 2002; Mavondo et al. 2005; Kim and Pennings 2009; Sandvik and Sandvik 2003; Raju et al. 2011; Haugland et al. 2007; Megicks and Warnaby 2008; Ramaseshan et al. 2002; Wood et al. 2000; Lonial et al. 2008; Porter 2004	• *Customer satisfaction and market performance* • *Financial performance* • *Innovation performance*

coordination. Bigne et al. (2004) argued that price policy, market tendency, segments identification, new products success, stimulation for including the changes to the strategy and fluently information between customers and companies in market, are really critical factors of being market-oriented. Customer orientation, competitor orientation and interfunctional coordination are the mostly used components in the literature. Furthermore, it seems that market orientation is well related to organizational performance. However, business performance is not a component of market orientation, it is one of its consequences.

Market orientation is an important factor of creating and delivering superior customer value and improvement of organizational performance. Considering the fact that customer needs change rapidly, a market orientation claims the understanding of both the present and future demand dynamism of target customers. The market orientation dimensions (i.e., customer and competitor orientation), seem to be two crucial strategic orientations of achieving a competitive advantage (Sørensen 2009; Zhou et al. 2005; Slater and Narver 1994; Dawes 2000). Market orientation forms a primary success factor for organizational performance (Tsiotsou and Vlachopoulou 2011; Panigyrakis and Theodoridis 2007; Narver et al. 2004; Raju et al. 2011). Moreover, it enhances a firm's performance, providing differentiation and cost advantages (Li and Zhou 2010; Salavou et al. 2004; Zolfagharian 2010). Therefore, market orientation affects performance measures and impacts a number of different levels of organizational performance (Liao et al. 2011).

2.3.2 Innovation Management

Apart from market orientation and its rich knowledge enhancement, organizations should have a constantly running routine called innovation management. If firms want to survive, they should be market driven. However, if firms want to succeed, they should be innovative. Organizational innovation is a means of response to the market. Organizational innovation is influenced by the market orientation. Jiménez-Jiménez et al. (2008) mention that innovation is the organization's means of response to the market. Additionally, innovation has to be developed and executed as an internal part of the business strategy (Gelard and Emamisaleh 2014). In addition, they support the importance of innovation to the organizational performance, indicating that it should be followed by the appropriate action plan. Market orientation has been proven to influence the firm's innovation management (Jaworski and Kholi 1993; Dibrell et al. 2011; Nasution et al. 2011; Naidoo 2010). Additionally, innovation influences organizational performance (Damanpour and Evan 1984; Damanpour et al. 1989; Naidoo 2010).

Innovation is not only restricted to innovations on products or services but also includes ways of improving management (Tajeddini et al. 2006; Naidoo 2010; Kitsios et al. 2015a, b). Successful innovations endow firms with a competitive edge in changing environment (Kim and Pennings 2009). According to Damanpour and Evan (1984), innovation is vital for coping with environmental changes or launching changes within the firm. Therefore, firms should develop and/or introduce their new organizational applications by successfully integrating technical changes (product, process or organizational innovation or the operationalization of a service) or administrative innovation (related to procedural, structural or authority tasks), in order to advance their level of goal success (Damanpour et al. 1989; Damanpour and Evan 1984). Han et al. (1998) and Mavondo et al. (2005) conclude that market-oriented firms are essentially more innovative, leading to their better organizational performance. This conclusion is explained by the fact that is composed by three major components which are intelligence generation, dissemination and response related to environmental changes (Jaworski and Kholi 1993; Kohli and Jaworski 1990; Jiménez-Jiménez et al. 2008). Therefore, firms are obliged to facilitate the necessary organizational, process and product, innovations to meet the continuous changes of the increasingly inconsistent competitive environment so as to meet present and future needs, identified by market intelligence. Dibrell et al. (2011), who refer to the same situation but on the three components of market orientation from Narver and Slater (1990). A market orientation philosophy aims at meeting customer constantly changing needs, therefore, organizations have to be innovative in order to be more flourishing in the market. Market-oriented companies tend to satisfy current and future customers and are able to use their customer and competitors' knowledge in order to change products and processes, meet customer's requests, develop new products and identify potential new customers (Dibrell et al. 2011; Salavou et al. 2004). Liao et al. (2011) and Aldas-Manzano et al. (2005) argue a positive relationship between market orientation and organizational innovation.

Both industrial and service sector provides reach literature of the relationship between market orientation and innovation. Sondergaard (2005) states that market orientation positively influences the development of new products in industries. Dibrell et al. (2011) reveals the importance of product innovation which influences market orientation in industries. Matear et al. (2002), analyzing service firms, examined the mediating effect of innovation in the relationship of market orientation and organizational performance. Furthermore, Laforet (2009) indicates that high-tech organizations advantage most from greater market orientation, when they are compared to low-tech organizations. Low et al. (2007), examining small and medium enterprises, indicate that innovation is positively correlated to market orientation. Keskin (2006) concludes that market orientation, whether of industrial or services firms, has a direct impact on their innovation.

Market orientation has been widely related with innovation and its outcomes. A variety of research studies expose a positive impact of market orientation on new products/service development specifically at the early stages of the product life cycle (Kohli and Jaworski 1990; Vazquez et al. 2002; Farrelly and Quester 2003; Kohli et al. 1993; Jaworski and Kohli 1996; Liu 1995; Narver and Slater 1990; Kumar and Subramanian 2000; Slater and Narver 1994; Gima 1995; Farrell and Oczkowski 2002; Avlonitis and Gounaris 1999; Akimova 2000; Dawes 2000; Farrell 2002; Varela and Rio 2003; Bigne et al. 2004; Laforet 2009). The organizational innovation process is to extensively depended on the amount of information gained in the market. Firms needs to be market-oriented in order to be aware of the changes in the customers' needs and behaviours, and watchfully monitor competitors' and suppliers' behaviors (Kohli and Jaworski 1990; Vazquez et al. 2002; Farrelly and Quester 2003; Kohli et al. 1993; Narver and Slater 1990; Kumar and Subramanian 2000; Slater and Narver 1994; Avlonitis and Gounaris 1997). A very rich literature supports the impact of market orientation on firm's innovativeness both in service and manufacturing companies (Gunday et al. 2011; Salavou et al. 2004; Han et al. 1998; Avlonitis and Gounaris 1999; Jiménez-Jiménez et al. 2008; Yang 2011; Tidd 2001; Cepeda and Vera 2007; Avlonitis et al. 2001; Jiménez-Jiménez and Sanz-Valle 2011; Kumar et al. 2011). Innovative firms are likely to be more efficient, achieving higher performance, and ensuring their survival (Damanpour and Schneider 2006; Leal-Rodríguez et al. 2014; Kitsios et al. 2009). Organizations that promote creativity and innovation are more likely to identify and attract opportunities that might lead to better results. Innovation always include a certain degree of risk and its success is never guaranteed. Most research studies hypothesize the existence of a positive relationship between innovation and performance (Ferraresi et al. 2012; Hassim et al. 2011; Teece 2007; Erdil et al. 2004; Herath and Mahmood 2013; Al-alak and Tarabieh 2011; Kotler 2010; Gunday et al. 2011; Salavou et al. 2004; Han et al. 1998; Jiménez-Jiménez et al. 2008; Li and Zhou 2010; Zhou et al. 2009; Cano et al. 2004; Wong 2013). An innovative approach enables firms to deal with a turbulent and dynamic environment, helping them to achieve and sustain long-term competitive advantages (Leal-Rodríguez et al. 2014; Mitroulis and Kitsios 2014; Salavou et al. 2004; Han et al. 1998; Jiménez-Jiménez et al. 2008).

Creating innovation capabilities and giving emphasis to every innovation type (organizational, administrative, process, production, technical and marketing innovation) is going to make a successful introduction of competitive advantage. Competition is the force of innovation (Salavou et al. 2004). Innovation management is the means of creating innovation capabilities (Salavou et al. 2004; Han et al. 1998; Avlonitis and Gounaris 1997; Jiménez-Jiménez et al. 2008). Innovation management process, includes acquisition, dissemination, and use of new knowledge (Jiménez-Jiménez et al. 2008; Salavou et al. 2004; Kitsios and Sindakis 2014;

Yang 2011; Damanpour 1996; Kitsios et al. 2015a, b; Mitroulis and Kitsios 2015). Every innovation type (administrative process, production, technical and marketing innovation) must be managed in the appropriate way so that the organization could enrich its capabilities (Gunday et al. 2011). Innovation management is the process of creating innovation capabilities (Salavou et al. 2004; Han et al. 1998; Avlonitis and Gounaris 1999; Jiménez-Jiménez et al. 2008; Yang 2011; Tidd 2001; Kitsios et al. 2009). In fact, innovation capabilities are the foundation of competitive advantage, combining new resources and information (Cepeda and Vera 2007; Avlonitis et al. 2001; Erdil et al. 2004; Jiménez-Jiménez and Sanz-Valle 2011; Kitsios and Sindakis 2014; Kumar et al. 2011). Differentiation is about offering superior, different and unique product/service to the customers. Organizations which adopt differentiation usually charge a higher price for their offerings than their competitors because of the unique features that they offer (Porter 1980). Finally, as Damanpour and Evan (1984) have proven, innovation improves the organization's outcome.

2.3.3 Organizational Performance

Organizational performance is a very crucial factor for a successful business. It depends on customer satisfaction, financial, market and innovation performance. The measurement of performance deals with market share, sales of new products, the rates of return on investment and evaluation of internal factors such as operational improvements (Ferraresi et al. 2012; Hassim et al. 2011, 2011; Teece 2007; Erdil et al. 2004; Herath and Mahmood 2013; Mitroulis and Kitsios 2014). Researchers have approached differently the measurement of business performance. Two basic approaches of business performance measurement have been applied, concluding the distinction of two categories: subjective and objective measures. Most objective measures are based on financial data, reports and information which, on the other hand, may be subjectively constructed. Obtaining objective data is hard to gain which leads to the use of subjective measures (Cano et al. 2004; Hooley et al. 2003) Managers are usually very busy and there is a risk that the questions, concerning organizational performance, might not be answered and managers may be reluctant to provide exact numerical values of analyzed indicators. It might be impossible to compare organizations of dissimilar sizes, operating in different industries, using different financial standards and defining their markets in diverse ways (Avlonitis and Gounaris 1997; Hooley et al. 2003).

Market orientation and organizational performance seems to be an important research area, regarding market orientation studies. Narver and Slater (1990) were the first who empirically verified the effects of market orientation on organizational performance, focusing fully on this relationship. Ramaseshan et al. (2002) also conclude a strong positive influence between market orientation and performance, implementing the development of new products or services, both in the industrial and the service sectors. Ge and Ding (2005) investigated the relationship of market orientation and organizational performance, showing a positive relationship between them. Both financial measures of performance such as, customer and employee satisfaction, service quality, and non-financial measures of performance, including market segment, occupancy rate and operating profit. A strong positive connection between market orientation and all forms of organizational performance is established. Haugland, et al. (2007) support a well-built positive relationship between market orientation and organizational performance, including as indicators, the return on assets, the perception of profitability compared to competitors and productivity, in the hotel industry. Megicks and Warnaby (2008), supported a strong positive correlation linking market orientation and organizational performance, in the retail sector, analyzing the return on investment of the firm and customer retention in a three-year period. Additionally, Wood et al. (2000) shared the same view in public hospitals. From another point of view, Lonial et al. (2008) prove that there is a positive relationship between market orientation and the development of new services, however, they do not prove any relationship between market orientation and financial performance.

Research linking innovation and organizational performance has found a positive relation existing between them (Damanpour and Evan 1984; Damanpour et al. 1989; Caves and Ghemawat 1992; Brown and Eisenhard 1995; Roberts 1999; Thornhill 2006; Kitsios and Grigoroudis 2014). Innovation has a variety of conceptualizations in the literature (Damanpour et al. 1989; Wolfe 1994; Damanpour and Gopalakrishnan 1998); but, most definitions of innovation indicate that innovation involves the adoption of a new ideas or behaviors. Considering that this research analyzes the how management innovation could influence the organizational performance. This study assumes a broad concept of innovation that includes the adoption of (a) organizational innovation, consisting of any new product/service, process, marketing and administrative innovation, (b) the ability of managing acquisition, dissemination and use of new knowledge, utilizing market-oriented inflows, (c) creating useful capabilities and (d) trying to differentiate through their innovation management creating competitive advantages. Innovation assists the firms to cope with changes in their external environment and it is one of the key drivers of a long-term successful organization, particularly in dynamic markets (Baker and Sinkula 2002; Balkin et al. 2000; Lyon and Ferrier 2002; Wolfe 1994). For these reasons, innovative firms are more capable of facing the challenges faster and taking advantage of new products/services and market opportunities better than non-innovative ones (Brown and Eisenhard 1995). Moreover, it is easy to recognize in the literature that innovation has a positive contribution to organizational performance. Damanpour et al. (1989), who firstly associated innovation and organizational performance, concluded that changes in the structure and their consequent innovations, either technical or administrative, lead to a better performance. Jiménez-Jiménez and Sanz-Valle 2011 found that the innovation-performance relationship is stronger in larger organizations, considering their capability of have more resources to invest. They, also, claim that the impact of innovation on performance is greater in industrial sector than in service sector. The age of the company also enhances innovation, mainly due to lack of organizational routines of new companies. Given the innovation-performance literature found, it could be concluded that innovation is the mediator of market orientation and organizational performance relationship (Sandvik and Sandvik 2003).

Consequently, the ultimate success of any business lies within its ability to serve its customers. This means that firms should implement market-based strategies, such as market orientation, in order to improve their overall performance (Zhou et al. 2009; Li and Zhou 2010). Kotler (2010) concluded that in order to measure an organization's performance, firms should consider customer satisfaction, customer preference, share of customer mind, customer perception, and so on. Eventually, market orientation is described as marketing's explanation of performance differentials between firms. Organizational performance is supposed to be the firm's ability to meet the needs of its stakeholders and its own needs for survival and growth (Al-alak and Tarabieh 2011; Sindakis and Kitsios 2014). The harmonious existence of innovation strategy within the general strategy of the organization could provide a stable base for more successful business strategy, operations and higher levels of performance (Wong 2013; Porter 2004).

2.3.4 Competitive Differentiation Strategy

Achievement of competitive advantage is critical for organizations. Competitive advantage is defined as a product or service of higher firm's customers value than similar offerings of its competitors (Baltzan and Phillips 2010; Porter 2004). In other words, the firm owns useful products, services, capabilities that its competitors do not. Competitive advantages are usually temporary as competitors often search for ways to duplicate the competitive advantage (Baltzan and Phillips 2010). In order to stay to the lead of competition, organizations have to constantly develop new competitive advantages. An organization should be capable of analyzing, identifying and building up competitive advantages using tools such as Porter's Five Forces, implementation of the three generic strategies and value chains (Porter 1985, 2004). Porter states that in order to analyze a firm's environment, five competitive forces must be engaged (Baltzan and Phillips 2010). These are the rivalry of competitors within its environment, threat posed by alternate products or services which might influence market share, threat of new entrants into its markets, negotiating power of customers and negotiating power of suppliers. These five competitive forces could be well-related to market orientation.

To endure and succeed, organizations are obliged to develop and implement strategies to effectively respond to the above five competitive forces. Differentiation, cost leadership and focused strategy are three competitive strategies that Porter introduced in 1985. Building a differentiation-based advantage, firms seek differentiation designing, producing and offering a highly distinctive or unique product or service features that create high value for their customers. Differentiation-based sources of competitive advantage in the process of adding value activities could be built inside the organization (O'Brien and Marakas 2011). It is a very important strategic consideration that managers must recognize whether differentiation does not lead to the ignorance of the firm's cost structure. Considering that low unit cost is less important than distinctive product/service features to firms differentiation strategy, the firm's total cost structure is really substantial (Tampoe 1994; Hitt et al. 1998). Generally, the costs of pursuing differentiation cannot be so high. Firms pursuing differentiation must be in control of expenses to balance higher costs in key activities. The cost structure of a business fostering a differentiation strategy still needs to be managed with awareness. Strategic differentiation must, therefore, aim at achieving cost equality or at least, cost proximity relatively to competitors, through keeping low costs in areas not related to differentiation and not spending too much to achieve differentiation (Beal 2000). On the other hand, differentiation is not an end in itself. Organizations should continue to search for new ways to advance the distinctive and unique features of their products/services and their ability to improve performance and competitive advantages. Gaining a high degree of customer satisfaction could improve profit and create loyalty leading to a substantial increase in repeat purchases of the firm's products (Tampoe 1994; Hitt et al. 1998). Additionally, another advantage is that a competitive differentiation strategy is based on high quality might increase the firm's potential market share (Pitts and Lei 2003).

In fact, competitive differentiation strategies based on high product/service quality increase market share resulting a significantly increased profitability. Product/service quality regularly leads to reputation increase and high demand which is translated into higher market share. Moreover, differentiation procedures could overcome substantial loyalty barriers that firms front. Highly distinctive or unique products/services make it difficult for new entrants to compete with the reputation and capabilities that existing organizations already possess. Differentiation could seriously increase the competitive advantages and profitability of firms (Pitts and Lei 2003; Bansal 2005).

On the other hand, implementing a differentiation strategy could lead to some disadvantages. To begin with, other organizations might try to imitate organizations that already have distinctive products, providing a similar or superior product. Although, differentiation strategies are effective in generating customer loyalty and higher prices, they do not completely block the market form other entrants Another disadvantage of differentiation is the difficulty in supporting a premium product/service price, which becomes familiar to the market. Whenever a product/service becomes more mature, customers become smarter and claim more requirements such as, genuine value and what they are willing to pay. Premium prices become difficult to defend as customers gain more knowledge about the product. A high cost structure of a firm practicing differentiation could become a true weakness when product/service imitations or substitutes, of lower price, strike the market. Differentiation leaves a firm vulnerable to its product/service offering or value concept when new competitors enter the market or when customers become more

Fig. 3 Conceptual framework

well-informed. As a result, firms that are incapable of sustaining their initial competitive differentiation strategy, are led to future product or service innovations, finding themselves in a dangerous cost disadvantage when large numbers of customers eventually choose those firms that offer a similar product or service at lower cost (Pitts and Lei 2003). Finally firms also face risks of overdoing differentiation that overextend the firm's resources.

Consequently, a competitive differentiation strategy seems to be well related with market orientation. In order to successfully implement such a strategy, organizations should always be informed about customers' wishes, competitors' movements, intentions and strategies. Focusing in gaining market-based knowledge and trying to set interfunctional coordination among the organizational units, firms are capable of being more successful in the market competition (Kotler 2010). The above assumptions lead to the positive relationship of market orientation and competitive differentiation strategy. Moreover, there is no sustainable differentiation strategy without an innovation focused intention. Differentiation competitive strategy is positively related to innovation management due to the unique features and capabilities that need to be utilized, by the firm, in order to successfully achieve competitive advantages. Eventually, although implementing a differentiation strategy might lead to some disadvantages, related to high cost, high prices, market barriers, customers intentions and knowledge, and competitors intentions and strategies, there is a well-based positive relationship with organizational performance (Fig. 3, Table 4).

Table 4 Concept-centric matrix (Webster and Watson 2002)

Authors	MOxIM	MOxOP	IMxOP
Kohli and Jaworski 1990; Vazquez et al. 2002; Farrelly and Quester 2003; Kohli et al. 1993; Jaworski and Kohli 1996; Liu 1995; Narver and Slater 1990; Kumar and Subramanian 2000; Slater & Narver 1994; Gima 1995; Farrell and Oczkowski 2002; Avlonitis and Gounaris 1999; Akimova 2000; Dawes 2000; Farrell 2002; Varela and Rio 2003; Bigné et al. 2004; Laforet 2009; Nasution et al. 2011; Naidoo 2010; Damanpour et al. 1989; Keskin 2006; Low et al. 2007; Dibrell et al. 2011; Sondergaard 2005; Aldas-Manzano et al. 2005; Tajeddini et al. 2006; Pumphrey 2004	Positive relationship		
Deng and Dart (1999); Gray et al. 1998; Cadogan et al. 1999; Harrison-Walker 2001; Hooley et al. 2003; Helfert et al. 2002; Avlonitis and Gounaris 1997; Pulendran et al. 2003; Kumar et al. 2011; Cano et al. 2004; Raju et al. 2011; Haugland, et al. 2007; Megicks and Warnaby 2008; Ramaseshan et al. 2002; Wood et al. 2000; Lonial et al. 2008; Ge and Ding 2005		Positive relationship	
Jiménez-Jiménez et al. 2008; Salavou et al. 2004; Han et al. 1998; Hult 2011; Zhou et al. 2005; Tsiotsou and Vlachopoulou 2011; Panigyrakis and Theodoridis 2007; Li and Zhou 2010; Salavou et al. 2004; Zolfagharian 2010; Liao et al. 2011; Damanpour and Schneider 2006; Mitroulis and Kitsios 2014; Gunday et al. 2011; Avlonitis et al. 2001; Narver et al. 2004; Sindakis and Kitsios 2014; Zhou et al. 2009; Al-alak and Tarabieh 2011; Matear et al. 2002; Mavondo et al. 2005; Sandvik and Sandvik 2003	Positive relationship	Positive relationship	Positive relationship
Yang 2011; Damanpour 1996; Damanpour and Evan 1984; Tidd 2001; Kitsios et al. 2009; Cepeda and Vera 2007; Jiménez-Jiménez and Sanz-Valle 2011; Teece 2007; Ferraresi et al. 2012; Hassim et al. 2011; Kotler 2010; Wong 2013; Herath and Mahmood 2013; Damanpour and Schneider 2006; Leal-Rodríguez et al. 2014; Nasution et al. 2011; Naidoo 2010; Damanpour et al. 1989; Tajeddini et al. 2006; Kim and Pennings 2009; Roberts 1999; Kitsios and Sindakis 2014; Thornhill 2006; Brown and Eisenhard 1995; Caves and Ghemawat 1992; Damanpour and Gopalakrishnan 1998; Wolfe 1994; Baker and Sinkula 2002; Balkin et al. 2000; Lyon and Ferrier 2002; Weerawardenaa and O'Cassb 2004			Positive relationship

3 Measurement, Scales and Questionnaire

In order to evaluate the theoretical research a questionnaire has been created. The literature review in the previews section, provides the basis for a survey design. This study collects and adapts scales from previous works according to which the items and responses appear on a ten-point Likert scale ranging from 1 (completely disagree) to 10 (completely agree), a seven-point Likert scale ranging from 1 (completely disagree) to 7 (completely agree) and five-point Likert scale ranging from 1 (completely disagree) to 5 (completely agree). The questionnaire consists of 32 questions which use a 5-point Likert scale, due to having more studies using it.

- The questions to measure the "market orientation" are those of the scale developed in the literature and appropriately modified. 6 questions are related to "customer orientation"; 4 related to the "competitor orientation"; and 4 related to "interfunctional coordination", totaling 14 questions.
- The questions to measure the "innovation management" are those of the scale developed in the literature and appropriately modified. The total number of questions is 14.
- To measure the organizational performance, 5 questions were included in the questionnaire. All the questions were adopted and combined from existing scales in the literature.

The questionnaire has been examined by a group of experts. It is addressed to general managers, marketing directors and functional managers due to their better knowing of the internal functions of the organizations (Table 5).

4 Discussion, Managerial and Academic Implications

The literature focuses on the role of firm innovativeness as a source of competitive advantages for organizations. Recently, variables such as market orientation, generally strategic orientations, knowledge management and organizational learning are also being studied as drivers of business performance. This study maintains that market orientation is a key antecedent of innovation and that they affect performance (Jiménez-Jiménez et al. 2008; Salavou et al. 2004; Han et al. 1998; Hult 2011; Zhou et al. 2005; Tsiotsou and Vlachopoulou 2011; Panigyrakis and Theodoridis 2007; Li and Zhou 2010; Zolfagharian 2010; Liao et al. 2011; Damanpour and Schneider 2006; Gunday et al. 2011; Avlonitis et al. 2001; Narver et al. 2004; Zhou et al. 2009; Al-alak and Tarabieh 2011; Matear et al. 2002; Mavondo et al. 2005; Sandvik and Sandvik 2003). Therefore, this work simultaneously assesses the direct link between market orientation and organizational performance in addition to the mediating role of innovation management formulating a differentiation competitive strategy. This study contributes to enhancing a recent literature research on the firm's strategic efforts on market orientation,

Table 5 Questionnaire

Second-order variables	First-order variables	Items
Market orientation	Customer orientation	1. Monitor the development in demand of existing and future customers.
		2. Understanding of customer needs, as a basis for competitive business strategies.
		3. Collecting information on customer preferences and customer satisfaction.
		4. Use of research techniques for gathering information about customers.
	Competitors orientation	1. Collection and analysis of information about products, strategies and actions of competitors.
		2. Response to competitors movements.
		3. Understanding of strengths and weaknesses of current and future competitors.
		4. Effective anticipation of local and foreign in the main part of the market.
	Interfunctional coordination	1. Effective process of collecting, analyzing and publishing information on customer satisfaction within the company.
		2. Cooperation of the marketing department with other departments (e.g. meeting etc.), to predict the future behaviour of customers and competitors.
		3. When a problem arises on a customer or market share, or find out information about competitors, how quickly is this information disseminated to all parts of the organization.
Innovation management	Differentiation	1. Launch of new products and services to market before their competitors.
		2. Novelty in introducing new or renewal of existing processes compared to that of competitors.
		3. There are innovations which copyright is owned or innovations that differentiate the firm from competitors.
		4. Offer after-sales services and customer support.
	Innovation	1. Technological changes, affect the activity and stability of the organization.
		2. Provide additional resources, such as time, money, etc, to predict a technological change.
		3. Search for new management systems, compared to the corresponding research of the competitors.
		4. Rate the employees innovation abilities.
		5. Rate the level of introduction of new products / services which respond to the customers' wishes.
		6. Focus on gaining customers, for whom we can achieve a competitive advantage

(continued)

Table 5 (continued)

Second-order variables	First-order variables	Items
Organizational performance	Financial outcomes	1. Return on assets (ROA).
		2. Return on investments (ROI).
		3. Return on sales (ROS).
		4. Increased sales in key products and markets of company.
		5. Efficiency of business in achieving financial targets.
	Quality and satisfaction	1. Market share
		2. Customer Satisfaction
		3. Productivity
		4. Quality of new products/service
		5. Employee Satisfaction
		6. Flexibility

innovation management in their endeavour of improving firm's performance results.

To begin with, we realize a positive correlation market orientation and organizational performance. This result is aligned with previews studies presenting additional evidence to maintain the relevance of market orientation as a driver of organizational performance enhancement and consequently as a source of competitive advantage (Narver and Slater 1990; Kumar et al. 2011; Cano et al. 2004; Raju et al. 2011; Haugland, et al. 2007; Megicks and Warnaby 2008; Ramaseshan et al. 2002; Wood et al. 2000; Lonial et al. 2008; Kohli and Jaworski 1990). Secondly, our results support the relationship of market orientation and innovation management. This finding is related to previous studies concluding that firms, in order to be innovative, must rely on mechanisms of acquisition and leveraging of external knowledge from customers, competitors, suppliers, etc., achieving an interfunctional coordination (Weerawardenaa and O'Cassb 2004; Jiménez-Jiménez et al. 2008). Finally, our results support the direct effects of innovation management on financial factors, quality and satisfaction factors of organizational performance.

This work has some important academic implications. It examines an exploratory research model of the relationships between market orientation, innovation management and organizational performance, offering an aggregated perception of the above elements. Additionally, there is a managerial impact in this research. It shows the importance of supporting a market oriented strategy along with effective innovation management, to improve the organizational performance. This strategy could enchase firms with competitive advantages and sustainability. For managers this research provides some clear implications. With the intention of reaching better outcomes and performance, different kinds of information sources should be utilized. Most notably if the company is targeting to differentiate in the market through fostering innovation, focus should be put into market information that is currently available or exists in the firm's internal and external environment. Mainly, while our study highlights how market orientation and innovation management lead to differentiation strategy could enhance firms with competitive advantages and therefore, with sustainability.

This study has a number of contributions to make to marketing theory development, as it makes available useful approaching of the literature to managers. Academically, this research has a contribution to a better understanding of the interaction of the concepts of market orientation, innovation management, competitive differentiation strategy and organizational performance. It seems that there is a strong relationship between them, in the literature. Firstly, it is necessary to deepen concepts in order to test valid scales for industrial and service sectors, for large, small and medium-sized enterprises, as supported by several researchers (Liao et al. 2011; Lings 2004; Van Raaij and Stoelhorst 2008). Secondly, this study has some managerial implications concluding that a greater investment is recommended on marketing strategies to support the firms' market orientation so as to strengthen their market position, in return. Additionally, this implications, could lead managers to broaden the market orientation and organizational performance relationship, since innovation management has a key factor role in harmonizing the external turbulence that may concern them. Therefore, companies could overcome change in the market and technology so as to improve their performance. Organizational performance is influenced by market orientation and innovation management. Obviously, this study added to the previous studies, focusing on a new questionnaire of measuring market orientation, innovation management and organizational performance, considering the managers who quest for a sustainable competitive advantage. Market orientation is crucial to achieving greater levels of innovation, raising the level of firms performance.

5 Conclusion

This research endeavoured into exploring the firms' ability use to a sustainable performance strategy. Based this research, we portray three main conclusions in order to answer our research questions. To begin with, the most important implementation is the ability to generate the knowledge to overlap the changes in the market. Our analysis reveals that market orientation is the main means of gaining market intelligence as a source of capabilities and differentiation. Moreover, using innovation management as a process of generating new product/service, novelty, differentiation, business process renewal and organizational innovation could achieve better performance. The innovation management process could be more successful being the result of market orientation. Secondly, market information appear to have crucial role not only in innovation management but also on creating a differentiation strategy. Third, organizational performance could be enriched with satisfaction implications such as, customer satisfaction, human resource satisfaction, apart from financial and non-financial results. Finally, the implementation of the research concept into the organization's value chain creating a successful strategic impact.

For further research, it would be really interesting to apply this concept in Greece. Greek SMEs seem to be suffocated by the economic crisis and they have lost their ability to perform well. As a matter of fact, they should become more competitive and more customer oriented, in order to gain the needed information and knowledge for better organizational performance, using innovation as a mediator to success. This research could have an important managerial impact for Greek SMEs, considering the importance of the sustainable growth and performance, as they are implied in the literature review. Organizations do their best for the purpose of achieving profits that ensure the covering of their expenses, their survivals and maximization of their values (Kumar et al. 2011). Therefore, it is vital that managers identify and understand strategic orientations such as market orientation to enable a firm to achieve sustainable competitive advantages and improved organizational performance.

References

Akimova I (2000) Development of market orientation and competitiveness of Ukrainian firms. Eur J Mark 34(9/10):1128–1148

Al-alak BA, Tarabieh S (MZ) A (2011) Gaining competitive advantage and organizational performance through customer orientation, innovation differentiation and market differentiation. Int J Econ Manag Sci 1(5):80–91

Aldas-Manzano J, Kuster I, Vila N (2005) Market orientation and innovation: an inter-relationship analysis. Eur J Innov Manag 8(4):437–452

Avlonitis GJ, Gounaris SP (1997) Marketing orientation and company performance: industrial vs consumer goods companies. Ind Mark Manag 26:385–402

Avlonitis GJ, Gounaris SP (1999) Marketing orientation and its determinants: an empirical analysis. Eur J Mark 33(11/12):1003–1037

Avlonitis G, Papastahopoulou P, Gounaris S (2001) An empirically-based typology of product innovativeness for new financial services: success and failure scenarios. J Prod Innov Manag 18(5):324–342

Baker W, Sinkula J (2002) Market orientation, learning orientation and product innovation: delving into the organization's black box. J Mark-Focus Manag 5(1):5–23

Balkin DB, Markaman GD, Gomez-Mejia LR (2000) Is CEO pay in high-technology firms related to innovation? Acad Manag J 43(6):1118–1129

Baltzan P, Phillips A (2010) Business driven technology, 4th edn. McGraw-Hill Irwin, Boston

Bansal P (2005) Evolving sustainability: a longitudinal study of corporate sustainable development. Strateg Manag J 26:197–218

Beal RM (2000) Competing effectively: environmental scanning, competitive strategy, and organizational performance in small manufacturing firms. J Small Bus Manag 38(1):27–47

Bigne JE et al (2004) Market orientation: an antecedent to the industrial manufacturer's power. Eur J Mark 38(1/2):175–193

Bondarouk T, Friebe C-M (2014) Shared services—standardization, formalization, and control: a structured literature review. In: Bondarouk T (ed) Shared services as a new organizational

form. Advanced series in management, vol 13. Emerald Group Publishing Limited, Bingley, pp 39–65
Brown SL, Eisenhard KM (1995) Product development: past research, present findings, and future directions. Acad Manag Rev 20(2):343–378
Cadogan JW, Diamantopoulos A, Siguaw J (1999) Export market-oriented behavior, its antecedents, performance consequences and moderating factors: evidence from the U.S. and the U.K. RP9916
Cano CR, Carrillat FA, Jaramillo F (2004) A meta-analysis of the relationship between marketing orientation and business performance: evidence from five continents. Int J Mark Res 21 (2):179–200
Cavera A, Molla A, Sanchez M (2001) Antecedents and consequence of market orientation in public organizations. Eur J Mark 35(11/12):1259–1286
Caves RE, Ghemawat P (1992) Identifying mobility barriers. Strateg Manag J 13(1):1–12
Cepeda G, Vera D (2007) Dynamic capabilities and operational capabilities: a knowledge management perspective. J Bus Res 60:426–437
Damanpour F (1996) Organizational complexity and innovation: developing and testing multiple contingency models. Manag Sci 42(5):693–716
Damanpour F, Evan W (1984) Organizational innovation and performance: the problem of organizational lag. Adm Sci Q 29:392–409
Damanpour F, Schneider M (2006) Phases of the adoption of innovation in organizations: effects of environment, organization and top managers. Br J Manag 17(3):215–236
Damanpour F, Szabat K, Evan W (1989) The relationship between types of innovation and organizational performance. J Manag Stud 26(6):587–601
Damanpour FY, Gopalakrishnan S (1998) Theories of organizational structure and innovation adoption: the role of environmental change. J Eng Technol Manag 15:1–24
Dawes J (2000) Market orientation and company profitability: further evidence incorporating longitudinal data. Aust J Manag 25(2):173–199
Deng S, Dart J (1999) The market orientation of Chinese enterprises during a time of transaction. Eur J Mark 33(5/6):631–654
Deshpande R, Farley JU (1998) Measuring market-orientation: generalization and synthesis. J Mark Focus Manag 2(3):213–232
Dibrell C, Craig J, Hansen E (2011) How managerial attitudes toward the natural environment affect market orientation and innovation. J Bus Res 64(4):401–407
Erdil S, Erdil O, Keskin H (2004) The relationships between market orientation, firm innovativeness and innovation performance. J Global Bus Technol:1–11
Farrell MA (2002) Critique of the development of alternative measures of market orientation. Mark Bull: 31-13
Farrell MA, Oczkowski E (2002) Are market orientation and learning orientation necessary for superior organizational performance? Working Paper, vol. 52/02, pp 1–40
Farrelly F, Quester P (2003) The effects of MO on trust and commitment: the case of the sponsorship business–to-business relationship. Eur J Mark 37(3/4):530–553
Ferraresi AA, Quandt CO, dos Santos SA, Frega JR (2012) Knowledge management and strategic orientation: leveraging innovativeness and performance. J Knowl Manag 16(5):688–701
Fritz W (1996) Market orientation and corporate success: findings from Germany. Eur J Mark 30 (8):59–74
Ge GL, Ding DZ (2005) Market orientation, competitive strategy and firm performance: an empirical study of Chinese firms. J Glob Mark 18(3/4):115–142
Gelard P, Emamisaleh K (2014) The evolution of innovation types towards production performance. Int Bus Manag 8(4):222–228
Gima KA (1995) An exploratory analysis of the impact of market orientation on new product performance. J Prod Innov Manag 12:275–293
Gray B, Matear S, Boshoff C, Matheson P (1998) Developing a better measure of market orientation. Eur J Mark 32(9–10):884–903

Gunday G, Ulusoy G, Kilic K, Alpkan L (2011) Effects of innovation types on firm performance. Int J Prod Econ 133(2):662–676
Hajjat M (2002) Customer orientation: construction and validation of the CUSTOR scale. Mark Intell Plann 20(7):428–441
Han JK, Kim N, Srivastava RK (1998) Market orientation and organizational performance: is innovation a missing link? J Mark 64(October):30–45
Harris LC, Ogbonna E (2001) Leadership style and market orientation: an empirical study. Eur J Mark 35(5/6):744–764
Harrison-Walker JL (2001) The measurement of a market orientation and its impact on business performance. J Qual Manag 6:137–172
Hassim AA, Asmat-Nizam A-T, Abdul Rahim AB (2011) The effects of entrepreneurial orientation on firm organisational innovation and market orientation towards firm business performance. In: International conference on sociality and economics development, vol 10(2011). IACSIT Press, Singapore, pp 280–284
Haugland S, Myrtveit I, Nygaard A (2007) Market orientation and performance in the service industry: a data envelopment analysis. J Bus Res 60(11):1191–1197
Helfert G, Ritter T, Walter A (2002) Redefining market orientation from a relationship perspective. Eur J Mark 36(9/10):1119
Herath HMA, Mahmood R (2013) Strategic orientation based research model of SME performance for developing countries. Rev Integr Bus Econ Res 2(1):430–440
Hitt MA, Keats BW, DeMarie SM (1998) Navigating in the new competitive landscape: building strategic flexibility and competitive advantage in the 21st century. Acad Manag Exec 12
Hooley G, Fahy J, Greenley G, Beracs J, Fonfara K, Snoj B (2003) Market orientation in the service sector of the transition economies of the central Europe. Eur J Mark 37(1/2):86–106
Hult GT (2011) Market-focused sustainability: market orientation plus. J Acad Mark Sci 39:1–6
Jaworski B, Kholi A (1993) Market orientation: antecedents and consequences. J Mar Sci 57(3):53
Jaworski BJ, Kohli AK (1996) Market orientation: review, refinement, and roadmap. J Mark-Focus Manag 1:119–135
Jiménez-Jiménez D, Sanz-Valle R (2011) Innovation, organizational learning, and performance. J Bus Res 64(2011):408–417
Jiménez-Jiménez D, Sanz-Valle R, Hernandez-Espallardo M (2008) Fostering innovation, the role of market orientation and organizational learning. Eur J Innov Manag 11(3):389–412
Keskin H (2006) Market orientation, learning orientation, and innovation capabilities in SMEs: an extended model. Eur J Innov Manag 9(4):396–417
Kim HE, Pennings JM (2009) Innovation and strategic renewal in mature markets. Organ Sci 20:368–383
Kitsios F, Grigoroudis E (2014) Evaluating new service development effectiveness in tourism: An ordinal regression analysis approach. In: 3rd international symposium and 25th National Conference on operational research, June 26–28, (HELORS 2014), Volos
Kitsios F, Sindakis S (2014) Analysis of innovation strategies in hospitality industry: developing a framework for the evaluation of new hotel services in Thailand. In: 2nd international 16 conference on innovation and entrepreneurship (ICIE 2014), Bangkok
Kitsios F, Doumpos M, Grigoroudis E, Zopounidis C (2009) Evaluation of new service development strategies using multicriteria analysis: predicting the success of innovative hospitality services. Oper Res 9(1):17–33
Kitsios F, Champipi E, Grigoroudis E (2015) New service development: Strategy and innovation process in cultural and creative industries using MCDA. In: Proceedings of the 4th international symposium and 26th national conference on operational research, June 4–6, (HELORS 2015), Chania, Greece
Kitsios F, Grigoroudis E, Giannikopoulos K, Doumpos M, Zopounidis C (2015b) Strategic decision making using multicriteria analysis: new service development in Greek hotels. Int J Data Anal Tech Strat 7(2):187–202

Kohli AK, Jaworski BJ (1990) Market orientation: The construct, research propositions, and managerial implications. J Mark 54:1–18
Kohli AK, Jaworski BJ, Kumar A (1993) MARKOR: a measure of market orientation. J Mark Res 30:467–477
Kotler P (2010) Principles of marketing: a South Asian perspective, 13/E. Pearson Education India
Kumar K, Subramanian R (2000) Navigating the external environment through a market orientation. SAM Adv Manag J 65:16
Kumar V, Jones E, Venkatesan R, Leone RP (2011) Is market orientation a source of sustainable competitive advantage or simply the cost of competing? J Mar Sci 75(1):16–30
Lado N, Maydeu-Olivares A, Rivera J (1998) Measuring market orientation in several populations: a structural equations model. Eur J Mark 32(1/2):23–39
Laforet S (2009) Effects of size, market and strategic orientation on innovation in non-hightech manufacturing SMEs. Eur J Mark 43(1/2):188–212
Leal-Rodríguez AL, Ariza-Montes JA, Roldán JL, Leal-Millán AG (2014) Absorptive capacity, innovation and cultural barriers: a conditional mediation model. J Bus Res 67:763–768
Li JJ, Zhou KZ (2010) How foreign firms achieve competitive advantage in the Chinese emerging economy: Managerial ties and market orientation. J Bus Res 63(8):856–862
Liao SH, Chang WJ, Wu CC, Katrichis JM (2011) A survey of market orientation research (1995-2008). Ind Mark Manag 40(2011):301–310
Lings IN (2004) Internal market orientation. Construct and consequences. J Bus Res 57(4):405–413
Liu H (1995) Market orientation and firm size: an empirical examination in UK firms. Eur J Mark 29(1):57–71
Lonial S, Tarim M, Tatoglu E, Zaim S, Zaim H (2008) The impact of market orientation on NSD and financial performance of hospital industry. Ind Manag Data Syst 108(6):794–811
Low D, Chapman R, Sloan T (2007) Inter-relationships between innovation and market orientation in SMEs. Manag Res News 30(12):878–891
Lyon D, Ferrier W (2002) Enhancing performance with product-market innovation: the influence of the top management team. J Manag Issues 14(14):452–469
Matear S, Osborne P, Garrett T, Gray B (2002) How does market orientation contribute to service firm performance?: an examination of alternative mechanisms. Eur J Mark 36 (9/10):1058–1075
Mavondo F, Chimhanzi J, Stewart J (2005) Learning orientation and market orientation: relationship with innovation, human resource practices and performance. Eur J Mark 39 (11/12):1235–1263
Megicks P, Warnaby G (2008) Market orientation and performance in small independent retailers in the UK. Int Rev Retail Distrib Consum Res 18(1):105–119
Mitroulis D, Kitsios F (2014) Innovation management strategies for organizational performance. In: 3rd international symposium and 25th National Conference on operational research, June 26–28, (HELORS 2014), Volos
Mitroulis D, Kitsios F (2015) Competitive innovation strategies improving the organizational performance. In: Proceedings of the 4th international symposium & 26th national conference on operational research, June 4–6 (HELORS 2015), Chania, Greece
Naidoo V (2010) Firm survival through a crisis: the influence of market orientation, marketing innovation and business strategy. Ind Mark Manag 39:1311–1320
Narver JC, Slater SF (1990) The effect of a market orientation on business profitability. J Mark 54:20–35
Narver JC, Slater SF, Maclachlan DL (2004) Responsive and proactive market orientation and new product success. J Prod Innov Manag 21(5):334–347
Nasution H, Mavondo F, Matanda M, Ndubisi N (2011) Entrepreneurship: its relationship with market orientation and learning orientation and as antecedents to innovation and customer value. Ind Mark Manag 40(3):336–345
O'Brien JA, Marakas GM (2011) Management Information Systems, 10th edn. McGraw-Hill, New York, USA

Oczkowski E, Farrell FA (1998) Discriminating between measurement scales: the case of market orientation. Int J Res Mark 15:349–66
Panigyrakis GG, Theodoridis PK (2007) Market orientation and performance: An empirical investigation in the retail industry in Greece. J Retail Consum Serv 14(2):137–149
Pitts RA, Lei D (2003) Strategic management: building and sustaining competitive advantage, 3rd edn. Thomson Learning/South-Western College Publishing, Ohio
Porter ME (1980) Competitive strategy. Free Press, New York
Porter ME (1985) Competitive advantage. Creating and sustaining superior performance. The Free Press, New York
Porter M (2004) Competitive advantage: creating and sustaining superior performance (Newth ed.). The Free Press, New York
Pulendran S, Speed R, Widing RE (2003) Marketing planning, market orientation and business performance. Eur J Mark 37(3/4):476–497
Pumphrey A (2004) Developing a market orientated service organization–a theoretical model. Mark Strat 551:1–15
Raju P, Lonial S, Crum M (2011) Market orientation in the context of SMEs: a conceptual framework. J Bus Res 64(12):1320–1326
Ramaseshan B, Caruana A, Pang L (2002) The effect of market orientation on new product performance: a study among Singaporean firms. J Prod Brand Manag 11(6):399–409
Roberts PW (1999) Product innovation, product-market competition and persistent profitability in the U.S. pharmaceutical industry. Strateg Manag J 20(7):655–670
Salavou H, Baltas G, Lioukas S (2004) Organizational innovation in SMEs: the importance of strategic orientation and competitive structure. Eur J Mark 38(9/10):1091–1112
Sandvik I, Sandvik K (2003) The impact of market orientation on product innovativeness and business performance. Int J Res Mark 20(4):355–376
Sindakis S, Kitsios F (2014) Entrepreneurial dynamics and patient involvement in service innovation: developing a model to promote growth and sustainability in mental health care. J Knowl Econ. (Online publication date: 5-Dec-2014) Springer US: 1–20
Slater S, Narver J (1994) Does competitive environment moderate the market orientation—performance relationship? J Mar Sci 58(1):46–55
Sondergaard H (2005) Market-oriented new product development: how can a means-end chain approach aff ect the process? Eur J Innov Manag 8(1):79–90
Sørensen HE (2009) Why competitors matter for market orientation. Eur J Mark 43(5/6):735–761
Tajeddini K, Trueman M, Larsen G (2006) Examining the effect of market orientation on innovativeness. J Mark Manag 22:529–551
Tampoe M (1994) Exploiting the core competences of your organization. Int J Strat Manag Long Range Plann 27(4):66–77
Teece DJ (2007) Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strateg Manag J 28:1319–1350
Thornhill S (2006) Knowledge, innovation and firm performance in high- and low-technology regimes. J Bus Ventur 21(5):687–703
Tidd J (2001) Innovation management in context: environment organization and performance. Int J Manag Rev 3(3):169–197
Tomášková IE (2009) The current methods of measurement of market orientation. Eur Res Stud 12 (3):136–150
Tsiotsou RH, Vlachopoulou M (2011) Understanding the effects of market orientation and e-marketing on service performance. Mark Intell Plan 29(2):141–155
Van Raaij EM, Stoelhorst JM (2008) The implementation of market orientation—a review and integration of the contributions to date. Eur J Mark 42(11/12):1265–1293
Varela JA, Rio M (2003) Market orientation behaviour: an empirical investigation using MARKOR. Mark Intell Plan 21(1):6–15
Vazquez R, Rio AB, Iglesias V (2002) Consumer-based brand equity: development and validation of a measurement instrument. J Mark Manag 18(1/2):27–48

Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Quart 26(2):xiii–xxiii
Weerawardenaa J, O'Cassb A (2004) Exploring the characteristics of the market-driven firms and antecedents to sustained competitive advantage. Ind Mark Manag 33:419–28
Wolfe RA (1994) Organizational innovation: review, critique and suggested research directions. J Manag Stud 31(3):405–431
Wong SKS (2013) The role of management involvement in innovation. Manag Decis 51 (4):709–729
Wood J, Bhuian S, Kiecker P (2000) Market orientation and organizational performance in notfor-profit hospitals. J Bus Res 48:213–226
Yang J (2011) Innovation capability and corporate growth: an empirical investigation in China". J Eng Technol Manag 29:34–46
Zhou KZ, Kin C, Tse DK (2005) The effects of strategic orientations on technology-and market-based breakthrough innovations. J Mar Sci 69(2):42
Zhou KZ, Brown JR, Dev CS (2009) Market orientation, competitive advantage, and performance: a demand-based perspective. J Bus Res 62(11):1063–1070
Zolfagharian MA (2010) Linking market orientation to strategy through segmentation complexity. J Bus Econ Res 8(9):79–92

Decision Aiding Process in the Frame of the Strategic Farm Management

Evangelia Krassadaki and Nikolaos F. Matsatsinis

Abstract This chapter is focused on strategy tools adapted for the needs of the agricultural sector. The increasing complexity and high level of changes in the agricultural context and economic sector lead to the development of new tools and a new specialty, the strategic adviser. An adviser based on specific knowledge, skills and tools offers services at small or very small farm exploitations on strategy issues.

Strategic Analysis is the process of conducting research on the business environment within which an organisation operates and on the organisation itself, in order to formulate strategy. Definitions of strategic analysis often differ, but the following attributes are commonly associated with it, like (a) the identification and evaluation of data relevant to strategy formulation, (b) the definition and analysis of the external and internal environment and (c) a range of analytical methods that can be employed in the analysis.

Commonly used analytical methods in strategic analysis are SWOT, PEST, Porter's five forces, four corner's analysis, value chain, early warning scans, war gaming, etc. The subsequent sections of this chapter focus on strategic analysis for farms, as the business context. The PerfEA and the Risk Wheel are presented, as two useful tools for a strategic adviser. The tools are proposed in the European-level Leonardo da Vinci project (STRAT-Training), and are adapted for the farm sector. The tools support the adviser for (a) exploring farm environment, (b) managing risks and (c) setting up an integrated strategy.

In addition, some aspects in the context of decision aiding process are discussed. Thus, as a twofold effort herein, we discuss issues related with complex situation in agro-businesses, like risks handling, and via presenting farm strategy tools we make links with decision science and multi-criteria decision making.

Keywords Farm management • Strategic planning • Strategic management • Business analysis • Decision aid

---

E. Krassadaki (✉) • N.F. Matsatsinis
Decision Support Systems Laboratory, Technical University of Crete, Chania, Greece
e-mail: lia@ergasya.tuc.gr; nikos@ergasya.tuc.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_6

1 Introduction

Decisions are the coin of the realm in businesses, even farm exploitations. Every success or opportunity seized or missed is the result of a decision that someone made or failed to make. Making good decisions and making them happen quickly are the hallmarks of high-performing organisations, irrespective of the business size, how well-known the firm is or how clever the strategy is.

Strategic analysis is the process of conducting research on the business environment within which an organisation operates and on the organisation itself. Definitions of strategic analysis often differ, nevertheless the following attributes are commonly associated with it, like (a) the identification and evaluation of data relevant to strategy formulation, (b) the definition of the external and internal environment to be analysed and (c) a range of analytical tools that can be employed in the analysis.

In this framework, strategic planning involves the development of long (medium)-term strategies to increase the profitability and competitiveness of a business. In case of farm exploitations, either family-driven (small or very small size) or of bigger size, strategic planning may consider for example the development of a new activity (such as organic production), the merging of farms, the on-farm processing, the direct marketing of products to consumers, the energy production for the needs of a farm, a semi-industrial energy production unit, the efficient production of traditional commodities, etc.

In addition, the purpose of strategic planning process is to form a farm exploitation that allows individuals involved to achieve their personal goals. We often think of making money as our primary personal goal. However, people are complex individuals. Personal goals may include finishing a college degree, spending more time with family, creating a college fund for children, buying a vacation home, getting involved in an organization, doing charitable work, practicing sport, etc.

Personal goals may also focus on activities such as producing safe and nutritious products for consumers, providing employment for the entire family, providing an opportunity for the next generation farm members, etc. Subsequently, personal goals may differ among members of a family, due to their gender (Bańkowska 2005) or age (Harman et al. 1972) or education (Sahin et al. 2013) or experience (Van Kooten et al. 1986). All or part of the above goals express what in decision science we call preferences.

Preferences are ‘rational’ desires or are the basis of any decision aiding activity. There are no activities without preferences expressed as goals, opinions, likelihoods, judges, beliefs, etc. In the sense of decision theory, the first question is who is involved in decision making in farms or who the decision-maker in agro-business is. It is reported that even in cases that decisions are made by a single decision-maker (owner of a farm), they are made under the influence of other people’s opinions like members of a family, often called ‘*significant others*’ (Solano et al. 2001). Of course, there are cases that decisions are made on a group basis, where each member equally participates, where this last case is mostly applied in unions of farmers or even in family farms.

In this framework, setting goals for the strategy of a farm is a complicated issue. Decisions about agricultural production usually involve multiple goals, sometimes conflicting. In addition, the context that a farm operates is risky. There are numerous studies that focus on the multiplicity of agricultural producers' goals, motives and criteria (i.e. Hayashi 2000; Gasson 1973; Wallace and Moss 2002). Usually, there are various strategies (alternatives or actions) to implement, which i.e. can be ranked according to a set of criteria or assessed. It can be stated that the performance evaluation and selection of a strategy has multilevel and multi-criteria features, so it can be regarded as a Multi-Criteria Decision Analysis (MCDA) task or a Multi-Criteria Decision Making (MCDM) process. MCDM or MCDA, are disciplines of operations research aimed at supporting decision-makers who have to make numerous and conflicting evaluations. MCDA aims at highlighting these conflicts and deriving a way to come to a compromise in a transparent procedure.

The chapter is organized as follows: second section discusses the decision aiding process in the frame of farm management. In Sect. 3, first sub-section, an overview of the most commonly used strategic analysis tools is presented, while in second sub-section new tools specialized in farm strategy formulation are discussed. In Sect. 4 a pilot implementation is discussed briefly. In the last section, some concluding remarks are presented.

2 Decision Aiding Process

Usually, a strategic adviser guides work, rather than executing it. A career as a strategic adviser is consulting in nature, requiring expert skills and knowledge to advise companies on what they need to do and how to do it. After an adviser gives a detailed how-to plan to a business, he might remain on board to monitor, track and support the new initiative.

Taking into account farm exploitations, as the operational context, strategic advisers are consultants who help farmers to explore by themselves what they need to do and how to do it. A strategic adviser, specialized in farm exploitations, acts completely different in comparison with the usual aforementioned case of business advisers. In this last case, strategic advisers are simultaneously experts—analysts—facilitators and mentors who guide farmers.

A strategic adviser is considered as an expert because he/she knows and applies specific methods and tools adjusted for strategy formulation of the small and very small exploitations. A strategic adviser is considered as an analyst because he/she is exploring the current situation of a farm. A strategic adviser is considered as a facilitator because he/she supports the actively involvement of farm members during several meetings. Finally, a strategic adviser is considered as a mentor because he/she encourages participants and builds mutual respect relationships during a strong cognitive procedure. In contrast to business consultants, a strategic adviser never expresses personal opinions, ideas, proposals, preferences, etc. Mainly, his/her effort focuses on the exploration of the personal goals of farm members and on the conclusion of a strategy which is adjusted to their needs.

Management support of agricultural activities, although rare in most Southern European countries, like in Greece, is a promising field in Northern and Central European countries. In France such services are offered to farmers by specialized consultants, members of the Agricultural Chamber. In Denmark, such services are a 'must' for farmers either because the last are obliged to present an action-plan for taking a loan from a bank or because they are used to ask this kind of services. Similar services are offered at farmers in Germany, Holland, UK, etc. Behind the aforementioned initiatives is the idea that farm management is a complicated task, which requires other knowledge and skills beyond that of a farmer's. Even in countries where someone to become a farmer has to follow a specific educational path, like in France, specialized management services are offered to exploitations as these are offered to any other kind of enterprise.

In this framework, a strategic adviser offers his services locally during a series of workshops with farmers. The relationship between the adviser and decision-maker (farmer) is close and on a continuous basis for a specific period of time. The workshops are useful and initially focus on the identification and definition of the main '*problem*' of a farm.

As Dewey proposes, problem-solving procedure consists of five consecutive stages: (a) a felt difficulty, (b) the definition of the character of that difficulty, (c) suggestion of possible solutions, (d) evaluation of the suggestion, and (e) further observation and experiment leading to acceptance or rejection of the suggestion. Herbert Simon (1960) modified Dewey's list of five stages to make it suitable for the context of decisions in organisations. According to Simon, decision-making consists of three principal *phases*: finding occasions for making a decision, finding possible courses of action, and choosing among courses of action. The first of these phases is called '*intelligence phase*', the second '*design*' and the third '*choice*'. Another subdivision of the decision process was proposed by Brim et al. (1962, p. 9). They divided the decision process into the following five steps: (1) Identification of the problem, (2) Obtaining necessary information, (3) production of possible solutions, (4) evaluation of such solutions, (5) selection of a strategy for performance, and (6) implementation of the decision.

