Fausto Pedro García Márquez
Benjamin Lev *Editors*

Advanced Business Analytics

Springer

Advanced Business Analytics

Fausto Pedro García Márquez • Benjamin Lev
Editors

Advanced Business Analytics

Springer

Editors
Fausto Pedro García Márquez
ETSI Industriales de Ciudad Real
University of Castilla-La Mancha
Ciudad Real, Spain

Benjamin Lev
Decision Sciences Department,
LeBow College of Business
Drexel University
Philadelphia, Pennsylvania, USA

ISBN 978-3-319-11414-9 ISBN 978-3-319-11415-6 (eBook)
DOI 10.1007/978-3-319-11415-6
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014959092

© Springer International Publishing Switzerland 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

The Advanced Business Analytics synthesizes analytic principles with business and provides an interface between the main disciplines of engineering/economics and the organizational, administrative, and planning abilities of management. It is also complementary to other disciplines such as finance, marketing, decision, and risk analysis. In this book, each chapter discusses different topics in Advanced Business Analytics.

This book is intended for engineers, economists, and researchers who wish to know more about the new discipline of Business Analytics or who apply Business Analytics tools in their profession. The authors who contributed the various 12 chapters of this edited volume describe their original work in these areas or provide material for case studies, which have been successfully implemented in real-life cases.

The first chapter presents Decision Making (DM) with a set of criteria for choosing among several alternatives. The diagram of a DM is performed by logical decision trees (LDTs). This is achieved by having a graphical representation of those cases that need to be improved. Applying Binary Decision Diagrams (BDDs) reduces the computational cost of quantitative analysis. LDT graphically describes the sources of certain problems and their interrelations. The size of the BDDs depends on the number of variables, thus affecting the computational cost of the analysis. A real case study is discussed here, which turns out to be an NP-hard nonlinear programming problem. The necessary conditions of optimality are defined by Karush–Khun–Tucker conditions and we find the optimal allocation when resources are limited.

Chapter 2 focuses on the design of an efficient algorithm to reduce inventory costs and its eventual integration as an add-on in SAP Business One platform. The problem arises in those firms with storage capacity that replenish orders for several items along a finite planning horizon. After 2 years, the algorithm is finally implemented in Microsoft C# and .NET. A comprehensive computational experiment is carried out considering a wide range of random cases. The experimental results revealed that heuristic solutions are on average 5 % better than CPLEX optimizers.

The rising wind energy industry and the increasing number of failures of the large wind turbines necessitate the reduction of costs in this industry and increase competition in this sector. For this reason, the wind energy industry is focused on the reduction of operation and maintenance (O&M) costs, which is discussed in Chap. 3. Condition Monitoring Systems (CMS) are probably the most effective approach to minimize O&M costs and substantially improve the availability, reliability, and safety of wind turbines by early detection of the problems. CMS requires knowledge and expertise to analyze the large volume of data collected from the sensors located in the wind turbines. The main objective of Chap. 3 is the development of a life cycle cost (LCC) model of the CMS for a wind turbine and the analysis of its economic feasibility. The LCC model has been applied to a real case study in Germany.

Multiple-Attribute Decision Making (MADM) refers to making decisions when there are multiple alternatives and multiple criteria. MADM considers a problem where management needs to prioritize or rank order alternative choices: identify key nodes in a business network, pick a contractor/sub-contractor, choose airports, rank recruiting efforts, rank banking facilities, or rank schools/colleges. How does one proceed to accomplish this analytically? In Chap. 4, four methodologies are presented to rank order alternatives based upon multiple criteria. These four methodologies include Data Envelopment Analysis (DEA), Simple Average Weighting (SAW), Analytical Hierarchy Process (AHP), and Technique of Order Preference by Similarity to Ideal Solution (TOPSIS). We describe all four methods and their uses, discuss the strengths and limitations of each method, present suggestions for sensitivity analysis, and present illustrative examples.

Business Analytics and Intelligence tools (BAI) are spreading across all industries. As the amount of business data grows exponentially, it is critical to have appropriate tools that make it possible to comprehend and utilize this digital era. Even though BAI tools have positively evolved in this direction, meaningful and productive use of data still remains a major challenge for many organizations. It has been demonstrated that BAI technologies should evolve toward a more holistic approach in which business users can focus on business concepts and questions, reducing time for data manipulation. Chapter 5 proposes a Business Analytics Architecture (BAA) as the infrastructure supporting "smart" enterprise BAI operations. It enables users to define the business concepts they want to focus on, as well as connecting them with data at storage-level. Analytical and data-mining algorithms are intensively exploited, all guided by the "semantic layer" previously depicted by business users. BAA integrates up-to-date data-mining and artificial intelligence techniques as well as some well-known business practices such as Balanced Scorecard and Strategy Maps.

Considering the reliability of an area traffic control road network, most travel delays are directly dependent on correct operation of signal settings. The purpose of Chap. 6 is to devise an efficient scheme to evaluate the reliability of a signal-controlled road network. A min–max complementarity problem is proposed to characterize user equilibrium flow in the presence of a worst-case disruption of a given link capacity loss. A computationally tractable solution scheme is proposed to identify important links whose disruption could cause a substantial increase in travel delays to all road users. Numerical computations are conducted in a medium-sized road network to demonstrate the feasibility of the proposed solution scheme. The results indicate that the most critical signal-controlled traffic streams can be conveniently identified and, when failed to perform their normal functions, would give rise to maximum travel delays to all road users.

Business development requires technologically reliable products developed through excellence in project management. This excellence is obtained with advanced project control that should include Earned Value Management (EVM) methodology and project objectives accomplishment orientation. Chapter 7 presents an analytical model to implement the project objectives control with the EVM. The objectives are periodically measured along the project life with their weights and impact at the end of the project affecting the earned value. This provides early indications for the project progress and results and highlights the possible need for corrective actions. The analytical model also considers the forecasting of the objectives and their limits at project completion. The model is applied to a case study of engineering product development focusing on obtaining technical objectives. The generality of the model permits to apply it to any type of project, small or large, in any industry and with any kind of objectives either technical, commercial, economic, etc.

Chapter 8 describes a new point of view for the concept of a maturity model, introducing new evaluation elements beyond the concept of process. Maturity models can be defined as approaches to improving the business process management of any organization, or, from a global system-wide perspective, the organizational processes. The rationale behind the concept of maturity is that the organizations with experience carry out tasks systematically, where immature organizations achieve their outcomes as a result of considerable efforts of individuals using approaches that they create more or less spontaneously. Nowadays, maturity models are indissolubly linked with the process maturity models, as conceptual models compare the maturity of an organization's current practices against an industry standard. These models support any organization in setting priorities for improving the product/service operations in order to find the optimum levels of efficiency and effectiveness. On the other hand, excellence models provide a holistic view of the organization and can be used to determine how these different methods fit together and complement each other. Any excellence model can therefore be used in conjunction with any number of these tools, including Business Process Management, based on the needs and function of the organization, as an overarching framework for developing sustainable excellence. In fact, one of the eight fundamental concepts of the European Foundation for Quality Management (EFQM) excellence model is "managing with agility," a principle based on how excellent organizations are managed through structured and strategically aligned processes using fact-based decision making to create balanced and sustained results. The structure of excellence models, in the case of the EFQM excellence model, includes the following criteria: fundamental concepts of excellence, model criteria, and RADAR Logic. They allow introducing a more complex concept about maturity in the management, and not only in processes.

Chapter 9 gives an overview of the Analytic Hierarchy Process (AHP) and Intuitionistic Fuzzy TOPSIS (IFT) methods. It deals with an evaluation methodology based on the AHP-IFT where the uncertainties are handled with linguistic values. First, the supplier selection problem is formulated using AHP and, then, it is used to determine the weights of the criteria. Later, IFT is used to obtain full ranking among the alternatives based on the opinion of the Decision Makers. The present model provides an accurate and easy classification in supplier attributes by chains prioritized in the hybrid model. A numerical example is given to clarify the main result that is developed in this chapter.

Chapter 10 intends to be a small guide to investing in hedge funds or a guide to consider when trying to understand the complexities of investing in hedge funds. It presents a general overview of the field of alternative investments and introduces its complexities, since investing in hedge funds is not a simple task. Alternative investment vehicles have taken an important role, not only in the diversification of portfolios but also as standalone investments. Capturing the entire risk dimensions implied in hedge fund investment strategies is paramount in understanding alternative investments.

The chapter also analyzes the problem of the lack of benchmarks. Should the performance of a hedge fund follow a benchmark? As absolute return vehicles, alternative investment organizations should not be benchmarked. However, investors need to compare returns with other assets in order to properly assess the opportunity cost implied in hedge fund investing. This chapter considers benchmarking among peers as a new and latest strategy.

Chapter 11 introduces functional data analysis, a relatively new technique for data analysis in business analytics. The distinct feature of this technique is that it deals with smooth functions or processes, which generate a discretized data sample that can be observed. The functional approach allows the user to flexibly model system dynamics, to analyze observations with measurement errors, and to fit data with sparse observations. This chapter illustrates the application of functional data analysis in the capital structure of California hospitals. It also points out the functional data analysis application in business research and future research directions.

Finally, Chap. 12 presents a software application developed to optimize the annual planning of visits of a brewing company sales/marketing staff to their customers. Each of these customers must be annually visited a provided number of times. Therefore, each salesperson is assigned to a set of customers that must be visited each week. The application assigns all visits of a salesperson to each customer so that all weeks should have roughly the same number of visits. By virtue of this approach, the brewing company diminished their marketing operating costs, as well as improved their customer relationships.

Ciudad Real, Spain — Fausto Pedro García Márquez
Philadelphia, Pennsylvania — Benjamin Lev

Contents

About the Editors

Fausto Pedro García Márquez got the European Doctorate on Engineering at the University of Castilla-La Mancha (UCLM), 2004. He received his Engineering Degree from the University of Murcia, Spain, in 1998; Technical Engineer Degree at UCLM in 1995; and the degree in Business Administration and Management at UCLM 2006. He also holds the titles of Supper Technician in Labor Risks Prevention by UCLM 2000 and Transport Specialist at the Polytechnic University of Madrid, Spain, 2001. He has been Senior Manager in Accenture 2013/2014. He is Senior Lecture (Full professor accredited) at UCLM, and Honorary Senior Research Fellow at Birmingham University UK, Lecturer in Instituto Europeo de Postgrado since 2014, and Director of Ingenium Research Group. He has been principal investigator in 3 European Projects, and 60 national and company research projects. He has international and national patents, more than 110 international papers, 12 books, and 3 International Awards in Engineering Management and Management Science (www.uclm.es/profesorado/fausto).

Benjamin Lev is Professor and Head of Decision Sciences at LeBow College of Business. He holds a PhD in Operations Research from Case Western Reserve University. Prior to joining Drexel University, Dr. Lev held academic and administrative positions at Temple University, University of Michigan-Dearborn, and Worcester Polytechnic Institute. He is the Editor-in-Chief of OMEGA-The International Journal of Management Science, is the Co-Editor-in-Chief of International Journal of Management Science and Engineering Management, and serves on several other journal editorial boards. He has published ten books and numerous articles and has organized many national and international conferences.

Decision Making Approach for Optimal Business Investments

Alberto Pliego Marugán and Fausto Pedro García Márquez

1 Introduction

Decision Making (DM) is defined as a criteria selection method, e.g. when a person wakes up in the morning and decides to bike instead of driving to work. These decisions can be made in the short, medium and long term. All the firms need to discern about the decisions that are really important or not from a particular perspective. A correct decision-making in any firm can become a competitive advantage.

In a DM scenario there is an event that is, or not, desired, which entail a path among the different alternatives that lets to the firm get the objective [1]. Generally it is required the optimal situation, i.e. the one which will provide the best results from the profit, cost, safety,... point of view [2]. It is possible employing a correct criterion for choosing the best scenario [3].

DM is defines by Harris as [4]:

> *The research of identifying and choosing alternatives based on the decision-maker weighs/ values and preferences. Make a decision entail there are several alternatives to be considered, and not only identify the alternatives is sought but also to choose the one that best fits the aims, restrictions, etc. is desired.*

DM consists on the transformation process from data to proceedings [5]. Data collection becomes a strategy task for DM in order to get the proceedings, and they help to keep improving the DM.

There are several alternatives to classify the decisions performed in a business. Figure 1 shows a classification of the decisions that can be made in a business environment.

Operational decisions aim to reach the strategic decisions. Wrong operational decisions have not far-reaching implications for the future and may be fixed easily.

Tactical decisions occur with more frequency than the operational decisions. They are control by a procedure and routines. Any wrong tactical decisions may bring troubles to the business.

A.P. Marugán (✉) • F.P.G. Márquez
Ingenium Research Group, University of Castilla-La Mancha, La Mancha, Spain
e-mail: Alberto.Pliego@uclm.com; FaustoPedro.Garcia@uclm.es

© Springer International Publishing Switzerland 2015

F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_1

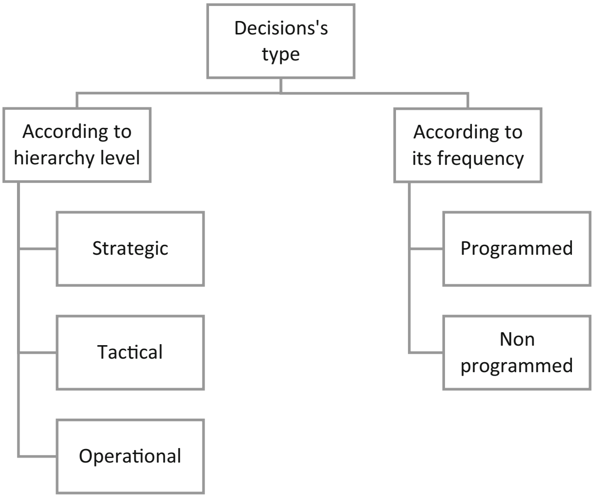


Fig. 1 Decision's classification

Strategic decisions are done in long period, where there is not enough dataset and it is critical in the future of the business, having a wrong strategic decisions fatal consequences.

According to the time period, programmed decisions are those repeated frequently in a business, where a procedure is developed to carry out every time it occurs. On the other hand, non-programmed decisions are those emerged unexpectedly. Usually involve a high degree of difficulty and they must be tackled by experts in the field.

The advances in the technology and information help to the firms to develop their own software in order to find good DM. Nevertheless it is recommended to consider algorithms in order to find the optimal DM.

2 DM Process

The diagram of a DM can be performed by decision trees. This is achieved having a graphical representation of those situations that need to be improved. Although measures of importance can be applied to decision trees, the use of Binary Decision Diagrams (BDD) involves a reduction in the computational cost for quantitative resolution, among other improvements. In addition, the cut sets obtained from the BDD will be the basis for the construction of the heuristic method developed for the analysis proposed in this chapter.

The following main scenarios can be distinguished according to the information available in the DM process:

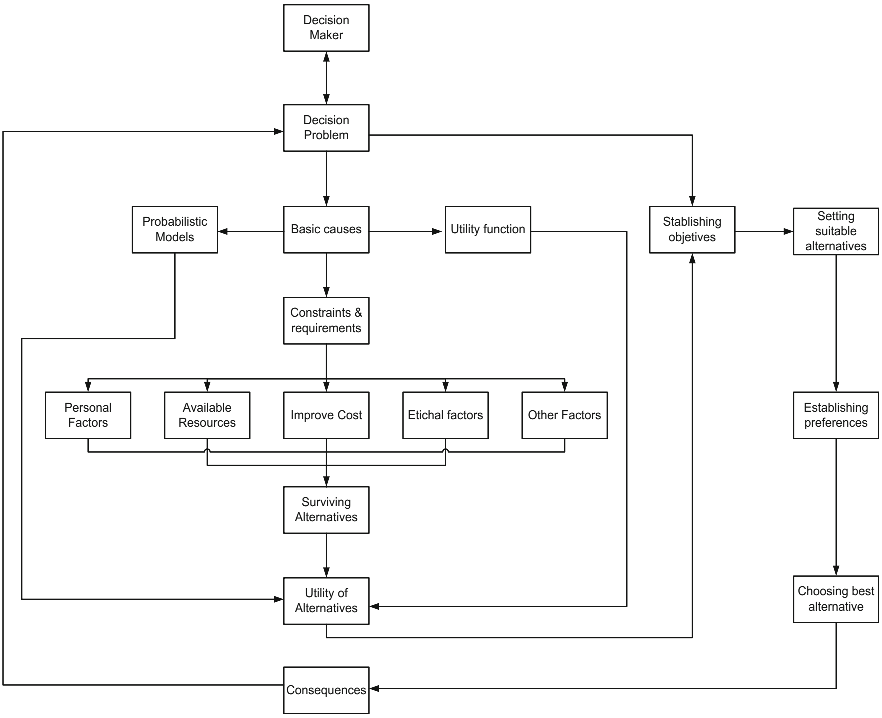


Fig. 2 DM process

– DM under certainty: The problem is entirely known, i.e. all possible states for a basic cause (BC) are known and any consequences of each decision can be completely achieved.
– DM under risk: Implies partial information and some are stochastic. This will be the scenario considered in this chapter.
– DM under uncertainty: Information about the main problem (MP) and BCs is not complete and part of the information is missed.

Figure 2 shows a flow chart regarding to the main steps for a decision maker considering rational and logical arguments that support their decision.

2.1 *The Decision Maker*

The decision maker is the person, system or organization that takes a decision. All decisions and assessments will be influenced by the decision makers. Any decision-maker should be essential to have some skills that can be resumed in experience, good judgment, creativity and quantitative skills. The first three skills are personals, and the last one is supported by existing methods and support systems for DM in order to choose, considering different scenarios, and to help to the decision maker to decide about best DM. In this field the paper presents and describes a quantitative method to support the DM process.

2.2 Constraints and Requirements

In any DM process there are constrains or requirements to consider, e.g. existing resources, available budget, environmental precautions, social issues, legal provisions, etc. Generally the constraints are exogenous and not access in order to incorporate to the mathematical or empirical models, i.e. an endogenization of constraints should be carried out in order to consider them. This endogenization involves conceptualizing the constraints as goals of the decision maker, i.e. it is possible to reformulate a constraint as the main objective [6].

This chapter is focused on expected-utility DM under constraints that can be considered as a process that provides a solution for a decision maker with different objectives: To satisfy the constraints and rule out unfeasible solutions, and to maximize utility functions among the surviving options.

2.3 The Utility Function

For a quantitative stochastic DM case, the utility function provides a value that determines the quality of the solution that is being considered. It is formulated as an analytical expression derived from LDT to BDD conversion. Thresholds might set to be establishing in order to determine the solutions that are most suitable for the objectives of the decision maker.

2.4 Results

Once a decision is made, then it is necessary to choose the best alternative or scenario considering variables as feasibility and costs for implementing the decisions. None of DM processes is completely reliable due to the possibility to take into account the total range of events involved in the solution, and also in an a posteriori evaluation of consequences. Particularly, evaluation of consequences is essential for improving those DM processes with data from forecasting studies. The results derived from a decision can affect the complex structure of the problem, or to modify some features of constraints and requirements. Feedback is necessary in order to determine the quality of the decision because the decision maker requires knowing if the system responds as expected. Moreover, there are some decisions that need to be done periodically and a feedback is needed in order to improve the new decision quality according to the previous decision.

3 Logical Decision Trees

DM process is carried out when a certain problem occurs, with the objective of discerning whether there is a real problem [7]. Logical Decision Trees (LDT) describes graphically the roots and causes of a certain problem and their interrelation. The logical operators 'AND' and 'OR' are employed in order to connect the events considered [8].

Figure 3 shows a LDT composed of seven non-basic causes and nine basic causes. BCs are those causes that are not possible to be broken down into simpler causes. All these causes are linked by logical gates, in particular by 1 'OR' gates and 3 'AND' gates. LDT provides information about the critical states of BCs and how MP is usually generated. Figure 3 shows that BC7 is one of the most important causes, i.e. if BC7 occurs then MP will occur.

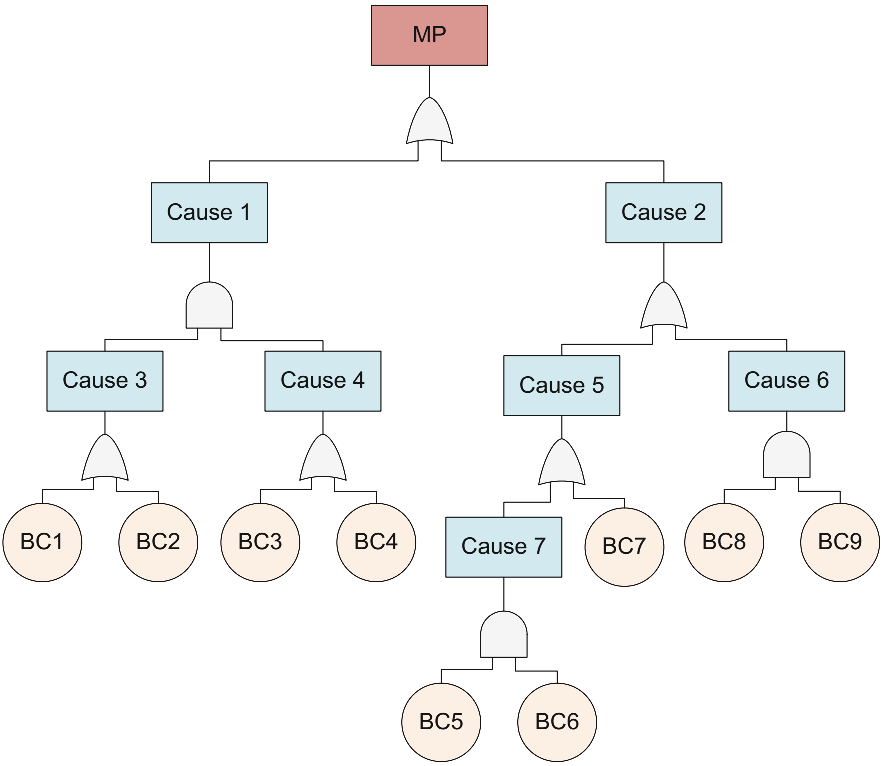


Fig. 3 Logical decision tree

When LTD is larger, e.g. composed of hundreds or thousands of BCs, there are also larger alternatives for a direct decision tree analysis. A direct analysis can be done by a LDT to BDD conversion. An "if-then-else" conversion approach is described in reference [9] in order to carry out this conversion.

4 Binary Decision Diagrams

BDDs have been successfully found in the constant search for an efficient way to simulate LDTs. BDDs were introduced by Lee [10], and further popularized by Akers [11], Moret [12] and Bryant [13]. The BDD is used in order to analyze the LDT. They are composed by a data structure that represents the Boolean functions. They provide a mathematical approach to the problem by Boolean algebra, such as Karnaugh maps or truth tables, being less complex than its truth table.

BDD is a directed graph representation of a Boolean function where equivalent Boolean sub-expressions are uniquely represented [14]. A directed acyclic graph is a directed graph, i.e. to each vertex v, there is no possible directed path that starts and finishes in v. It is composed of some interconnected nodes in a way that each node has two vertices. Each vertex is possible to be a terminal or non-terminal vertex. BDD is a graph-based data structure whereby the occurrence probability of a certain problem in a DM is possible to be achieved. Each single variable has two branches: 0-branch corresponds to the cases where the variable is 0 and it is graphically represented by a dashed line (Fig. 7); on the other hand, 1-branch cases are those where the event is being carried out and corresponds to the variable with a value of 1, and it is represented by a solid line (Fig. 7).

It will allow obtaining an analytical expression depending on the occurrence probability and the logical structure of the tree of every single basic cause. Paths starting from the top BC to a terminal one provide a certain state in which MP will occur. These paths are named cut-sets (CS).

4.1 Ranking of the Basic Causes

The size of the BDD, as well as CPU runtime, have an awfully dependence on the variable ordering. Different ranking methods can be used in order to reduce the number of cut-set, and consequently, to reduce the CPU runtime. It must be emphasized that there is not any method that provide the minimum size of BDD in all cases [22]. The main methods are described in this section.

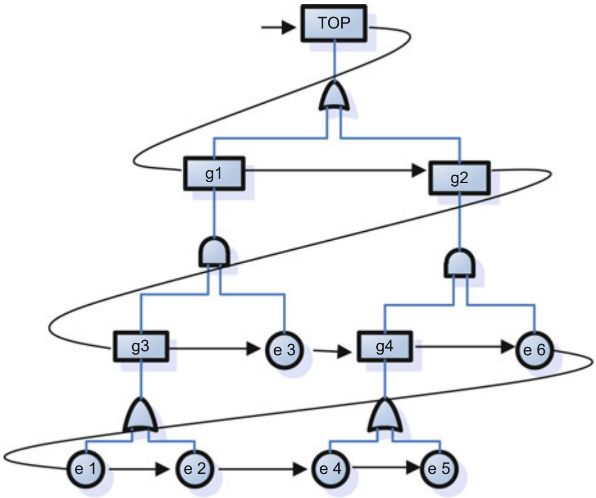


Fig. 4 TDLR ranking method

4.1.1 Top-Down-Left-Right (TDLR)

This method generates a ranking of the events by ordering them from the original FT structure in a top-down and then left-right manner [15]. The listing of the events is initialized, at each level, in a left to right path adding the basic events found in the ordering list. In case that any event had been considered previously and it was located higher up the then it is ignored (Fig. 4).

The ranking for the example showed in Fig. 3 using the TDLR method is:

$$BC_1 > BC_2 > BC_3 > BC_4 > BC_7 > BC_8 > BC_9 > BC_5 > BC_6$$

4.1.2 Depth First Search (DFS)

This approach goes from top to down of a root and each sub-tree from left to right (see Fig. 5). This procedure is a non-recursive implementation and all freshly expanded nodes are added as last-input last-output process [16].

The ranking for the example presented in Fig. 3 is:

$$BC_1 > BC_2 > BC_3 > BC_4 > BC_7 > BC_5 > BC_6 > BC_8 > BC_9$$

4.1.3 The Breath First Search (BFS)

This algorithm begins ordering all the basic events obtained expanding from the standpoint by the first-input first-output procedure (Fig. 6). The events not considered are added in a queue list named “open” that is recalled “closed” list when the all the events are studied [17].

The ranking for the example of Fig. 3 is:

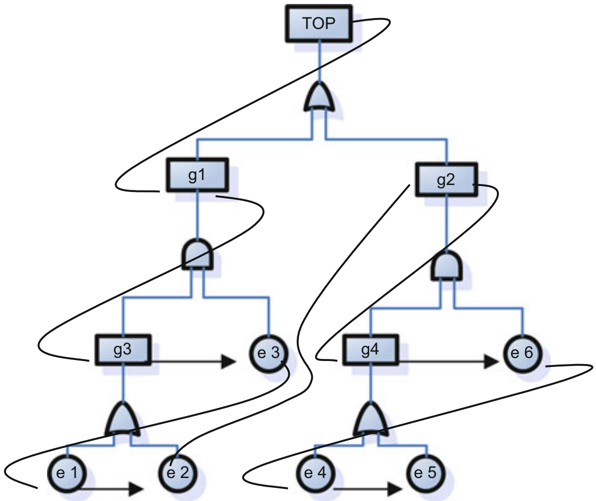


Fig. 5 DFS ranking method

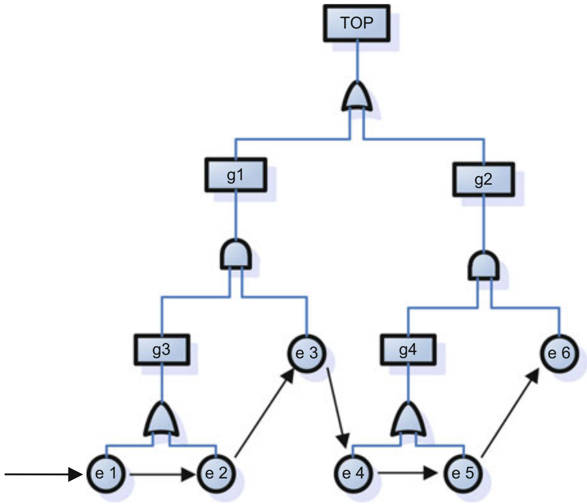


Fig. 6 BFS ranking methods

$$BC_1 > BC_2 > BC_3 > BC_4 > BC_5 > BC_6 > BC_7 > BC_8 > BC_9$$

4.1.4 Level Method

The level of any event is understood as the number of the gates that has higher up a tree until the top event. The "level" method creates the ranking of the events according to the level of them. In case that two or more events have the same

Table 1 Level method

Events.	Level
BC_1	3
BC_2	3
BC_3	3
BC_4	3
BC_5	4
BC_6	4
BC_7	3
BC_8	3
BC_9	3
BC_{10}	3

level, the event will have higher priority if it appears early in the tree [15]. Table 1 shows the level of the events of the LDT showed in Fig. 3.

Therefore, according to the Table 1 and the Level method, the ranking obtained is:

$$BC_1 > BC_2 > BC_3 > BC_4 > BC_7 > BC_8 > BC_9 > BC_5 > BC_6$$

It can be observed that in this case the ranking proposed by this method is the same that TDLR method, therefore, the CSs obtained will be the same as well.

4.1.5 AND Method

Xie et al. [18] suggest by the AND criterion that the importance of the basic event is based on the "and" gates that are between the k event and the top event, because in FTA the "and" gates imply that there are redundancies in the system. Consequently, basic events under an "and" gate can be considered less important because it is independent to other basic events occurring for the intermediate events [18]. Furthermore:

- Basic events with the highest number of "and" gates will be ranked at the end.
- In case of duplicated basic events, the event with less "and" gates has preference.
- Basic events with the same number of "and" gates can be ranked as the TDLR method approach.

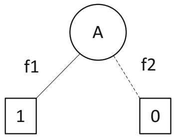
The ranking for the example showed in Fig. 3 using the AND method is:

$$BC_7 > BC_1 > BC_2 > BC_3 > BC_4 > BC_8 > BC_9 > BC_5 > BC_6$$

4.2 BDD Conversions

Small BDDs for calculated the MP occurrence probability is possible to be done manually, but when larger LDTs have to be converted it is almost impossible.

Fig. 7 ite applied to BDD



Ite (If-Then-Else) conditional expression is employed in this research work as an approach for the BDD's cornerstones, based on the approach presented in [19]. Figure 7 shows an example of an ite done in a BDD.

Which could be described as: "**If A** variable occur, Then **f1**, Else **f2**" [20]. The solid line always belongs to the ones as well as the dashed lines to the zeros, above explained.

Considering the Shannon's theorem is obtained the following expression from Fig. 7.

$$f = b_i \cdot f_1 + \overline{b}_i \cdot f_2$$

where

$$f = b_i \cdot f_1 + \overline{b}_i \cdot f_2 = ite(b_i, f_1, f_2)$$

Table 2 shows the different CSs obtained using the abovementioned ranking methods. A comparison between the CSs in done in order to analyze the efficiency of the methods ordering the basic causes showed in Fig. 3.