The approaches by Dewey, Simon and Brim et al., are all sequential in the sense that they divide decision processes into parts, which always come in the same order or sequence. Empirical surveys indicate that stages are performed in parallel than in sequence (Witte 1972). We can claim the same in the frame of the strategic farm management plus the unexpected circumstances of operations context. Indicatively, when an adviser cooperates with a farmer, they usually conclude on a set of problems, goals, risks, visions or challenges that formulate the core of a strategy-plan or the set of potential actions. If a sequential procedure is followed the initial '*intelligence phase*' will be followed by the '*design phase*', which comprises the problem modelling. In the specific case, as the whole procedure is interactive, the adviser supports every stage of strategy formulation, follows-on the whole process (during the process) and follows-up the implementation (after the process). Thus, the strategy formulation and implementation constitutes a dynamic activity, which is adjusted to known or unknown, expected or unexpected, circumstances. Therefore, the relation between *phases* is circular rather linear, not sequential as Mintzberg et al. (1976) proposed.

Usually, at high-risk contexts the follow-on services are appropriate for either regular support of activities or in case of unexpected circumstances. Additionally, the follow-up services are necessary for the effective implementation of a strategy for a specific period of time. These follow-on and/or follow-up services are essential when new issues emerge that may lead to a completely new strategy, the modification of an existent strategy or may cause changes of priorities, preferences, goals, initiatives, strategic actions, etc. Therefore, unknown or unexpected events (i.e. production destruction due to weather conditions) may lead to modifications in a dynamic way. This specific effort for follow-on and/or follow-up constitutes what the PerfEA methodology adopts, called as '*Continuous Improvement Cycle*' (for PerfEA see Sect. 3.2).

Moreover, it is noteworthy as Simon (1960, p. 2) claimed that executives spend a large fraction of their time in *intelligence activities*, an even larger fraction in *design activity* and a small fraction in *choice activity*. Mintzberg et al. (1976) concluded that the *intelligence phase* dominates the other two. In most businesses, strategic planning isn't about making decisions as Mankins and Steele (2006) claim. It's about documenting choices, often haphazardly. In contrast to this, by far the largest part of today's scientific research has focused on multi-stage analysis by using a combination of SWOT, MCDM and fuzzy sets theory (Celik and Peker 2009; Zaerpour et al. 2008; Zavadskas et al. 2011).

Furthermore, by far the largest part of literature on decision-making has focused on the *evaluation-choice* routine. Although many empirical decision studies have taken the whole decision process into consideration, decision theory has been exclusively concerned with the *evaluation-choice* routine. In this framework, many operations research scientists have tried to combine strategic analytical tools with multi-criteria methods. For this purpose SWOT and Analytic Hierarchy Process (AHP) appear mostly in literature. Usually, the data mined by SWOT analysis are applied as criteria values for MCDM. Multi-criteria methods like AHP (Saaty 1980), the integrated AHP—ELECTRE method (Kaya and Kahraman 2011), the AHP and Data Envelopment Analysis (DEA) as utilized by Sinuany-Stern et al. (2011), are usually applied.

Other interesting combinations of SWOT analysis with multi-criteria methods is the application of the Permutation method of feasible alternatives (Paelnick 1976) in combination with AHP, as these are applied for the selection of construction enterprises management strategy (Zavadskas et al. 2011). The evaluation criteria are selected by taking into consideration the objectives and interests of stakeholders. The algorithm is a feasible tool to aid in decision making for ranking alternatives.

3 Tools and Techniques for Strategic Analysis

One of the key skills of a strategic adviser is his/her capability to realise which analytical tools or techniques are most appropriate to the objectives of the analysis. In this sense, the most commonly used tools, like SWOT, PEST and Porter's five forces analysis, Four Corner's analysis, Value Chain analysis, Early Warning scans and War Gaming, are briefly presented in the following subsection. For a detailed description on these tools, see CIMA (2007).

Furthermore, tools like PerfEA and Risk Wheel, as these are proposed in the European-level Leonardo da Vinci project STRAT–Training focusing on farms, are discussed in Sect. 3.2. These tools are useful as the means for (a) exploring the context in which farms operate, (b) managing risks and (c) setting up an integrated strategy. For a detailed presentation of the STRAT-Training project, see https://strattrainingproject.wordpress.com/.

3.1 Commonly Used Tools and Techniques

SWOT analysis has been widely used in all areas of business management and strategic management. SWOT analysis is a simple tool that helps in understanding the *strengths*, *weaknesses*, *opportunities* and *threats* involved in a project or business activity. It starts by defining the objective of the project or business activity and identifies the internal and external factors that are important to achieving that objective. Strengths and weaknesses are usually internal to the organization, while opportunities and threats are usually external. Often these are plotted on a simple 2×2 matrix.

PEST analysis is a scan of the external macro-environment in which an organization exists. It is a useful tool for understanding the Political, Economic, Sociocultural and Technological environment that an organization operates in. It can be used for evaluating market growth or decline, and as such the position, potential and direction for an enterprise.

The Five Forces of competitive position analysis, as developed in 1979 by M. Porter of Harvard Business School, is a framework for assessing and evaluating the competitive strength and position of an organization. Strategic analysts use Porter's Five Forces to understand whether new products or services are potentially profitable. The five forces are: supplier power (an assessment of how easy it is for suppliers to drive up prices), buyer power (an assessment of how easy it is for buyers to drive prices down), competitive rivalry (number of competitors and capabilities), threat of substitution and threat of new entry.

The Four corner's analysis, developed by M. Porter, is a useful tool for analyzing competitors. It emphasizes that the objective of competitive analysis should always be on generating insights into the future. The Four corners' refers to four diagnostic components, which are essential to competitor analysis, like: future goals, current strategy, assumptions and capabilities.

Before making a strategic decision, it is important to understand how activities within the organization create value for customers. One way to do this is to conduct a Value Chain analysis. The analysis is based on the principle that organisations exist to *create value* for their customers. Thus, the organisation's activities are divided into separate sets, which add value. Each value adding activity is considered as a source of competitive advantage. The steps for the analysis are: Separate the organisation's activities into primary and support activities, allocate cost to each activity, identify the activities that are critical to customer's satisfaction and market success.

The purpose of strategic Early Warning systems is to detect and predict strategically important events as early as possible. They are often used to identify the first step of attack from a competitor or to access the likelihood of a given scenario becoming reality. The seven key points of an Early Warning system are: market definition, open systems, filtering, predictive intelligence, communicating intelligence, contingency planning and finally cyclical process.

War Games are a useful technique for identifying competitive vulnerabilities and misguided internal assumptions about competitors' strategies. Simulations of these scenarios are used to explore the implications of changes in strategy in a no-risk environment. This technique is particular useful for organisations facing critical strategic decisions.

3.2 STRAT Proposed Methodologies and Tools

3.2.1 The PerfEA Methodology

The PerfEA methodology (Perfomance globale des Exploitations Agricoles), was developed as part of a French research project, driven by VetAgro Sup in partnership with SupAgro Florac and APCA-Resolia. The method was tested and adjusted via a pilot application in seven farms in the agricultural schools of the Auvergne region (France). For the PerfEA, see http://perfea.org/wakka.php?wiki=Accueil.

The main consideration for the development of PerfEA is that agriculture is a complex activity, which leans on natural and human resources aiming to create added value in a constantly evolving context. Farmers have to adapt their activities to several issues, like climate changes, agricultural policies, disappearance of market regulations, etc. From this point of view, strategic management is considered as a relevant approach, which can support farmers to manage effectively their property.

The basic idea behind PerfEA is that farm members have to address challenges for a sustainable development of their unit. In order to achieve this they have to think in a strategic manner, which is based on a team effort between members and an expert (strategic adviser).

The analytical framework of PerfEA is based on a *Continuous Improvement Cycle*. The *Continuous Improvement Cycle* is based on Deming *Plan-Do-Check-Act* model. This four-step model as a circle has no end. It is repeated again and again for continuous improvement. Thus, the adviser offers services adjusted to regular or irregular circumstances. He/she follows-on and follows-up the whole process and adjusts the strategy formulation.

Based on the diagnostic analysis that farmers perform of their situation and of the changes in the operational context, this analytical framework allows them to establish and implement a continuous improvement project. Therefore, PerfEA aims to provide the necessary means to manage effectively the farm-activity and involve participants in a process of continuous improvement.

The whole procedure is supported by an experienced adviser. The adviser is actively involved in a series of meetings with farm members, organized in their property. On a team-based activity the adviser following PerfEA methodology tries to find-out answers during open discussions with farmers. On average, seven meetings are appropriate for concluding in a strategy formulation. The follow-up services are appropriate for a certain pre-agreed period of time. Usually, a period of 1 year is enough.

The PerfEA is a three-step methodology which comprises the following: (a) analysis and strategic thinking, (b) strategy formulation and (c) strategy implementation (see Fig. 1). Several tools may utilized during a strong cognitive procedure between members of a farm and an adviser, like the Global Performance Wheel, the SWOT analysis, the cognitive maps and the Balanced Scorecard; all of them focusing on the exploration and formulation of strategy.

The first step of PerfEA initially focuses on identification of stakeholders of the exploitation; those individuals or entities directly or indirectly involved in the farm. The '*Plan of Borders*', a visualized representation like in Fig. 2, is used to determine stakeholders and existent relations between them and the farm. This effort focuses on the clarification of the context in which the farm evolves. Two circles separate the internal and external border. The internal border includes owners, decision-makers of the exploitation. Between the internal and external border there are the employees, who are actively involved, not as decision-makers. While beyond the external border there are agents, organisations or persons that affect the farm, like customers, suppliers, etc. Therefore, the '*Plan of Borders*' aims to diversify stakeholders by the roles they really have and to clarify the existing relations between the internal and external environment.

Whenever the stakeholders are clarified, an analysis of current situation follows. It aims at checking the sustainability of the farm exploitation. Farmers in cooperation with the adviser define past successes and failures, values that lead the organization, consider the enterprise's missions and owners' vision for the near future in a 3–4 years period. All relevant information is collected by the adviser and is organized into a draft cognitive map.

More specifically the first step of PerfEA, as a strong cognitive procedure and time-consuming is analysed into three separate phases:

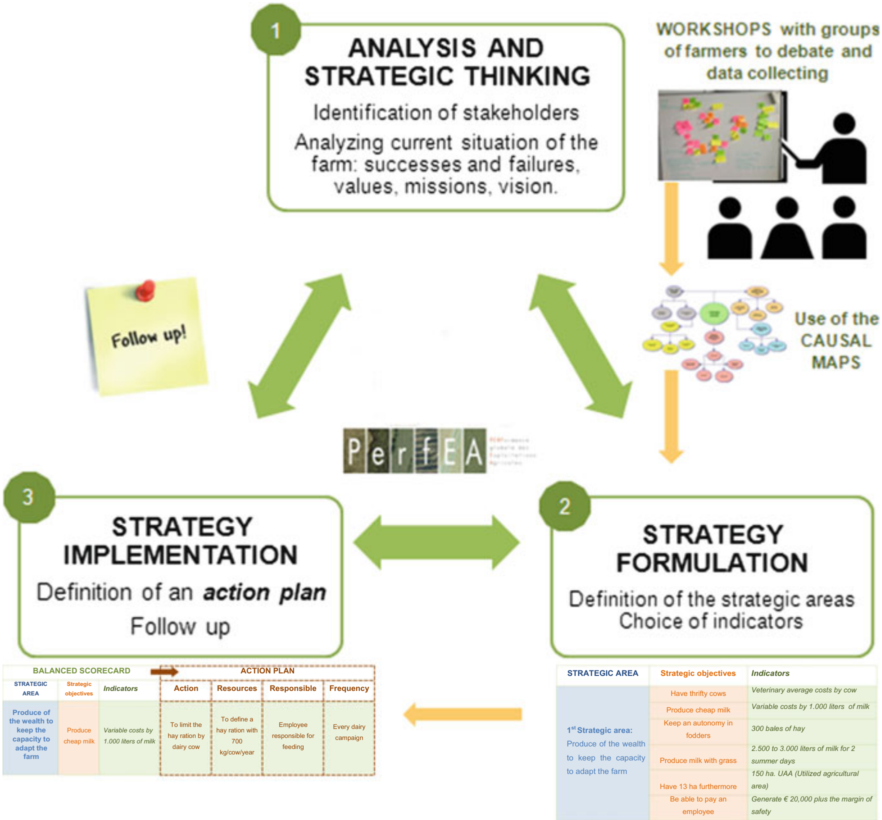


Fig. 1 The three steps of PerfEA (*source*: STRAT project)

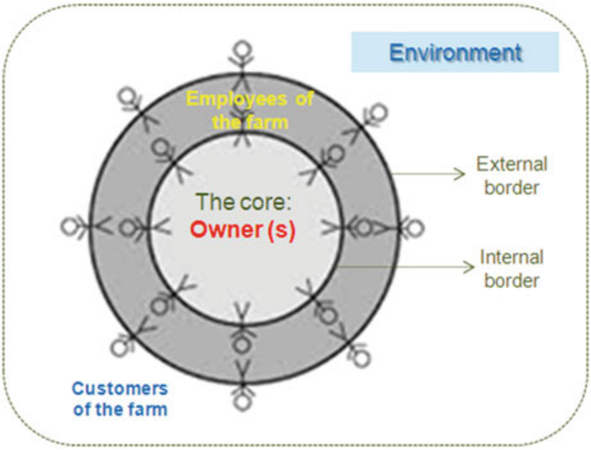


Fig. 2 The plan of borders (*source*: STRAT project)

- Phase 1. Identification of stakeholders, as it is already discussed.
- Phase 2. Analysis of current situation and durability of the exploitation by using a Global Performance approach.
- Phase 3. Data collection and organization of the information.

In this framework a series of debates, discussions, and exploration activities take place. The analysis of current situation and durability of the exploitation is based on the Bossel's (2001) principles for sustainable development, which are the following:

- Existence
- Effectiveness
- Security
- Adaptability
- Freedom of action
- Coexistence
- Transmissibility
- Psychological needs
- Responsibility

The specific effort aims at highlighting the main processes of management, favorable or unfavorable, which guarantee the global performance of the exploitation. For this purpose, the adviser raises a number of questions related to each of the Bossel principles, as indicatively some of them are presented in the Appendix. Based on the answers, the adviser prepares a specific *Global Performance Wheel* (see an example in Fig. 3). The *Global Performance Wheel* is a visual representation of the nine Bossel's principles, adapted in the farm under consideration. An additional objective behind this effort is the development of a strategic way of thinking of farmers, when they make by themselves a diagnosis of the current situation of the farm.

The description of the current situation includes incentives and perspectives of farmers, successes and failures of farm, as well as values and missions. The values constitute major principles, which structure the identity of the farm exploitation, like i.e. *farm profitability versus productivity*. The missions justify the existence of the exploitation (i.e. *why does the farm exist? To keep the family farm*); while the vision describes what farmers want to be in the future. Therefore, the vision as a '*signal*' clarifies the purpose of the whole strategic project. For example, a vision could be something like '*to become a benchmark exploitation in the local market*'. In this respect, successes and failures, values and missions constitute a SWOT 2×2 matrix (see Fig. 4). The SWOT analysis, as a useful and simple tool, is utilized by the adviser for organizing information and for prioritizing *strategic objectives* (i.e. *production of cheap milk, secure the farm*).

All the above information is utilized for the development of draft cognitive maps, which allow a visual representation of concepts. Maps are designed by the adviser on the basis of the information he/she has gathered during meetings and the analysis made. Thus, taking into account several *key strategic elements* like

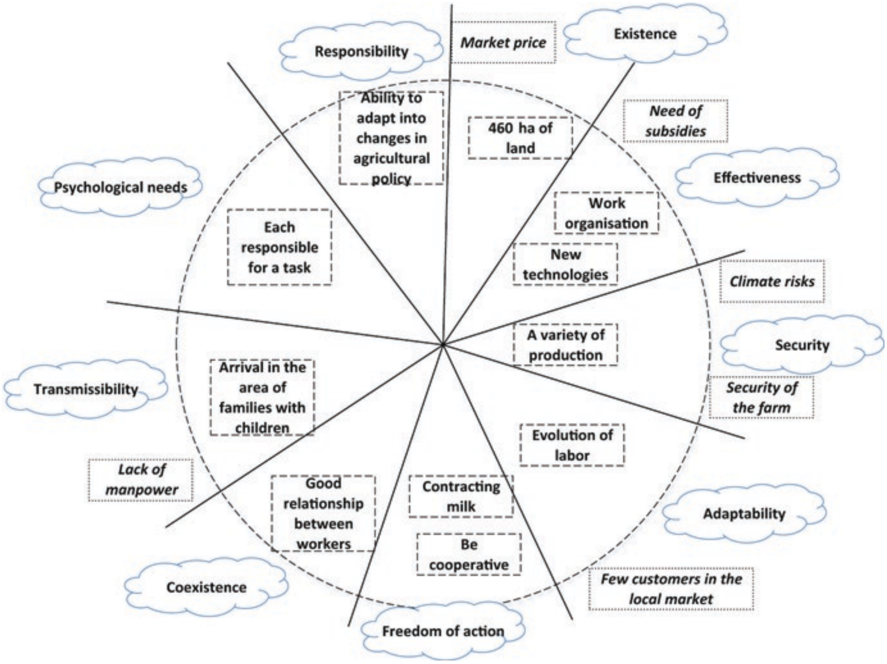


Fig. 3 The global performance wheel (*source*: STRAT project)

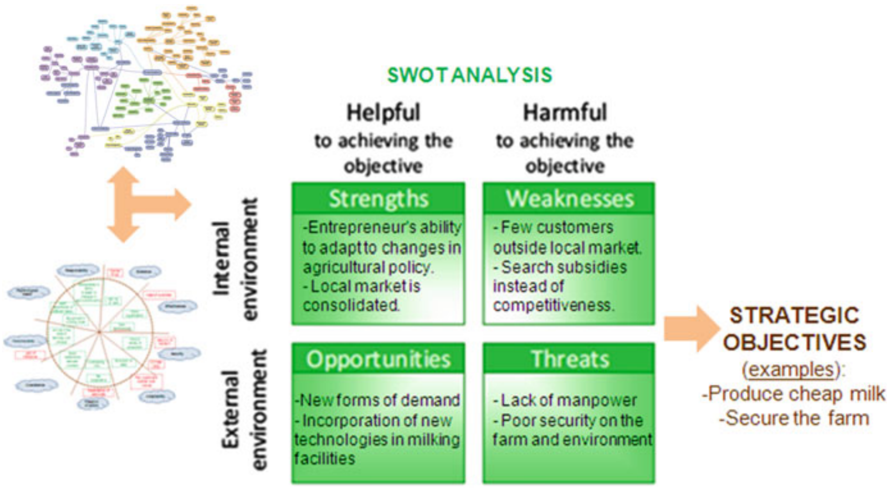


Fig. 4 A SWOT analysis example

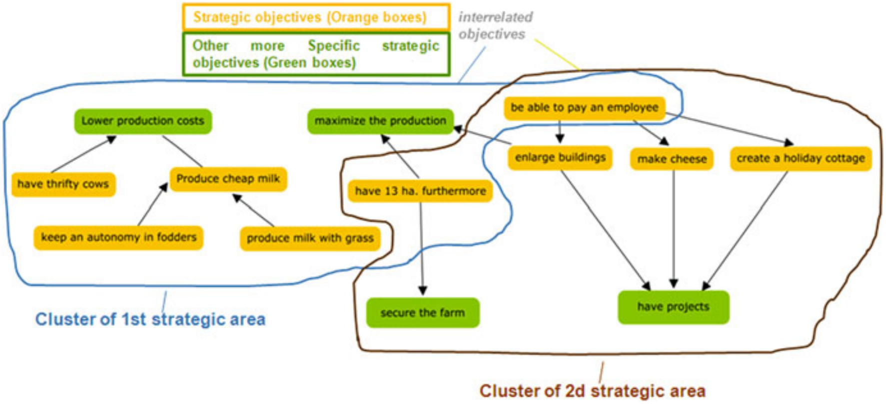


Fig. 5 Part of a cognitive map presenting two strategic areas (*source*: STRAT project)

successes and failures, missions and vision the adviser initially represent graphically the whole set of *strategic objectives* of a farm.

The second step of PerfEA focuses on the strategy formulation. The formulation comprises the following:

(a) Definition of *strategy areas*, as well as
(b) Identification of measurement *indicators*.

For the identification of *strategic areas* (or final strategy) a new cognitive map is developed, based on the initial draft map. The final strategy consists of few *strategic areas*. As it is presented in Fig. 5 two *strategic areas* are identified. Each one is based on a number of *objectives*. The various *Strategic objectives* (orange boxes) are linked to each other and are synthesized to few *Specific strategic objectives* or goals of the farmer (green boxes). For example, the strategic objectives '*enlarge buildings*', '*make cheese*' and '*create a holiday cottage*' are linked together focusing on a specific strategic objective '*have projects*'. Or the strategic objectives '*have thrifty cows*' and '*produce cheap milk*' constitute the specific strategic objective '*lower production cost*'. The whole process is based on a synthesis of various strategic objectives to goals that farmers would like to achieve (reduce the production cost, maximize the production, secure the farm, have projects).

Based on the goals of the farmers, as these are represented by green-color boxes, the strategic areas can be defined, which constitute the strategy of the exploitation. According to the example of Fig. 5 the first *strategic area* may be something which is derived by the two specific objectives '*lower production cost*' and '*maximise the production*'. Therefore, the farmer and the adviser can describe this strategic area as '*produce wealth in order to adapt the farm*'. Similarly, the second *strategic area* that combines both specific strategic objectives '*secure the farm*' and '*have projects*' it could be described as '*towards a mastered, secure, adaptive and innovative farm*'. It is worthwhile to mention that different display formats in a cognitive map, like different colors, communicate patterns within the model, like in the example (green-color for specific strategic objectives or goals of the farmer, orange-color for detailed strategic objectives).

Since each *strategic area* has been analysed in a set of detailed *objectives* (orange-color items in the cognitive map) the aim is to choose appropriate *indicators* for measuring them. Therefore, for each detailed objective an *indicator* is chosen based on a discussion between decision-makers (owners) of a farm-unit. Some features of these kind of *indicators* are:

- Indicators can be financial or not, short or long term, qualitative or quantitative, retrospective or prospective.
- Indicators represent a certain forward-looking vision.

For example, for the first or second *strategic area,* which both contain the detailed *strategic objective* '*be able to pay an employee*', the selected *indicator* may be something like '*income of 20,000 €, plus the margin of safety'.*

Subsequently, a Balanced Scorecard approach can be utilized at the end of the second step of PerfEA. Balanced Scorecard, as a strategy performance management tool, is adapted herein for the needs of PerfEA. Consequently, the tool describes for each *strategic area* the detailed *strategic objectives* and for each of them an *indicator*. An example is presented at Table 1. Furthermore, the adapted Balanced Scorecard as it is proposed herein, constitutes a useful tool for the next third step of PerfEA in order to develop an *Action Plan*.

The last third step of PerfEA focuses on the strategy implementation. This step includes two actions. The first action is associated with the development of an *Action Plan*, which is based on the adapted Balanced Scorecard (see Table 2), while the second action constitutes the *follow-up* procedure.

Table 1 Example of a balanced scorecard-like table

Strategic area	Strategic objectives	Indicators
First strategic area: Produce wealth in order to adapt the farm	Have thrifty cows	Veterinary average costs per cow
	Produce cheap milk	Variable costs by 1000 l of milk
	Keep an autonomy in fodders	300 bales of hay
	Produce milk with grass	2500–3000 l of milk for 2 summer days
	Have 13 ha furthermore	150 ha UAA (utilized agricultural area)
	Be able to pay an employee	Income 20,000 €, plus the margin of safety

Table 2 An example of an action plan combined with a balanced scorecard-like table

Balanced scorecard			Action plan			
Strategic area	Strategic objectives	Indicators	Action	Resources	Responsible	Frequency
Produce wealth to keep the capacity to adapt the farm	Produce cheap milk	Variable costs by 1000 l of milk	To limit the hay ration per dairy cow	To define a hay ration 700 kg/cow/year	Employee responsible for feeding	Every dairy period

An *Action Plan* is a set of coherent actions for achieving one or several objectives, which can improve the farm's Global Performance. Therefore, farmers in cooperation with the adviser develop an action plan. The adviser in order to facilitate the task of identifying actions, it is necessary to find answers in the following questions:

- Action: type, purpose.
- Resources: which are the necessary resources to implement an action.
- Responsible: who is responsible for the realization of an action.
- Frequency of control: when this action is realized.

The *follow-up* action is a team-based responsibility of workshops' participants, farmers and adviser. It is related with the evaluation of performance and continuous improvement. For ensuring the implementation of a strategy it is necessary to track the proposed actions. The steps are:

- At the end of the time period established for each of the proposed action, the responsible person must check that every action is made properly. If actions are not completed, he/she will have to think of other measures for achieving selected objectives (i.e. produce cheap milk).
- The changes must be communicated to the rest of the organization. In this way all members of the farm know the achievement level of each strategy.

The main advantage of the PerfEA is the full exploration of an exploitation and its context, and via the step-by-step procedure the conclusion of major strategies. Strategy areas that influence the decision making are related to an understanding of the reality of organizational life. This merely cognitive procedure for formulating a strategy in a group-basis deserves some more comments and links with decision science and group-decision making.

The adviser's role as a facilitator for mediating, helping, coaching, mentoring, etc. the whole project is completely distinct and appropriate as well as necessary for both the follow-on (during the process) and follow-up (after the process) period. This role is time consuming and can be described as '*neutral*'. This means that he/she usually listens carefully, facilitates or manages discussion, handles objections, enforces dialogue, and tries to explore strategies in cooperation with participants. He/she may come back to the participants for verification of workshops' conclusions but never expresses personal opinions, judges, preferences, critiques, etc.

The two roles of the main actors, adviser and decision-maker, are completely distinct and differentiated. Sometimes the 'client' may be an official (manager) of the agro-business, not necessarily being a decision-maker; he might not have decision power and may be for instance himself an analyst of the owner(s) part. For simplicity purposes we can claim that PerfEA lies on a pair of actors: the adviser and the decision maker(s). This last distinction is not always the case at decision making process, where both roles of the analyst and the decision-maker may be undertaken by the same person.

The two actors are in a continuous interaction for a period of time. The results of this interaction are based on the methodological knowledge, communication and technical skills of the adviser as well as the domain knowledge of the participants-farmers. As Tsoukias (2007) claims this decision aiding process is a special type of decision process called as '*consensual construction of a client's concerns representation*'. The difference between decision making and decision aiding has already been discussed by Roy (1993), although Roy considers these as two different approaches and not as different situations as Tsoukias (2007) claims. The whole constructive approach for the strategy formulation and the handling of each strategic objective by choosing measurable indices is a step ahead of those decision aiding approaches focusing either on the problem formulation or the evaluation model.

As it is briefly discussed herein, we can claim from a practical point of view that the three-step methodology integrates both, the *problem formulation* and the *evaluation model*. For the last, although nonspecific modelling or algorithm is applied, the decision-maker(s) throughout the process explores or assesses alternatives, expresses preferences, makes choices or proposals, rejects or accepts actions, concludes on objectives in a rational-verbal-interactive way. Therefore, the final product of this effort is a representation of the decision-maker's concerns and decisions.

Many issues emerge during workshops, like concerns for the future, available resources, etc. all of a descriptive nature. For the needs of problem structuring the cognitive maps are used. A cognitive map is a representation of the perceptions and beliefs of an individual about his own subjective world. Cognitive maps are initially proposed by Eden (1988, 1994), as a visual interactive method for managing complexity in settings where the decision making effort is rather client-oriented than solution-oriented. In this case, the concepts of a map do not have any formal properties and do not necessarily define a structure, like as a hierarchy. Simply concepts represent what participants know or wish to achieve.

Recently, attention has been increasingly directed towards problems of strategic decision-making and decision making in turbulent environments. It is recognized that purely analytical approaches are inadequate to deal with such problems. Furthermore, Klein and Cooper (1982) use cognitive mapping for analyzing decisions during a research wargame. It is worthwhile to mention herein that PerfEA uses cognitive maps, which express the 'subjective' world of an individual (in cases of one-person farming) or of a group (in case of family-farming or agro-businesses). This last case, considered as a '*consensual cognitive mapping*', is the more often applied since one-person farm units are more rare.

All the effort evolves on a group-basis from the early beginning until the last step of strategy implementation and follow-up. During this process, decision-makers, may be either single persons (one-person farming) or many (i.e. family-farming, agro-business). In cases of family-farming PerfEA suggests that decisions are made on a *consensus basis* of the equally participated owners of the farm. Of course in real contexts, it is not rare that decisions are made by either the man of the family or the older person of the holding, or whoever has more influence on the others ('*significant others*'). In respect the agro-businesses, where the organizational structure is more complicated, PerfEA suggests a commitment (contract) of the organisation with the adviser (or the organization he/she works for). This commitment tries to 'solve' issues like: who decides for what problem, what is the commitment of the staff members on the strategy formulation, what is the role of each one within the process, etc. Or tries to 'solve' issues like who participates in what step of the process, the strength they may have, the consequences of their opinion, etc.

Therefore, we cannot generally claim that members of an agro-business equally participate or all of them participate, or to declare who participates in what step. In this sense, we cannot for sure claim that group decision-making is a clear mapping of the reality for strategy formulation in agro-businesses, since is not clear if members participate during all the steps of strategy formulation, as is the case in group decision-making.

In contrast to the aforementioned claims, one well-known group decision support system like the Delphi (Linstone and Turoff 1975), supports a group of anonymous experts, who equally express their opinion during several rounds of communications with a director. The experts' responses are collected and evaluated by the director. The result of the processed information will again be distributed among the panel members. Each member receives new information and comment on them or revise his own opinion. The director will decide when to stop the process, based on the level of desired consensus. In conclusion, even the Delphi step-wise process presupposes equal participation of experts (decision-makers) and the advanced role of a director. Nevertheless, the strategy formulation, as it is described in the PerfEA method, is a more complicated issue so we claim that partially could be represented as a group decision-making problem. Indicatively, for group decision-making methods, tools and algorithms the reader is referred on Matsatsinis and Samaras (2001) as well as Matsatsinis et al. (2005).

3.2.2 The Risk Wheel Tool

The most common sources of risk in farming are production, marketing, financial, institutional and human. Crop and livestock performance depend on biological processes that are affected by the weather, and by pests and diseases (production risks). When farmers plant crops or commit resources to raising livestock, they do not know for certain what prices they will obtain for their products. In situations of low rainfall, production of crops is often reduced and, as a result, prices rise (marketing risks). Financial risk occurs when money is borrowed to finance the farm business. Smallholder farmers, who borrow money at high interest rates may have particular difficulty making debt repayments. Institutional risks occur when there are unpredictable changes in the provision of services from institutions that support farming. Such institutions can be banks, cooperatives, marketing organizations, etc. Human risks refer to health problems that affect farm activities, which are usually caused by illness or death.

In general, managing risks as a step-wise process requires the following: identify the possible sources of risk, identify possible outcomes that could occur as a result of risks, decide on alternative strategies available, assess consequences or results of each possible outcome for each strategy, and evaluate trade-offs between the cost of risk and gains that can be achieved.

Good risk management decisions depend on accurate information, which in turn, requires reliable data. The sources for such data are usually the following:

- Farm records, like data on production, yields, use of fertilizers or seeds, labor.
- Off-farm information, which i.e. comes from statistics, newsletters, other farmers, traders, suppliers.
- Other information, which is related to issues for food safety or water quality or animal health, or environmental legislation, etc.

Indicatively, actions that prevent from production risks, in general, may be something from the following or combination of them:

- Diversification. The production diversification is a useful way to prevent farm income from unexpected events, which may have negative effects in production or operations. Such activities may be the following: the engaging in the same farm enterprise in different physical locations, the managing of multiple farm enterprises together at any one time, the engaging in the same farm enterprise over successive periods of time, the generating of income that comes from off-farm activities, etc.
- Share leases. Under this scheme the owner of the land pays part of the operating expenses and, in return, receives from the tenant a portion of the crop or livestock. In this way all production risks are shared between the tenant and the owner.
- Custom farming. Custom farming involves a farmer entering into an agreement with an operator to carry out various farm operations. It is also referred to as 'contractor' farming. The advantage is that operational costs can be fixed. For

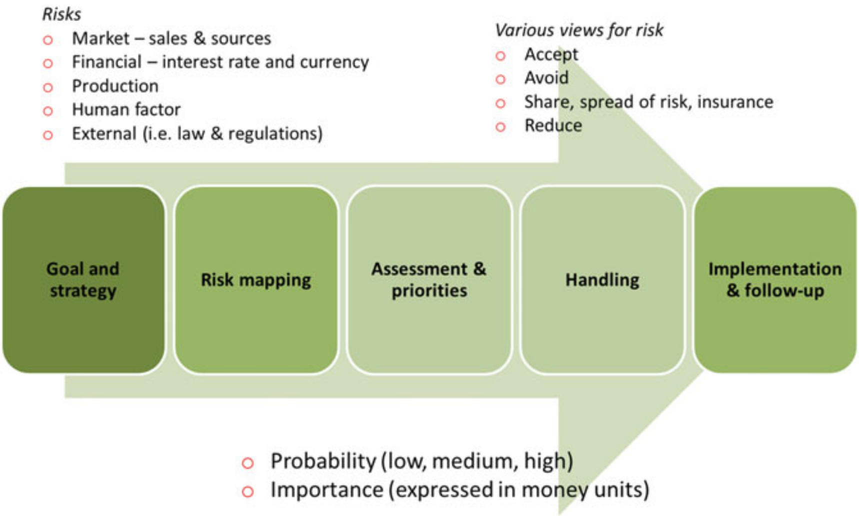


Fig. 6 The five steps of the Risk Wheel tool

instance, instead of facing the risk of high equipment costs, the farmer contracts someone else to do this work.

In particular, the Risk Wheel tool provides an overview of risk factors and suggests a prioritization of efforts in a farm exploitation. It is based on an Excel Add-in software and is developed by DAAS—Danish Agricultural Advisory Service (Denmark) for the needs of the STRAT project. The Risk Wheel tool allows the collection, presentation and evaluation of information towards a Holistic Risk Management approach.

The tool is effective when farm owners have already begun the implementation of a specific strategy and would like to have an overview of possible risk factors that may influence their target or may have a negative effect on their goals. Such a strategy may focuses on a new serious change like a new production or expansion of a farm or application of a new technology, or merge of farms.

Risks may emerge from sources like, prices' fluctuation, climate changes, weather condition, production, interest rates, exchange rates, new legislation, human factors, etc. The objective is to assume every risk on a '*gainful*' manner and adapt the farm strategy effectively via an action plan. Moreover, as Kahan (2008) suggests the time is another source that may include risk. In particular, according to Kahan "*the time between when a decision is made and when the outcome or consequence of that decision is experienced also affects risk. The farmer often needs to integrate what are called short-term tactical decisions with longer-term strategic decisions. Time also influences the usefulness of information used in decision-making. The ability of the farmer to respond to events is also affected by time. These aspects of time make assessing risk more complex*".

The Risk Wheel tool comprises the following five steps (see Fig. 6):

1. Definition of goal and strategy
2. Risk mapping
3. Assessment the impact of risk factors (importance) and their probability to happen
4. Handling risks (accept, avoid, transfer-share, reduce), and
5. Implementation and follow-up.

Initially, the goal and strategy are declared. The tool is applied when a strategy has already confirmed and the farmer's goals have already specified like i.e. '*towards* a *profitable and sustainable agriculture plus a good workplace environment*'.

The tool proposes five major Risk Categories: Market, Financial, Production, Human, and External. The farmer in cooperation with a strategic adviser concludes on specific risk issues per category. For this reason the farmer answers a series of questions raised by the adviser. In preparing for an in-depth discussion of risk, the adviser should have a clear picture of the farm and farm households in the area, in addition to an understanding of farming activities and practices.

For example, changes in prices are beyond the control of any individual farmer. The price of farm products is affected by the supply of a product, demand for a product, and the cost of production. In addition, a market risk may be caused by the rising prices of raw materials, a financial risk could be the interest rates or current liquidity, while a production risk may be related with an animal disease. A human risk may be caused by an accident, illness or death of farm members, which can disrupt performance, while an external risk may be the elimination of production quotas due to EU regulations. For this reason, via an exploratory procedure the farmer in cooperation with a strategic adviser describes each risk issue that may affect the existent strategy of the farm. This step is called Risk mapping.

During the third step (Assessment and Priorities) a probability is assigned for each risk factor using a three-level scale, '*low*'—'*medium*'—'*high*'. The probabilities are expressed by the farmer and constitute a prioritization of risks factors. Thus, the probability-scale describes how likely the risk event or condition is to occur. In this sense, factors which are characterized of *high* probability are considered as of primary interest in respect with others of *medium* or *low* probability.

At the same step, a second assessment is necessary. This is related to the *significance level* for each risk factor in the sense of the *impact* it causes. The *importance* of each risk factor is expressed in money units. This quantification of *impact* offers evidence, facilitates decision making and creates a sense of confidence to overcome a risk. Let's say that a farm member will face a health problem. This risk is characterized as of '*medium*' probability to occur. However, if a replacement is necessary by hiring a new employee in the farm, the economic impact is estimated 12,000 € per year. In this way, probability and significance level (economic impact) are assigned for each risk factor, respectively.

The forth step of the tool deals with the Handling of risks. The proposed choices for handling risks are either to (a) *accept* or (b) *avoid* or (c) *transfer-share* or (d) *reduce* a risk, respectively. For example, a life insurance may be a choice to

Table 3 An example of risks per category

Strategic objective		Evaluate risks of merging two farms			
Risk taking		High willingness to take risks			
Risk capacity		Limited due to previous investments			
Risk		**Description**	**Probability**	**Importance** (€)	**Action**
Market	Volatility in prices	Raw material prices	*High*	*21,000*	*Reduce*
Financial	Loans	Payment capacity	*High*	*15,000*	*Accept*
Production	Crops	Rent new land	*High*	*4000*	*Accept*
Human	Service replacement	Replace a partner due to health problems with an employee	*Medium*	*12,000*	*Accept*
External	Environmental restrictions	Elimination of production quotas	*Medium*	*50,000*	*Reduce*

transfer-share a specific risk. At this step, the farmer's risk tolerance is a useful information. The risk tolerance may be something from the following:

1. Low willingness to take risks (risk-aversion). The risk-averse farmers try to avoid taking risks.
2. Neutrality on taking risks (risk-neutrality). Risk-neutral lies between the risk-averse and risk-taking position.
3. High willingness to take risks (risk-taking). The risk-takers prefer to take a chance to achieve higher income.

Therefore, the personal willingness to take risks affects the several ways of handling.

Simultaneously, the Risk Capacity of the exploitation is taken into account. The Risk Capacity is based on whether risk taking is limited or not, and is depended on factors such as: if farmers already have significant past investments, if they have got several loans, etc. A short example follows in Table 3. As it is presented in the example the strategic objective is related with the evaluation of risks due to the merging of two farms. In this case the farmer is willing to take risks but the risk capacity is limited due to previous investments. The detailed risks for the farm-unit are presented per category at the second and third columns of the Table. Indicatively, the *volatility in raw material prices* is considered by the farmer as a risk of *high probability*, which may causes an expenditure of 21,000 €. This is proposed as a *reduced*-type of risk. In other words, the farmer concludes to follow some actions, which can *reduce* the specific risk. For this reason, he may either look for new type of raw-materials, instead of those already used, or/and new markets (suppliers). Despite the fact that the farmer is willing to take risks, the capacity of the exploitation is limited, so specific actions will be undertaken in order to reduce as many risks as possible.

Headline	Description	Economic effect DKK.	Assesment of probability	Management of risks and actions
Milk	Drop in milk yield	300	Medium	Accept
	Diseases among the herd	300	Medium	Reduce
Crops	Dividend and quality of roughage	300	Medium	Accept
	Key employee leaves	300	Low	Reduce
Price	Volatility in prices of concentrated feed and raw materials +/- 25 DKK/hkg	220	High	Reduce
	Volatility in prices of sales crops +/- 25 DKK/hkg	150	High	Reduce
	Volatility in prices of milk +/- 10 cent DKK	295	Medium	Accept
Human	Dead, testimony has not been written yet		Low	Share/prevent
	Incapable of working as a result of an accident or attrition	500	Low	Share/prevent
Financial	35 mio DKK in floating-rate loans DKK og EUR +/- 1 % interest rate	350	Medium	Accept

Fig. 7 Risk Wheel screen shot of a spreadsheet

Another similar example is presented in Fig. 7, which is a screen shot from the Excel Add-in spreadsheet. For example, the production risk '*Drop in milk yield*' is considered as of a *medium* probability to occur, which if it occurs the cost will be 300 Danish Coronas (or around 2240 €) per year. This risk is considered as '*accepted*'. This means that the farmer will accept the specific risk because he cannot do something for it. Similarly, all risks are assessed, as are presented in the example.

At the last step of the Risk Wheel, Implementation and Follow-Up, the adviser prepares an Action Plan, which is a graphical representation of all possible risks, designed as a wheel called Risk Wheel. For this purpose, all relevant information is inserted in a spreadsheet, the Excel Add-in tool. Entering all possible risks and relevant data, the tool designs in a separate sheet the Risk Wheel, which constitutes the Action Plan for a farm-unit. The final product of this effort, the Risk Wheel, is presented to the farmer, confirmed by him and constitutes the plan for further actions. A Risk Wheel screen shot is presented in Fig. 8. The follow-up services are related with the implementation of agreed actions. Therefore, for a pre-agreed period of time, a few months to a year, the strategic adviser guides and supports the implementation of the action plan.

In general, using this tool it is necessary to consider: goal and applied strategy, risk taking, risk capacity, type of risks, economic effect (importance), probabilities and management of risk. All previous milestones are represented in the Risk Wheel, which coincides with the analyzed case in a farm-unit.

The Risk Wheel tool is useful for collecting, presenting and assessing information in a Holistic Risk Management approach of a farm exploitation. During at least one meeting in the farmer's property the adviser discusses with the owner, collects useful information and develops via a spreadsheet the Risk Wheel. His effort, based

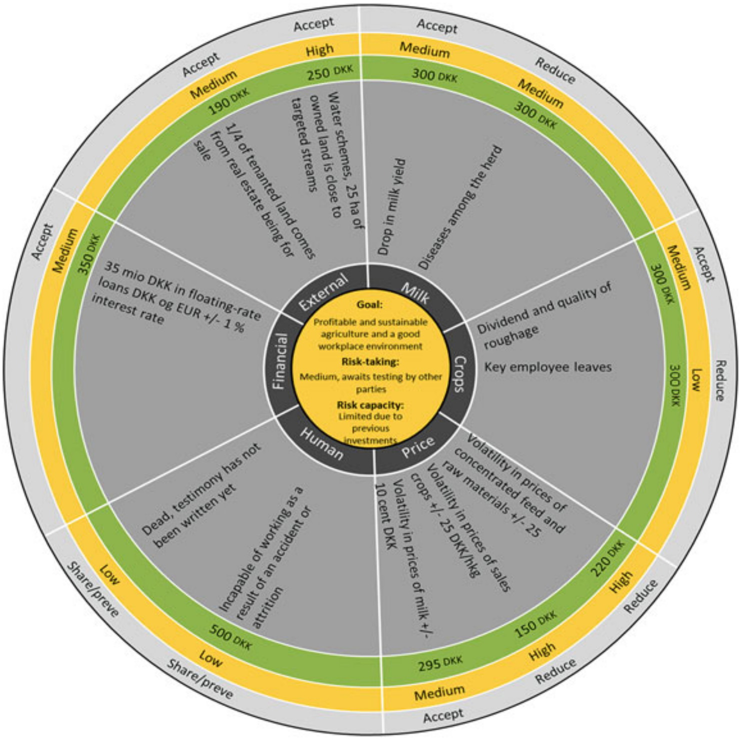


Fig. 8 The action plan as a graph (Risk Wheel screen shot)

on a comprehensive dialogue, aims to gather appropriate information and translate it into risks. At the same time the farmer has the chance to explore all potential risks he may face in near future, give them a priority and decide how to deal with them. For a more comprehensive analysis of the risks in farming, the reader is referred to Kahan (2008).

Nonetheless, more specialised methods could be applied for the estimation of either probabilities, or significance level (importance of risk factors) or for classification of risk factors. This last point, reflects on what is well-known in decision theory as *problematic β*, a typology for expressing global preference when potential actions are to be sorted in classes. In this case, four classes are proposed by the Risk Wheel. These classes are the '*accept*', the '*avoid*', the '*transfer-share*' and the '*reduce*', respectively. It could be of further research interest, if the specific tool is used in combination with outranking method Electre-Tri or any classification method like UTADIS (Devaud et al. 1980). It is worth noting that UTADIS has been implemented in multi-criteria decision support systems, such as the FINCLAS (Zopounidis and Doumpos 1998) and the PREFDIS system (Zopounidis et al. 1996).

Table 4 Quality aspects of a range of analytical tools and techniques

Tool	Focus on external/internal environment
SWOT	E/I
PEST	E
Porter's 5	E/I
Four Corner's	E
Value Chain	I
Early Warning	E/I
War Gaming	E
PerfEA	E/I
Risk Wheel	E/I

3.3 Some Quality Aspects

A range of analytical tools and techniques, as briefly described in this chapter, can be employed in the strategic analysis. All of them, even those presented for farm activities, try to identify and evaluate relevant data to form a strategy. It is worthy to note, that evaluation herein is used in its broader meaning, not strictly the quantitative one, like in the *choice phase* of Simon where the analyst translate desires, preferences, opinions, goals, etc., in numbers.

Furthermore, as it is presented in Table 4, some methods and tools mostly focus on external and/or internal environment of an enterprise. The SWOT analysis focuses both on internal and external environment, like the PerfEA and the Risk Wheel. The PEST, Four Corner's and War Gaming focus on external environment while Value Chain focus on internal environment.

In addition, all the tools, directly or indirectly, take into account risks. For example, Risk Wheel or War Gaming are directly related with risks, while SWOT or PerfEA, among other issues, incorporate any kind of risk or threat for an enterprise or exploitation.

4 Transfer of Innovation: Pilot Implementation

The transfer of innovative methods and tools in the Greek agricultural sector took place in 2014 at six local farms, located in the Crete Island. For the needs of the pilot implementation six post-graduate students of the School of Production Engineering and Management attended a special training course at farm management strategy issues, lasting one academic semester. After the training the pilot implementation followed for 6 more months. The PerfEA method applied in five cases, while the Risk Wheel tool in one case. The details concerning the participating exploitations in the implementation are presented in Table 5. As it is presented, most exploitations are family-driven, as it is the usual case in Greece, of small or very small size. The owners are both decision-makers and workers in most of them, either full- or part-time farmers. At the olive-oil enterprise there is permanent staff

Table 5 Details of farm exploitations

Type	Farmer—owner details	Property—plant—equipment or livestock	Production per year or net income per year
Exporting olive-oil enterprise	Family enterprise (two persons). Seven permanent employees and some more on a seasonal basis (Full-time entrepreneurs)	A modern industry with high quality machinery like mills, bottling equipment, tractors, etc.	600 tonnes of standardised olive-oil exported in China, S. Africa, UK, etc.
Greenhouse tomato production farm	Four persons—family farm (Full-time farmers)	3500 m^2 of greenhouse, warehouse, etc.	16–18 tonnes tomatoes per 1000 m^2 of greenhouse (ideally) or about 56 tonnes of tomatoes
Avocado farm	Two persons—family farm. Some more employees on a seasonal basis (Part-time farmers)	5000 m^2, 90 avocado trees, etc.	5.5 tonnes of fruits
Potato farming	One person. Some more employees on a seasonal basis (Full-time farmer)	15,000 m^2 owned and rented land, plant tools, tractor, etc.	20 tonnes of potatoes. Gross income: 6000–7000 €
A unit of rabbits	One person. Some help from two family members (Full-time farmer)	Ten female and one male rabbits. Building of 35 m^2, 18 cages, etc.	450 rabbits per year. Net income 3000 €
Beekeeping farm	Two persons—family farm. Some more employees on a seasonal basis (Full-time farmers)	68,000 m^2 owned land and 10,000 m^2 rented land. 525 beehives. Specialised machinery, tracks, vans, etc.	12,000 Kg of honey Gross income 70,000 €. Net income 30,000 €.

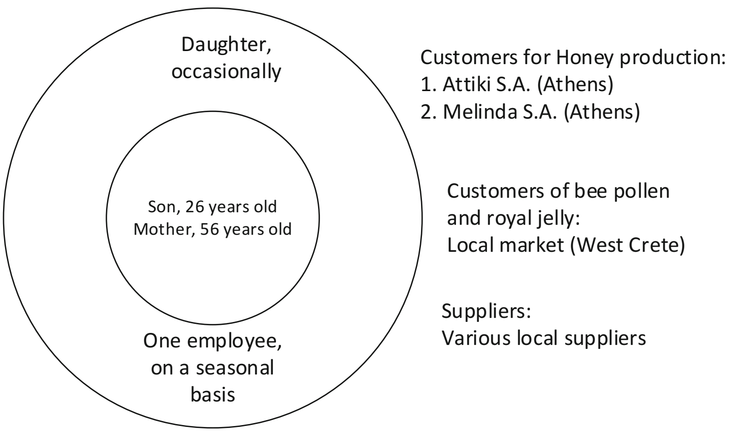


Fig. 9 The plan of borders for the thyme-honey producing unit

of seven employees, while at the rest exploitations the employed staff is on a seasonal basis. The common problems that the owners experience are production risks, lack of liquidity, loans and low net income, according to their opinion.

Indicatively, the case of the beekeepers unit is presented. In the specific case the strategic adviser applied the PerfEA method in cooperation with the owners, son and mother, respectively. The unit is transferred from the father to the son, when the first had serious health problems. The Plan of Borders (Fig. 9) presents the decision-makers in the internal circle, the two extra persons who offer services occasionally in the outer circle and customers/suppliers beyond the external border. The customers of the unit are two packaging companies in Athens that buy the thyme-honey production of 12,000 Kg per year.

The current situation of the exploitation is described as follows:

- Land: 6.8 ha plus 1 ha, owned and rented land respectively. Some beehives are hosted in this land but mostly beehives are hosted in land all around the Island. This circulation is necessary for feeding purposes. Thus, mountain areas where thyme herb grows are chosen.
- Livestock: 295 beehives (son) plus 230 beehives (mother). Controlled fertilization by the queen. Yearly honey production 12,000 kg in total. All the quantity, unpackaged, is sold at two big factories-companies in Athens. The specific production is considered as very high and the owners are among the leader honey-producers of the area.
- Labor: Two persons plus two more occasionally. Hours: 12 h per day, usually 07:00–19:00. Sometimes, the timetable changes and is depended on activities like transfer of beehives during the night, or drive for more than 3 h to reach different locations.
- Equipment: four hundred new beehives, two vans, one track, one trailer, four specialized machineries for honey extraction, four metal warehouses, specialized machinery for feeding during winter-time, uniforms, etc.

The owners discussed with the adviser a series of aspects about the durability and the existence of the exploitation, the most basic of them are: the increase the bees' population, lack of sleep when they are obliged to drive car during nights, hiring a permanent employee, lack of free time, financing problems due to late payments by customers, thoughts about taking a loan and hard working conditions.

The SWOT analysis which comprises strengths and weaknesses as well as opportunities and threats is concluded after a series of meetings (see Fig. 10). Family experience and the existent equipment constitute basic strengths. In addition, the thyme-honey produced in Crete Island, is of high quality, very tasty, and very well paid to producers in comparison to other types of similar products of different geographical areas. However, the weaknesses of the exploitation are the financing problems, the lack of staff and the inability to package the product. Thus, they sell the honey in bulk earning less income. The opportunities which exist are related with the ability to get trained, especially the son of the unit, as younger than his mother. Furthermore, for young farmers there is a special subsidy programme of 20,000 €, if the farmer applies for it. The use of internet is another opportunity for

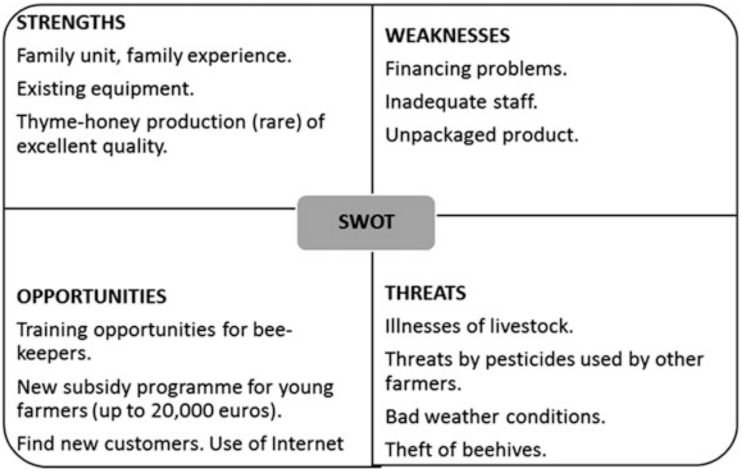


Fig. 10 A SWOT analysis in a beekeeping unit

promoting final products either honey or more specialized products, like bee pollen and royal belly, which are very healthy and very expensive. The threats of the unit include illnesses of the bees, which are common last years and cause sometimes death of a large portion of the population. Another threat for bees' life, except the bad weather conditions, is the used pesticides by other farmers, when they spray trees or plants that bees may visit. Last, another problem is the destruction or missing beehives, when they are located many kilometers far away from the owners' control.

The strategic adviser based on the vision of the son of the family and his mother to keep the exploitation, to expand and improve it, represented everything on a detailed cognitive map by using the Decision Explorer© software. The cognitive maps gave the opportunity to the owners to understand complex and interrelated issues they face and have an insight on a series of aspects for their exploitation. The maps as very detailed, are not presented herein. The final strategic areas that the owners concluded are the following:

– Quality improvement
– Income enhancement

The Action Plan combined with the Balanced Scorecard for both strategic areas, are presented below in Tables 6 and 7. The first strategic area (quality improvement, see Table 6) is analysed into three objectives: honey packaging, quality control and training. The honey packaging (and the using of an effective product label) is limited at 20 % of the total quantity produced. The appropriate packaging equipment and necessary materials cost 4500 €, which in turn can be covered by the higher retail prices achieved. The quality control of honey and royal jelly production,

Table 6 Quality improvement

Balanced scorecard				Action plan			
Strategic area	Strategic objectives	Indicators (today)	Indicators (future)	Action	Resources	Responsible	Frequency
Quality improvement	Honey packaging	No	At least 20 %	Look for appropriate equipment	Cost 4500 € will be covered by the extra income (5280 €)	Son	6 months
	Quality control	No	Quality assurance system	Cooperate with experts	Part of subsidy, if application is approved	Son	1 year
	Training	No	Yes	Participating in training	Free participation for farmers	Son	6 months

standards of equipment and installations for beekeeping, constitute another strategic objective. The expenses of this optimistic goal will be covered by a subsidy (if the relevant application is approved). Furthermore, for quality purposes training opportunities will be exploited. Seminars for beekeepers are usually organized by union of producers, cooperatives, etc. The son of the family holds a university degree, however his knowledges in beekeeping are limited in family experience.

The second strategic area (income enhancement, see Table 7) is analysed into the following objectives: produce cheap honey, find new customers, apply for subsidy and develop an internet site. Today cost of honey is 3.3 €/kg. The target is to decline to 3 €. There are several different ways to achieve the specific goal. The approved one is to increase the production rate of the beehive by moving the installation more often than today in areas which are rich in thyme and flowers. The better the feeding, the higher the quantity and quality produced. The estimated honey production is 15,000 kg instead of today 12,000 kg, which implies a lower cost at 3 €/kg. Moreover, the owners will find new customers by packaging and selling in retail. In this case, the income will be 19,200 € if they sell 8 € each kg, instead of a lower price in bulk. Thus, they will achieve an extra amount of 5280 €, which in turn plan to use it for covering the expenses for packaging equipment. Once the initial cost is covered, the extra income will be a plus to the existent income. Another source for enhancing income of the family is to apply for a subsidy. There are special subsidies up to 20,000 € for new farmers. In this case, the son of the family can apply for such subsidy. In case of approval, a small amount will go to the payment of the adviser, while the rest will cover the expenses for the quality assurance system. Moreover, the development of an internet-site will promote the high quality products to new customers, thus will positively affect family income. The development cost is low (500 €). All the above strategic objectives are stated in the Action Plan with details about resources, responsible persons and frequency of control.

In conclusion, the advising process is a new issue for Greek farmers, as opposed to other European countries, therefore a period of time is needed and some motivation for cultural change. Farmers are usually suspicious to share information of their activity with strategic advisers and skeptical about the usefulness of strategic approach. The initial experience of the advisers in most pilot implementations is that farmers have a difficulty to express their future plans. However, at the end of the advising process the farmers have achieved to think in a strategic manner, set goals and made plans for the future of their exploitation. In the presented aforementioned case of the beekeepers, although the owners were very busy with daily activities, the cooperation with them was very effective. The son of the family as younger, well-educated, and of similar age with the adviser realized from the early beginning the usefulness of the strategic approach. The pilot implementation stopped when the Action Plan was presented to the owners after a cooperation period lasting 4 months. In another case, the strategic adviser follows-up the implementation of the strategy for a certain period.

Table 7 Income enhancement

Balanced scorecard				Action plan			
Strategic area	Strategic objectives	Indicators (today)	Indicators (future)	Action	Resources	Responsible	Frequency
Income enhancement	Produce cheap honey	Cost 3.3 €/kg (40,000 € / 12,000 kg)	Cost 3 €/kg (45,000 € / 15,000 kg)	Move beehives on more flower-rich areas for better feeding	Higher productivity. Production of 15,000 kg instead of 12,000 kg today	Both members of the unit	Every period
	New retail customers. Sell packaged honey 2400 kg × 8 €/kg, instead of a lower price to companies	No retail customers	2400 kg × 8 € = 19,200 €	more income of 5280 €	Internet site and public relations	Both members of the unit	Every year
	Apply for subsidy	No subsidy	Up to 20,000 €	Contact a finance adviser	A payment for the adviser, percentage of the approved amount	Son	6 months
	Internet site	No	Yes	Look for an expert	500 €	Son	6 months

5 Conclusions

A range of analytical tools, methods and techniques can be employed in strategic analysis by specialized advisers. Some well-known and some new tools proposed in a European-level project are presented. All of them elaborate with the initial issue for strategy planning, to explore—identify and evaluate relevant data. According to Simon's decision-making phases, most tools elaborate with the *intelligence* and the *design* phase, while for the *choice* phase, as operational research scientists proceed on choices, almost nothing is proposed. This gap may be eliminated by applying interdisciplinary approaches, by combining analytical tools for strategic analysis, like SWOT, with multi-criteria methods. Tools like PerfEA and Risk Wheel, focused on strategy formulation for farm exploitations, can be integrated in conjunction with multi-criteria methods. In this framework, aspects like the ranking of conflicting goals, the weights of a set of goals or actions, the preference assessment of a group of farm members, the sorting of actions in different risk classes, are all issues that can be addressed by multi-criteria algorithms and techniques. Furthermore, group decision-making methods and algorithms can be incorporated to support expressing of preferences of *'significant others*', members of family farms. However, all the above issues constitute a part of the whole effort, the strategy formulation and implementation; especially when this dynamic activity has to be adjusted to known or unknown, expected or unexpected, circumstances.

Acknowledgment This research, training and test-site applications have been co-financed by the European Union, Leonardo da Vinci Programme, Transfer of Innovation, no. 2013-2-FR1-LEO05-48380. The authors would like to thank the following post-graduate students of the School of Production Engineering and Management of the Technical University of Crete for providing the utilized data from local farm exploitations: Adontakis N., Agapitou E., Bourbakis Y., Diafonidi H., Kalafati C., Lykakis Y.

Appendix

Principles of sustainability and related questions (*source*: STRAT project pedagogical guide).

Existence	What are the major constraints which the structure has to face? The resources which are necessary for farm existence are they available?
Effectiveness	What are the rare resources of the exploitation? What can favor or force the optimal use of these resources?
Security	What changes, frequent hazards, can return the fragile exploitation, destabilize it? How can the exploitation adapt itself to these frequent changes, face the hazards of the context? What can help the exploitation to face the hazards of the context?

(continued)

Adaptability	What elements can facilitate or limit the changes of orientation, organization or functioning of the exploitation to adapt itself to the long-lasting modifications of the context? What favors the adaptation of the exploitation to these changes? What disadvantages the adaptation of the exploitation to these changes?
Freedom of action	On what is the exploitation dependent? In what is autonomous? What favors its autonomy? What can, in a changing context, facilitate or prevent choices, autonomy of decision? What choices are realized under duress?
Coexistence	What are the main actors in interaction with the exploitation? What common objectives they have with the exploitation? What positive/negative impacts can the exploitation produce on these various actors?
Transmissibility	What are the essential elements which allow perpetuating the exploitation and its activities?
Psychological needs	To what extent expectations and needs for the persons which work on the exploitation are they taken into account? To what extent expectations and needs for the partners of the exploitation are taken into account?
Responsibility	How do the choices of the exploitation impact on the current, future generations and on the environment of the exploitation? For you, for whom is the exploitation responsible?

References

Bańkowska K (2005) Podstawy decyzji zawodowych rolników w dobie polaryzacji struktury gospodarstw. In: Woźniak MG (ed) Nierówności społeczne a wzrost gospodarczy. Kapitał ludzki i intelektualny, część 2. Rzeszów, Poland, pp 465–471

Bossel H (2001) Assessing viability and sustainability: a systems-based approach for deriving comprehensive indicator sets. Conserv Ecol 5(2):12

Brim OG et al (1962) Personality and decision processes, studies in the social psychology of thinking. Stanford University Press, Stanford

Celik Y, Peker K (2009) Benefit/cost analysis of mushroom production for diversification of income in developing countries. Bulg J Agric Sci 15(3):228–236

CIMA (The Chartered Institute of Management Accountants) (2007) Strategic analysis tools. Topic gateway series no. 34

Devaud JM, Groussaud G, Jacquet-Lagreze E (1980) UTADIS: Une methode de construction de fonctions d'utilite additives rendant compte de jugements globaux. European Working Group on Multicriteria Decision Aid, Bochum

Eden C (1988) Cognitive mapping. Eur J Oper Res 36(1):1–13

Eden C (1994) Cognitive mapping and problem structuring for system dynamics model building. Syst Dyn Rev 10(2–3):257–276

Gasson R (1973) Goals and values of farmers. J Agric Econ 24(3):521–524

Harman WL, Eidman VR, Hatch RE, Claypool PL (1972) Relating farm and operator characteristic to multiple goals. South J Agric Econ 4(1):215–220

Hayashi K (2000) Multicriteria analysis for agricultural resource management: a critical survey and future perspectives. Eur J Oper Res 122(2):486–500

Kahan D (2008) Managing risk in farming (Farm management extension guide). Food and Agriculture Organization of the United Nations, Rome

Kaya T, Kahraman C (2011) A fuzzy approach to e-banking website quality assessment based on an integrated AHP-ELECTRE method. Technol Econ Dev Econ 17(2):313–334

Klein J, Cooper D (1982) Cognitive maps of decision-makers in a complex game. J Oper Res Soc 33:63–71
Linstone HA, Turoff M (1975) The Delphi method: techniques and applications. Addison-Wesley Publishing Company, Reading, MA
Mankins M, Steele R (2006) Stop making plans. Start making decisions. Business Magazine, Decision Making, Harvard Business Review (January issue): 75–76
Matsatsinis NF, Samaras A (2001) MCDA and preference disaggregation in group decision support systems. Eur J Oper Res 130:414–429
Matsatsinis NF, Grigoroudis E, Samaras A (2005) Aggregation and disaggregation of preferences for collective decision-making. Group Decis Negot 14:217–232
Mintzberg H, Raisinghani D, Théorêt A (1976) The structure of 'unstructured' decision processes. Adm Sc Q 21:246–275
Paelnick JHP (1976) Qualitative multiple criteria analysis, environmental protection and multiregional development. Pap Reg Sci Assoc 36:59–74
Roy B (1993) Decision science or decision-aid science? Eur J Oper Res 66:184–203
Saaty TL (1980) The analytic hierarchy process: planning, priority setting, resource allocation. McGraw-Hill, New York
Sahin A, Gunden C, Cankurt M, Miran B, Yildirim I (2013) Determination of farmers' goal statements priorities and factors effective on decision (an application of fuzzy pairwise comparison analysis). Proceedings of the 1st Annual International Interdisciplinary Conference (AIIC 2013), Azores, Portugal, pp 537–544
Simon H (1960) The new science of management decision. Harper, New York
Sinuany-Stern Z, Mehrez A, Hadad Y (2011) An AHP/DEA methodology for ranking decision making units. Int Trans Oper Res 7:109–124
Solano C, León H, Pérez E, Herrero M (2001) Characterising objective profiles of Costa Rican dairy farmers. Agric Syst 67(3):153–179
Tsoukias A (2007) On the concept of decision aiding process: an operational perspective. Ann Oper Res 154:3–27
van Kooten GC, Schoney RA, Hayward KA (1986) An alternative approach to the evaluation of goal hierarchies among farmers. West J Agric Econ 11(1):40–49
Wallace MT, Moss JE (2002) Farmer decision-making with conflicting goals: a recursive strategic programming analysis. J Agric Econ 53(1):82–100
Witte E (1972) Field research on complex decision-making processes—the phase theorem. International Studies of Management and Organization 2(2):156–182
Zaerpour N, Rabbani M, Gharehgozli AH, Tavakkoli-Moghaddam R (2008) Make-to-order or make-to-stock decision by a novel hybrid approach. Adv Eng Inform 22(2):186–201
Zavadskas EK, Turskis Z, Tamosaitiene J (2011) Selection of construction enterprises management strategy based on the SWOT and multi-criteria analysis. Arch Civ Mech Eng XI(4): 1063–1082
Zopounidis C, Doumpos M (1998) Developing a multicriteria decision support system for financial classification problems: the FINCLAS system. Optim Methods Softw 8:277–304
Zopounidis C, Matsatsinis NF, Doumpos M (1996) FINEVA: a multicriteria knowledge-based decision support system for the assessment of corporate performance and viability. In the International Conference on Intelligent Technologies in Human-Related Sciences, pp 175–182

Exploring Population Drift on Consumer Credit Behavioral Scoring

Dimitris Nikolaidis, Michael Doumpos, and Constantin Zopounidis

Abstract Behavioral credit scoring models are a specific kind of credit scoring models, where time-evolving data about delinquency pattern, outstanding amounts, and account activity, is used. These data have a dynamic nature as they evolve over time in accordance with the economic environment. On the other hand, scoring models are usually static, implicitly assuming that the relationship between the performance characteristics and the subsequent performance of a customer will be the same under the current situation as it was when the information on which the scorecard was built was collected, no matter what economic changes have occurred in that period. In this study we investigate how this assumption affects the predictive power of behavioral scoring models, using a large data set from Greece, where consumer credit has been heavily affected by the economic crisis that hit the country since 2009.

Keywords Behavioral credit scoring • Consumer credit • Population drift • Risk management

1 Introduction

Credit scoring, as a principal tool to identify good prospective borrowers in presence of information asymmetry [with its well studied consequences, see Besanko and Thakor (1987) and Stiglitz and Weiss (1981), etc] and support lending organizations' decision making processes, began as early as 1941 (Durand 1941). However, the automated and thus widespread application of credit scoring did not

D. Nikolaidis • M. Doumpos (✉)
Financial Engineering Laboratory, Technical University of Crete, Chania, Greece
e-mail: mdoumpos@dpem.tuc.gr

C. Zopounidis
Financial Engineering Laboratory, Technical University of Crete, Chania, Greece

Audencia Business School, Nantes, France
e-mail: kostas@dpem.tuc.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_7

take place until the 1980's, when computing power to perform sophisticated statistical calculations became affordable. One definition of credit scoring is "the use of statistical models to transform relevant data into numerical measures that guide credit decisions" (Anderson 2007). According to Thomas et al. (2002) credit scoring has been vital in the "...phenomenal growth in the consumer credit over the last five decades. Without (credit scoring techniques, as) an accurate and automatically operated risk assessment tool, lenders of consumer credit could not have expanded their loan (effectively)".

The techniques utilized in building credit scoring models rely mostly on classification methods and can be categorized into the following groups (Yu et al. 2008a, b):

1. Statistical models: Logistic, probit or linear regression, linear discriminant analysis (LDA), classification trees, k-nearest neighbor etc.
2. Mathematical programming methods: linear programming, integer programming, etc.
3. Artificial intelligence approaches (also referred as machine learning or data mining or soft computing techniques): artificial neural networks, support vector machines, genetic algorithms, fuzzy logic, rough sets, etc.
4. Hybrid approaches: The hybridization approach is based on combining two (or more) different machine learning techniques, but only one single predictor is applied (Verikas et al. 2010). For example, a hybrid classification model can be composed of one unsupervised learner (clustering method) to pre-process the training data and one supervised learner (classifier) to learn the clustering result or vice versa (Tsai and Chen 2010).
5. Ensemble methods: Similar to hybrid approaches, an ensemble of classifiers uses more than one predictors but (unlike hybrid methods) the final prediction aggregates in some way the outputs of them.

Conventionally, the most widely applied method for credit scoring is logistic regression (Thomas et al. 2005) followed by other linear methods, such as LDA. This preference is not without a good reason since linear models provide in practice a very good compromise between classification accuracy (compared with soft computing methods) and simplicity and interpretability (Yu et al. 2008a, b). Especially financial institutions are more reluctant to adopt less intuitive, "black box" approaches (Sousa et al. 2013) since their legislative and operational framework[1] imposes constraints on data availability, transparency, verifiability and interpretability of their risk evaluation methods and processes. Nevertheless, soft computing techniques have gained much academic interest in the later years (see

[1] Just to name a few constraints the European Consumer Credit Directive 2008/48/EC (http://ec.europa.eu/consumers/financial_services/consumer_credit_directive/index_en.htm) stipulates among other that an applicant has the right to be comprehensively informed about the reasons of a rejection; The Basel II Accord (http://www.bis.org/publ/bcbsca.htm) imposes specific requirements for risk evaluation that have to be accredited.

(Stefan Lessmann et al. 2013) for a comparative study of several classification algorithms).

Traditionally we can distinguish two[2] broad types of credit scoring (Bijak and Thomas 2012):

- Application scoring: Refers to the assessment of the credit worthiness usually for new applicants. Application scoring models quantify the probability of default, by taking characteristics found in loan applications eg demographic attributes (such as age and family status), salary etc. This is historically the first type of credit scoring developed and by far the most researched and widely applied.
- Behavioral scoring: Behavioral scoring models are a natural extension from application scoring since they are used after a credit has been granted. The extra information in behavioral models is data based on the credit lines' repayment performance. We can score an existing customer using behavioral models (as part of her or his credit risk assessment) in order to provide a new credit line or to monitor and manage existing ones (i.e. coordinate customer-level decision making). For example, a bank may decide to reduce or advance credit limits on customers depending on their risk profiles. We shall note here that the distinction between behavioral and application scoring is not clear-cut in the sense that if an existing customer applies for a new credit line all available information (behavior and application data) will be used.

Since credit scoring models (as happens in all predictive models (Shmueli 2010)) assume that the future is like the past, the accuracy of the prediction depends on the available past data. This highlights one of the most important and frequently occurred problems in learning algorithms: **concept drift**, ie the evolution of data distributions over time in a dynamic, non stationary environment (Tsymbal 2004; Widmer and Kubat 1996; Žliobaitė 2009). Specifically when the population distributions change over time then we refer to *population drift*, a very common phenomenon in economy. Thus static credit scoring models based on historical data may fail to accommodate the inherent cyclicality of banking business (in accordance with the economic cycles of recession and expansion) and the shift this entails to the entire loss distribution (Allen and Saunders 2002; Niklis et al. 2014). On a first glance this thesis seems almost axiomatic. However, there are subtle points for consideration such as: what constitutes a static model, save including macroeconomic variables

[2] We shall note here that there is an expanding research in credit scoring –and especially behavioral scoring- to support decisions in areas such as marketing, through the use of *propensity scores* (Bijak 2011; Thomas 2003; Thomas et al. 2005): there are response models (will the consumer respond to marketing offers), usage models (will the consumer use a credit line) and attrition models (will a customer continue with the lender). A recent trend is also *profit scoring*, that is the use of scorecards to maximize profit (Andreeva et al. 2007; Crook et al. 2007; Finlay 2010). Additional areas in which scoring models find application are *collection scoring* (dividing insolvent customers into groups, separating those who require decisive actions from those who don't need to be attended to immediately) and *fraud detection* (ranking applicants according to the relative likelihood that their application may be fraudulent) (Phua et al. 2010).

(Bellotti and Crook 2014; Breeden et al. 2007; Crook and Bellotti 2010; Saha and Siddiqi 2011; Sousa et al. 2014; 2013), which presents specific challenges and has not yet been widely applied or adopted (at least on a commercial level)? Is there a clear cut point where to assume that population drift has occurred and therefore the classification fails or this is a more gradual and therefore less obvious process? What depth of historical data is required to model the drift?

This research will put under scrutiny all these issues by evaluating the impact of severe population drift in the context of behavioral scoring for consumer credit (i.e. credit extended to households) with actual data from the financial sector.

The rest of the paper is structured as follows: Sect. 2 provides a short introduction to credit and behavioral scoring. Section 4 examines population drift in the context of consumer credit scoring and sets the economic background of the available data for examining its effects. Section 4 presents the data used and the models that were developed and Sect. 5 presents and discusses the results. Finally Sect. 6 concludes the paper.

2 Credit Scoring Framework

2.1 Credit Scoring Formalization

Credit scoring models in general classify customers[3] into dichotomous *good* or *bad* (non-default/default) risk classes[4] which indicate the probability of a credit line to be repaid or not respectively. The *default definition* quantifies this categorization by identifying what constitutes a "bad" credit line or customer (e.g. missing 3 or more consecutive payments in the next 12 months).

Formally, let $X = (X_1, X_2, \ldots, X_n)$ be a set of n random variables (also referred as *characteristics*) that describe an applicant. Examples of typical characteristics are: (a) demographic data (such as age, employment, salary etc found in a typical loan application form), (b) account information (such as type of credit used, outstanding amounts etc) and (c) bureau data: current and past payment behavior (such as delinquencies, adverse data etc). Let also $x = (x_1, x_2, \ldots, x_n)$ denote the actual value of characteristics for a specific applicant and $p(G)$ and $p(B)$ the proportion of applicants who are good or bad respectively (prior probabilities). Thus we are interested in calculating the probability of an applicant being good given its characteristics x:

$$p(G|x) = \frac{p(x|G)p(G)}{p(x)} \tag{1}$$

[3] A customer may be an existing one who has already a credit history or a new one applying for first time. This distinction is important in the context of behavioral score.

[4] Multi-class credit risk classification has not being extensively studied or applied in practice (see (Chen 2012; Hsieh et al. 2010; Tang and Qiu 2012) for an example of multi-class SVM for credit scoring).

where $p(x|G)$ and $p(x|B)$ refer to the conditional probabilities of the risk classes (the distributions of the classes) and the outcome $p(G|x)$ refers to the posterior probabilities of classes.

In terms of modeling, the relationship between the characteristics x of an applicant (inputs) and the outcome $y = p(G|x)$ can be described as follows:

$$y = f(x_1, x_2, \ldots, x_n) \tag{2}$$

Credit scoring models are usually presented in the form of *scorecards*, where the characteristics, their attributes and their relative weights are encoded in a tabular format (when intended for human audience). Hence the terminology "scoring model" and "scorecard" can be used interchangeably.

2.2 Behavioral Scoring

Credit behavioral scoring is used by financial organizations to guide lending decisions for customers in cases such as: credit limit management strategies; managing debt collection and recovery; retaining future profitable customers; predicting accounts likely to close or settle early; cross-selling; offering new interest rates; predicting fraudulent activity etc (Breeden et al. 2007; Hand and Henley 1997; Thomas and Malik 2010).

Table 1 presents sources and typical characteristics of behavioral scorecards.[5] The choice of which of these are going to be used depends on their availability as well as their appropriateness. For example promotions history is more appropriate for an attrition scorecard than for a behavioral model of retail loans (Kennedy et al. 2013).

Whereas, application scoring can be considered as static, behavioral scoring is updated constantly through the current and most recent performance of the borrower (Thomas et al. 2001). This has a very significant implication in the context of our study. A comparison between these different types of scoring has not been researched in the literature, however behavioral scorecards appear to be more predictive of future loan quality than application ones (Hand and Henley 1997).

Medema et al. (2009) indicate two main types of behavioral models: classification and duration models. We have already mentioned classification models aiming at predicting the event of default, most often through logistic regression or

[5] We shall note here that these data reflect the traditional and well-established academic and financial industry perspective. However, there is an increasing fad (eg FICO score with use of alternative data, see Andriotis (2015) and similar approaches from Equifax and TransUnion) for using alternative data (mainly from sources such as utilities & telecommunication bill payments, but also from data from social networks, rentals etc). The aim is to outreach those with very thin data or serious delinquencies, sometimes mentioned as "credit invisibles" [see http://www.perc.net (Turner et al. 2015; 2009)].

Table 1 Behavioral scoring characteristics and their sources

Data source	Feature example
Delinquency history	Max/current delinquency
Usage history	Balance-to-limit/amount ratio
Static information	Age, demographic data
Payment history	Frequency of payments
Collections activity	Outcomes, contact frequency
Type of credit	Number of transactions, Type
Customer service	Inbound/outbound contact
Promotions history	Number/outcome of offers
Bureau data	Scores, account information

discriminant analysis. In duration models the focus is on predicting the time to default using survival analysis (Andreeva et al. 2007; Baesens et al. 2005; Bellotti and Crook 2008a, b, 2014; Giambona 2012; Im et al. 2012; Sarlija et al. 2009; Stepanova and Thomas 2002).

The methodology to develop behavioral models is similar to the general credit scoring model development:

1. A random sample of customers/credit lines is selected and their performance at either side of an arbitrary *observation point* is measured.
2. The period before observation point is called as *performance window* (usually between 6 and 24 months) and using the available characteristics during this period the performance of the customers/credit lines is determined.
3. In the period (usually 12 months) after the observation point (*outcome window*) the customer/credit line is classified as good or bad, based on the predetermined default definition.
4. We then use the outcome period classification to estimate the parameters of the models for the characteristics obtained in the observation period.

Figure 1 illustrates this longitudinal aspect of the data in behavioral scoring.

Notice that there are three important parameters to be decided in practice: The lengths of the outcome and performance window respectively and the definition of default. Credit scoring literature has not provided strong answers to these issues and there is no standard way to define the window length or the default definition. The recommendations in the literature vary the length of performance and outcome windows from 6 to 24 months (Mays 2005; Thomas et al. 2002; Thomas and Malik 2010). Practitioners use a trial-and-error approach to identify the best candidate. Regarding default definition, R. Anderson in (Anderson 2007) designated that financial institutions choose between: (a) a *current status* definition that classifies an account as good or bad based on its status at the end of the outcome window and (b) a *worst status* approach that uses a time-period during the outcome window. A 90-dpd (days past due) worst status approach is commonly used in practice in behavioral scorecards and complies with the Basel II Capital Accord.[6]

[6] http://www.bis.org/publ/bcbsca.htm

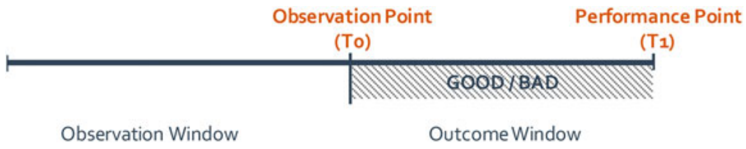


Fig. 1 Performance and outcome windows

Kennedy et al. (2013) have made a comparative study of various values for these parameters. Their results indicated that behavioral scoring models using:

- default definitions based on a worst status approach outperformed those with current status
- a 12-month performance window outperformed the ones with 6- and 18-month windows in combination with shorter (12 months or less) outcome windows
- 6-months outcome windows (or less) and a current status default definition outperformed longer outcome windows; 12-month (or less) outcome windows combined with worst status default definition outperformed longer ones.

3 Population Drift and Evidence from Current Situation

As mentioned above, an outstanding issue in credit scoring is population drift: the change of population distribution over time.[7] Evolution of population can occur in three ways (Kelly et al. 1999; Pavlidis et al. 2012):

(a) change of risk classes prior probabilities $p(G)$ and $p(B)$,
(b) change in the class conditional probabilities $p(x|G)$, $p(x|B)$ and
(c) change in the posterior probabilities $p(G|x)$, $p(B|x)$.

It's worth mentioning that changes in class priors and/or class conditional probabilities do not necessarily lead to change in posterior probabilities, in which case the predictive decision will remain unaffected (Gama et al. 2014; Kelly et al. 1999). However, in reality we could only observe the changes in the joint probability $p(x,G)$ or $p(x,B)$ making it hard to distinguish whether $p(x)$ or $p(G,B|x)$ has changed (Gao et al. 2007).

One of the disadvantages of credit scoring is that one typically needs at least two years history to build a scorecard and thus both the population characteristics and the economic environment may have changed. The latter problem is heightened because credit scorecards tend to have no external economic characteristics in them (Thomas et al. 2001), in the sense that they generally do not relate aspects of credit risk to determining external factors that vary over time (Crook and Bellotti 2010).

[7] Concept drift is a wider phenomenon than population drift; it can refer eg to situations where the classes of the classification problem change over time (Kelly and Hand 1999).

However, especially behavioral credit scoring models include by design time-changing characteristics (such as delinquency pattern, account activity etc) assuming implicitly that the relationship between the performance characteristics and the subsequent delinquency status of a customer will be the same under current situation as it was two to three year ago when the information on which the scorecard was built was collected, no matter what economic changes have occurred in that period. **Our study will further investigate this assumption**.

The conventional approach to handle the problem of population drift in practice is to manually rebuild or recalibrate the scorecards. This is costly, sub-optimal in terms of potential model quality and slow to adapt to changes due to verification latency (Hofer and Krempl 2013; Marrs et al. 2010), and needs constant manual intervention. In fact scorecards left un-calibrated, is considered as one of the reasons for the US subprime mortgage crisis (Thomas and Jung 2013; Rona-Tas and Hiss 2008).

3.1 Evidence from Greek Market

There is considerable evidence in literature that macroeconomic factors affect the risk of borrowers and the probability of default (Crook and Banasik 2005; Crook and Bellotti 2010; Ingolfsson and Elvarsson 2010; Thomas et al. 2002). Greece has undergone a rapid and significant (in term of volume in Euros) expansion of its consumer credit (Fig. 2), followed since 2010 from a total collapse: Figs. 3 and 4 indicate the extend of this deep crisis and the impact of it in the NPLs (non-performing loans), where a loan is characterized as nonperforming “when payments of interest and/or principal are past due by 90 days or more, or interest payments equal to 90 days or more have been capitalized, refinanced, or delayed by agreement, or payments are less than 90 days overdue, but there are other good

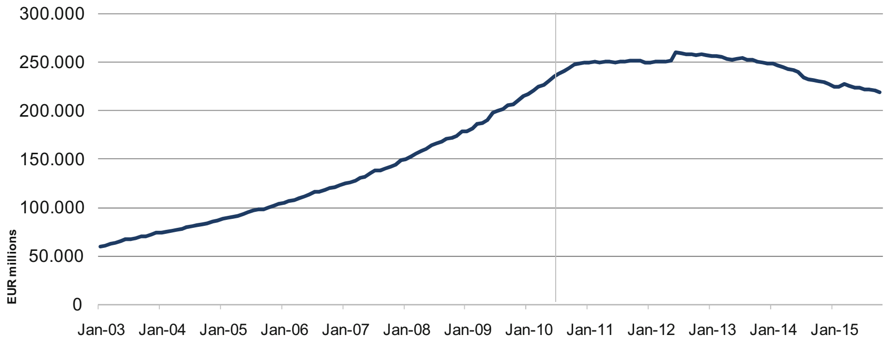


Fig. 2 Outstanding amounts to households (Credit to domestic non-MFI residents by domestic MFIs)—*Grey line* indicates the official onset of the economic crisis (May 2010) by entering into the 1st economic adjustment program. Source: Bank of Greece

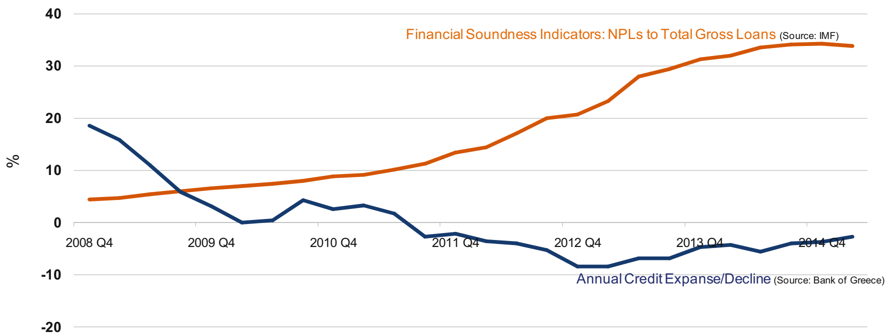


Fig. 3 NPLs & Credit growth/decline (annual change of outstanding amounts). Source: IMF, Bank of Greece

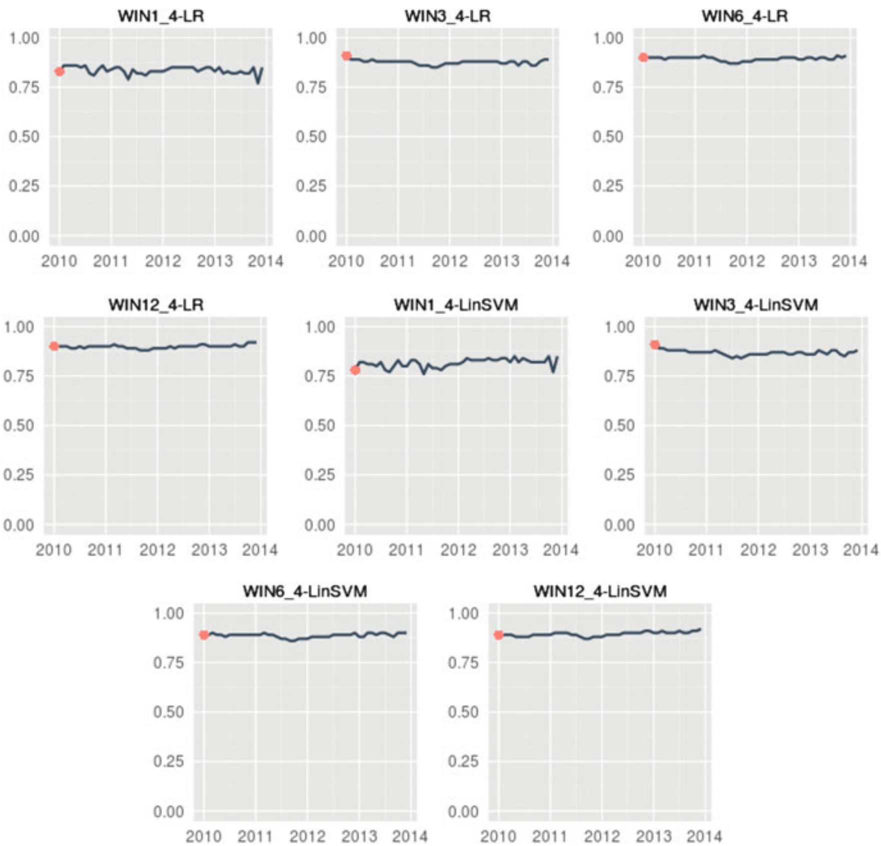


Fig. 4 Trend of AUC of logistic regression and linear SVM

reasons—such as a debtor filing for bankruptcy—to doubt that payments will be made in full”.[8]

As unfortunate as this on-going downturn may be, it enables us, nevertheless, to examine the impact of a severe shock in the macroeconomic environment to the performance of behavioral scoring models.

4 Data and Models

Our data came from a proprietary database containing behavioral variables for more than 3 million credit lines on household accounts (i.e. consumer loans, credit cards and mortgages) covering the entire financial sector of the country for a period from January 2010 (just on the onset of the macroeconomic crisis) to December 2014. The database is updated on a monthly basis. From this database we extracted a random sample of 20,000 borrowers extant in December 2014 (86,082 credit lines) and followed their credit lines “backwards” up to January 2010. To avoid survivorship bias (Elton et al. 1996) we included all credit lines that these borrowers have had at any time during our sampling period, keeping closed ones e.g. paid off, defaulted etc) up to December 2014 and including new credit lines that were approved at any time during our sampling period.

4.1 Behavioral Scoring Parameters

The following parameters, necessary for creating a behavioral scoring model, where defined, taking into account also the specifications of the Basel II accord:

- Observation Period(s): Time windows of 1,3,6 and 12 months prior to each observation point. Our initial observation point was at January 2010 and every subsequent month thereafter.
- Performance Period: a 12 month window after observation point.
- Observation Point (T0) default definition: The classification of credit lines in GOOD/BAD/INDETERMINATE/OTHER classes using their **current** delinquency (see Appendix 1—Table 3) at each observation point. These definitions were used to specify the scorable sample at the given point.
- Performance Point (T1) default definition: The classification of credit lines in GOOD/BAD/INDETERMINATE/OTHER classes using a **worst delinquency** approach (see Appendix 1—Table 4) at each performance point. These definitions served as the output (depended variable) to train the classification models.

[8] IMF: Compilation Guide on Financial Soundness Indicators, https://www.imf.org/external/np/sta/fsi/eng/2004/guide/index.htm

4.2 Modeling Methodology

As classifiers we used logistic regression and a linear SVM. We based our decision for the selection of these models to the following criteria:

- Logistic regression is the financial industry standard for credit scoring. We have the opportunity to observe its performance under stress conditions and evaluate the results of population drift. It will be used as a baseline in our study.
- There is growing and important evidence in the literature (see for example (Baesens et al. 2003; Chang and Yeh 2012; Hand 2006; Stefan Lessmann et al. 2013; Van Gestel et al. 2010) etc) that simple methods (such as logistic regression or LDA) perform very close to more advanced methods (such as neural networks, SVMs, extreme learning machines, ensemble classifiers etc); a phenomenon that seems to not pertain only in credit scoring (Holte 1993). For example, a very comprehensive study (Stefan Lessmann et al. 2013) compared 41 different classification algorithms/methodologies on 7 datasets. The difference of performance between logistic regression compared to the absolutely best classifier for each dataset in terms of the AUC (Area under receiver operating characteristic curve), varied from 0.2 % to 2.2 % in all but one datasets (in which the performance difference was 14.9 %).
- Finally simple classifiers, probably due to the *flat maximum* effect (Hand 2006; Overstreet et al. 1992), combined with domain knowledge "can do surprisingly well in classifying the points" (Hand 2006).

To examine the performance of the models in the presence of the population drift we trained them using data from January 2010 and cross-validated them on every subsequent month. Specifically, our model building process was the following (see also (Anderson 2007; Siddiqi 2005) for a practitioners' approach into building credit scoring models).

For each T0 starting from January 2010 up to December 2013 (recall that we need a 12 month performance period to observe the performance)

1. First we excluded credit lines with not enough data or that were already defaulted (BAD) at T0, deriving thus the scorable sample
2. Subsequently only those credit lines that were characterized as GOOD/BAD at performance point T1 (12 months after T0) were kept from the scorable sample.
3. The model was cross-validated with out-of-time data for every T0 subsequent to January 2010. Do to the very reach validation data and the nature of the problem there was no was no need for hold-out validation.

For measuring the performance of the models we will use the area under ROC (AUC), a well-established classifier metric (Fawcett 2006) in credit scoring literature and practice (Clémençon and Vayatis 2010; Kraft et al. 2002; Thomas 2007; Rezac and Rezac 2011; Siarka 2012; Lessmann et al. 2013). It is also explicitly mentioned in the Basel II capital accord. AUC values range from 0 to 1, where 1 and 0.5 correspond to perfect and random classification, respectively. Specifically, the

AUC equals the probability that a randomly chosen positive example will be ranked higher than a randomly chosen negative example.

Despite its advantages, the AUC as a performance measure has drawn valid criticism (Hand 2009; Hand and Anagnostopoulos 2013), basically as an incoherent metric to compare different classifiers. Since we are mainly interested in examining the effects of population drift on the *same classifier* (may it be logistic regression, or a linear SVM) this incoherence will not affect our conclusions.

4.3 Variables

The available behavioral variables per credit line fall into these dimensions:

- Type of credit (consumer loan, mortgage, revolving credit -such as overdrafts-, credit card etc)
- Current delinquency (months in arrears)
- Amounts (Outstanding balance, delinquent balance, approved amount or credit limit -for credit cards or revolving credit-)
- Age of credit line (months since approval)
- Status (active, paid-off, defaulted)

The monthly snapshots of the variables were taken for 60 periods (January 2010–December 2014) and derivative characteristics were also calculated (see Appendix 2). Variable (or feature) selection is per se an active research topic [see e.g. Bellotti and Crook (2008a, b), Hajek and Michalak (2013), Liu and Schumann (2005), Maldonado et al. (2011), Shi et al. (2013), Siami et al. (2013), Song et al. (2010), Tsai (2009), Waad et al. (2013), Wang et al. (2012), Yao (2009) for specific application of feature selection approaches in credit scoring]. However, we decided to an ad-hoc/expert approach based on results from univariate and bivariate analysis for the following reasons:

- Data were not of high dimensionality. We aimed for simplicity as we mentioned in Sect. 4.1 above.
- Variable interdependence/multicolineraity was checked using variance inflation factor (VIF) which, nevertheless, has been criticized as a practical measure for multicolinerarity (O'brien 2007). Also problems caused by multicollinearity when fitting standard logistic regression models in small samples do not seem to hold for the large samples that are frequently encountered in a credit scoring context (De Jongh et al. 2015).
- Automated variable selection methods, specifically stepwise logistic regression, which would suit our needs, has been heavily criticized (Harrell 2015; Shtatland et al. 2001).
- We do not need to be concerned about overfitting since by design we will actually perform extensive cross-validation with out-of-time data.

- Most importantly domain knowledge was available by the authors enabling thus an expert feature selection (Guyon and Elisseeff 2003).

Various combination of variables were tested. Group averages for the final variables used in the model (TYPE, MAXIMUM DELINQUENCY, MAXIMUM UTILIZATION) are presented in Appendix 2.

5 Results and Discussion

Fitting logistic regression and a linear SVM on January 2010 and cross-validating it across 47 months up to December 2013, yielded the results presented in Fig. 4. The charts display the trend of the AUC of each model for every data point (month) where the training point (January 2010) is displayed with a red dot. WIN1 indicates that the variables of the model used only one month (current month at T0), WIN3 indicates a 3-month window for training e.g. the maximum delinquency was calculated across 3 months) etc. Table 2 displays also the statistics of the results.

The immediate conclusions arising from these results are the following:

- First and foremost in both models we tried (LR, linear SVM) the drift occurred to the population due to the macroeconomic crisis did not affect the models at all. Their performance remained stable throughout the entire cross-validation period.
- The predictive ability itself of the models is quite high (see Table 2) taking into consideration that we used data from the real world and not artificially constructed.
- Bigger time-windows (up to 12 months) seem to yield better (if only slightly) performance.
- There seems no advantage of SVM over logistic regression.

Table 2 Key statistics of models performances

Model	Mean. AUC	Max. AUC	Min. AUC	stddev
LR-1 month	0.84	0.86	0.77	0.02
LR-3 month	0.88	0.89	0.85	0.01
LR-6 month	0.89	0.91	0.87	0.01
LR-12 month	0.90	0.92	0.88	0.01
LinSVM-1 month	0.82	0.85	0.76	0.02
LinSVM-3 month	0.87	0.89	0.84	0.01
LinSVM-6 month	0.89	0.90	0.86	0.01
LinSVM-12 month	0.89	0.92	0.87	0.01

5.1 Discussion

On the basis of the obtained results a number of interesting findings emerge:

1. Behavioral variables are extremely predictive and separate very well good from bad population (see Figs. 6 and 7, Appendix 2). In a sense it's like a "self-fulfilling prophecy" to have a severe delinquency currently and predict that this credit line will default.
2. Since we scored the credit line and not the borrower in total, point 1 above is amplified: a borrower facing financial difficulties (i.e. having the willingness but not the ability to pay all of her debts) may choose not to pay eg a credit card over her mortgage. This would be reflected in model that would a borrower and all her credit data as a distinct entity.
3. Exploring our data in more depth we realized that new credit approvals were actually diminished as evidenced by the decrease by 18 % of total outstanding balance of loans (Fig. 2) but also dictated by common logic (financial institutions in Greece have undergone 2 recapitalizations in the last 3 years). Instead of new loans, it took place only restructuring or defaulting of existing credit by those having financial difficulties and repayment by those capable of. This is further evidenced by Fig. 5. An examination reveals that: (a) the percentage of good population has significantly decreased whereas that of bad increased. In absolute numbers BAD increased by 187 % whereas GOOD by a mere 6 %; (b) NOT ENOUGH DATA increased (in absolute numbers) by 1356 % from January 2010 to December 2013. This was caused by closing (paid-off) credit lines which being closed did not have enough data in the next period to be scored and thus characterized as "thin".
4. Furthermore, the above trend highlighted an inevitable survivor bias amplified by the fiscal environment: removing defaulted credit lines from the system and not adding new ones led to keeping only those borrowers with very robust

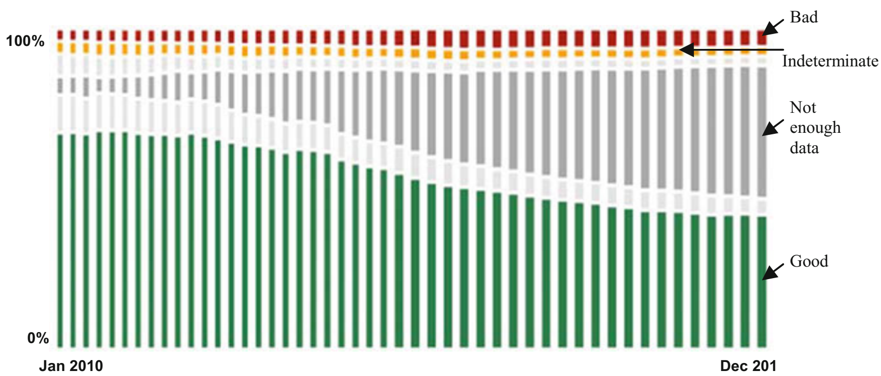


Fig. 5 Evolution (stacked % of total for each month) of distribution of population classification (GOOD/BAD/INDETERMINATE/NOT ENOUGH DATA) over time

financial position. In turn they continue to pay their financial obligations and thus enhancing the correlation between behavioral variables (payment, delinquency) and outcome (default/non-default).

6 Conclusions

Using a large database from Greece, this study showed that, on a credit line level, behavioral credit scoring based on data from financial institutions is unaffected by severe deterioration in the macroeconomic environment. In terms of modeling terminology, and despite structural changes in the credit market, analytical models developed using past data still perform well after the emergence of the Greek crisis. An interesting extension to this work would be to expand our line of view and test this result on a borrower level accounting thus for all financial obligations. Extensions to corporate default data could also be explored.

Appendix 1: Scoring Parameters

Table 3 Observation point (T0) definitions

Classification at T0	Definition
Bad	**Current Delinquency** >= 4 months (90 days past due) with Delinquent Balance > 50 €
Indeterminate	**Current** Delinquency < 4 months AND >= 2 months
Other	Not enough data
Good	Not ALREADY BAD or INDETERMINATE and **current delinquency** 0 or 1 month

Table 4 Performance point (T1) definitions

Classification at T1	Definition
Bad	**Worst Delinquency** in performance period >= 4 months (90 days past due) with Delinquent Balance > 50 €
Indeterminate	**Worst Delinquency** in performance period < 4 months AND >= 2 months
Other	Not enough data
Good	Not ALREADY BAD or INDETERMINATE and **Worst Delinquency** <= 1 month

Appendix 2: Variables

Table 5 Behavioral scoring variables

Variable (Characteristic)	Values
TYPE	CONS=Consumer loan CARD = credit card HOUS = housing loan
DELIQINDOCCUR X X=1,3,6,12 months	Worst delinquency in last X months, if X=1, then current delinquency at T0
CURBALOCCUR X X=1,3,6,12,24 months	Maximum outstanding balance in last X months, if X=1, then balance at T0
CURBALDELOCCUR X X=1,3,6,12,24 months	Maximum delinquent balance in last X months, if X=1, then delinquent balance at T0
DELIQINDOCCUR12SUM	Sum of delinquencies in last 12 months
UTILIZATIONMAX X X=1,3,6,12,24 months	Maximum utilization in last X months, where UTILIZATION = ratio of current balance to approved amount (or limit for credit cards) if X=1, then utilization at T0
CURBALOCCUR1_DIFFX, X=3,6,12,24 months	Difference between CURRENT BALANCE at T0 and at T0-X, e.g. CURBALOCCUR1_DIFF3= [CURBALOCCUR1 at T0 - CURBALOCCUR1 at (T0-3)]/ CURBALOCCUR1 at (T0-3)
UTILIZATION_DIFFX, X=3,6,12,24 months	Difference between UTILIZATION at T0 and at T0-X
CURBALDELOCCUR1_DIFFX, X=3,6,12,24 months	Difference between DELINQUENT BALANCE at T0 and at T0-X
AGEOFACCOUNT	Months since approval

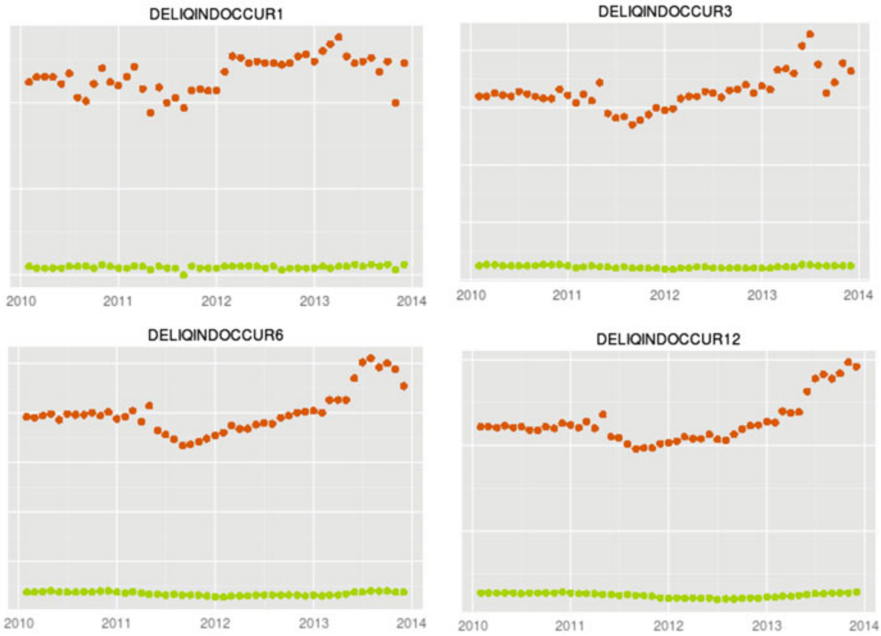


Fig. 6 MAXIMUM DELINQUENCY (months in arrears) group average values (*dots* on *top* indicate BAD credit lines, *bottom* indicate GOOD ones)

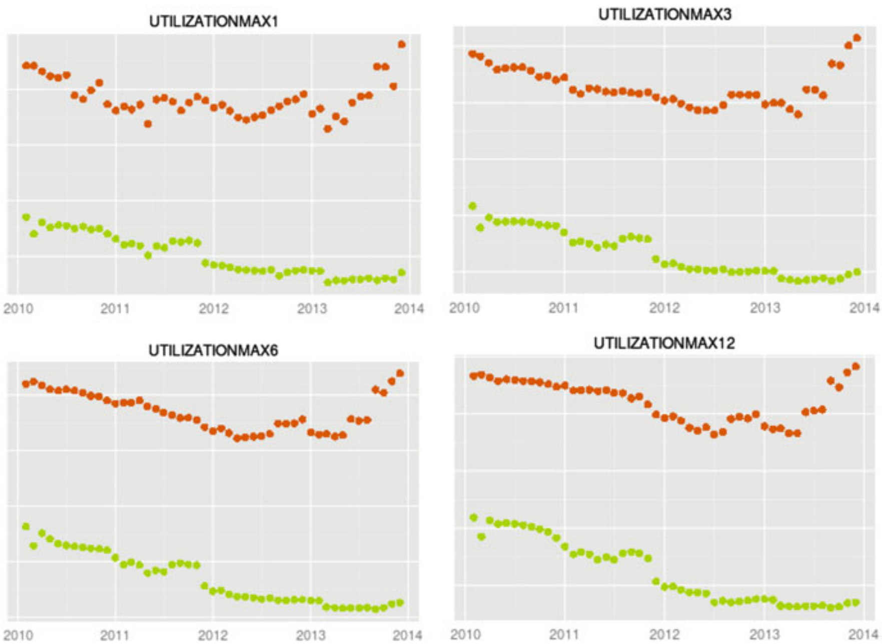


Fig. 7 MAXIMUM UTILIZATION (ratio of current balance to approved amount) group average values (*dots* on *top* indicate BAD credit lines, *bottom* indicate GOOD ones)

References

Allen L, Saunders A (2002) A survey of cyclical effects in credit risk measurement models (SSRN Scholarly Paper No. ID 315561). Social Science Research Network, Rochester, NY

Anderson R (2007) The credit scoring toolkit: theory and practice for retail credit risk management and decision automation. Oxford University Press, Oxford

Andreeva G, Ansell J, Crook J (2007) Modelling profitability using survival combination scores. Eur J Oper Res 183:1537–1549. doi:10.1016/j.ejor.2006.10.064

Andriotis A (2015) FICO announces new credit score based on alternative data. Wall Street J. http://www.wsj.com/articles/fico-announces-new-credit-score-based-on-alternative-data-1427989748

Baesens B, Gestel TV, Stepanova M, Van den Poel D, Vanthienen J (2005) Neural network survival analysis for personal loan data. J Oper Res Soc 56:1089–1098. doi:10.1057/palgrave.jors.2601990

Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J, Vanthienen J (2003) Benchmarking state-of-the-art classification algorithms for credit scoring. J Oper Res Soc 54:627–635

Bellotti T, Crook J (2014) Retail credit stress testing using a discrete hazard model with macroeconomic factors. J Oper Res Soc 65:340–350. doi:10.1057/jors.2013.91

Bellotti T, Crook J (2008a) Credit scoring with macroeconomic variables using survival analysis. J Oper Res Soc 60:1699–1707

Bellotti T, Crook J (2008b) Support vector machines for credit scoring and discovery of significant features. Expert Syst Appl 36:3302–3308

Besanko D, Thakor AV (1987) Competitive equilibrium in the credit market under asymmetric information. J Econ Theory 42:167–182. doi:10.1016/0022-0531(87)90108-6

Bijak K (2011) Kalman filtering as a performance monitoring technique for a propensity scorecard. J Oper Res Soc 62:29–37. doi:10.1057/jors.2009.183

Bijak K, Thomas LC (2012) Does segmentation always improve model performance in credit scoring? Expert Syst Appl 39:2433–2442. doi:10.1016/j.eswa.2011.08.093

Breeden J, Thomas L, McDonald J III (2007) Stress testing retail load portfolios with dual-time dynamics. J Risk Model Valid 2:1–19

Chang S-Y, Yeh T-Y (2012) An artificial immune classifier for credit scoring analysis. Appl Soft Comput 12:611–618. doi:10.1016/j.asoc.2011.11.002

Chen Y (2012) Research on multi-classification of credit rating of small and medium-sized enterprises in growth enterprises board based on fuzzy ordinal regression support vector machine. Int J Econ Finance 4:248–252. doi:10.5539/ijef.v4n3p248

Clémençon S, Vayatis N (2010) Overlaying classifiers: a practical approach to optimal scoring. Constr Approx 32:619–648. doi:10.1007/s00365-010-9084-9

Crook J, Banasik J (2005) Explaining aggregate consumer delinquency behaviour over time. Working Paper Series-University of Edinburgh Management School 5.

Crook J, Bellotti T (2010) Time varying and dynamic models for default risk in consumer loans. J R Stat Soc A 173:283–305. doi:10.1111/j.1467-985X.2009.00617.x

Crook JN, Edelman DB, Thomas LC (2007) Recent developments in consumer credit risk assessment. Eur J Oper Res 183:1447–1465

De Jongh P, De Jongh E, Pienaar M, Gordon-Grant H, Oberholzer M, Santana L (2015) The impact of pre-selected variance inflation factor thresholds on the stability and predictive power of logistic regression models in credit scoring. ORiON 31:17. doi:10.5784/31-1-162

Durand D (1941) Credit-rating formulae. In: Durand D (Ed) Risk elements in consumer installment financing. NBER, pp. 83–91.

Elton EJ, Gruber MJ, Blake CR (1996) Survivor bias and mutual fund performance. Rev Financ Stud 9:1097–1120

Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27:861–874

Finlay S (2010) Credit scoring for profitability objectives. Eur J Oper Res 202:528–537. doi:10.1016/j.ejor.2009.05.025

Gama J, Žliobait\e I, Bifet A, Pechenizkiy M, Bouchachia A (2014) A survey on concept drift adaptation. ACM Comput Surv 46, 44:1–44:37. doi:10.1145/2523813
Gao J, Fan W, Han J, Yu PS (2007) A general framework for mining concept-drifting data streams with skewed distributions. In: Proceedings of SDM'07
Giambona F (2012) A discrete-time hazard model for loans: some evidence from Italian banking system. Am J Appl Sci 9:1337–1346
Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182
Hajek P, Michalak K (2013) Feature selection in corporate credit rating prediction. Knowl-Based Syst 51:72–84. doi:10.1016/j.knosys.2013.07.008
Hand DJ (2009) Measuring classifier performance: a coherent alternative to the area under the ROC curve. Mach Learn 77:103–123
Hand DJ (2006) Classifier technology and the illusion of progress. Statist Sci 21:1–14. doi:10.1214/088342306000000060
Hand DJ, Anagnostopoulos C (2013) When is the area under the receiver operating characteristic curve an appropriate measure of classifier performance? Pattern Recogn Lett 34:492–495. doi:10.1016/j.patrec.2012.12.004
Hand DJ, Henley WE (1997) Statistical classification methods in consumer credit scoring: a review. J R Statist Soc A 160:523–541
Harrell F (2015) Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer, Heidelberg
Hofer V, Krempl G (2013) Drift mining in data: a framework for addressing drift in classification. Comput Stat Data Anal 57:377–391. doi:10.1016/j.csda.2012.07.007
Holte RC (1993) Very simple classification rules perform well on most commonly used datasets. Mach Learn 11:63–90. doi:10.1023/A:1022631118932
Hsieh H-I, Lee T-P, Lee T-S (2010) Data mining in building behavioral scoring models. In: 2010 international conference on computational intelligence and software engineering (CiSE). IEEE, pp 1–4
Im J, Apley DW, Qi C, Shan X (2012) A time-dependent proportional hazards survival model for credit risk analysis. J Oper Res Soc 63:306–321. doi:10.1057/jors.2011.34
Ingolfsson S, Elvarsson BT (2010) Cyclical adjustment of point-in-time PD. J Oper Res Soc 61:374–380. doi:10.1057/jors.2009.136
Kelly MG, Hand DJ (1999) Credit scoring with uncertain class definitions. IMA J Manag Math 10:331–345. doi:10.1093/imaman/10.4.331
Kelly, M.G., Hand, D.J., Adams, N.M., 1999. The impact of changing populations on classifier performance. In: Proceedings of the 5th ACM SIGKDD international conference on knowledge discovery and data mining, pp 367–371
Kennedy K, Mac Namee B, Delany SJ, O'Sullivan M, Watson N (2013) A window of opportunity: assessing behavioural scoring. Expert Syst Appl 40:1372–1380. doi:10.1016/j.eswa.2012.08.052
Kraft H, Kroisandt G, Müller M (2002) Assessing the discriminatory power of credit scores. Discussion papers. Interdisciplinary Research Project 373: quantification and simulation of economic processes
Liu Y, Schumann M (2005) Data mining feature selection for credit scoring models. J Oper Res Soc 56:1099–1108. doi:10.1057/palgrave.jors.2601976
Thomas L (2007) Measuring the discrimination quality of suites of scorecards : ROCs Ginis, bounds and segmentation. Presented at the Credit Scoring and Credit Control X, Edinburgh
Thomas L, Jung KM (2013) When to rebuild or when to recalibrate scorecards. Presented at the Credit Scoring and Credit Control XIII, Edinburgh
Maldonado S, Weber R, Basak J (2011) Simultaneous feature selection and classification using kernel-penalized support vector machines. Inf Sci 181:115–128. doi:10.1016/j.ins.2010.08.047

Marrs GR, Hickey RJ, Black MM (2010) The impact of latency on online classification learning with concept drift. In: Bi Y, Williams M-A (eds) Knowledge science, engineering and management. Lecture notes in computer science. Springer, Berlin, Heidelberg, pp 459–469

Mays E (2005) Handbook of credit scoring. Publishers Group Uk, [S.l.]