There is not a significant difference between these methods because the sizes of all the CSs are similar (Table 2). The main reason is because the LDT that is being analyzing does not have a large number of events. In Table 2 can be observed that the method that generates a lowest number of CSs is the AND method. Therefore it will be chose in this research work.

Figure 8 shows the BDD obtained from the LDT (Fig. 3) to BDD conversion using AND method.

4.3 Analytical Expression

The probability of occurrence must be assigned to each BC. $P(BCi)$ is the probability of occurrence of the ith BC. $P\left(\overline{BCi}\right)$ is the probability of non-occurrence of the ith BC. Therefore:

$$P\left(\overline{BCi}\right) = 1 - P(BCi)$$

The probability of occurrence of jth CS ($P(CSj)$) can be calculated as the product of $P(BCi)$ and $P\left(\overline{BCi}\right)$ that compose the CS. Probability of occurrence of the MP (QMP) is given by:

Table 2 Cut sets from ranking methods for the LDT shows in Fig. 3

TDLR method
$CS_1 : BC_1 \cdot BC_3$
$CS_2 : BC_1 \cdot \overline{BC_3} \cdot BC_4$
$CS_3 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_7$
$CS_4 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$
$CS_5 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_8 \cdot \overline{BC_9} \cdot BC_5 \cdot BC_6$
$CS_6 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \cdot \overline{BC_7} \cdot \overline{BC_8} \cdot BC_5 \cdot BC_6$
$CS_7 : \overline{BC_1} \cdot BC_2 \cdot BC_3$
$CS_8 : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot BC_4$
$CS_9 : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_7$
$CS_{10} : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$
$CS_{11} : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_8 \cdot \overline{BC_9} \cdot BC_5 \cdot BC_6$
$CS_{12} : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot \overline{BC_8} \cdot BC_5 \cdot BC_6$
$CS_{13} : \overline{BC_1} \cdot \overline{BC_2} \cdot BC_7$
$CS_{14} : \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$
$CS_{15} : \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_7} \cdot BC_8 \cdot \overline{BC_9} \cdot BC_5 \cdot BC_6$
$CS_{16} : \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_7} \cdot \overline{BC_8} \cdot BC_5 \cdot BC_6$
DFS method
$CS_1 : BC_1 \cdot BC_3$
$CS_2 : BC_1 \cdot \overline{BC_3} \cdot BC_4$
$CS_3 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_7$
$CS_4 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_5 \cdot BC_6$
$CS_5 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_5 \cdot \overline{BC_6} \cdot BC_8 \cdot BC_9$
$CS_6 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot \overline{BC_5} \cdot BC_8 \cdot BC_9$
$CS_7 : \overline{BC_1} \cdot BC_2 \cdot BC_3$
$CS_8 : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot BC_4$
$CS_9 : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_7$
$CS_{10} : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_5 \cdot BC_6$
$CS_{11} : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_5 \cdot \overline{BC_6} \cdot BC_8 \cdot BC_9$
$CS_{12} : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot \overline{BC_5} \cdot BC_8 \cdot BC_9$
$CS_{13} : \overline{BC_1} \cdot \overline{BC_2} \cdot BC_7$
$CS_{14} : \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_7} \cdot BC_5 \cdot BC_6$
$CS_{15} : \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_7} \cdot BC_5 \cdot \overline{BC_6} \cdot BC_8 \cdot BC_9$
$CS_{16} : \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_7} \cdot \overline{BC_5} \cdot BC_8 \cdot BC_9$
BFS method
$CS_1 : BC_1 \cdot BC_3$
$CS_2 : BC_1 \cdot \overline{BC_3} \cdot BC_4$
$CS_3 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_5 \cdot BC_6$
$CS_4 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_5 \cdot \overline{BC_6} \cdot BC_7$
$CS_5 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_5 \cdot \overline{BC_6} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$
$CS_6 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_5} \cdot BC_7$
$CS_7 : BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_5} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$
$CS_8 : \overline{BC_1} \cdot BC_2 \cdot BC_3$
$CS_9 : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot BC_4$
$CS_{10} : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_5 \cdot BC_6$
$CS_{11} : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_5 \cdot \overline{BC_6} \cdot BC_7$
$CS_{12} : \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_5 \cdot \overline{BC_6} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$

(continued)

Table 2 (continued)

TDLR method
$CS_{13}: \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_5} \cdot BC_7$
$CS_{14}: \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_5} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$
$CS_{15}: \overline{BC_1} \cdot \overline{BC_2} \cdot BC_5 \cdot BC_6$
$CS_{16}: \overline{BC_1} \cdot \overline{BC_2} \cdot BC_5 \cdot \overline{BC_6} \cdot BC_7$
Level method
$CS_1: BC_1 \cdot BC_3$
$CS_2: BC_1 \cdot \overline{BC_3} \cdot BC_4$
$CS_3: BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_7$
$CS_4: BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$
$CS_5: BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_8 \cdot \overline{BC_9} \cdot BC_5 \cdot BC_6$
$CS_6: BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot \overline{BC_8} \cdot BC_5 \cdot BC_6$
$CS_7: \overline{BC_1} \cdot BC_2 \cdot BC_3$
$CS_8: \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot BC_4$
$CS_9: \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_7$
$CS_{10}: \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$
$CS_{11}: \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot BC_8 \cdot \overline{BC_9} \cdot BC_5 \cdot BC_6$
$CS_{12}: \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_7} \cdot \overline{BC_8} \cdot BC_5 \cdot BC_6$
$CS_{13}: \overline{BC_1} \cdot \overline{BC_2} \cdot BC_7$
$CS_{14}: \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_7} \cdot BC_8 \cdot BC_9$
$CS_{15}: \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_7} \cdot BC_8 \cdot \overline{BC_9} \cdot BC_5 \cdot BC_6$
$CS_{16}: \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_7} \cdot \overline{BC_8} \cdot BC_5 \cdot BC_6$
AND method
$CS_1: BC_7$
$CS_2: \overline{BC_7} \cdot BC_1 \cdot BC_3$
$CS_3: \overline{BC_1} \cdot BC_1 \cdot \overline{BC_3} \cdot BC_4$
$CS_4: \overline{BC_1} \cdot BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_8 \cdot BC_9$
$CS_5: \overline{BC_7} \cdot BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_8 \cdot \overline{BC_9} \cdot BC_5 \cdot BC_6$
$CS_6: \overline{BC_7} \cdot BC_1 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_8} \cdot BC_5 \cdot BC_6$
$CS_7: \overline{BC_7} \cdot \overline{BC_1} \cdot BC_2 \cdot BC_3$
$CS_8: \overline{BC_7} \cdot \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot BC_4$
$CS_9: \overline{BC_7} \cdot \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_8 \cdot BC_9$
$CS_{10}: \overline{BC_7} \cdot \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot BC_8 \cdot \overline{BC_9} \cdot BC_5 \cdot BC_6$
$CS_{11}: \overline{BC_7} \cdot \overline{BC_1} \cdot BC_2 \cdot \overline{BC_3} \cdot \overline{BC_4} \cdot \overline{BC_8} \cdot BC_5 \cdot BC_6$
$CS_{12}: \overline{BC_7} \cdot \overline{BC_1} \cdot \overline{BC_2} \cdot BC_8 \cdot BC_9$
$CS_{13}: \overline{BC_7} \cdot \overline{BC_1} \cdot \overline{BC_2} \cdot BC_8 \cdot \overline{BC_9} \cdot BC_5 \cdot BC_6$
$CS_{14}: \overline{BC_7} \cdot \overline{BC_1} \cdot \overline{BC_2} \cdot \overline{BC_8} \cdot BC_5 \cdot BC_6$

$$Q_{MP} = \sum_{j=1}^{n} P(CSj)$$

where n is the total number of CSs.

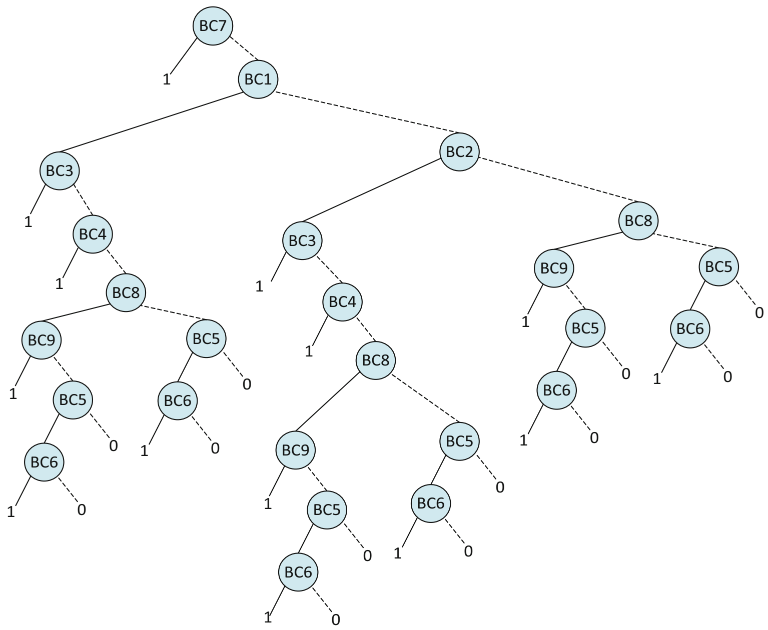


Fig. 8 BDD for LDT shows in Fig. 3

5 Optimization Approach in DM Process

When the probability of occurrence of the MP is achieved, the new objective is to minimize it [21]. It is assumed that LDT will be fixed, and therefore the reduction of the QMP will be performed by the different BCs. Given a BC, the objective is to determinate the investment on it in order to reduce its probability of occurrence, considering the rest of BCs probabilities and the total investment, being the objective function to minimize the probability of occurrence of the top event QMP. A new variable that consider this reduction is defined, being:

$$\mathbf{Imp(BC)} = [Imp(BC_1), Imp(BC_2), \ldots Imp(BC_i) \ldots Imp(BC_n),]$$

The ith component of **Imp**(**BC**) provides the reduction of the probability of occurrence when some resources are assigned to the ith BC. In addition, a probability vector is defined as:

$$\mathbf{P(BC)} = [P(BC_1), P(BC_2), \ldots P(BC_i) \ldots P(BC_n)]$$

The ith component of **P**(**BC**) provides the probability of occurrence of the ith BC. Once the BCs have been improved, the new probability assignment is calculated as the difference between its probability of occurrence and its Imp:

$$\mathbf{P}^*(\mathbf{BC}) = P^*(BC_1), P^*(BC_2), \ldots P^*(BC_i) \ldots P^*(BC_n)$$
$$= \left[P(BC_1) - Imp(BC_1), P(BC_2) - Imp(BC_2), \ldots P(BC_i)\right.$$
$$\left. - Imp(BC_i) \ldots P(BC_n) - Imp(BC_n)\right]$$

The BDD evaluated using **P**(**BC**) provides the value of QMP. If it is being evaluated using, **P***(**BC**), the data obtained will be termed as Q^*_{MP}. It would be desired that $Q_{MP} \geq Q^*_{MP}$, otherwise the optimization procedure is producing wrong outcomes.

The analytic expression provided by BDD will be an optimization function when is evaluated employing **P***(**BC**). The optimization function will be calculated by $Q_{MP}(Imp)$.

BCs are not necessarily corrigible but almost always improvable, therefore **Imp** (**BC**) will range between 0 and a certain threshold. The first constrain is defined as:

$0 \leq Imp(BC_i) \leq a_i$, where a indicates the maximum improvement that can be done in the ith BC. The a_i values will be:

$$0 \leq a_i \leq P(BC_i)$$

$a_i = P(BC_i)$ means that the ith BC is capable to be totally corrected due to this BC allows its own probability of occurrence to be zero (in this case BC_i will not continue contributing to the MP occurrence). If $a_i = 0$ then the improvements in the ith BC are not possible.

The improvement cost (IC) is defined for each BC in order to adequate a quantitative analysis to the nature of BCs, where a high IC for a BC means that a large amount of resources must be invested in order to reduce the probability of occurrence of BC.

IC(**BC**) = $[IC(BC_1), IC(BC_2) \ldots IC(BC_i) \ldots IC(BC_n)]$, where **IC**(**BC**$_\mathrm{i}$) indicates the amount of resources invested in BCi for reducing the probability of occurrence of BCi from 1 to 0.

A new variable is defined as the total amount of resources at the time of the investment operation (Bg), where:

$$\sum_{i=n}^{i=1} IC(BC_i) \cdot Imp(BC_i) \leq Bg$$

The optimization problem is defined as:

$$
\begin{aligned}
&\textit{minimize} \ \ QMP(Imp) \\
&\textit{subject to} \quad \sum_{i=n}^{i=1} IC(BC_i) \cdot Imp(BC_i) \leq Bg \\
&\qquad\qquad Imp(BC_i) - a_i \leq 0 \\
&\qquad\qquad - Imp(BC_i) \ \leq 0
\end{aligned}
$$

This is a Non-Linear Programming Problem (NLPP) and NP-hard. The necessary conditions of optimality are defined by Karush-Khun-Tucker (KKT) conditions [4].

6 Case Study

A case study is presented in this section. The LDT showed in Fig. 3 is related to a real case study of a confidential firm. Table 3 details the initial conditions.

The probability of occurrence of the MP is 0.4825, calculated by the analytic expression obtained from the BDD in Fig. 8, and according to the probabilities of occurrence showed in Table 3. And the maximum investment (MI) would be:

Table 3 Initial conditions

Description	Notation	Probability of occurrence	*IC* (Monetary units)	Improvement limit (*a*)
BUDGET 350				
Customer complaints	MP	–	–	–
Errors from the quality control department	Cause 1	–	–	–
Lack of training	Cause 2	–	–	–
No internal training is provided	BC1	40 %	500	90 %
Lack of necessary equipment	BC2	40 %	150	50 %
Failures in protocols	Cause 3	–	–	–
Review protocols are not well defined	BC3	20 %	400	40 %
Absenteeism training courses	BC4	30 %	400	90 %
Errors from the logistics department	Cause 4	–	–	–
Problems in the distribution	Cause 5	–	–	–
Poorly optimized distribution	Cause 6	–	–	–
Poorly established routes	BC5	10 %	150	80 %
Forecast Problems	BC6	20 %	200	50 %
Driving inappropriate for fragile products	BC7	25 %	500	60 %
Problems in stores	Cause 7	–	–	–
Improperly stored products	BC8	20 %	300	30 %
Obsolete warehouse	BC9	10 %	900	95 %

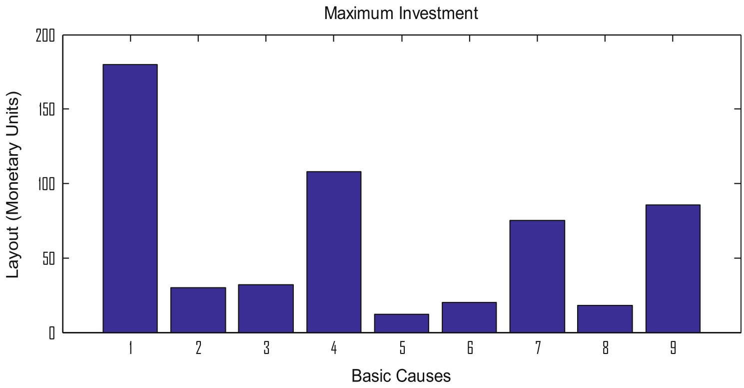


Fig. 9 Maximum investment allowed

$$\begin{aligned} MI &= 0.4 * 0.9 * 500 + 0.4 * 0.5 * 150 + 0.2 * 0.4 * 400 + 0.3 * 0.9 * 400 + 0.1 \\ &\quad * 0.8 * 150 + 0.2 * 0.5 * 200 + 0.25 * 0.6 * 500 + 0.2 * 0.3 * 300 + 0.1 \\ &\quad * 0.95 * 900 \\ &= 560.5 \end{aligned}$$

Figure 9 shows the maximum investment for each BC.

If this maximum budget were available, then the minimum probabilities of occurrence (P_{min}) of the BCs would be:

$$P_{min} = [0.04 \quad 0.20 \quad 0.12 \quad 0.03 \quad 0.020 \quad 0.10 \quad 0.10 \quad 0.14 \quad 0.005]$$

And, consequently, the minimum QMP is:

$$Q_{MP} = 0.1329$$

Therefore, when the available budget is less than 560.5 monetary units (µm), it will be necessary to choose which BCs are the best to be improved in order to minimize the QMP.

The budget considered in the following example is 350 µm. For this purpose, the objective function, subjected to the constraints defined by budget and improvement limits, will be:

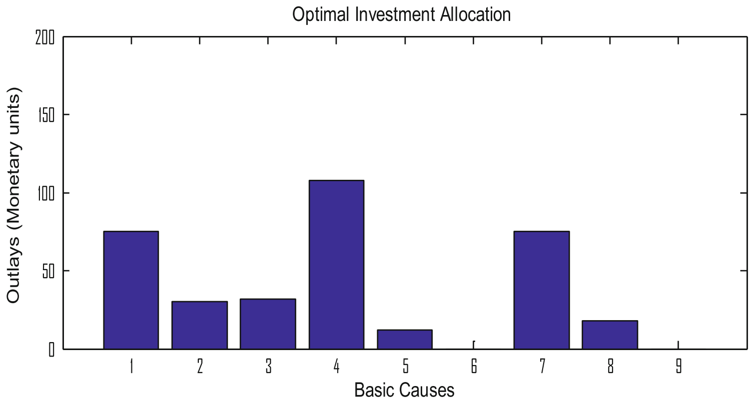


Fig. 10 Optimal investment

$$
\begin{aligned}
&\textit{minimize}\ \ QMP(Imp(BC)) \\
&\text{subject to } (500 \cdot Imp(BC_1) + 150 \cdot Imp(BC_2) + 400 \cdot \\
&\quad Imp(BC_3) + 400 \cdot Imp(BC_4) + 150 \cdot Imp(BC_5) + 200 \cdot Imp(BC_6) + 500 \cdot \\
&\quad Imp(BC_7) + 300 \cdot Imp\ (BC_8) + 900 \cdot Imp(BC_9)) \cdot 10 \leq 350 \\
&\quad Imp(BC_2) - 0.2 \leq 0; \quad - Imp(BC_2)\ \leq 0 \\
&\quad Imp(BC_3) - 0.12 \leq 0; \quad - Imp(BC_3)\ \leq 0 \\
&\quad Imp(BC_4) - 0.27 \leq 0; \quad - Imp(BC_4)\ \leq 0 \\
&\quad Imp(BC_5) - 0.08 \leq 0; \quad - Imp(BC_5)\ \leq 0 \\
&\quad Imp(BC_6) - 0.1 \leq 0; \quad - Imp(BC_6)\ \leq 0 \\
&\quad Imp(BC_7) - 0.2 \leq 0; \quad - Imp(BC_4)\ \leq 0 \\
&\quad Imp(BC_7) - 0.2 \leq 0; \quad - Imp(BC_4)\ \leq 0 \\
&\quad Imp(BC_8) - 0.06 \leq 0; \quad - Imp(BC_5)\ \leq 0 \\
&\quad Imp(BC_9) - 0.095 \leq 0; \quad - Imp(BC_6)\ \leq 0
\end{aligned}
$$

Figure 10 shows the optimal investment allocation subject to a budget of 350 μm in order to minimize QMP.

There are two BCs (Fig. 10) that are not improved due to the budget is limited, and there are one BCs where the maximum investment has not been completed due to lack of budget. The missing amounts to invest are:

$$\mathrm{BC}_1 = 105\,\mu\mathrm{m},\ \mathrm{BC}_6 = 20\,\mu\mathrm{m},\ \mathrm{BC}_9 = 85\,\mu\mathrm{m}$$

Adding all these quantities a value of 250, that is precisely the difference between the maximum budget and the available budget, is obtained. Figure 11 shows the

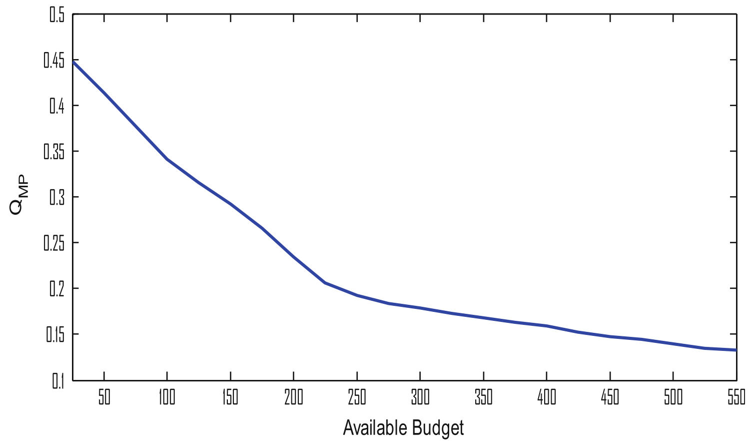


Fig. 11 Optimal investment allocation

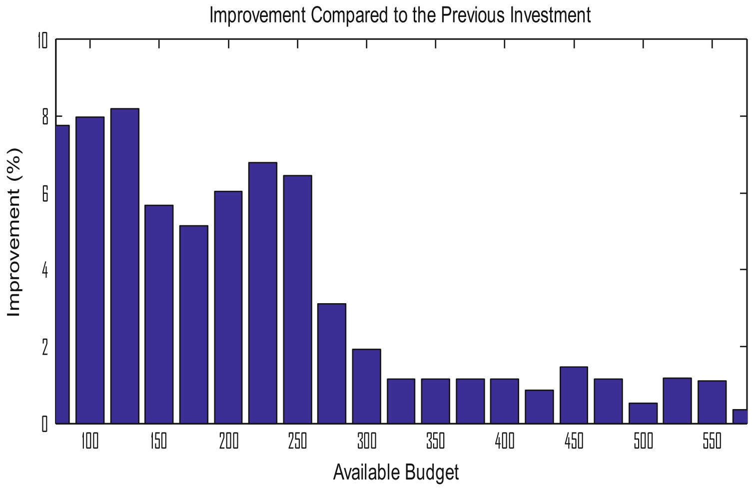


Fig. 12 Percentage improvement

behavior of the QMP reduction according to the available budget. The QMP presents a non-linear behavior.

The slope is bigger at the beginning (Fig. 11), i.e. the investments are more useful until reach to a certain budget. Figure 12 shows the improvement of each investment compared to the previous one. The investment considered rises with a step of 50 monetary units.

The firsts 250 µm (Fig. 12) are more useful than the rest of the investment. This is a useful information when the availability of budget is limited.

7 Conclusions

Decision Making (DM) is a criteria selection method for choosing a good alternative. The diagram of a DM can be performed by logical decision trees (LTDs). This is achieved having a graphical representation of those situations that need to be improved. Employing binary decision diagrams (BDDs) can be reduced the computational cost for quantitative analysis. LDT describes graphically the roots of a certain problem and their interrelations.

The size of the BDDs depends of the variable ordering, and therefore the computational cost for the qualitatively analysis. TDLR, DFS, BFS Level and AND methods has been employed in order to find the optimal variable ordering.

A NP-hard and non-linear programming problem (NLPP) is considered in a real case study in this study. The necessary conditions of optimality are defined by Karush-Khun-Tucker (KKT) conditions. It has been found the optimal allocation when resources are limited.

Acknowledgements The work reported herewith has been financially supported by Spanish Ministerio de Economía y Competitividad, under Research Grant DPI2012-31579.

References

1. Umm-e-Habiba, Asghar S (2009) A survey on multi-criteria decision making approaches. International conference on emerging technologies, Pakistan
2. Ekárt A, Németh SZ (2005) Stability analysis of tree structured decision functions. Eur J Oper Res 160:676–695
3. Baker D, Donald B, Hunter R, et al (2001) Guidebook to decision-making methods. Department of energy WSRC-IM-2002-00002. December 2001
4. Harris R (2013) Introduction to decision making. Virtualsalt. Retrieved from http://www.virtualsalt.com/crebook5.htm. 9 June 2012, 6 September 2013
5. Forrester JW (1993) System Dynamic and the Lessons of 35 years. A systems-based approach to policymaking. pp 199–240
6. Goertz G (2004) Constraints, compromises and decision making. Department of Political Science. University of Arizona. J Confl Resol 48(1):14–37
7. Huber GP (2007) Toma de decisiones en la gerencia, 2nd edn. Trillas, México
8. Lopez D, Van Slyke WJ (1977) Logic tree analysis for decision making. Omega, Int J Manag Sci 5(5):614–617
9. Pliego A (2012) Estudio cuantitativo y cualitativo de fallos en sistemas complejos. July 2012. Ciudad Real, Spain
10. Lee CY (1959) Representation of switching circuits by binary-decision programs. Bell Syst Tech J 38:985–999
11. Akers SB (1978) Binary decision diagrams. IEEE Trans Comput C-27(6):509–516, June 1978
12. Moret BME (1982) Decision trees and diagrams. Comput Surv 14:413–416
13. Bryant RE (1986) Graph-based algorithms for Boolean functions using a graphical representation. IEEE Trans Comput C-35(8):677–691
14. Masahiro F, Hisanori F, Bobuaki K (1988) Evaluation and improvements of Boolean comparison. Method based on binary decision diagrams. Fujitsu Laboratories Ltd

15. Malik S, Wang AR, Brayton RK, Vincentelli AS (1988) Logic verification using binary decision diagrams in logic synthesis environment. In: Proceedings of the IEEE international conference on computer aided design
16. Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) Introduction to algorithms, 2nd edn. MIT Press and McGraw-Hill. Section 22.3: Depth-first search, pp 540–549. ISBN 0-262-03293-7
17. Jensen R, Veloso MM (2000) OBDD-based universal planning for synchronized agents in non-deterministic domains. J Artif Intel Res 13:189–226
18. Xie M, Tan KC, Goh KH, Huang XR (2000) Optimum prioritisation and resource allocation based on fault tree analysis. Int J Qual Reliab Manag 17(2):189–199
19. Artigao E (2009) Análisis de árboles de fallos mediante diagramas de decisión binarios. November 2009
20. Brace KS, Rudell RL, Bryant RE (1990) Efficient implementation of a BDD package. 27th ACM/IEEE design automation conference
21. Garcia F, Pliego A, Lorente J, Trapero J (2014) A new ranking approach for decision making in maintenance management. Proceedings of the seventh international conference on management science and engineering management. Lecture notes in electrical engineering, vol 241. pp. 27–38
22. Bartlett LM (2003) Progression of the binary decision diagram conversion methods. Proceedings of the 21st international system safety conference; August 4–8, 2003. Ottawa, Westin Hotel, pp 116–125

Integration of a Heuristic Method into an ERP Software: A Successful Experience for a Dynamic Multi-item Lot Sizing Problem

José M. Gutiérrez, Marcos Colebrook, Cristian Rivero, and Teresa Pestana

1 Introduction

Nowadays, logistics costs (among those are the inventory carrying and ordering costs) as a percentage of gross domestic product are about 10 % in OECD countries [1]. However, these logistics expenditures in Latin America and Caribbean, as a percentage of the final product value, are between 18 and 35 % higher than those of OECD countries [2]. Accordingly, in the current economic scenario, where the different operations in the supply chain are globalized, it is imperative that at a microeconomic level small and medium enterprises, or SMEs for short, define efficient strategies that make them more competitive with rivals. One of these approaches consists of determining either optimal or near-optimal ordering policies, which significantly reduce the inventory and replenishment costs. Precisely, Inventory Management, as a field of Operations Research, provides an appropriate methodology to deal with these challenges from a scientific prism.

Moreover, we are interested in designing a solution method to determine a joint replenishment plan for multiple items that are to be stored at either a wholesaler/

J.M. Gutiérrez (✉)
Departamento de Matemáticas, Estadística e Investigación Operativa, Universidad de La Laguna, Tenerife, Spain
e-mail: jmgrrez@ull.edu.es

M. Colebrook
Departamento de Ingeniería Informática, Universidad de La Laguna, Tenerife, Spain
e-mail: mcolesan@ull.edu.es

C. Rivero
Urban Science, Madrid, Spain
e-mail: crivero@urbanscience.com

T. Pestana
ITOP MC, Tenerife, Spain
e-mail: teresapestana@itop.es

© Springer International Publishing Switzerland 2015
F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_2

retailer, when items are finished goods, or an in-process storage area in case of raw materials, work-in-process goods, components or spare parts. Therefore, we assume that a capacitated warehouse/depot within the supply chain is replenished from external suppliers and/or internal manufacturing processes to thereafter satisfy the deterministic time-varying demands for multiple items of end-customers, retailers or ERP (*Enterprise Resource Planning*), which is an information system that integrates and automates key business processes within the company.

Thus, we consider that each time an order for an item is placed a setup cost (fixed and independent of the order quantity) and a linear replenishment cost are incurred. Additionally, a carrying cost dependent on the ending inventory level is applied for each item and period. Furthermore, the demand for any item and period must be fulfilled on time and instantly, i.e. stockouts are not permitted and leadtimes are negligible. The problem can be mathematically described using Mixed Integer Programming (MIP) and solved by any commercial optimization software. However, since the number of items can be huge in practice, even the most powerful optimization program would consume too much time in determining an optimal solution. Accordingly, ad-hoc heuristic methods become interesting tools to obtain compromise solutions.

The first contribution to determine an optimal replenishment plan under deterministic time-varying parameters is credited to Wagner and Whitin [3]. It is a fact that the model considered by the authors is extremely idealistic. In particular, a single-item is produced/replenished in an uncapacitated single-facility and the demand varies with time through a finite planning horizon and should be met without shortages in each period. Furthermore, the cost structure consists of the sum of linear carrying cost and fixed production/replenishment cost (regardless of the order quantity) for all periods. However, a key result was derived in this paper, which has been extensively used afterward in other more complex dynamic models. A family of optimal solutions called Zero Inventory Ordering (ZIO) policies were introduced for the first time. These policies define a simple decision rule: a production/replenishment order is placed only when the inventory level is zero. Thereafter, Veinott [4] and Zangwill [5] proved that this rule can be also applied even when the cost structure is concave in general.

In the early seventies, Love [6] proposed an extension of the Wagner-Whitin model to include both limited storage capacities and general concave cost structures for inventory carrying, placing orders and backlogging. Moreover, a polynomial solution method based on the Dynamic Programming (DP) paradigm was derived to construct optimal plans. Although the theoretical complexity of the algorithm by Love was not improved, Gutiérrez et al. [7, 8] introduced a new characterization of the optimal plans that allows devising a DP algorithm with faster running times and linear expected time complexity.