Medema L, Koning RH, Lensink R (2009) A practical approach to validating a PD model. J Bank Financ 33:701–708

Turner M, Varghese R, Walker P (2015) Research consensus confirms benefits of alternative data (Technical Report). PERC

Niklis D, Doumpos M, Zopounidis C (2014) Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines. Appl Math Comput 234:69–81. doi:10.1016/j.amc.2014.02.028

O'brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41:673–690

Overstreet GA, Bradley EL, Kemp RS (1992) The flat-maximum effect and generic linear scoring models: a test. IMA J Manag Math 4:97–109. doi:10.1093/imaman/4.1.97

Pavlidis NG, Tasoulis DK, Adams NM, Hand DJ (2012) Adaptive consumer credit classification. J Oper Res Soc 63:1645–1654. doi:10.1057/jors.2012.15

Phua C, Lee V, Smith K, Gayler R (2010) A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.

Rezac M, Rezac F (2011) How to measure the quality of credit scoring models. Czech J Econ Finance (Finance a uver) 61:486–507

Rona-Tas A, Hiss S (2008) Consumer and corporate credit ratings and the subprime crisis in the US with some lessons for Germany. SCHUFA, Wiesbaden

Saha A, Siddiqi N (2011) Survival analysis workflow: assessing the impact of macro-economic shocks on credit portfolios and predicting the time of default

Sarlija N, Bensic M, Zekic-Susac M (2009) Comparison procedure of predicting the time to default in behavioural scoring. Expert Syst Appl 36:8778–8788. doi:10.1016/j.eswa.2008.11.042

Shi J, Zhang S, Qiu L (2013) Credit scoring by feature-weighted support vector machines. J Zhejiang Univ Sci C 14:197–204

Shmueli G (2010) To explain or to predict? Stat Sci 25:289–310. doi:10.1214/10-STS330

Shtatland ES, Cain E, Barton MB (2001) The perils of stepwise logistic regression and how to escape them using information criteria and the output delivery system. In: 26th Annual SAS users group international conference, Long Beach, California

Siami M, Gholamian MR, Basiri J (2013) An application of locally linear model tree algorithm with combination of feature selection in credit scoring. Int J Syst Sci 1–10. doi:10.1080/00207721.2013.767395

Siarka P (2012) Quality measures of scoring models. J Risk Manag Financ Instit 5:432–446

Siddiqi N (2005) Credit risk scorecards: developing and implementing intelligent credit scoring (Wiley and SAS Business Series), 1st ed. Wiley

Song X, Ding Y, Huang J, Ge Y (2010) Feature selection for support vector machine in financial crisis prediction: a case study in China. Expert Syst 27:299–310. doi:10.1111/j.1468-0394.2010.00546.x

Sousa MR, Gama J, Brandão E (2013) Introducing time-changing economics into credit scoring (No. 513), FEP Working Papers. Universidade do Porto, Faculdade de Economia do Porto

Sousa MR, Gama J, Gonçalves MJS (2014) A two-stage model for dealing with temporal degradation of credit scoring. arXiv:1406.7775 [q-fin]

Lessmann S, Thomas LC, Seow H-V, Baesens B (2013). Benchmarking state-of-the-art classification algorithms for credit scoring: A ten-year update. Presented at the Credit Scoring and Credit Control XIII

Stepanova M, Thomas L (2002) Survival analysis methods for personal loan data. Oper Res:277–289

Stiglitz JE, Weiss A (1981) Credit rationing in markets with imperfect information. Am Econ Rev 71:393–410
Tang B, Qiu SB (2012) Multi-class support vector machine for credit scoring. Appl Mech Mater 235. doi:10.4028/www.scientific.net/AMM.235.419
Thomas LC (2003) Consumer credit modelling: context and current issues, Workshop Paper Presented on the Banff Credit Risk conference
Thomas LC, Edelman DB, Crook JN (2002) Credit scoring & its applications (monographs on mathematical modeling and computation), 1st ed. Society for Industrial and Applied Mathematics
Thomas LC, Ho J, Scherer WT (2001) Time will tell: behavioural scoring and the dynamics of consumer credit assessment. IMA J Manag Math 12:89
Thomas LC, Malik M (2010) Comparison of credit risk models for portfolios of retail loans based on behavioural scores. In: Rausch D, Scheule H (Eds) Model risk in financial crises. Risk Books, pp 209–232.
Thomas LC, Oliver RW, Hand DJ (2005) A survey of the issues in consumer credit modelling research. J Oper Res Soc 56:1006–1015. doi:10.1057/palgrave.jors.2602018
Tsai C-F (2009) Feature selection in bankruptcy prediction. Knowl-Based Syst 22:120–127. doi:10.1016/j.knosys.2008.08.002
Tsai C-F, Chen M-L (2010) Credit rating by hybrid machine learning techniques. Appl Soft Comput 10:374–380. doi:10.1016/j.asoc.2009.08.003
Tsymbal A (2004) The problem of concept drift: definitions and related work. Computer Science Department, Trinity College Dublin
Turner MA, Walker PD, Dusek K (2009) New to credit from alternative data. Information Policy Institute/Political and Economic Research Council
Van Gestel T, Baesens B, Martens D (2010) From linear to non-linear kernel based classifiers for bankruptcy prediction. Neurocomput Int J 73:2955–2970. doi:10.1016/j.neucom.2010.07.002
Verikas A, Kalsyte Z, Bacauskiene M, Gelzinis A (2010) Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. Soft Comput 14:995–1010. doi:10.1007/s00500-009-0490-5
Waad B, Ghazi BM, Mohamed L (2013) A three-stage feature selection using quadratic programming for credit scoring. Appl Artif Intell 27:721–742. doi:10.1080/08839514.2013.823327
Wang J, Hedar A-R, Wang S, Ma J (2012) Rough set and scatter search metaheuristic based feature selection for credit scoring. Expert Syst Appl 39:6123–6128. doi:10.1016/j.eswa.2011.11.011
Widmer G, Kubat M (1996) Learning in the presence of concept drift and hidden contexts. Mach Learn 23:69–101
Yao P (2009) Fuzzy rough set and information entropy based feature selection for credit scoring. In: 6th international conference on fuzzy systems and knowledge discovery, 2009. FSKD'09. Presented at the 6th International Conference on Fuzzy Systems and Knowledge Discovery, 2009. FSKD'09, pp. 247–251. doi:10.1109/FSKD.2009.713
Yu L, Wang S, Lai KK (2008a) Credit risk assessment with a multistage neural network ensemble learning approach. Expert Syst Appl 34:1434–1444. doi:10.1016/j.eswa.2007.01.009
Yu L, Wang S, Lai KK, Zhou L (2008b) Bio-inspired credit risk analysis: computational intelligence with support vector machines. Springer, Heidelberg
Žliobaitė I (2009) Learning under concept drift: an overview (Technical Report). Faculty of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania

Solving Portfolio Optimization Problems Using AMPL

Alexis Karakalidis and Angelo Sifaleras

Abstract This work presents a new optimization software library which contains a number of financial optimization models. Roughly speaking, the majority of these portfolio allocation models aim to compute the optimal allocation investment weights, and thus they are particularly useful for supporting investment decisions in financial markets. Algebraic modeling languages are very well suited for prototyping and developing optimization models. All the financial optimization models have been implemented in AMPL mathematical programming modeling language and solved using either Gurobi Optimizer or Knitro (for those models having general nonlinear objectives). This proposed software library includes several well-known portfolio allocation models, such as the Markowitz mean-variance model, the Konno-Yamazaki absolute deviation model, the Black-Litterman model, Young's minimax model and others. These models aim either to minimize the variance of the portfolios, or maximize the expected returns subject to a number of constraints, or include portfolios with a risk-free asset, transaction costs, and others. Furthermore, we also present a literature review of financial optimization software packages and discuss the benefits and drawbacks of our proposed portfolio allocation model library. Since this is a work in progress, new models are still being added to the proposed library.

Keywords Financial optimization • Mathematical programming • AMPL

1 Introduction

In today's complex environment of globalization, constantly increased competition, liberalization of markets and rapid changes in the international economic environment, the use of information technology in the field of finance is of vital importance for minimizing risks. Portfolio optimization mainly focus on asset allocation

A. Karakalidis • A. Sifaleras (✉)
Department of Applied Informatics, School of Information Sciences, University of Macedonia, Thessaloniki, Greece
e-mail: sifalera@uom.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_8

(Ibbotson 2010), so the research field of financial optimization utilizes the capabilities of the information technology in order to ensure the investor's portfolio assets.

Due to inherent risks of economic-financial investments, various researchers tackle these problems by developing mathematical models for financial optimization. The novel work of Harry Markowitz on the well-known mean-variance model (Markowitz 1952, 1959) has set the basis for the development of this field. Moreover, six decades later several studies are still presented that derive from the Markowitz mean-variance model (Zopounidis et al. 2014; Kolm et al. 2014; Adame-García et al. 2015). A number of mathematical models based on the classic Markowitz model, have been also recently proposed in electricity investments and energy (Zhu and Fan 2010). More specifically, the variance of the returns of the portfolio's stocks is a measure of risk so the majority of these mathematical models aim to minimize the variance (risk).

Despite the fact that there is a large number of efficient financial optimization software packages, more research efforts are still required in order to develop a more user-friendly non-commercial software package. This is the motivation of this research work; to develop a software library with a large number of the most well-known financial optimization models. The benefits of the proposed financial optimization software library are: first, it utilizes the rich features of the AMPL modelling language (Fourer et al. 2002), second the codes of the financial optimization models can be easily extended, and third that the majority of the operational research (OR) scientists are more familiar to general-purpose optimization modelling languages, rather than other programming languages. However, this is a work in progress so new models are still added to enrich the existing library.

The chapter is organised as follows. The next section presents a literature review on financial optimization software packages. Section 3 presents the financial optimization models that are currently implemented in the proposed optimization software library. Furthermore, some implementation details regarding AMPL and examples of two indicative financial optimization models with solution interpretations are also given in this section. Finally, the last Section concludes this work and discusses some future research directions.

2 Literature Review and Existing Portfolio Optimization Software Packages

There are several optimization software packages for the solution of financial optimization problems. Apart from their price, all these financial optimization software packages differ in the number of the mathematical models they offer, in the variety of produced reports, in the solution time, in reliability, etc. An indicative list of some of the most well-known financial optimization software packages is presented in this section. Table 1 presents some brief information about these

Table 1 Portfolio optimization packages—general information

Package	Access	Add in package	Operating system
Hoadley portfolio optimizer	Commercial	Microsoft Excel	Microsoft Windows
StockPortfolio, Quadprog	Open-source	R	Microsoft Windows, Mac OS
Financial toolbox	Commercial	Matlab	Microsoft Windows, Linux, Mac OS X
Smartfolio	Commercial	Microsoft Excel	Microsoft Windows
MVO Plus	Commercial	–	Microsoft Windows
NUOPT	Commercial	S+	Microsoft Windows, Linux, Solaris
Mvport	Non-commercial	Stata	Windows, Mac OS X, Unix, Linux
iOptima	Commercial	–	Microsoft Windows
Zephyr AllocationADVISOR	Commercial	–	Microsoft Windows

software packages, regarding their availability, required operating system, and whether they constitute add-in packages or not.

Following in Table 2, size limitations and data downloading/importing and reporting capabilities of these software packages, are also reported.

Microsoft Excel[1] Microsoft Excel is one of the simplest software packages for financial optimization. The add-in package called Solver tool can be used for modeling and solving several financial optimization problems, such us a quadratic problem (Çetin and Göktas 2009). Using the Solver tool, one can easily define the cells that contain the optimal stock allocations, define the cell containing the objective function to be optimized, and also add the necessary constraints. The most classical example of a quadratic problem in portfolio optimization is Markowitz mean-variance model and a practical approach is discussed in (Myles and Mangram 2013). Consequently, data tables of the, e.g., stock returns or covariance matrices have to be read in order to make the above computations. Polak et al. (2010) have also used Microsoft Excel Solver for solving minimax portfolio optimization problems. Additionally, Livingston (2013) showed that it is important to use Excel's built-in MMULT function for finding efficient portfolios. Recently, Lee (2015) also reported the use of the MMULT function for solving risk loan portfolio optimization model based on CvaR risk measure. Although, Microsoft Excel does not have any specific financial optimization package pre-installed, the commercial Hoadley Portfolio Optimizer[2] add-in for Microsoft Excel already exists

[1] http://www.solver.com

[2] http://www.hoadley.net/options/develtoolsoptimize.htm

Table 2 Portfolio optimization packages—data management capabilities

Package	Data downloading/ importing	Reporting capabilities	Size limitation
Hoadley portfolio optimizer	Historic stock prices from any web source, from external data	Excel spread-sheets, graphic visual charts	Up to 1,048,576 rows and 16,384 columns
StockPortfolio, Quadprog	Historic stock prices from Yahoo Finance	Text output	Up to 2^31-1 elements (2,137,483,647)
Financial toolbox	Historic stock prices from numerous web sources, Importing external databases	Graphic-visual reports	Up to 2^48-1 elements and 8 TB maximum size of matrix (depending on the OS and software's version)
Smartfolio	Historic stock prices from Yahoo Finance, External databases	Excel spread-sheets, graphic-visual charts	Up to 1,048,576 rows and 16,384 columns, up to 64 portfolios for individual license
MVO Plus	Historic stock prices from external databases	Graphic-visual reports	Up to 20 total assets-stocks (Five in trial version)
NUOPT	Historic stock prices from external databases	Text output, graphic-visual reports	–
Mvport	Historic stock prices from Yahoo Finance	Text output	Up to 32,767 variables
iOptima	Historic stock prices from external databases	Graphic-visual reports	Undisclosed
Zephyr AllocationADVISOR	Historic stock prices from external databases	Graphic-visual reports	Undisclosed

in the market. Hoadley Portfolio Optimizer mostly uses Markowitz mean variance model in order to find the optimal portfolio. Additionally, it uses the Sharpe model and it has also indicative tutorials for both the Black-Litterman model and the general use of the package.

R[3] R is an open-source software which is mostly used for statistical computing. However, it is also efficient on dealing with financial optimization problems (Pfaff 2012). A number of packages implemented in the R programming language assist the writing of the code for financial optimization models. Some of these packages are StockPortfolio and Quadprog. StockPortfolio can be used to download stock data from Yahoo Finance, build mathematical models and calculate the optimal asset allocations. By using the GetReturn function, the returns of the stocks can be downloaded. Also, by using the stockModel and the OptimalPort functions a model

[3] https://www.r-project.org/

can be built and the optimal asset allocations can be calculated, respectively. The Quadprog package constitutes a solver for quadratic programming problems for computing the optimal portfolio using the Markowitz mean-variance model.

LINGO[4] LINGO is a general-purpose mathematical optimization language and, thus, it can be used for financial optimization problems as well. Soleimani et al. recently reported an application of a genetic algorithm for Markowitz-based portfolio selection with minimum transaction lots, cardinality constraints and regarding sector capitalization and compared it with the LINGO exact solver (Soleimani et al. 2009).

MATLAB[5] MATLAB (MATrix LABoratory) is one of the most well-known computational tools for scientific computing applications. Nevertheless, it is also very-well suited for solving financial optimization problems (Brandimarte 2002). MATLAB also contains the financial toolbox (Mathworks 2000), which incorporates Portfolio Optimization and Asset Allocation, thus, providing useful tools for portfolio optimization theory, mean-variance portfolio optimization, Value at Risk analysis, and portfolio analysis. Also, Chen et al. in 2010 demonstrated mean-variance spanning tests using MATLAB. Furthermore the validity of the maximum entropy method for the minimax portfolio selection problem with short sale restriction has been tested in MATLAB and illustrated in (Wu et al. 2009) using real stock data from Shanghai Stock Exchange.

SmartFolio 3[6] SmartFolio is a user-friendly analytical tool and add-in to Microsoft Excel for assisting investors regarding their portfolio. It enables users to manage data and historical returns of the stocks by using four optimization criteria (maximization of expected utility, minimization of target shortfall probability, maximization of Sharpe ratio, and benchmarking). Another interesting feature is that it supports robust portfolio optimization where, in scenario-based optimization, under the worst case scenario the resultant portfolios demonstrate optimal behavior.

GAMS[7] The General Algebraic Modelling System is a mathematical modeling language for the solution of mathematical optimization problems. GAMS handles linear, non-linear and mixed integer optimization problems. A practical application of financial optimization problems and a library of a large number of financial optimization models are presented in (Zenios et al. 2009). This library contains several models such as the Markowitz mean-variance model, the mean absolute deviation model of Konno and Yamazaki (1991), Sharpe's model and others. Moreover a construction of a portfolio model in GAMS and an application on the Athens Stock Exchange is made in (Xidonas et al. 2010).

4 http://www.lindo.com/

5 http://www.mathworks.com/

6 http://www.smartfolio.com/

7 http://www.gams.com/

MVO Plus[8] MVO (Mean-Variance Optimizer) Plus is a commercial tool for mean-variance optimization, based on the classical Markowitz model. Some of its benefits is that it effectively uses both geometric and algebraic means as input to solve Markowitz's model. Furthermore, MVO Plus simulates historic data of stocks in order to gauge their effectiveness (back testing).

S+[9] S+ is a commercial programming environment of Solution Metrics Company for the solution of large-scale optimization problems. It is used for a wide range of applications such as circuit optimization, linear and non-linear problems, statistical analysis and last but not least for financial optimization. S+ includes a tool called NUOPT (Numerical OPTimization) which combines statistical and graphical environments. NUOPT provides user with solvers that efficiently solve linear, mixed integer, quadratic and non-linear optimization problems. As far as its applications in financial optimization are concerned, it mostly makes use of the Markowitz mean-variance model finding the optimal asset allocation even for large-scale problems. Another library of financial optimization models is also presented in (Scherer and Martin 2005).

Stata[10] Stata is a commercial, general purpose software for data management, statistical analysis, and simulation and regression analysis. It is being used as a research tool in several fields like sociology, political science, biomedicine and economics. Due to the non-commercial add-in package mvport (Dosamantes 2013), it is also used in financial optimization too. Mvport consists a Stata package for mean-variance portfolio optimization. Also, it has a number of functions that collect online updated data for stocks and analyze them, in order to find the minimum variance of the portfolio and thus the optimal asset allocations.

iOptima[11] iOptima is a commercial software of Finvent company for portfolio optimization. It offers a user friendly environment and a number of efficient solvers for several financial problems. As far as the portfolio optimization is concerned, it uses the classical Markowitz mean-variance approach.

Zephyr AllocationADVISOR[12] Zephyr AllocationADVISOR is a commercial optimization software, which is specialized in portfolio simulation and optimization. Some of its key benefits are the creation of efficient portfolios custom to the investor's view and his profile, detailed comparison of different portfolios, portfolio projection, and graphical presentation of the results. The portfolio optimization and asset allocation utilities of Zephyr AllocationADVISOR, use either the classic Markowitz mean-variance model or the Black Litterman model.

---

[8] http://www.effisols.com/

[9] http://www.solutionmetrics.com.au/

[10] http://www.stata.com/

[11] http://www.finvent.com/

[12] http://www.styleadvisor.com/

Table 3 Library of portfolio optimization models implemented in AMPL

Name	Aim	Objective	Type of problem	Solver
Markowitz (MVO)	Minimize	Variance	Quadratic	Gurobi
Markowitz with upper bound	Minimize	Variance	Quadratic	Gurobi
Markowitz with risk free asset	Minimize	Variance	Quadratic	Gurobi
Markowitz with transaction costs	Minimize	Variance	Quadratic	Gurobi
Sharpe	Maximize	Sharpe ratio	Non-linear	KNITRO
Factor	Minimize	Variance	Non-linear	KNITRO
Black-Litterman	Minimize	Variance	Quadratic	KNITRO
Konno-Yamazaki(MAD)	Minimize	Mean absolute deviation	Linear	Gurobi
Conditional Value at Risk (VaR)	Minimize	Conditional value at risk	Linear	Gurobi
Young	Maximize	Minimum returns	Linear	Gurobi

3 Financial Optimization Mathematical Models

The proposed software optimization library consists of ten models as shown in Table 3. All these models are either linear (Mansini et al. 2014), non-linear, or quadratic; thus, a suitable solver is called in each case. Furthermore each one model has a specific aim and objective. More specifically this software optimization library consists of the most well-known Markowitz's mean-variance model (Markowitz 1952, 1991; Steinbach 2001), which aims to minimize the risk of the portfolio subject to a number of constraints. In this case, the variance between the stocks is the measure of risk that has to be minimized.

Furthermore, there are some variations of the classic Markowitz model such as the Markowitz's mean-variance model with upper bound, where the optimal allocations do not exceed an upper bound. Other variants include the Markowitz's mean-variance model with risk free asset, where a risk free asset (Pang 1980) is added to the portfolio to examine its effectiveness in relation with the other risky assets. Moreover, the Markowitz's mean-variance model with transaction costs where the model, apart from seeking the optimal asset allocations, offers also the optimal amounts which should be additionally bought and sold. All the variations of Markowitz's models and the classical model are examples of quadratic programming models.

Additionally, the Sharpe model (Sharpe 1989, 1992, 1994) aims to maximize the portfolio's Sharpe ratio subject to budget constraint, and the Factor model which minimizes the risk given the fact that a factor has affected the portfolio.

Furthermore, the formulation of the Black Litermans' model (Black and Litterman 1992) is almost the same as that of Markowitz classic mean-variance model, however, it also considers the market equilibrium in combination with investor's view. Roughly speaking, the difference lies in the expected returns of the stocks where in classical Markowitz's model these are just the arithmetic means of the individual stocks and in Black Littermans' model they are affected also by the market equilibrium and the investor's view as mentioned before.

Moreover, the Konno-Yamazaki mean absolute deviation (MAD) model (Konno and Yamazaki 1991) aims to minimize the mean absolute deviation subject to a number of constraints.

The Conditional Value at Risk model (CVaR) (Rockafellar and Uryasev 2000) aims to minimize the conditional value which is at risk or correspondingly the probability of large losses of the portfolio. Last but not least, the Young's minimax model (Young 1998) uses games theory methodology and aims to maximize the minimum returns of the portfolio subject to some constraints.

Despite the fact that Black-Litterman's model is a quadratic problem, a non-linear solver has been used in this case, due to the fact that non-linear computations had to be preprocessed in order to obtain the expected returns of the stocks. This is due to the fact that, a self-contained/native AMPL code was used for the formulation of the library models, without any link to other callable software to make extra computations.

3.1 AMPL Implementation Details

AMPL (A Mathematical Programming Language) is a mathematical modelling language for the solution of large optimization problems. Financial optimization constitutes an important optimization problem, and thus AMPL is able not only to efficiently model such problems but also to solve them by calling appropriate state-of-the-art solvers. The architecture of the proposed library is illustrated with analytical flow charts for each step. Initially, the general idea of the library and the way AMPL works is depicted in Fig. 1.

The user has to use three files; the model file which contains the portfolio optimization model, the data file, and finally the script file which has to be run in order to get the optimal asset allocation. The architecture of these files is shown below. More specifically in Figs. 2 and 3 the architectures of the model file and the script file are illustrated, respectively. To begin with the model file as shown in Fig. 2, the formulation of the model file consists of some simple steps as defining the parameters of the model, the variables which denote the asset allocations, the formulation of the objective function, and the constraints.

The next step is the preparation of the data file which contains all the data required by the model. Firstly the returns of the stocks have to be downloaded and included in this file, then the portfolio's required return and the available budget have to be declared which vary according to the decision maker (investors) and requirements. Finally, an upper bound of investment may also be declared. Afterwards, the script file loads the model, reads the data, and sets various options (e.g.,

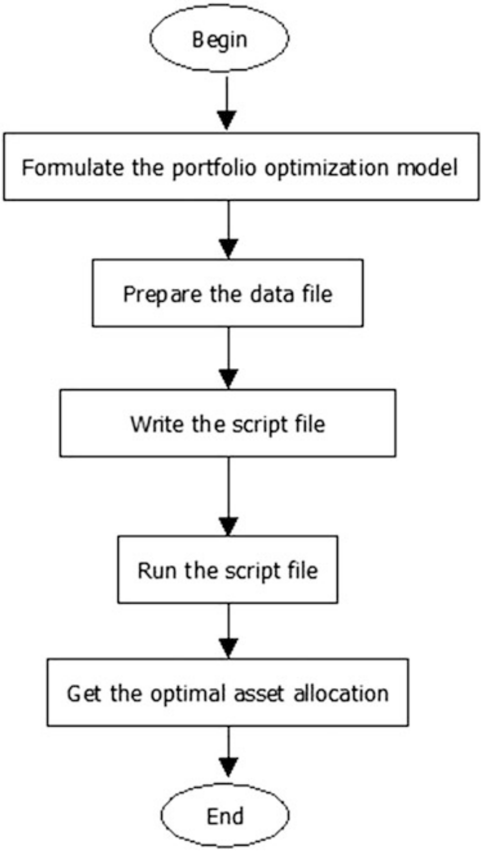


Fig. 1 General architecture of the library

selection of solver, time limit, format of the results) and returns the optimal asset allocations computed by the solver.

Gurobi or CPLEX solvers have been used, since the majority of these models have linear objectives. However some more complex models due to their non-linear objectives, force us to use other solvers such as KNITRO or Minos. Following in the next subsection, two indicative financial optimization models of the proposed library will be presented.

3.2 Konno-Yamazaki Mean Absolute Deviation Model

In 1991 Hiroshi Konno and Hiroaki Yamazaki introduced the Mean Absolute Deviation (MAD) model (Konno and Yamazaki 1991). The central idea of this model is that since the measure of risk in this case is considered the mean absolute deviation, there is no need to compute the covariance matrix as in classic

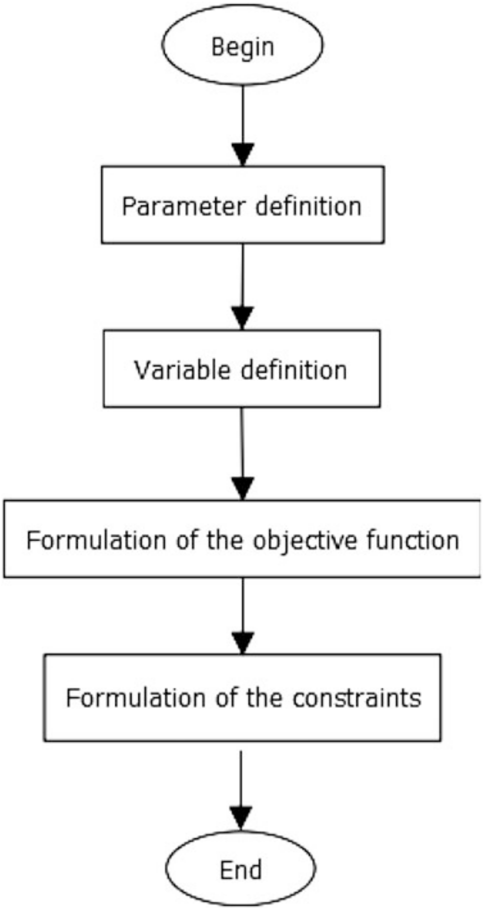


Fig. 2 Architecture of the model file

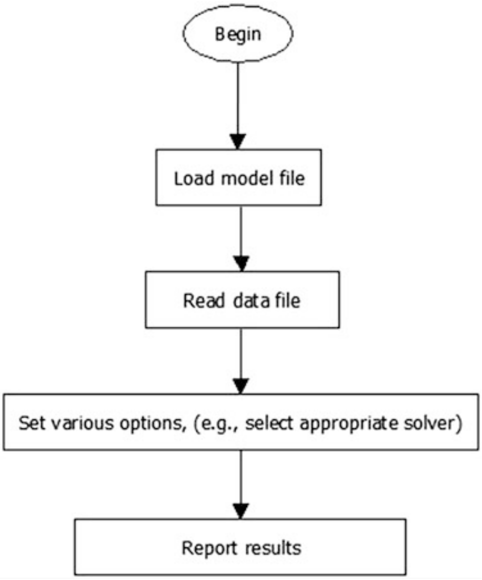
Markowitz mean variance model. Therefore the objective function seeks to minimize the mean absolute deviation of the portfolio subject to some constraints. The mathematical formulation of the Konno Yamazaki MAD model is as follows:

$$\min \sum_{t=1}^{T} \left| \sum_{j=1}^{n} a_{jt} x_j \right| / T$$

subject to

$$\sum_{j=1}^{n} r_j x_j \geq \rho M_0$$

Fig. 3 Architecture of the script file



$$\sum_{j=1}^{n} x_j = M_0$$

$$0 \leq x_j \leq u_j \quad j = 1 \ldots n$$

Which is also equivalent to its linear form:

$$\min \sum_{t=1}^{T} y_t / T$$

$$s.t.$$

$$y_t + \sum_{j=1}^{n} \alpha_{jt} x_i \geq 0 \quad t = 1, \ldots T,$$

$$y_t - \sum_{j=1}^{n} \alpha_{jt} x_i \geq 0 \quad t = 1, \ldots T,$$

$$\sum_{j=1}^{n} r_j x_j \geq \rho M_0$$

$$\sum_{j=1}^{n} x_j = M_0$$

$$0 \leq x_j \leq u_i \quad \text{j} = 1, \ldots n$$

- The target of the objective function is the minimization of the portfolio's mean absolute deviation (MAD)

  n : number of the assets
  x_j: optimal allocation
  T: time period e.g. years
  a_{jt}: mean absolute deviation (MAD) of the stocks
  r_j: expected returns of the assets
  M_0: total budget
  ρ: target return
  u_j: upper bound of optimal allocations

  Example Konno and Yamazaki's MAD model aims to minimize the mean absolute deviation of the portfolio, given the historical annual returns retrieved from Yahoo Finance. For each one of the 5 years (2010–2015) for eighteen (18) stocks as shown in Table 4, the minimum target return, that a moderate investor requires from the portfolio (0.095 per annum), and the total budget he holds ($150,000). More specifically the expected returns of each one of the three stocks are the arithmetic average of the historic returns for all 5 years. The mean absolute deviation is the absolute value of the subtraction of the returns of the stocks[i,j] from the expected return of stock[j]. Therefore according to Konno and Yamazaki, this quantity has to be minimized, provided that:

- The sum of the expected returns of the assets multiplied by the optimal allocations is no less than the goal value.
- The sum of invested capital should be equal to a pre-specified investor's budget (M_0).
- The optimal allocations should be non-negative and should not exceed the upper bound of investment (if there is one).

3.2.1 Results Interpretation

The optimal asset allocations and the set of stocks are analytically shown in Table 4:

3.3 Young's Minimax Model

During the last 15 years game theory had a large impact in portfolio optimization. Many studies have been reported that combined these two fields, resulting to some important portfolio optimization models based on the minimax solution. Young's

Table 4 Konno and Yamazaki MAD model optimal asset allocation

Stock name	Full stock name	Stock exchange	Optimal asset allocation ($)	Optimal asset allocation (%)
CAT	Caterpillar Inc.	NYSE	3389.47	2.26
ORCL	Oracle Corporation	NYSE	6435.8	4.291
RIG	Transocean Ltd.	NYSE	1602.7	1.068
VDSI	VASCO Data Security International Inc.	NasdaqCM	18735.6	12.49
XOM	Exxon Mobil Corporation	NYSE	4798.37	3.199
FCEL	FuelCell Energy Inc.	NasdaqGM	3163.84	2.109
MSFT	Microsoft Corporation	NasdaqGS	18401.2	12.267
INL. DE	Intel Corporation	XETRA	14827.2	9.885
SNE	Sony Corporation	NYSE	3024.13	2.016
BTX	BioTime, Inc.	NYSE MKT	2191.36	1.461
QCOM	QUALCOMM Incorporated	NasdaqGS	4190.68	2.794
RIG	Transocean Ltd.	NYSE	1596.79	1.065
KNDI	Kandi Technologies Group, Inc.	NasdaqGS	9651.89	6.435
HOT	Starwood Hotels & Resorts Worldwide Inc.	NYSE	5281.46	3.521
AIZ	Assurant Inc.	NYSE	44244.4	29.496
VNO	Vornado Realty Trust	NYSE	4860.33	3.24
TDW	Tidewater Inc.	NYSE	1548.4	1.032
AAPL	Apple Inc.	NasdaqGS	2056.33	1.371
Total			150,000	100

Minimax model (Young 1998) was the novel work in the combination of game theory with portfolio optimization, and a few years later other similar models such as Cai's model (Cai et al. 2000), Teo's model (Teo and Yang 2001) and Deng's model (Deng et al. 2005) followed. Also a comparison test has been implemented to prove that the optimal allocations of Konno-Yamazaki MAD model are very similar to the classic Markowitz's Mean-Variance model whereas Young's Minimax model and Markowitz's MVO differ somehow (Hoe et al. 2010). In this subsection Young's Minimax model is presented.

In 1998 Martin Young introduced the minimax model in portfolio optimization (Young 1998). The essence of this model lies in the minimax formulation of game theory, so the objective function is to maximize the minimum returns of the portfolio subject to some constraints. The mathematical formulation of Young's minimax model is:

$$\max_{M_p, w} M_p$$

$$\textit{subject to}$$

$$\sum_{j=1}^{N} w_j y_{jt} - M_p \geq 0, \quad t = 1 \ldots T$$

$$\sum_{j=1}^{N} w_j \bar{y}_j \geq G$$

$$\sum_{i=1}^{N} w_j \leq W$$

$$0 \leq w_j \leq u, \quad j = 1 \ldots N$$

- The target of the objective function is the maximization of the portfolio's minimum returns (Mp)

  w_j: Optimal allocation
  M_p: Portfolio's minimum returns
  y_{jt}: Historic monthly returns of the shares
  $\bar{y}_j$: Expected returns of the assets
  W: Investor's budget
  G: Target return
  N: Number of the assets
  u: Upper bound of optimal allocations
  T: Time period e.g. months

The AMPL model file as also the corresponding script file are as follows (Table 5):

Example Young's minimax model aims to maximize the minimum returns of the portfolio given the same historical annual returns for the stocks used in the previous example, the minimum target return that the same investor requires from the portfolio (0.095), the total budget he holds ($150,000) and the upper bound of investment in a single stock ($120,000). More specifically the expected returns of each one of the three stocks are the arithmetic average of the historic returns for all 12 months. The minimum returns are the result of subtraction of standard deviation of stock[j] from the expected return of stock[j].

3.3.1 Results Interpretation

The optimal asset allocations and the set of stocks are analytically shown in Table 6:

Table 5 AMPL model and script files, for the Young's minimax model

AMPL model file

```
# parameter definition
param n > 0;                 # number of shares
param T > 0;                 # number of months
param W;                     # budget
param RetMat{1..T, 1..n};    # historic monthly returns for selected shares
param u;                     # Upper limit for investing in a single share
param G;                     # target return of the portfolio
param Mp{1..n};              # minimum portfolio
param ExpRet{1..n};          # expected returns of shares
param stdv{1..n};            # standard deviation of shares

# variable definition
var w{1..n} >=0;

# objective function
maximize MinimumReturn:
sum {j in 1..n} w[j]*Mp[j];

# constraints
subject to TargetReturn:
sum {j in 1..n} ExpRet[j]*w[j] >= G;

subject to Budget:
sum {j in 1..n} w[j] <= W;

subject to bounds {j in 1..n}:
0 <= w[j] <= u;
```

AMPL script file

```
# script file for the Young Minimax model

model Young.mod;
data Young.dat;

let {j in 1..n} ExpRet[j] := sum{i in 1..T} RetMat[i,j]/T;
let {j in 1..n} stdv[j] := sqrt((sum{i in 1..T} (RetMat[i,j] -ExpRet[j])^2)/T);
let {j in 1..n} Mp[j] := ExpRet[j]-stdv[j];

option solver gurobi_ampl;
solve;
display w;
```

Table 6 Young's minimax model optimal asset allocation

Stock name	Full stock name	Stock exchange	Optimal asset allocation ($)	Optimal asset allocation (%)
CAT	Caterpillar Inc.	NYSE	0	0.00
ORCL	Oracle Corporation	NYSE	30,000	20.00
RIG	Transocean Ltd.	NYSE	0	0.00
VDSI	VASCO Data Security International Inc.	NasdaqCM	0	0.00
XOM	Exxon Mobil Corporation	NYSE	0	0.00
FCEL	FuelCell Energy Inc.	NasdaqGM	0	0.00
MSFT	Microsoft Corporation	NasdaqGS	120,000	80.00
INL.DE	Intel Corporation	XETRA	0	0.00
SNE	Sony Corporation	NYSE	0	0.00
BTX	BioTime, Inc.	NYSE MKT	0	0.00
QCOM	QUALCOMM Incorporated	NasdaqGS	0	0.00
RIG	Transocean Ltd.	NYSE	0	0.00
KNDI	Kandi Technologies Group, Inc.	NasdaqGS	0	0.00
HOT	Starwood Hotels & Resorts Worldwide Inc.	NYSE	0	0.00
AIZ	Assurant Inc.	NYSE	0	0.00
VNO	Vornado Realty Trust	NYSE	0	0.00
TDW	Tidewater Inc.	NYSE	0	0.00
AAPL	Apple Inc.	NasdaqGS	0	0.00
Total			150,000	100

4 Conclusions and Future Work

Finance and decision making theory are among the most useful research fields in today's market. Thus, the focus of this work was to show that the combination of these two research fields, i.e., financial optimization has a plethora of interesting real-world applications. The proposed software library that contains these financial optimization models assists the investor to make the best decision for asset allocations in finance in order to both ensure his investments and have a satisfactory return of his portfolio by giving him a plethora of models which use different methodologies.

It is well-known that algebraic modelling languages are ideal tools for rapid prototyping and optimization model development. Thus, our proposed portfolio optimization models software package utilizes the flexibility and convenience of the AMPL modelling language. A strong point of the proposed work is the variety of the state-of-the-art portfolio optimization models which aim is to advice the investor about the optimal asset allocation.

Generally speaking, only a few financial optimization packages use the mean absolute deviation model of Konno and Yamazaki and/or the game theoretic Young's minimax model. Furthermore this software library of portfolio optimization models is non-commercial (apart of course from AMPL) and every researcher, or investor may use these models. Furthermore, the code of these models can be easily extended or modified, since the majority of the operational and financial researchers are rather more familiar with mathematical modelling languages than with common programming languages.

This work is in progress and more portfolio optimization models (especially game theoretic portfolio optimization models) are still being added to our library. Moreover, as a future research direction we plan to implement all these models, using the open-source GNU MathProg modeling language. The open-source GNU MathProg modeling language features a syntax similar to the one of AMPL. Finally it is worth exploring how multicriteria decision analysis (MCDA) has vitally aided portfolio optimization in the past decades (Zopounidis and Doumpos 2013; Zopounidis et al. 2015), leading us to further enrich our library.

References

Adame-García MV, Rodríguez FF, Sosvilla Rivero S (2015) Portfolios in the Ibex 35 index: alternative methods to the traditional framework, a comparative with the naive diversification in a pre- and post- crisis context. Documentos de Trabajo del Instituto Complutense de Análisis Económico (ICAE) 1(7):1–33

Black F, Litterman R (1992) Global portfolio optimization. Financ Anal J 48(5):28–43

Brandimarte P (2002) Numerical methods in finance a MATLAB-based introduction. Wiley, New York

Cai XQ, Teo KL, Yang XQ, Zhou XY (2000) Portfolio optimization under a minimax rule. Manag Sci 46:957–972

Çetin N, Göktas Ü (2009) An excel solver model for a blending type optimization problem in mining with quadratic programming. Proceedings of the first International Symposium on Sustainable Development, pp 364–368

Chen WP, Chung H, Ho KY, Hsu TL (2010) Portfolio optimization models and mean–variance spanning tests. In: Lee C-F, Lee AC (eds) John Lee handbook of quantitative finance and risk management. Springer, New York, pp 165–184

Deng XT, Li ZF, Wang SY (2005) A minimax portfolio selection strategy with equilibrium. Eur J Oper Res 166:278–292

Dosamantes CAD (2013) Commands for financial data management and portfolio optimization. Presented in Mexican Stata Users Group meeting in Mexico City, México

Fourer R, Gay DM, Kernighan BW (2002) AMPL: a modeling language for mathematical programming. Duxbury Press, Pacific Grove

Hoe LW, Hafizah JS, Zaidi I (2010) An empirical comparison of different risk measures in portfolio optimization. BEH—Bus and Econ Horiz 1(1):39–45

Ibbotson RG (2010) The importance of asset allocation. Financ Anal J 66(2):18–20

Kolm PN, Tütüncü R, Fabozzi FJ (2014) 60 Years of portfolio optimization: practical challenges and current trends. Eur J Oper Res 234(2):356–371

Konno H, Yamazaki H (1991) Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. Manag Sci 37(5):519–531

Lee MC (2015) Risk loan portfolio optimization model based on CvaR risk measure. Ecoforum 4 (2):56–64
Livingston LS (2013) Adding Markowitz and Sharpe to portfolio investment projects. Bus Educ Accredit 5(2):79–91
Mangram ME (2013) A simplified perspective of the Markowitz portfolio theory. Glob J of Bus Res 7(1):59–70
Mansini R, Ogryczak W, Speranza G (2014) Twenty years of linear programming based portfolio optimization. Eur J Oper Res 234(1):518–535
Markowitz H (1952) Portfolio selection. J Financ 7(1):77–91
Markowitz H (1959) Portfolio selection. Wiley, New York
Markowitz H (1991) Foundations of portfolio theory. J Financ 46(2):469–477
MathWorks (2000) Financial toolbox: for use with Matlab: user's guide. MathWorks, New York
Pang J (1980) A parametric linear complementary technique for optimal portfolio selection with a risk-free asset. Oper Res 28(4):927–941
Pfaff B (2012) Financial risk modelling and portfolio optimization with R. Willey, Germany
Polak GG, Rogers DF, Sweeney DJ (2010) Risk management strategies via minimax portfolio optimization. Eur J Oper Res 207(1):409–419
Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J Risk 2:21–41
Scherer B, Martin D (2005) Introduction to modern portfolio Optimization with NuOpt, S-Plus and S+Bayes. Springer, Chicago
Sharpe WF (1989) Mean-variance analysis in portfolio choice and capital markets by Harry M Markowitz. J Financ 44(2):531–535
Sharpe WF (1992) Asset allocation: management style and performance measurement. J Portf Manag 18(2):7–19
Sharpe WF (1994) The Sharpe ratio. J Portf Manag 21(1):49–58
Soleimani H, Golmakani HR, Salimi MH (2009) Markowitz-based portfolio selection with minimum transaction lots, cardinality constraints and regarding sector capitalization using genetic algorithm. Expert Syst Appl 36(1):5058–5063
Steinbach MC (2001) Markowitz revisited: mean-variance models in financial portfolio analysis. SIAM Rev 43(1):31–85
Teo KL, Yang XQ (2001) Portfolio selection problem with minimax type risk function. Ann Oper Res 101:333–349
Wu ZW, Song XF, Xu YY, Liu K (2009) A note on a minimax rule for portfolio selection and equilibrium price system. Appl Math Comput 208(1):49–57
Xidonas P, Mavrotas G, Psarras J (2010) Portfolio construction on the Athens stock exchange: a multiobjective optimization approach. Optimization 59(8):1211–1229
Young MR (1998) A Minimax portfolio selection rule with linear programming solution. Manag Sci 44(5):673–683
Zenios S, Nielson S, Consiglio A (2009) Practical financial optimization: a library of GAMS models. Wiley, New York
Zhu L, Fan Y (2010) Optimization of China's generating portfolio and policy implications based on portfolio theory. Energy 35(1):1391–1402
Zopounidis C, Doumpos M (2013) Multicriteria decision systems for financial problems. TOP 21 (2):241–261
Zopounidis C, Doumpos M, Fabozzi FJ (2014) Preface to the Special Issue: 60 years following Harry Markowitz's contributions in portfolio theory and operations research. Eur J Oper Res 234(2):343–345
Zopounidis C, Galariotis E, Doumpos M, Sarri S, Andrisopoulos K (2015) Multiplecriteria decision aiding for finance: an updated bibliographic survey. Eur J Oper Res 247(1):339–348

Approximating Throughput of Small Production Lines Using Genetic Programming

Konstantinos Boulas, Georgios Dounias, and Chrissoleon Papadopoulos

Abstract Genetic Programming (GP) has been used in a variety of fields to solve complicated problems. This paper shows that GP can be applied in the domain of serial production systems for acquiring useful measurements and line characteristics such as throughput. Extensive experimentation has been performed in order to set up the genetic programming implementation and to deal with problems like code bloat or over fitting. We improve previous work on estimation of throughput for three stages and present a formula for the estimation of throughput of production lines with four stations. Further work is needed, but so far, results are encouraging.

Keywords Production lines • Genetic programming • Symbolic regression • Throughput

1 Introduction

For the design of serial production lines two types of models are used, the generative models and the evaluative models (Papadopoulos et al. 2009). The evaluative models are used to calculate the various performance measures such as throughput, average work-in-process (WIP), whereas, the generative models are used to optimize the performance measures e.g., to maximize throughput or minimize the average WIP. A K-station production line with $K-1$ intermediate buffers is a system in which, each part enters the system from the first station, passes in order from all the stations and the intermediate buffer locations and exits the line from the

K. Boulas • G. Dounias (✉)
Management and Decision Engineering Laboratory (MDE-Lab), Department of Financial and Management Engineering, Business School, University of the Aegean, Chios, Greece
e-mail: kboulas@aegean.gr; g.dounias@aegean.gr

C. Papadopoulos
Department of Economics, Aristotle University of Thessaloniki, Thessaloniki, Greece
e-mail: hpap@econ.auth.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_9

last station. The description and the assumptions of the model are given in Papadopoulos et al. (2002). The basic performance measure in the analysis of production lines is the mean production rate or throughput. Another useful parameter of the system is utilization which is defined as the ratio of mean arrival rate to mean service rate. The domain of serial production lines is complicated due to combinatorial explosion, depending on the number of workstations involved in the examined line, the capacity of buffers existing within the workstations, etc. (Dallery and Gershwin 1992). The number of feasible allocations of B buffer slots among the K−1 intermediate buffer locations increases dramatically with B and K and is given by the formula (1):

$$\binom{B-K+2}{K-2} = \frac{(B+1)(B+2)\ldots(B+K-2)}{(K-2)!} \tag{1}$$

Consequently general formulas for acquiring useful measurements and line characteristics, such as throughput, either can be found for only very short lines with simple characteristics (Hunt 1956), (see Appendix 1) or do not exist at all. The number of the states is an indication of the difficulties one should overcome in order to find an exact solution for more complex models.

The determination of unknown parameters is usually realized through decomposition algorithms Dallery and Frein (1993) and techniques (Lim et al. 1990; Heavy et al. 1993). These methods presuppose the existence of computer assistance to the production engineer in order to be able to dynamically setup the production line in the optimal way. The overall design of a production line's problem is complicated and difficult. As a result, a few approximation methods (Blumenfeld 1990; Martin 1993; Muth 1987; Li et al. 2015) and algorithms mainly related to computational intelligence (Papadopoulos et al. 2002) have been used in the past, in an attempt to acquire general formulas.

This work attempts to approach the problem by applying genetic programming techniques (Koza 1992, 1994; Angeline and Kinnear 1996; Poli et al. 2008) in order to extract useful general formulas. The basic *methodological steps* are described as follows: The problem is treated by increasing gradually the operating parameters i.e. the number of the stations. A combination of servers which follow the exponential distribution is produced and with the appropriate method an accurate training data set for a dependent characteristic is obtained. So far, we have examined four different algorithms to form our training data sets for different line settings. Using the implementation as presented in Papadopoulos et al. (2009) we concluded that our training set will be formed firstly by the available till now exact formulas, secondly by applying the decomposition method beyond the exact formulas. With the available data set we apply a genetic programming algorithm in order to acquire a formula for the given system. The aim is to find an accurate formula for the characteristic of the system using a simple function set with terminals, parameters of the system, such as mean service rate. Some accurate formulas from Hunt's work (Hunt 1956) based on **Markovian** analysis are confirmed for short lines with simple characteristics and now experiments are extended in order to acquire more complex formulas. At the same time the influence of genetic programming setup parameters is investigated in order to improve the effectiveness of the proposed approach.

The rest of the paper is organized as follows: Sect. 2 briefly describes the basic methodological steps of the current approach. Then a reference is made to genetic programming. In Sect. 3 some results are given. In Sect. 4 the paper concludes with the discussion of results and future research.

2 Data and Methodology

As already stated above, the objective of this work is to obtain accurate or approximate formulas that characterize the examined system in terms of the production line parameters (i.e., the number of stations, size of buffers, mean processing time), assuming there are sufficient jobs at the beginning of the line to ensure that the first station is never starved of jobs (saturated line) and that the last station is never blocked.

In order to cope with the complexity of the parameter settings which characterize a system, genetic programming approaches are used. In an attempt to improve the outputs of the genetic programming, we used both C language and Python lan-guage for code implementation. An embodiment using lists instead of strings was constructed, developing an alternative method to accelerate the calculations by storing in memory the trees' fitness. More sophisticated techniques for initializing the trees in order to achieve differentiation are used. Experimentation was per-formed with the values of the parsimony coefficient and the covariant parsimony pressure was also tried. In addition, the application of uniform crossover was at-tempted in order to explore better the solution space. Experiments were extended to those cases showing superior performance by increasing the number of generations for them, or by trying alternative configurations (for example by increasing considerably the mutation). Finally, genetic programming was fed with new training sets whose values were extracted from analysis of real systems.

Hunt's formulas were used in order to generate the complete training data set which was used later as input to the genetic programming scheme for the acquisition of the accurate or approximate formulas for the throughput. A training set for the throughput was obtained by the decomposition algorithm DECO-2 (Papadopoulos et al. 2009).

2.1 *Generation of Training Data*

To identify accurate or approximate generalized formulas for calculating characteristics of short serial production lines with $K = 2, 3, 4$ stations and no intermediate buffers, the methodological steps, given below, were followed for each value of number of workstations K:

If K = 2, 3, 4:

1. STEP 1: Identify the exact throughput values for this K, using the Hunt's accurate formulas or use the DECO-2 algorithm to estimate throughput.
2. STEP 2: Initiate an iterative application of a genetic programming scheme in the full data set acquired from Step 1, to obtain an approximate formula containing basic algebraic operations.
3. STEP 3: Stop the genetic programming process, when an accurate formula has been obtained for the training data set or another parameter of genetic programming has reached the limits set (i.e., maximum number of generation, maximum length of solution).
4. STEP 4: If the genetic programming execution has identified a solution, then transform this solution into a useful infix notation.

2.2 Genetic Programming Implementation

Genetic programming techniques are applied in this paper, to approximate the calculation of throughput or maximum possible utilization in short serial exponential production lines. Due to the complexity of the solution space of the problem, an encoding in a hierarchical expression is required like the tree representation of a solution. In genetic programming, a population of random trees is initially generated, representing programs, i.e., candidate solutions of the given problem. Then, the genetic operations, viz., crossover and mutation are performed on these trees. We use the standard flattened (linear) representation for trees in a prefix notation. The number of generations in an execution varies from 50 to 1000 and the population ranges from 2000 to 40,000 or more, up to 200,000. The initialization method is the growth method, full method, the half and half ramped method and a method in which the maximum depth of the tree is changed following a triangular probability distribution; the selection is performed by using tournament selection. Crossover happens in range from 30 to 90 % at each run. The used point mutation is set at about 10–20 %. When crossover does not happen, then a mutation takes place.

A generation is completed when for the size of the examined population all crossover and mutation operations have been performed. In each generation statistics are extracted as well as a formula with the best fitness of the population which is transformed in infix representation. If the formula satisfies the accuracy criterion then is treated further i.e. is simplified in order to improve the readability and to be easily manageable.

In order to create a population, a function set F and a terminal set T, are primarily defined. The terminal set contains variables as the mean service rate, μ_i or the buffer size, B_i for each stage, i of the process. The functions must be able to "pass" information among each other. The term describing this need is called closure achievement. The function set includes addition, subtraction, multiplication and protected division. In order to create a random tree, we select randomly from the function set, until all tree branches end in terminals. A generation is considered completed, when the number of crossover/mutation operations that have been performed is equal to the number of individuals which compose the population. The selection of each candidate individual for the participation of genetic operation is tournament selection. Sub tree crossover is used. The selection of crossover points is uniform, so every node is chosen equally likely. Point mutation is used. The candidate nodes for mutation in the tree are randomly chosen. The grow method or the half and half ramped initialization method is used to create the initial population. After considering the initialization of a random population and the operators selection, the next step is to determine a fitness function, which will be used for the evaluation of the candidate solutions, and therefore for the selection of the individuals for the crossover and other operations. The fitness function is the sum of the absolute differences between the actual program output calculated with an iterative procedure which exploits the prefix notation of tree representation and the desired output for each fitness case. For each generation, statistics are calculated and collected for the evaluation of the overall procedure.

The implementation of the genetic programming system creates LISP-like symbolic expressions. The structure of the final solution is not predetermined and is created by performing the genetic programming operators. Therefore, a problem is searched through all the candidate combinations of symbolic expressions, which are defined initially by the solver. The genetic programming process followed in our work can be divided into the following five steps:

1. STEP G1: Create a random population of programs using the symbolic expressions provided.
2. STEP G2: Evaluate each program assigning a fitness value according to a pre-specified fitness function, which actually corresponds to the ability of the program to solve the problem.
3. STEP G3: Use reproduction techniques to copy existing programs into the new generation.
4. STEP G4: Recombine genetically the new population with the crossover or mutation operation from a randomly based chosen set of parents after tournament selection.
5. STEP G5: Repeat steps 2–4 until a termination criterion has been reached..
6. To evaluate a program without error (i.e., to achieve closure) it is imperative that the symbolic expressions operated by the genetic programming mechanism are compatible. Therefore, in order to avoid division by zero during the evaluation process, the standard division is substituted by the following expression (called protected division): double div(double x, double y) {if (y == 0) return x; else return (x/y);}

As an initial step in our experiments, to ensure sufficiency, the four basic functions were selected for the considered function set F = {+,−,*,%}. As during the training phase the algorithm allocates a large proportion of computer memory, the authors adopted a steady-state genetic process. In steady state genetic programming, the parents to be recombined are selected from the current population using some criteria, but a child is also chosen from the same population. The recombined genetic program is evaluated and it takes the position of the selected child. A generation is considered as completed when the number of created children is equal to the size of the population. In some experiments at some point the average program size starts growing at a rapid pace (see Fig. 1) with no accompanied by any corresponding increase in fitness (see Fig. 2). This phenomenon is the well known code-bloat Langdon (2000) and in order to restrict this, some empirical techniques have been implemented. A type of parsimony pressure method is adopted (see Fig. 3). According to this, the fitness measure is defined to be: fitness = error + size

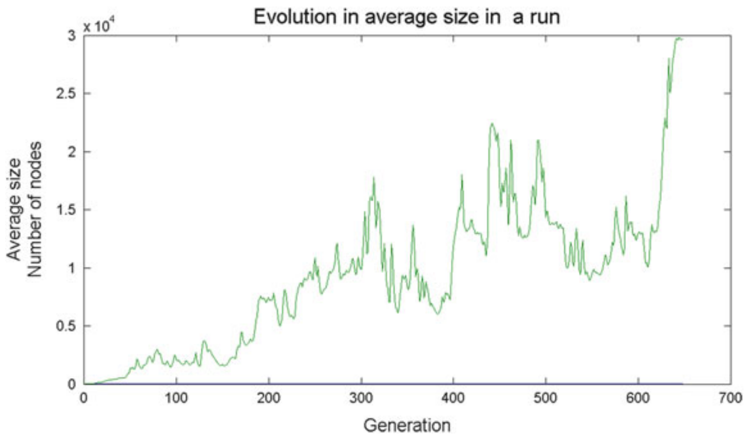


Fig. 1 Average individual size without parsimony fitness measure (max generation is 648)

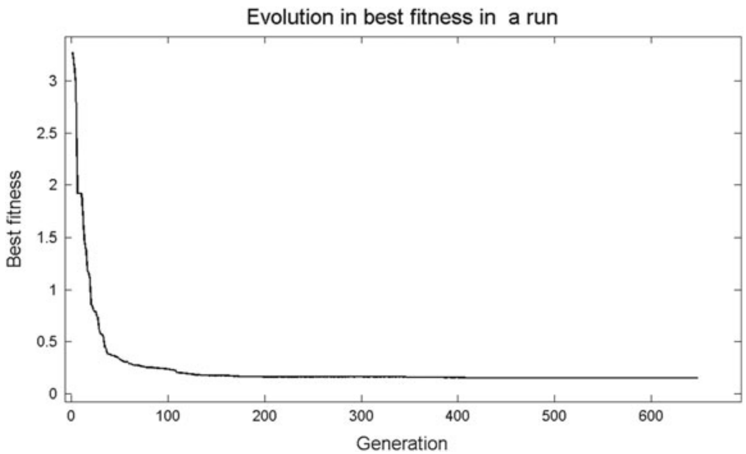


Fig. 2 Best fitness without parsimony fitness measure (max generation is 648)

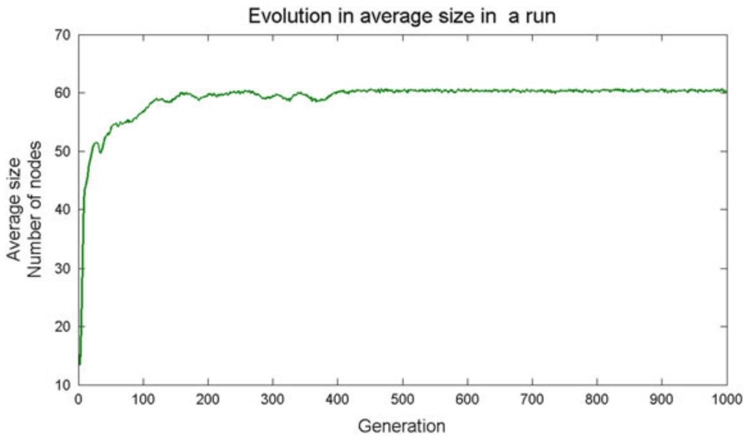


Fig. 3 Average individual size with parsimony fitness measure (max generation is 1000)

of program/ weight. The value of weight is the result of a trial and error method (Lai 2003).