Unfortunately, references to the model with multiple items are quite sparse in the literature. Possibly, this lack of interest is due to research efforts have focused primarily on the Capacitated Lot Sizing Problems (CLSP, family of problems where the constraints are imposed on the production capacity rather than storage limitations). One reason for this disinterest could be that researchers have traditionally considered that what applies to the CLSP also applies to the other model. However, this approach is not entirely true since whilst the CLSP is NP-Complete in general [9], the Wagner-Whitin model with storage capacities is not. Conversely, the latter model is as complex as the CLSP when multiple products are involved, and hence heuristics become attractive methods to generate good approximate solutions thereby offsetting the high running times consumed by commercial solvers.

At the end of the last decade, Minner [10] carried out a comprehensive study in which three different heuristic strategies were implemented and compared. These approaches are based on different methodologies. On the one hand, the *smoothing approach* by Dixon and Poh [11] consists of determining independently the replenishment policy for each item first by using any of the efficient algorithms proposed for the original dynamic lot sizing problem (namely, [12–14]). In a second step, infeasibilities in one period are fixed either postponing to the next period (PUSH) or advancing to the previous period (PULL) the replenishment quantity that minimizes the marginal cost. Secondly, the *constructive approach* devised by Günther [15] determines first lot-for-lot policies to solely satisfy the demand for each product in each period. After that, the replenishment quantities for each item and period are increased using an economic criterion as a discriminant. Finally, the *savings approach* developed by Axsäter [16] for the vehicle routing problem consists of reducing the ordering costs by combining consecutive replenishments in one batch whenever this operation represents a saving. The computational experiment in Minner [10] suggests that the last method is more robust than the other two heuristics when demand variability increases; and shows smaller increases of deviations for variations of costs and capacity parameters.

The rest of this chapter is organized as follows. In Sect. 2 we formulate the model in terms of a Mixed Integer Programming (MIP) problem. The key results to improve the approach proposed by Dixon and Poh [11] and the heuristic method are described in detail in Sect. 3. Moreover, the integration of the heuristic method as an add-on in the commercial software SAP Business One is explained in Sect. 4. In Sect. 5, we introduce a visual example to illustrate both how the input parameters should be entered and how the final solution is reported to the decision maker. Finally, in Sect. 6 we provide some conclusions and final remarks.

2 Problem Formulation

We assume a set of N items independently demanded and a planning horizon with T periods. Besides, we define the following parameters (see Table 1):

Moreover, we denote by $D_{n,t}$ the accumulated demand of item n from period t to T, namely $D_{n,t} = \sum_{k=t}^{T} d_{n,k}$. Additionally, let S_t be the total dynamic inventory capacity at the warehouse in period t and let w_n be the unit capacity (volume) of item n. On the other hand, we define the following variables (see Table 2):

Table 1 The set of parameters

Parameter	Description
$d_{n,t}$	Demand for item n in period t
$f_{n,t}$	Fixed setup cost for item n in period t
$p_{n,t}$	Replenishment cost for item n in period t
$h_{n,t}$	Carrying cost for item n in period t

Table 2 The set of variables

Variable	Description
$x_{n,t}$	Order quantity replenished at the beginning of the period t for item n
$y_{n,t}$	Binary variable related to the order quantity, which is equal to 1 if an order for item n is placed in period t, and 0 otherwise
$I_{n,t}$	Inventory level of item n at the end of period t

Moreover, we assume that both the initial and final inventory level for each item is zero, i.e. $I_{n,0} = I_{n,T} = 0$ for all items n. However, these assumptions can be dropped off without loss of generality since the case of positive initial and/or final inventory can be adapted to the formulation below just allocating initial inventories to demands of the first periods and/or adding a required final inventory to the demand of the last period. Accordingly, we can state the following MIP problem called the *Multi-Item Ordering/Production Problem with Storage Constraints* (MIOPPSC):

$$
\begin{aligned}
&(\text{MIOPPSC}) \ \min \sum_{n=1}^{N} \sum_{t=1}^{T} f_{n,t} y_{n,t} + p_{n,t} x_{n,t} + h_{n,t} I_{n,t} \\
&s.t.: \\
&\sum_{n=1}^{N} w_n (I_{n,t-1} + x_{n,t}) \le S_t, t = 1, \ldots, T \\
&I_{n,t-1} - I_{n,t} + x_{n,t} = d_{n,t}, t = 1, \ldots, T; n = 1, \ldots, N \\
&x_{n,t} \le y_{n,t} D_{n,t}, t = 1, \ldots, T; n = 1, \ldots, N \\
&x_{n,t}, I_{n,t} \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}; y_{n,t} \in \{0, 1\}, t = 1, \ldots, T; n = 1, \ldots, N
\end{aligned}
$$

The terms in the objective function represent, respectively, the total setup cost, the total ordering cost and the total holding cost. The first set of constraints ensures that the inventory level at the beginning of each period does not exceed the warehouse capacity. The next constraint set are the well-known material balance equations and the third set states the relationship between the order quantity and its binary variable for each item and period. Finally, the last set of constraints defines the character of each variable.

3 The Heuristic Approach

Inspired by the smoothing approach by Dixon and Poh [11], we have proposed in Gutiérrez et al. [17] a heuristic method to determine near-optimal ordering schedules for a set of items, which share a common warehouse within a demand system not necessarily independent (horizontal). In case of dependent demand systems, we confine ourselves to consider only the final requirements for those items (raw material, sub-assemblies, intermediate assemblies and sub-components) sharing the same depot/container after the Bill of Materials (BOM) explosion is accomplished. The new method is based on the natural extension of the characterization of the optimal plans for the single-item case in Gutiérrez et al. [7] to consider multiple items.

As in Dixon and Poh [11], the first step of our heuristic consists of determining independent plans by solving N single-product, dynamic uncapacitated lot-sizing problems. If infeasibility occurs at the end of period t, the excess warehouse capacity requirements can be reduced either by postponing an item replenishment batch (PUSH operation) or by advancing a future replenishment (PULL operation). Nevertheless, Dixon and Poh claimed that on average the PUSH operation is preferred to the PULL strategy since the impact of the move on total costs is predictable. Thus, our heuristic differs from that given by Dixon and Poh in that we confine ourselves to analyze the PUSH operation but extending the strategy to consider both postponements of orders for a subset of items to nonconsecutive periods (i.e., from period t to period $t+k$, $k \in \{1,\ldots,T-t\}$, instead of $t+1$ only) and replenishment quantities different from a sum of demands of consecutive periods for an item.

Precisely, let $\hat{x}_{n,t}$ denote the optimal replenishment quantity in period t of the independent plan for item n, then when the exceeded capacity $W_t = \sum_{n=1}^{N} w_n(I_{n,t-1} + \hat{x}_{n,t}) - S_t$ is positive in period t, the replenishment quantity in period t for a selection of items must be partially postponed to later periods. The items are chosen from those minimizing the marginal cost of shifting one unit of item n from a period $i \le t$ to a later period $j > t$. Obviously, the items must be selected assuring the feasibility at period t, hence the total storage capacity should be not exceeded and the total replenishment must satisfy at least the demand of each item in that period.

For the sake of clarity, let us assume that the storage capacity constraint at period t is violated ($W_t > 0$). Moreover, let us admit that it is convenient shifting forward from period i to period j a certain replenishment quantity of item n. Note that although the infeasibility occurs at period t, we do not confine ourselves to consider only those items that were replenished at that period in the corresponding independent optimal plans. Instead, we also determine the marginal costs $c_{i,j}^{n}$ for those items n replenished in a period i prior to t. Accordingly, let $succ(n,t)$ and $pred(n,t)$ denote the production periods successor and predecessor, respectively, to period t in the independent optimal plan for item n. It should be noticed that period j is strictly less than the production period $succ(n,t)$ successor to t in the independent optimal plan for item n. Consequently, period j was not a production period in that plan and so the corresponding setup cost must be included to determine the actual cost of the postponement, which is defined as:

$$r^n_{i,j} = f_{n,j} + c^n_{i,j} q^n_{t,j}$$

where $q^n_{t,j}$ represents the actual quantity that is postponed from period i to period j for item n. Observe that $q^n_{t,j}$ can be either a sum of demands of consecutive periods for item n or a quantity enough to remove the exceeded inventory in period t. However, in order to make period t feasible, withdrawing a sum of demands for a single item could not be sufficient, and hence we should select a subset of items N_t such that the sums of demands $\sum_{k \in N_t} q^k_{t,j_k}$ with minimum marginal costs are equal to or greater than W_t. Consequently, for each unfeasible period, we consider two types of sorted lists, namely, a list $L[n]$ for each item n, which contains increasing marginal costs corresponding to sums of demands of consecutive periods and a unique list G including the marginal costs of those items for which W_t is a multiple of their unit capacities.

Once these lists are completed, we should compare the sum of accumulated marginal costs (CLW) for a selection of items obtained from lists $L[n]$ with the minimum marginal cost (CW) in G. Remark that W_t should be updated each time the replenishment of one item is partially shifted to a subsequent period. If after the postponement the storage capacity remains to be violated, then we must select a different item to reduce the excess warehouse capacity requirements. We show below a sketch of the heuristic method to determine near-optimal solutions.

Algorithm MIOPPSC

```
For the sake of clarity, we denote the data structures as: W(t), w(n),
etc.
for all N items do
  Solve each uncapacitated plan independently using
    Wagelmans et al. (1992) algorithm
end for
for each period t where the capacity S(t) is violated in the previous
plans
  for all items n ← 1 to N do
    Compute list G and lists L[n]
  end for
  CW ← minimum cost value of list G
  CLW ← 0
  while W(t) > 0 do
    CLW ← CLW + minimum cost of all lists L[n], n = 1 to N
    if CLW < CW then
```

```
      Decrease W(t) by the corresponding quantity q multiplied by
      the item's weight w(n)
      Recalculate lists G and L[n]
    else
      W(t) ← 0
    end if
  end while
  Update properly plan x and all subsequent variables
end for
```

The reader is referred to Gutiérrez et al. [17] for a comprehensive description of both the algorithm and the computational experiment, which reports that the solutions provided by the heuristic are on average 5 % above the best solution achieved by CPLEX. Furthermore, the heuristic has been afterward encapsulated into an ERP. In this sense, Iris and Yenisey [18] recently addressed a very similar model to the one presented in this chapter, and they plan to embed it also into an ERP system.

4 Development and Implementation Within SAP Business One

SAP Business One is an ERP developed by SAP that is focused on small and medium enterprises (SME), and that covers various fields: accounting, procurement, sales, inventory management, production and manufacturing, customer relation management (CRM), etc.

The development process along with the subsequent deployment was carried out as a joint collaboration with ITOP Management Consulting, a company created in 2006 in Tenerife (Spain), that offers its services on information technology primarily to SME. ITOP holds a certification as an authorized SAP partner, as well as an UNE 166002 certification on R&D Management granted by AENOR (Spanish Association for Standardization and Certification), being one of the first national companies in obtaining this type of certification.

The algorithm described in a previous section that solves the MIOPPSC problem was programmed in C# using Microsoft Visual Studio [19] and the SAP Business One SDK (Software Development Kit) [20] as an *add-on*. These SAP Business One add-ons are basically programs developed to extend the functionality in order to satisfy the final user requirements.

The main components of the system are:

- SAP Business One Server: it holds the database that contains all the company's data such as information about the business partners, products, invoices, etc.
- SAP Business One Clients: the MIOPPSC add-on is installed on them, and they connect to the server to get the required information.

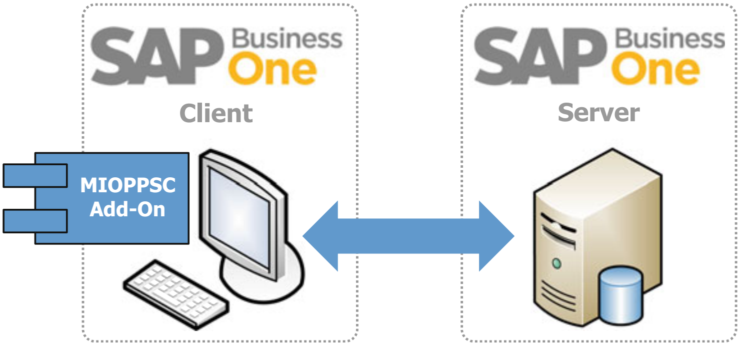


Fig. 1 System architecture integration between the MIOPPSC add-on and SAP Business One

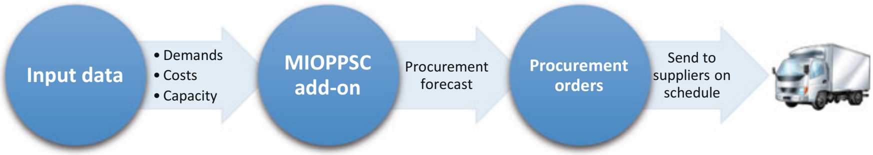


Fig. 2 Data flow from the initial input to the final output

The following Fig. 1 summarizes the integration between SAP Business One and the add-on.

The flow of information inside SAP Business One from the initial input to the final solution is as follows (see Fig. 2):

1. The algorithm takes, for each period $t = 1, \ldots, T$, the input data for the demands $d_{n,t}$, the setup costs $f_{n,t}$, and the inventory capacity of the warehouse S_t.
2. This input data is then used by the MIOPPSC algorithm to compute a procurement forecast.
3. Finally, this forecast is used by the MRP (*Material Requirements Planning*) module to determine the procurement orders to send to the suppliers in order to satisfy the customers' demands over each period.

To illustrate this workflow, we show in the next section a visual example running on a real case.

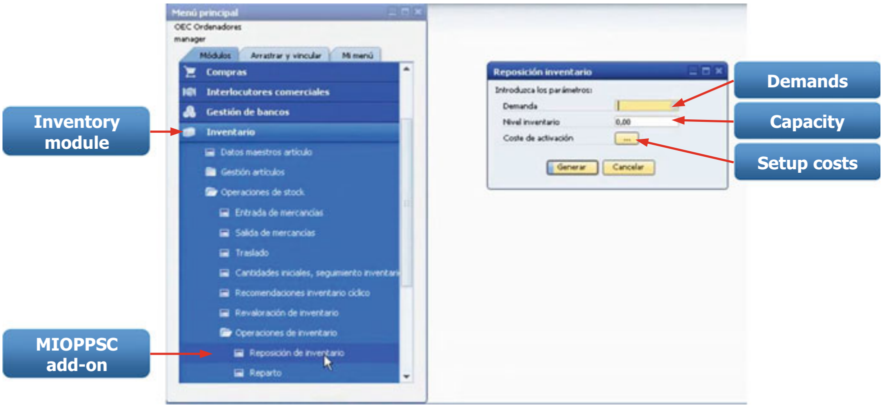


Fig. 3 Input parameters of the MIOPPSC add-on running within SAP Business One

5 Visual Example

First of all and, and in order to show the example, we need to run the MIOPPSC add-on that was installed under the *Inventory module* of SAP Business One (see Fig. 3).

The input parameters needed by MIOPPSC are explained in the following sections.

5.1 *Demands*

The demands for each item will be taken from a pre-defined object in SAP Business One called *forecasts* within the MRP module, which are the known demands in future periods (see Fig. 4).

In this particular case, the demands for the two items were identical:

$$d_{1,t} = d_{2,t} = \{69, 29, 36, 61, 61, 26, 34, 67, 45, 67, 79, 56\}, \quad t = 1, \ldots, 12$$

The forecast panel has several parts:

- *Forecast information*: such as the name and the description.
- *Forecast dates*: initial and final dates.
- *Month view*: in our example, the periods are expressed in months within a year range (from April to March which corresponds to $t = 1, \ldots, 12$).
- *Item information*: such as item code and its description. In this example, we only have two items ($n = 1, 2$).
- *Item demands*: for each period (months).

The rest of these two items' parameters ($n = 1, 2$), such as the unit capacity (volume, w_n), and both the production costs (p_t) and the holding/carrying costs (h_t),

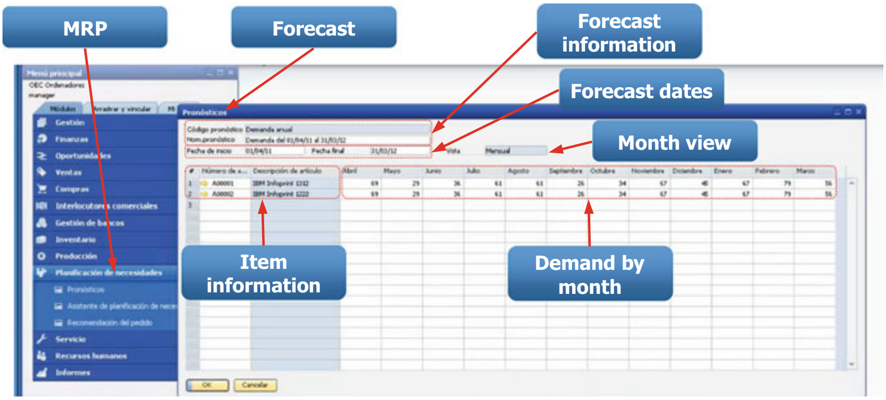


Fig. 4 Demand forecast of two items for each period (month) as the input to the add-on

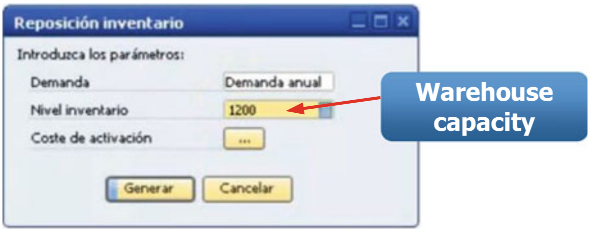


Fig. 5 The capacity (inventory level) of the warehouse is set for all the periods

with $t = 1, \ldots 12$, can be taken from each item's master data. In this particular example, the values of these parameters are the following:

- Unit capacity: $w_1 = 5$, $w_2 = 10$.
- Production costs: $p_{1,t} = p_{2,t} = 0$, for all periods $t = 1, \ldots 12$.
- Holding/carrying costs: $h_{1,t} = h_{2,t} = 1$, for all periods $t = 1, \ldots 12$.

5.2 Capacity

In this example, the capacity (inventory level of the warehouse) will be set to 1,200 for all periods $t = 1, \ldots 12$ (see Fig. 5).

5.3 Setup Costs

For each period $t = 1, \ldots 12$, the setup costs of each item must be entered manually as we show in Fig. 6. The values are the following:

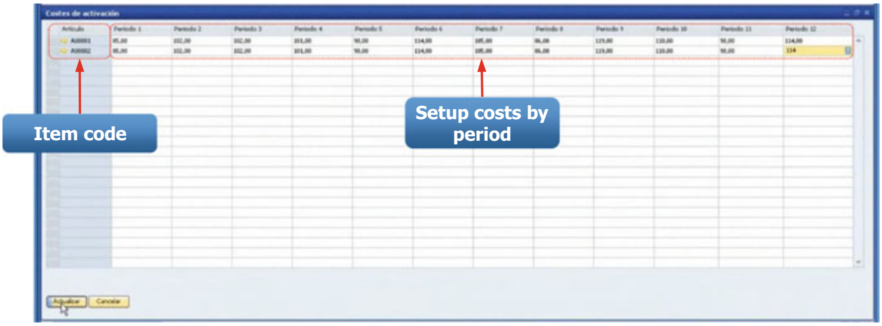


Fig. 6 Setup costs for the two items and the twelve periods considered in the example

Fig. 7 Example ready to be solved

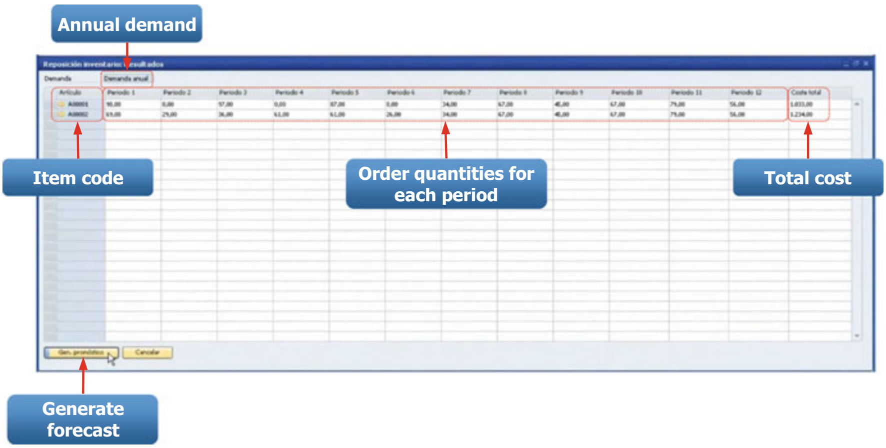


Fig. 8 Solution to the example showing the order quantities to be replenished at the beginning of each period

$$f_{1,t} = f_{2,t} = \{85, 102, 102, 101, 98, 114, 105, 86, 119, 110, 98, 114\}$$

Once all these data is set up, we can press the corresponding button to solve the problem (see Fig. 7).

The solution output will point out the order quantities of each item ($n = 1, 2$) to be replenished at the beginning of the corresponding period (months, $t = 1, \ldots 12$), in order to minimize the total cost (see Fig. 8). The solution values are the following:

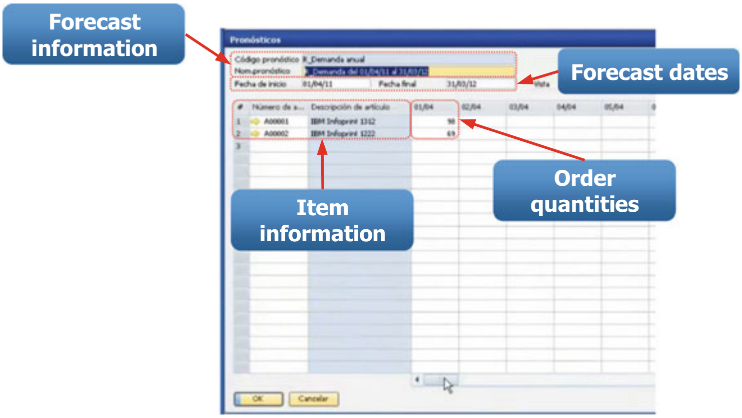


Fig. 9 Forecast generated by the MIOPPSC algorithm that will be used in the MRP module

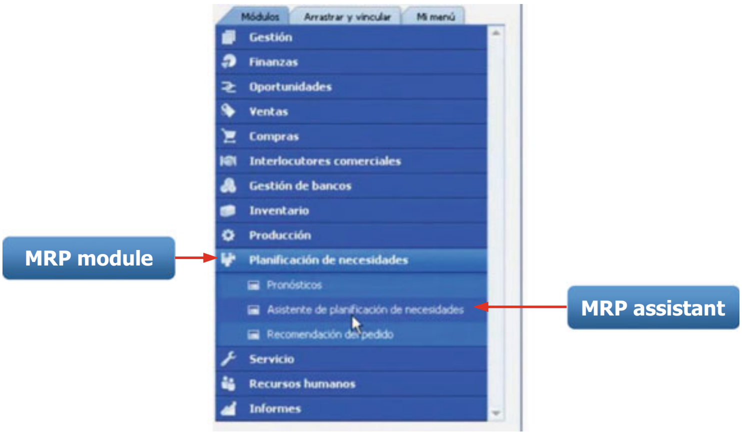


Fig. 10 MRP assistant module within SAP Business One

$x_{1,t} = \{98, 0, 97, 0, 87, 0, 34, 67, 45, 67, 79, 56\}$, with a total cost of 1,033 €.
$x_{2,t} = \{69, 29, 36, 61, 61, 26, 34, 67, 45, 67, 79, 56\}$, with a total cost of 1,234 €.

Once the solution is obtained, we now can generate a new forecast (optimized with the algorithm solution) that contains the order quantities to be replenished at the beginning of each period, that is, the first day of each month (see Fig. 9). This forecast will be used as an input by the MRP module.

The next step involves using the MRP module of SAP Business One. This module has an assistant that will guide the user through the process to obtain a procurement planning based on the previous forecast (Fig. 10).

The first step of the process consists of creating a new scenario with a name and a description (see Fig. 11). This scenario will return the procurement planning for a whole year based on the forecast computed by the MIOPPSC algorithm.

In the next step, the user has to enter all the parameters regarding the horizon planning, the range of item codes, and the visualization options (see Fig. 12):

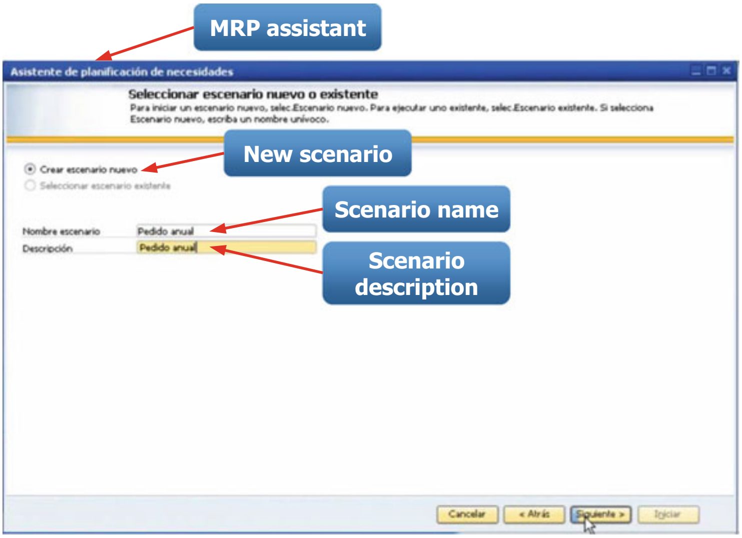


Fig. 11 Creating a new scenario using the MRP assistant

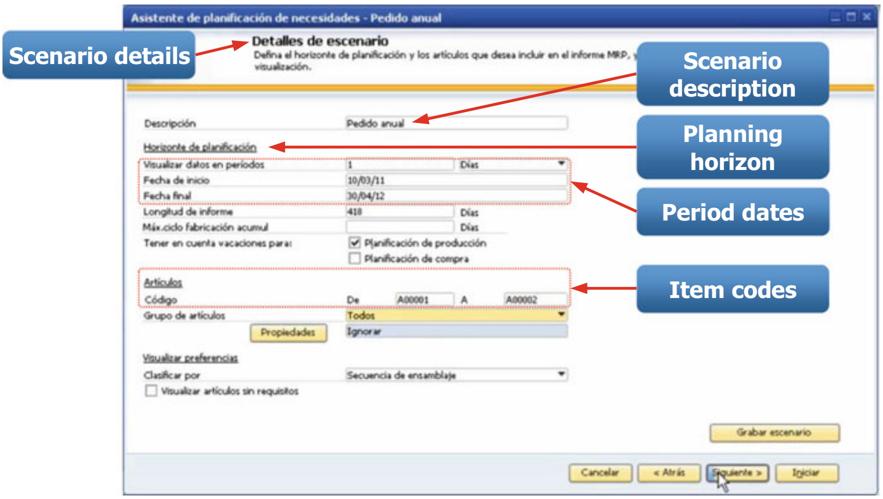


Fig. 12 Scenario parameters concerning horizon planning, item codes and visualization options

- *Period dates*: the initial and final dates of the planning. These parameters are only needed for visualization purposes of the final results.
- *Item codes*: range of items to be considered by the MRP.

Finally, the last step involves providing the data source that will be used by the MRP to generate the procurement planning. As it is shown in Fig. 13, the user can select among several choices, with most of them left unselected:

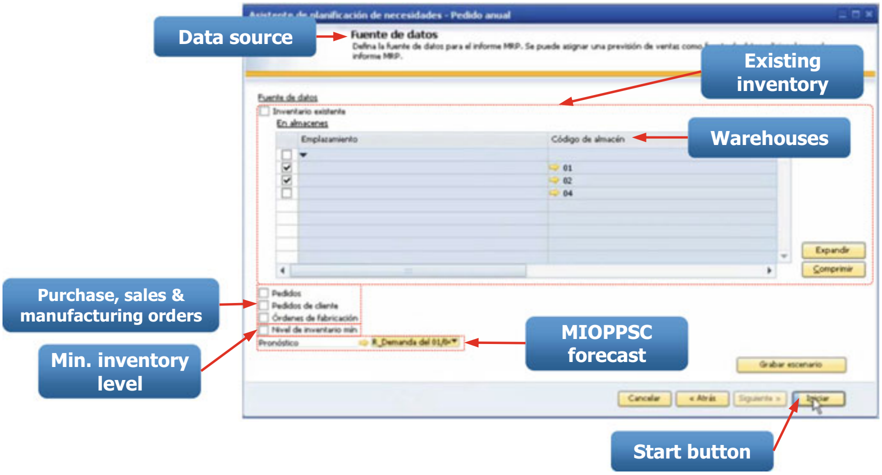


Fig. 13 Different data sources that can be used as inputs by the MRP

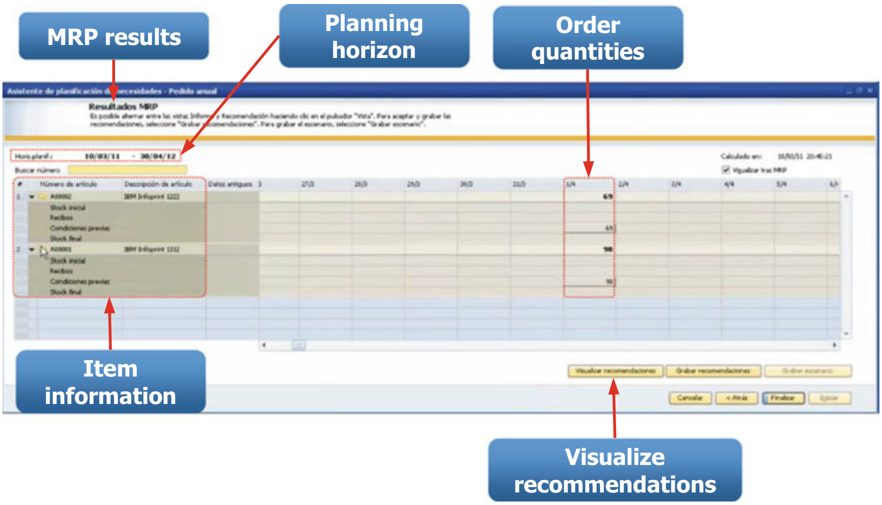


Fig. 14 MRP results using the forecast of the MIOPPSC algorithm

- *Existing inventory* (omitted): current inventory held in the different warehouses.
- *Orders of purchase, sales or manufacturing* (omitted): current orders that are already stored in the system.
- *Min. inventory level* (omitted): use the minimal inventory level as the input.
- *Forecast*: in this example, the input is sourced from the previous forecast computed by the MIOPPSC algorithm.

Once all the information needed by the MRP module has been supplied, the user can press the *Start* button to generate the MRP output.

Fig. 14 shows the MRP results for our example:

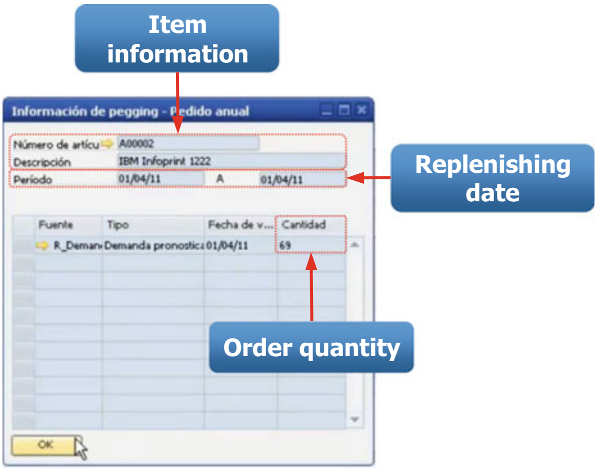


Fig. 15 Detailed view of the first item's replenishing order

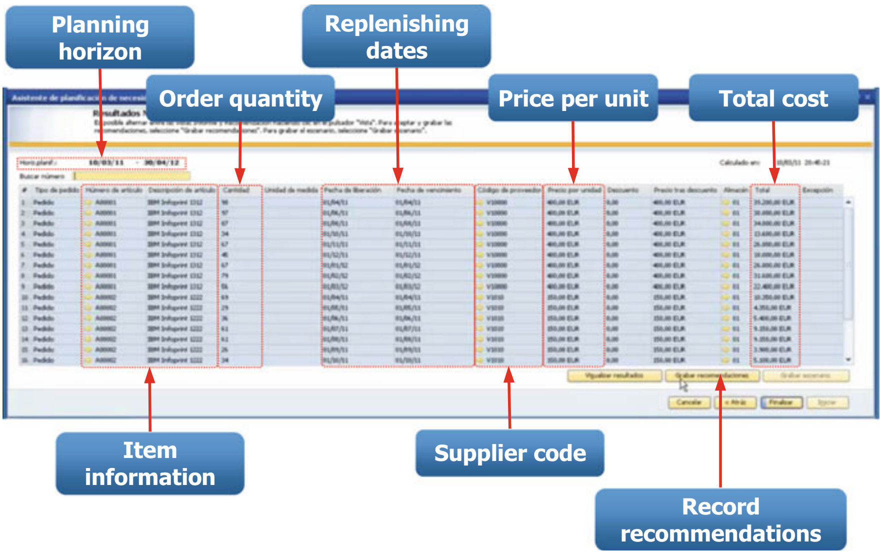


Fig. 16 Procurement recommendations generate by the MRP module

- *Planning horizon*: is the same that we set up in the scenario details.
- *Order quantities*: these orders match the ones obtained or each item by the MIOPPSC algorithm, and must be replenished at the beginning of each period (as shown in the figure).
- *Item information*: such as the item code, the description, and the stock levels.

If the user clicks on any order quantity, a new window will display showing a detailed view of the replenishing information. Fig. 15 shows the replenishing information for the first item (69 units).

In case the user presses the *Visualize recommendations* button (Fig. 14), a new window will show a full list containing all the procurement recommendations generated by the MRP module (see Fig. 16).

These recommendations can be recorded in the system to make them effective at any time by pressing the *Record recommendations* button.

6 Conclusions

We have presented a successful collaboration between a university's research group and a SAP partner company to both develop a solution in order to reduce the inventory costs in the SME, and to integrate it as an add-on in SAP Business One. As a result of this collaboration, an efficient algorithm was devised and embedded into the ERP system. Besides, the experimental results reported that the heuristic solutions were only on average a 5 % above the best solution given by a commercial solver.

The major benefits of this implementation are:

- Total integration with a high level ERP as SAP Business One.
- Provides a procurement schedule in advance to avoid stock shortages as well as overstocks.
- Optimizes the management of product costs with regards to stock holding, replenishment and delivery service to the customers.

Acknowledgments This work is partly supported by the former Spanish Ministry of Science and Innovation through the Research Project with reference MTM2010-18591 and by the new Spanish Ministry of Economy and Competitiveness through the Research Project with reference MTM2013-43396-P.

References

1. Christopher M (2011) Logistics and supply chain management. Pearson Education, Great Britain
2. Guasch JL, Kogan J (2006) Inventories and logistic costs in developing countries: levels and determinants—A red flag for competitiveness and growth. Revista de la Competencia y de la Propiedad Intelectual. Lima, Perú
3. Wagner HM, Whitin TM (1958) Dynamic version of the economic lot size model. Manag Sci 5 (1):89–96
4. Veinott AF Jr (1969) Minimum concave-cost solution of leontief substitution models of multifacility inventory systems. Oper Res 7:262–290
5. Zangwill WI (1966) A deterministic multiproduct multifacility production and inventory model. Oper Res 4:486–507
6. Love SF (1973) Bounded production and inventory models with piecewise concave costs. Manag Sci 20(3):313–8

7. Gutiérrez J, Sedeño-Noda A, Colebrook M, Sicilia J (2003) A new characterization for the dynamic lot size problem with bounded inventory. Comput Oper Res 30:383–395
8. Gutiérrez J, Sedeño-Noda A, Colebrook M, Sicilia J (2007) A polynomial algorithm for the production/ordering planning problem with limited storage. Comput Oper Res 34(4):934–937
9. Florian M, Lenstra JK, Rinnooy Kan AHG (1980) Deterministic production planning: algorithms and complexity. Manag Sci 26(7):669–679
10. Minner S (2009) A comparison of simple heuristics for multi-product dynamic demand lot-sizing with limited warehouse capacity. Int J Prod Econ 118:305–310
11. Dixon PS, Poh CL (1990) Heuristic procedures for multi-item inventory planning with limited storage. IIE Trans 22(2):112–123
12. Federgruen A, Tzur M (1991) A simple forward algorithm to solve general dynamic lot sizing models with n periods in $O(n \log n)$ or $O(n)$ time. Manag Sci 37(8):909–925
13. Wagelmans A, Hoesel SV, Kolen A (1992) Economic lot sizing: an $O(n \log n)$ algorithm that runs in linear time in the Wagner–Whitin case. Oper Res 40(1):145–156
14. Aggarwal A, Park JK (1993) Improved algorithms for economic lot size problems. Oper Res 41(3):549–571
15. Günther HO (1991) Bestellmengenplanung aus logistischer sicht. Z Betriebswirtschaft 51:541–555
16. Axsäter S (1980) Economic lot sizes and vehicles scheduling. Eur J Oper Res 4:395–398
17. Gutiérrez J, Colebrook M, Abdul-Jalbar B, Sicilia J (2013) Effective replenishment policies for the multi-item dynamic lot-sizing problem with storage capacities. Comput Oper Res 40:2844–2851
18. Iris C, Yenisey MM (2012) Multi-item simultaneous lot sizing and storage allocation with production and warehouse capacities. In: Hu H, Shi X, Stahlbock R, Voß S (eds) ICCL'12: Third international conference on computational logistics, Shanghai, China, September 2012. Lecture notes in computer science, vol 7555. Springer, Berlin, p 129–141
19. MVS (2014) MS Visual Studio. http://msdn.microsoft.com/en-us/vstudio/aa718325. Accessed 1 Sep 2014
20. SAP BO SDK (2014) SAP Business One SDK General. http://scn.sap.com/docs/DOC-28739. Accessed 1 Sep 2014

Economic Viability Analytics for Wind Energy Maintenance Management

Jesús María Pinar Pérez, Eva Segura Asensio, and Fausto Pedro García Márquez

1 Introduction

The climate change have motivated the study and developed of the renewable energies, where the new cleaner and efficient energy technologies have an important role in the sustainable development in the future energy scenario [1–3].

The rising in the wind energy industry has done that firms will be focusing in more competitiveness costs. For example, for a 20-year life, the operations and maintenance (O&M) costs of 750 kW turbines might account for about 25–30 % of the overall energy generation cost [4] or 75–90 % of the investment costs [5].

Reference [6] suggested that larger wind turbines fail more frequently and thus require more maintenance. Reducing inspection and maintenance costs has become increasingly important as wind turbine size and numbers have continued to rise. Condition Monitoring Systems (CMS) are probably the most effective approaches to minimize O&M costs and to improve the availability of wind turbines by early detection of the faults. On the other hand, CMS is usual a complex task for any firm because it requires a set of sensors and data acquisition systems to monitor different parameters of the wind turbines. It also requires knowledge and expertise to interpret the large volume of data collected from the wind turbines. In this research work is studied the economic feasibility of a CMS in a wind turbine. The main objective of this work is the development of a life cycle cost (LCC) model for a CMS on wind turbines, being applied to a real case study.

The work has been structured as follows: The introduction is presented in Sect. 1, and a review of life cycle costs for CMS in wind turbines is described in

J.M. Pinar Pérez (✉)
CUNEF-Ingenium, Madrid, Spain
e-mail: jesusmaria.pinar@cunef.edu

E. Segura Asensio • F.P. García Márquez
Ingenium Research Group, University of Castilla-La Mancha, Ciudad Real, Spain
e-mail: Eva.segura.asensio@hotmail.com; FaustoPedro.Garcia@uclm.es

© Springer International Publishing Switzerland 2015

F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_3

Sect. 2. The wind turbine condition monitoring strategies is introduced in Sect. 3, where the state of the art in the literature is presented in Sect. 4. The case of study is described in Sect. 5. The novel life cycle cost approach is presented in this work is proposed in Sect. 6. Finally, conclusions are presented in Sect. 7.

2 Life Cycle Costs

Life cycle costs (LCC) can be defined as the sum of all recurring and one-time (non-recurring) costs over the full life span, or a specified period, of a particular solution. LCC includes direct and initial costs plus any periodic or continuing costs for operation and maintenance [7]. The International Electromechanical Commission (IEC) divides the LCC into two different subcategories (IEC 300-3-3, 1996):

1. *Investment (or acquisition) cost*, i.e. the initial costs that will be produced before operation.
2. *Cost of ownership*, or life support cost once operational, i.e. operating, maintenance and repair costs. These terms must be discounted to their present value since they arise in subsequent years.

The main contribution of this research work is the application of LCC of CMS of wind turbines, where it has been carry out for a real case study from Schleswig Holstein in Germany (named LKW).

3 Wind Turbines and Wind Turbine Condition Monitoring

3.1 Wind Turbines

Generally, a wind turbine is a rotary system in charge to extract the energy from the wind. The main parts of a wind turbine are depicted in Fig. 1 [8] and the operation of wind turbines is resumed as follows: The blades, connected to the rotor via the hub are moved by the wind. The blades, connected to the rotor via the hub, are designed to capture the energy from the fast and strong wind. The rotor, designed to capture the maximum surface of the wind, transmits the mechanical energy via the low speed shaft through the gearbox to the high speed shaft, ending in the generator which produces electricity from the rotation of the rotor. The low speed shaft is supported by the main bearing. The alignment to the direction of the wind is controlled by a yaw system that turns the housing (or "nacelle") for that purpose. The nacelle seals and protects the generator, the gearbox, the converter, etc. and is mounted at the top of a tower, and the tower support the nacelle and the blades and is assembled on a base or foundation. The pitch system (mounted in each blade) is a mechanism that turns the blade to controls the wind power captured, and it can be

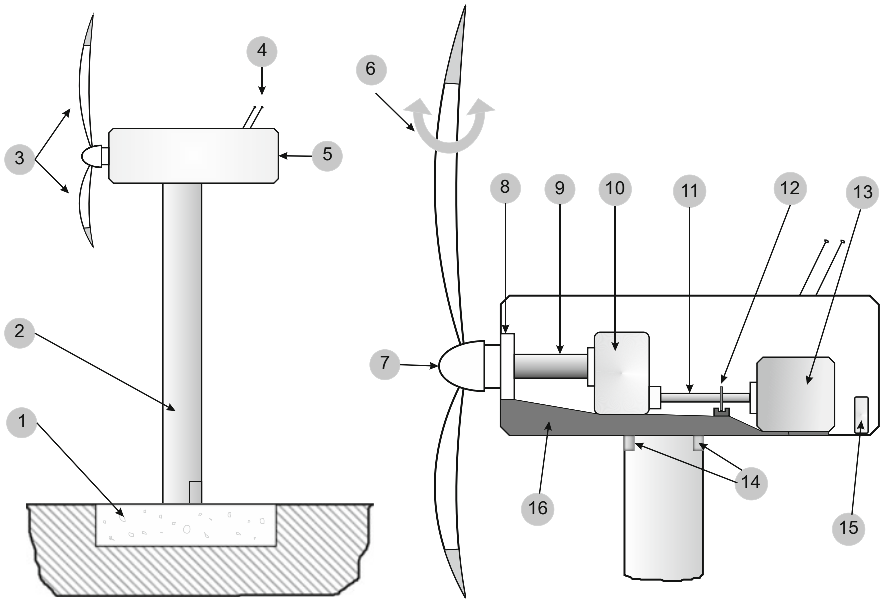


Fig. 1 Components of the WT: (1) base/foundations; (2) tower; (3) blades; (4) meteorological unit (vane and anemometry); (5) nacelle; (6) pitch system; (7) hub; (8) main bearing; (9) low speed (main) shaft; (10) gearbox; (11) high speed shaft; (12) brake system; (13) generator; (14) yaw system; (15) converter; (16) bedplate. N.B. Drive train = 9 + 11

employed as an aerodynamic brake. The WT has also a hydraulic brake to stop the WT. The meteorological unit, or weather station, provides the weather data (e.g. wind speed and direction) to the control system and it leads to control the pitch system, the brake, the yaw system, etc.

A description of the different wind turbine configurations can be found in [9].

3.2 Condition Monitoring for Wind Turbines

Under the assumption that a "significant change is indicative of a developing failure" [10], CMS have recently emerged as a new technique employed by the wind energy industry that consist in monitoring the state of the components of the wind turbine via combinations of sensors and signal processing equipment. CMS together with a detection of any deterioration of these components, based on the parameters/features obtained by measurements, is called Fault Detection and Diagnosis (FDD) [11]. CMS can be divided into two subcategories [12]: On-line CMS provide instantaneous feedback of condition; off-line CMS the data are collected at regular time intervals using measurement systems that are not integrated with the equipment [13].

CMS help to significantly reduce the maintenance tasks resulting in increased reliability, availability, maintainability and safety (RAMS), while downtimes and O&M cost are substantially reduce [14, 15]. The benefits of these modern techniques have been especially effective in offshore wind farms owing to the high cost of O&M at sea and the high dimensions of the turbines [16]. On the other hand, the implementation of CMS requires the use sensor and data acquisition systems to collect and store the information of the system, a processing step to interpret the data and a fault detection step to implement effective maintenance policies. This implies that CMS are more complex than other maintenance strategies such as run-to-failure maintenance (RFM) or scheduled maintenance (SM) [17].

The effectiveness of CMS depends on three main different elements:

- The technique applied,
- The number and type of sensors,
- The associated signal processing to extract the useful information from the different signals.

Different studies can be found in the scientific literature regarding to the techniques applied. Vibration analysis is the most popular and known for CMS, being usually applied for rotating equipment [18–20]. Acoustic emission is another popular non-destructive CMS that has been applied for the diagnostics of bearings and bears [21] and for the structural health of the blades [22]. Ultrasonic testing are also non-destructive techniques extensively used in the detection of internal defects in towers and blades [23, 24]. Oil analysis is usually used to monitor the status of the components oil lubricated in wind turbines. These techniques can be applied off-line [25] or on-line [26]. Strain measurements are used for lifetime forecasting and to avoid critical structure stress levels [27, 28]. Thermography is often used for monitoring electric and electronic components [29], but applications have been reported for the detection of damages in wind turbines [30]. Other important techniques applied to wind turbines are shock pulse method (SPM), performance monitoring or radiographic inspection, among others [31–33].

On the other hand, different options can be found in the scientific community regardless of the technique. Trend analysis has been applied in the monitoring of pitch mechanisms. It collects data from different sensors and searching for trends [34]. Time-domain analysis studies variations in signals and trends to detect possible faults in wind turbines [35]. With amplitude modulation is extracted very low-amplitude and low-frequency periodic signals that might be masked by other higher energy vibration as in wind turbine gearboxes [36]. Wavelet transformations has been successfully applied to monitor the vibration level caused by misalignment or bearing and can be used as a general indication of a fault produced in a wind turbine [37, 38]. It provides a time-frequency 3D map of the signal being studied and its decomposition into a set of sub-signals with different frequencies. Hidden Markov Models (HMV) has been successfully applied in bearing fault detection [39] and vibration signals analysis in the machine [40]. Other promising techniques such as artificial intelligence have been applied in the fault detection of mechanical equipment [41].

Finally, for an extended explanation of the condition monitoring of wind turbines can be found in [42].

4 State of the Art

There are different economic studies in wind turbines or wind farms in the literature. The main objective of these analysis is to find how the costs can be reduced. The definition of the different costs and their link with the different elements and sub-elements of a machine or a group of machines is very important. For this purpose, an analysis based on the cost breakdown structure (CBS) is presented in [43]. Based on CBS, [44] shows a novel theoretical methodology process to study the life cycle cost of floating offshore wind farms considering technical and economic issues and their relations.

Another aspect to be considered in the calculation of these costs are the application of different rates such as the inflation rate. A general economic analysis about repowering wind farms in Spain is reported in [45]. The study describes the net present value (NPV) of the repowering process for the wind farms. A life cycle cost model for offshore wind farms that takes into account the breakdown costs and revenues expected by the construction of a wind farm is developed in [46]. Different simulations are presented in order to study the relation with failure rate, inflation rate, interest rate and discount rate.

The production losses must be converted in terms of costs. Reference [47] revealed a calculation scheme to quantify wind farm production losses reached by scheduled and unscheduled downtimes.

The wind industry is focused on the operation and maintenance reduction costs. Therefore, the O&M costs and availability of the elements of a wind farm are identified in [48].

In [49] is presented a reliability-centred asset maintenance (RCAM) strategy for maintenance optimization of wind farms. The corrective and condition based maintenance strategies were compared by carrying out a LCC analysis showing the cost-benefit of CMS with respect to the reliability and availability of wind turbines today. Furthermore, in [50] different features of CMS are studied in order to enhance that the cost of CMS design and installation is substantial in comparison to other maintenance approaches in short-term. But in long-term CMS provides benefits surpassing the costs. The reported results depicts that CMS are an excellent and viable option or increasing the production rate and reducing the downtimes in wind turbines.

A LCC analysis with different hypothetical strategies using CMS is developed in [16]. The analysis of different real case studies found that CMS improves maintenance planning in offshore and onshore farms. In the same vein, [51] shows the application of LCC with probabilistic methods and sensitivity analysis to identify the benefit of using CMS. Two approaches are used to analyze how the random behavior of the failures can affect the LCC and what are the critical parameters (subject to uncertainty). The results of the study depict a high economic benefit of using CMS and substantial benefits on the risk.

The literature does not collect any study about the LCC of CMS using different annual rates of return and providing a numerical solution for a concrete wind turbine. A life cycle cost model using the net present value is developed in this work taken into account the breakdown costs of the different elements of the CMS related with the different components of the WT. O&M costs for CM system are incorporated to the model. In addition, a bank credit or amortization is included for a real point of view.

5 Case of Study

This work is based on the data from a wind farm in Schleswig Holstein in Germany (LKW). LKW statistic data compound data from 729 wind turbines of 20 different types in a period of 5 years.

The number of failures of these 729 wind turbines collected from 2005 to 2009 reveal that only four components reached the 45.5 % of the accumulated failures. These components were electrical system, blades, gearbox and generator (see Fig. 2).

This research work considers a more complex analysis of the data a failure rate. The average failure rate is given by the number of failures per turbine per year, i.e.:

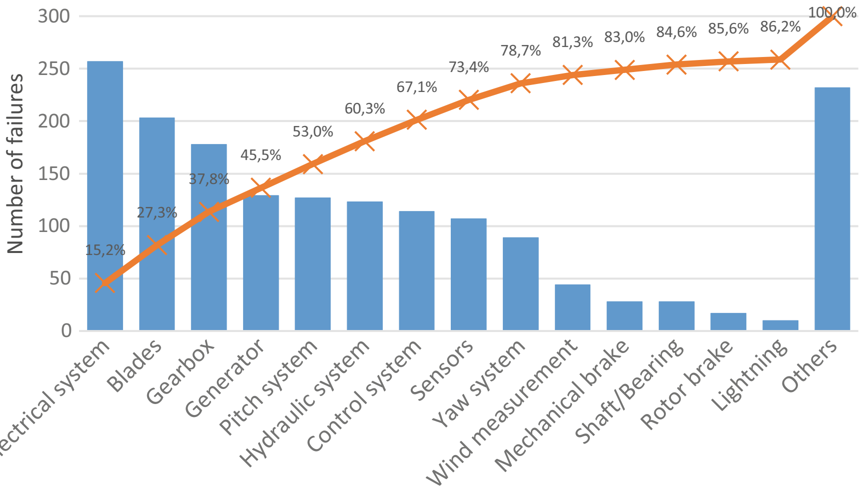


Fig. 2 Pareto chart for the case study

$$f = \frac{\sum_{i=1}^{I} N_i}{\sum_{i=1}^{I} X_i \cdot T_i}$$

where:

f: Failure rate [failures per turbine per year]
N_i: Number of failures that occurred during the time interval T_i
T_i: Time interval (I in total of 1 year each one)
X_i: Number of WTs reported for the time interval T_i.
i: 1,2,. . .,I. (years)

The downtime is the time during which a WT is not operating mainly for any maintenance task. The downtime is composed typically by:

- Diagnosing the failure (in the case of non-condition monitoring systems),
- Gathering repair equipment and spare parts,
- Accessing the mechanism, and
- Repairing and restarting the WT (usually the longest);

It is calculated as:

$$d = \frac{\sum_{i=1}^{I} d_i}{\sum_{i=1}^{I} X_i \cdot T_i}$$

where

d: Downtime due to failures per WT per year [hours per turbine per year]
d_i: Productive hours lost during the time interval T_i due to failures.

The data of LKW collect information about different types of wind turbines. This work is focused on Vestas V66 of 1.65 MW of power. The failure rate and downtime of Vestas V-66 wind turbine have been calculated. Table 1 shows these calculations.

Table 1 Failure rate and downtime for the components of Vestas V66

Component	Failure rate	Downtime
Blades	0.617	36.9
Tower	0.02	1
Pitch system	0.375	11.175
Mechanical brake	0.5	2.5
Shaft/Bearing	0.1	2
Gearbox	0.625	136.925
Generator	0.4	101.625
Hydraulic system	0.5	19.075
Yaw system	0.15	5.5
Wind measurement	0.217	11.725
Control system	0.507	12.567
Sensors	0.327	11
Electrical system	0.7	33.907
Others	0.433	7.2

6 Application of the LCC Model

6.1 Model Definition

The LCC model presented in this manuscript is based in the model described in [52]. The following assumptions are taken into account:

- Property tax, value added tax, etc., and general inflation are constant and included within the annual discount rate.
- Cash required for investment is provided by the enterprise (rather than being borrowed) so the equity rate is 1; the standard LCC model can be written as

$$\mathbf{Y} = \sum_{i=1}^{n} \mathbf{y}_i = \lambda \sum_{i=1}^{n} \mathbf{a}_i \cdot \mathbf{c}_i^T$$

Where $\mathbf{Y}$ is the total cost, $\mathbf{Y} = [Y_1, \ldots, Y_t, \ldots, Y_T]$, Y_t denotes the cumulative cost in year t, the subscript T is the total number of years, y_i indicates the cost of breakdown in category i, and λ signifies the net present value (NPV) factor vector.

In the case study considered in this work, it has been appropriated to select $n = 5$ (CMS investment: $i = 1$; CMS operation: $i = 2$; CMS maintenance: $i = 3$; maintenance reductions by CMS: $i = 4$ and; energy production and energy losses by CMS: $i = 5$). Now, if the term a_{ij} indicates the number of times that the unit cost c_{ij} is incurred, the matrix $\mathbf{A}$ and $\mathbf{C}$ can be described as $\mathbf{A} = [\mathbf{a_1}, \ldots, \mathbf{a_5}]$, and $\mathbf{C} = [\mathbf{c_1}, \ldots, \mathbf{c_5}]$, where $\mathbf{a_i}$ and $\mathbf{c_i}$ are one dimensional arrays of length 4, i.e., $\mathbf{a_i} = [\mathbf{a_{i1}}, \ldots, \mathbf{a_{i4}}]$ and $\mathbf{c_i} = [\mathbf{c_{i1}}, \ldots, \mathbf{c_{i4}}]$.

6.2 CMS Investment Costs

The condition monitoring system investment cost $\mathbf{c_1}$ is related to the costs of capital. These costs include the general investment costs of CMS (capital, installation, regulatory approval, initial testing, software, power and communications), and the costs of the different parts of the CMS installed in the wind turbine (tower, nacelle and blades). The elements of $\mathbf{c_1}$ are therefore described as follows:

c_{11} (General CMS investment costs) = −66,600 €; $a_{11} = 1$;
c_{12} (Tower CMS investment costs) = −7,051.96 €; $a_{12} = 1$
c_{13} (Nacelle CMS investment costs) = −8,900.91 €; $a_{13} = 1$.
c_{14} (Blades CMS investment costs) = −35,151.94 €; $a_{14} = 1$.

6.3 CMS Operation Costs

The CMS operation cost $\mathbf{c_2}$ is the cost incurred by the technical operation process in a period of time. These costs are collected in the CMS general operation costs c_{21} (taken into account the costs of: data acquisition and transmission, software, testing, power consume and human resources). The costs above described are calculated by month, therefore $a_{21} = 12$.

6.4 CMS Maintenance Costs

The CMS maintenance costs $\mathbf{c_3}$ are the cost for condition monitoring system maintenance management processes. The general CMS maintenance cost c_{31} collects the costs of corrective and preventive CMS maintenance costs being $c_{31} = 2,500$ € per year, and $c_{32} = c_{33} = c_{34} = 0$.

6.5 Maintenance Reduction Costs by CMS

An analysis of the cost of overall WT maintenance tasks, with and without CMS, is presented in this sub-section. These costs include preventive and corrective maintenance costs. Preventive and inspection costs are reduced in 75 % with CMS. Corrective maintenance costs are reduced in 40 % with CMS.

c_{41} is related with the general costs and includes preventive costs (ground inspections, cleaning, road network maintenance, general inspections…) and corrective costs (other failures, see failures rates in Table 1).

The tower maintenance reduction cost is represented by c_{42}.

Table 2 Assumptions for calculations

Assumptions	
Hours by year	8,760
Power (kW)	1,600
Efficiency (%)	0.6
Electricity price (€/kWh)	0.05
% Failures reduction by CM	0.4

The nacelle maintenance reduction cost (c_{43}) includes the preventive and corrective maintenance cost of different elements such as generator, gearbox, shaft/bearings, brake system, electrical system.

Finally, c_{44} is related with the maintenance reduction cost in blades.

The values obtained are:

$c_{41} = 946.67$ € per year
$c_{42} = 1{,}306$ € per year
$c_{43} = 34{,}572.17$ € per year
$c_{44} = 7{,}633.33$ € per year

6.6 Energy Production and Energy Losses by CMS

The cost due to production losses is given by the costs difference of production losses with and without CMS. Some assumptions are taken into account for this study and showed in Table 2.

The elements of $\mathbf{c}_5$ are defined as the electricity price (0.05 €/kWh).

The elements of $\mathbf{a}_5$ are the difference of power loss (kWh) between the different elements with and without CMS. The downtime of each component (see Table 1) of the wind turbine is used to calculate the following items.

$a_{51} = 2{,}764.8$ kWh (general production losses)
$a_{52} = 384$ kWh (production losses by the tower)
$a_{53} = 129{,}340$ kWh (production losses by the nacelle)
$a_{54} = 14{,}169.6$ kWh (production losses by the blades)

6.7 Net Present Value

The net present value (NPV) is defined as the sum of the present values (PVs) of the individual cash flows of the same entity. The NPV is defined by the following expression:

$$NPV_t = \frac{CF_t}{(1+k)^{t-1}}$$

where CF_t is the cash flow in year t assuming that all costs are defined using base-year prices, and k is the annual rate of return on investment, referred to as the cost of capital by some researchers. The total discounted cost over t years of life is therefore achieved by the following expression:

$$NPV = -I_0 + \sum_{t=1}^{T} \frac{CF_t}{(1+k)^{t-1}}$$

being I_0 the initial investment. Additionally, assuming that the previous costs remains constant throughout the lifetime of the project, $CF_1 = CF_2 = \ldots = CF_T = CF$, the NPV factor λ is given as follows:

$$\lambda = \sum_{t=1}^{T} \frac{1_t}{(1+k)^{t-1}} = \frac{1}{k}\left[1 - (1+k)^{-T}\right]$$

Considering the above expressions, the NPV value can be calculated as

$$NPV = -I_0 + CF \cdot \lambda = -I_0 + \frac{CF}{k}\left[1 - (1+k)^{-T}\right]$$

6.8 Calculation of LCC

In this study the base year has been set 0. The values that compose the matrixes **A** and **C** are the following:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \\ a_{51} & a_{52} & a_{53} & a_{54} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \\ c_{51} & c_{52} & c_{53} & c_{54} \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} -66600 & -7051.962 & -8900.906 & -35151.94 \\ -2650 & 0 & 0 & 0 \\ -2500 & 0 & 0 & 0 \\ 946.6667 & 1306 & 34572.17 & 7633.333 \\ 0.05 & 0.05 & 0.05 & 0.05 \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 12 & 12 & 12 & 12 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 2764.8 & 384 & 129340.2 & 14169.6 \end{bmatrix}$$

The values obtained from LCC and NPV are illustrated in Fig. 3. This graph shows different curves depending on the annual rate of return (k). The initial

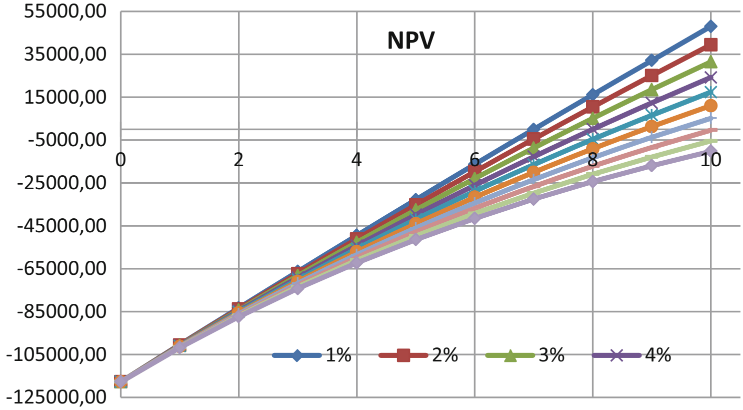


Fig. 3 NPV for different annual rates of return

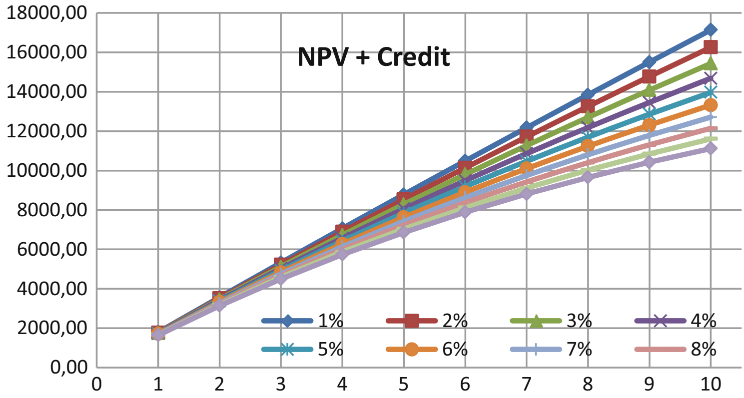


Fig. 4 NPV with the amortization of a credit for initial investment

investment does not recover for larges values of k (8, 9 and 10 %) in the 10 years of operation. For k = 5 and 6 % the payback will be in the 8th year. And curves of low rates (1–4 %) will be recovered in the 7th year. The NPV for a k = 1 % in the 10th year will be 40,000 € and 17,357 € for k = 5 %.