In order to control bloat and to find a better fitness value we utilize some special techniques:

- In order to look at the calculation time we save trees in a data structure. Each tree is paired with a unique ID of the nodes and the value of fitness. When an offspring formed after crossover or mutation a check for the registration of the offspring in memory takes place. If the result is positive GP takes directly the value of fitness from memory. If the result is negative GP figures out the offspring's fitness. A unique ID hooks up the offspring and its fitness value. Then the new ID and the fitness value are stored in memory. Gradually the number of individuals in memory increase. Because the ID consists of an alphanumeric variable that is a symbol for each tree node, memory consumes in proportion to the length of the tree. This procedure functions effectively in the beginning of an experiment, but when the number of trees increasing and is saved in memory, the process proves inefficient. The above procedure is suitable for monitoring the evolution of the population. In general in the experiments conducted, we noticed that in each run about 60 % of unique individuals were present. The requirements of memory were large for long runs or large numbers of population.
- To deal with the code bloat problem we performed the covariant parsimony pressure described in Poli and McPhee (2008). This method was able to held the average size within the imposed limits but we received not so accurate results in fittness.
- To perform a global search of the search space we tried the uniform crossover (Poli and Langdon 1998). This kind of genetic operation needs far more

computations than standard crossover. So we performed this operation in a smaller number of population (maximum number: 2000).

- We performed a series of runs of 10 or 20 with generation numbers up to 50, in order to see how the seed of random generator affects the specific experiment. When good quality results were received, we increased the number of generations up to 1000 to achieve a fitness closer to zero.

3 Numerical Results

We consider the case of a line with two or three stations and exponentially distributed processing times. For the case of two identical stations with finite number of buffers in front of second station the genetic programming algorithm gave the exact solution as this is described in Hunt (1956) and Blumenfeld (1990).

Two-Station Lines

We have considered this problem in Boulas et al. (2015). For the case of two stations with $\mu_1 \neq \mu_2$ we aim at validating Eq. (19) of Hunt (1956) i.e. the maximum possible utilization. After performing the GP execution, the requested formula is obtained during the third (and last) generation. The prefix notation for this formula according to the corresponding tree is: $(/(*/\mu2\mu1)\mu1)(+\mu2\,(*\mu1(/\mu1(+\mu2\,\mu1)))))$. We write again in infix notation and it becomes: $((\mu2\,/\mu1)\,{*}\mu1)/(\mu2 + (\mu1{*}\,(\mu1/(\mu2 + \mu1)))))$ which actually is the right side of the Eq. (19) of Hunt (i.e. the exact formula):

$$\rho_{\max,K=2} = \frac{\mu_2(\mu_1 + \mu_2)}{{\mu_1}^2 + \mu_1\mu_2 + {\mu_2}^2} \tag{2}$$

Three-Station Lines

For the case of three stations with $\mu_1 \neq \mu_2 \neq \mu_3$ and without intermediate buffers, the maximum possible utilization is given from Eqs. 21, 22 and 23 of Hunt (1956). So far the genetic programming implementation has approximated the exact solution in range [0.95, 1.06] where 108 training cases were used. An approximation is presented in the following Eq. (3) and the results from this are presented in Fig. 4 where one can find the improvement compared against Boulas et al. (2015). This formula has been obtained from the 999th generation of a GP execution. We note that the fitness is equal to 0.000762522 and is the same after the 824th generation of the run. This improvement shows that genetic programming has the ability to find a solution in a difficult problem if the setup is proper.

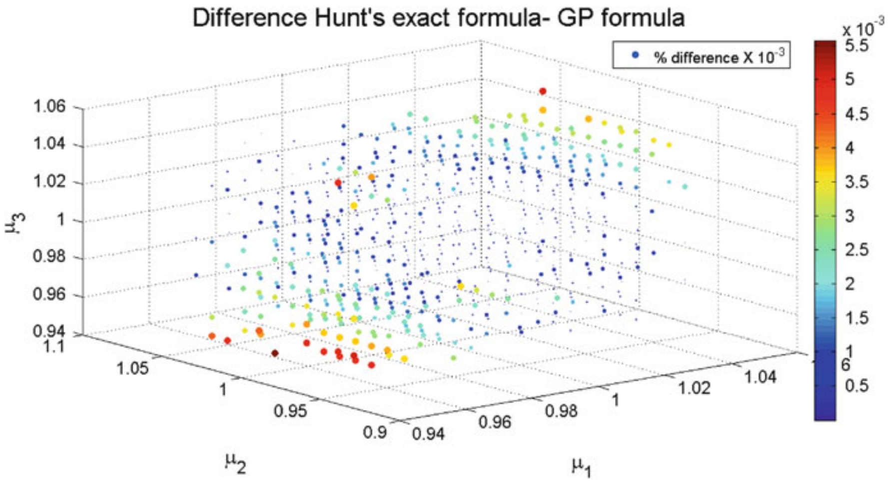


Fig. 4 Difference between Hunt's formula and the proposed GP formula results for three stations in estimation of throughput within the area of the training set. The max % difference is 5.5×10^{-3} for 1000 random points

$$\rho_{\max,k=3} = \frac{\mu_2}{\mu_2 + \frac{\mu_1}{\begin{array}{r} \mu_2 - \frac{\mu_2(\mu_2-\mu_3)}{\mu_1 - \frac{\mu_1-\mu_3}{\mu_2}} \\ \frac{\mu_3}{\mu_1 - \frac{\mu_1-\mu_3}{\frac{2\mu_2}{\mu_1}+\frac{\mu_2^2}{\mu_1^2}+\frac{1}{\mu_2\mu_3}+\frac{\mu_2^2}{\mu_1(\mu_3^2+\mu_1)}}} + \frac{\mu_2}{2\mu_3 + \frac{\mu_2^4}{2\mu_1 - \mu_2(\mu_2-\mu_3) + \frac{\mu_1\mu_2^2\mu_3^2}{\mu_1+\mu_2}} + \mu_2^2} \end{array}}} \tag{3}$$

For the same case the throughput is figured out with DECO-2 algorithm. An approximate formula, same as the one given in Boulas et al. (2015) is presented below and results are shown in Fig. 5. This formula (4) has been obtained from the 998th generation of a GP execution:

$$X_{K=3} = \frac{2}{2\mu_2 + \mu_3 + \frac{\mu_1}{\mu_2 + \frac{\mu_2}{\mu_1\left(\mu_3 + \mu_1^2\mu_2 - \frac{\mu_1^2\mu_2^4\mu_3}{(2\mu_2\mu_1^2+\mu_3)(3\mu_1+\mu_3)}\right)}}} \tag{4}$$

Four-Station Lines

For the case of four stations with $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ and without intermediate buffers, the throughput is figured out with the use of data taken from the DECO-2 algorithm. We use 81 training cases which are the combination of three values for μ_i, $\mu_i \epsilon \{0.98,1,1.03\}$ An approximate formula is presented below and results from

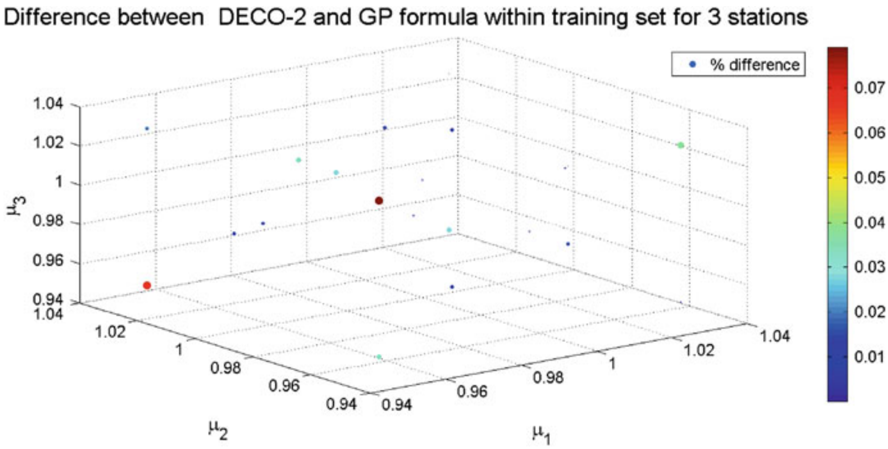


Fig. 5 Difference between DECO-2 algorithm results and the proposed GP formula results for the case of three stations in throughput estimation within training set

Table 1 Comparison of results with DECO-2 for the case of four (4) workstations

M	$X_{DECO\text{-}2}$	X_{GP}	$X_{DECO\text{-}2} - X_{GP}$	% Error
(0.985,1.02,1.01,1.025)	0.480541	0.480521	0.000020	0.00413
(0.982,1.005,0.982,0.982)	0.491324	0.491241	0.000083	0.01687
(0.985,0.995,0.995,0.985)	0.490284	0.490179	0.000105	0.02141
(0.982,1.02,1.005,1.005)	0.483536	0.483484	0.000052	0.01066
(0.995,0.985,1.005,1.005)	0.487225	0.487122	0.000103	0.02112
(1.025,0.982,1.02,0.985)	0.484486	0.484527	−0.000041	−0.00848
(1.01,1.005,0.985,1.02)	0.484215	0.484157	0.000058	0.01193
(1.02,0.982,1.025,1.005)	0.482305	0.482292	0.000013	0.00269
(1.02,1.02,1.025,0.982)	0.479226	0.479279	−0.000053	−0.01114
(1.01,1.01,1.01,1.02)	0.480016	0.479944	0.000072	0.01490

this formula are shown in Table 1, for ten random modulations of the system within the area of the training set. This formula (5) has been obtained from the 350th generation of a GP execution and the corresponding tree is illustrated in Fig. 6:

$$X_{K=4} = \frac{\dfrac{\mu_4}{\mu_1 + \mu_4 + \dfrac{\mu_1}{\mu_4} + \mu_3\left(\mu_3 + \dfrac{\mu_1}{\mu_4 + \dfrac{\mu_2\mu_3}{\mu_3 + \frac{\mu_1}{\mu_3}}}\right)} + 1}{\mu_2 + \mu_4 - \dfrac{\mu_2}{\mu_1{\mu_3}^2 + \mu_2} + 1} \tag{5}$$

Equation (5) is useful for making computations in the following form: X = (((μ1/μ1) + (μ4/(μ1 + ((μ4 + (μ1/μ4)) + (((μ1/(((μ3*μ2)/(μ3 + (μ1/μ3))) + μ4)) + μ3)*μ3)))))/((((μ1/μ1)-(μ2/(((μ1*μ3)*μ3) + μ2))) + μ2) + μ4)) in order to achieve some intuition on how the system could behave in the area of the training set or near it. If we try to analyze Eq. (5), then we obtain the following form of the Eq. (6):

$$X_{K=4} = \frac{N_1 + N_2 + N_3 + N_4 + N_5}{D_{1+}D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_9} \tag{6}$$

Where:

$$N_1 = {\mu_1}^2\mu_2{\mu_4}^2 + {\mu_1}^2\mu_2{\mu_3}^4 + {\mu_1}^3{\mu_3}^2\mu_4 + {\mu_1}^2{\mu_3}^4\mu_4 + {\mu_1}^3{\mu_3}^3\mu_4 + {\mu_1}^2{\mu_3}^5\mu_4$$

$$N_2 = {\mu_1}^3{\mu_3}^2{\mu_4}^2 + 2{\mu_1}^2{\mu_3}^4{\mu_4}^2 + {\mu_1}^2\mu_2\mu_4 + {\mu_1}^2\mu_2\mu_3\mu_4 + {\mu_1}^2\mu_2{\mu_3}^4\mu_4 + 2{\mu_1}^2{\mu_3}^2{\mu_4}^3$$

$$N_3 = \mu_1{\mu_2}^2{\mu_3}^2\mu_4 + \mu_1{\mu_2}^2{\mu_3}^2 + \mu_1\mu_2{\mu_3}^6\mu_4 + 2\mu_1\mu_2{\mu_3}^4{\mu_4}^2 + \mu_1\mu_2{\mu_3}^3\mu_4 + {\mu_2}^2{\mu_3}^4\mu_4$$

$$N_4 = 2{\mu_2}^2{\mu_3}^2{\mu_4}^2 + \mu_2{\mu_3}^4{\mu_4}^2 + 2\mu_2{\mu_3}^2{\mu_4}^3 + 2\mu_1\mu_2{\mu_3}^2{\mu_4}^2 + \mu_1\mu_2{\mu_3}^2\mu_4 + 2\mu_1\mu_2{\mu_4}^3$$

$$N_5 = \mu_1{\mu_3}^6{\mu_4}^2 + 2\mu_1{\mu_3}^4{\mu_4}^3$$

$$D_1 = {\mu_1}^3{\mu_3}^2\mu_4 + {\mu_1}^3{\mu_3}^3\mu_4 + 2{\mu_1}^3{\mu_3}^2{\mu_4}^2 + {\mu_1}^3{\mu_3}^2{\mu_4}^3 + {\mu_1}^3{\mu_3}^3{\mu_4}^2 + {\mu_1}^3\mu_2{\mu_3}^2\mu_4$$

$$D_2 = {\mu_1}^3\mu_2{\mu_3}^3\mu_4 + {\mu_1}^3\mu_2{\mu_3}^2{\mu_4}^2 + {\mu_1}^2{\mu_2}^2{\mu_3}^4\mu_4 + {\mu_1}^2{\mu_2}^2{\mu_3}^4 + {\mu_1}^2{\mu_2}^2\mu_3\mu_4$$

$$D_3 = {\mu_1}^2{\mu_2}^2{\mu_4}^2 + {\mu_1}^2{\mu_2}^2\mu_4 + {\mu_1}^2\mu_2{\mu_3}^5\mu_4 + 3{\mu_1}^2\mu_2{\mu_3}^4{\mu_4}^2 + {\mu_1}^2{\mu_3}^5{\mu_4}^2 + {\mu_1}^2{\mu_3}^5\mu_4 + 2{\mu_1}^2{\mu_3}^4{\mu_4}^3$$

$$D_4 = 3\mu_2{\mu_1}^2{\mu_3}^4\mu_4 + \mu_2{\mu_1}^2{\mu_3}^4 + \mu_2{\mu_1}^2{\mu_3}^2{\mu_4}^3 + \mu_2{\mu_1}^2\mu_3{\mu_4}^2 + \mu_2{\mu_1}^2{\mu_4}^3 + \mu_2{\mu_1}^2{\mu_4}^2$$

$$D_5 = 3{\mu_1}^2{\mu_3}^4{\mu_4}^2 + {\mu_1}^2{\mu_3}^4\mu_4 + {\mu_1}^2{\mu_3}^2{\mu_4}^4 + {\mu_1}^2{\mu_3}^2{\mu_4}^3 + \mu_1{\mu_2}^3{\mu_3}^2\mu_4 + \mu_1{\mu_2}^3{\mu_3}^2$$

$$D_6 = \mu_1{\mu_2}^2{\mu_3}^6\mu_4 + \mu_1{\mu_2}^2{\mu_3}^4{\mu_4}^2 + \mu_1{\mu_2}^2{\mu_3}^3\mu_4 + 3\mu_1{\mu_2}^2{\mu_3}^2{\mu_4}^2 + 2\mu_1{\mu_2}^2{\mu_3}^2\mu_4$$

$$D_7 = \mu_1{\mu_2}^2{\mu_4}^3 + 2\mu_1\mu_2{\mu_3}^6{\mu_4}^2 + \mu_1\mu_2{\mu_3}^6\mu_4 + 2\mu_1\mu_2{\mu_3}^4{\mu_4}^3 + \mu_1\mu_2{\mu_3}^4{\mu_4}^2 + \mu_1\mu_2{\mu_3}^3{\mu_4}^2$$

$$D_8 = {\mu_2}^3{\mu_3}^4\mu_4 + {\mu_2}^3{\mu_3}^2{\mu_4}^2 + 2{\mu_2}^2{\mu_3}^4{\mu_4}^2 + 2{\mu_2}^2{\mu_3}^2{\mu_4}^3 + \mu_2{\mu_3}^4{\mu_4}^3 + \mu_2{\mu_3}^2{\mu_4}^4 + 2\mu_1\mu_2{\mu_3}^2{\mu_4}^3$$

$$D_9 = \mu_1{\mu_3}^6{\mu_4}^3 + \mu_1{\mu_3}^6{\mu_4}^2 + \mu_1{\mu_3}^4{\mu_4}^4 + \mu_1{\mu_3}^4{\mu_4}^3 + \mu_1\mu_2{\mu_3}^2{\mu_4}^2 + \mu_1\mu_2{\mu_4}^4$$

According to this analysis the throughput for four-station lines is a ratio of two polynomials where the nominator has 26 terms and the denominator has 54 terms.

4 Conclusions and Further Research

In this work we have studied the behavior of a genetic programming implementation for extracting useful accurate or approximate formulas for serial production lines. After extensive experimentation we have observed that GP-programs are growing rapidly in size which means consumption of computational resources, are needed for additional processing time and difficulties in fitness improvement. In this sense an effort has been given in order to restrict the code bloat phenomenon. The introduction of the parsimony factor leads to smaller and more manageable programs, while it increases comprehensibility and reduces the time of an execution. In all GP-experiments the search process evolves rather slowly.

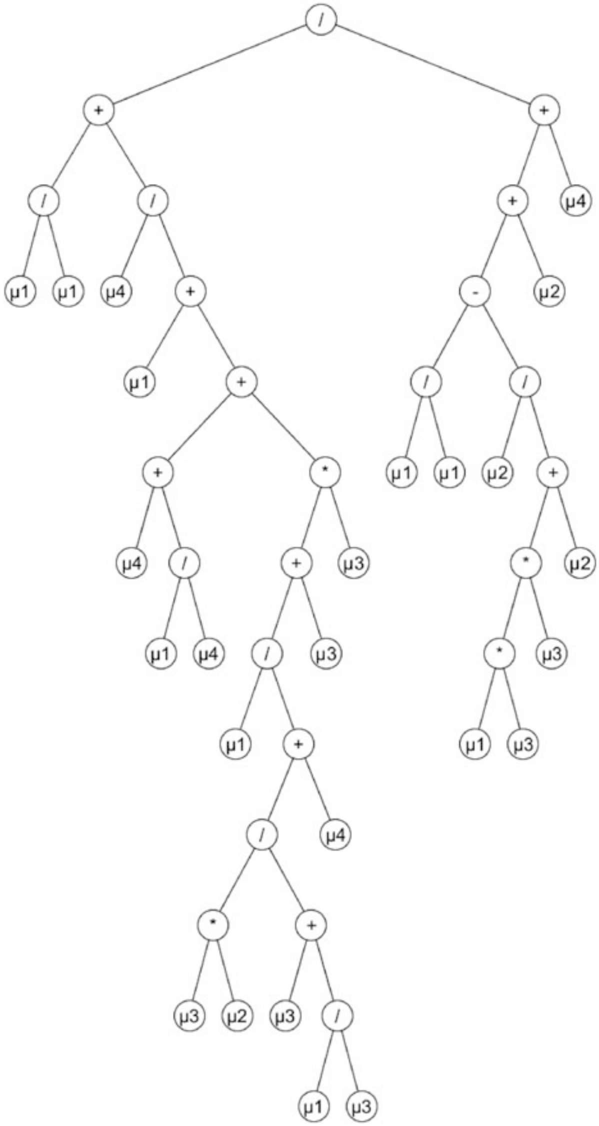
The existence of accurate formulas is very important. The fact that we can induce an accurate formula allows the investigation of a very large area like the one represented in Fig. 5. We are experimenting with methods to prevent code bloat, to improve the search of solution space, to implement different genetic programming operations, to test other function sets or GP setups. Our aim far ahead is to increase the value of fitness and to obtain solutions for larger systems with more than four stations.

Note that this is one of the few times that in practice GP-approaches verify existing accurate formulas at any real world application domain. Moreover, it is the first time that this verification concerns exact analytical formulas in production lines. This happens for short serial production lines with two stations. For three-station lines highly accurate approximate formulas have been proposed so far, performing at an acceptable level of accuracy also out of the range of the training set (out of sample performance), with the use of a rather limited training set.

In our experiments we managed to increase the level of accuracy of the formula and to get closer to the theoretical results. The whole formula discovery process seems to be dependent on the quality and specific characteristics of the initialization process (a high quality solution obtained during the initialization of the GP-tree, leads quickly to more accurate or even exact solutions). In this direction we plan for the future the design and application of GP-ensemble models, to be able to perform several parallel GP-executions, in order to ensure the best possible initialization of the evolution process.

Another issue encountered has to do with the complexity of genetic programming. Genetic programming is a complex system that can expand candidate solutions in length and shape. From the tree of Fig. 6 we observe for example that term μ_1 appears 11 times in a total of 49 knots. Four times of these appearances are used to construct the constant one. If we compare the number of instances of μ_1 in Eq. (6), one may notice we find that it appears 70 times. This observation indicates that genetic programming hides an internal complexity resulting in side effects. For example every time that a mutation is performed in a good candidate solution, results are unpredictable. If we represent the Eq. (3) in a form similar to Eq. (6) it will be noted that the value of the maximum utility is a ratio of two polynomials. The numerator has 150 terms and the denominator has 217 terms. These terms of the equation are similar to Hunt's term however they are rather noisy i.e., displayed powers of order 9 or 10 and coefficients such as 2, 6, 10, 20, positive or negative. This noise seems to cause the error, and acts as a factor which is preventing fitness to be closer to zero. These terms resemble the Hunt's terms in the sense that it seems to reflect underlying Markov processes. We plan to examine if there are ways to give all these pieces of information to GP in order to obtain more accurate results with less effort.

Fig. 6 The tree to estimate throughput for four stations



Acknowledgement This research has been co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: Thales (ASPASIA). Investing in knowledge society through the European Social Fund.

Appendix 1: Exact Formula for Three-Station Line Without Buffers

A K-station serial procuction line is presented in Fig. 7. A system with three stages (see Fig. 8) under the assumptions in Hunt (1956) has eight states in the corresponding state transition rate diagram (see Fig. 9). A state is represented by a node which contains a vector of three numbers. This vector represents the state of the entire line, i.e, first number represents the state of the first station, second number represents the state of the second station and so on. Every number in the vector is interpreted as 0 for idle station, 1 for busy station and 2 for blocked station. The interconnections are represented by the mean service rate μ_i of each station (Muth 1984). According to this analysis the numbers of the states of a K-station line n_K are shown in Table 2 (Muth 1984).

For the case of $K = 3$ the throughput is expressed in Eq. (7), (Hunt 1956). In Hunt's work the term maximum possible utilization ρ_{max} was used.

$$\rho_{\max} = \frac{N}{D} \tag{7}$$

Where

$$\begin{aligned} N = {} & \mu_2\mu_3(\mu_2+\mu_3)\left(\mu_1{}^4+2\mu_1{}^3\mu_2+3\mu_1{}^3\mu_3+\mu_1{}^2\mu_2{}^2\right. \\ & +4\mu_1{}^2\mu_2\mu_3+3\mu_1{}^2\mu_3{}^2+\mu_1\mu_2{}^2\mu_3+4\mu_1\mu_2\mu_3{}^2 \\ & \left.+\mu_1\mu_3{}^3+\mu_2{}^2\mu_3{}^2+\mu_3{}^3\mu_2\right) \end{aligned}$$

$$\begin{aligned} D = {} & \mu_1{}^5(\mu_2{}^2+\mu_2\mu_3+\mu_3{}^2) \\ & +\mu_1{}^4(2\mu_2{}^3+5\mu_2{}^2\mu_3+5\mu_2\mu_3{}^2+3\mu_3{}^3) \\ & +\mu_1{}^3(\mu_3{}^4+5\mu_2{}^3\mu_3+8\mu_2{}^2\mu_3{}^2+7\mu_2\mu_3{}^3+3\mu_3{}^4) \\ & +\mu_1{}^2\left(\mu_2{}^4\mu_3+5\mu_2{}^3\mu_3{}^2+8\mu_2{}^2\mu_3{}^3+5\mu_2\mu_3{}^4+\mu_3{}^5\right) \\ & +\mu_1\left(\mu_2{}^4\mu_3{}^2+5\mu_2{}^3\mu_3{}^3+5\mu_2{}^2\mu_3{}^4+\mu_2\mu_3{}^5\right) \\ & +\left(\mu_2{}^4\mu_3{}^3+2\mu_2{}^3\mu_3{}^4+\mu_2{}^2\mu_3{}^5\right) \end{aligned}$$

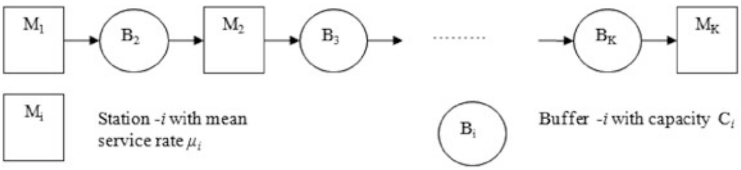


Fig. 7 A K-station production line with $K-1$ intermediate buffers

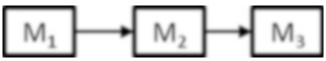


Fig. 8 A three station production line with no intermediate buffers

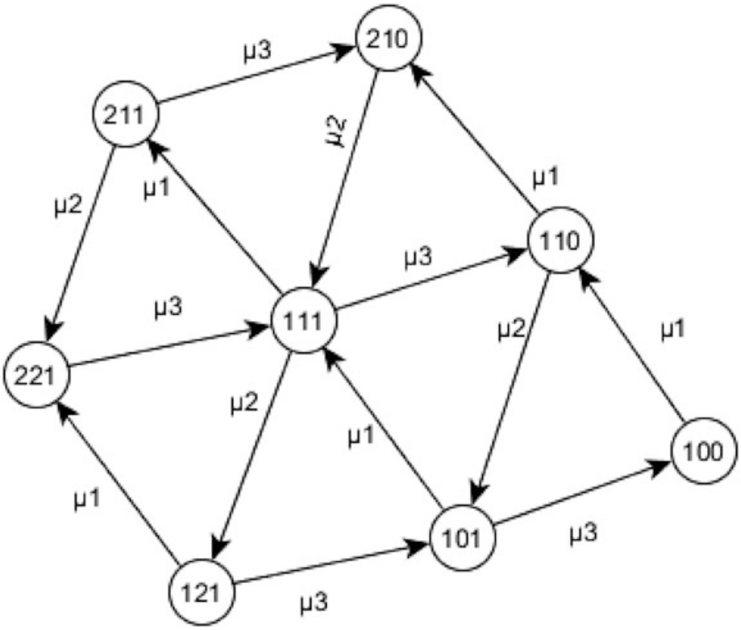


Fig. 9 State transition rate diagram for three station line with no intermediate buffers

Table 2 Number of states for a K-station line without buffers

Number of station K	Corresponding number of states n_K
2	3
3	8
4	21
5	55
6	144
7	377
8	987
9	2584
10	6765

Appendix 2: The DECO-2 Algorithm

The DECO-2 algorithm (Diamantidis et al. 2007) is capable of handling saturated lines (with over 1000 stations in series) with exponential service times, parallel identical machines at each station and finite intermediate buffers using a decomposition methodology and estimates the throughput of the specified production line. Figure 10 shows the decomposition scheme for a K-stations line. Details of the algorithm are presented in Papadopoulos et al. (2009). The steps of the algorithm are presented as follow:

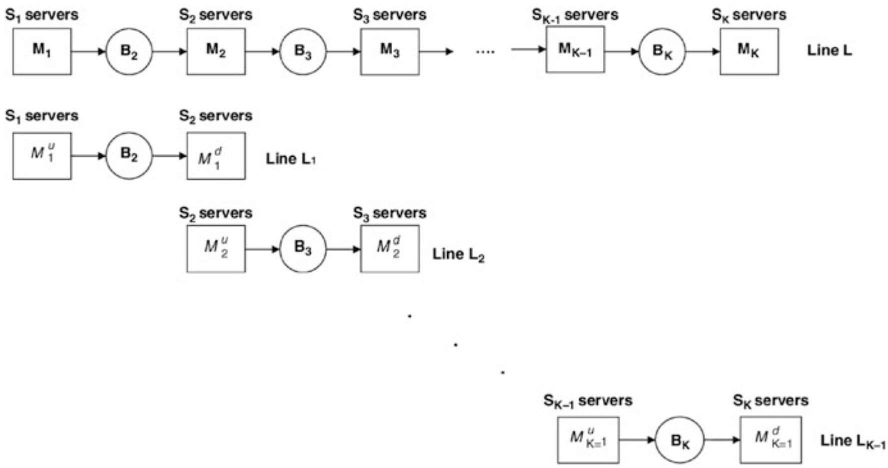


Fig. 10 Flow line with K parallel-machine work-stations, $K-1$ intermediate buffers (Line L) and decomposition scheme (Lines $L_1,\ldots,L_{K-1}$) (Papadopoulos et al. 2009). In the present work S = 1

Table 3 Glossary for decomposition approach

Symbol	Meaning
C	Total storage capacity of the system
L	Original production line that is decomposed in the decomposition approach L_i Sub-line i, $i = 1, 2, \ldots, K-1$ in the decomposition approach
M_i^u	The part of the original line, L, upstream buffer B_{i+1}. It is a pseudo work- station for $i = 2, \ldots, K-1$. For i = 1 it holds: $M_1^u = M_1$
M_{i-1}^d	The part of the original line, L, downstream buffer B_i. It is a pseudo work- station for $i = 2, \ldots, K-1$. Special case: It holds: $M_K^d = M_K$
μ_i	The mean service rate of station i, $i = 1, 2, \ldots, K$ in the original line, L
μ_i^u	The mean service (processing) rate of the upstream station of buffer $B_{i+1}, i = 1, \ldots, K-1$ in the decomposition approach
μ_i^d	The mean service (processing) rate of the downstream station of buffer $B_i, i = 2, \ldots, K$ in the decomposition approach
P	Transition probabilities
S	Number of server in each stage. In the present work S = 1
X_{DECO}	Throughput or mean production (output) rate of a K -station production line obtained from application of the decomposition method

{Step 1: Initialization}
for $i = 1$ **to** $K - 1$ **do**

$\mu_i^u = \mu_i$
$\mu_i^d = \mu_{i+1}$

ε = small positive number for terminating condition
end for

$\left\{ \textbf{Step 2: Calculate } \mu_i^u \textbf{ and } \mu_j^d \right\}$
for $i = 2$ **to** $K - 1$ **do**
Calculate μ_i^u using the following equation

$$\mu_i^u = \frac{1}{\frac{1}{\mu_i} + \frac{s_i}{X_{i-1}}} - \frac{1}{\mu_{i-1}^d}, \; i = 2, \ldots, K - 1$$

Evaluate the two-work-station, one buffer sub-line L_{i-1}, using the most recent values of μ_{i-1}^u and μ_{i-1}^d in the algorithm for generating the transition matrix
end for
for $i = 2$ **to** **K − 1** **do**
$j = K - i$
Calculate μ_j^d using the following equation

$$\mu_i^d = \frac{1}{\frac{1}{\mu_{i+1}} + \frac{s_{i+1}}{X_{i+1}}} - \frac{1}{\mu_{i+1}^u}, \; i = K - 2, \ldots, 1$$

Evaluate the two-work-station, one buffer sub-line L_{i+1}, using the most recent values of μ_{i+1}^u and μ_{i+1}^d in the algorithm for generating the transition matrix
end for
{**Step3: Terminating Conditions**}
if $\left| X_i^L - X_1^L \right| < \varepsilon, \; i = 2, \ldots, K - 1$ **then**
  GOTO *Step* 4
else
  GOTO *Step* 2
end if
{**Step4: Output Results**}
$X = X_i^L, i = 1, \ldots, K - 1$

The algorithm generates the transition probabilities in three stages (Diamantidis et al. 2007)

1. transition probabilities of the lower boundary states
2. transition probabilities of the internal states
3. transition probability of the upper boundary state

The steps of the algorithm are presented as follow:

```
{Lower boundary states}
P_{0,0} = 1 − S_1 μ_1
P_{0,1} = S_1 μ_1
for c = 2 to C do
 P_{0,c} = 0.0
end for
{Internal states}
for i = 1 to C − 1 do
for j = 0 to j = C do
if i > j and i − j = 1 and i < S_2 then
        P_{i,j} = i μ_2
end if
if i > j and i − j = 1 and i ≥ S_2 then
        P_{i,j} = S_2 μ_2
end if
if i = j and j < S_2 and i < S_2 + B + 1 then
        P_{i,j} = 1 − S_1 μ_1 − j μ_2
end if
if i = j and j ≥ S_2 and i < S_2 + B + 1 then
          P_{i,j} = 1 − S_1 μ_1 − S_2 μ_2
end if
if i = j and j ≥ S_2 and i ≥ S_2 + B + 1 then
          K = C − i
          P_{i,j} = 1 − K μ_1 − S_2 μ_2
end if
if j > i and j − i = 1 and i < S_2 + B + 1 then
        P_{i,j} = S_1 μ_1
endif
if j > i and j − i = 1 and i ≥ S_2 + B + 1 then
          m = C − i
          P_{i,j} = m μ_1
end if
if i > j and i − j > 1 then
          P_{i,j} = 0.0
end if
if j > i and j − i > 1 then
          P_{i,j} = 0.0
end if
end for
end for
```

{Upper boundary states}
$P_{C,C-1} = S_2\mu_2$
$P_{C,C} = 1 - S_2\mu_2$
for $c = 0$ **to** $C - 2$ **do**
$P_{C,c} = 0.0$
end for

References

Angeline PJ, Kinnear KE (1996) Advances in genetic programming, vol 2. MIT Press, Cambridge

Blumenfeld DE (1990) A simple formula for estimating throughput of serial production lines with variable processing times and limited buffer capacity. Int J Prod Res 28(6):1163–1182

Boulas K, Dounias G, Papadopoulos C, Tsakonas A (2015) Acquisition of accurate or approximate throughput formulas for serial production lines through genetic programming. Proceedings of 4th international symposium and 26th national conference on operational research, Hellenic Operational Research Society, pp 128–133

Dallery Y, Frein Y (1993) On decomposition methods for tandem queueing networks with blocking. Oper Res 41(2):386–399

Dallery Y, Gershwin SB (1992) Manufacturing flow line systems: a review of models and analytical results. Queueing Syst 12(1–2):3–94

Diamantidis AC, Papadopoulos CT, Heavey C (2007) Approximate analysis of serial flow lines with multiple parallel-machine stations. IIE Trans 39(4):361–375

Heavy C, Papadopoulos H, Browne J (1993) The throughput rate of multistation unreliable production lines. Eur J Oper Res 68(1):69–89

Hunt GC (1956) Sequential arrays of waiting lines. Oper Res 4(6):674–683

Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection, vol 1. MIT Press, Cambridge

Koza JR (1994) Genetic programming II- automatic discovery of reusable programs. MIT Press, Cambridge

Lai T (2003) Discovery of understandable math formulas using genetic programming. In: Genetic algorithms and genetic programming at Stanford. Stanford Bookstore, Stanford, pp 118–127

Langdon WB (2000) Quadratic bloat in genetic programming. In: Whitley DL (ed) GECCO. Morgan Kaufmann Publishers, San Francisco, pp 451–458

Li L, Qian Y, Du K, Yang Y (2015) Analysis of approximately balanced production lines. Int J Prod Res [online]: 1–18. doi: 10.1080/00207543.2015.1015750. Accessed 26 Apr 2015

Lim J-T, Meerkov S, Top F (1990) Homogeneous, asymptotically reliable serial production lines: theory and a case study. IEEE Trans Autom Control 35(5):524–534

Martin G (1993) Predictive formulae for un-paced line efficiency. Int J Prod Res 31(8):1981–1990

Muth EJ (1984) Stochastic processes and their network representations associated with a production line queuing model. Eur J Oper Res 15(1):63–83

Muth EJ (1987) An update on analytical models of serial transfer lines. Department of Industrial and Systems Engineering, University of Florida, Gainesville

Papadopoulos C, Tsakonas A, Dounias G (2002) Combined use of genetic programming and decomposition techniques for the induction of generalized approximate throughput formulas in short exponential production lines with buffers. Proceedings of the 30th international conference on computers and industrial engineering, Tinos Island, Greece, vol II, pp 695–700

Papadopoulos CT, O'Kelly ME, Vidalis MJ, Spinellis D (2009) Analysis and design of discrete part production lines, vol 31. Springer, New York

Poli R, Langdon WB (1998) On the search properties of different crossover operators in genetic programming. In: Koza JR et al (eds) Genetic programming 1998: proceedings of the 3rd annual conference, Universtity of Wisconsin, pp 293–301

Poli R, McPhee NF (2008) Covariant parsimony pressure in genetic programming. Technical report. Citeseer

Poli R, Langdon WB, McPhee NF (2008) A field guide to genetic programming (with contributions by Koza JR). In: A field guide to genetic programming. www.gp-field-guide.org.uk. Available via http://dces.essex.ac.uk/staff/rpoli/gp-field-guide/A_Field_Guide_to_Genetic_Programming.pdf. Accessed 1 Nov 2014

An Island Memetic Algorithm for Real World Vehicle Routing Problems

Ioannis Rogdakis, Magdalene Marinaki, Yannis Marinakis, and Athanasios Migdalas

Abstract In this paper, a new algorithm is presented which is applied to a real world Vehicle Routing Problem (VRP) of a provision company in the island of Crete in Greece. The company serves 116 customers located in Crete. This real world problem is solved effectively by a hybrid Island Memetic Algorithm (IMA) which employs Greedy Randomized Adaptive Search Procedure (GRASP) and Iterated Local Search (ILS). The proposed algorithm is also compared to five other approaches both on the real world problem and on classic benchmark instances from the literature. Methods such as GRASP, local search and Iterated Local Search (ILS) are employed as subroutines with certain probabilities in the algorithms. Furthermore, it is also demonstrated how premature convergence can be prevented by adopting specific strategy. Computational results show the superiority of the proposed hybrid Island Memetic Algorithm.

Keywords Vehicle Routing Problem • Island Memetic Algorithms • Greedy Randomized Adaptive Search Procedure • Iterated Local Search

1 Introduction

The difficulty in preserving the nutritional characteristics of fresh provisions during transportation makes the optimization of such processes entirely necessary. An effective solution of this problem is achieved using a new Island Memetic Algorithm (IMA) hybridized with a Greedy Randomized Adaptive Search Procedure

I. Rogdakis • M. Marinaki • Y. Marinakis
School of Production Engineering and Management, Technical University of Crete, Chania, Greece
e-mail: irogdakis@yahoo.com; magda@dssl.tuc.gr; marinakis@ergasya.tuc.gr

A. Migdalas (✉)
Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Department of Business Administration, Industrial Logistics, Technology and Social Sciences, Luleå University of Technology, Luleå, Sweden
e-mail: samig@auth.gr; athmig@ltu.se

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_10

(GRASP) (Feo and Resende 1995; Marinakis et al. 2005a, b, 2009; Resende and Ribeiro 2003) and an Iterated Local Search (ILS) (Besten et al. 2001; Lourenco et al. 2002; Martin et al. 1991; Talbi 2009) algorithm. Memetic algorithms resemble to genetic ones (to the fact that they use processes such as crossover, mutation, evolution of populations, etc) and they differ from them as they use a local search phase. Thus, memetic algorithms are more capable in searching the solution space as they possess both exploration abilities by utilizing evolutionary properties and exploitation abilities by employing a local search phase.

In this paper, it is demonstrated that Island Memetic Algorithms can be effectively applied to a real life Vehicle Routing Problem (VRP). In an Island Memetic Algorithm there are seven main steps: the initialization of each island's population, the selection of the parents, the crossover operator, the local search phase, the migration phase, the mutation operator and the replacement of each generation. The parents try to help their offsprings in order to learn and evolve and, thus, to become more competitive and increase their possibilities to survive and to become themselves parents for the next generations. One of the goals of this paper is to introduce an effective way for the evolution of each individual (solution) with the use of the Island Memetic Algorithm combined with GRASP and ILS. In each generation produced, the two best solutions are combined using a Path-Relinking strategy (Glover et al. 2003). The best solutions of the whole island survive and move to the next generation of the memetic algorithm, i.e. to the step of parent selection. The reason that a memetic algorithm (Moscato and Cotta 2003) is used instead of a classic genetic one is due to the difficulties encountered by the latter in effectively exploring the solution space. A combination of a global search optimization method with a local search optimization method usually improves the performance of the former. The reason that an island approach is used instead of a classic memetic algorithm is that it is desired to explore thoroughly any region of the solution space. By using a number of different islands, the algorithm has different locally best solutions to exploit (the best solution of each island) in each iteration. The migration phase of the algorithm helps the information about good solutions found in each island to be diffused between the islands and, thus, improve the probability of the overall algorithm to find good solutions. In this paper, a combination of four local search methods (two-opt, three-opt, Swap, and Relocate) is used to improve each individual separately. GRASP is used to speed up the algorithm by giving quickly new solutions of good quality. GRASP, local search and Iterated Local Search are used with certain probabilities for each step of the algorithm. This way, premature convergence is prevented.

The rest of the paper is organized as follows: In the next section a description of the Vehicle Routing Problem is presented. The proposed algorithm, the Island Memetic Algorithm with GRASP and ILS, is presented and analyzed thoroughly, in the third section. In the fourth section, the computational results of benchmark problems of the literature are presented and analyzed, and in the fifth section, the real world problem of the provision company is introduced and its results by using the proposed algorithm are presented. Finally, the conclusions are given in the last section.

2 The Vehicle Routing Problem

The **Capacitated Vehicle Routing Problem (CVRP)** is the problem in which vehicles based at a depot are required to visit a number of customers in order to fulfill known customer demands. Let $G = (V, E)$ be a graph where $V = \{i_1, i_2, \cdots i_n\}$ is the node set (i_1 refers to the depot and the customers are indexed $i_2, \cdots, i_n$) and $E = \{(i_l, i_m) : i_l, i_m \in V\}$ is the edge set. The customers must be assigned to the k vehicles (each one of them to exactly one vehicle) taking into account that the total size of deliveries to customers assigned to each vehicle must not exceed the vehicle's capacity (Q_k). When vehicles are homogeneous, then, they are of the same capacity which is denoted by Q. A demand q_l and a service time st_l are associated with each customer i_l. The demand q_1 and the service time st_1 which are referred to the demand and service time of the depot are set equal to zero. The travel cost and the travel time between customers i_l and i_m is c_{lm} and tt^k_{lm}, respectively, and T_k is the maximum time allowed for a route of vehicle k. The problem is to construct a low cost, feasible set of routes—one for each vehicle. The variable x^k_{lm} is equal to 1 if the arc (i_l, i_m) is traversed by vehicle k and 0 otherwise. The vehicle must start and finish its tour at the depot. In the following we present the mathematical formulation of the VRP (Bodin et al. 1983; Marinakis and Marinaki 2010):

$$J = \min \sum_{l=1}^{n} \sum_{m=1}^{n} \sum_{k=1}^{K} c_{lm} x^k_{lm} \tag{1}$$

s. t.

$$\sum_{i_l=1}^{n} \sum_{k=1}^{K} x^k_{lm} = 1, i_m = 2, \cdots, n \tag{2}$$

$$\sum_{i_m=1}^{n} \sum_{k=1}^{K} x^k_{lm} = 1,\ i_l = 2, \cdots, n \tag{3}$$

$$\sum_{i_l=1}^{n} x^k_{lf} - \sum_{i_m=1}^{n} x^k_{fm} = 0, \quad k = 1, \cdots, K \qquad i_f = 1, \cdots, n \tag{4}$$

$$\sum_{i_l=1}^{n} q_l \sum_{i_m=1}^{n} x^k_{lm} \leq Q_k, k = 1, \cdots, K \tag{5}$$

$$\sum_{i_l=1}^{n} st^k_l \sum_{i_m=1}^{n} x^k_{lm} + \sum_{i_l=1}^{n} \sum_{i_m=1}^{n} tt^k_{lm} x^k_{lm} \leq T_k, k = 1, \cdots, K \tag{6}$$

$$\sum_{i_m=2}^{n} x_{1m}^{k} \leq 1, k = 1, \cdots, K \tag{7}$$

$$\sum_{i_l=2}^{n} x_{l1}^{k} \leq 1, k = 1, \cdots, K \tag{8}$$

$$x_{lm}^{k} = 0 \text{ or } 1, \text{for all } i_l, i_m, k \tag{9}$$

Objective function (1) states that the total distance is to be minimized. Eqs. (2) and (3) ensure that each demand node is served by exactly one vehicle. Route continuity is represented by Eq. (4). Equation (5) are the vehicle capacity constraints and Eq. (6) are the total elapsed route time constraints. Equations (7) and (8) guarantee that vehicle availability is not exceeded.

The vehicle routing problem was first introduced by Dantzig and Ramser (1959). As it is an NP-hard problem, a large number of approximation techniques have been proposed. These techniques are classified into two major categories: Classical heuristics that were developed mostly between 1960 and 1990 (Altinkemer and Gavish 1991; Bodin and Golden 1981; Bodin et al. 1983; Christofides et al. 1979; Clarke and Wright 1964; Desrochers and Verhoog 1989; Fisher and Jaikumar 1981; Foster and Ryan 1976; Gillett and Miller 1974; Lin 1965; Lin and Kernighan 1973; Mole and Jameson 1976; Wark and Holt 1994) and metaheuristics that were developed in the last 25 years. Metaheuristic algorithms are classified in categories based on the used strategy, i.e. Tabu Search strategy (Barbarosoglu and Ozgur 1999; Cordeau et al. 2002; Gendreau et al. 1994; Osman 1993; Rego 1998, 2001; Taillard 1993; Toth and Vigo 2003; Xu and Kelly 1996), Adaptive Memory (Rochat and Taillard 1995; Tarantilis 2005; Tarantilis and Kiranoudis 2002), Simulated annealing (Osman 1993), record to record travel (Golden et al. 1998; Li et al. 2005), greedy randomized adaptive search procedure (Prins 2008), threshold accepting algorithms (Tarantilis et al. 2002a, b). In the last 10 years a number of nature inspired metaheuristic algorithms have been applied for the solution of the Vehicle Routing Problem. The most commonly used nature inspired methods for the solution of this problem are genetic algorithms (Baker and Ayechew 2003; Berger and Barkaoui 2003; Marinakis et al. 2007; Prins 2004), ant colony optimization (Bullnheimer et al. 1999; Reimann et al. 2002, 2004), particle swarm optimization (Marinakis and Marinaki 2010; Marinakis et al. 2013, 2010), honey bees mating optimization (Marinakis et al. 2008) and other evolutionary techniques (Cordeau et al. 2005; Mester and Braysy 2005, 2007). The reader can find more detailed descriptions of these algorithms in the survey papers (Bodin and Golden 1981; Bodin et al. 1983; Fisher 1995; Gendreau et al. 1997, 2002; Laporte et al. 2000; Laporte and Semet 2002; Marinakis and Migdalas 2002; Tarantilis 2005) and in the books (Golden and Assad 1988; Golden et al. 2008; Pereira and Tavares 2008; Toth and Vigo 2002, 2014).

3 Island Memetic Algorithm Using GRASP and ILS

3.1 General Description of Island Memetic Algorithm with the Use of GRASP and ILS (IMGL)

Memetic Algorithms are essentially Genetic Algorithms equipped with a local search technique. While the positions of the customers can be represented by nodes with specific coordinates, in the algorithm, each customer node corresponds to a serial number. That is, the solutions are coded in path representation, i.e., in the form of vectors which indicate the sequence in which the customers are served. Each customer is represented by a number in the solution's main vector. A secondary vector indicates the vehicle by which each customer is being served.

Initially, in the Island Memetic Algorithm (IMGL), a population (*pop*) is created. Half of the initial solutions are created randomly while the rest are created by using the Greedy Randomized Adaptive Search Procedure (GRASP) (see Sect. 3.3) for each island. Then, local search is applied on the initial solutions in order to improve them. Local search at this point includes the methods of 2-opt, 3-opt, Swap and Relocate (see Sect. 3.4). Then, contrary to the classic memetic algorithm (MA), the initial population in the IMGL algorithm is divided into a number of islands (isl_{max}), depending on the selection of the user (the number of islands is one of the parameters of the algorithm). Each island represents a different population, where the MA's operators are applied independently from the other islands. The basic difference of the IMGL algorithm from the classic MA is the **migration policy** of the population. The migration policy is used in order to exchange information between different islands. There is a number of different ways to realize the migration strategy (Engelbrecht 2007). In the algorithm of the present paper, the user gives the percentage of the population that will migrate to another island and also the number of times that a migration policy will be applied. A ring topology is used for the migration of the population, which means that the migrants migrate to a neighboring island. The migration occurs a number of times depending on the selection of the user. At the end of each step of the main algorithm, the r_1 best solutions of the i_{th} island replace the r_1 worst solutions of the $(i+1)_{th}$ island. This circular replacement constitutes the migration and helps the diversification of the solutions in each island. There is a parallel evolution of the solutions in every island converging to different local optima. This way, extended diversification and better quality of solutions may be achieved.

Afterwards, the $\frac{pop}{2}$ best initial solutions are used in the process of crossover (see Sect. 3.2). In the i_{th} step of this process, the i_{th} solution is combined with another randomly chosen solution. Almost always the solution that results from the procedure of crossover is not feasible. This is due to the path representation of each solution. Thus, in order to restore feasibility of the solution, the mutation process is applied to it. Subsequently local search is applied (described in Sect. 3.4) in order to improve the quality of the solution. If the new solution is better than the i_{th} one, it is stored in the $\left(\frac{pop}{2}+i\right)_{th}$ place of the population. Otherwise, the new solution is deleted and a random one is generated on which iterated local search is applied. The final result is then stored in the $\left(\frac{pop}{2}+i\right)_{th}$ place of the population. This procedure is repeated for all the $2\,\frac{pop}{2}$ solutions. Finally, $\frac{pop}{2}$ new solutions have been created, so the total number of solutions is *pop*. Subsequently Path-Relinking is applied (see Sect. 3.5) to the two best solutions of each island. If these solutions are similar, then the best and some other, randomly selected, solution are used instead to initiate the method. If the resulting solutions differ from the initials, they are stored in the $\left(\frac{pop}{2}-1\right)_{th}$ and the $\left(\frac{pop}{2}\right)_{th}$ position respectively of the island's solution set, where there are certainly solutions of worse quality. Local search is then applied on these. If the resulting two solutions differ from the initial ones, they are stored in $(pop-2)_{th}$ and in the $(pop-1)_{th}$ position respectively of the island's solution set. Finally, the *pop* solutions are sorted from the best to worst.

As the repeated application of the described procedure progresses, convergence may start to appear. That is, solutions of an island may start resemble or even be identical to other solutions of the same island. But the similar or identical solutions are sorted close together because they have the similar or identical transfer costs. So, if a solution resembles another up to $(70+a)\%$ (where a is a continuous random variable from uniform distribution on the interval (0,4]), it is deleted. If the number of deleted solutions exceeds $\frac{pop}{2}$, each solution that is being removed (after the $\left(\frac{pop}{2}\right)_{th}$) is replaced by a random solution on which a local search is applied. The main algorithm is terminated when a maximum number of iterations ($iter_{max}$) has been reached or all the islands have converged. Then, the best solution over all islands is selected and an extra local search is applied. For all the aforementioned methods the maximum number of iterations is $lsiter_{max}$. A pseudocode of the overall method is presented next.

Initialization

For $i=1$ to isl_{max}
    For $j=1$ to pop
        Creation of a Random Variable (rv) from uniform distribution in the interval (0,1)
        If ($rv > 0.5$) **then**
            Creation of a Random Solution (probability = 50 %)
        else
            Creation of a Solution with GRASP method (probability = 50 %)
        Endif
    Endfor
Endfor

Main Algorithm

Do while (Convergence OR Completion of $iter_{max}$)
    For $i=1$ to isl_{max}
        For $l=1$ to $iter_{max}$
            For $j=1$ to $\frac{pop}{2}$
                Call Crossover Operator
                Call Mutation Operator
                Call Iterated Local Search

Endfor
Deletion of same or similar solutions
Selection of $\frac{pop}{2}$ best solutions (for the next generation)
Endfor
Application of circular migration
Endfor
Enddo
Selection of best solution
Extra application of Local Search on best solution

3.2 Crossover: Mutation Operators

The genetic procedure of the evolutionary and consequently of the memetic algorithms is the crossover operator: Two solutions (the parents) are combined in order to produce a third solution (an offspring) which includes attributes of both parents. In a one-point crossover, two such offsprings can be produced: The first section of the vector representing the first offspring is copied in from the corresponding section of the first parent while the second section of the vector is copied in from the second parent. If a second offspring is produced, the parent contributions to it are in reversed order. In order for the most positive solution attributes to be passed to the next generations, the parents are splitted in places which correspond to expensive transfer costs. An effective way to mutate an offspring vector is to remove a node corresponding to expensive transfer costs and place in its position another node corresponding to lower transfer costs from its predecessor and to its successor on the route sequence.

In this paper, a new stochastic crossover operator is used which combines two different ways of selecting the splitting points of the two parents. With a probability $pop \times 2\%$, the two sequence vectors are divided in places that correspond to the $split \pm y$ most expensive transfer costs, where y is a random integer variable from a uniform distribution on the interval $[-3,3]$. Otherwise, the splitting points are chosen randomly. The reason for using this combine operator is because it is desired to delete the most unpromising edges of the parents and replace them by better edges in the offspring. The randomization in the splitting procedure provides the crossover operator with better exploration abilities.

If the offspring solution is not feasible, it is converted to a feasible one by applying a mutation operator. Each customer must be served in exactly one route. Therefore if a customer node appears twice in the infeasible solution, one of its occurrence must be replaced by a customer node that does not appear in the non-feasible solution. In this mutation process, all the nodes that do not appear in the infeasible solution are initially detected and they are placed in a set S. Furthermore, the positions of the nodes that appear multiple times in the infeasible solution vector are also detected and registered in a set Q. In the first step of the mutation process, a node is selected from the set S in order to be placed in the first position (i) of the set Q. With probability $pop \times 4\%$, the node that is selected is one of them that correspond to the $2 \times split \pm 5$ lowest transfer costs from the predecessor of i and to the successor of i on the route. Otherwise, the node is chosen randomly. Then, the selected node is placed in the i_{th} position of the solution vector and it is removed from the set S. Also, the position i is removed from the set Q. This process is continued until both sets, S and Q, become void and the solution becomes feasible.

3.3 GRASP Algorithm

Greedy Randomized Adaptive Search Procedure (GRASP) (Feo and Resende 1995; Resende and Ribeiro 2003) is an iterative two phase search method which has been used extensively in combinatorial optimization. For each iteration two phases exist, the **construction phase** and the **local search phase**. In the construction phase, a randomized greedy function is used to create step by step an initial solution. This randomized technique always produces a feasible solution to which a local search algorithm is subsequently applied. The final solution produced by the algorithm is simply the best over all such iterations.

In the construction phase, one element at a time is added iteratively to the incumbent partial solution until a complete, feasible solution is produced. The selection of the next element to be added is not completely greedy, e.g., is neither the nearest neighboring nor the nearest insertion element. It is instead determined by ordering all the possible candidates in a list, called Restricted Candidate List and denoted by RCL, based on a greedy function. The method is probabilistic since one of the best candidates in this list is selected randomly in order to be added the partial solution. one. The adaptiveness of the method is realized by updating the list after each such selection. The algorithm is of greedy-type as is based on greediness for the list construction and since it only performs a simple pass in order to create the initial solution.

Marinakis et al. (2005a) proposed an algorithm for the solution of the TSP that adds new features to the original GRASP algorithm and has been proved very efficient for the solution of that problem.