Nowadays the operator firms assume the initial investments with a bank credit. The following figure shows the inclusion in the LCC of an amortization of the initial investment with a rate of 6 % and 10 years. The NPV reaches 17,143 € and 11,122 € for k = 1 % and 10 % respectively. If the credit rate increases the NPV will decrease. The NPV is negative for an interest rate of 9 % being 8.4717 % the value when the NPV is practically zero (Fig. 4).

The LCC model is very sensitive to small variations of the variable reduction of failures using CM. Figure 5 shows the NPV for a 45 % of a reduction of failures

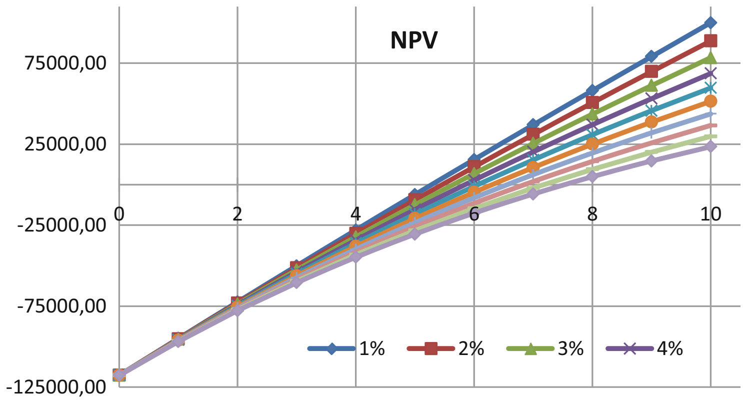


Fig. 5 NPV for different annual rate of return and a 45 % of reduction of failures using CM

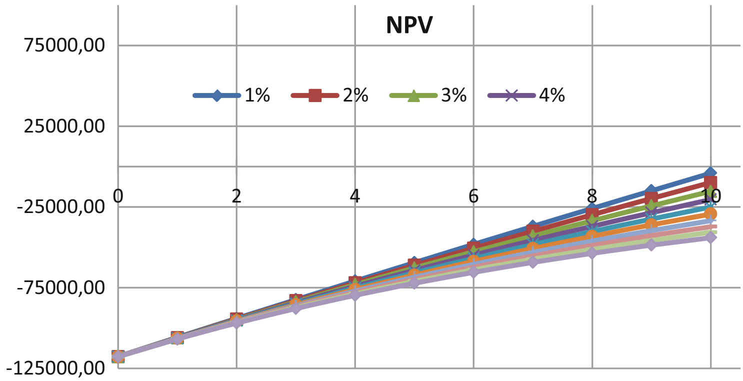


Fig. 6 NPV for different annual rate of return and a 35 % of reduction of failures using CM

using CMS. The payback will be between years five and seven reaching a NPV of 100,000 € or 23,539 € for an annual rate of return of 1 or 10 % respectively (Fig. 5).

Figure 6 shows the NPV for a 35 % of a reduction of failures using CMS. The payback does not recover before the 10th year for all the annual rate of return.

7 Conclusions

The rising wind energy in addition to the increasing number of failures of the larger wind turbines makes necessary the reduction of costs in this industry to make it more competitive in this sector. For this propose the wind energy industry is focused on the reduction of the operation and maintenance (O&M) costs. Condition Monitoring Systems (CMS) are probably the most effective approach to minimize O&M cost and substantially improve the availability, reliability and safety of wind turbines by early detection of the faults. CMS requires knowledge and expertise to analyze the large volume of data collected from the sensors located in the wind turbines. The main objective of this work is the development of a life cycle cost (LCC) model of the CMS for a wind turbine and the analysis of their economic feasibility. The LCC model have been applied to a real case study in Germany finding that the return of the investment will be after the 7th year of operation for an annual rate of return of 1 %. Note that the payback for $k = 8$, 9 and 10 % will not done before the 10th year of operation. The operators usually face the investment by a bank credit to implement a CMS in the WT. The NPV depends on the interest rate to a large degree. Finally, the LCC model is very sensitive to small variations of the reduction of failures using CM.

References

1. Ghenai C (2012) Life cycle analysis of wind turbine, sustainable development. In: Ghenai C (ed) Energy, engineering and technologies - manufacturing and environment, Chap. 2. InTech, ISBN: 978-953-51-0165-9, pp 19–32
2. Fung KT, Scheffler RL, Stolpe J (1981) Wind energy – a utility perspective. IEEE Trans Power Apparatus Syst 100:1176–1182
3. Ezio S, Claudio C (1998) Exploitation of wind as an energy source to meet the world's electricity demand. J Wind Eng Ind Aerod 74–76:375–387
4. Milborrow D (2006) Operation and maintenance costs compared and revealed. Wind Stats 19(3):3
5. Vachon W (2002) Long-term O&M cost of wind turbines based on failure rates and repair costs. In: Proceedings WINDPOWER, American Wind Energy Association annual conference, Portland, OR, June 2002, pp 2–5
6. Tavner PJ, Spinato F, van Bussel GJW, Koutoulakos E (2008) Reliability of different wind turbine concepts with relevance of offshore application. In: European wind energy conference, Brussels, April 2008
7. (2004) Life cycle costing guideline. Total asset management, New South Wales Treasury, ISBN 0 7313 3325 X
8. de Novaes Pires G, Alencar E, Kraj A (2010) Remote conditioning monitoring system for a hybrid wind diesel system-application at Fernando de Naronha Island, Brasil, http://www.ontario-sea.org (19-07-10)
9. Pinar Pérez JM, García Márquez FP, Tobias AM, Papaelias M (2013) Wind turbine reliability analysis. Renew Sustain Energy Rev 23:463–472
10. Wiggelinkhuizen E, Verbruggen T, Xian J, Watson SJ, Giebel G, Norton E et al (2007) CONMOW: condition monitoring for offshore wind farms. In: Proceedings of the 2007 EWEA European wind energy conference (EWEC2007), Milan, Italy, May 2007, pp 7–10
11. Tongdan J, Mechehoul M (2010) Minimize production loss in device testing via condition-based equipment maintenance. IEEE Trans Autom Sci Eng 7(4):958–963
12. Rumsey MA, Paquette JA (2008) Structural health monitoring of wind turbine blades. In: Proceedings of SPIE 2008, vol 6933. 69330E-1-69330E-15. SPIE-2008-6933-14A
13. Scarf PA (2007) A framework for condition monitoring and condition based maintenance. Qual Technol Quant Manag 4(2):301–312
14. Yang WX, Tavner PJ, Crabtree CJ, Wilkinson M (2010) Cost effective condition monitoring for wind turbines. IEEE Trans Ind Electron 57(1):263–271
15. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20(7):1483–1510

16. Nilsson J, Bertling L (2007) Maintenance management of wind power systems using condition monitoring systems – life cycle costs analysis for two case studies. IEEE Trans Energy Conv 22(1):223–229
17. Orsagh RF, Lee H, Watson M, Byington CS, Powers J (2006) Advanced vibration monitoring for wind turbine health management, impact technologies, http://www.impact-tek.com
18. Amirat Y, Benbouzid MEH, Bensaker B, Wamkeue R (2007) Condition monitoring and fault diagnosis in wind energy conversion systems: a review. In: Electric machines and drives conference 2007 (IEMDC07), Antalya, Turkey, pp 1434–1439
19. Wakui T, Yokoyama R (2013) Wind Speed Sensorless performance monitoring based on operating behaviour for stand-alone vertical axis wind turbine. Renew Energy 53(1):49–59
20. Kusiak A, Zhang Z, Verma A (2013) Prediction, operations, and condition monitoring in wind energy. Energy 60(1):1–2
21. Li L, Wenxiu L, Fulei C (2010) Application of AE techniques for the detection of wind turbine using Hilbert-Huang transform. In: PHM'10 prognostics and health management conference, Macao, China, pp 1–7
22. Tsopelas N, Kourousis D, Ladis I, Anastasopulos A, Lekou DJ, Mouzakis F (2012) Health monitoring of operating wind turbine blades with acoustic emission. In: Paipetis AS et al (eds) Emerging technologies in non-destructive testing V. Taylor and Francis, London, pp 347–352
23. Knezevic J (1993) Reliability, maintainability and supportability engineering: a probabilistic approach. McGraw Hill, New York
24. Endrenyi J, McCauley J, Shing C (2001) The present status of maintenance strategies and the impact of maintenance and reliability. IEEE Trans Power Syst 16(4):638–646
25. Hameed Z, Hong YS, Choa YM, Ahn SH, Song CK (2009) Condition Monitoring and fault detection for wind turbines and related algorithms: a review. Renew Sustain Energy Rev 13:1–39
26. Wiesent BR, Schardt M, Koch AW (2012) Gear oil condition monitoring for offshore wind turbines. Available online at http://www.machinerylubrication.com/Read/28782/gear-oil-condition-monitoring (12-06-14)
27. Morfiadakis E, Papadopoulos K, Philippidis TP (2000) Assessment of the stain gauge technique for measurement of wind energy turbine blade loads. Wind Energy 3(1):35–65
28. Schroeder K, Ecke W, Apitz J, Lembke E, Lenschow G (2006) A fiber Bragg grating sensor system monitors operational load in a wind turbine rotor blade. Meas Sci Technol 17(5): 1167–1172
29. Smith BM (1978) Condition monitoring by thermography. NDT Int 11(3):121–122
30. Rumsey MA, Musial W (2001) Application of infrared thermography non destructive testing during wind turbine blade tests. J Solar Energy Eng 123(4):271
31. Zhen L, Zhengjia H, Yanyang Z, Xuefeng C (2008) Bearing condition monitoring based on shock pulse method and improved redundant lifting scheme. Math Comput Simulat 79(3): 318–338
32. Sorensen BF, Lading L, Sendrup P, McGugan M, Debel CP, Kristensen OJD et al (2002) Fundamentals for remote structural health monitoring of wind turbines blades – a preproject. In: Risø-R-1336(EN). Risø National Laboratory, Roskilde, Denmark, May 2002
33. Raisutis R, Jasiuniene E, Sliteris R, Vladisauskas A (2008) The review of non-destructive testing techniques suitable for inspection of the wind turbines blades. Ultragarsas (Ultrasound) 63(1):26–30
34. Caselitz P, Giebhardt J (2003) Fault prediction techniques for offshore wind farm maintenance and repair strategies. In: Proceedings of the EWEC2003
35. Cheng J, Yang Y, Yu D (2010) The envelope order spectrum based on generalized demodulation time-frequency analysis and its application to gear fault diagnosis. Mech Syst Signal Process 24:508–521
36. Tandon N, Nakra BC (1992) Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings. Tribol Int 25(3):205–212

37. Staszewski WJ, Tomlinson GR (1994) Application of the wavelet transform to fault detection in a Spur gear. Mech Syst Signal Process 8:289–307
38. Luo GY, Osypiw D, Irle M (2003) Online vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing. J Vib Control 9:931–947
39. Ocak H, Loparo KA (2001) A new bearing fault detection and diagnosis scheme based on Hidden Markov modeling of vibrations signals. IEEE ICASSP 5:3141–3144
40. Miaoa Q, Makis V (2007) Condition monitoring and classification of rotating machinery using wavelets and Hidden Markov Models. Mech Syst Signal Process 21:840–855
41. Yam RCM, Tse PW, Li L, Tu P (2001) Intelligent predictive decision support system for condition-based maintenance. Int J Adv Manuf Technol 17(5):383–391
42. García Márquez FP, Tobias AM, Pinar Pérez JM, Papaelias M (2012) Condition monitoring of wind turbines: techniques and methods. Renew Energy 46:169–178
43. Fabrycky WJ, Blanchard BS (1991) Life-cycle cost and economic analysis. Prentice Hall, Englewood Cliffs, NJ
44. Castro-Santos L, Prado García G, Diaz-Casas V (2013) Methodology to study the life cycle cost of floating offshore wind farms. In: 10th deep sea wind R&D conference
45. Castro-Santos L, Filgueira Vizoso A, Muñoz Camacho E, Piegiari L (2012) General economic analysis about the wind farms repowering in Spain. J Energy Power Eng 6:1158–1162
46. Nordahl M (2011) The development of a life cycle cost model for an offshore wind farm, Göteborg: Chalmers tekniska högskola. Diploma work – Department of Applied Mechanics, Chalmers University of Technology, Göteborg, Sweden. ISSN 1652-8557
47. Krokoszinski H-J (2003) Efficiency and effectiveness of wind farms – keys to cost optimized operation and maintenance. Renew Energy 28:2165–2178
48. Walford CA (2006) Wind turbine reliability: understanding and minimizing wind turbine operation and maintenance costs. Sandia Report, SAND2006-1100. Sandia National Laboratories, Albuquerque, NM and Livermore, CA
49. Besnard F, Fischer K, Bertling L (2010) Reliability-centred asset maintenance – a step towards enhanced reliability, availability, and profitability of wind power plants. In: 2010 I.E. PES Innovative smart grid technologies conference Europe (ISGT Europe), pp 1–8
50. Hameed Z, Ahn SH, Cho YM (2010) Practical aspects of a condition monitoring system for a wind turbine with emphasis on its design, system architecture, testing and installation. Renew Energy 35:879–894
51. Besnard F, Nilsson J, Bertling L (2010) On the economic benefits of using condition monitoring systems for maintenance management of wind power systems. In: 2010 I.E. 11th international conference on probabilistic methods applied to power systems, pp 160–165
52. García Márquez FP, Lewis RW, Tobias AM, Roberts C (2008) Life cycle costs for railway condition monitoring. Transport Res E 44:1175–1187

Introduction to Multi-attribute Decision Making in Business Analytics

William P. Fox

1 Introduction

Multiple-attribute decision making (MADM) refers to making decisions when there are multiple but a finite list of alternatives and multiple criteria.

Consider a problem where management needs to prioritize or rank order alternative choices: identify key nodes in a business network, pick a contractor or sub-contractor, choose airports, rank recruiting efforts, ranks banking facilities, rank schools or colleges, etc. How does one proceed to accomplish this analytically?

In this chapter we will present four methodologies to rank order or prioritize alternatives based upon multiple criteria. These four methodologies include:

Data Envelopment Analysis (*DEA*)
Simple Average Weighting (*SAW*)
Analytical Hierarchy Process (*AHP*)
Technique of Order Preference by Similarity to Ideal Solution (*TOPSIS*)

For each method, we describe the method and its uses, discuss some strengths and limitations to the method, discuss tips for conducting sensitivity analysis, and present several illustrative examples.

W.P. Fox (✉)
Naval Postgraduate School, Monterey, CA, USA
e-mail: wpfox@nps.edu

© Springer International Publishing Switzerland 2015

F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_4

2 Data Envelopment Analysis (DEA)

2.1 Description and Uses

Data envelopment analysis (DEA) is a relatively new "data input-output driven" approach for evaluating the performance of entities called decision making units (DMUs) that convert multiple inputs into multiple outputs [1]. The definition of a DMU is generic and very flexible. It has been used to evaluate the *performance* or *efficiencies* of hospitals, schools, departments, US Air Force wings, US armed forces recruiting agencies, universities, cities, courts, businesses, banking facilities, countries, regions, etc. According to Cooper [1], DEA has been used to gain insights into activities that were not obtained by other quantitative or qualitative methods.

Charnes et al. [2] described DEA as a mathematical programming model applied to observational data—providing a new way of obtaining empirical estimates of relations. It is formally defined as a methodology directed to frontiers rather than central tendencies.

2.2 Methodology

The model, in simplest terms, may be formulated and solved as a linear programming problem [3, 4]. Although several formulations for DEA exist, we seek the most straight forward formulation in order to maximize an efficiency of a DMU as constrained by inputs and outputs as shown in Eq. (1). As an option, we might normalize the metric inputs and outputs for the alternatives if poorly scaled. Otherwise, we will call this matrix, **X**, with entries x_{ij}. We define an efficiency unit as E_i *for* $i = 1,2,\ldots,nodes$. We let w_i be the weights or coefficients for the linear combinations. Further, we restrict any efficiency from being larger than one. This gives the following linear programming formulation for single outputs but multiple inputs:

$$
\begin{aligned}
&Max\, E_i \\
&subject\, to \\
&\sum_{i=1}^{n} w_i x_{ij} - E_i = 0, j = 1, 2, \ldots \\
&E_i \leq for \quad all \quad i
\end{aligned}
\tag{1}
$$

For multiple inputs and outputs, we recommend the formulations provided by Winston [3] and Trick [5] using Eq. (2). For any DMU_0, let X_i be the inputs and Y_i be the outputs. Let X_0 and Y_0 be the *DMU* being modeled.

$$\begin{array}{l} Min\,\theta \\ subject\,to \\ \Sigma\lambda_i X_i \leq \theta X_0 \\ \Sigma\lambda_i Y_i \leq Y_0 \\ \qquad\qquad \lambda i \geq 0 \\ Non\text{-}negativity \end{array} \tag{2}$$

2.3 Strengths and Limitations to DEA

DEA can be a very useful tool when used wisely. A few of the strengths that make it extremely useful are provided by Trick [6]:

- DEA can handle multiple input and multiple output models.
- DEA doesn't require an assumption of a functional form relating inputs to outputs.
- DMUs are directly compared against a peer or combination of peers.
- Inputs and outputs can have very different units. For example, X_1 could be in units of lives saved and X_2 could be in units of dollars without requiring any *a priori* tradeoff between the two.

The same characteristics that make DEA a powerful tool can also create limitations. An analyst should keep these limitations in mind when choosing whether or not to use DEA.

- Since DEA is an extreme point technique, noise in the data such as measurement error can cause significant problems.
- DEA is good at estimating "relative" efficiency of a DMU but it converges very slowly to "absolute" efficiency. In other words, it can tell you how well you are doing compared to your peers but not compared to a "theoretical maximum."
- Since DEA is a nonparametric technique, statistical hypothesis tests are difficult and are the focus of ongoing research.
- Since a standard formulation of DEA with multiple inputs and outputs creates a separate linear program for each DMU, large problems can be computationally intensive.
- Linear programming does not ensure all weights are considered. We find that the value for weights are only for those that optimally determine an efficiency rating. If having all criteria weighted (inputs, outputs) is essential to the decision maker then do not use DEA.

2.4 Sensitivity Analysis

According to Neralic [7], an increase in any output cannot worsen an already achieved efficiency rating nor can a decrease in inputs alone worsen an already achieved efficiency rating. As a result in our examples we only decrease outputs and increase inputs [7]. We will illustrate some sensitivity analysis, as applicable, in our examples.

2.5 Illustrative Examples

Example 1. Manufacturing Consider the following manufacturing process from Winston [3] where we have three DMUs each of which has two inputs and three outputs as shown in the data table.

DMU	Input #1	Input #2	Output #1	Output #2	Output #3
1	*5*	*14*	*9*	*4*	*16*
2	*8*	*15*	*5*	*7*	*10*
3	*7*	*12*	*4*	*9*	*13*

Since no units are given and the scales are similar so we decide not to normalize the data. We define the following decision variables:

t_i = value of a single unit of output of *DMU i*, for $i = 1,2,3$

w_i = cost or weights for one unit of inputs of *DMU i*, for $i = 1,2$

$efficiency_i = DMU_i$ = (total value of *i*'s outputs)/(total cost of *i*'s inputs), for $i = 1,2,3$

The following modeling assumptions are made:

1. No DMU will have an efficiency of more than 100 %.
2. If any efficiency is less than 1, then it is inefficient.
3. We will scale the costs so that the costs of the inputs equals 1 for each linear program. For example, we will use $5w_1 + 14w_2 = 1$ in our program for DMU_1.
4. All values and weights must be strictly positive, so we use a constant such as 0.0001 in lieu of 0.

To calculate the efficiency of DMU_1, we define the linear program using Eq. (2) as:

Maximize $DMU_1 = 9t_1 + 4t_2 + 16t_3$
Subject to
$-9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0$
$-5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0$
$-4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0$
$5w_1 + 14w_2 = 1$
$t_i \geq 0.0001,\ i = 1,2,3$
$w_i \geq 0.0001,\ i = 1,2$
Non-negativity

To calculate the efficiency of DMU_2, we define the linear program as:

Maximize $DMU_2 = 5t_1 + 7t_2 + 10t_3$
Subject to
$-9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0$
$-5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0$
$-4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0$
$8w_1 + 15w_2 = 1$
$t_i \geq 0.0001,\ i = 1,2,3$
$w_i \geq 0.0001,\ i = 1,2$
Non-negativity

To calculate the efficiency of DMU_3, we define the linear program as

Maximize $DMU_3 = 4t_1 + 9t_2 + 13t_3$
Subject to
$-9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0$
$-5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0$
$-4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0$
$7w_1 + 12w_2 = 1$
$t_i \geq 0.0001,\ i = 1,2,3$
$w_i \geq 0.0001,\ i = 1,2$
Non-negativity

The linear programming solutions show the efficiencies as $DMU_1 = DMU_3 = 1$, $DMU_2 = 0.77303$.

Interpretation DMU_2 is operating at 77.303 % of the efficiency of DMU_1 and DMU_3. Management could concentrate some improvements or best practices from DMU_1 or DMU_3 for DMU_2. An examination of the dual prices for the linear program of DMU_2 yields $\lambda_1 = 0.261538$, $\lambda_2 = 0$, and $\lambda_3 = 0.661538$. The average output vector for DMU_2 can be written as:

$$0.261538\begin{bmatrix} 9 \\ 4 \\ 16 \end{bmatrix} + 0.661538\begin{bmatrix} 4 \\ 9 \\ 13 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 12.785 \end{bmatrix}$$

and the average input vector can be written as

$$0.261538\begin{bmatrix} 5 \\ 14 \end{bmatrix} + 0.661538\begin{bmatrix} 7 \\ 12 \end{bmatrix} = \begin{bmatrix} 5.938 \\ 11.6 \end{bmatrix}.$$

In our data, output 3 is 10 units. Thus, we may clearly see the inefficiency is in output #3 where 12.785 units are required. We find that they are short 2.785 units (12.785 – 10 = 2.785). This helps focus on treating the inefficiency found for output #3.

Sensitivity Analysis Sensitivity analysis in a linear program is sometimes referred to as "what if" analysis. Let's assume that without management engaging some additional training for DMU_2 that DMU_2 output #3 dips from 10 to 9 units of output while the input 2 h increases from 15 to 16 h. We find that these changes in the *technology coefficients* are easily handled in resolving the LPs. Since DMU_2 is affected, we might only modify and solve the LP concerning DMU_2. We find with these changes that DMU's efficiency is now only 74 % as effective as DMU_1 and DMU_3.

Example 2. Social Networks Consider the Kite Social Network from Krackhardt [8] shown in Fig. 1.

ORA [9], a social network software, was used to obtain the metrics for this network. A subset of the output is shown in Table 1. We restricted the metrics presented: Total Centrality (TC), Eigenvector Centrality (EC), In-Closeness (IC), Out-Closeness (OC), Information Centrality (INC), and Betweenness (Betw), whose definitions can be found in recent social network literature [10, 11].

We formulate the linear program from Eq. (1) to measure the efficiency of the nodes. We define the decision variables:

u_i = efficiency of node i, $i = 1,2,3,\ldots 10$

w_j = weight of input j, $j = 1,2,3,4,5$

Maximize u_1

Subject to

$\boldsymbol{A} = \boldsymbol{0}$

$u_i \leq = 1$ for $i = 1,2,3,\ldots,10$

where

$\boldsymbol{A} =$

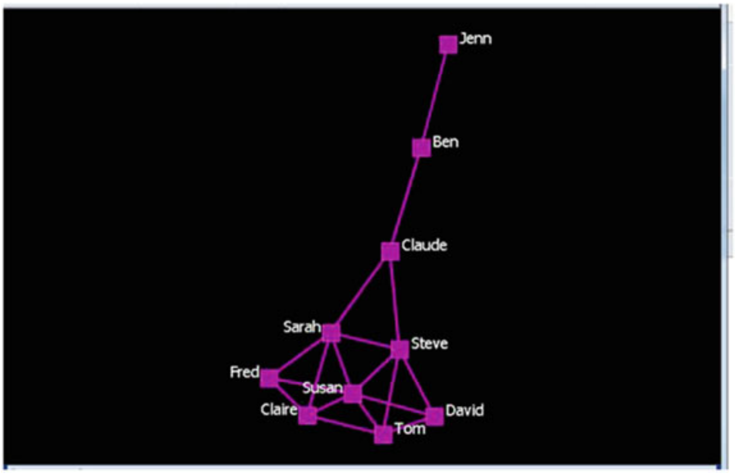


Fig. 1 Kite network diagram from ORA (Carley, 2011)

Table 1 ORA outputs for the Kite network

TC	EC	IC	OC	INC	Betw
0.1806	0.1751	0.0920	0.1081	0.1088	0.2022
0.1389	0.1375	0.0997	0.1003	0.1131	0.1553
0.1250	0.1375	0.1107	0.0892	0.1131	0.1042
0.1111	0.1144	0.0997	0.1003	0.1009	0.0194
0.1111	0.1144	0.0997	0.1003	0.1009	0.0194
0.0833	0.0938	0.0997	0.1003	0.0975	0.0000
0.0833	0.0938	0.0997	0.1003	0.0975	0.0000
0.0833	0.1042	0.0997	0.1003	0.1088	0.3177
0.0556	0.0241	0.0997	0.1003	0.0885	0.1818
0.0278	0.0052	0.0997	0.1003	0.0707	0.0000

$$\begin{bmatrix}
0.180555556\,w1 + 0.175080826\,w2 + 0.091993186\,w3 + 0.10806175\,w4 + 0.108849307\,w5 + 0.202247191\,w6 - u1 \\
0.138888889\,w1 + 0.137527978\,w2 + 0.099659284\,w3 + 0.100343053\,w4 + 0.113090189\,w5 + 0.15526047\,w6 - u2 \\
0.125\,w1 + 0.137527978\,w2 + 0.110732538\,w3 + 0.089193825\,w4 + 0.113090189\,w5 + 0.104187947\,w6 - u3 \\
0.111111111\,w1 + 0.114399403\,w2 + 0.099659284\,w3 + 0.100343053\,w4 + 0.100932994\,w5 + 0.019407559\,w6 - u4 \\
0.111111111\,w1 + 0.114399403\,w2 + 0.099659284\,w3 + 0.100343053\,w4 + 0.100932994\,w5 + 0.019407559\,w6 - u5 \\
0.083333333\,w1 + 0.093757772\,w2 + 0.099659284\,w3 + 0.100343053\,w4 + 0.097540288\,w5 - u6 \\
0.083333333\,w1 + 0.093757772\,w2 + 0.099659284\,w3 + 0.100343053\,w4 + 0.097540288\,w5 - u7 \\
0.083333333\,w1 + 0.104202935\,w2 + 0.099659284\,w3 + 0.100343053\,w4 + 0.108849307\,w5 + 0.317671093\,w6 - u8 \\
0.055555556\,w1 + 0.024123352\,w2 + 0.099659284\,w3 + 0.100343053\,w4 + 0.088493073\,w5 + 0.181818182\,w6 - u9 \\
0.027777778\,w1 + 0.005222581\,w2 + 0.099659284\,w3 + 0.100343053\,w4 + 0.070681368\,w5 - u10
\end{bmatrix}$$

The linear programming solution is

	DV	
Susan	DMU1	1
Steven	DMU2	0.785511
Sarah	DMU3	0.785511
Tom	DMU4	0.653409
Claire	DMU5	0.653409
Fred	DMU6	0.535511
David	DMU7	0.535511
Claudia	DMU8	0.59517
Ben	DMU9	0.137784
Jennifer	DMU10	0.02983

w1	0
w2	5.711648
w3	0
w4	0
w5	0
w6	0

Interpretation We interpret the linear programming solution as follows: Player 1, $u_1 =$ Susan, is rated most influential followed closely by Sarah and Steven. Additionally, we see the most important criterion in solving the optimal problem was the eigenvector centrality, w_2, of the network. The solution, translated back into the original variables is found as:

$Susan = 1$, $Sarah = 0.78551$, $Steven = 0.78551$, $Claire = 0.6534$, $Tom = 0.6534$, $Fred = 0.5355$, $David = 0.5355$, $Claudia = 0.5951$, $Ben = 0.1377$, and $Jennifer = 0.02983$ while $w1 = w3 = w4 = w5 = w6 = 0$ and $w2 = 5.7116$.

Since the output metrics are network metrics calculated from ORA we do not recommend any sensitivity analysis for this type problem unless your goal is to improve the influence (efficiency) of another member of the network If so, then the finding the *dual prices* (*shadow prices*) would be required as shown in the first example.

3 Simple Additive Weighting (SAW) Method

3.1 Description and Uses

This is also called the weighted sum method [12] and is the simplest, and still one of the widest used of the MADM methods. Its simplistic approach makes it easy to use. Depending on the type relational data used, we might either want the larger average or the smaller average.

3.2 Methodology

Here, each criterion (attribute) is given a weight, and the sum of all weights must be equal to 1. Each alternative is assessed with regard to every criterion (attribute). The overall or composite performance score of an alternative is given simply by Eq. (3) with *m* criteria.

$$P_i = \left(\sum_{j=1}^{m} w_j m_{ij}\right)/m \tag{3}$$

Previously, it was argued that SAW should be used only when the decision criteria can be expressed in identical units of measure (*e.g.*, only dollars, only pounds, only seconds, *etc.*). However, if all the elements of the decision table are normalized, then this procedure can be used for any type and any number of criteria. In that case, Eq. (3) will take the following form still with *m* criteria shown as Eq. (4):

$$P_i = \left(\sum_{j=1}^{m} W_j m_{ijNormalized}\right)/m \tag{4}$$

where ($m_{ijNormalized}$) represents the normalized value of m_{ij}, and P_i is the overall or composite score of the alternative A_i. The alternative with the highest value of P_i is considered the best alternative.

3.3 Strengths and Limitations

The strengths are the ease of use and the normalized data allow for comparison across many differing criteria. Limitations include larger is always better or smaller is always better. There is not the flexibility in this method to state which criterion should be larger or smaller to achieve better performance. This makes gathering useful data of the same relational value scheme (larger or smaller) essential.

3.4 Sensitivity Analysis

Sensitivity analysis should be applied to the weighting scheme employed to determine how sensitive the model is to the weights. Weighting can be arbitrary for a decision maker or in order to obtain weights you might choose to use a scheme to perform pairwise comparison as we show in AHP that we discuss later. Whenever subjectivity enters into the process for finding weights, then sensitivity analysis is recommended. Please see later sections for a suggested scheme for dealing with sensitivity analysis for individual criteria weights.

3.5 Illustrative Examples

Example 1. Car Selection (data from Consumer's Report [13] and US News and World Report on-line data) We are considering six cars: Ford Fusion, Toyota Prius, Toyota Camry, Nissan Leaf, Chevy Volt, and Hyundai Sonata. For each car we have data on seven criteria that were extracted from Consumer's Report and US News and World Report data sources. They are *cost*, *mpg city*, *mpg highway*, *performance*, *interior and style*, *safety*, and *reliability*. We provide the extracted information in the Table 2.

Table 2 Raw data

Cars	Cost ($000)	MPG City	MPG HW	Performance	Interior and style	Safety	Reliability
Prius	27.8	44	40	7.5	8.7	9.4	3
Fusion	28.5	47	47	8.4	8.1	9.6	4
Volt	38.668	35	40	8.2	6.3	9.6	3
Camry	25.5	43	39	7.8	7.5	9.4	5
Sonata	27.5	36	40	7.6	8.3	9.6	5
Leaf	36.2	40	40	8.1	8.0	9.4	3

Initially, we might assume all weights are equal to obtain a baseline ranking. We substitute the rank orders (1st to 6th) for the actual data. We compute the average rank attempting to find the best ranking (smaller is better). We find our rank ordering is Fusion, Sonata, Camry, Prius, Volt, and Leaf.

3.6 SAW Using Rank Ordering of Data by Criteria

	Cost	MPG City	MPG HW	Perf	Interior style	Safety	Reliability		Rank
Prius	3	2	2	6	1	2	4	**2.857143**	**4**
Fusion	4	1	1	1	3	1	3	**2**	**1**
volt	6	6	2	2	6	1	4	**3.857143**	**6**
Camry	1	3	3	4	5	2	1	**2.714286**	**3**
Sonata	2	5	2	5	2	1	1	**2.571429**	**2**
Leaf	5	4	2	2	4	2	4	**3.285714**	**5**

Next, we apply a scheme to the weights and still use the ranks 1–6 as before. Perhaps we apply a technique similar to the pairwise comparison that we will discuss in the AHP Sect. 4. Using the pairwise comparison to obtain new weights, we obtain a new ordering: Camry, Sonata, Fusion, Prius, Leaf, and Volt. The changes in results of the rank ordering differ from using equal weights shows the sensitivity that the model has to the given criteria weights. We assume the criteria in order of importance are: cost, reliability, MPG City, safety, MPG HW, performance, interior and style.

Cars	Weighted Cost	Weighted MPG City	Weighted MPG HW	Weighted Perf	Weighted Interior style	Weighted Safety	Reliability	Average	Ranking
Prius	0.933468	0.267228	0.191572	0.330411637	0.049997	0.258743	17.77726	**0.33857**	**4**
Fusion	1.244624	0.133614	0.095786	0.055068606	0.149991	0.129372	13.33294	**0.301409**	**2**
Volt	1.866936	0.801684	0.191572	0.110137212	0.299982	0.129372	17.77726	**0.566614**	**6**
Camry	0.311156	0.400842	0.287359	0.220274424	0.249985	0.258743	4.444315	**0.28806**	**1**
Sonata	0.622312	0.66807	0.191572	0.275343031	0.099994	0.129372	4.444315	**0.331111**	**3**
Leaf	1.55578	0.534456	0.191572	0.110137212	0.199988	0.258743	17.77726	**0.475113**	**5**

We use pairwise comparisons to obtain weights:

Unequal weights

Weights	0.311156	0.133614	0.095786	0.055069	0.049997069	0.129372	0.225007
	Cost	MPG City	MPG HW	Perf	Interior style	Safety	Reliability

Using these weights and applying to the previous ranking we obtain values that we average and we select the smaller average.

SAW Using Raw Data

We could use the raw data directly from Table 2 except cost. Now, only *cost* represents a value where smaller is better so we can replace cost with its reciprocal. So *1/cost* represents a variable where larger is better. If we use the criteria weights from the previous results and our raw data replacing *cost* with *1/cost*, we obtain a final ranking based upon larger values are better: Fusion (1.972), Camry (1.805), Prius (1.780), Leaf (1.703), Sonata (1.693), Volt (1.599)

Sensitivity Analysis

We suggest employing sensitivity analysis on the criteria weights that are used in the model as presented and described in Sects. 4 and 5.

Example 2. Kite Network We revisit the Kite Network described earlier. Here we present two methods that will work on the data from example 2 from the previous section. Method I representing transforming the output data into rankings from 1st to last place. Then we apply the weights and average all the values. We rank them smaller to larger to represent the alternative choices. We present only results using the pairwise compare criteria to obtain the weighted criteria.

Weights	0.153209	0.144982	0.11944	0.067199	0.157688	0.357482			
Susan	1	1	10	1	3	2			
Steve	2	2	2	2	1	4			
Sarah	3	2	1	10	1	7			
Tom	4	4	2	2	5	5			
Claire	4	4	2	2	5	5			
Fred	6	7	2	2	7	8			
David	6	7	2	2	7	8			
Claudia	6	6	2	2	3	1			
Ben	9	9	2	2	9	3			
Jennifer	10	10	2	2	10	8			
Susan	0.153209	0.144982	1.194396	0.067199	0.473064	0.714965	0.457969	Steve	0.426213
Steve	0.306418	0.289964	0.238879	0.134398	0.157688	1.42993	0.426213	Susan	0.457969
Sarah	0.459627	0.289964	0.11944	0.67199	0.157688	2.502377	0.700181	Claudia	0.498828
Tom	0.612835	0.579928	0.238879	0.134398	0.78844	1.787412	0.690316	Tom	0.690316
Claire	0.612835	0.579928	0.238879	0.134398	0.78844	1.787412	0.690316	Claire	0.690316
Fred	0.919253	1.014875	0.238879	0.134398	1.103816	2.859859	1.04518	Sarah	0.700181
David	0.919253	1.014875	0.238879	0.134398	1.103816	2.859859	1.04518	Ben	0.924772
Claudia	0.919253	0.869893	0.238879	0.134398	0.473064	0.357482	0.498828	Fred	1.04518
Ben	1.37888	1.304839	0.238879	0.134398	1.419192	1.072447	0.924772	David	1.04518
Jennifer	1.532089	1.449821	0.238879	0.134398	1.576879	2.859859	1.298654	Jennifer	1.298654

Method I rankings: Steve, Susan, Claudia. Tom, Claire, Sarah, Ben Fred, David, and Jennifer.

Method II uses the raw metrics data and the weights as above where larger values are better.

Susan	0.027663	0.025384	0.010988	0.007262	0.017164	0.0723	0.026793	Claudia	0.029541
Steve	0.021279	0.019939	0.011903	0.006743	0.017833	0.055503	0.0222	Susan	0.026793
Sarah	0.019151	0.019939	0.013226	0.005994	0.017833	0.037245	0.018898	Steve	0.0222
Tom	0.017023	0.016586	0.011903	0.006743	0.015916	0.006938	0.012518	Sarah	0.018898
Claire	0.017023	0.016586	0.011903	0.006743	0.015916	0.006938	0.012518	Ben	0.018268
Fred	0.012767	0.013593	0.011903	0.006743	0.015381	0	0.010065	Tom	0.012518
David	0.012767	0.013593	0.011903	0.006743	0.015381	0	0.010065	Claire	0.012518
Claudia	0.012767	0.015108	0.011903	0.006743	0.017164	0.113562	0.029541	Fred	0.010065
Ben	0.008512	0.003497	0.011903	0.006743	0.013954	0.064997	0.018268	David	0.010065
Jennifer	0.004256	0.000757	0.011903	0.006743	0.011146	0	0.005801	Jennifer	0.005801

The results are Claudia, Susan, Steven, Sarah, Ben, Tom, Claire, Fred, David, and Jennifer. Although the top three are the same their order is different. The model is sensitive both to the inputs format and the weights.

Sensitivity Analysis

We can apply sensitivity analysis to the weights, in controlled manner, and determine each changes impact on the final rankings. We recommend using Eq. (1) to modify the weights. This is discussed in Sects. 4 and 5.

4 Analytical Hierarchy Process (AHP)

4.1 Description and Uses

AHP is a multi-objective decision analysis tool first proposed by Saaty [14]. It is designed when either subjective and objective measures or just subjective measures are being evaluated in terms of a set of alternatives based upon multiple criteria, organized in a hierarchical structure. At the top level, the criteria are evaluated or weighted, and at the bottom level the alternatives are measured against each criterion. The decision maker assesses their evaluation by making pairwise comparisons in which every pair is subjectively or objectively compared. The subjective method involves a 9 point scale that we present later in Table 3.

We only desire to briefly discuss the elements in the framework of AHP. This can be described as a method to decompose a problem into sub-problems. In most decisions, the decision maker has a choice among many alternatives. Each alternative has a set of attributes or characteristics that can be measured, either subjectively or objectively. We will call these attributes, criteria. The attribute elements of the hierarchal process can relate to any aspect of the decision problem—tangible or intangible, carefully measured or roughly estimated, well- or poorly-understood—anything at all that applies to the decision at hand.

Table 3 Saaty's 9-point scale

Intensity of importance in pair-wise comparisons	Definition
1	Equal Importance
3	Moderate Importance
5	Strong Importance
7	Very Strong Importance
9	Extreme Importance
2, 4, 6, 8	For comparing between the above
Reciprocals of above	In comparison of elements *i* and *j* if *i* is 3 compared to *j*, then j is 1/3 compared to *i*
Rationale	Force consistency; measure values available

We state simply that in order to perform AHP we need an objective and a set of alternatives, each with criteria (attributes) to compare. Once the hierarchy is built, the decision makers systematically evaluate the various elements pairwise (by comparing them to one another two at a time), with respect to their impact on an element above them in the hierarchy. In making the comparisons, the decision makers can use concrete data about the elements, but they typically use their judgments about the elements' relative meaning and importance. It is the essence of the AHP that human judgments, and not just the underlying information, both can be used in performing the evaluations.

The AHP converts these evaluations to numerical values that can be processed and compared over the entire range of the problem. A numerical weight or priority is derived for each element of the hierarchy, allowing diverse and often incommensurable elements to be compared to one another in a rational and consistent way. This capability distinguishes the AHP from other decision making techniques.

In the final step of the process, numerical priorities are calculated for each of the decision alternatives. These numbers represent the alternatives' relative ability to achieve the decision goal, so they allow a straightforward consideration of the various courses of action.

While it can be used by individuals working on straightforward decisions, the Analytic Hierarchy Process (AHP) is most useful where teams of people are working on complex problems, especially those with high stakes, involving human perceptions and judgments, whose resolutions have long-term repercussions. It has unique advantages when important elements of the decision are difficult to quantify or compare, or where communication among team members is impeded by their different specializations, terminologies, or perspectives.

Decision situations to which the AHP can be applied include the following where we desire ranking:

- Choice—The selection of one alternative from a given set of alternatives, usually where there are multiple decision criteria involved.
- Ranking—Putting a set of alternatives in order from most to least desirable

- Prioritization—Determining the relative merit of members of a set of alternatives, as opposed to selecting a single one or merely ranking them
- Resource allocation—Apportioning resources among a set of alternatives
- Benchmarking—Comparing the processes in one's own organization with those of other best-of-breed organizations
- Quality management—Dealing with the multidimensional aspects of quality and quality improvement
- Conflict resolution—Settling disputes between parties with apparently incompatible goals or positions

4.2 Methodology of the Analytic Hierarchy Process

The procedure for using the AHP can be summarized as:

Step 1. Build the hierarchy for the decision

Goal	*Select the best alternative*
Criteria	$c_1, c_2, c_3, \ldots, c_m$
Alternatives:	$a_1, a_2, a_3, \ldots, a_n$

Step 2. Judgments and Comparison

Build a numerical representation using a 9-point scale in a pairwise comparison for the attributes criterion and the alternatives. The goal, in AHP, is to obtain a set of eigenvectors of the system that measures the importance with respect to the criterion. We can put these values into a matrix or table based on the values from Table 3.

We must ensure that this pairwise matrix is consistent according to Saaty's scheme to compute the Consistency Ratio, *CR*. The value of *CR* must be less than or equal to 0.1 to be considered consistent. Saaty's computed the random index, *RI*, for random matrices for up to ten criteria is given in Table 4.

Next, we approximate the largest eigenvalue, λ, using the power method [15]. We compute the consistency index, *CI*, using Eq. (5):

$$CI = (\lambda - n)/(n - 1) \tag{5}$$

Then we compute the *CR* using:

$$CR = CI/RI$$

If $CR \leq 0.1$, then our pairwise comparison matrix is consistent and we may continue the AHP process. If not, we must go back to our pairwise comparison and fix the

Table 4 Random indexes

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.52	0.89	1.1	1.24	1.35	1.4	1.45	1.49

inconsistencies until the $CR \leq 0.1$. In general, the consistency ensures that if $A > B$, $B > C$, that $A > C$ for all A, B, and C all of which can be criteria or alternatives related by pairwise comparisons.

Step 3. Finding all the eigenvectors combined in order to obtain a comparative ranking.

Step 4. After the $m \times 1$ criterion weights are found and the $n\ x\ m$ matrix for n alternatives by m criterion, we use matrix multiplication to obtain the $n \times 1$ final rankings.

Step 5. We order the final ranking.

4.3 Strengths and Limitations of AHP

Like all modeling methods, the AHP has strengths and limitations. The main advantage of the AHP is its ability to rank choices in the order of their effectiveness in meeting conflicting objectives. If the judgments made about the relative importance of criteria and those about the alternatives' ability to satisfy those objectives, have been made in good faith and effort, then the AHP calculations lead to the logical consequence of those judgments. It is quite hard, but not impossible, to manually change the pairwise judgments to get some predetermined result. A further strength of the AHP is its ability to detect inconsistent judgments in the pairwise comparisons using the CR value.

The limitations of the AHP are that it only works because the matrices are all of the same mathematical form—known as a positive reciprocal matrix. The reasons for this are explained in Saaty's book, which is not for the mathematically daunted, so we will simply state that point. To create such a matrix requires that, if we use the number 9 to represent 'A is absolutely more important than B', then we have to use 1/9 to define the relative importance of B with respect to A. Some people regard that as reasonable; others do not.

Some suggest a drawback is in the possible scaling. However, understanding that the final values obtained simply say that one alternative is relatively better than another alternative. For example, if the AHP values for alternatives $\{A,B, C\}$ found were (0.392, 0.406, 0.204) then they only imply that alternatives A and B are about equally good at approximately 0.4, while C is worse at 0.2. It does not mean that A and B are twice as good as C.

In less clear-cut cases, it would not be a bad thing to change the rating scale and see what difference it makes. If one option consistently scores well with different scales, it is likely to be a very robust choice.

In short, the AHP is a useful technique for discriminating between competing options in the light of a range of objectives to be met. The calculations are not complex and, while the AHP relies on what might be seen as a mathematical trick, you don't need to understand the mathematics to use the technique. Be aware that it only shows relative values.

Although AHP has been used in many applications of the public and private sectors, Hartwich [16] noted several limitations. First and foremost, AHP was criticized for not providing sufficient guidance about structuring the problem to be solved, forming the levels of the hierarchy for criteria and alternatives, and aggregating group opinions when team members are geographically dispersed or are subject to time constraints. Team members may carry out rating items individually or as a group. As the levels of hierarchy increase, so does the difficulty and time it takes to synthesize weights. One remedy in preventing these problems is by conducting “AHP Walk-throughs” (i.e., meetings of decision-making participants who review the basics of the AHP methodology and work through examples so that concepts are thoroughly and easily understood).

Another critique of AHP is the “rank reversal” problem, i.e., changes in the importance ratings whenever criteria or alternatives are added-to or deleted-from the initial set of alternatives compared. Several modifications to AHP have been proposed to cope with this and other related issues. Many of the enhancements involved ways of computing, synthesizing pairwise comparisons, and/or normalizing the priority and weighting vectors. We mention now that TOPSIS corrects this rank reversal issue.

4.4 Sensitivity Analysis

Since AHP, at least in the pairwise comparisons, is based upon subjective inputs using the *9*-point scale then sensitivity analysis is extremely important. Leonelli [17] in his master’s thesis, outlines procedures for sensitivity analysis to enhance decision support tools including numerical incremental analysis of a weight, probabilistic simulations, and mathematical models. How often do we change our minds about the relative importance of an object, place, or thing? Often enough that we should alter the pairwise comparison values to determine how robust our rankings are in the AHP process. We suggest doing enough sensitivity analysis to find the “break-point” values, if they exist, of the decision maker weights that change the rankings of our alternatives. Since the pairwise comparisons are subjective matrices compiled using the Saaty method, we suggest as a minimum a “trial and error” sensitivity analysis using the numerical incremental analysis of the weights.

Chen [18] grouped sensitivity analysis into three main groups: numerical incremental analysis, probabilistic simulations, and mathematical models The numerical incremental analysis, also known as One-at-a-time (OAT) or “trial and error” works by incrementally changing one parameter at a time, finding the new solution and showing graphically how the ranks change. There exist several variations of this method by Barker [19] and Hurley [20]. Probabilistic simulations employs Monte Carlo simulation by Butler [21] that allows random changes in the weights and simultaneously explores the effect on the ranks. Modeling may be used when it is possible to express the relationship between the input data and the solution results. We used Eq. (6) by Alinezhad et al. [22] for adjusting weights:

$$w_j^{'} = \frac{1 - w_p^{'}}{1 - w_p} w_j \tag{6}$$

where w_j' is the new weight and w_p is the original weight of the criterion to be adjusted and w_p' is the value after the criterion was adjusted. We found this to be an easy method to adjust weights to reenter back into our model.

4.5 Illustrative Examples

Example 1. Car Selection We revisit Car Selection with our raw data presented in Table 2 to illustrate AHP in selecting the best alternative based upon pairwise comparisons of the decision criteria.

Step 1. Build the hierarchy and prioritize the criterion from your highest to lower priority.

Goal	*Select the best car*
Criteria	$c_1, c_2, c_3, \ldots, c_m$
Alternatives:	$a_1, a_2, a_3, \ldots, a_n$

For our cars example we choose the priority as follows: Cost, MPG City, Safety, Reliability, MPG Highway, Performance, and Interior and Style. Putting these in a priority order allows for an easier assessment of the pairwise comparisons. We used an Excel template prepared for these pairwise comparisons.

Step 2. Perform the pairwise comparisons using Saaty's 9-point scale. We use an Excel template created to organize the pairwise comparisons.

	Element A		Element B	More Important	Intensity (1-9)
1	Cost	compared with	MPG_city	A	2
2			MPG_HW	A	2
3			Safety	A	3
4			Reliability	A	4
5			Performance	A	5
6			Interior & Style	A	6
7					
1	MPG_city	compared with	MPG_HW	A	2
2			Safety	A	3
3			Reliability	A	4
4			Performance	A	5
5			Interior & Style	A	5
6					
1	MPG_HW	comp. with	Safety	A	2
2			Reliability	A	2
3			Performance	A	3
4			Interior & Style	A	3
5					
1	Safety	comp. with	Reliability	A	1
2			Performance	A	2
3			Interior & Style	A	3
4					
1	Reliability	vs	Performance	A	2
2			Interior & Style	A	3
3					
1	Performance	vs	Interior & Style	A	2
2					

This yields the following decision criterion matrix,

	Cost	MPG_City	MPG_HW	Safety	Reliability	Performance	Interior and Style
Cost	1	2	2	3	4	5	6
MPG_City	1/2	1	2	3	4	5	5
MPG-HW	1/2	1/2	1	2	2	3	3
Safety	1/3	1/3	1/2	1	1	2	3
Reliability	1/4	1/4	1/2	1	1	2	3
Performance	1/5	1/5	1/3	1/2	1	1	2
Interior and Style	1/6	1/5	1/3	1/3	1/3	1/2	1

We check the *CR*, the consistency ratio, to ensure it is less than 0.1. For our pairwise decision matrix the $CR = 0.00695$. Since the $CR < 0.1$, we continue. We find the *eigenvector* for the decision weights:

Cost	0.342407554
City	0.230887543
hw	0.151297361
Safety	0.094091851
Reliability	0.080127732
Performance	0.055515667
Interior and style	0.045672293

Step 3. For the alternatives, we either have the data as we obtained it for each car under each decision criterion or we can use pairwise comparisons by criteria for how each car fares versus its competitors. In this example, we take the raw data from before except now we will use *1/cost* to replace *cost* before we normalize the columns.

We have other options for dealing with a criteria and variable like *cost*. Thus, we have three courses of action, COA, (1) use *1/cost* to replace *cost*, (2) use a pairwise comparison using the 9 point scale, or (3) remove *cost* from a criteria and a variable, run the analysis, and then do a *benefit/cost* ratio to re-rank the results.

Step 4. We multiply the matrix of the normalized raw data from Consumer Reports and the matrix of weights to obtain the rankings. Using COA (1) in step 3, we obtain the following results:

Car	Value
Fusion	0.180528
Camry	0.178434
Prius	0.171964
Sonata	0.168776
Leaf	0.154005
Volt	0.146184

Fusion is our first choice, followed by Camry, Prius, Sonata, Leaf and Volt.

If we use method COA (2) in step 3 then within the final matrix, we replace the actual costs with these pairwise results ($CR = 0.031$):

Prius	0.107059
Fusion	0.073259
Volt	0.046756
Camry	0.465277
Sonata	0.256847
Leaf	0.050802

Then we obtain the ranked results as:

Camry	0.270679
Sonata	0.203471
Fusion	0.146001
Prius	0.141518
Leaf	0.12178
Volt	0.116551

If we do COA (3) in step 3 then this method requires us to redo the pairwise criterion matrix without the cost criteria. These weights are:

City MPG	0.363386
HW MPG	0.241683
Safety	0.159679
Reliability	0.097000
Performance	0.081418
Interior/Style	0.056834

Prius	0.168552
Fusion	0.184414
Volt	0.152874
Camry	0.169553
Sonata	0.161423
Leaf	0.163184

We normalize the original costs from Table 2, and divide these ranked values by the normalized *cost* to obtain a *cost/benefit* value. These are shown in ranked order:

Camry	1.211261
Fusion	1.178748
Prius	1.10449
Sonata	1.06931
Leaf	0.821187
Volt	0.759482

Sensitivity Analysis for Cars

We alter our decision pairwise values to obtain a new set of decision weights to use in COA (1) from step 3 to obtain new results: Camry, Fusion, Sonata, Prius, Leaf, and Volt. The new weights and model's results are:

Cost	0.311155922
MPG City	0.133614062
MPG HW	0.095786226
Performance	0.055068606
Interior and style	0.049997069
Safety	0.129371535
Reliability	0.225006578

Alternatives	Values	Ranks
Prius	0.10882648	4
Fusion	0.11927995	2

(continued)

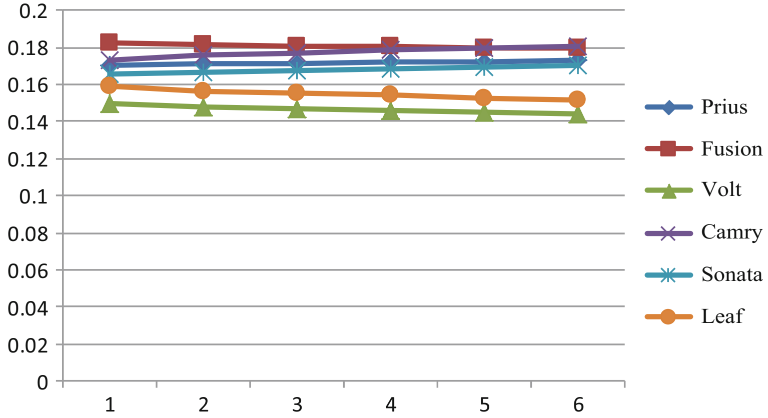


Fig. 2 Camry overtakes Fusion as the top alternative as we change the weight of Cost

Alternatives	Values	Ranks
Volt	0.04816882	5
Camry	0.18399172	1
Sonata	0.11816156	3
Leaf	0.04357927	6

The resulting values have changed but not the relative rankings of the cars. Again, we recommend using sensitivity analysis to find a "break point", if one exists.

We systemically varied the cost weights using Eq. (6) with increments of (±) 0.05. We potted the results to show the approximate break point of the criteria cost as weight of cost + 0.1 as seen in Fig. 2.

Example 2. Kite Network revisited Assume all we have are the outputs from ORA which we do not show here due to the volume of output produced. We take the metrics from ORA and normalize each column. The columns for each criterion are placed in a matrix X with entries, x_{ij}. We define wj as the weights for each criterion.

Next, we assume we can obtain pairwise comparison matrix from the decision maker concerning the criterion. We use the output from ORA and normalize the results for AHP to rate the alternatives within each criterion. We provide a sample pairwise comparison matrix for weighting the criterion from the Kite example using Saaty's 9-point scale. The *CR* is 0.0828, which is less than 0.1, so our pairwise matrix is consistent and we continue.

Pairwise Comparison Matrix

	Central	Eigenvector	In-degree	Out-degree	Information centrality	Betweenness
Central	*1*	*3*	*2*	*2*	½	*1/3*
Eigenvector	*1/3*	*1*	*1/3*	*1*	*2*	½
In-degree	½	*3*	*1*	½	½	¼
Out-degree	½	½	*1*	*1*	¼	¼
Information Centrality	*2*	*2*	*4*	*4*	*1*	*1/3*
Betweenness	*3*	*2*	*4*	*4*	*3*	*1*

We obtain the steady state values that will be our criterion weights, where the sum of the weights equals 1.0. There exist many methods to obtain these weights. The methods used here are the power method from numerical analysis [15] and discrete dynamical systems [23, 24].

0.1532	0.1532	0.1532	0.1532	0.1532	0.1532
0.1450	0.1450	0.1450	0.1450	0.1450	0.1450
0.1194	0.1195	0.1194	0.1194	0.1194	0.1194
0.0672	0.0672	0.0672	0.0672	0.0672	0.0672
0.1577	0.1577	0.1577	0.1577	0.1577	0.1577
0.3575	0.3575	0.3575	0.3575	0.3575	0.3575

These values provide the weights for each criterion: *centrality* = *0.1532*, *eigenvectors* = *0.1450*, *in-centrality* = *0.1194*, *out-centrality* = *0.0672*, *information centrality* = *0.1577*, and *betweenness* = *0.3575*.

We multiply the matrix of the weights and the normalized matrix of metrics from ORA to obtain our output and ranking:

Node	AHP Value	Rank
Susan	0.160762473	2
Steven	0.133201647	3
Sarah	0.113388361	4
Tom	0.075107843	6
Claire	0.075107843	6
Fred	0.060386019	8
David	0.060386019	8
Claudia	0.177251415	1
Ben	0.109606727	5
Jennifer	0.034801653	10

For this example with AHP Claudia, *cl*, is the key node. However, the bias of the decision maker is important in the analysis of the criterion weights. The criterion, "*Betweenness*", is two to three times more important than the other criterion.

Sensitivity Analysis for the Kite network Changes in the pairwise decision criterion cause fluctuations in the key nodes. We change our pairwise comparison so that "*Betweenness*" is not so dominant a criterion. With these slight pairwise changes, we now find Susan is now ranked first, followed by Steven and then Claudia. The AHP process is sensitive to changes in the criterion weights. We vary *betweenness* in increments of 0.05 to find the break point.

	Centrality	IN	OUT	Eigen	EIGENC	Close	IN-Close	Betw	INFO Cen.
t	0.111111	0.111111	0.111111	0.114399	0.114507	0.100734	0.099804	0.019408	0.110889
c	0.111111	0.111111	0.111111	0.114399	0.114507	0.100734	0.099804	0.019408	0.108891
f	0.083333	0.083333	0.083333	0.093758	0.094004	0.097348	0.09645	0	0.097902
s	0.125	0.138889	0.111111	0.137528	0.137331	0.100734	0.111826	0.104188	0.112887
su	0.180556	0.166667	0.194444	0.175081	0.174855	0.122743	0.107632	0.202247	0.132867
st	0.138889	0.138889	0.138889	0.137528	0.137331	0.112867	0.111826	0.15526	0.123876
d	0.083333	0.083333	0.083333	0.093758	0.094004	0.097348	0.107632	0	0.100899
cl	0.083333	0.083333	0.083333	0.104203	0.104062	0.108634	0.107632	0.317671	0.110889
b	0.055556	0.055556	0.055556	0.024123	0.023985	0.088318	0.087503	0.181818	0.061938
j	0.027778	0.027778	0.027778	0.005223	0.005416	0.070542	0.069891	0	0.038961
10 alternatives and 9 attributes or criterion									
Criterion weights									

w1	0.034486
w2	0.037178
w3	0.045778
w4	0.398079
w5	0.055033
w6	0.086323
w7	0.135133
w8	0.207991

With these slight pairwise changes, we now find Susan is now ranked first, followed by Steven and then Claudia. The AHP process is sensitive to changes in the criterion weights. We vary *betweenness* in increments of 0.05 to find the break point.

Tom	0.098628	Susan	0.161609
Claire	0.098212	Steven	0.133528
Fred	0.081731	Claudia	0.133428
Sarah	0.12264	Sarah	0.12264
Susan	0.161609	Tom	0.098628
Steven	0.133528	Claire	0.098212
David	0.083319	David	0.083319
Claudia	0.133428	Fred	0.081731
Ben	0.0645	Ben	0.0645
Jennifer	0.022405	Jennifer	0.022405

Further, sensitivity analysis of the nodes is provided in Fig. 3.

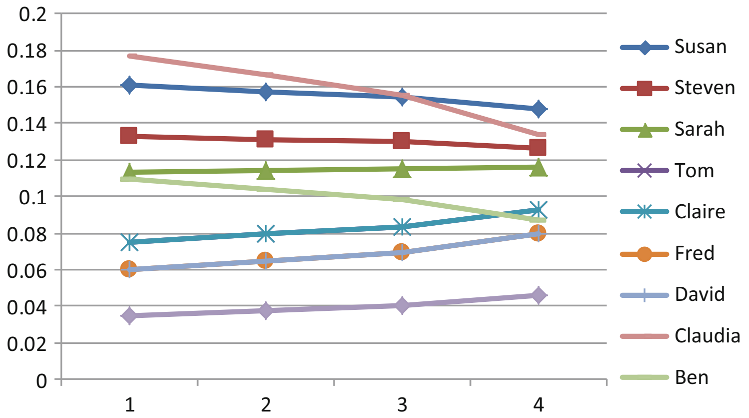


Fig. 3 Sensitivity Analysis for Nodes varying only *Betweenness*

We varied the weight of the criterion *Betweenness* by lowering it by 0.05 each iteration and increasing the other weights using Eq. (6). We see the Claudia and Susan change as the top node when we reduce *Betweenness* by 0.1.