In this paper, in each step of the construction phase, a new customer node is added to the partial solution vector. The node added during a step is with high probability among those corresponding to the lower transfer costs from the node added in the previous step. This is achieved by selecting in each step a node that belongs to the $RCL \pm x$ nearest nodes to the node added in the previous step, where x is a random integer variable from the uniform distribution on the interval $[-3,3]$. In the last 15 steps, nodes are placed randomly in the solution vector. The new solution is thus created quickly and is generally of good quality.

3.4 Local Search

Local search is executed until $\frac{lsiter_{max}}{2}$ iterations are completed or the solution does not improve for $\frac{lsiter_{max}}{20} \pm 1$ iterations. Only in the cases of improving the initial and final solutions is the maximum number of iterations equal to $lsiter_{max}$. The initial solution vector is split with a certain probability at positions that correspond to expensive transfer costs. With the same probability a solution is split at some places randomly while other splitting points correspond to worst transfer costs. The number of worst transfer costs is not kept fixed. It fluctuates around a constant *split*. The variation from the constant *split* is a random number with uniform distribution on the interval $[-v, v]$. Thus, the number of the most expensive costs that are detected each time is $split \pm v$. In order to increase the search breath, the solution is split most of the time randomly. The local search algorithm to be applied is selected among 2-opt, 3-opt, Relocate and Swap randomly. How these algorithms have been applied is described next.

In the 2-opt method the solution vector is split in two places. Then, the node sequence order in the middle section of the path representation of the tour is reversed. The solution is split at one of the $split \pm 2$ places that correspond to the most expensive transfer costs with probability 3 % while the second point is chosen with a probability 3 % in a similar fashion or randomly. Both splitting points are chosen randomly with probability 94 %.

In the 3-opt method, the solution vector is split in three places. Then, the node sequence order is reversed in each of the two middle sections. With probability 6 % the first splitting point belongs to the $split \pm 2$ most expensive transfer costs. The other two splitting points can be chosen in a similar fashion however with probability 3 % each. Otherwise the splitting points are chosen randomly.

In the Swap method, two nodes that correspond to the $split \pm 2$ most expensive transfer costs are selected with probability 3 %. Then, their positions are swapped. Also with probability 3 % one node may be selected among those corresponding to expensive costs while the second node is selected randomly. Otherwise both nodes are chosen randomly.

In the Relocate method, a subroute from some node i to another node j is selected with 3 % probability among the $split \pm 2$ most expensive routes connecting the two nodes. Also with probability 6 % a node that corresponds to the $split \pm 2$ highest transfer costs is selected. Then, this node is moved between the nodes i and j. Otherwise, there is 94 % probability for which both the node and the route are chosen randomly.

3.5 Path Relinking

This approach generates new solutions by exploring trajectories that connect high-quality solutions—by starting from one of these solutions, called the *starting* *solution* and generating a path in the neighborhood space that leads towards the other solution, called the *target solution* (Glover et al. 2003). Here the Path-Relinking method is applied as follows: In the i_{th} step of the method the i_{th} element of the starting solution is replaced by the i_{th} element of the target solution. Supposing that this element preexisted in the j_{th} position of the starting solution, swap must be applied on the i_{th} and j_{th} nodes of the starting solution in order to retain feasibility. After $n-1$ such steps (where n is the number of customers) the starting solution is transformed to the target one. In each step of this process a new solution is created and its cost is computed. At the end of the procedure, the two best solutions over all steps are stored. If two high quality solutions are used in this method, it is probable that the results of the process will be solutions of very good quality. Consequently, the two best solutions of each generation are used to initialize the process. However, in case these solutions are similar, then the best one of the two and a second randomly chosen solution are used to initialize the process.

3.6 Iterated Local Search

Local search is one of the most important components of the proposed algorithm. However, protracted use of it will almost surely result in the algorithm being trapped in the neighborhood of a local optimum. This drawback is counterpoised by the method of iterated local search (Besten et al. 2001; Lourenco et al. 2002; Martin et al. 1991; Talbi 2009). The main objective of this method, as implemented in the present study, is to expand the solution search and cover the attraction regions of as many as possible local optima in order to increase the probability of obtaining a better solution by combining the different local optima. The iterated local search obtains solutions by applying a sequence of perturbation methods. Firstly, a local optimum is obtained by the 2-opt method. Then diversification is applied to it. It is possible that the new solution so obtained is of moderate quality but diversified enough to be redeemed from the attraction region of the local optimum. On this solution local search is applied in the form of the Relocate method. The final solution is compared with the initial one and if it is better, then, it replaces it. The maximum diversification percentage is 22 %. Every time the perturbation method is used, the solution will be diversified per $a\,\%$, where a is a continuous random variable from uniform distribution on the interval $(0, 22]$. Thus, $n * a\,\%$ nodes will change positions in the solution vector, where n is the total number of customers. There is 2 % probability that the nodes will be chosen to correspond to the most expensive transfer costs. Otherwise, the nodes are chosen randomly.

3.7 Deletion of Similar Solutions

In the main algorithm of this paper the deletion of similar solutions is used as a way of avoiding premature convergence. In each memetic algorithm, as the number of generations grows, the probability of convergence increases. Indication of convergence is the appearance of similar or even identical solutions withing individual islands. When the solutions are sorted by their costs, the identical or similar solutions occupy positions close to each other in the sorted set since they have the identical or similar transfer costs. Sorting in ascending cost order is assumed. A solution is deleted completely when it resembles its predecessor in the sorted list by $(70 + a)\%$, where a is a continuous random variable uniformly distributed on the interval (0,4]. After the pop_{th} removed solution from each generation, each deleted solution is replaced by a random one which has been improved by the methods of local search as described above.

4 Computational Results

In order to assess the effectiveness of the algorithm, its performance was tested on the 14 classic benchmark instances proposed by Christofides (Christofides et al. 1979). The instances of this set consist of 51 up to 200 nodes including the depot. They are all capacitated while the instances numbered 6–10, 13 and 14 are also subjected to maximum route length restrictions (m.t.l.) and non zero service times (s.t.). For the first ten instances the nodes are randomly located over a square, while for the remaining ones the nodes are distributed in clusters and the depot is not centered. The quality of the solution produced by the proposed IMGL algorithm is computed in terms of the relative deviation from the best known solution for the corresponding instance, that is $\omega = \frac{(c_{IMGL} - c_{BKS})}{c_{BKS}}\%$, where c_{IMGL} denotes the cost of the solution found by IMGL and c_{BKS} is the cost of the best known solution. The algorithm parameters were selected after a thorough testing process and the best values selected are shown in Table 1. Once the parameter selection was complete, ten different runs of the IMGL algorithm were performed for each instance.

The most important characteristics (number of nodes (Nod.), Capacity of Vehicles (Cap), maximum route length restrictions (m.t.l.) and service times (s.t.)) of each data set are presented in Table 2 under columns 1–4. In the last three columns of the table, the results of the IMGL algorithm (column 5), the best known solutions (BKS—column 6) from the literature, and the quality measurements of the IMGL solutions (ω—column 7) are given. It can be seen from Table 2, that the IMGL algorithm has reached the best known solutions for four instances while for the remaining instances the quality of the solutions is between 0.06 % and 4.82 %, averaging over all 14 instances to 1.29 %.

In order to evaluate the impact on the final result of the incorporation of the GRASP and Iterated Local Search in the hybridized Memetic Algorithm, five more

Table 1 Best parameter values

Parameter	Selected value
Maximum iterations' number ($iter_{max}$)	1000
Maximum number of local search iterations ($lsiter_{max}$)	100
Number of islands (isl_{max})	5
Islands' population (pop)	40
Migrated members of the population (r_1)	10
Restricted Candidate List (RCL)	12
split	7

Table 2 Computational results for the 14 Christofides benchmark instances

Nod.	Cap.	m.t.l.	s.t.	IMGL	BKS	ω
51	160	∞	0	524.61	524.61 (Rochat and Taillard 1995)	0.00
76	140	∞	0	835.77	835.26 (Rochat and Taillard 1995)	0.06
101	200	∞	0	832.27	826.14 (Rochat and Taillard 1995)	0.74
151	200	∞	0	1057.24	1028.42 (Rochat and Taillard 1995)	2.80
200	200	∞	0	1353.51	1291.29 (Rochat and Taillard 1995)	4.82
51	160	200	10	555.43	555.43 (Rochat and Taillard 1995)	0.00
76	140	160	10	920.17	909.68 (Rochat and Taillard 1995)	1.15
101	200	230	10	871.71	865.94 (Rochat and Taillard 1995)	0.67
151	200	200	10	1186.36	1162.55 (Rochat and Taillard 1995)	2.05
200	200	200	10	1452.39	1395.85 (Rochat and Taillard 1995)	4.05
121	200	∞	0	1042.11	1042.11 (Rochat and Taillard 1995)	0.00
101	200	∞	0	819.56	819.56 (Rochat and Taillard 1995)	0.00
121	200	720	50	1563.80	1541.14 (Rochat and Taillard 1995)	1.47
101	200	1040	90	867.17	866.37 (Rochat and Taillard 1995)	0.09

algorithms, mainly variants of the basic approach were implemented and compared to IMGL. The purpose is to investigate the necessity of each of the phases of the proposed algorithm. Two of the additional algorithms, the Greedy Randomized Adaptive Search Procedure (GRASP) (columns 9 and 10 in Table 3) and the Iterated Local Search (ILS) (columns 7 and 8 in Table 3), are not of evolutionary nature. The remaining three are evolutionary algorithms; a Memetic Algorithm incorporating GRASP (MG) (column 5 and 6 in Table 3), a Memetic Algorithm incorporating ILS (MGL) in addition to GRASP (column 3 and 4 in Table 3), and an Island Memetic Algorithm incorporating only GRASP (IMG) (columns 1 and 2 in Table 3). The main difference between the IMG and the IMGL is the incorporation of the Iterated Local Search (ILS) in the latter. The main difference between the MGL and IMGL is the utilization of the island concept in the latter. Finally, the most significant differences between the MG and the IMGL is that the MG does not utilize either the ILS or the islands concept. For all algorithms the parameter values were chosen so that the equal number of function evaluations were made. Any additional required parameter values were set equal to those used in the IMGL. The

Table 3 Comparison of the proposed algorithm (IMGL) with IMG, MGL, MG, ILS and GRASP for the 14 Christofides benchmark instances

IMG	ω	MGL	ω	MG	ω	ILS	ω	GRASP	ω	IMGL	ω
524.61	0.00	524.61	0.00	524.61	0.00	524.61	0.00	524.61	0.00	524.61	0.00
838.93	0.44	839.35	0.49	840.91	0.68	864.86	3.54	857.00	2.60	835.77	0.06
833.21	0.86	843.23	2.07	849.02	2.77	855.69	3.58	852.72	3.22	832.27	0.74
1060.77	3.15	1071.62	4.20	1072.89	4.32	1082.04	5.21	1078.75	4.89	1057.24	2.80
1355.69	4.96	1356.63	5.06	1359.95	5.32	1367.99	5.94	1362.88	5.54	1353.51	4.82
555.43	0.00	555.43	0.00	555.43	0.00	555.43	0.00	555.43	0.00	555.43	0.00
921.09	1.25	926.92	1.90	929.06	2.13	930.41	2.28	927.81	1.99	920.17	1.15
875.38	1.09	886.41	2.36	892.77	3.10	902.93	4.27	904.02	4.40	871.71	0.67
1213.27	4.36	1220.92	5.02	1235.38	6.26	1248.83	7.42	1240.45	6.70	1186.36	2.05
1477.50	5.85	1475.89	5.73	1492.16	6.90	1522.74	9.09	1497.53	7.28	1452.39	4.05
1048.16	0.58	1050.50	0.81	1048.64	0.63	1066.56	2.35	1056.37	1.36	1042.11	0.00
819.56	0.00	819.56	0.00	819.56	0.00	819.56	0.00	819.56	0.00	819.56	0.00
1582.38	2.68	1572.00	2.00	1591.58	3.27	1650.04	7.07	1590.68	3.21	1563.80	1.47
868.42	0.24	869.02	0.31	869.02	0.31	880.61	1.64	872.80	0.74	867.17	0.09

same implementation of the 2-opt, 3-opt, Swap, Relocate, and Path-relinking methods were incorporated in all algorithms.

In Table 3, the cost and the quality of the solutions for each implemented algorithm are given. It is clear that the IMGL algorithm produces improved results. More specifically, the improvement obtained by IMGL over the ILS algorithm ranges between 0.00 % and 5.37 %. The improvement obtained by the IMGL algorithm over the GRASP algorithm is between 0.00 % and 4.65 %. The improvement obtained by the IMGL over the MG algorithm is in the range 0.00–4.21 % with an average improvement of 1.27 %. The improvement in the quality of the results obtained by IMGL over the MGL algorithm is between 0.00 % and 2.97 % with an average improvement of 0.86 %. Finally, the improvement obtained by IMGL over the IMG algorithm ranges between 0.00 % and 2.31 % with an average improvement of 0.54 %. It is therefore clear that the addition of each characteristic to the IMGL algorithm improves the final results. Also, it is important to note that the use of GRASP algorithm improves the results significantly.

5 The Provision Company

The IMGL algorithm is applied to solve the distribution problem of a provision company in Greece. This provision company produces and distributes frozen products in Crete with a privately owned fleet of vehicles. The main purpose of the problem under consideration is to select the order in which the company customers are served while satisfying their demand at a minimum total transportation cost.

The evolutionary approach provided by the IMGL algorithm is able to utilize all the information provided by the company about its customers and its fleet in order to attack such an important optimization problem in its supply chain. The information provided for the customers includes their demand and the distances between them based on the road network of the region. Some of the difficulties of the problem originate from the structure of the road network of Crete. There is only one motorway that connects the four prefectures of Crete and consequently an optimization plan is devised separately within each prefecture. Optimization is also performed with respect to the allocation of customers to vehicles during a week schedule in order to derive an improved clustering of customers that decreases the total cost by avoiding visits to some prefectures during specific days of the week. The total number of customers is 116. The company owns five trucks and has at its disposal 13 drivers for 6 days a week and 8 h per day. The trucks are inhomogeneous and their different capacities are shown in Table 4.

The average speed of a loaded truck is 50 km/h while its average speed when unloaded is 80 km/h. According to the current company policy the 116 clients are served by 11 routes and the total transportation cost is 3567 Euros per week. More specifically, the total truck load and the duration of each route are shown in the

Table 4 Truck capacities

R/N	Payload (kg)
1	1080
2	2100
3	4250
4	5482
5	6360

Table 5 Truck load and duration of each route

Route R/N	Load (kg)	Time (min)
1	4064.00	449.93
2	2300.60	472.27
3	774.06	435.98
4	422.97	475.07
5	416.42	462.90
6	782.96	474.60
7	1354.40	464.5
8	51.85	468.6
9	28.34	456.85
10	8.05	355.15
11	602.40	418.15

Table 6 Vehicle allocation to each week day

Day of the week	Vehicle R/N
Monday	1,2
Tuesday	1,2
Wednesday	1,2
Thursday	1,3
Friday	1,3
Saturday	1

Table 5 while the vehicle allocation for each day of the week is shown in the Table 6.

The result from application of IMGL to this problem are presented in Table 7 where it is compared to the results obtained by the other algorithmic variants. These results concern a week schedule. The company's empirically determined distribution plan results into a cost of 3567 Euros per week (the CS column of the Table). The results of the proposed algorithm and the other tested variants appear in the columns 3–8 under the acronyms previously introduced. As it can be seen, the IMGL algorithm improves the distribution plan significantly. Indeed, the improvement in the quality of the week schedule equals to 36.11 %, and the total cost is reduced to 2279 Euros. The results obtained by the simpler algorithmic variants are analogous to those obtained for the classic benchmark instances, that is, they are less favorable when compared to IMGL signifying the importance of each of the incorporated algorithmic characteristics.

Table 7 Cost and deviations for the company's problem

	CS	IMGL	MGL	IMG	MG	GRASP	ILS
Costs	3567	2279	2561	2523.1	2645.5	2702.5	2723
ω (%)	–	−36.11	−28.20	−29.27	−25.83	−24.24	−23.66

6 Conclusion

In this paper, an effective method is proposed in order to solve a real world Vehicle Routing problems, the Island Memetic Algorithm which employs Greedy Randomized Adaptive Search Procedure (GRASP) and Iterated Local Search (IMGL). This methodology combines the advantages of the Island Memetic Algorithms and the positive characteristics of GRASP and ILS in the local search phase. The improvement in the results for the real world problem were significant, as much as 36.11 %. It should be noted that the IMGL algorithm of this paper is designed for the solution of general real world problems. The algorithm is suitable not only for the classic Vehicle Routing Problem but for any variant of it. Furthermore, this method performs well on the 14 benchmark instances of Christofides as it was shown. These results would have been even better if the IMGL method was tailored to the problem instead of being implemented for general problems. The proposed method was compared to five other algorithmic variants and it was demonstrated that each characteristic incorporated into the IMGL algorithm plays an important role in obtaining the improved quality of the final results.

References

Altinkemer K, Gavish B (1991) Parallel savings based heuristics for the delivery problem. Oper Res 39(3):456–469

Baker BM, Ayechew MA (2003) A genetic algorithm for the vehicle routing problem. Comput Oper Res 30(5):787–800

Barbarosoglu G, Ozgur D (1999) A tabu search algorithm for the vehicle routing problem. Comput Oper Res 26:255–270

Berger J, Barkaoui M (2003) A hybrid genetic algorithm for the capacitated vehicle routing problem. Proceedings of the genetic and evolutionary computation conference, Chicago, IL, pp 646–656

Besten M, Stutzle T, Dorigo M (2001) A hybrid genetic algorithm for the capacitated vehicle routing problem. In: Boers EJW et al (eds) Design of iterated local search algorithms, EvoWorkshop LNCS 2037. Springer, Berlin, pp 441–451

Bodin L, Golden B (1981) Classification in vehicle routing and scheduling. Networks 11:97–108

Bodin L, Golden B, Assad A, Ball M (1983) The state of the art in the routing and scheduling of vehicles and crews. Comput Oper Res 10:63–212

Bullnheimer B, Hartl PF, Strauss C (1999) An improved ant system algorithm for the vehicle routing problem. Ann Oper Res 89:319–328

Christofides N, Mingozzi A, Toth P (1979) The vehicle routing problem. In: Christofides N, Mingozzi A, Toth P, Sandi C (eds) Combinatorial optimization. Wiley, Chichester

Clarke G, Wright J (1964) Scheduling of vehicles from a central depot to a number of delivery points. Oper Res 12:568–581

Cordeau JF, Gendreau M, Laporte G, Potvin JY, Semet F (2002) A guide to vehicle routing heuristics. J Oper Res Soc 53:512–522

Cordeau JF, Gendreau M, Hertz A, Laporte G, Sormany JS (2005) New heuristics for the vehicle routing problem. In: Langevine A, Riopel D (eds) Logistics systems: design and optimization. Wiley, Chichester, pp 279–298

Dantzig GB, Ramser JH (1959) The truck dispatching problem. Manage Sci 6(1):80–91

Desrochers M, Verhoog TW (1989) A matching based savings algorithm for the vehicle routing problem. Les Cahiers du GERAD G-89-04, Ecole des Hautes Etudes Commerciales de Montreal

Engelbrecht AP (2007) Computational intelligence: an introduction. Wiley, Chichester

Feo TA, Resende MGC (1995) Greedy randomized adaptive search procedure. J Glob Optim 6:109–133

Fisher ML (1995) Vehicle routing. In: Ball MO, Magnanti TL, Momma CL, Nemhauser GL (eds) Network routing, vol 8, Handbooks in operations research and management science. Elsevier, Amsterdam, pp 1–33

Fisher ML, Jaikumar R (1981) A generalized assignment heuristic for vehicle routing. Networks 11:109–124

Foster BA, Ryan DM (1976) An integer programming approach to the vehicle scheduling problem. Oper Res 27:367–384

Gendreau M, Hertz A, Laporte G (1994) A tabu search heuristic for the vehicle routing problem. Manage Sci 40:1276–1290

Gendreau M, Laporte G, Potvin J-Y (1997) Vehicle routing: modern heuristics. In: Aarts EHL, Lenstra JK (eds) Local search in combinatorial optimization. Wiley, Chichester, pp 311–336

Gendreau M, Laporte G, Potvin JY (2002) Metaheuristics for the capacitated VRP. In: Toth P, Vigo D (eds) The vehicle routing problem, Monographs on discrete mathematics and applications. Siam, Philadelphia, PA, pp 129–154

Gillett BE, Miller LR (1974) A heuristic algorithm for the vehicle dispatch problem. Oper Res 22:240–349

Glover F, Laguna M, Marti R (2003) Scatter search and path relinking: advances and applications. In: Glover F, Kochenberger GA (eds) Handbook of metaheuristics. Kluwer, Boston, MA, pp 1–36

Golden BL, Assad AA (1988) Vehicle routing: methods and studies. Elsevier, Amsterdam

Golden BL, Wasil EA, Kelly JP, Chao IM (1998) The impact of metaheuristics on solving the vehicle routing problem: algorithms, problem sets, and computational results. In: Crainic TG, Laporte G (eds) Fleet management and logistics. Kluwer, Boston, MA, pp 33–56

Golden B, Raghavan S, Wasil E (2008) The vehicle routing problem: latest advances and new challenges. Springer, New York, NY

Laporte G, Semet F (2002) Classical heuristics for the capacitated VRP. In: Toth P, Vigo D (eds) The vehicle routing problem, Monographs on discrete mathematics and applications. Siam, Philadelphia, PA, pp 109–128

Laporte G, Gendreau M, Potvin J-Y, Semet F (2000) Classical and modern heuristics for the vehicle routing problem. Int Trans Oper Res 7:285–300

Li F, Golden B, Wasil E (2005) Very large-scale vehicle routing: new test problems, algorithms and results. Comput Oper Res 32(5):1165–1179

Lin S (1965) Computer solutions of the traveling salesman problem. Bell Syst Tech J 44:2245–2269

Lin S, Kernighan BW (1973) An effective heuristic algorithm for the traveling salesman problem. Oper Res 21:498–516

Lourenco HR, Martin O, Stutzle T (2002) Iterated local search. In: Glover F, Kochenberger GA (eds) Handbook of metaheuristics, vol 57, Operations research and management science. Kluwer, Norwell, MA, pp 321–353

Marinakis Y, Marinaki M (2010) A hybrid genetic—particle swarm optimization algorithm for the vehicle routing problem. Expert Syst Appl 37:1446–1455

Marinakis Y, Migdalas A (2002) Heuristic solutions of vehicle routing problems in supply chain management. In: Burkard R, Pardalos PM, Migdalas A (eds) Combinatorial and global optimization. World Scientific Publishing, Singapore, pp 205–236

Marinakis Y, Migdalas A, Pardalos PM (2005a) Expanding neighborhood GRASP for the traveling salesman problem. Comput Optim Appl 32:231–257

Marinakis Y, Migdalas A, Pardalos PM (2005b) A hybrid Genetic-GRASP algorithm using Langrangean relaxation for the traveling salesman problem. J Comb Optim 10:311–326

Marinakis Y, Migdalas A, Pardalos PM (2007) A new bilevel formulation for the vehicle routing problem and a solution method using a genetic algorithm. J Glob Optim 38:555–580

Marinakis Y, Marinaki M, Dounias G (2008) Honey bees mating optimization algorithm for the vehicle routing problem. In: Krasnogor N, Nicosia G, Pavone M, Pelta D (eds) Nature Inspired Cooperative Strategies for Optimization—NICSO 2007, vol 129, Studies in Computational Intelligence. Springer, Berlin, pp 139–148

Marinakis Y, Migdalas A, Pardalos PM (2009) Multiple phase neighborhood search GRASP based on Lagrangean relaxation and random backtracking Lin Kernighan for the traveling salesman problem. J Comb Optim 17:134–156

Marinakis Y, Marinaki M, Dounias G (2010) A hybrid particle swarm optimization algorithm for the vehicle routing problem. Eng Appl Artif Intel 23:463–472

Marinakis Y, Iordanidou G, Marinaki M (2013) Particle swarm optimization for the vehicle routing problem with stochastic demands. Appl Soft Comput 13(4):1693–1704

Martin O, Otto SW, Felten EW (1991) Large-step Markov chains for the traveling salesman problem. Complex Syst 5(3):299–326

Mester D, Braysy O (2005) Active guided evolution strategies for the large scale vehicle routing problems with time windows. Comput Oper Res 32:1593–1614

Mester D, Braysy O (2007) Active guided evolution strategies for large scale capacitated vehicle routing problems. Comput Oper Res 34:2964–2975

Mole RH, Jameson SR (1976) A sequential route-building algorithm employing a generalized savings criterion. Oper Res Q 27:503–511

Moscato P, Cotta C (2003) A gentle introduction to memetic algorithms. In: Glover F, Kochenberger GA (eds) Handbooks of metaheuristics. Kluwer, Dordrecht, pp 105–144

Osman IH (1993) Metastrategy simulated annealing and tabu search algorithms for combinatorial optimization problems. Ann Oper Res 41:421–451

Pereira FB, Tavares J (2008) Bio-inspired algorithms for the vehicle routing problem, vol 161, Studies in computational intelligence. Springer, Berlin

Prins C (2004) A simple and effective evolutionary algorithm for the vehicle routing problem. Comput Oper Res 31:1985–2002

Prins C (2008) A GRASP X evolutionary local search hybrid for the vehicle routing problem. In: Pereira FB, Tavares J (eds) Bio-inspired algorithms for the vehicle routing problem, SCI 161. Springer, Berlin, pp 35–53

Rego C (1998) A subpath ejection method for the vehicle routing problem. Manage Sci 44:1447–1459

Rego C (2001) Node-ejection chains for the vehicle routing problem: sequential and parallel algorithms. Parallel Comput 27(3):201–222

Reimann M, Stummer M, Doerner K (2002) A savings based ant system for the vehicle routing problem. Proceedings of the genetic and evolutionary computation conference, New York, pp 1317–1326

Reimann M, Doerner K, Hartl RF (2004) D-Ants: savings based ants divide and conquer the vehicle routing problem. Comput Oper Res 31(4):563–591

Resende MGC, Ribeiro CC (2003) Greedy randomized adaptive search procedures. In: Glover F, Kochenberger GA (eds) Handbook of metaheuristics. Kluwer, Boston, MA, pp 219–249

Rochat Y, Taillard ED (1995) Probabilistic diversification and intensification in local search for vehicle routing. J Heuristics 1:147–167
Taillard ED (1993) Parallel iterative search methods for vehicle routing problems. Networks 23:661–672
Talbi E-G (2009) Metaheuristics: from design to implementation. Wiley, Hoboken, NJ
Tarantilis CD (2005) Solving the vehicle routing problem with adaptive memory programming methodology. Comput Oper Res 32(9):2309–2327
Tarantilis CD, Kiranoudis CT (2002) BoneRoute: an adaptive memory-based method for effective fleet management. Ann Oper Res 115(1):227–241
Tarantilis CD, Kiranoudis CT, Vassiliadis VS (2002a) A backtracking adaptive threshold accepting metaheuristic method for the Vehicle Routing Problem. Syst Anal Model Simul 42 (5):631–644
Tarantilis CD, Kiranoudis CT, Vassiliadis VS (2002b) A list based threshold accepting algorithm for the capacitated vehicle routing problem. Int J Comp Math 79(5):537–553
Toth P, Vigo D (2002) The vehicle routing problem, Monographs on discrete mathematics and applications. Siam, Philadelphia, PA
Toth P, Vigo D (2003) The granular tabu search (and its application to the vehicle routing problem). INFORMS J Comput 15(4):333–348
Toth P, Vigo D (2014) Vehicle routing: problems, methods and applications, 2nd edn, MOS-Siam series on optimization. Siam, Philadelphia, PA
Wark P, Holt J (1994) A repeated matching heuristic for the vehicle routing problem. J Oper Res Soc 45:1156–1167
Xu J, Kelly JP (1996) A new network flow-based tabu search heuristic for the vehicle routing problem. Transp Sci 30:379–393

Three-Dimensional Multiple-Bin-Size Bin Packing: A Case Study with a New MILP-Based Upper Bound

Mariem Baazaoui, Saïd Hanafi, and Hichem Kamoun

Abstract In our research, we are interested in a practical problem closely related to the three-dimensional multiple-bin-size bin packing problem. We deal with the real word application of cutting mousse blocks proposed by a Tunisian industrial company. First, we present the general context related to this optimization problem. Second, for solving this practical problem, we propose an upper bound based on a MILP formulation (mixed integer linear programming). Finally, computational and comparative results are presented to evaluate the performance of the proposed bound by testing a large instance from the same industrial company.

Keywords Multiple bin size bin packing problem • Upper bound • Mixed integer linear program

1 Introduction

3D Bin Packing problem (3D BPP) consists in minimizing the number of bins containing the pieces ordered by the clients without overlapping. In the case when the bins have the same dimensions, it is the classical 3D BPP. In the other case when the bins have different dimensions and costs we have the multiple-bin-size bin packing problem (MBSBPP), with the objective of minimizing the total cost of the used bins. Both problems have many real applications in different industries such that wood, glass and steel industry. There are many studies in the literature considering the classical problem as example (Verweij 1996; Lodi et al. 2002; Crainic et al. 2008; Fekete and van der Veen 2007; Crainic et al. 2009; Faroe et al. 2003; Parreño et al. 2008; Martello et al. 2000; Bortfeldt and Wascher 2012; deQueiroz et al. 2012; Hifi et al. 2014) but

M. Baazaoui (✉) • H. Kamoun
MODILS Laboratory, Faculty of Economics and Management of Sfax, University of Sfax, Sfax, Tunisia
e-mail: baazaoui.mariem@gmail.com; hichem.kamoun@fsegs.rnu.tn

S. Hanafi
LAMIH Laboratory, University of Valenciennes and Hainaut-Cambrésis, Famars, France
e-mail: said.hanafi@univ-valenciennes.fr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_11

there are only a few recent papers concerning the MBSBPP even for one and two-dimensional cases. For the three dimensional case, in the paper Chen et al. (1995) the authors presented an analytical model, and formulated the problem as a mixed integer linear programming (MILP) model including the consideration of multiple containers, multiple object sizes, object orientations, and the overlapping of pieces in the container. For the two dimensional case, the authors in Pisinger and Sigurd (2005) presented an integer-linear formulation of the two-dimensional variable size bin packing problem (2DVSBPP) and introduced several lower bounds for the considered problem. In the paper Correia et al. (2008), the authors proposed several formulations for the 1D-MBSBPP and developed new Lower bounds for three dimensional MBSBPP and valid inequalities. For solving this type of problems in the two dimensions we can find also the paper of Hopper (2000) who combined a simple heuristic with a genetic algorithm, we can note also the papers of Pisinger and Sigurd (2005) and Ortmann et al. (2010), who proposed a new and improved level heuristics for the rectangular strip packing and variable-sized bin packing problems.

In this paper, we focus on a practical case of an industrial company of cutting blocks mousse. It can be conceived as a combination of optimization problems such the assignment problem, the 3D strip packing, the 3DMBSBPP and the Three dimensional cutting stock problem multiple stock size (CSPMSS), but it also considers the constraints that related to the cutting specifications. The definition of these problems can be founded in the typology developed by Wascher et al. (2007). The MBSBPP can be defined as follows: Given a set N of n three-dimensional rectangular boxes, where each box $j \in N$ has a width wj, a height hj and a length lj, and a set M of m three-dimensional bin types, where each bin $i \in M$ has a width Wi, height Hi, length Li and cost Ci. The objective is to find an orthogonal packing (the edges of the boxes parallel to the edges of the bin), without overlapping, of all the boxes into the bins at minimum cost. According to the same typology in Wascher et al. (2007), the Cutting Stock Problem with Multiple Stock Sizes (CSPMSS) is related to the MBSBPP which is defined by a set of classes of pieces, each class with a given demand, have to be cut from a set of types of stock sheets, while minimizing the total number of sheets used. For the MBSBPP the pieces to be packed are strongly heterogeneous while for the CSPMSS the pieces are weakly heterogeneous. We note also that in the cutting stock there are industrial constraints such the guillotine constraint. The paper Furini and Malaguti (2013) deals with two-dimensional cutting stock problem where stock of different sizes is available, and a set of rectangular items has to be obtained through two-staged guillotine cuts. They defined a MILP model for the solution of this knapsack problem, as well as a heuristic procedure based on dynamic programming. The industrial practical problem related to this type of problem is studied by Tae Park et al. (2013), where the authors developed a heuristic for glass cutting process optimization: A case of two-dimensional two-stage guillotine cutting with multiple stock sizes. Some works in the literature have been devoted to industrial cases studies such as Lu et al. (2013) for cutting stock problems in the thin-film transistor liquid crystal display industry and Ertek and Kilic (2006), who propose a beam search algorithm, a greedy algorithm and a tree search enumeration algorithm to solve a packing problem in a major automobile manufacturer in Turkey. Another industrial application was investigated in Joaquim et al. (2014), the author presented a tabu search algorithm for the 3D bin packing problem in the steel industry, after they proved the performance of their method compare with an heuristic and ant colony optimization (ACO) algorithms.

The paper is organized as follows, in the first section we introduce the practical studied case whereas in the second section, we present an upper bound based on a mathematical formulation.

2 Case Study: Problem Description

In Baazaoui et al. (2014), we defined the practical problem through two phases: the production phase of the bins called a sizing phase and the cutting phase. In the sizing phase, the bins (blocks) are not initially available in stock but their dimensions are to be decided within the decision making process. In fact, the company has a number of possible types characterized by height and width for each tunnel (succession of block number with same height and width but different length) and its main aim is to decide (assign) the best type for each tunnel. The length (L) of each block of each tunnel must be between a minimum and maximum value $L_{\min}$ and $L_{\max}$ respectively. The total length of tunnel (L_T) must be minimized (see Fig. 1).

The cutting phase provides us with finished pieces ready for sale to customers. The cutting must respect the guillotine constraint in order to meet the characteristics of the cutting blades. The cut is done by several parallel horizontal knives for cutting the sub-blocks and other knives for the final extraction of the pieces (all cutting are guillotine: A guillotine cut is a cut that is parallel to one of the sides of the bin and goes from one side to the opposite one) (see Fig. 2). In our problem, the decision should be taken on two levels; the cutting level with the objective of seeking the best configuration and the sizing level of the bins (blocks) with the objective of assigning the best type of (height and width) to the best configuration and determine the length that belongs to an interval. We do mention that both levels of decision are interdependent from each other.

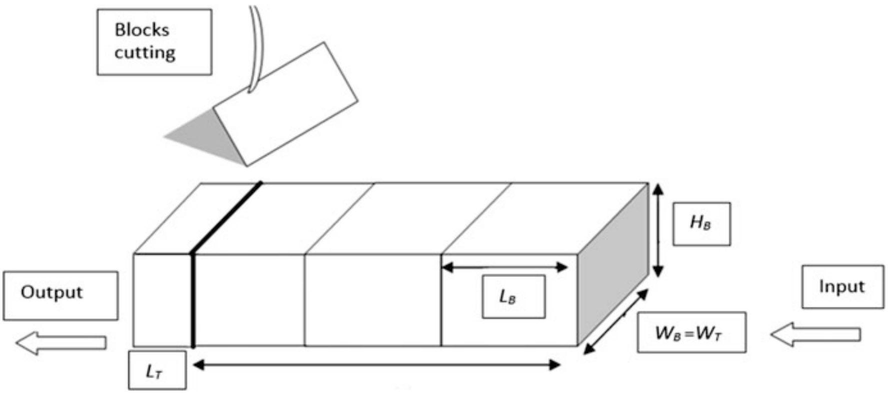


Fig. 1 Cut blocks of a tunnel

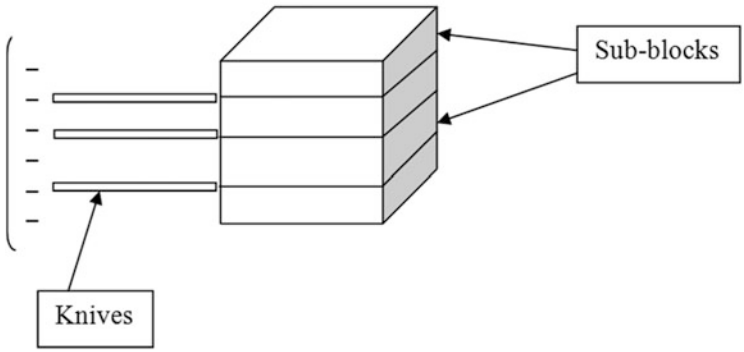


Fig. 2 Cut blocks of a tunnel

The practical problem of SOTIM Company considered in Baazaoui et al. (2014) can be described as follows. A set N of pieces is given, where each piece i has different size (h_i: height, l_i: length and w_i: width) that must be cut out of blocks of dimensions H, W and L. The production of the blocks is done by tunnel; the blocks of the same tunnel must have the same parameters: (H: height; W: width). The guillotine constraints must be respected. The objective is to find a configuration that places the pieces in the tunnels with minimal trim loss (cutting) while respecting the guillotine constraint. To find (H, W) of the tunnels in question and the length L_i associated with each block (sizing), we proposed in Baazaoui et al. (2014) a MILP model and a lower bound, while in this paper we propose an upper bound that constitutes a feasible solution for the industrial problem.

3 Upper Bound Deduced From Exact Formulation

In this section we propose an upper bound address the above problem. Our idea is based on some modifications of the lower bound proposed in Ertek and Kilic (2006) so we have the same parameters:

N	set of n items
M	set of m tunnels
N_k	set of maximal number of n_k blocks in a tunnel k
$L_{\min}$, $L_{\max}$	minimum and maximum length of item i
$B_{\min}$, $B_{\max}$	minimum and maximum number of sub-block for each tunnel
l_i	length of item i
w_i	width of item i
h_i	height of item i
W_{jk}	width of sub-block j of tunnel k
H_{jk}	height of sub-block j of tunnel k

Decision Variables:

$S_k = 1$	if the tunnel k is used; else 0.
$Y_{jk} = 1$	if the sub-block j is assigned to tunnel k; else 0 (is used)
$X_{ijk} = 1$	if the item i is assigned to sub-block j of tunnel k; else 0.
L_{jk}	integer variable which corresponds to length for sub-block j of tunnel k
$Z_{jk} = L_{jk} Y_{jk}$	variable used to linearize the objective function.

The proposed upper bound to the industrial problem can be obtained using the following formulation:

$$\text{Min z} = \sum_{k=1}^{m} \sum_{j=1}^{n_k} Z_{jk} \quad (1)$$

s.t.

$$\sum_{j=1}^{n_k} \sum_{k=1}^{m} X_{ijk} = 1 \qquad \forall\, i \in N \quad (2)$$

$$\sum_{i=1}^{n} X_{ijk} \leq n Y_{jk} \qquad \forall\, k \in M,\ j \in N_k \quad (3)$$

$$X_{i\,jk} \leq S_k \qquad \forall\, i \in N,\ k \in M,\ j \in N_k \quad (4)$$

$$S_{k+1} \leq S_k \qquad \forall\, k \in M \quad (5)$$

$$Y_{j+1,k} \leq Y_{jk} \qquad \forall\, k \in M,\ j \in N_k \quad (6)$$

$$B_{min} S_k \leq \sum_{j=1}^{n_k} Y_{jk} \leq B_{\max} S_k \qquad \forall\, k \in M \quad (7)$$

$$L_{\min} \leq L_{j\,k} \leq L_{\max} \qquad \forall\, k \in M,\ j \in N_k \quad (8)$$

$$\sum_{i=1}^{n} l_i\, w_i h_i X_{ijk} \leq H_{j\,k} W_{j\,k} L_{j\,k} \qquad \forall\, k \in M,\ j \in N_k \quad (9)$$

$$l_i X_{ijk} \leq L_{j\,k} \qquad \forall\, i \in N,\ k \in M,\ j \in N_k \quad (10)$$

$$\sum_{i=1}^{n} h_i\, X_{ijk} \leq H_k \qquad \forall\, k \in M,\ j \in N_k \quad (11)$$

Or

$$\sum_{i=1}^{n} w_i\, X_{ijk} \leq W_k \qquad \forall\, k \in M,\ j \in N_k \quad (12)$$

$$Z_{jk} \geq L_{jk} - M\left(1 - \mathrm{Y}_{jk}\right) \qquad \forall\, k \in M,\ j \in N_k \quad (13)$$

$$Z_{jk} \leq M Y_{jk} \qquad \forall\, k \in M,\ j \in N_k \quad (14)$$

$$Z_{jk} \leq L_{j\,k} \qquad \forall\, k \in M,\ j \in N_k \quad (15)$$

$$L_{jk} \geq 0 \qquad \forall\, k \in M,\ j \in N_k \quad (16)$$

$$S_k, Y_{jk}, X_{ijk} \in \{0, 1\} \qquad \forall\, i \in N,\ k \in M,\ j \in N_k \quad (17)$$

The objective function (1) minimizes the total volume of the tunnels used. Constraint (2) ensures that each item (piece) must be packed. Constraint (3) ensures that no item is assigned an unused sub-block. Constraint (4) ensures that an item may be packed only if the tunnel is used. Constraint (5) ensures that the tunnel $k + 1$ is used only if the tunnel k is used. Constraint (6) ensures that sub-block $j + 1$ is used only if the sub-block j is used. Constraint (7) ensures that the number of sub-block in each tunnel must be between two bound values $B_{\max}$ and $B_{\min}$. Constraint (8) ensures that no sub-block has length less than $L_{\min}$ value: the minimum length of all pieces and no sub-block has length bigger than $L_{\max}$: the maximum length of all pieces. Constraint (9) states that the sum of volume of all items assigned to a sub-block must be lower than the volume of this sub-block. Constraint (10) ensures that the length of the item should not exceed that of the sub-block.

Indeed, constraint (11) ensures that if we have at maximum one item along the width and all the items in this sub-block are superimposed (one upon the other). So the sum of the height of the items should not exceed that of the sub-block (see Fig. 3). Constraint (12) ensures that if we have at maximum one item along the height and all the items in this sub-block are Contiguous (one side by side) (see Fig. 4). So the sum of the width of the item should not exceed that of the sub-block. Constraints (14) and (15) introduce a linearization of the quadratic terms: $Z_{jk} = L_{jk}Y_{jk}$.

This mathematical formulation gives an example of one of the feasible solution. In fact, it respects all constraints including positioning and not overlapping pieces but it doesn't represent them in the general case. Therefore, this formulation considers the two possible extreme cases (in the model presented by the constraints 11 and 12 which are illustrated both in Figs. 3 and 4).

To test this formulation, we took all the possibilities of H and W (both in cm), which are as follows: $H = 140$, 150, 160, 170, 180, 190 and $W = 160$, 170, 180, 190, 200, 210. We respect the cutting with made by sub-blocks and each sub-block corresponds to a single layer of superimposed pieces.

Under these conditions, we tested our formulation respectively for 10, 20 and 30 items, which represent a random list of demands obtained from the industry. In the following, we test 50, 70, 100 pieces. The associated results are given in Table 1.

For the first three instances (10, 20 and 30), one tunnel with $H = 140$ and $W = 160$ is sufficient to contain all the pieces. For the same instances, the execution time is less than 3 s. With this formulation, we can have results for 50 pieces with execution time less than 2 min. We don't have results for 70 and 100 items. This bound gives results in the case of small instances within a reasonable time. All these pieces are taken randomly from the industry demand. In the table below, in order to test our bounds better, we will firstly focus on other types of pieces. We will test respectively long items where the difference between the three dimensions (l_i: length of item i, w_i: width of item i, h_i: height of item i) is important, square items where the difference between the three dimensions (l_i, w_i, h_i) is low. Finally, we will test instance of items aggregating the two previous types of items. The comparative study is given in Table 2.

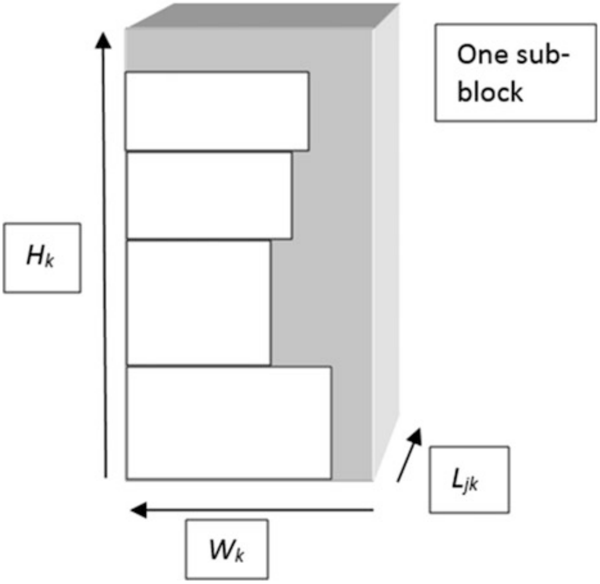


Fig. 3 The first case: the pieces are the one upon the other

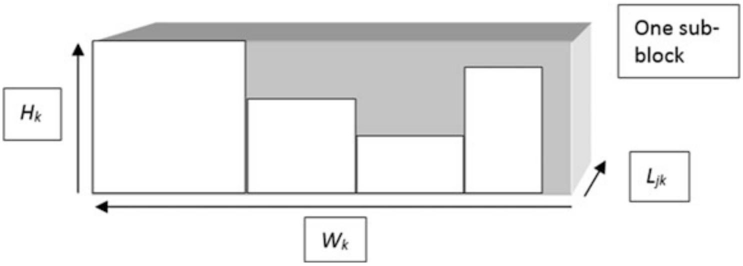


Fig. 4 The second case: the pieces are the one beside the other

For all instances cited in the above table the value of the upper bound and that of lower are equal but in different times of execution. This experiment has shown that we may have an optimal solution for up to 10 items regardless of their types.

Secondly, to test the effectiveness of the proposed upper bound, we calculate for the same instances the Gap between the proposed bound and the lower bound tested in Ertek and Kilic (2006). We note the different results in the following table (Table 3).

We remark for the first instance; we have the same results in UB and LB, so that corresponds exactly to the optimal solution of the industrial problem. This result confirms the previous results that for the instances composed of less than 10 pieces, the lower bound provides the optimal solution. This bound remains ineffective for the instances composed of more than 50 pieces.

Table 1 Upper bound

Instances	Number of tunnels used	Number and type of sub-blocks used	Length of sub-blocks used L_j (cm)	Execution time in seconds
10	One tunnel: (140; 160)	Two sub-block	170; 80	1.1
20	One tunnel: (140; 160)	Three sub-blocks	81; 80; 170	2.02
30	One tunnel: (140; 160)	Two sub-block	160; 100	2.42
50	Five tunnels: (140; 160); (150; 170); (160; 180); (140; 160); (170; 190); (180; 200);	One sub-block in each tunnel	60; 100; 100; 110; 160	107
70		–	–	–
100		–	–	–

Table 2 Comparative results for different types of items

	Length (cm)		Time (s)		
Instances	LB	UB	LB	UB	Gap = (UP-LB)/LB
5	390	390	0.78	0.50	0
5	200	200	0.88	1.03	0
5	180	180	0.85	0.56	0
10	390	390	0.95	0.92	0
10	200	200	0.77	0.85	0
10	180	180	0.84	0.98	0

Table 3 Comparatives results

	Length (cm)		Time (s)		
Instances	LB	UB	LB	UB	Gap = (UP-LB)/LB
10	250	250	0.9	1.10	0
20	250	331	1.2	2.02	0.324
30	215	260	1.81	2.42	0.200
50	267	530	7.1	107	0.980
70		–			–
100		–			–

4 Conclusions

In this paper, we propose an upper bound for an industrial problem. It is deduced by adding specifications (constraints) to the lower bound already proposed in our previous work. This upper bound constitutes a feasible solution for our problem.

The result from the upper bound shows that we can obtain solutions for 50 items (pieces) in a lap of time less than 2 min and provide an optimal solution in the case of 10 pieces. In future, we plan to propose a metaheuristic method to adapt the solution for a larger number of pieces. We plan on combining our MILP with one of metaheuristic techniques presented in the literature. We can also emphasize in the nature of the industry demand by clustering the different demands in groups to improve the results.

References

Baazaoui M, Hanafi S, Kamoun H (2014) A mathematical formulation and a lower bound for the three-dimensional multiple-bin-size bin packing problem (MBSBPP): a Tunisian industrial case. In: International conference on control, decision and information technologies (CoDIT), janvier. IEEE, Metz, pp 219–224

Bortfeldt A, Wascher G (2012) Container loading problems—a state- of-the- art-review. Working Paper 7

Chen CS, Lee SM, Shen QS (1995) An analytical model for the container loading problem". Eur J Oper Res 80(1):68–76

Correia I, Gouveia L, Saldanha-da-Gama F (2008) Solving the variable size bin packing problem with discretized formulations". Comput Oper Res 35:03–13

Crainic TG, Perboli G, Tadei R (2008) Extreme point-based heuristics for three dimensional bin packing. INFORMS J Comput 20:368–384

Crainic TG, Perboli G, Tadei R (2009) TS2 PACK: a two-level Tabu search for the three-dimensional bin packing problem". Eur J Oper Res 195:744–760

deQueiroz TA, Miyazawa FK, Wakabayashi Y, Xavier EC (2012) Algorithms for 3D guillotine cutting problems: unbounded knapsack, cutting stock and strip packing. Comput Oper Res 39:200–212

Ertek G, Kilic K (2006) Decision support for packing in warehouses". Lect Notes Comput Sci 4263:115–124

Faroe O, Pisinger D, Zachariasen M (2003) Guided local search for the three dimensional bin-packing problem". INFORMS J Comput 15:267–283

Fekete SP, van der Veen JC (2007) PackLib2: an integrated library of multidimensional packing problems. Eur J Oper Res 183:1131–1135

Furini F, Malaguti E (2013) Models for the two-dimensional two-stage cutting stock problem with multiple stock size. Comput Oper Res 40:1953–1962

Hifi M, Negre S, Wu L (2014) Hybrid greedy heuristics based on linear programming for the three-dimensional single bin-size bin packing problem. Int Trans Oper Res 21:59–79

Hopper E (2000) Two-dimensional packing utilising evolutionary algorithms and other meta-heuristic methods, PhD dissertation. University of Wales, Cardiff

Joaquim V, Susana V, Elsa H, João S (2014) A Tabu search algorithm for the 3D bin packing problem in the steel industry. In: CONTROLO'2014—Proceedings of the 11th Portuguese conference on automatic control, Volume 321 of the series Lecture Notes in Electrical Engineering, pp 355–364

Lodi A, Martello S, Vigo D (2002) Heuristic algorithms for the three-dimensional bin packing problem. Eur J Oper Res 141:410–420

Lu H, Huang Y, Tseng K (2013) An integrated algorithm for cutting stock problems in the thin-film transistor liquid crystal display industry". Comput Ind Eng 64:1084–1092

Martello S, Pisinger D, Vigo D (2000) The three-dimensional bin packing problem. Oper Res 48:256–267

Ortmann FG, Ntene N, Van Vuuren JH (2010) New and improved level heuristics for the rectangular strip packing and variable-sized bin packing problems. Eur J Oper Res 203 (2):306–315

Parreño F, Alvarez-Valdes R, Oliveira JF, Tamarit JM (2008) A hybrid GRASP/VND algorithm for two- and three-dimensional bin packing. Ann Oper Res 179:203–220

Pisinger P, Sigurd M (2005) The two-dimensional bin packing problem with variable bin sizes and costs". Discret Optim 2:154–167

Tae Park K, Ryu J, Lee H, Lee I (2013) Developing a heuristics for glass cutting process optimization: a case of two-dimensional two-stage guillotine cutting with multiple stock sizes. Korean J Chem Eng 30(2):278–285

Verweij B (1996) Multiple destination bin packing. Technical report CS-1996-39. Department of Information and Computing Sciences, Faculty of Science, Utrecht University 3508 TC Utrecht, Netherlands

Wascher G, Haussner H, Schumann H (2007) An improved typology of cutting and packing problems. Eur J Oper Res 183:1109–1130

The Effects of Quality on Market Share and Profitability in Single Stage Make-to-Stock Production Systems

Dimitrios Konstantas, Stratos Ioannidis, Evangelos Grigoroudis, and Vassilis S. Kouikoglou

Abstract We develop simple models for understanding how the dynamics of quality may affect customer satisfaction and profitability in make-to-stock manufacturing systems. We study a Markovian, single stage system facing random demand. Any demand not satisfied immediately from stock is lost to competitors. The market is assumed to be finite and comprises both *regular* and *occasional* customers. Regular customers have higher mean demand rates than occasional customers. Each outgoing product is inspected and classified as *high quality*, *medium quality*, or *nonconforming*. The customer who purchases an item joins the regular or the occasional class, with corresponding probabilities which depend on the quality level and on past customer state. The higher the quality level, the higher the probability for a customer to remain or become a regular customer. Our goal is to investigate the structure of the optimal production, order satisfaction and quality control policy in order to maximize the average profit rate of the system. We investigate numerically the structure of the optimal policy using stochastic dynamic programming.

Keywords Production control • Quality control • Markov decision processes • Make-to-stock production systems • Customer satisfaction

1 Introduction

Customer satisfaction is considered a key determinant of profitability and sustainability of manufacturing firms, due to its direct impact on market share and, consequently, on sales and revenue rates. Customer satisfaction is a complex, dynamic process, which depends on many factors, such as product quality, product

---

D. Konstantas (✉) • S. Ioannidis • E. Grigoroudis • V.S. Kouikoglou
School of Production Engineering and Management, Technical University of Crete, Chania, Greece
e-mail: dkonstantas1@isc.tuc.gr; efioan@dpem.tuc.gr; vangelis@ergasya.tuc.gr; kouik@dpem.tuc.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_12

price, lead time in filling customer orders, and so on. Its nature as well as its antecedents and consequences have been subjects of extensive study (Anderson et al. 1994; Oliver 1997). Rust and Zahorik (1993) developed the first method for assessing the effect of satisfaction on customer retention and on market share. In the same year, Anderson and Sullivan (1993) published a study on the factors affecting customer satisfaction and its effect on the financial performance of a firm. The same study also observed asymmetries in the effects of opposite variations (satisfaction, dissatisfaction) on profit, which is in agreement with expert estimates and empirical studies suggesting that new customer attraction is 3 to 30 times more costly than customer retention.

Recently, Kostami and Rajagopalan (2014) examine the trade-offs between quality, speed, and price in a monopoly setting, using queueing theory and convex optimization while Deng et al. (2014) develop models for estimating optimal inventory levels in systems with service-dependent demand.

Although customer satisfaction has received considerable attention by marketing researchers, it has not yet been studied to the same extend jointly with the problems of production and quality control. Towards this end, Kouikoglou and Phillis (2002), Gershwin and Schick (2007) among others have studied problems in which production design or control decisions are coupled with quality control strategies. In this paper we take the interplay of quality and production optimization one step further by also considering the effects of these decisions on the market size of a company. In the next section we describe a single stage, make-to-stock manufacturing system whose market comprises regular and occasional customers. We then investigate through numerical experiments the optimal production and quality control policy which maximizes the average profit rate of the system based on exact knowledge of the current state and finally we present the conclusions.

2 System Description

Consider a single-stage production facility producing a single product, which fills incoming orders from a stock of already produced items (make-to-stock system). Each outgoing product has a quality level i ($i = 1, 2, 3$ with quality decreasing as i increases) with corresponding probability p_i. If an outgoing product is of quality level 3, then it is automatically scrapped, otherwise we keep it in stock along with products of the same quality level. Suppose that the market size M is known and constant. The M customers are divided in two distinct classes: the class of regular customers, denoted $k = 1$, and the class of occasional customers, $k = 2$. The state of the system can be described by three variables:

x inventory level of products of quality level 1
y inventory level of products of quality level 2
z number of regular customers in orbit.

Customers arrive according to Poisson processes and each customer requests one unit of product. Each customer has a class-dependent mean demand rate, λ_k, $k = 1, 2$, where $\lambda_1 > \lambda_2$. Item processing times are independent random variables exponentially distributed with mean $1/\mu$. If a product of quality level i is sold to a customer of class k, then this customer will eventually be satisfied with (conditional) probability s_i^k or dissatisfied with the complementary probability. Satisfied customers join the regular class while dissatisfied ones become occasional customers. We assume that high quality items ($i = 1$) correspond to higher satisfaction probabilities, ie, $s_1^k > s_2^k$ for every $k = 1, 2$.

The overall performance measure of the system is the mean profit rate. This quantity depends on the sales rate (which equals the system's throughput), the costs of rejected items, and the holding costs. We consider three economic parameters:

r revenue from selling a product to a customer
c unit rejection cost (cost for scrapping an item)
h unit holding cost per item and per time unit.

The state of the system is described by the variable set (x, y, z) with state space $S = \{(x, y, z) \mid x = 0, 1, \ldots, K, y = 0, \ldots, K$ and $z = 0, \ldots, M\}$. Even though there is no obvious limit for the values of x and y, it can be easily proved that in order to keep the system profitable the overall inventory should not exceed $K = M\lambda_1$r/h, and thus $x + y \leq K$. At each production epoch we make a production decision $\pi_\mu(x, y, z)$, where $\pi_\mu(x, y, z) = 1$ if we continue producing in state (x, y, z) and $\pi_\mu(x, y, z) = 0$ otherwise. When an item is scrapped the state of the system does not change. When an order from a class 1 customer arrives, we make a selling decision for regular customers $\pi_{\lambda 1}(x, y, z)$, where $\pi_{\lambda 1}(x, y, z) = 1$ if we decide to fill the order with a quality level 1 product or $\pi_{\lambda 1}(x, y, z) = 2$ if we decide to sell quality level 2 product. If the customer is satisfied with the item's quality then it remains regular and the system state changes to $(x - 1, y, z)$ or $(x, y - 1, z)$ respectively. In case of dissatisfaction, the customer becomes occasional and the system state changes to $(x - 1, y, z - 1)$ or $(x, y - 1, z - 1)$ respectively. When $x = 0$ we have an extra option of not servicing the order and consequently dissatisfying the customer, thus changing the system state to $(x, y, z - 1)$. When we have an order from class 2 customers, we make a selling decision for occasional customers $\pi_{\lambda 2}(x, y, z)$, where $\pi_{\lambda 2}(x, y, z) = 1$ if we decide to fill the order with a product of quality level 1, or $\pi_{\lambda 2}(x, y, z) = 2$ if a product of quality level 2 is sold, or finally $\pi_{\lambda 2}(x, y, z) = 0$ if we decide to reject the incoming order. A customer who is satisfied with the item's quality becomes a regular customer and the system state changes to $(x - 1, y, z + 1)$ or $(x, y - 1, z + 1)$ respectively. In case of dissatisfaction the customer remains occasional and the system state changes to $(x - 1, y, z)$ or $(x, y - 1, z)$ respectively. In case of rejecting the order the state of the system doesn't change.

3 Dynamic Programming

The system evolution is driven by two types of events: order placement by a type k customer and production of a quality i item. By using the uniformization technique (Bertsekas 1995; Sennott 1999) we formulate the problem as a Markov decision process. The fixed size M of the market as well as the overall inventory limit K, bound the state space and ensure stability of the system. The continuous time Markov chain is uniformized using the maximum transition rate $v = M\lambda_1 + \mu$, which is a uniform rate of event occurrences that permits the conversion of continuous time into a sequence of stages. Following the standard stochastic dynamic programming approach, we define $V_k(x, y, z)$ as the total expected profit over the first k events (or decision stages) when the initial state is (x, y, z). The Bellman equations for the optimal total expected profit over the first $k+1$ events are:

$$\begin{aligned} V_{k+1}(0,0,0) &= \frac{1}{v}\{\mu\max[p_3(V_k(0,0,0)-c)+p_1V_k(1,0,0) \\ &\quad + p_2V_k(0,1,0), V_k(0,0,0)] + M\lambda_1 V_k(0,0,0)\} \end{aligned} \tag{1}$$

$$\begin{aligned} V_{k+1}(0,0,z) &= \frac{1}{v}\{\mu\max[p_3(V_k(0,0,z)-c)+p_1V_k(1,0,z) \\ &\quad + p_2V_k(0,1,z), V_k(0,0,z)] + z\lambda_1 V_k(0,0,z-1) \\ &\quad + (M-z)\lambda_1 V_k(0,0,z)\}, 1 \le z \le M \end{aligned} \tag{2}$$

$$\begin{aligned} V_{k+1}(x,0,0) &= \frac{1}{v}\{\mu\max[p_3(V_k(x,0,0)-c)+p_1V_k(x+1,0,0) \\ &\quad + p_2V_k(x,1,0), V_k(x,0,0)] + M\lambda_2\max[s_1^2V_k(x-1,0,1) \\ &\quad + (1-s_1^2)V_k(x-1,0,0)+r, V_k(x,0,0)] \\ &\quad + M(\lambda_1-\lambda_2)\ V_k(x,0,0) - xh\},\ 0<x<K \end{aligned} \tag{3}$$

$$\begin{aligned} V_{k+1}(0,y,0) &= \frac{1}{v}\{\mu\max[p_3(V_k(0,y,0)-c)+p_1V_k(1,y,0) \\ &\quad + p_2V_k(0,y+1,0),\ V_k(0,y,0)] + M\lambda_2\max[s_2^2V_k(0,y-1,1) \\ &\quad + (1-s_2^2)V_k(0,y-1,0)+r, V_k(0,y,0)] \\ &\quad + M(\lambda_1-\lambda_2)V_k(0,y,0) - yh\},\ 0<y<K \end{aligned} \tag{4}$$

$$\begin{aligned} V_{k+1}(x,y,0) &= \frac{1}{v}\{\mu\max[p_3(V_k(x,y,0)-c)+p_1V_k(x+1,y,0) \\ &\quad + p_2V_k(x,y+1,0),V_k(x,y,0)] + M\lambda_2\max[s_1^2V_k(x-1,y,1) \\ &\quad + (1-s_1^2)V_k(x-1,y,0)+r, s_2^2V_k(x,y-1,1) \\ &\quad + (1-s_2^2)V_k(x,y-1,0)+r, V_k(x,y,0)] \\ &\quad + M(\lambda_1-\lambda_2)V_k(x,y,0)-(x+y)h\},\ x>0,\ y>0,\ x+y<K \end{aligned} \tag{5}$$

$$\begin{aligned} V_{k+1}(0,y,z) &= \frac{1}{v}\{\mu\max[p_3(V_k(0,y,z)-c)+p_1V_k(1,y,z)+p_2V_k(0,y \\ &\quad + 1,z),V_k(0,y,z)] + z\lambda_1\max[s_2^1V_k(0,y-1,z) \\ &\quad + (1-s_2^1)V_k(0,y-1,z-1)+r,\ V_k(0,y,z-1)] \\ &\quad + (M-z)\lambda_2\max[s_2^2V_k(0,y-1,z+1)+(1-s_2^2)V_k(0,y-1,z) \\ &\quad + r, V_k(0,y,z)] + (M-z)(\lambda_1-\lambda_2)V_k(0,y,z)-yh\}, \\ &\quad 0<y<K,\ 0<z<M \end{aligned} \tag{6}$$

$$\begin{aligned} V_{k+1}(0,y,M) &= \frac{1}{v}\{\mu\max[p_3(V_k(0,y,M)-c)+p_1V_k(1,y,M)+p_2V_k(0,y \\ &\quad + 1,M),V_k(0,y,M)] + M\lambda_1\max[s_2^1V_k(0,y-1,M) \\ &\quad + (1-s_2^1)V_k(0,y-1,M-1)+r,V_k(0,y,M-1)]-yh\}, 0<y<K \end{aligned} \tag{7}$$

$$\begin{aligned} V_{k+1}(x,0,M) &= \frac{1}{v}\{\mu\max[p_3(V_k(x,0,M)-c)+p_1V_k(x+1,0,M) \\ &\quad + p_2V_k(x,1,M),V_k(x,0,M)] + M\lambda_1[s_1^1V_k(x-1,0,M) \\ &\quad + (1-s_1^1)V_k(x-1,0,M-1)+r]-xh\},\ 0<x<K \end{aligned} \tag{8}$$

$$\begin{aligned} V_{k+1}(x,0,z) &= \frac{1}{v}\{\mu\max[p_3(V_k(x,0,z)-c)+p_1V_k(x+1,0,z) \\ &\quad + p_2V_k(x,1,z),V_k(x,0,z)] + z\lambda_1[s_1^1V_k(x-1,0,z) \\ &\quad + (1-s_1^1)V_k(x-1,0,z-1)+r] \\ &\quad + (M-z)\lambda_2\max[s_1^2V_k(x-1,0,z+1) \\ &\quad + (1-s_1^2)V_k(x-1,0,z)+r,V_k(x,0,z)] \\ &\quad + (M-z)(\lambda_1-\lambda_2)V_k(x,0,z)-xh\},\ 0<x<K,\ 0<z<M \end{aligned} \tag{9}$$

$$
\begin{aligned}
V_{k+1}(x,y,M) = &\frac{1}{v}\{\mu \max[p_3(V_k(x,y,M)-c)+p_1V_k(x+1,y,M)\\
&+p_2V_k(x,y+1,M),V_k(x,y,M)]\\
&+M\lambda_1\max[s_1^1\ V_k(x-1,y,M)+(1-s_1^1)V_k(x-1,y,M-1)\\
&+r,s_2^1V_k(x,y-1,M)+(1-s_2^1)V_k(x,y-1,M-1)\\
&+r]-(x+y)h\},\ x>0,y>0,x+y<K
\end{aligned} \tag{10}
$$

$$
\begin{aligned}
V_{k+1}(K,0,0) = &\frac{1}{v}\{M\lambda_2\max[s_1^2V_k(K-1,0,1)+(1-s_1^2)V_k(K-1,0,0)\\
&+r,V_k(K,0,0)]+[\mu+M(\lambda_1-\lambda_2)]V_k(K,0,0)-Kh\}
\end{aligned} \tag{11}
$$

$$
\begin{aligned}
V_{k+1}(0,K,0) = &\frac{1}{v}\{M\lambda_2\max[s_2^2V_k(0,K-1,1)+(1-s_2^2)V_k(0,K-1,0)\\
&+r,V_k(0,K,0)]+[\mu+M(\lambda_1-\lambda_2)]V_k(0,K,0)-Kh\}
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
V_{k+1}(K,0,M) = &\frac{1}{v}\{M\lambda_1[s_1^1V_k(K-1,0,M)+(1-s_1^1)V_k(K-1,0,M-1)+r]\\
&+\mu V_k(K,0,M)-Kh\}
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
V_{k+1}(0,K,M) = &\frac{1}{v}\{M\lambda_1\max[s_2^1V_k(0,K-1,M)+(1-s_2^1)V_k(0,K-1,M\\
&-1)+r,\ V_k(0,K,M-1)]+\mu V_k(0,K,M)-Kh\}
\end{aligned} \tag{14}
$$

$$
\begin{aligned}
V_{k+1}(0,K,z) = &\frac{1}{v}\{z\lambda_1\max[s_2^1V_k(0,K-1,z)+(1-s_2^1)V_k(0,K-1,z-1)\\
&+r,\ V_k(0,K,z-1)]+(M-z)\lambda_2\max[s_2^2V_k(0,K-1,z\\
&+1)+(1-s_2^2)V_k(0,K-1,z)+r,\ V_k(0,K,z)]\\
&+[\mu+(M-z)(\lambda_1-\lambda_2)]V_k(0,K,z)-Kh\},0<z<M
\end{aligned} \tag{15}
$$

$$
\begin{aligned}
V_{k+1}(K,0,z) = &\frac{1}{v}\{z\lambda_1[s_1^1V_k(K-1,0,z)+(1-s_1^1)V_k(K-1,0,z-1)+r]\\
&+(M-z)\lambda_2\max[s_1^2V_k(K-1,0,z+1)+(1-s_1^2)V_k(K\\
&-1,0,z)+r,\ V_k(K,0,z)]\\
&+[\mu+(M-z)(\lambda_1-\lambda_2)]V_k(K,0,z)-Kh\},\ 0<z<M
\end{aligned} \tag{16}
$$

$$
\begin{aligned}
V_{k+1}(x, K-x, 0) = \frac{1}{v}\{ & M\lambda_2 \max[s_1^2 V_k(x-1, K-x, 1) \\
& + (1-s_1^2)V_k(x-1, K-x, 0) + r,\ s_2^2 V_k(x, K-x-1, 1) \\
& + (1-s_2^2)V_k(x, K-x-1, 0) + r, V_k(x, K-x, 0)] \\
& + [\mu + M(\lambda_1 - \lambda_2)]V_k(x, K-x, 0) - Kh\},\ 0 < x < K
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
V_{k+1}(x, K-x, M) = \frac{1}{v}\{ & M\lambda_1 \max[s_1^1 V_k(x-1, K-x, M) \\
& + (1-s_1^1)V_k(x-1, K-x, M-1) \\
& + r, s_2^1 V_k(x, K-x-1, M) \\
& + (1-s_2^1)V_k(x, K-x-1, M-1) + r] \\
& + \mu V_k(x, K-x, M) - Kh\},\ 0 < x < K
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
V_{k+1}(x, K-x, z) = \frac{1}{v}\{ & z\lambda_1 \max[s_1^1 V_k(x-1, K-x, z) \\
& + (1-s_1^1)V_k(x-1, K-x, z-1) \\
& + r, s_2^1 V_k(x, K-x-1, z) + (1-s_2^1)V_k(x, K-x-1, z-1) \\
& + r + (M-z)\lambda_2 \max[s_1^2 V_k(x-1, K-x, z+1) \\
& + (1-s_1^2)V_k(x-1, K-x, z) + r,\ s_2^2 V_k(x, K-x-1, z+1) \\
& + (1-s_2^2)V_k(x, K-x-1, z) + r, V_k(x, K-x, z)] \\
& + [\mu + (M-z)(\lambda_1 - \lambda_2)]V_k(x, K-x, z) - Kh\}, \\
& 0 < x < K,\ 0 < z < M
\end{aligned}
\tag{19}
$$

$$\begin{aligned} V_{k+1}(x,y,z) =& \frac{1}{v}\{\mu \max[p_3(V_k(x,y,z)-c)+p_1V_k(x+1,y,z)+p_2V_k(x,y\\ &+1,z),V_k(x,y,z)]\\ &+z\lambda_1\max[s_1^1V_k(x-1,y,z)+(1-s_1^1)V_k(x-1,y,z-1)\\ &+r,s_2^1V_k(x,y-1,z)+(1-s_2^1)V_k(x,y-1,z-1)+r]\\ &+(M-z)\lambda_2\max[s_1^2V_k(x-1,y,z+1)\\ &+(1-s_1^2)V_k(x-1,y,z)+r,s_2^2V_k(x,y-1,z+1)\\ &+(1-s_2^2)V_k(x,y-1,z)+r,V_k(x,y,z)]\\ &+(M-z)(\lambda_1-\lambda_2)V_k(x,y,z)-(x+y)h\}, \qquad (20)\\ &\ 0<x<K,\ 0<y<K,\ 0<z<M \end{aligned}$$

for $k = 1, 2, 3, \ldots$ and $(x, y, z) \in S$, and initial values $V_0(x, y, z) \equiv 0$ for every $(x, y, z) \in S$. The optimal long-run average profit function can be approximated numerically by the value iteration algorithm as stated in the following proposition (Puterman 1994).

Proposition 1 Suppose that every average optimal stationary deterministic policy has an aperiodic transition matrix, then the long-run average profit rate J^* is given by

$$J^* = v\lim_{k\to\infty}[V_k(x,y,z)-V_{k-1}(x,y,z)]$$

for every $(x, y, z) \in S$ and for any $V_0(x, y, z)$.

A state is aperiodic if returns to this state can occur at irregular times. In our system the transition matrix of every policy is aperiodic because the probability that the system remains in the current state is positive regardless of the decision made.

4 Numerical Experiments

In order to investigate the structure of the optimal policy we have performed a number of numerical experiments applying the value iteration algorithm in the Eqs. (1)–(20). The test case uses the following parameter values: $M = 15$, $\lambda_1 = 1$, $\lambda_2 = 0.1$, $\mu = 6$, $r = 1$, $c = 2$ and $h = 0.2$. The outgoing quality and satisfaction probabilities, used in the numerical experiments, are presented in Table 1.

Table 1 Production and satisfaction probabilities for the test case examined

Quality level i:	1	2
p_i	0.682689	0.271810
s_i^1	0.619317	0.195921
s_i^2	0.577068	0.145328

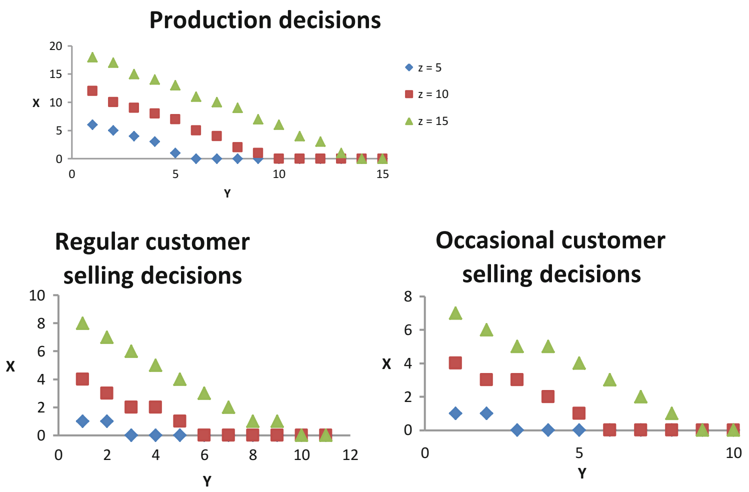


Fig. 1 Optimal policies for three different levels of regular customers

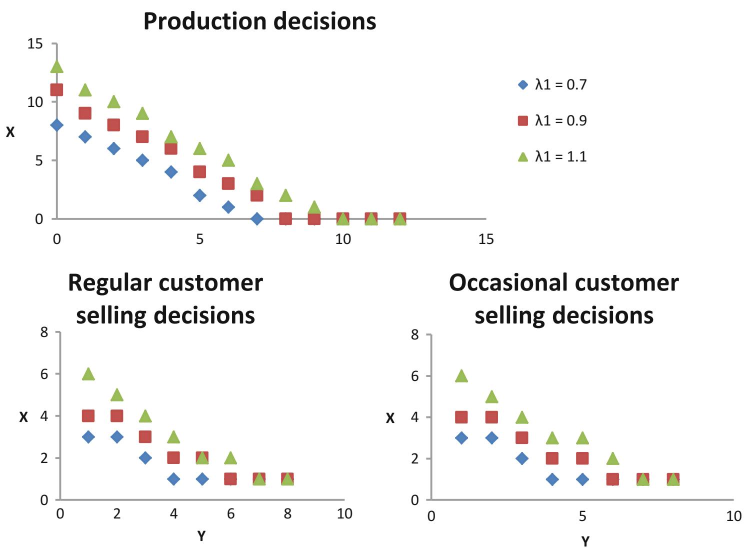


Fig. 2 Optimal policies for different levels of demand rate λ_1, when $z = 10$

In Fig. 1 we present the optimal policies for this system when the number of regular customers, z is 5, 10 or 15. Due to the state space complexity and in order to have a better presentation we have chosen to present the optimal policies for certain characteristic values of z. The marker points lie on the borderlines separating the optimal production decisions 1 (continue) and 0 (stop producing) and the optimal customer selling decisions for quality 1 or quality 2 items. Production decisions

1 are optimal for small values of x and y and therefore they are chosen for points (x, y) in the lower left half planes of the first scatter plot of Fig. 1. For both regular and occasional customers, it is optimal to sell high quality items when there is enough stock and medium quality items when the inventory level is low. Therefore $\pi_{\lambda 1}(x, y, z) = 1$ and $\pi_{\lambda 2}(x, y, z) = 1$ are optimal for points (x,y) in the upper right half planes defined by the borderlines shown in the last two scatter plots of Fig. 1. Rejecting an occasional customer, ie, the decision $\pi_{\lambda 2}(x, y, z) = 0$, is never optimal for the parameters given above. Of course, in state $x = y = 0$ in which there is no stock at all, the customer order is rejected by default.

These results are quite reasonable. As the number of regular customers increases so does the total demand rate and, therefore, a higher inventory is necessary to satisfy the incoming orders. Furthermore, in such cases, it may be more profitable to sell products of lower quality ($i = 2$) more frequently than when the demand rate is low. It is worth noting that, the value iteration algorithm returns $J^* = 1.6891$ as the optimal mean profit rate for the initial test case.

Next, we examine how sensitive the optimal policies are to changes in the demand rate λ_1 of regular customers. In Fig. 2 we present the optimal decisions returned when the number of regular customers is $z = 10$, and for three different values of λ_1. The results are again quite reasonable, since a higher demand rate calls for higher inventory levels and, since the outgoing quality probabilities and the system production rate are both fixed, this is only possible by increasing the amount of items in stock, thus deciding $\pi(x, y, z) = 1$ more frequently, and relaxing the quality requirements and making decisions $\pi_{\lambda 1}(x, y, z) = 2$ and $\pi_{\lambda 2}(x, y, z) = 2$ also more frequently.

5 Conclusions

We have developed a model of a make to stock production system which incorporates customer satisfaction and market dynamics into production and quality control decisions. The model can serve as a tool for investigating optimal policies for production systems. Moreover, it highlights the interactions between production, sales and quality, by showing the effects of quality in customer satisfaction and therefore in system profitability.

References

Anderson EW, Sullivan M (1993) The antecedents and consequences of customer satisfaction for firms. Mark Sci 12(2):125–43

Anderson EW, Fornell C, Lehmann DR (1994) Customer satisfaction, market share and profitability. J Mark 58(2):53–66

Bertsekas D (1995) Dynamic programming and optimal control, vol 2. Athena Scientific, Belmont, MA

Deng T, Shen Z-JM, Shanthikumar JG (2014) Statistical learning of service-dependent demand in a multiperiod newsvendor setting. Oper Res 62(5):1064–1076

Gershwin S, Schick I (2007) A taxonomy of quality/quantity issues in manufacturing systems. In: 6th conference on the analysis of manufacturing systems. Lunteren, The Netherlands

Kostami V, Rajagopalan S (2014) Speed-quality trade-offs in a dynamic model. Manuf Serv Oper Manag 16(1):104–118

Kouikoglou VS, Phillis YA (2002) Design of product specifications and control policies in a single-stage production system. IIE Trans 34:590–600

Oliver R (1997) Satisfaction: a behavioral perspective on the customer. McGraw-Hill, New York, NY

Puterman ML (1994) Markov decision processes: discrete stochastic dynamic programming. Wiley, New York, NY

Rust RT, Zahorik AJ (1993) Customer satisfaction, customer retention, and market share. J Retail 57(3):25–48

Sennott LI (1999) Stochastin dynamic programming and the control of queueing systems. Wiley, New York, NY

Two-Warehouse Inventory Systems for Seasonal Deteriorating Products with Permissible Delay in Payments

Iris-Pandora Krommyda, Konstantina Skouri, Ioannis Konstantaras, and Ioannis Ganas

Abstract In this paper two inventory systems are presented assuming general ramp type demand rate, constant deterioration rate and partial backlogging of unsatisfied demand. Since the capacity of the Owned Warehouse (OW) is usually limited and, under some special circumstances, the procurement of a large amount of items can be decided, a Rented Warehouse (RW) can be used to store the excess quantity. In addition, we assume, that the supplier offers the retailer a credit scheme, which provides a fixed delay period for settling his account. Sales revenue generated during this credit period is deposited to an interest bearing account. At the end of this period the retailer settles the account. Thereafter, capital opportunity cost for the value of items still in stock is charged. The study of the inventory system requires exploring the feasible ordering relations between the time parameters. Consequently the following cases, which lead to two inventory models, must be examined: (1) the offered credit period is less than the time point when the demand rate is stabilized, (2) the offered credit period is longer than the time point when the demand rate is stabilized and less than the planning horizon. The single warehouse inventory problem is also examined as a special case of the model. The optimal

I.-P. Krommyda (✉)
Department of Business Administration of Food and Agricultural Enterprises, University of Patras, Agrinio, Greece
e-mail: ikrommyd@cc.uoi.gr

K. Skouri
Department of Mathematics, University of Ioannina, Ioannina, Greece
e-mail: kskouri@uoi.gr

I. Konstantaras
Department of Business Administration, School of Business Administration, University of Macedonia, Thessaloniki, Greece
e-mail: ikonst@uom.gr

I. Ganas
Department of Accounting, Technological Educational Institute of Epirus, Preveza, Greece
e-mail: ganas@teiep.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_13

replenishment policy for each model is determined. The results obtained are highlighted by suitably selected examples.

Keywords Ramp type demand rate • Deterioration • Partial backlogging • Delay in payments • Two warehouses

1 Introduction

There are types of products, like seasonable products, whose demand increases with time up to a certain moment and then stabilizes to a constant rate. The term "ramp type" is used to represent such a demand pattern. Hill (1995) proposed an inventory model with increasing demand (general power of time) followed by a constant demand. Research on this field continues with Mandal and Pal (1998), Wu et al. (1999), Wu and Ouyang (2000) and Wu (2001), who studied inventory models with linearly increasing, up to its stabilization point, demand rate under various assumptions. Deng et al. (2007) reconsidered the inventory models of Mandal and Pal (1998) and Wu and Ouyang (2000). Skouri et al. (2009) extend the work of Deng et al. (2007), by (i) introducing a general ramp type demand rate, with its variable part being any positive function of time, (ii) considering Weibull distributed deterioration rate, and (iii) allowing backlogging rate to be any decreasing function of the waiting time up to the next replenishment.

All the above models were developed for a single warehouse. However, when the item under consideration is a seasonal product, the buyer may purchase more goods than can be stored in its own warehouse. Therefore, these excess quantities are stored in a rented warehouse RW. Agrawal and Banerjee (2011), developed an inventory model for both—the two-warehouse and the single-warehouse system when demand rate is a general ramp-type function of time. Shortages are allowed and a constant fraction of them is assumed to be backlogged. Agrawal et al. (2013) extended the work of Agrawal and Banerjee (2011), by allowing deterioration of stored items. Skouri and Konstantaras (2013), studied an order level two-warehouse inventory model for deteriorating seasonal products with a general ramp type demand rate, where shortages are allowed and partially backlogged at a rate, which is any non-increasing function of the waiting time up to the next replenishment.

In the traditional inventory models, it was tacitly assumed that the purchaser pays for the items ordered as soon as the items are received in his warehouse. However, in practice, the supplier offers a certain fixed credit period to settle the account, for stimulating demand. Revenues obtained from sales during the trade credit period are invested and earn interest. At the end of the credit period the retailer is charged (at some interest) for the items still in stock. Goyal (1985), was the first who developed an inventory model for determining the economic ordering quantity in the case that the supplier offers to retailer a fixed credit period. Darzanou and Skouri (2011) extended the inventory system of Skouri et al. (2009) assuming constant deterioration rate and a two-level trade credit scheme. Skouri et al. (2011) proposed an inventory model with general ramp type demand rate, constant deterioration rate, partial backlogging of unsatisfied demand and conditions of permissible delay in payments.

In this paper we extend the work of Skouri and Konstantaras (2013), to allow for delay in payments. The following cases, which lead to two different models, must be considered: the offered credit period is (i) less than the time point when the demand rate is stabilized and (ii) longer than the time point when the demand rate is stabilized and less than the planning horizon. Note that the case where the offered credit period is longer than the planning horizon could also be examined, however this case is not considered realistic. The paper is organized as follows: The notation and assumptions used are given in Sect. 2. In Sects. 3 and 4 the model formulation is presented for the two warehouses and single warehouse inventory problems, respectively. The optimal solutions to the problems are derived in Sects. 5 and 6. A solution procedure is proposed in Sect. 7. Numerical examples highlighting the results obtained are given in Sect. 8 and a sensitivity analysis on the basic parameters of the problem is conducted in Sect. 9. The paper closes with concluding remarks in Sect. 10.

2 Notation and Assumptions

2.1 Notation

T: constant scheduling period (cycle)	c_{1O}: holding cost per unit per time unit in OW
t_1: time at which the inventory level reaches zero in OW	c_{1R}: holding cost per unit per time unit in RW
x_1: time at which the inventory level reaches zero in RW	c_2: shortage cost per unit per time unit
μ: parameter of the ramp type demand function	c_3: deterioration cost
M: offered credit period	c_4: opportunity cost due to the lost sales
θ_1: deterioration rate in RW	c_t: transportation cost
θ_2: deterioration rate in OW	P: unit selling price
W: capacity of owned warehouse (OW)	I_e: interest rate earned
c_p: unit purchase cost	I_c: interest rate charged

2.2 Assumptions

The inventory model is developed under the following assumptions:

- The system operates for a prescribed period of T units of time (planning horizon), the replenishment rate is infinite and lead time is zero.

- Shortages are backlogged at a rate $\beta(x)$ which is a non-increasing function of x with $0 < \beta(x) \le 1$, $\beta(0) = 1$ and x is the waiting time up to the next replenishment. Moreover it is assumed that $\beta(x)$ satisfies the relation $\beta(x) + T\beta'(x) \ge 0$, where $\beta'(x)$ is the derivate of $\beta(x)$. The cases with $\beta(x) = 1$ correspond to complete backlogging model.
- The supplier offers the retailer a credit scheme, which provides a fixed delay period. Sales revenue generated during this credit period is deposited to an interest bearing account. At the end of this period the retailer settles his account. Thereafter he is charged interest (capital opportunity cost) for the value of the items still in stock (Goyal (1985)).
- The on hand inventory deteriorates at a constant rate θ_1 in RW and at θ_2 in OW, where $0 < \theta_2 < \theta_1 < 1$. The deteriorated items are withdrawn immediately from the warehouses and there is no provision for repair or replacement.
- The holding costs in RW are higher than those in OW, i.e., $c_{1R} > c_{1O}$, hence it is cost-effective to consume the goods of the RW first.
- The owned warehouse (OW) has a fixed capacity of W units and the rented warehouse (RW) has unlimited capacity.
- The demand rate $D(t)$ is a ramp type function of time given by:

$$D(t) = \begin{cases} f(t), & t < \mu \\ f(\mu), & t \ge \mu \end{cases} \tag{1}$$

where $f(t)$ is a positive, continuous and increasing function of $t \in (0, \ T]$.

3 Model Formulation for the Two Warehouse Problem

During the interval $(0, x_1)$, the inventory level in RW decreases due to demand and deterioration and drops to zero at $t = x_1$. In OW, the inventory level decreases during $(0, x_1)$ due to deterioration only, but during $(x_1, \ t_1)$ the inventory is depleted due to both demand and deterioration. By the time t_1 both warehouses are empty and thereafter shortages are allowed to occur. The shortage quantity is supplied to customers at the beginning of the next cycle. For the development of the model the following three cases: (i) $x_1 < t_1 \le \mu$, (ii) $x_1 \le \mu < t_1$ and (iii) $\mu \le x_1 < t_1$, must be considered. The inventory levels and the corresponding costs for these three cases, without taking into consideration the delay in payments, are derived in Skouri and Konstantaras (2013) and are presented below.

3.1 Cost Calculation Excluding Interest Earned and Charged

The total inventory cost (holding cost, deterioration cost, shortage cost, opportunity cost due to lost sales and transportation cost) during the time interval [0, T], for the case $x_1 < t_1 \leq \mu$, is given by:

$$\begin{aligned} A(t_1,x_1) \;=\; & \left(\frac{c_{1R}+c_3\theta_1}{\theta_1}\right)\left[\int_0^{x_1}\left(e^{\theta_1 x}-1\right)f(t)dt\right] \\ & +\left(\frac{c_{1O}+c_3\theta_2}{\theta_2}\right)\int_{x_1}^{t_1}\left(e^{\theta_2 t}-1\right)f(t)dt \\ & +c_2\left[\int_{t_1}^{\mu} f(t)\beta(T-t)(-t)dt+f(\mu)\int_{\mu}^{T}\beta(T-t)(T-t)\,dt\right] \\ & +c_4\left[\int_{t_1}^{\mu}(1-\beta(T-t))f(t)\,dt+f(\mu)\int_{\mu}^{T}(1-\beta(T-t))\,dt\right] \\ & +c_t\int_0^{x_1} f(t)dt. \end{aligned} \tag{2}$$

The total inventory cost during the time interval [0, T], for the case $x_1 \leq \mu < t_1$, is given by:

$$\begin{aligned} B(t_1,x_1)= & \left(\frac{c_{1R}+c_3\theta_1}{\theta_1}\right)\left[\int_0^{x_1}\left(e^{\theta_1 t}-1\right)f(t)dt\right] \\ & +\left(\frac{c_{1O}+c_3\theta_2}{\theta_2}\right)\left[\int_{x_1}^{\mu}\left(e^{\theta_2 t}-1\right)f(t)dt+f(\mu)\int_{\mu}^{t_1}\left(e^{\theta_2 t}-1\right)dt\right] \\ & +c_2 f(\mu)\int_{t_1}^{T}(T-t)\times\beta(T-t)dt+c_4 f(\mu)\int_{t_1}^{T}(1-\beta(T-t))dt+c_t\int_0^{x_1} f(t)dt. \end{aligned} \tag{3}$$

The total inventory cost during the time interval [0, T], for case $\mu \leq x_1 < t_1$, is given by:

$$\begin{aligned} \varGamma(t_1,x_1) \;=\; & \left(\frac{c_{1R}+c_3\theta_1}{\theta_1}\right)\left[\int_0^{\mu}\left(e^{\theta_1 t}-1\right)f(t)dt+f(\mu)\int_{\mu}^{x_1}\left(e^{\theta_1 t}-1\right)dt\right] \\ & +\left(\frac{c_{1O}+c_3\theta_2}{\theta_2}\right)f(\mu)\int_{x_1}^{t_1}\left(e^{\theta_2 t}-1\right)dt+c_2 f(\mu)\int_{t_1}^{T}(T-t)\beta(T-t)dt \\ & +c_4 f(\mu)\int_{t_1}^{T}(1-\beta(T-t))\,dt+c_t\left[\int_0^{\mu} f(t)dt+\int_{\mu}^{x_1} f(\mu)dt\right]. \end{aligned} \tag{4}$$

To proceed and formulate the objective function for each model, the ordering relations between μ, M and T must be taken into account. So we are led to consider the two cases $M \leq \mu$ and $\mu < M \leq T$.

3.2 The Inventory Model for the Case $M \leq \mu$

To build up the cost function for this case (credit period M, is less than the demand stabilizing time point μ), the ordering relations between x_1, t_1, μ, M *and* T must be considered. The total cost in the interval $[0, \ T]$, for the sub-cases that arise, is given as:

$$TC_1(t_1, x_1) = \begin{cases} TC_{1.A}(t_1, x_1), x_1 < t_1 \leq \mu \\ TC_{1.B}(t_1, x_1), x_1 \leq \mu < t_1 \\ TC_{1.\Gamma}(t_1, x_1), \mu \leq x_1 < t_1 \end{cases}, \tag{5}$$

where

$$TC_{1.A}(t_1, x_1) = \begin{cases} A(t_1, x_1) - I_{P_1} + P_{T_1}, \ x_1 < t_1 < M < \mu < T \\ A(t_1, x_1) - I_{P_2} + P_{T_2}, \ x_1 < M < t_1 < \mu < T \\ A(t_1, x_1) - I_{P_3} + P_{T_3}, \ M < x_1 < t_1 < \mu < T \end{cases}, \tag{6}$$

$$TC_{1.B}(t_1, x_1) = \begin{cases} B(t_1, x_1) - I_{P_4} + P_{T_4}, \ x_1 < M < \mu < t_1 < T \\ B(t_1, x_1) - I_{P_5} + P_{T_5}, \ M < x_1 < \mu < t_1 < T \end{cases}, \tag{7}$$

$$TC_{1.\Gamma}(t_1, x_1) = \Gamma(t_1, x_1) - I_{P_6} + P_{T_6}, \ M < \mu < x_1 < t_1 < T, \tag{8}$$

I_{P_i} and P_{T_i} represent the interest earned and interest charged for each case respectively and are defined as follows.

Interest earned:

$$I_{P_1} = pI_e\left(\int_0^{t_1}\int_0^t f(x)dxdt + (M - t_1)\int_0^{t_1} f(x)dx\right)$$

$$I_{P_i} = pI_e\int_0^M\int_0^t f(x)dxdt, \quad i = 2, 3, 4, 5, 6.$$

Interest charged:

$$P_{T_1} = 0,$$

$$P_{T_2} = \frac{C_p I_c}{\theta_2}\int_M^{t_1}\left(e^{\theta_2(t-M)} - 1\right)f(t)dt,$$

$$P_{T_3} = \frac{C_p I_c}{\theta_1} \int_M^{x_1} \left(e^{\theta_1(t-M)} - 1\right) f(t)dt + \frac{C_p I_c}{\theta_2} \int_{x_1}^{t_1} \left(e^{\theta_2(t-M)} - 1\right) f(x)dx,$$

$$P_{T_4} = \frac{C_p I_c}{\theta_2} \left(\int_M^{\mu} \left(e^{\theta_2(t-M)} - 1\right) f(t)dt + \int_{\mu}^{t_1} \left(e^{\theta_2(t-M)} - 1\right) f(\mu)dt \right),$$

$$P_{T_5} = \frac{C_p I_c}{\theta_1} \int_M^{x_1} \left(e^{\theta_1(t-M)} - 1\right) f(t)dt + \frac{C_p I_c}{\theta_2} \left(\int_{x_1}^{\mu} \left(e^{\theta_2(t-M)} - 1\right) f(t)dt + \int_{\mu}^{t_1} \left(e^{\theta_2(t-M)} - 1\right) f(\mu)dt \right),$$

$$P_{T_6} = \frac{C_p I_c}{\theta_1} \left(\int_M^{\mu} \left(e^{\theta_1(t-M)} - 1\right) f(t)dt + \int_{\mu}^{x_1} \left(e^{\theta_1(t-M)} - 1\right) f(\mu)dt \right) + \frac{C_p I_c}{\theta_2} \int_{x_1}^{t_1} \left(e^{\theta_2(t-M)} - 1\right) f(\mu)dt.$$

3.3 *The Inventory Model for the Case $\mu < M < T$*

In this case, the offered credit period M is longer than the demand stabilizing point μ and shorter than the planning horizon T. The total cost in the interval [0, T], for the sub-cases that arise, is given as:

$$TC_2(t_1, x_1) = \begin{cases} TC_{2.A}(t_1, x_1), x_1 < t_1 \leq \mu \\ TC_{2.B}(t_1, x_1), x_1 \leq \mu < t_1, \\ TC_{2.\Gamma}(t_1, x_1), \mu \leq x_1 < t_1 \end{cases} \tag{9}$$

where

$$TC_{2.A}(t_1, x_1) = A(t_1, x_1) - I_{P_1} + P_{T_1},\ x_1 < t_1 < \mu < M < T, \tag{10}$$

$$TC_{2.B}(t_1, x_1) = \begin{cases} (t_1, x_1) - I_{P_2} + P_{T_2},\ x_1 < \mu < t_1 < M < T \\ (t_1, x_1) - I_{P_3} + P_{T_3},\ x_1 < \mu < M < t_1 < T \end{cases}, \tag{11}$$

$$TC_{2.\Gamma}(t_1, x_1) = \begin{cases} \Gamma(t_1, x_1) - I_{P_4} + P_{T_4},\ \mu < x_1 < t_1 < M < T \\ \Gamma(t_1, x_1) - I_{P_5} + P_{T_5},\ \mu < x_1 < M < t_1 < T, \\ \Gamma(t_1, x_1) - I_{P_6} + P_{T_6},\ \mu < M < x_1 < t_1 < T \end{cases} \tag{12}$$

I_{P_i} and P_{T_i} represent the interest earned and interest charged for each case respectively and are defined as follows.

Interest earned:

$$I_{P_1} = pI_e\left(\int_0^{t_1}\int_0^{t} f(x)dxdt + (M - t_1)\int_0^{t_1} f(x)dx\right),$$

$$\begin{aligned} I_{P_2} &= I_{P_4} \\ &= pI_e\left(\int_0^{\mu}\int_0^{t} f(x)\,dx\,dt + \int_{\mu}^{t_1}\int_{\mu}^{t} f(\mu)\,dx\,dt + (M-\mu)\int_0^{\mu} f(x)\,dx\right. \\ &\left.+(M - t_1)\int_{\mu}^{t_1} f(\mu)\,dx\right), \end{aligned}$$

$$I_{P_3} = I_{P_5} = I_{P_6} = pI_e\left(\int_0^{\mu}\int_0^{t} f(x)\,dx\,dt + (M-\mu)\int_0^{\mu} f(x)\,dx + \int_{\mu}^{M}\int_{\mu}^{t} f(\mu)\,dx\,dt\right).$$

Interest charged:

$$P_{T_1} = P_{T_2} = P_{T_4} = 0,$$

$$P_{T_3} = P_{T_5} = \frac{C_p I_c}{\theta_2}\int_M^{t_1}\left(e^{\theta_2(t-M)} - 1\right)f(\mu)dt,$$

$$P_{T_6} = \tfrac{C_p I_c}{\theta_1}\int_M^{x_1}\left(e^{\theta_1(t-M)} - 1\right)f(\mu)dt + \tfrac{C_p I_c}{\theta_2}\int_{x_1}^{t_1}\left(e^{\theta_2(t-M)} - 1\right)f(\mu)dt.$$

4 Model Formulation for the Single Warehouse Problem

Next we examine the single warehouse inventory problem. This is a special case of the two warehouse inventory problem studied above, for $x_1 = 0$. This model has been examined by Skouri et al. (2011), by using a different approach at calculating the interest earned.

4.1 Cost Calculation Excluding Interest Earned and Charged

The total inventory cost during the time interval $[0,\ T]$, for the case $0 < t_1 \le \mu$ is given by:

$$\begin{aligned}A_0(t_1) = &\left(\frac{c_{1O}+c_3\theta_2}{\theta_2}\right)\int_0^{t_1}\left(e^{\theta_2 t}-1\right)f(t)dt\\&+c_2\left[\int_{t_1}^{\mu}f(t)\beta(T-t)(-t)dt+f(\mu)\int_{\mu}^{T}\beta(T-t)(T-t)dt\right]\\&\quad+c_4\left[\int_{t_1}^{\mu}(1-\beta(T-t))f(t)dt+f(\mu)\int_{\mu}^{T}(1-\beta(T-t))dt\right].\end{aligned} \tag{13}$$

The total inventory cost during the time interval [0, T], for the case $0 \le \mu < t_1$ is given by:

$$\begin{aligned}B_o(t_1) = &\left(\frac{c_{1O}+c_3\theta_2}{\theta_2}\right)\left[\int_0^{\mu}\left(e^{\theta_2 t}-1\right)f(t)dt+f(\mu)\int_{\mu}^{t_1}\left(e^{\theta_2 t}-1\right)dt\right]\\&+c_2 f(\mu)\int_{t_1}^{T}(T-t)\beta(T-t)dt+c_4 f(\mu)\int_{t_1}^{T}(1-\beta(T-t))\,dt.\end{aligned} \tag{14}$$

As in the two warehouse model, taking into account the ordering relations between μ, M and T, we are led to consider the two cases $M \le \mu$ and $\mu < M \le T$.

4.2 The Inventory Model for the Case $M \le \mu$

The total cost in the interval [0, T], for the sub-cases that arise, is given as:

$$TC_{O1}(t_1) = \begin{cases} A_O(t_1) - I_{P_1} + P_{T_1}, \ 0 < t_1 < M < \mu < T \\ A_O(t_1) - I_{P_2} + P_{T_2}, \ 0 < M < t_1 < \mu < T\,, \\ B_O(t_1) - I_{P_3} + P_{T_3}, \ 0 < M < \mu < t_1 < T \end{cases} \tag{15}$$

I_{P_i} and P_{T_i} represent the interest earned and interest charged for each case respectively and are defined as follows.

Interest earned:

$$I_{P_1} = pI_e\left(\int_0^{t_1}\int_0^{t} f(x)dxdt + (M-t_1)\int_0^{t_1} f(x)dx\right),$$

$$I_{P_i} = pI_e\int_0^{M}\int_0^{t} f(x)dxdt, \quad i = 2,3.$$

Interest charged:

$$P_{T_1} = 0,$$

$$P_{T_2} = \frac{C_p I_c}{\theta_2} \int_M^{t_1} \left(e^{\theta_2(t-M)} - 1\right) f(t)dt,$$

$$P_{T_3} = \frac{C_p I_c}{\theta_2} \left(\int_M^{\mu} \left(e^{\theta_2(t-M)} - 1\right) f(t)dt + \int_{\mu}^{t_1} \left(e^{\theta_2(t-M)} - 1\right) f(\mu)dt \right).$$

4.3 The Inventory Model for the Case $\mu < M < T$

The total cost in the interval [0, T], for the sub-cases that arise, is given as:

$$TC_{02}(t_1) = \begin{cases} A_O(t_1) - I_{P_1} + P_{T_1}, \; 0 < t_1 < \mu < M < T \\ _O(t_1) - I_{P_2} + P_{T_2}, \; 0 < \mu < t_1 < M < T \\ _O(t_1) - I_{P_3} + P_{T_3}, \; 0 < \mu < M < t_1 < T \end{cases}, \tag{16}$$

I_{P_i} and P_{T_i} represent the interest earned and interest charged for each case respectively and are defined as follows.

Interest earned:

$$I_{P_1} = pI_e \left(\int_0^{t_1} \int_0^t f(x)dxdt + (M - t_1) \int_0^{t_1} f(x)dx \right),$$

$$I_{P_2} = pI_e \left(\int_0^{\mu} \int_0^t f(x)\,dx\,dt + \int_{\mu}^{t_1} \int_{\mu}^t f(\mu)\,dx\,dt + (M - \mu) \int_0^{\mu} f(x)\,dx \right.$$

$$\left. + (M - t_1) \int_{\mu}^{t_1} f(\mu)\,dx \right),$$

$$I_{P_3} = pI_e \left(\int_0^{\mu} \int_0^t f(x)\,dx\,dt + (M - \mu) \int_0^{\mu} f(x)\,dx + \int_{\mu}^{M} \int_{\mu}^t f(\mu)\,dx\,dt \right),$$

Interest charged:

$$P_{T_1} = P_{T_2} = 0,$$

$$P_{T_3} = \frac{C_p I_c}{\theta_2} \int_M^{t_1} \left(e^{\theta_2(t-M)} - 1\right) f(\mu)dt.$$

5 Optimal Solution for the Two Warehouse Inventory Model

5.1 Optimal Solution for the Case $M \leq \mu$

Our aim is to minimize the total cost function $TC_1(x_1, t_1)$ with respect to (x_1, t_1). Hence the three problems stated below must be solved.

$$P_1 : \begin{cases} \min TC_{1.A}(t_1, x_1) \\ s.t. \ \int_{x_1}^{t_1} e^{\theta_2 x} f(x) dx - W = 0 \end{cases},$$

$$P_2 : \begin{cases} \min TC_{1.}(t_1, x_1) \\ s.t. \ \int_{x_1}^{\mu} e^{\theta_2 x} f(x) dx + f(\mu) \int_{\mu}^{t_1} e^{\theta_2 x} dx - W = 0 \end{cases},$$

$$P_3 : \begin{cases} \min TC_{1.\Gamma}(t_1, x_1) \\ s.t. \ f(\mu) \int_{x_1}^{t_1} e^{\theta_2 x} dx - W = 0 \end{cases}.$$

Note that the ordering relations between x_1, t_1, μ, M and T are temporally ignored, however, these relations must be checked for the feasibility of the solution.

The constraints given above are derived by Skouri and Konstantaras (2013), from the continuity of the inventory level. From these equations we can consider x_1 as a function of t_1.

$$\int_{x_1}^{t_1} e^{\theta_2 x} f(x) dx - W = 0 \Leftrightarrow x_1 = h_A(t_1), \tag{17}$$

$$\int_{x_1}^{\mu} e^{\theta_2 x} f(x) dx + f(\mu) \int_{\mu}^{t_1} e^{\theta_2 x} dx - W = 0 \Leftrightarrow x_1 = h_B(t_1), \tag{18}$$

$$f(\mu) \int_{x_1}^{t_1} e^{\theta_2 x} dx - W = 0 \Leftrightarrow x_1 = h_\Gamma(t_1). \tag{19}$$

Consequently, the cost function becomes a function of one variable

$$TC_{1.i}(t_1, x_1) = TC_{1.i}(t_1, x_1 = h_i(t_1)), \quad i = A, B, \Gamma.$$

Next we find the first order conditions for the optimal solution to the problem P_1:

$$\frac{dTC_{1.A.1}}{dt_1} = (b(t_1, x_1 = h_A(t_1)) - pI_e(M - t_1)) f(t_1) = 0, \text{ for } x_1 < t_1 < M < \mu < T, \tag{20}$$

$$\frac{dTC_{1.A.2}}{dt_1} = \left(b(t_1, x_1 = h_A(t_1)) + \frac{C_p I_c}{\theta_2}\left(e^{\theta_2(t_1-M)} - 1\right)\right) f(t_1) \\ = 0, \quad \text{for} \quad x_1 < M < t_1 < \mu < T, \tag{21}$$

$$\frac{dTC_{1.A.3}}{dt_1} \\ = \left(b(t_1, x_1 = h_A(t_1)) + C_p I_c\left[\frac{1}{\theta_1}e^{\theta_2(t_1-x_1)}\left(e^{\theta_1(x_1-M)} - 1\right) + \frac{1}{\theta_2}\left(e^{\theta_2(t_1-x_1)} - 1\right)\right]\right) \\ \times f(t_1) = 0, \quad \text{for} \quad M < x_1 < t_1 < \mu < T. \tag{22}$$

The first order conditions for the optimal solution to the problem P_2:

$$\frac{dTC_{1.B.1}}{dt_1} = \left(b(t_1, x_1 = h_B(t_1)) + \frac{C_p I_c}{\theta_2}\left(e^{\theta_2(t_1-M)} - 1\right)\right) f(\mu) \\ = 0, \quad \text{for} \quad x_1 < M < \mu < t_1 < T \tag{23}$$

$$\frac{dTC_{1.B.2}}{dt_1} \\ = \left(b(t_1, x_1 = h_B(t_1)) + C_p I_c\left[\frac{1}{\theta_1}e^{\theta_2(t_1-x_1)}\left(e^{\theta_1(x_1-M)} - 1\right) + \frac{1}{\theta_2}\left(e^{\theta_1(t_1-x_1)} - 1\right)\right]\right) \\ \times f(\mu) = 0, \quad \text{for} \quad M < x_1 < \mu < t_1 < T. \tag{24}$$

The first order conditions for the optimal solution to the problem P_3:

$$\frac{dTC_{1.\Gamma}}{dt_1} \\ = \left(b(t_1, x_1 = h_\Gamma(t_1)) + C_p I_c\left[\frac{1}{\theta_1}e^{\theta_2(t_1-x_1)}\left(e^{\theta_1(x_1-M)} - 1\right) + \frac{1}{\theta_2}\left(e^{\theta_1(t_1-x_1)} - 1\right)\right]\right) \\ \times f(\mu) = 0, \quad \text{for} \quad M < \mu < x_1 < t_1 < T, \tag{25}$$

where

$$b\left(t_1, x_1 = h_i(t_1)\right) \\ = \frac{c_{1R} + c_3\theta_1}{\theta_1}e^{\theta_2(t_1-x_1)}(e^{\theta_1 x_1} - 1) + \frac{c_{1O} + c_3\theta_2}{\theta_2}(e^{\theta_2(t_1-x_1)} - 1) - c_2(T-t_1)\beta(T-t_1) \\ - c_4\left(1 - \beta(T-t_1)\right) + c_t e^{\theta_2(t_1-x_1)}, \; i = A, B, \Gamma.$$

Theorem 1
If $\frac{c_{1R}}{\theta_1} > \frac{c_{1O}}{\theta_2} + c_t$ and $c_2 T\beta(T) + c_4(1 - \beta(T)) \geq c_t$, then the cost function $TC_1(t_1, x_1)$ has a unique minimum.

Proof. See Appendix 1.

5.2 Optimal Solution for the Case $\mu < M < T$

Our aim is to minimize the total cost function $TC_2(x_1, t_1)$ with respect to (x_1, t_1). Hence we must solve the three problems stated below.

$$P_4 : \begin{cases} \min TC_{2.\mathrm{A}}(t_1, x_1) \\ s.t. \displaystyle\int_{x_1}^{t_1} e^{\theta_2 x} f(x)dx - W = 0 \end{cases},$$

$$P_5 : \begin{cases} \min TC_{2.}(t_1, x_1) \\ s.t. \displaystyle\int_{x_1}^{\mu} e^{\theta_2 x} f(x)dx + f(\mu)\int_{\mu}^{t_1} e^{\theta_2 x}dx - W = 0 \end{cases},$$

$$P_6 : \begin{cases} \min TC_{2.\Gamma}(t_1, x_1) \\ s.t.\ f(\mu)\displaystyle\int_{x_1}^{t_1} e^{\theta_2 x}dx - W = 0 \end{cases}.$$

Note that the ordering relations between x_1, t_1, μ, M and T are temporally ignored but they will be checked later for the feasibility of the solution.

Again, from the above constraints, x_1 is an increasing function of t_1. Consequently, the cost function becomes a function of one variable $TC_{2.i}(t_1, x_1) = TC_{2.i}(t_1, x_1 = h_i(t_1)),\ i = \mathrm{A}, , \Gamma$.

Next we find the first order conditions for the optimal solution to the problem P_4:

$$\frac{dTC_{2.A}}{dt_1} = (b(t_1, x_1 = h_A(t_1)) - pI_e(M - t_1))f(t_1) = 0, \text{ for } x_1 < t_1 < \mu < M < T. \quad (26)$$

The first order conditions for the optimal solution to the problem P_5:

$$\frac{dTC_{2.B.1}}{dt_1} = (b(t_1, x_1 = h_B(t_1)) - pI_e(M - t_1))f(\mu) = 0, \text{ for } x_1 < \mu < t_1 < M < T, \quad (27)$$

$$\begin{aligned} \frac{dTC_{2.B.2}}{dt_1} &= \left(b(t_1, x_1 = h_B(t_1)) + \frac{C_p I_c}{\theta_2}\left(e^{\theta_2(t_1 - M)} - 1\right)\right) f(\mu) \\ &= 0, \quad \text{for} \quad x_1 < \mu < M < t_1 < T\ . \end{aligned} \quad (28)$$

The first order conditions for the optimal solution to the problem P_6:

$$\frac{dTC_{2.\Gamma.1}}{dt_1} = (b(t_1, x_1 = h_\Gamma(t_1)) - pI_e(M - t_1))f(\mu) = 0, \text{ for } \mu < x_1 < t_1 < M < T, \quad (29)$$

$$\begin{aligned}\frac{dTC_{2.\Gamma.2}}{dt_1} &= \left(b(t_1, x_1 = h_\Gamma(t_1)) + \frac{C_p I_c}{\theta_2}\left(e^{\theta_2(t_1 - M)} - 1\right)\right)f(\mu) \\ &= 0, \quad \text{for} \quad \mu < x_1 < M < t_1 < T,\end{aligned} \quad (30)$$

$$\begin{aligned}&\frac{dTC_{2.\Gamma.3}}{dt_1} \\ &= \left(b(t_1, x_1 = h_\Gamma(t_1)) + C_p I_c\left[\frac{1}{\theta_1}e^{\theta_2(t_1 - x_1)}\left(e^{\theta_1(x_1 - M)} - 1\right) + \frac{1}{\theta_2}\left(e^{\theta_2(t_1 - x_1)} - 1\right)\right]\right) \\ &\times f(\mu) = 0, \quad \text{for} \quad \mu < M < x_1 < t_1 < T.\end{aligned} \quad (31)$$

Theorem 2
If $\frac{c_{1R}}{\theta_1} > \frac{c_{1O}}{\theta_2} + c_t$ and $c_2 T\beta(T) + c_4(1 - \beta(T)) \geq c_t$, then the cost function $TC_2(t_1, x_1)$ has a unique minimum.

Proof. Similar to the proof of Theorem 1.

Remark 1
From (17) by setting $x_1 = 0$, a unique value for t_1, say t_{10}, is obtained. Obviously, the domain of t_1 is the interval $[t_{10}, T]$ and any cases that do not satisfy this constraint must not be examined (for example, if $t_{10} > \mu$ then the problems P_1 and P_4 can be excluded from the solution procedure). Also, if by setting $x_1 = \mu$ in relation (19), the value that is obtained, say $t_{1\mu}$, is $t_{1\mu} \geq T$ then the problems P_3 and P_6 can be excluded from the solution procedure.

Remark 2
It is easy to prove, that the cost function is a continuously differentiable function with respect to t_1. Hence, the optimal solution is obtained by the first order condition that provides a feasible solution. If no feasible solution can be obtained the boundary of the interval $[t_{10}, T]$ must be examined, where t_{10} is obtained by setting $x_1 = 0$ in relation (17).