5 Technique of Order Preference by Similarity to the Ideal Solution (TOPSIS)

5.1 Description and Uses

The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a multi-criteria decision analysis method, which was originally developed in a dissertation from Kansas State University [25]. It has been further developed by others [26, 27]. TOPSIS is based on the concept that the chosen alternative should have the shortest geometric distance from the positive ideal solution and the longest geometric distance from the negative ideal solution. It is a method of compensatory aggregation that compares a set of alternatives by identifying weights for each criterion, normalizing the scores for each criterion and calculating the geometric distance between each alternative and the ideal alternative, which is the best score in each criterion. An assumption of TOPSIS is that the criteria are monotonically increasing or decreasing. Normalization is usually required as the parameters or criteria are often of incompatible dimensions in multi-criteria problems. Compensatory methods such as TOPSIS allow trade-offs between criteria, where a poor result in one criterion can be negated by a good result in another criterion. This provides a more realistic form of modeling than non-compensatory methods, which include or exclude alternative solutions based on hard cut-offs.

We only desire to briefly discuss the elements in the framework of TOPSIS. TOPSIS can be described as a method to decompose a problem into sub-problems. In most decisions, the decision maker has a choice among many alternatives. Each alternative has a set of attributes or characteristics that can be measured, either subjectively or objectively. The attribute elements of the hierarchal process can relate to any aspect of the decision problem—tangible or intangible, carefully measured or roughly estimated, well- or poorly-understood—anything at all that applies to the decision at hand.

5.2 Methodology

The TOPSIS process is carried out as follows:

Step 1. Create an evaluation matrix consisting of m alternatives and n criteria, with the intersection of each alternative and criterion given as x_{ij}, giving us a matrix $(\boldsymbol{X}_{ij})_{m\ x\ n}$.

$$D = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ \cdot \\ \cdot \\ \cdot \\ A_m \end{array} \begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_3 & \cdot\ \cdot\ \cdot & x_n \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdot\ \cdot\ \cdot & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdot\ \cdot\ \cdot & x_{2n} \\ x_{31} & x_{32} & x_{33} & \cdot\ \cdot\ \cdot & x_{3n} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ x_{m1} & x_{m2} & x_{m3} & \cdot\ \cdot\ \cdot & x_{mn} \end{bmatrix} \end{array}$$

Step 2. The matrix shown as $\boldsymbol{D}$ above then is normalized to form the matrix $\boldsymbol{R} = (\boldsymbol{R}_{ij})_{m\ x\ n}$ as shown using the normalization method

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum x_{ij}^2}}$$

for $i = 1,2\ldots,m$; $j = 1,2,\ldots n$

Step 3. Calculate the weighted normalized decision matrix. First we need the weights. Weights can come from either the decision maker or by computation.

Step 3a. Use either the decision maker's weights for the attributes $x_1, x_2, \ldots x_n$ or compute the weights through the use of Saaty's [14] AHP decision maker weights method to obtain the weights as the eigenvector to the attributes versus attribute pairwise comparison matrix.

$$\sum_{j=1}^{n} w_j = 1$$

The sum of the weights over all attributes must equal 1 regardless of the method used.

Step 3b. Multiply the weights to each of the column entries in the matrix from *Step 2* to obtain the matrix, ***T***.

$$T = \left(t_{ij}\right)_{mxn} = \left(w_j r_{ij}\right)_{mxn,} i = 1, 2, \ldots, m$$

Step 4. Determine the worst alternative (A_w) and the best alternative (A_b): Examine each attribute's column and select the largest and smallest values appropriately. If the values imply larger is better (profit), then the best alternatives are the largest values, and if the values imply smaller is better (such as cost), then the best alternative is the smallest value.

$$\begin{aligned} A_w &= \left\{\left\langle \max\left(t_{ij} | i = 1, 2, \ldots, m | j \in J_-\right\rangle, \left\langle \min\left(t_{ij}\middle| i = 1, 2, \ldots, m\right) \middle| j \in J_+\right\rangle\right\} \\ &\equiv \left\{t_{wj} | j = 1, 2, \ldots, n\right\}, \\ A_{wb} &= \left\{\left\langle \min\left(t_{ij} | i = 1, 2, \ldots, m | j \in J_-\right\rangle, \left\langle \max\left(t_{ij}\middle| i = 1, 2, \ldots, m\right) \middle| j \in J_+\right\rangle\right\} \\ &\equiv \left\{t_{bj} | j = 1, 2, \ldots, n\right\}, \end{aligned}$$

where,

$J_+ = \{j = 1, 2, \ldots n | j\}$ associated with the criteria having a positive impact, and
$J_- = \{j = 1, 2, \ldots n | j\}$ associated with the criteria having a negative impact.

We suggest that if possible make all entry values in terms of positive impacts.

Step 5. Calculate the L2-distance between the target alternative i and the worst condition A_w

$$d_{iw} = \sqrt{\sum\nolimits_{j=1}^{n} \left(t_{ij} - t_{wj}\right)^2}, i = 1, 2, \ldots, m$$

and then calculate the distance between the alternative i and the best condition A_b

$$d_{ib} = \sqrt{\sum\nolimits_{j=1}^{n} \left(t_{ij} - t_{bj}\right)^2}, i = 1, 2, \ldots, m$$

where d_{iw} and d_{ib} are L2-norm distances from the target alternative i to the worst and best conditions, respectively.

Step 6. Calculate the similarity to the worst condition:

$$s_{iw} = \frac{d_{iw}}{(d_{iw} + d_{ib})}, 0 \leq s_{iw} \leq 1, i = 1, 2, \ldots, m$$

$S_{iw} = 1$ if and only if the alternative solution has the worst condition; and
$S_{iw} = 0$ if and only if the alternative solution has the best condition.

Step 7. Rank the alternatives according to their value from S_{iw} ($i = 1, 2, \ldots, m$).

5.2.1 Normalization

Two methods of normalization that have been used to deal with incongruous criteria dimensions are linear normalization and vector normalization. Normalization can be calculated as in *Step 2* of the TOPSIS process above. Vector normalization was incorporated with the original development of the TOPSIS method [26], and is calculated using the following formula:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum x_{ij}^2}} \quad \text{for } i = 1, 2 \ldots, m; j = 1, 2, \ldots n$$

In using vector normalization, the non-linear distances between single dimension scores and ratios should produce smoother trade-offs [25].

Let's suggest two options for the weights in Step 3. First, the decision maker might actually have a weighting scheme that they want the analyst to use. If not, we suggest using Saaty's *9*-point pairwise method developed for the Analytical Hierarchy Process (AHP) [14]. We refer the reader to our discussion in the AHP section for the decision weights using the Saaty's *9*-point scale and pairwise comparisons. In TOPSIS, we have the following scheme.

Objective Statement ← This is the decision desired
Alternatives: $1, 2, 3, \ldots, n$
For each of the alternatives there are criteria (attributes) to compare:
Criteria (or Attributes): $c_1, c_2, \ldots, c_m$

Once the hierarchy is built, the decision maker(s) systematically evaluate its various elements pairwise (by comparing them to one another two at a time), with respect to their impact on an element above them in the hierarchy. In making the comparisons, the decision makers can use concrete data about the elements, but they typically use their judgments about the elements' relative meaning and importance. It is the essence of the TOPSIS that human judgments, and not just the underlying information, can be used in performing the evaluations. TOPSIS converts these evaluations to numerical values that can be processed and compared over the entire range of the problem. A numerical weight or priority is derived for each element of the hierarchy, allowing diverse and often incommensurable elements to be compared to one another in a rational and consistent way. This capability distinguishes the TOPSIS from other decision making techniques.

In the final step of the process, numerical priorities or ranking are calculated for each of the decision alternatives. These numbers represent the alternatives' relative ability to achieve the decision goal, so that they allow a straightforward consideration of the various courses of action.

While it can be used by individuals working on straightforward decisions, TOPSIS is most useful where teams of people are working on complex problems, especially those with high stakes, involving human perceptions and judgments, whose resolutions have long-term repercussions. It has unique advantages when important elements of the decision are difficult to quantify or compare, or where communication among team members is impeded by their different specializations, terminologies, or perspectives.

Decision situations to which the TOPSIS might be applied are identical to what we presented earlier for AHP:

- Choice—The selection of one alternative from a given set of alternatives, usually where there are multiple decision criteria involved.
- Ranking—Putting a set of alternatives in order from most to least desirable
- Prioritization—Determining the relative merit of members of a set of alternatives, as opposed to selecting a single one or merely ranking them
- Resource allocation—Apportioning resources among a set of alternatives
- Benchmarking—Comparing the processes in one's own organization with those of other best-of-breed organizations
- Quality management—Dealing with the multidimensional aspects of quality and quality improvement
- Conflict resolution—Settling disputes between parties with apparently incompatible goals or positions

5.3 Strengths and Limitations

TOPSIS is based on the concept that the chosen alternative should have the shortest geometric distance from the positive ideal solution and the longest geometric distance from the negative ideal solution. It is a method of compensatory aggregation that compares a set of alternatives by identifying weights for each criterion, normalizing scores for each criterion and calculating the geometric distance between each alternative and the ideal alternative, which is the best score in each criterion. An assumption of TOPSIS is that the criteria are monotonically increasing or decreasing. Normalization is usually required as the parameters or criteria are often of incongruous dimensions in multi-criteria problems. Compensatory methods such as TOPSIS allow trade-offs between criteria, where a poor result in one criterion can be negated by a good result in another criterion. This provides a more realistic form of modeling than non-compensatory methods, which include or exclude alternative solutions based on hard cut-offs. TOPSIS corrects the rank reversal that was a limitation in strictly using the AHP method. TOPSIS also allows the user to state which of the criteria are maximized and which are minimized for better results. In the late 1980s TOPSIS was a department of defense standard for performing selection of systems across all branches in tight budget years.

5.4 Sensitivity Analysis

The decision weights are subject to sensitivity analysis to determine how they affect the final ranking. The same procedures discussed in Sect. 4.4 are valid here. Sensitivity analysis is essential to good analysis. Additionally, Alinehad [22] suggests sensitivity analysis for TOPSIS for changing an attribute weight. We will again use Eq. (1) in our sensitivity analysis.

5.5 Illustrative Examples

Example 1. Car Selection We might assume that our decision maker weights from the AHP section are still valid for our use and we revisit Table 2.

Weights from before:

Cost	0.38960838
MPG City	0.11759671
MPGHW	0.04836533
Performance	0.0698967
Interior	0.05785692
Safety	0.10540328
Reliability	0.21127268

	Cost	MPG_city	MPG_HW	Perf.	Interior	Safety	Reliability	N/A
Cost	1	4	6	5	6	4	2	0
MPG_city	0.25	1	6	3	5	1	0.333333333	0
MPG_HW	0.166667	0.166667	1	0.5	0.5	0.333333	0.25	0
Perf.	0.2	0.333333	2	1	2	0.5	0.333333333	0
Interior	0.166667	0.2	2	0.5	1	0.5	0.333333333	0
Safety	0.25	1	3	2	2	1	0.5	0
Reliability	0.5	3	4	3	3	2	1	0
N/A	0	0	0	0	0	0	0	1

We use the identical data from the car example from AHP but we apply steps 3–7 from TOPSIS to our data. We are able to keep the cost data and just inform TOPSIS that a smaller cost is better. We obtained the following rank ordering of the cars: Camry, Fusion, Prius, Sonata, Volt, and Leaf.

4	0.82154128	Camry
2	0.74622988	Fusion
1	0.72890117	Prius
5	0.70182382	Sonata
6	0.15580913	Leaf
3	0.11771999	Volt

It is critical to perform sensitivity analysis on the weights to see how they affect the final ranking. This time we work toward finding the break point where the order of cars actually changes. Since cost is the largest criterion weight, we vary it suing Eq. (1) in increments of 0.05. We see from Fig. 4, the Fusion overtakes Camry when cost is decreased by about 0.1, which allows reliability to overtake cost as the dominate-weighted decision criterion.

Example 2. Social Networks We revisit the Kite Network with TOPSIS to find influences in the network. We present the output from ORA that we used in Table 5.

We use the decision weights from AHP (unless a decision maker gives us their own weights) and find the eigenvectors for our eight metrics as:

w1	0.034486
w2	0.037178
w3	0.045778
w4	0.398079
w5	0.055033
w6	0.086323
w7	0.135133
w8	0.207991

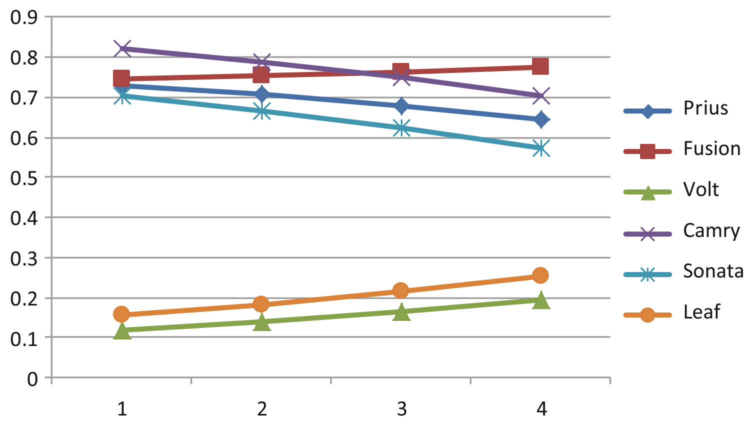


Fig. 4 TOPSIS Values of the cars by varying the weight for cost incrementally by −0.05 each of four increments along the x-axis

Table 5 ORA output for Kite network

	IN	OUT	Eigen	EigenL	Close	IN-Close	Betweenness	INF Centr
Tom	0.4	0.4	0.46	0.296	0.357	0.357	0.019	0.111
Claire	0.4	0.4	0.46	0.296	0.357	0.357	0.019	0.109
Fred	0.3	0.3	0.377	0.243	0.345	0.345	0	0.098
Sarah	0.5	0.4	0.553	0.355	0.357	0.4	0.102	0.113
Susan	0.6	0.7	0.704	0.452	0.435	0.385	0.198	0.133
Steven	0.5	0.5	0.553	0.355	0.4	0.4	0.152	0.124
David	0.3	0.3	0.377	0.243	0.345	0.385	0	0.101
Claudia	0.3	0.3	0.419	0.269	0.385	0.385	0.311	0.111
Ben	0.2	0.2	0.097	0.062	0.313	0.313	0.178	0.062
Jennifer	0.1	0.1	0.021	0.014	0.25	0.25	0	0.039

We take the metrics from ORA and perform steps 2–7 of TOPSIS to obtain the results:

S+	S−	C	
0.0273861	0.181270536	0.86875041	SUSAN
0.0497878	0.148965362	0.749499497	STEVEN
0.0565358	0.14154449	0.714581437	SARAH
0.0801011	0.134445151	0.626648721	TOM
0.0803318	0.133785196	0.624822765	CLAIRE
0.10599	0.138108941	0.565790826	CLAUDIA
0.1112243	0.12987004	0.538668909	DAVID
0.1115873	0.128942016	0.536076177	ERED
0.1714404	0.113580988	0.398499927	BEN
0.2042871	0.130399883	0.389617444	JENNIFER

We rank order the final output from TOPSIS as shown in the last column above. We interpret the results as follows: The key node is *Susan* followed by *Steven*, *Sarah*, *Tom*, and *Claire*.

Sensitivity Analysis

We used Eq. (6) and systemically altered the value of the largest criteria weight, *EigenL*, shown in Fig. 5.

We note that Susan remain the most influential node.

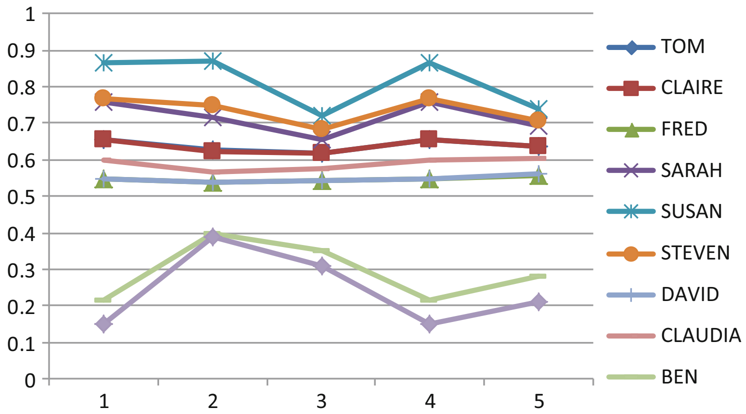


Fig. 5 Sensitivity Analysis plot as a function of varying *EigenL* weight in increments of −0.05 units

6 Comparison of Results for the Kite Network

We have also used the two other MADM methods to rank order our nodes in previous work in SNA [10]. When we applied data envelopment analysis and AHP to compare to TOPSIS, we obtained the results displayed in Table 6 for the Kite Network.

It might be useful to use this table as input for another round of one of these presented methods and then use sensitivity analysis.

Table 6 MADM applied to Kite network

Node	SAW	TOPSIS value (rank)	DEA efficiency value (rank)	AHP value (rank)
Susan	0.046 (1)	0.862 (1)	1 (1)	0.159 (2)
Sarah	0.021 (4)	0.675 (3)	0.786 (2)	0.113 (4)
Steven	0.026 (3)	0.721 (2)	0.786 (2)	0.133 (3)
Claire	0.0115 (7)	0.649 (4)	0.653 (4)	0.076 (6)
Fred	0.0115 (7)	0.446 (8)	0.653 (4)	0.061 (8)
David	0.031 (2)	0.449 (7)	0.536 (8)	0.061 (8)
Claudia	0.012 (8)	0.540 (6)	0.595 (6)	0.176 (1)
Ben	0.018 (5)	0.246 (9)	0.138 (9)	0.109 (5)
Jennifer	0.005 (10)	0 (10)	0.030 (10)	0.036 (10)
Tom	0.0143 (6)	0.542 (5)	0.553 (7)	0.076 (6)

7 Technologies Available

We provide a brief summary of computer technologies that are available although it is not difficult to set up an Excel spreadsheet to do all these methods that we presented.

We refer the reader to the internet to find out the current list of MADM software and the details. These are generally user-friendly packages that make estimating efficiency and capacity utilization relatively straightforward. These include DEA-Solver by Cooper, Seiford and Tone [28], which is an add-on to Microsoft Excel, User guides and examples are also provided when downloading the software. Some additional free and commercial software is listed for reference.

TOPSIS

Name	Key features	Compatibility
TOPSIS Solver 2012	20 criteria; 20 alternatives; automatically normalizes; user indicated positive and negative criteria	Windows; MS-EXCEL
Triptych SDI Tools	200 Criteria; 200 alternatives; each criterion can be weighted	Windows; MS-EXCEL

AHP or ANP

Name	Key features	Compatibility
AHP.NET	Performs Saaty's AHP method	Windows;
AHP Template in Excel by BPMSG	Performs Saaty's AHP method	Windows; MS-EXCEL
PriEst	AHP	LINUX and Windows
Super Decisions	AHP and ANP	Windows
Expert Choice	AHP	Windows

DEA and LP Software

Name	Features	Compatibility
DEAFrontier™	The software is developed based upon Professor Zhu's years of DEA research and teaching experience. The software is written by Professor Zhu in an effort to minimize the possibility of misrepresentation of DEA models during coding	MS_EXCEL Add-In; Windows XP, Windows 8
GAMS	Linear programming package	Various platforms (consult GAMS)
LINDO & LINGO	Linear programming package	Windows
MS-EXCEL Solver	Linear Programming option	Windows

Acknowledgements We thank all the reviewers of this and other material used to produce this chapter. These views are the views of the author and not of the Department of Defense, the Department of the Navy, or the Naval Postgraduate School.

References

1. Cooper W, Seiford L, Tone K (2000) Data envelopment analysis. Kluwer Academic Press, London, UK
2. Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decisión making units. Eur J Oper Res 2:429–444
3. Winston W (1995) Introduction to mathematical programming. Duxbury Press, Belmont, CA, pp 322–325
4. Callen J (1991) Data envelopment analysis: practical survey and managerial accounting applications. J Manag Account Res 3:35–57
5. Trick MA (2014) Data envelopment analysis, Chapter 12. http://mat.gsia.cmu.edu/classes/QUANT/NOTES/chap12.pdf. Accessed Apr 2014
6. Trick MA (1996) Multiple criteria decision making for consultants. http://mat.gsia.cmu.du/classes/mstc/multiple/multiple.html. Accessed Apr 2014
7. Neralic L (1998) Sensitivity analysis in models of data envelopment analysis. Math Commun 3:41–45
8. Krackhardt D (1990) Assessing the political landscape: structure, cognition, and power in organizations. Sci Q 35:342–369
9. Carly K (2011) Organizational risk analyzer (ORA). Center for Computational analysis of social and organizational systems (CASOS), Carnegie Mellon University, Pittsburgh, PA
10. Fox W, Everton S (2013) Mathematical modeling in social network analysis: using TOPSIS to find node influences in a social network. J Math Syst Sci 3(10):531–541
11. Fox W, Everton S (2014) Mathematical modeling in social network analysis: using data envelopment analysis and analytical hierarchy process to find node influences in a social network. J Def Model Simul 2(2014):1–9
12. Fishburn PC (1967) Additive utilities with incomplete product set: applications to priorities and assignments. Oper Res Soc Am 15:537–542
13. Consumer's Reports Car Guide (2012). The Editors of Consumer Reports
14. Saaty T (1980) The analytical hierarchy process. McGraw Hill, United States
15. Burden R, Faires D (2013) Numerical analysis, 9th edn. Cengage, Boston, MA
16. Hartwich F (1999) Weighting of agricultural research results: strength and limitations of the analytic hierarchy process (AHP)t, Universitat Hohenheim. Retrieved https://entwicklungspolitik.unihohenheim.de/uploads/media/DP_09_1999_Hartwich_02.pdf
17. Leonelli R (2012) Enhancing a decision support tool with sensitivity analysis. Thesis, University of Manchester
18. Chen H, Kocaoglu D (2008) A sensitivity analysis algorithm for hierarchical decision models. Eur J Oper Res 185(1):266–288
19. Baker T, Zabinsky Z (2011) A multicriteria decision making model for reverse logistics using Analytical Hierarchy Process. Omega 39:558–573
20. Hurly WJ (2001) The Analytical Hierarchy Process: a note on an approach to sensitivity which preserves rank order. Comput Oper Res 28:185–188
21. Butler J, Jia J, Dyer J (1997) Simulation techniques for the sensitivity analysis of multi-criteria decision models. Eur J Oper Res 103:531–546
22. Alinezhad A, Amini A (2011) Sensitivity analysis of TOPSIS technique: the results of change in the weight of one attribute on the final ranking of alternatives. J Optim Ind Eng 7:23–28
23. Fox WP (2012) Mathematical modeling of the analytical hierarchy process using discrete dynamical systems in decision analysis. Comput Educ J 22:27–34

24. Giordano FR, Fox W, Horton S (2013) A first course in mathematical modeling, 5th edn. Brooks-Cole, Boston
25. Hwang CL, Yoon K (1981) Multiple attribute decision making: methods and applications. Springer, New York
26. Yoon K (1987) A reconciliation among discrete compromise situations. J Oper Res Soc 38: 277–286
27. Hwang CL, Lai Y, Liu TY (1993) A new approach for multiple objective decision making. Comput Oper Res 20:889–899
28. Cooper W, Li S, Seiford L, Thrall RM, Zhu J (2001) Sensitivity and stability analysis in DEA: some recent developments. J Prod Anal 15(3):217–246

How Business Analytics Should Work

Marco Antonio Villegas-García, Fausto Pedro García Márquez, and Diego José Pedregal Tercero

1 A World in Transformation

We are now approaching the so called Information Era. The globalization came to world and the countries connected together, the people's life changed as well as their cultures. As Information and Communication Technologies (ICT) spread all around the world, the digital universe grows astonishingly. It is estimated that the amount of data produced in the world grows at an astounding annual rate of 60 % [1].

The volume of transactions and interchanged data is reaching astronomical scale [2]: IBM estimates that humanity creates 2.4 quintillion bytes (a billion billion) of data everyday. Much of this data is created by digital systems usually linked to internet. International Data Corporation estimates that the digital universe will double in size through 2020 and reach 40 ZB (zetabytes), which means 5,247 GB for every person on earth in 2020. The *digital behavioural universe* is being created from the clickstream and the digital footprints of every person across Earth interacting, participating and consuming this data. At this respect, one of the bigger trends that most drive the behavioural dimension of the digital universe is the mobile computing. Actually, six billion of the world's seven billion people have access to mobile phone, what means that, by far, it is the largest service infrastructure across the world.

Some authors refer to this revolution as a hinge of history [3], highlighting the fact that the humanity is entering in a new scope, a different world, where the game is driven by different rules:

> The combination of massive computing power, massive expansion in data management tools and practices, and exponentialized increases in customer expectations have created a

M.A. Villegas-García (✉) • F.P. García Márquez • D.J. Pedregal Tercero
Ingenium Research Group, University of Castilla-La Mancha, Ciudad Real, Spain
e-mail: marco.villegas@uclm.es; faustopedro.garcia@uclm.es; diego.pedregal@uclm.es

© Springer International Publishing Switzerland 2015
F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_5

> world so complex and a customer so demanding that unaugmented human cognition—by this I mean making decisions without the assistance of a robust business analytic tool set—is no longer good enough. What you need to know, whom you need to know, how you come to know and the very abbreviated time window available for making efficacious use of knowledge are transforming.

However, the promise of digital analytics still remains largely unrealized. EMC estimates that the majority of new data is largely untagged, file-based, and unstructured data, little is known about it. Only 3 % of the data being created today is useful for analyses, whereas only 0.05 % of that data is actually being analyzed. Therefore, 99.95 % of useful data available today for analysis is not being analyzed. IDC estimates a 67 % increase in data available for analysis by 2020.

1.1 Some Lasting Stumbling Blocks

In spite of the enormous changes experienced by the society in the last years, a very striking truth is the fact that the way we manage information today is not much different from how we have done for millennia [3]. One of the most ancient form of recorded information are the prehistoric cave drawings. Someone decided to create this record, probably ignoring that he had discover a data storage technique, far more perdurable in time than oral transmission. We realize that for us human, the problem is not storing data, it is in access and meaning. Following the example of the cave drawing, accessing the information means getting there and to stand up in front of it. But once you get there, what do you see? Is the picture a message? Is it perhaps a lyric? Or is it just art? It is evident that the sole data is not enough for us. We need a semantic context in which such bunch of data takes up meaning and becomes valuable information. This enriched information is the required base for building wisdom, which will let us to make better decisions.

The information media evolved with the centuries. The paper became the major support for data recording and communication. Vinyl discs made the miracle of storing audio, whereas electromagnetic tapes did so with video. Now, in the digital era, we are able to carry in the pocket an immense quantity of multimedia content, all with a vulgar USB stick. And the industry continues in this endless run. Definitely we are pretty good at storing data. However, the volume of accessible data far exceeds the human ability to consume it, even more in our days. Yesterday our offices were overflowing with papers, today with emails and digital documents. It is true that Information Management strategies have positively evolved, but they haven't gone very far. In spite of these enormous changes, we have still some of the same problems since the beginning. We could synthesize these with the following questions:

- How to build knowledge up from data?
- How to relate data? How to make it meaningful?

Two separated poles have remain unconnected. By one hand, the fast evolving hardware technology, and by the other the human-minded activity supported by that technology. There are enormous oceans of information to be exploited, but the actual benefits we obtain from them depends on how well connected these two poles are. And here is where Business Analytics come in action. This term refers to a pretty interesting assembly of several fields, like math and statistics sciences, computer science and management sciences, in which the methods, processes and methodologies are continuously enriched by all these knowledge areas. Although this has been a pretty good effort in bringing humans and data-oceans closer, there is still a long way to go.

2 Data Analysis and Synthesis

Another important characteristic of today's society is the *fragmentation*. Our daily experience of life is minced into little pieces, little facets or domains which we have to continually integrate and reconcile: professional and familiar life; business trip plan and restrictions in flights, airlines, budgets. They are labyrinth of decomposition [4]: Organizations are decomposed into regions, divisions, departments, products, and services, not to mention missions, visions, objectives, programs, budgets, and systems; likewise, agendas are decomposed into issues, and strategic issues are decomposed into strengths, weaknesses, threats, and opportunities.

Analysis consists primarily in breaking down a complex topic of problem into smaller parts to gain a better understanding of it. We believe that if we have all the data, and we are smart enough, we can solve any problem. In recruiting people, we test their analytical skills. Indeed, regardless their industry sector, many people work as 'analysts'. It has become like an obsession [5]:

> Why is everyone so obsessed with analysis? Analysis is only one style of solving problems. [...] We seem to have forgotten all about synthesis, the opposite approach. Take two or more ideas and combine them into a larger new idea. Tackling a problem in this way might lead to entirely new insights, where problems of the "old world" (before the synthesis) do not even occur anymore. Where analysis focuses on working within the boundaries of a certain domain [...], synthesis connects various domains.

Managers oversee all this chaos, and they are supposed to integrate the whole confusing mess, most of the times with the sole help of their intelligence and intuition: they make the synthesis on their own. Synthesis is the very essence of managing [4]:

> Putting things together, in the form of coherent strategies, unified organizations, and integrated systems. [...] It's not that managers don't need analysis; it's that they need it as an input to synthesis. [...] So how can a manager see the big picture amid so many little details?

As companies get large and complex, it is more evident that it is impossible for a single person to watch them conscientiously on detail, even a single department. In some companies, the business operations are carried out by pretty big teams; they work together to accomplish the business goals, keeping certain norms and protocols. Business Analytics give us the possibility of exploiting large volumes of data, extract value from them, and use this value to empower the processes and business decisions at every level in the company. Even though any incorporation of BA is positive to the company's performance, the major benefits come from a wider application of BA throughout the organization [6]. The solutions implemented at this level are very specific to the concrete industry and organization and involve an important amount of time-effort from a considerable number of IT and business experts. A consequence of this is that only big companies can afford these implementation costs, whereas the main population of companies remain far from these possibilities. Even for the BA leaders, at present there is no clear methodology for a comprehensive BA implementation [7]. The big challenge consist on finding a clear path to broadly implement BA without getting stuck (distracted) with "nasty" technical details, but center at the business logic and concepts.