Remark 3
The optimal ordering quantity is obtained by (see Skouri and Konstantaras 2013)

$$\begin{aligned}Q^* = \int_0^{x_1} e^{\theta_1 x} f(x)dx + \int_{x_1}^{t_1} e^{\theta_2 x} f(x)dx + f(\mu)\int_\mu^T \beta(T - x)dx + \int_{t_1}^{\mu} f(x)\beta(T - x)dx,\\ \text{for} \quad x_1 < t_1 \leq \mu,\end{aligned} \quad (32)$$

$$Q^* = \int_0^{x_1} e^{\theta_1 x} f(x)dx + \int_{x_1}^{\mu} e^{\theta_2 x} f(x)dx + f(\mu)\int_{\mu}^{t_1} e^{\theta_2 x}dx + f(\mu)\int_{t_1}^{T} \beta(T-x)dx,$$

$$\text{for} \quad x_1 \leq \mu < t_1, \tag{33}$$

$$Q^* = \int_0^{\mu} e^{\theta_1 x} f(x)dx + f(\mu)\int_{\mu}^{x_1} e^{\theta_1 x}dx + f(\mu)\int_{x_1}^{t_1} e^{\theta_2 x}dx + f(\mu)\int_{t_1}^{T} \beta(T-x)dx,$$

$$\text{for} \quad \mu \leq x_1 < t_1. \tag{34}$$

6 Optimal Solution for the Single Warehouse Inventory Model

6.1 Optimal Solution for the Case $M \leq \mu$

In this section the minimization of $TC_{O1}(t_1)$ with respect to t_1 will be presented. The first order conditions for the optimal solution to the problem are:

$$\frac{dTC_{O.1.1}}{dt_1} = (b_0(t_1) - pI_e(M - t_1))f(t_1) = 0, \quad \text{for} \quad 0 < t_1 < M < \mu < T, \tag{35}$$

$$\frac{dTC_{0.1.2}}{dt_1} = \left(b_0(t_1) + \frac{C_p I_c}{\theta_2}\left(e^{\theta_2(t_1 - M)} - 1\right)\right)f(t_1) = 0, \text{ for } 0 < M < t_1 < \mu < T, \tag{36}$$

$$\frac{dTC_{O.1.3}}{dt_1} = \left(b_0(t_1) + \frac{C_p I_c}{\theta_2}\left(e^{\theta_2(t_1 - M)} - 1\right)\right)f(\mu) = 0, \text{ for } 0 < M < \mu < t_1 < T, \tag{37}$$

where $b_0(t_1) = \frac{c_{1O} + c_3\theta_2}{\theta_2}\left(e^{\theta_2 t_1} - 1\right) - c_2(T - t_1)\beta(T - t_1) - c_4(1 - \beta(T - t_1))$.

Note that the ordering relations between t_1, μ, M and T must be checked for the feasibility of the solution.

Theorem 3
The cost function $TC_{O1}(t_1)$ has a unique minimum.

Proof. See Appendix 2.

6.2 Optimal Solution for the Case $\mu < M < T$

Our aim is to minimize the total cost function $TC_{O2}(t_1)$ with respect to t_1. The first order conditions for the optimal solution to the problem are:

$$\frac{dTC_{0.2.1}}{dt_1} = (b_o(t_1) - pI_e(M - t_1))f(t_1) = 0, \quad \text{for} \quad 0 < t_1 < \mu < M < T, \tag{38}$$

$$\frac{dTC_{O.2.2}}{dt_1} = (b_O(t_1) - pI_e(M - t_1))f(\mu) = 0, \quad \text{for} \quad 0 < \mu < t_1 < M < T, \tag{39}$$

$$\frac{dTC_{0.2.3}}{dt_1} = \left(b_0(t_1) + \frac{C_p I_c}{\theta_2}\left(e^{\theta_2(t_1 - M)} - 1\right)\right)f(\mu) = 0, \ \text{for} \ 0 < \mu < M < t_1 < T. \tag{40}$$

Note that the ordering relations between t_1, μ, M *and* T must be checked for the feasibility of the solution.

Theorem 4
The cost function $TC_{O2}(t_1)$ has a unique minimum.

Proof. Similar to the proof of Theorem 3.

Remark 4
The optimal ordering quantity for the single warehouse problem is obtained by (see Skouri et al. (2011)):

$$Q^* = \int_0^{t_1} e^{\theta_2 x} f(x)dx + \int_{t_1}^{\mu} f(x)\beta(T - x)dx + f(\mu)\int_{\mu}^{T} \beta(T - x)dx, \quad \text{for} \quad 0 < t_1 \leq \mu, \tag{41}$$

$$Q^* = \int_0^{\mu} e^{\theta_2 x} f(x)dx + f(\mu)\int_{\mu}^{t_1} e^{\theta_2 x}dx + f(\mu)\int_{t_1}^{T} \beta(T - x)dx, \quad \text{for} \quad 0 \leq \mu < t_1. \tag{42}$$

The optimal maximum inventory level at OW is obtained by (see Skouri et al. (2011)):

$$I_O(0)^* = \int_0^{t_1} e^{\theta_2 x} f(x)dx, \quad \text{for} \quad 0 < t_1 \leq \mu, \tag{43}$$

$$I_O(0)^* = \int_0^{\mu} e^{\theta_2 x} f(x)dx + f(\mu)\int_{\mu}^{t_1} e^{\theta_2 x} dx, \quad \text{for} \quad 0 \le \mu < t_1. \tag{44}$$

7 Solution Procedure

- First solve the single warehouse problem by using Eqs. (35)–(37) for the inventory problem corresponding to the case $M < \mu$ and Eqs. (38)–(40) for the inventory problem corresponding to the case $\mu < M < T$. The first order condition that gives a feasible solution is the optimal solution to the inventory problem , t_1^*, and the rest do not have to be examined.
- Calculate the optimal maximum inventory level using the appropriate relation from (43) and (44). If $I_0^*(0) \le W$, then the optimal solution has been obtained and the two warehouse problem does not have to be examined. Otherwise we study the two warehouse problem as follows.
- Set $x_1 = 0$ in relation (17) and obtain t_{10}. Set $x_1 = \mu$ in relation (19) and obtain $t_{1\mu}$.
- Solve the first order conditions using Eqs. (20)–(25) for the inventory problem corresponding to the case $M < \mu$ and Eqs. (26)–(31) for the inventory problem corresponding to the case $\mu < M < T$. Cases that do not satisfy the constraint $t_{10} \le t_1 \le \min\{T, t_{1\mu}\}$ are not examined.
- The first order condition that gives a feasible solution, is the optimal solution to the inventory problem, t_1^*, and the rest do not have to be examined.
- Obtain the optimal value for x_1, say x_1^*, using the appropriate relation from (17) to (19).
- Calculate the optimal ordering quantity by using the appropriate relation from (32) to (34).
- Obtain the optimal total cost TC^*, by using the appropriate relation from (6) to (8) for the inventory problem corresponding to the case $M < \mu$ and from relations (10)–(12) for the inventory problem corresponding to the case $\mu < M < T$.

8 Numerical Examples

The following examples illustrate the results obtained.

8.1 Example 1 ($M < \mu$)

For reasons of comparison we use the parameter values of Skouri and Konstantaras (2013), excluding the capital cost from the holding costs in the OW and RW: $c_{1R} = 3$, $c_{1O} = 2$, $c_2 = 15$, $c_3 = 5$, $c_4 = 20$, $c_t = 1.5$, $\mu = 0.9$ years, $\theta_1 = 0.002$, $\theta_2 = 0.001$, $T = 1$ years, $W = 20$, $f(t) = 3e^{4.5t}$, $\beta(x) = e^{-0.2x}$, $M = 0.4$, $p = 30$, $C_p = 10$, $I_e = 0.09$ and $I_c = 0.1$.

By solving the first order equations, the only feasible solution is derived from Eq. (21) and corresponds to the case $x_1 < M < t_1 < \mu < T$. Hence, the optimal solution to the problem is $\left(t_1{}^*, x_1{}^*\right) = (0.79, 0.36)$, where $x_1{}^*$ is derived from relation (17) . The minimum cost $TC_{1.A.2} = 92.68$ is derived from Eq. (6) and the optimal order quantity $Q^* = 54.2$, from Eq. (32). The minimum cost obtained by this model is improved (as it was expected) compared to the minimum cost $TC = 102.29$ ($t_1 = 0.78$, $x_1 = 0.24$), which is derived by the model of Skouri and Konstantaras (2013).

8.2 Example 2 ($\mu < M < T$)

As a second example we examine the case described in Sect. 3.3. Again, we use the same input parameters as in example 1 and set $M = 0.95$, so that $\mu < M < T$. For this case the only feasible solution is derived from Eq. (26) and corresponds to the case $x_1 < t_1 < \mu < M < T$. Hence, the optimal solution to the problem is $\left(t_1{}^*, x_1{}^*\right) = (0.82, 0.5)$, where $x_1{}^*$ is derived from relation (17) . The minimum cost $TC_{2.A} = 65.86$ is derived from Eq. (10) and the optimal order quantity $Q^* = 54.4$, from Eq. (32). Again the proposed model gives an improved outcome compared to Skouri and Konstantaras (2013).

9 Sensitivity Analysis

To study the effect, of under or over estimation of various parameters on the optimal replenishment policy and total cost, a sensitivity analysis is performed for both cases and the results are presented in Tables 1 and 2 respectively.

From both tables, we can observe, that the optimal cost is sensitive to changes in μ, W and c_{1O}, t_1^* is insensitive to any changes of the parameters, x_1^* is moderately sensitive to changes in the considered parameters and the optimal ordering quantity is sensitive to changes in μ.

It is worth noting the low cost values in Tables 1 and 2, when $\mu = 0.45$. This could be explained due to the small demand rate, as the exponential type demand rate is stabilized very soon, which leads to reduced holding costs. Also, since only one warehouse is used, the transportation cost is omitted.

Table 1 Sensitivity analysis for Case 1 (M < μ)

Parameter	Percentage of change	t_1^*	x_1^*	TC^*	Q^*	$\frac{TC^{*'}-TC^*}{TC^*}100\%$
M	−50 % (M = 0.2)	0.78	0.3	97.82	54.2	+5.5 %
	−25 % (M = 0.3)	0.79	0.33	95.36	54.2	+2.9 %
	+25 % (M = 0.5)	0.79	0.38	89.64	54.2	−3.3 %
	+50 % (M = 0.6)	0.8	0.4	86.06	54.3	−7.1 %
μ	−50 % (μ = 0.45)	0.88	0[a]	20	16.9	−78.4 %
	−25 % (μ = 0.675)	0.8	0.2	52.4	33.3	−43.5 %
W	−50 % (W = 10)	0.78	0.64	112.95	54.2	+21.8 %
	−25 % (W = 15)	0.78	0.54	102.48	54.2	+10.6 %
	+25 % (W = 25)	0.81	0	84.23	54.3	−9.1 %
	+50 % (W = 30)	0.85	0	79.2	54.5	−14.5 %
c_{1R}	−25 % (c_{1R} = 2.25)	0.8	0.43	92.04	54.3	−0.6 %
	+25 % (c_{1R} = 3.75)	0.78	0.3	93.03	54.2	+0.4 %
	+50 % (c_{1R} = 4.5)	0.78	0.25	93.23	54.2	+0.6
c_{1O}	−50 % (c_{10} = 1)	0.8	0.43	79.56	54.3	−14.2 %
	−25 % (c_{10} = 1.5)	0.79	0.4	86.2	54.3	−7 %
	+25 % (c_{10} = 2.5)	0.78	0.3	98.99	54.2	+6.8 %

[a]The single warehouse model is used to obtain the optimal solution

Table 2 Sensitivity analysis for Case 2 (μ < M < T)

Parameter	Percentage of change	t_1^*	x_1^*	TC^*	Q^*	$\frac{TC^{*'}-TC^*}{TC^*}100\%$
μ	−50 % (μ = 0.45)	0.9	0[a]	3.2	16.9	−95.1 %
	−25 % (μ = 0.675)	0.83	0.36	27.8	33.4	−43.5 %
W	−50 % (W = 10)	0.81	0.7	87.1	54.3	+32.3 %
	−25 % (W = 15)	0.82	0.63	76.2	54.4	+15.7 %
	+25 % (W = 25)	0.83	0.32	56.2	54.4	−14.7 %
	+50 % (W = 30)	0.85	0	48.1	54.5	−27 %
c_{1R}	−25 % (c_{1R} = 2.25)	0.84	0.57	63.9	54.4	−3 %
	+25 % (c_{1R} = 3.75)	0.81	0.46	67.2	54.3	+2 %
	+50 % (c_{1R} = 4.5)	0.8	0.4	68	54.3	+3.3 %
c_{1O}	−50 % (c_{10} = 1)	0.83	0.55	51.7	54.4	−21.5 %
	−25 % (c_{10} = 1.5)	0.83	0.53	58.8	54.4	−10.7 %
	+25 % (c_{10} = 2.5)	0.81	0.49	72.8	54.3	+10.5 %

[a]The single warehouse model is used to obtain the optimal solution

10 Concluding Remarks

In this paper an inventory model for seasonal products is considered, taking into account the warehouse capacity and credit period options. The study of the inventory system requires exploring the feasible ordering relations between the involved time parameters. Consequently, two cases, which lead to two inventory models, are examined: (1) the offered credit period is less than the time point when the demand rate is stabilized ($M < \mu$), (2) the offered credit period is longer than the time point when the demand rate is stabilized and less than the planning horizon($\mu < M < T$). The aim is to find the optimal time point at which the inventory level of the RW and the OW reaches zero, in order to minimize the total cost of each inventory system. The determination of these time instances lead to the determination of the optimal ordering quantity. Furthermore, in order to solve the problem, the respective single warehouse problem should be solved first, hence this case is also examined. We note that this problem can also be treated by considering profit maximization. In this case the values for lost sales and deterioration cost should be appropriately modified (see Teng et al. 2002). However, the optimization procedure remains unchanged, as the profit maximization problem can be easily transformed to a cost minimization one.

The proposed model could be extended assuming that demand is price and time dependent. In addition, vendor-buyer coordination could be considered.

Appendix 1

The solution of Eq. (20) is given by $a_{1.A.1}(t_1) = b(t_1, x_1 = h_A(t_1)) - pI_e(M - t_1) = 0$. Differentiating $a_{1.A.1}(t_1)$ with respect to t_1 we obtain:

$$a_{1.\mathrm{A}.1}{'}(t_1) = b'(t_1, x_1 = h_A(t_1)) + pI_e > 0,$$

where

$$b'(t_1,x_1=h_A(t_1))=\left[\left(\frac{c_{1O}}{\theta_2}-\frac{c_{1R}}{\theta_1}+c_t\right)\theta_2\left(1-\frac{dx_1}{dt_1}\right)+\frac{c_{1R}+c_3\theta_1}{\theta_1}(\theta_1-\theta_2)e^{\theta_1 x_1}\frac{dx_1}{dt_1}\right.$$
$$\left.+\frac{c_{1R}+c_3\theta_1}{\theta_1}\theta_2 e^{\theta_1 x_1}\right]e^{\theta_2(t_1-x_1)}+c_2\left[\beta(T-t_1)+(T-t_1)\beta'(T-t_1)\right]-c_4\beta'(T-t_1)>0,$$

since from Eq. (17) we obtain $\frac{dx_1}{dt_1} = e^{\theta_2(t_1-x_1)}\frac{f(t_1)}{f(x_1)} > 1$ and by the assumptions we have $\frac{c_{1R}}{\theta_1} > \frac{c_{1O}}{\theta_2} + c_t$, $\theta_1 > \theta_2$ and $\beta(T - t_1) + (T - t_1)\beta'(T - t_1) > 0$.

Also, we have

$$\begin{aligned} a_{1.A.1}(0) = {} & \frac{c_{1R} + c_3\theta_1}{\theta_1} e^{-\theta_2 x_1}\left(e^{\theta_1 x_1} - 1\right) + \frac{c_{1O} + c_3\theta_2}{\theta_2}\left(e^{-\theta_2 x_1} - 1\right) - c_2 T\beta(T) \\ & - c_4(1 - \beta(T)) + c_t e^{-\theta_2 x_1} - pI_e M \\ < {} & 0, \end{aligned}$$

if $c_2 T\beta(T) + c_4(1 - \beta(T)) \geq c_t$ (see Skouri and Konstantaras 2013), and $a_{1.A.1}(T) > 0$.

Hence Eq. (20) has a unique solution which satisfies the second order conditions for a minimum.

Accordingly from Eqs. (21) to (25) we have

$$\begin{aligned} a_{1.A.2}(t_1) &= b(t_1, x_1 = h_A(t_1)) + \frac{C_p I_c}{\theta_2}\left(e^{\theta_2(t_1-M)} - 1\right) = 0 \Rightarrow a_{1.\mathrm{A}.2}{}'(t_1) \\ &= b'(t_1, x_1 = h_A(t_1)) + C_p I_c e^{\theta_2(t_1-M)} > 0, \\ a_{1.A.3}(t_1) &= b(t_1, x_1 = h_A(t_1)) \\ &\quad + C_p I_c\left[\frac{1}{\theta_1} e^{\theta_2(t_1-x_1)}\left(e^{\theta_1(x_1-M)} - 1\right) + \frac{1}{\theta_2}\left(e^{\theta_2(t_1-x_1)} - 1\right)\right] \\ &\Rightarrow \end{aligned}$$

$$\begin{aligned} a_{1.\mathrm{A}.3}{}'(t_1) &= b'(t_1, x_1 = h_A(t_1)) \\ &\quad + C_p I_c\left[\frac{\theta_2 - \theta_1}{\theta_1}\left(e^{\theta_1(x_1-M)} - 1\right)\left(1 - \frac{dx_1}{dt_1}\right) + e^{\theta_1(x_1-M)}\right] e^{\theta_2(t_1-x_1)} \\ &> 0, \end{aligned}$$

$$\begin{aligned} a_{1..1}(t_1) &= b(t_1, x_1 = h_B(t_1)) + \frac{C_p I_c}{\theta_2}\left(e^{\theta_2(t_1-M)} - 1\right) \Rightarrow a_{1..1}{}'(t_1) \\ &= b'(t_1, x_1 = h(t_1)) + C_p I_c e^{\theta_2(t_1-M)} > 0, \\ a_{1..2}(t_1) &= b(t_1, x_1 = h_B(t_1)) \\ &\quad + C_p I_c\left[\frac{1}{\theta_1} e^{\theta_2(t_1-x_1)}\left(e^{\theta_1(x_1-M)} - 1\right) + \frac{1}{\theta_2}\left(e^{\theta_1(t_1-x_1)} - 1\right)\right] \\ &\Rightarrow \end{aligned}$$

$$\begin{aligned} a_{1..2}{}'(t_1) &= b'(t_1, x_1 = h(t_1)) \\ &\quad + C_p I_c\left[\frac{\theta_2 - \theta_1}{\theta_1}\left(e^{\theta_1(x_1-M)} - 1\right)\left(1 - \frac{dx_1}{dt_1}\right) + e^{\theta_1(x_1-M)}\right] e^{\theta_2(t_1-x_1)} \\ &> 0, \end{aligned}$$

$$\begin{aligned} a_{1.\Gamma.1}(t_1) &= b(t_1, x_1 = h_\Gamma(t_1)) \\ &\quad + C_p I_c\left[\frac{1}{\theta_1} e^{\theta_2(t_1-x_1)}\left(e^{\theta_1(x_1-M)} - 1\right) + \frac{1}{\theta_2}\left(e^{\theta_1(t_1-x_1)} - 1\right)\right] \\ &\Rightarrow \end{aligned}$$

$$a_{1.\Gamma.1}{}'(t_1) = b'(t_1, x_1 = h_\Gamma(t_1)) + C_pI_c\left[\frac{\theta_2-\theta_1}{\theta_1}\left(e^{\theta_1(x_1-M)}-1\right)\left(1-\frac{dx_1}{dt_1}\right)+e^{\theta_1(x_1-M)}\right]e^{\theta_2(t_1-x_1)} > 0,$$

taking into account the assumptions of Theorem 1.

The optimal value of x_1 is derived by using the appropriate relation from (17) to (19).

Obviously, the solution of each equation must also satisfy the constraints that are derived by the ordering relations between x_1, t_1, μ, M and T, for each case accordingly.

Appendix 2

The solution of Eq. (35) is given by $a_{0.1.1}(t_1) = b_0(t_1) - pI_e(M - t_1) = 0$. Differentiating $a_{O.1.1}(t_1)$ with respect to t_1 we obtain: $a_{0.1.1}{}'(t_1) = b_0{}'(t_1) + pI_e > 0$, where

$$b_O'(t_1) = (c_{1O} + c_3\theta_2)e^{\theta_2 t_1} + c_2\left[\beta(T-t_1) + (T-t_1)\beta'(T-t_1)\right] - c_4\beta'(T-t_1) > 0,$$

since $\beta(T-t_1) + (T-t_1)\beta'(T-t_1) > 0$.

Also, we have $a_{O.1.1}(0) = -c_2T\beta(T) - c_4(1-\beta(T)) - pI_eM < 0$ and $a_{0.1.1}(T) > 0$.

Hence Eq. (35) has a unique solution which satisfies the second order conditions for a minimum.

Accordingly from Eqs. (36) to (37) we have

$$a_{O.1.2}(t_1) = b_O(t_1) + \frac{C_pI_c}{\theta_2}\left(e^{\theta_2(t_1-M)}-1\right) = 0 \Rightarrow a_{O.1.2}{}'(t_1) = b_0'(t_1) + C_pI_ce^{\theta_2(t_1-M)} > 0,$$

$$a_{O.1.3}(t_1) = b_0(t_1) + \frac{C_pI_c}{\theta_2}\left(e^{\theta_2(t_1-M)}-1\right) \Rightarrow a_{O.1.3}{}'(t_1) = b_0'(t_1) + C_pI_ce^{\theta_2(t_1-M)}e^{\theta_2 t_1} > 0.$$

Obviously, the solution of each equation must also satisfy the constraints that are derived by the ordering relations between t_1, μ, M and T, for each case accordingly.

References

Agrawal S, Banerjee S (2011) Two-warehouse inventory model with ramp-type demand and partially backlogged shortages. Int J Syst Sci 42(7):1115–1126

Agrawal S, Banerjee S, Papachristos S (2013) Inventory model with deteriorating items, ramp-type demand and partially backlogged shortages for a two warehouse system. Appl Math Model 37(20–21):8912–8929

Darzanou G, Skouri K (2011) An inventory system for deteriorating products with ramp-type demand rate under two-level trade credit financing. Adv Decis Sci 2011:15. doi:10.1155/2011/761961, Article ID: 761961

Deng PS, Lin HJ, Chu P (2007) A note on the inventory models for deteriorating items with ramp type demand rate. Eur J Oper Res 178(1):112–120

Goyal SK (1985) Economic order quantity under conditions of permissible delay in payments. J Oper Res Soc 36(4):335–338

Hill RM (1995) Inventory model for increasing demand followed by level demand. J Oper Res Soc 46:1250–1259

Mandal B, Pal AK (1998) Order level inventory system with ramp type demand rate for deteriorating items. J Interdis Math 1:49–66

Skouri K, Konstantaras I (2013) Two warehouse inventory model for deteriorating seasonable products with ramp type demand rate. J Ind Manag Opt 9(4):855–883

Skouri K, Konstantaras I, Papachristos S, Ganas I (2009) Inventory models with ramp type demand rate, partial backlogging and Weibull deterioration rate. Eur J Oper Res 192(1):79–92

Skouri K, Konstantaras I, Papachristos S, Teng JT (2011) Supply chain models for deteriorating products with ramp type demand rate under permissible delay in payments. Expert Syst Appl 38(12):14861–14869

Teng JT, Chang HJ, Dye CY, Hung CH (2002) An optimal replenishment policy for deteriorating items with time-varying demand and partial backlogging. Oper Res Lett 30:387–393

Wu KS (2001) An EOQ inventory model for items with Weibull distribution deterioration, ramp type demand rate and partial backlogging. Prod Plan Control 12(8):787–793

Wu KS, Ouyang LY (2000) A replenishment policy for deteriorating items with ramp type demand rate. Proc Natl Sci Council ROC(A) 24(4):279–286

Wu JW, Lin C, Tan B, Lee WC (1999) An EOQ inventory model with ramp type demand rate for items with Weibull deterioration. Inform Manag Sci 10(3):41–51

Optimal Active Power Management in All Electric Ship Employing DC Grid Technology

Fotis D. Kanellos, John Prousalidis, and George J. Tsekouras

Abstract Extensive electrification and the use of dc distribution grid are recently proved to be very promising technologies for the development of more efficient and environmentally friendly ships. Onboard dc grids present several advantages such as, improved efficiency, easy integration of different types of power sources, reduced size and rating of switchboard, elimination of reactive power flow, increased reconfiguration capability etc. All electric ship (AES) concept, dc distribution grid and optimal power management can lead to a substantial improvement of ship efficiency and compliance with the environmental constraints. In this paper, a method for optimal demand side management and power generation scheduling is proposed for AES employing dc grid. Demand side management is based on the adjustment of the power consumed by ship electric propulsion motors. Dynamic programming algorithm subject to operation, environmental and travel constraints is used to solve the above problem.

Keywords All electric ship • Dynamic programming • GHG emissions limitation • Onboard DC grids • Optimal power management

F.D. Kanellos (✉)
School of Production Engineering and Management, Technical University of Crete, Chania, Greece
e-mail: fkanellos@dpem.tuc.gr

J. Prousalidis
School of Naval Architecture and Marine Engineering, National Technical University of Athens, Zografos, Athens, Greece
e-mail: jprousal@naval.ntua.gr

G.J. Tsekouras
Department of Electrical Engineering and Computer Science, Hellenic Naval Academy, Piraeus (Hatzikiriakio), Greece
e-mail: tsekouras@snd.edu.gr

E. Grigoroudis, M. Doumpos (eds.), *Operational Research in Business and Economics*, Springer Proceedings in Business and Economics,
DOI 10.1007/978-3-319-33003-7_14

1 Introduction

Several studies have shown that the extensive electrification of ship power systems, can lead to a “greener” and more efficient ship. To this end, novel concepts are explored, with an attractive one being the introduction of direct current (DC) in the power distribution system. Onboard DC grids present several advantages (Kanellos et al. 2015; Justo et al. 2013; Planas et al. 2013; Zahedi et al. 2014; Certuche-Alzate and Vélez-Reyes 2009; Sulligoi et al. 2013; Sudhoff and Crider 2011; Zadeh et al. 2013; Zhu et al. 2013; Chen et al. 2012; Tang and Ooi 2007) such as, improved efficiency, easy integration of different types of power sources, reduced size and rating of switchboard, elimination of reactive power flow, increased reconfiguration capability etc. Ship efficiency is improved mainly by the fact that the power generation system is not locked at a specific frequency while space savings and weight reduction have also a significant positive effect. The vast majority of ship electric generators prime movers are combustion engines, most fueled with liquid oil, some with liquid natural gas while some can use both liquid fuel and gas. Prime mover fixed speed operation in conventional ship power systems entails minimum fuel consumption within a narrow operating area (usually around 85 % of nominal power). In case of dc grid technology variable speed operation is possible and minimum fuel consumption can be obtained in a wider operating area while fuel consumption is significantly decreased as compared with fixed speed operation. Variable speed operation is especially beneficial for vessels with highly variable load like those with dynamic positioning operation capability.

The exploitation of electric motors for propulsion purposes, known as All Electric Ship (AES) technology, has become an appealing technology compared with conventional ship power system concepts. In this case, improvement in ship power system efficiency could be achieved if optimal power management and control techniques are adopted (Kanellos 2014; Kanellos et al. 2012; Tsekouras and Kanellos 2013; Tsekouras et al. 2015; McCoy and Amy Jr 2009; Prousalidis et al. 2012; Falahi et al. 2013; Gjengedal 1996; Singhal and Sharma 2011). Except from the maximization of ship energy efficiency another crucial issue for the future ship power systems is the limitation of greenhouse gas (GHG) emissions. Optimal power management can provide solutions also to this problem.

AES is a very complex system making the application of centralized optimal control systems quite difficult and challenging. Optimal power management in AES is a complex optimization problem, since next to the minimization of the operational cost and GHG emissions, several technical and operational constraints, like power production ramp rates, minimum–maximum power production, minimum generator operation time, ship speed limits, total route length, constraints stemming from calls at intermediate ports etc., must be satisfied. Moreover, several of the optimization goals might be antagonistic while the problems of demand side management and power generation scheduling are coupled and cannot be solved independently.

This paper aims to present in a coherent and methodical way an optimal active power management technique that could be applied to AES employing dc grids leading to the improvement of ship efficiency and the limitation of GHG emissions.

2 AES Power System

A generic single-line diagram of an AES with dc distribution grid is shown in Fig. 1. Electric power generation system supplies ship service load and the electric propulsion motors. The new concept in this type of ship power system is that the power produced by the generators if fed by ac-dc power converters to the dc distribution grid while dc-ac converters are used where power should be supplied to ac loads (Kanellos et al. 2015). Each main AC consumer is supplied by its own converter in order to ensure maximum reliability.

Onboard DC grid configurations can be categorized into the two major schemes: a centralized and a fully distributed system. Two major advantages of both centralized and distributed DC architectures are that the main AC switchboard and thruster transformers are omitted. In centralized configuration all converters are located in the same space while in the distributed one each converter is located near to the electric generator or the electric load it serves. In this case, each electric generator is equipped with a rectifier mounted directly on the generator or placed nearby the generator.

It is a common practice, especially in naval applications, to exploit distribution systems with zonal structure. Power generation units are typically installed in some of the electrical zones while in case of single bus architecture sufficient power generation should be installed in each zone.

The most significant benefits from onboard DC power distribution are synopsized next.

– Efficiency can be increased up to 20 % because of reduced fuel consumption, space and weight savings.

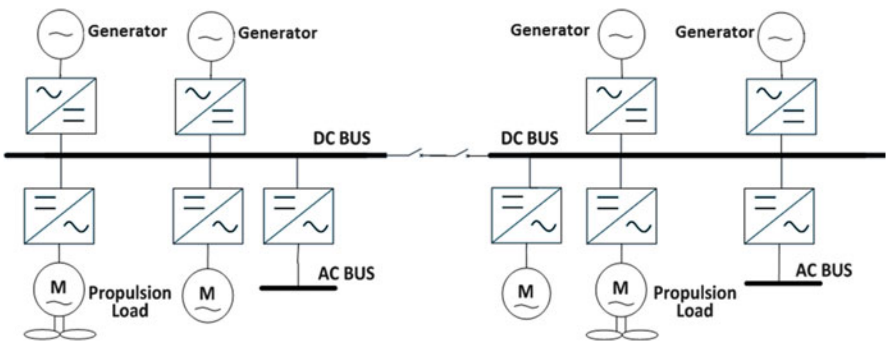


Fig. 1 Generic single-line diagram of AES with dc grid technology

- Reduced fuel consumption can be achieved because of the generator variable speed operation allowing it to operate at the optimal operation point for variable loading.
- Less maintenance is required for the electric generators as they are operated in a more efficient way (operation at variable speed).
- The footprint of the electrical equipment is decreased allowing savings in electrical equipment space requirements and weight by up to 30 %. This can be attributed to the reduced size and rating of the switchboard, the use of smaller high speed generators and the elimination of large low-frequency thruster transformers. It is noted that the exact savings will vary depending on ship type and application.
- Improved power flow control, better dynamic response and maneuverability are possible. Moreover, reactive power flows and respective voltage drops are eliminated.
- Fault currents could be limited and managed more effectively leading to the drastic reduction of short-circuit levels and lower DC switchboard class requirements.
- Reconfiguration capability is significantly increased.
- Simplified connection and disconnection of different types and sizes of power generation and energy storage systems can be easily integrated while plug and play capability for future installed power sources is provided. The need for phase angle synchronization of multiple sources and loads is eliminated.
- Higher power ratings for a given cable size are achieved.

2.1 Operation Constraints in AES Power System Operation

AES power system operation is subject to constraints stemming from power generation-consumption balance, generator loading, ramp rate limitation, blackout prevention and generator frequent start/stop avoidance. There are also some additional constraints related with ship operation listed next (Kanellos 2014; Kanellos et al. 2012; Tsekouras and Kanellos 2013; Tsekouras et al. 2015):

- Ship speed is upper and lower bounded by the respective maximum and minimum ship speed.
- The distance traveled by the ship when arriving at intermediate ports should be equal to the route length between ship departure port and the intermediate port.
- The distance traveled at the end of the last time interval of the optimization period should be equal to the total route length.
- GHG emissions should be limited below a certain upper limit.

2.2 Economic and Environmentally Friendly AES Operation

AES load is composed by ship service and electric propulsion load. Electric propulsion load includes propulsion and thruster drives while ship service load comprises the remaining part of the electric power consumption.

$$L = P_{\mathrm{Pr}op} + L_{Serv} \tag{1}$$

The flexibility of the electric load depends highly on the flexibility of the power demand of the electric propulsion motors. Propulsion power depends on ship speed which is in turn derived by travel scheduling. Ship electric propulsion system power demand can be adjusted with respect to ship service load in order to lead to more economic and environmentally friendly operation.

Specific fuel consumption function (*SFC*) is particularly useful for the determination of the most economical point of operation of a generator. *SFC* determines generator fuel consumption per MW and hour and it can be approximated as follows:

$$SFC_i\left(P_i, \omega^*\right) = \frac{a_{0i} + a_{1i} \cdot P_i\left(\omega^*\right) + a_{2i} \cdot P_i^2\left(\omega^*\right)}{P_i(\omega^*)}, \quad P_{\min,i} \leq P_i \leq P_{\max,i} \tag{2}$$

Where, P_i is the power produced by i-th generator and ω^* is the optimal rotating speed of the generator prime mover. In ship power systems with dc grid technology ω^* is not constant but it is continuously adjusted.

In the following analysis it is assumed that speed controller ensures successful tracking of optimal speed ω^* and the steady state operation is considered where *SFC* is decoupled from ω^*. However, it is noted that the shape of *SFC* curve of a variable speed engine is different from that of the respective fixed speed engine. In the case study presented in Sect. 4 typical *SFC* curves of variable speed engines are used. In case of variable speed operation minimum fuel consumption can be obtained in a wider operating area while fuel consumption is significantly decreased as compared with fixed speed operation. In this case active power sharing through active power—dc voltage droop characteristics in an onboard DC grid cannot guarantee minimum fuel consumption, as this depends on the amount of the produced power and the generator prime mover rotating speed. To deal with this major challenge smart power management techniques should be employed.

The total variable operation cost, *ToC*, of AES power system for a travel time period *T*, is calculated as:

$$\mathrm{To}C = \sum_{j=1}^{T} \sum_{i=1}^{N_g} \left(St_{ij} \cdot \left(Fuel\_Cost_i \cdot SFC_i(P_{ij}) + OC_i\right) \cdot P_{ij} \cdot \Delta T_j + SC_{ij}\right) \tag{3}$$

Where, OC_i is the variable operation cost per produced MWh including lubricants, maintenance etc., SC_{ij} is the start-up cost of the i-th generator at time interval, ΔT_{j} , $Fuel\_Cost_i$ is the unit cost of the fuel consumed by the i-th generator (m.u./fuel gram) and St_{ij} is a parameter being equal to 1 or 0 if the i-th unit is ON or OFF, respectively.

According to the International Maritime Organization policy ship operation efficiency should be evaluated via Energy Efficiency Operation Indicator (EEOI). EEOI can be slightly modified to facilitate the optimization of AES operation if it is referred to an arbitrary observation time interval ΔT_j, as follows:

$$\mathrm{EEOI}_{1,j} = \frac{mCO_2}{LF \cdot V_j \cdot \Delta T_j} = \frac{\sum_i c_i \cdot P_{ij} \cdot SFC_i\left(P_{ij}\right)}{LF \cdot V_j} \quad \left[{}^{g}CO_2/_{tn \cdot kn}\right] \tag{4}$$

Ship loading factor, LF, (Kanellos 2014) is calculated as:

$$LF = \frac{n_P' \cdot 0.1 + n_V'}{n_P \cdot 0.1 + n_V} GT \tag{5}$$

Where, GT is ship gross tonnage, n_V' is the number of vehicles onboard, n_V is the maximum number of vehicles, n_P' is the number of passengers onboard and n_P is the maximum number of passengers.

2.3 Propulsion Load Adjustment

Ship speed-propulsion power curve depends on hull resistance at specific conditions (loading of the ship, weather conditions etc.) and it is well-described by the following formula (Turbo 2012):

$$P_{Prop} = c_{P1} \cdot V^{c_{P2}} \tag{6}$$

Where, V is ship velocity, P_{Prop} is the required propulsion power to develop velocity V, c_{P1} is a coefficient used for propulsion power and ship velocity matching, c_{P2} is a constant depending on hull form ($c_{P2} = 3$ for conventional hull forms).

If during time interval ΔT_j ship speed is V_j' while the scheduled speed before optimization is V_j then the deviation of the propulsion power is:

$$\Delta P_{\mathrm{Prop},j} = c_{P1} \cdot \left(V_j'^{\,c_{P2}} - V_j^{c_{P2}}\right) \tag{7}$$

3 Optimal Active Power Management Problem

3.1 Formulation

In the examined optimization problem the total ship power system operation cost is minimized for the whole ship travel period subject to the operational and technical constraints of ship power system. The objective function and the applied constraints are contained in the following mathematic relations.

Minimize:

$$\mathrm{To}C = \sum_{j=1}^{T}\sum_{i=1}^{N_g}\left(St_{ij}\cdot\left(Fuel\_Cost_i\cdot SFC_i\left(P_{ij}\right)+OC_i\right)\cdot P_{ij}\cdot \Delta T_j + SC_{ij}\right) \tag{8}$$

Subject to:

Power balance constraint,

$$\sum_{i=1}^{N_g} P_{ij} = L_j + \Delta P_{\mathrm{Pr}opj}, \quad \forall j \tag{9}$$

Minimum and maximum generator loading constraints,

$$P_{i,\min} < P_{ij} < P_{i,\max} \quad \forall i, j \tag{10}$$

Generator ramp rate constraint,

$$\frac{\left|P_{ij} - P_{i,j-1}\right|}{\Delta T_j} \leq Rc_{i,\max}, \quad \forall i, j \tag{11}$$

Minimum operation time of generator,

$$t_{\rightarrow \mathrm{OFF,i}} - t_{\rightarrow \mathrm{ON,i}} \geq \mathrm{T}_{\mathrm{ON\_min,i}}, \quad \forall i \tag{12}$$

Minimum out of operation time of a generator,

$$t_{\rightarrow \mathrm{ON,i}} - t_{\rightarrow \mathrm{OFF,i}} \geq \mathrm{T}_{\mathrm{OFF\_min,i}}, \quad \forall i \tag{13}$$

Blackout prevention constraint,

$$\sum_i St_{ij}\cdot P_{i,\max} - L_j - \Delta P_{\mathrm{Pr}\,opj} \geq \max\left\{P_{i,\max}\right\}, \quad \forall j \tag{14}$$

GHG emissions constraint,

$$\frac{\sum_{i=1}^{N_g} c_i \cdot St_{ij} \cdot P_{ij} \cdot SFC_i\left(P_{ij}\right)}{LF \cdot V_j} \leq I_{\max} \tag{15}$$

Minimum—maximum ship speed constraint,

$$V_{\min} < V_j < V_{\max}, \quad \forall j \tag{16}$$

Initial condition for the deviation of the actual travel distance from the scheduled,

$$\Delta S_0 = 0 \tag{17}$$

Final condition for the deviation of the actual travel distance from the scheduled one,

$$\Delta S_{\mathrm{T}} = 0 \tag{18}$$

Deviation of the actual traveled distance from the scheduled one at the intermediate ports,

$$\Delta S_j = 0 \quad \forall j \in \Im \tag{19}$$

Minimum and maximum deviation of the actual traveled distance from the scheduled,

$$\Delta S_{\min,j} \leq \Delta S_j \leq \Delta S_{\max,j}, \quad \forall j \tag{20}$$

Where, i denotes *i-th* generator, j denotes *j-th* time interval, $\Im$ denotes the set containing time intervals corresponding to ship calls at intermediate ports and St_{ij} is a parameter being equal to 1 or 0 if the *i-th* generator is ON or OFF, respectively.

3.2 Solution Method

Dynamic programming is used to solve the above constraint nonlinear optimization problem (Kanellos 2014; Gjengedal 1996; Singhal and Sharma 2011). Grid stages correspond to the time intervals of the examined time period while grid states contain generator-state vector. State vector used for the examined problem is defined in (21).

$$State = \left[\underbrace{D_{G_1} \quad D_{G_2} \quad \ldots \quad D_{G_{Ng}}}_{Generators\ State} \underbrace{D_{\Delta S_1} \quad D_{\Delta S_2} \quad \ldots \quad D_{\Delta S_{Ns}}}_{\Delta S_{bin}} \right] \tag{21}$$

Where, D_{Gi} is binary digit used to define i-th generator state of operation while the deviation from the initially scheduled travelled distance is quantified with a string, $\Delta S_{bin,\ell}$, of binary digits, $D_{\Delta S}$, with total number of digits, N_s.

The deviation from the scheduled speed can be easily calculated by using the deviation from travelled distance that is contained into the grid-state. Power production cost, PC, is minimized at each grid point by applying classical economic power dispatch based on the well-known Lagrange method. Afterwards, total cost of state ℓ at time interval t, $ToC_{t,l}$, is minimized according to (22).

$$ToC_{t,\ell} = \min\left\{PC_{t,\ell} + SC_{(t-1,k),(t,\ell)} + ToC_{t-1,k}\right\}, \forall k\ state\ at\ t-1\, stage \tag{22}$$

The completion process of a grid point is summarized in the following steps.

1. Constraints are checked.
2. Ship speed is estimated for each possible state transition and it is checked if it is between its upper and lower limits.
3. Propulsion load is estimated for all acceptable ship speed values obtained in step 2.
4. Total ship electric load is estimated for all propulsion load values obtained in step 3.
5. Optimal power dispatch is performed for all possible state transitions from previous stage and respective production cost is estimated.
6. Total cost values (ToC) for production cost values obtained in step 5 are estimated and minimum total cost is obtained according to (22). State transition resulting in minimum total cost (ToC) is marked. Finally, the grid point at the last stage with the minimum total cost is selected, and moving backward, optimal operation path is formed.

4 Case Study

The proposed method is applied to the electric power system of a cruise-ferry with full electric propulsion and dc distribution grid. All the necessary technical parameters of the ship and the onboard power system are presented in Table 1. The examined cruise-ferry comprises 2 large electric propulsion motors supplied by the set of 4 electrical generators driven by different diesel engines. Two operation scenarios are examined next. In scenario 1, the proposed optimal power management method is applied while in the second scenario propulsion load is not adjusted and EEOI is not limited.

Table 1 Ship power system model data

	Gen 1	Gen 2	Gen 3	Gen 4
Ship electric system parameters				
Nominal power (MW)	10	10	15	15
Technical minimum (MW)	2	2	3	3
Minimum hours for generator being in operation or out of operation (h)	2/2	2/2	2/2	2/2
Generator startup/stop cost (m.u.)	200/0	200/0	200/0	200/0
Fuel Cost = SFC (monetary unit/MWh) × P(MW)				
Gen 1	$FC_1(P_1) = SFC_1(P_1) \cdot P_1 = 98.1 + 148.7 \cdot P_1 + 3.39 \cdot {P_1}^2$			
Gen 2	$FC_2(P_2) = SFC_2(P_2) \cdot P_2 = 86.5 + 145.8 \cdot P_2 + 3.69 \cdot {P_2}^2$			
Gen 3	$FC_3(P_3) = SFC_3(P_3) \cdot P_3 = 156.1 + 146.2 \cdot P_3 + 2.28 \cdot {P_3}^2$			
Gen 4	$FC_4(P_4) = SFC_4(P_4) \cdot P_4 = 137.7 + 143.2 \cdot P_4 + 2.485 \cdot {P_4}^2$			

Type	Nominal speed (kn)	Max. no. of vehicles (n_v)	Max. number of passengers	$EEOI_{max}(gCO_2/tn.kn)$	Gross registered tons
Ship parameters					
Cruise ferry	24.5	700	2500	15	60,000

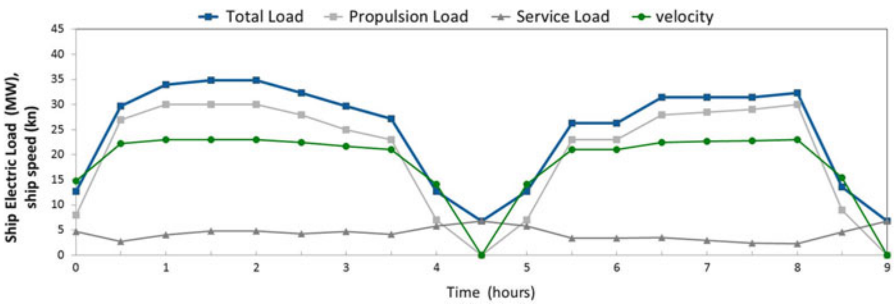


Fig. 2 Ship total load, propulsion load, service load for non-optimized operation

Moreover, it is assumed that the ship stops at one intermediate port. Ship loading factor (LF) values during the two parts of the examined travel are assumed 49,000 tn and 46,000 tn, respectively. For the estimation of LF, it has been assumed that every passenger weighs 100 kg with his luggage and every vehicle 1 tn on average. Total ship electric load and service load forecasts are shown in Fig. 2 for

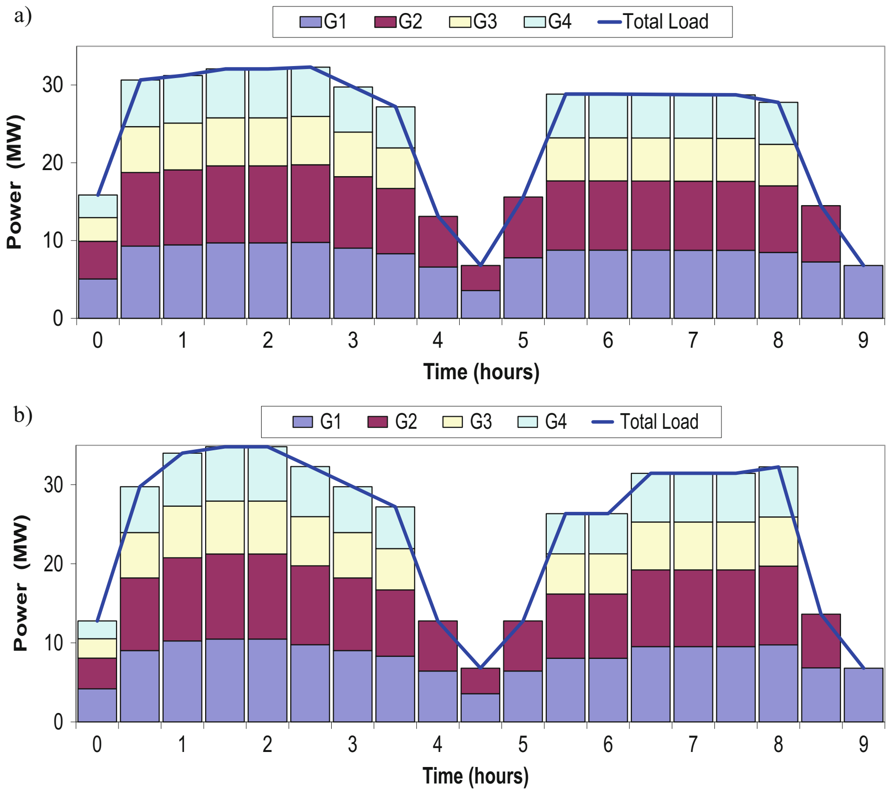


Fig. 3 Generators power production. (**a**) Scenario 1 and (**b**) scenario 2.

ship power system non-optimized operation. Also, the initial non-optimized schedules for the demand in propulsion power and the ship speed for the examined voyage are shown in the same figure.

The powers produced by the electric generators are shown in stack form in Fig. 3. In scenario 1, propulsion power is adjusted and smaller deviations of total electric load occur. Generators 1 and 2 are less expensive than generators 3 and 4; thus they are operated continuously while 3 and 4 are used during high load periods. EEOI is successfully limited in scenario 1 as it is shown in Fig. 4a while in scenario 2 it exceeds its upper limit. In scenario 1, ship speed is decreased when ship is in open sea, while it is increased, when ship is approaching or leaving from a port. In case of scenario 1 the total operation cost is 83,564 (m.u.) that is 3 % lower than that of scenario 2 (85,780 m.u.). The deviation of ship power system operation cost is shown in Fig. 5.

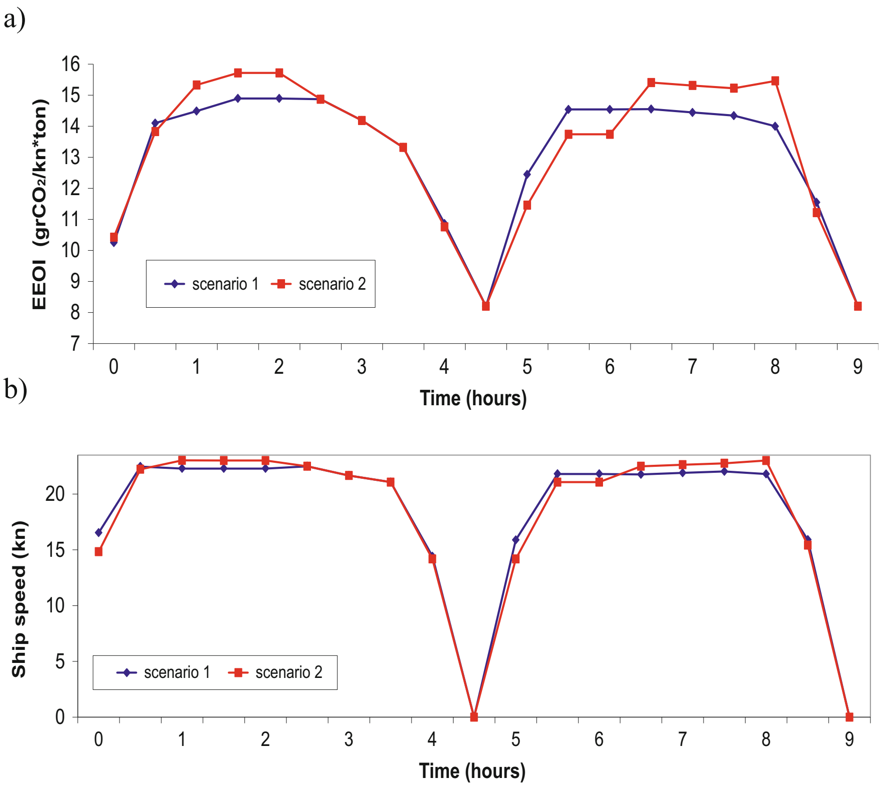


Fig. 4 (**a**) EEOI time evolution and (**b**) Ship speed time evolution

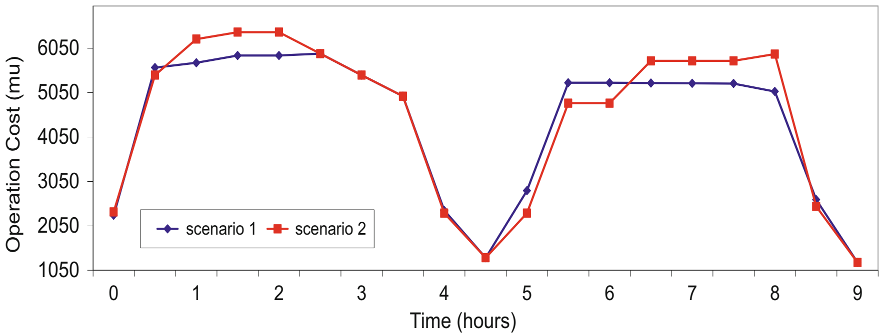


Fig. 5 Operation cost time evolution

5 Conclusions

In this paper an optimal power management method for AES employing dc grid technology is proposed. The proposed method aims at operation cost minimization, GHG emissions limitation and, satisfaction of all ship power system technical and operational constraints concerning power balance, blackout avoidance, travel schedule etc. Despite the complexity of the problem the proposed method is proved efficient in minimizing the operation cost and limiting GHG emissions. Results obtained by applying the method to a cruise-ferry showed that operation cost can be reduced by almost 3 % if propulsion adjustment together with EEOI limitation are applied with respect to operation without propulsion adjustment and GHG limitation. The benefits from the application of the proposed method together with the advantages of onboard dc grid technology prove that the examined ship electric power system can be a very promising solution for the future AES.

Acknowledgement The work presented in this paper has been financially supported within the framework of the "DC-Ship" project (ARISTEIA-EXCELLENCE-I contract No 987/2012 of the General Secretariat Research and Technology of the Hellenic Government) co-financed by the European Union (European Social Fund—ESF) and Greek National funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program "ARISTEIA" (EXCELLENCE).

References

Certuche-Alzate JP, Vélez-Reyes M (2009) A reconfiguration algorithm for a DC zonal electric distribution system based on graph theory method. In: IEEE Electric Ship Technologies Symposium, ESTS 2009, Art. no. 4906521, pp 235–241

Chen D, Xu L, Yao L (2012) DC network stability and dynamic analysis using virtual impedance method. In: IECON proceedings (industrial electronics conference), pp 5625–5630

Falahi M, Butler-Purry KL, Ehsani M (2013) Reactive power coordination of shipboard power systems in presence of pulsed loads. IEEE Trans Power Syst 28(4):3675–3682

Gjengedal T (1996) Emission Constrained Unit-Commitment (ECUC). IEEE Trans Energy Convers 11(1):132–138

Justo JJ, Mwasilu F, Lee J, Jung J-W (2013) AC-microgrids versus DC-microgrids with distributed energy resources: a review. Renew Sust Energ Rev 24:387–405

Kanellos FD (2014) Optimal power management with GHG emissions limitation in All Electric Ship power systems comprising energy storage systems. IEEE Trans Power Syst 29 (1):330–339

Kanellos FD, Tsekouras J, Prousalidis J (2012) Control system for fuel consumption minimization–gas emission limitation of full electric propulsion ship power systems. Proc Institut Mech Eng M J Eng Maritime Environ 228(1):1–12

Kanellos FD, Tsekouras GJ, Prousalidis J (2015) Onboard DC grid employing smart grid technology: challenges, state of the art and future prospects. IET Electr Syst Transp 5(1):1–11

McCoy TJ, Amy JV Jr (2009) The state-of-the-art of integrated electric power and propulsion systems and technologies on ships. In: IEEE Electric Ship Technologies Symposium, ESTS 2009, Art. no. 4906534, pp 340–344

Planas E, Gil-De-Muro A, Andreu J, Kortabarria I, Martínez De Alegría I (2013) General aspects, hierarchical controls and droop methods in microgrids: a review. Renew Sust Energ Rev 17:147–159

Prousalidis J, Patsios C, Kanellos F, Sarigiannidis A, Tsekouras N, Antonopoulos G (2012) Exploiting shaft generators to improve ship efficiency. In: Proceedings of the electrical systems for aircraft, railway and ship propulsion, pp 1–6

Singhal PK, Sharma RN (2011) Dynamic programming approach for solving power generating unit commitment problem. In: 2011 2nd international conference on computer and communication technology, ICCCT-2011, pp 298–303

Sudhoff SD, Crider JM (2011) Advancements in generalized immittance based stability analysis of DC power electronics based distribution systems. In: 2011 I.E. Electric Ship Technologies Symposium, ESTS 2011, pp 207–212

Sulligoi G, Tessarolo A, Benucci V, Millerani TA, Baret M, Luise F (2013) Shipboard power generation: design and development of a medium-voltage dc generation system. IEEE Ind Appl Mag 19(4):47–55

Tang L, Ooi B-T (2007) Locating and isolating DC faults in multi-terminal DC systems. IEEE Trans Power Delivery 22(3):1877–1884

Tsekouras GJ, Kanellos FD (2013) Optimal operation of ship electrical power system with energy storage system and photovoltaics: analysis and application. WSEAS Trans Power Syst 8 (4):145–155

Tsekouras GJ, Kanellos FD, Prousalidis J (2015) Simplified method for the assessment of ship electric power systems operation cost reduction from energy storage and Renewable energy sources integration. IET Electr Syst Transp 5(2):61–69

Man Diesel & Turbo (2012, August) Basic principles of ship propulsion [Online]. Available http://www.mandieselturbo.com/0000245/Press/Publications/Technical-Papers.html?page=3

Zadeh MK, Zahedi B, Molinas M, Norum LE (2013) Centralized stabilizer for marine DC microgrid. In: IECON proceedings, pp 3359–3363

Zahedi B, Norum LE, Ludvigsen KB (2014) Optimized efficiency of all-electric ships by dc hybrid power systems. J Power Sources 255(341–354):2014

Zhu L, Liu J, Cupelli M, Monti A (2013) Decentralized linear quadratic Gaussian control of multi-generator MVDC shipboard power system with constant power loads. In: 2013 I.E. Electric Ship Technologies Symposium, ESTS 2013, pp 308–313