We claim that business intelligence and analytics should evolve to a higher status in which the business users can 'navigate' more fluidly across the processes, data, resources, restrictions, goals... This implies that users should be able to interact directly with the BA systems via *business concepts* like revenue, costs, customer satisfaction, etc. The technical details on how data is collected and analysed, and how they are arranged to built more abstract artefacts, most be hidden for business users. The problem is that even IT and business analysts get frequently trapped into the high complexity of the BA details implementation.

We propose a BA architecture that facilitates both the implementation and exploitation of BA systems. Our idea is strongly based on business modelling techniques, in particular on the BIM proposed by Barone et al. in 2010 [8]. The following section explain the main components of this approach.

3 The Business Analytics Architecture (BAA)

The main objective of Business Intelligence tools is to transform raw data into meaningful and actionable information for business purposes. We understand by 'meaningful and actionable' information that which allow to the readers building a better knowledge about the reality in a given context, and give rise to making wise decisions driving the reality to the desired state.

At this point we can identify two main faces of all BI systems: the 'internal face' deals with the raw data that is going to be transformed into 'meaningful and actionable' information. That data is usually extracted from operational information systems which record everything happening in the company. There are many commercial solutions filling this section, most of them under the names ERP, MRP and CRM.[1] By the other hand, the 'external face' of BA systems is associated with the business purposes we are pursuing. These comprise the designed business strategy, usually stated at the level of enterprise/corporation; and the operational objectives, which are more related with a particular department.

In spite of seem disconnected, these two poles are tightly related. In fact, the enterprise's data is just a representation of the actual business' execution. They absolutely shouldn't be disconnected, as they are nowadays in many business information systems. By far, most of the efforts driven by BIA community could be condensed in this sublime goal: achieve a more fluid an natural connection between the data and the real-world. And this should be done in both directions: (1) from data to real-world, in order to make a fact-based tracking of the company's performance, and be able to drive it to success; and (2) from real-world to data, enabling meaningful data analysis powered by business semantic.

To achieve this goal, the BAA is compose of three main layers (see Fig. 1). The higher one contains the business logic and concepts, using the same terminology that business analysts use everyday. It is usually called 'the semantic layer', because it provides the logic framework that gives the appropriate meaning to all other elements. This is the 'external face' of the BI system, which is supposed to provide an user-friendly interface to embed the business logic into the system. In the next section we will drop a light about how the business users should interact with this layer and the easy-to-use functionalities it should provide.

This semantic layer needs to be connected with low-level data, and this is done through the mapping layer. It contains the conceptual mapping between the data entities and the business entities defined in the semantic layer. Most of the complexity of connecting abstract concepts with concrete facts are embedded in this layer; we follow a simplify version of the model proposed by Rizzolo et al. [9]. This connection is meant to be fully bidirectional, because any relevant change in row data (concrete facts) should be translated and presented in terms of business concepts (synthesis), and vice-versa: the logic depicted by business concepts is used to guide and sharpen the mining of concrete local data (analysis).

The bottom layer deals with the integration of multiple data sources in a unified repository. In most business IT infrastructures, the Data-warehouse is devoted to this kind of tasks, which today is becoming more challenging than ever. Some facts causing this are (1) the availability of huge amounts of data (2) the proliferation of new and unstructured data sources like multimedia, hypertext... (3) the very high requirements in terms of availability and response time.

The analytical machinery (algorithms and processes) cross over all the three layers. We must take profit of the analytical power at every level of abstraction. For example, we can use a predictive model to forecast the company sales for the next month, which could be an application in the semantic layer; by the other hand, we

[1] Enterprise Resource Planning, Material Resource Planning and Customer Relationship Management, respectively.

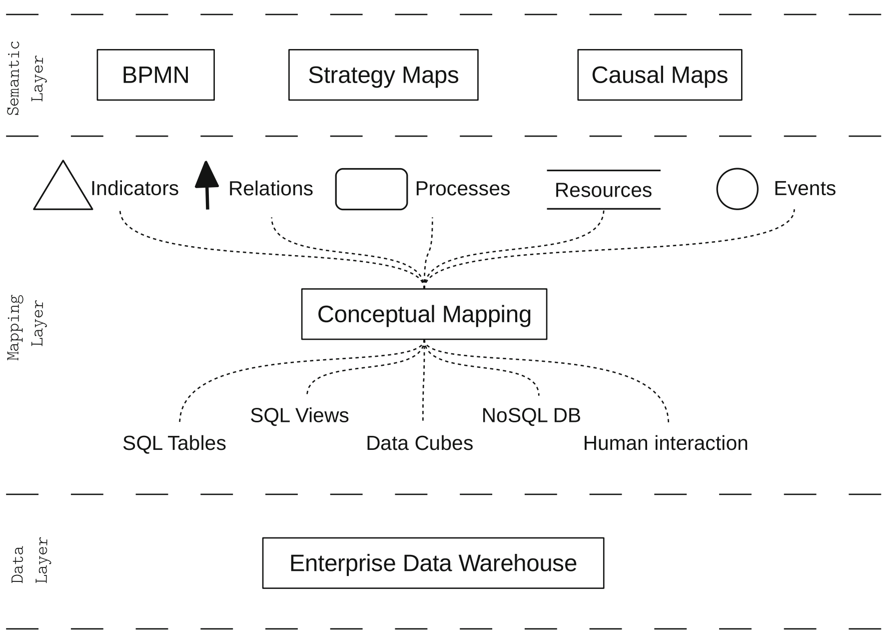


Fig. 1 Business intelligence architecture

could use a clustering model to clean up a dirty data source, before it becomes part of the data layer.

3.1 Semantic Layer: Business Modelling

The semantic layer allows business users to conceptualize their business operations and strategies using concepts that are familiar to them, including: Actor, Directive, Intention, Event, Situation, Indicator, Influence, Process, Resource, and Domain Assumption [7]. These concepts are synthesized from some well known management methodologies like Balanced Scorecard, Strategy Maps and Business Process Management. In this layer we will model the enterprise in different dimensions. The objective is to represent the business knowledge that people use in day to day work.

A *Balanced Scorecard* (BS) [10] is designed to align the work and priorities of employees with the strategic objectives that comprise an organization's defined mission. It allows managers to look at the business from four important perspectives: finance, customers, processes, innovation and learning. While keeping it simple, the BS meets several managerial key points: first, it brings together many seemingly disparate elements of a company's competitive challenges; second, it provides a tool for balancing the strategy across the fundamental business dimensions.

A *Strategy Map* is an illustration of an organization's strategy. It is extremely useful for simplify the translation of strategy into operational terms and to communicate to employees how their jobs relate to the organizations overall objectives. Strategy maps are intended to help organizations focus on their strategies in a comprehensive, yet concise and systematic way [11]. In fact, it works as the integration mechanism, in the sense that, the four BS perspectives, the associated strategic objectives, and the key performance indicators (KPI) are linked together as cause-and-effect relationships [12].

The main objective of *Business Process Management* (BPM) [13] is to align the processes with the business strategy. It essentially pursues the "achievement of the organization's objectives through the improvement, management and control of essential business processes". By the other hand, the Business Process Management Notation (BPMN) [14] is a notation standard for modelling business processes, probably the best known and established in the industry. The primary focus of BPM is in elements and processes, while BS and strategy maps focus on strategy and objectives.

Therefore, at this architectural level (the semantic layer) we make use of three different type of models. The strategy map (see Fig. 2) provides a way to depict a strategy that achieves a main goal [7]. A goal is split in several subgoals creating a hierarchy that clearly sets how the subgoals should be accomplished to achieve the main goal. To attain a particular goal, one or more processes most be performed, and a set of indicators are configured to measure the achievement level of every

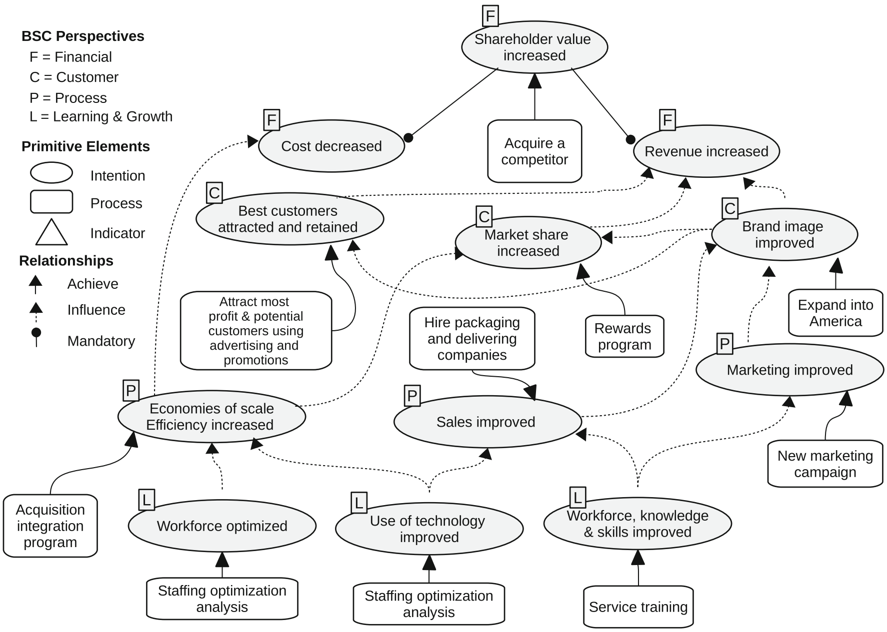


Fig. 2 Strategy map (adapted from [7])

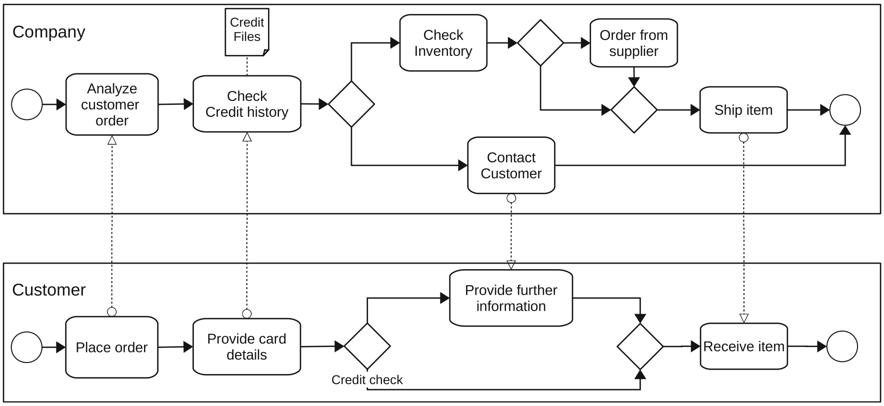


Fig. 3 BPMN model

goal (KPI). Notice that each goal has a label indicating which dimension of BSC corresponds to (financial, customer, process, learning & grow).

Within this modelling technique we regard the strategy, the metrics we are going to use to measure the accomplishment level, the key processes involved in this accomplishment, and the (causal) relation among these elements. Whereas the strategy components are all included in this diagram, not all the business process are depicted on it. For these other processes that we consider worth to monitor but are not included in the strategy map, we use the BPMN (Fig. 3).

The BPMN express the flow of processes, their relationships and interactions from the initial state to the final one. We widely adopt the standard BPMN 2.0 [14]. This type of models contain information about resources and how are they consumed/produced by the processes, as well as detailed information of each element like geographical location, organizational level, starting conditions, processing time. They provide an internal look of the organization, leaving aside the global picture of strategic goals.

In addition to these two modelling techniques, we consider the use of *Causal Maps* as a source of additional although different kind of business information. The causal maps help analysts to express the business intuitions that managers and operational employees hold about they work. These intuitions/ideas could be clearly deviated in any direction, according to the mental model hold by the people and the organizational culture. In spite of all these things, we claim that is worthy to construct such a diagram for many reasons: (1) they serve as a mean to unify and clarify the personal perception of the business; (2) they help to determine and focus on the factors that actually have impact on business; (3) they open the possibility of doing 'automatic' inference (reasoning) about the multiple influence and strengths that play a role in the day to day work (Fig. 4).

These three diagrams are connected to each other. The strategy plan provides the full map of goals, how they are related, how certain processes will help to achieve them. These 'certain' processes are fully specified in the business process model, in

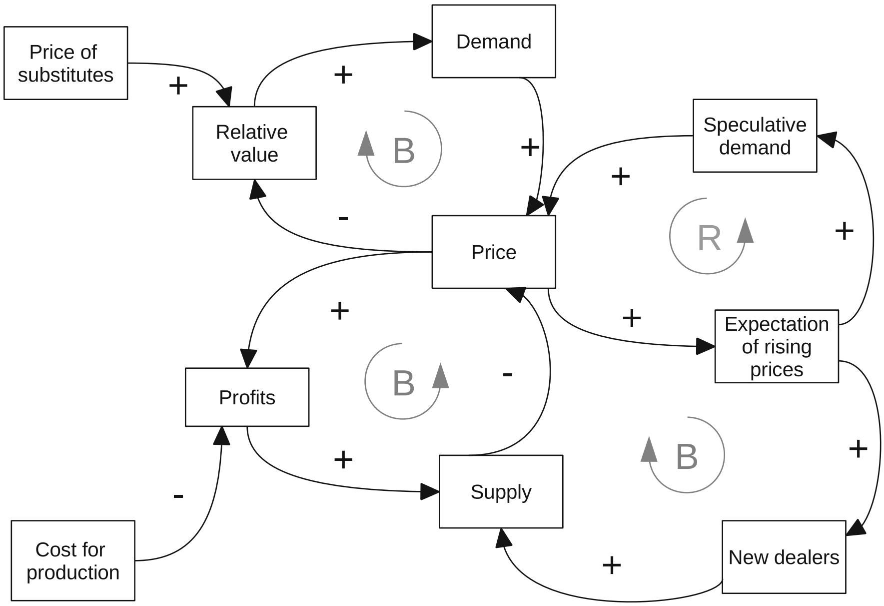


Fig. 4 Causal map

which we set the dynamic that arranges people, processes, resources, events, situations, restrictions… This 'dynamic' defines a set of clear relations or influences among the entities. For all other causal relations, which can not be derived from BPMN, but are contained in the knowledge formed by the people's experience, we use the causal maps.

The idea behind all this is to represent as much business knowledge as we can. It is in this layer where the business concepts acquire their meanings. It provides a semantic context in which the ideas are defined, and constitute the proper environment for direct interaction with the business user. From the users' perspective, this new 'semantic environment' will provide at least two big benefits. First, they will not be bothered with technical stuff, so they can center at business analysis and decision. Second, they are not left alone when analyse data: they are supported, guided and powered by the system, following the logic captured by the diagrams previously introduced.

3.2 Mapping Layer: Conceptual Mapping

One of the big challenges that information technology should face nowadays is bridging the gap between the *ideas* or *concepts* that humans use every, and the data elements that populate the enormous digital universe. As the society becomes more and more data-centered, and the data incrementally proliferate every day, the solution of this challenge gets more and more relevance.

Most of the available data is stored in relational format. It usually consists on a set of data tables, which are related by keys-columns following certain rules. The ERM [15] probably constitute the most well established de-facto standard for storing and representing data, even though some new and very promising approaches have emerged in the last years [16].

Many solutions have been proposed to close the representational gap between the storage layer (e.g., ERM) and the conceptual layer, which gives the 'user-semantic' meaning to the data. Some of them include EDM, Hibernate, Doctrine, RedBean, ActiveRecord. These mapping technologies are known as Object-Relational Mappings (ORM).

However, for BI applications these approaches are insufficient due to the huge volume of data involved. For a mid-size food company, it is not rare to handle 100 thousands billing transactions a day. The problems come when the managers and business users want to make sense of datasets that expand several years and interactively explore them, because the underling storage technology cannot timely support such a huge operations. That's why the so called Data-Warehouse technologies has come to existence, as well as the OLAP and DataMarts tools [17]. They intend to solve the problem by pre-aggregating the data into data-cubes, drastically reducing the system response time. In top of this data aggregation layer it is common to find data visualization tools, constituting the 'business intelligence' capability of such a systems. This kind of systems have become very popular in the IT industry, but they suffer some weaknesses. Their implementation usually involve important resources in terms of IT personnel effort and required time; they usually are pretty specific to the concrete industry and company involved; and being constructed following a bottom-up methodology, they can grow in size and complexity without real business necessity. For these reasons, some alternatives to Data-warehousing have been proposed. Some of them primarily focus on hardware optimization, like 'In-memory' technologies [18]. Other approaches to solve the gap between the conceptual and the storage facets of data, are based on conceptual modelling techniques. They help to raise the abstraction level, seamlessly connecting user's concepts with physical data.

The second layer of our Business Intelligence Architecture is the Conceptual Mapping, which is responsible of mapping the business concepts to the raw-data entities. Our approach is partially based on the Conceptual Integration Model (CIM) proposed by Rizzolo et al [9]. They extend MultiDim model [19], which in turn extends Chen's ER model to support multidimensionality (Fig. 5).

3.3 Data Layer: Data-Warehouses

Data are proliferating in volume and format, they are present throughout the organizations. Many data storage solutions have appeared in the last years, so

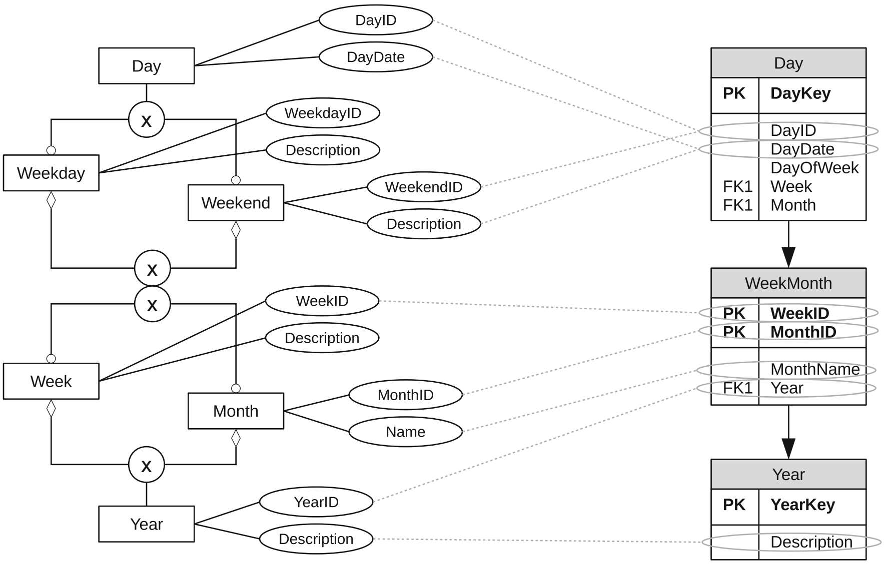


Fig. 5 Conceptual mapping excerpt

companies usually combine many of them. This layer is responsible for unifying all enterprise data sources available, in such a way it is possible to connect them with the Conceptual Mapping layer. We widely adopt ETL industry standard to this purpose, in addition to data warehousing technologies.

Data warehouses typically contain data from multiple organizational silos. This data is often more integrated, better understood, and cleansed more thoroughly. However, for building predictive analytic models they hide some drawbacks, as the problem of eliminating critical outlier data within the process of data cleansing. Despite this, data warehouses can be very useful for the construction of predictive analytic models, if built correctly. Doing so, you will regard at least the following tips [20]: (1) Data warehouses are not as space-constrained as operational databases, and mostly are used for historical analysis. As such, there is less pressure to delete unused records. So maintain as much data as you can. (2) The ability to store more data makes it more practical to store a new version of the record every time it is updated in the source system. It will avoid leaks from the future. (3) Many data warehouses are used to produce reports and analysis at a summary level. Taking wide profit of predictive analytic models imply storing transactional data as well as the roll-up and summary data. A well designed and implemented data warehouse is a great source of data that leverage the power of predictive analytics (Fig. 6).

Trying to gather in a single system all the information of a company, could become an arduous task, and sometimes impractical due to the volume of data involved. For that reason, some organizations have developed what has become to called data marts: data is extracted from operational databases of from an enterprise data warehouse and organized to focus on a particular solution area. Owned by a

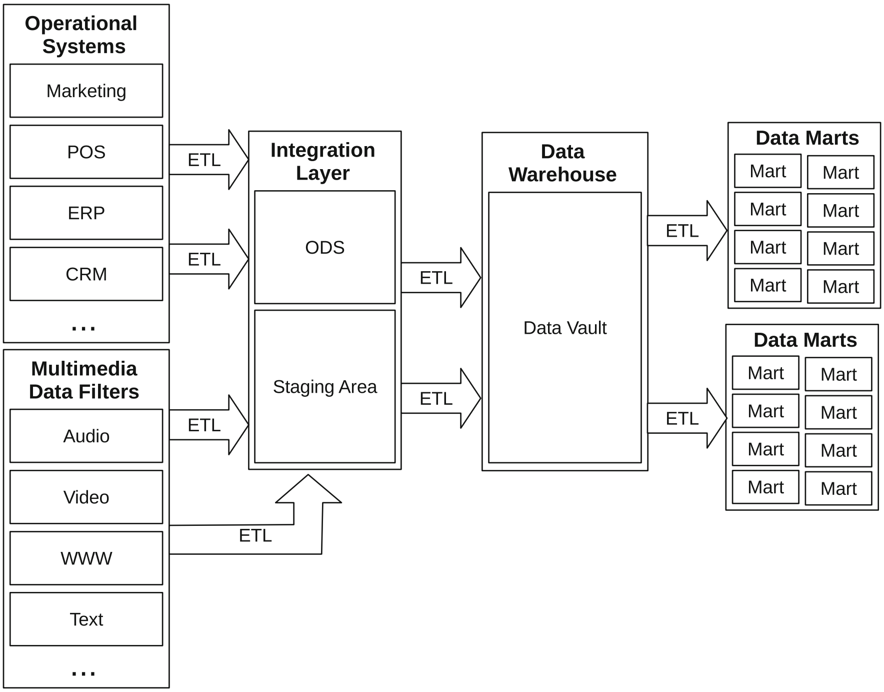


Fig. 6 Basic components of a data warehouse

single department of business area, data marts allow a particular group of user more control and flexibility when it comes to the data they need.

Another less common extension of data warehouses consist of connecting them with unstructured sources of data, like the Web, written documents, and any other media like images, audio and video. To do so, it is necessary to configure and specify how the raw data is transformed into regular data, what is the valid output range and what to do if something is wrong. This work is done by separated modules that we call Multimedia Data Filters (MDF). Basically an MDF receive a source media like a video and return a corresponding bunch of data. They focus on a particular type of source media and are powered by analytics, and need to be specify separately due to the involved complexity. Some example are sentiment analysis, web mining, video analysis and image decomposition.

4 Analytical Foundations

As we have already mentioned, at every level in the process of abstraction we take profit of analytical algorithms. They are usually gathered under the names of Pattern Recognition and Machine Learning algorithm, but is not unusual hear about Artificial Intelligence, Data Mining and Knowledge Discovery. What are all they really about?

Pattern Recognition deals with the problem of (automatically) finding and characterising patterns or regularities in data [21]. By patterns we understand any relations, regularities or structure inherent in a source of data. By detecting patterns, we can expect to build a system that is able to make predictions on new data extracted from the same source. If it works well, we say that the system has acquired *generalization power* by learning something from the data.

This approach is commonly called the *learning methodology*, because the process is focused on extracting patterns from the sample data that lead us to make generalizations about the population data [22]. In this sense, it is a *data driven* approach, in contrast with *theory driven* approach. However, it is extremely useful to tackle complex problems in which an exact formulation is not possible, for example, recognising a face in a photo or genes in a DNA sequence.

Consider a dataset containing thousands of observations of pea plants, in the same format of Gregor Mendel's observations. It is obvious that the characters (color and size, for example) of certain pea plant generation could be predicted by using the Mendel's laws. Therefore, the dataset contains an amount of redundancy, that is, information that could be reconstructed from other parts of the data. In such cases we say that the dataset is *redundant*.

This characteristic has an special importance for us, because the redundancy in the data leads us to formulate relations expressing such behaviours. If the relation is accurate and holds for all observations in the data, we refer to it as an *exact relation*. This is the case, for example, of the Laws of Inheritance: Mendel found that some patterns surprisingly held for all his experiments. For that reason, we say that this part of the data is also *predictable*: we can reconstruct it from the rest of the data, as well as predicting future data, like the color and size of new plants by using the current plants data.

Finding exact relations is not, by far, the general case for someone who analyses data. Certainly, the common case is finding patterns that hold with a certain probability. We call them *statistical relations*. Examples of such relations are: forecasting the total sales of a company for the next month, or inferring the credit score [23] of a new client in a bank by analysing his information.

The *science* of pattern analysis has considerably evolved from its early formulations. In the 1960s efficient algorithms for detecting linear relations were introduced. This is the case of the Perceptron algorithm [24], formulated in 1957. In the mid 1980s a set of new algorithms started to appear, making possible for the first time to detect nonlinear patterns. This group includes the backpropagation algorithm for multilayer neural networks and decision tree learning algorithms.

The emergence of the new pattern analysis approach known as kernel-based methods in mid 1990s, changed the field of pattern analysis towards a new and exciting perspective: the new approach enabled researchers to analyse nonlinear relations with the efficiency of linear algorithms via the use of kernel matrices. Kernel-based methods first appeared in the form of support vector machines (SVM), a classification algorithm that quickly gained great popularity in the community for its efficiency and robustness. Nowadays we have a variate and versatile toolbox composed by the algorithms developed by the scientific community during the short live of this research area.

5 Discussion and Related Work

Companies have traditionally adopted business intelligence solutions to support business decision making on a consistent daily basis, bringing data from disparate sources into a common data infrastructure or warehouse for reporting, analysis and creating analytic applications. By the other hand, emerging data discovery tools have focused instead on providing data-savvy user with free-form, more tactical analysis on single data sets. Definitely both are valuable in maximizing the value of data to an organization, but they do not use up the worth enclosed in the data. Moving forward, organizations will maximize value from BI and analytics by integrating those models and tools into a broader architecture that ultimately unify them. This approach will enable the "single source of truth" throughout the organization, as well as a more seamlessly and productive use of data by the staff.

This scenario provides several advantages in front of most current BA implementations:

- The final business users can interactively reason with the BA system by using business concepts, focusing on business issues.
- The business analysts can focus on business rules and logic. They can apply analytics' power directly into business reasoning, because the internal data-gears are transparent for them.
- The IT analysts can integrate multiple sources of data, and map them to 'business entities' to feed the whole architecture.

Nauman Sheikh [25] presents a nice guidebook on how to plan, design, and build analytics solutions to solve business problems. The approach is rather broad, dropping a light on every aspect of implementing analytics on organizations. For Sheikh, *analytics solution* means the process of collecting data, learning from it, anticipating scenarios, automating decisions and monitoring. One the big objectives of this work is the *simplification* of the full implementation of an analytics solution, as well as demystifying some topics associated to business analytics.

A very promising initiative in the industry of analytical software is the development of the Predictive Model Markup Language (PMML) [26] which provides a way for applications to define statistical and data mining models and to share models between PMML compliant applications. PMML provides applications a vendor-independent method of defining models so that proprietary issues and incompatibilities are no longer a barrier to the exchange of models between applications. It allows users to develop models within one vendor's application, and use other vendors' applications to visualize, analyze, evaluate or otherwise use the models [27].

Some efforts have been devoted to extend and profit the power of business modelling. Yu et al. [28] propose an approach that incorporate the intentional dimension of motivations, rationales and goals. While most Enterprise Architecture frameworks define components for data (what), function (how), network (where), people (who), time (when), the motivational aspect (why) has not received much attention, and is seldom supported by modelling. Nevertheless, understanding the motivation and intention of stakeholders is critical for architectural decisions and actions. Modelling the intentions will help to make them *transparent*, making it possible a *systematic analysis* of design implications, the exploration of possible strategies in a *rational way*, and the justification of activities by *tracing them back*.

References

1. Cukier K (2010) Data, data everywhere: a special report on managing information. Economist Newspaper
2. Phillips J (2014) Building a digital analytics organization
3. May T (2009) The new know: innovation powered by analytics, vol 23. Wiley, Hoboken, NJ
4. Mintzberg H (2009) Managing. Berrett-Koehler, San Francisco, CA
5. Buytendijk F (2010) Dealing with dilemmas: where business analytics fall short. Wiley, Hoboken, NJ
6. Davenport TH, Harris JG (2007) Competing on analytics: the new science of winning. Harvard Business Press
7. Barone D, Yu E, Won J, Jiang L, Mylopoulos J (2010) Enterprise modeling for business intelligence. In: The practice of enterprise modeling. Springer, Berlin, pp 31–45
8. Barone D, Mylopoulos J, Jiang L, Amyot D (2010) The business intelligence model: strategic modelling. Technical report, University of Toronto (April 2010)
9. Rizzolo F, Kiringa I, Pottinger R, Wong K (2010) The conceptual integration modeling framework. pp 1–18
10. Kaplan RS, Norton DP, Horv6th P (1996) The balanced scorecard, vol 6. Harvard Business School Press, Boston
11. Lawson R, Desroches D, Hatch T (2008) Scorecard best practices: design, implementation, and evaluation. Wiley, Hoboken, NJ
12. Kaplan RS, Norton DP (2004) Strategy maps: converting intangible assets into tangible outcomes. Harvard Business Press
13. Jeston J, Nelis J (2008) Business process management: practical guidelines to successful implementations. Routledge, London
14. Decker G, Dijkman R, Dumas M, García-Bañuelos L (2010) The business process modeling notation. In: Modern business process automation. Springer, pp 347–368.
15. Chen PPS (1976) The entity-relationship model—toward a unified view of data. ACM Trans Database Syst 1(1):9–36
16. Han J, Haihong E, Le G, Du J (2011) Survey on nosql database. In: Pervasive computing and applications (ICPCA), 2011 6th international conference on. IEEE, pp 363–366
17. Lenzerini M, Vassiliou Y, Vassiliadis P, Jarke M (2003) Fundamentals of data warehouses. Springer, Berlin
18. Plattner H (2009) A common database approach for oltp and olap using an in-memory column database. In: Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, pp 1–2
19. Malinowski E, Zimányi E (2008) Advanced data warehouse design: from conventional to spatial and temporal applications. Springer, Berlin

20. Taylor J (2011) Decision management systems: a practical guide to using business rules and predictive analytics. Pearson Education, Upper Saddle River, NJ
21. Shawe-Taylor J, Cristianini N (2004) Kernel methods for pattern analysis
22. Villegas-García MA (2013) An investigation into new kernels for categorical variables. Master Thesis, 2013. http://upcommons.upc.edu/handle/2099.1/17172
23. Huang C-L, Chen M-C, Wang C-J (2007) Credit scoring with a data mining approach based on support vector machines. Expert Systems with Applications 33(4):847–856
24. Rosenblatt F (1957) The perceptron—a perceiving and recognizing automaton. Technical Report 85-460-1
25. Sheikh N (2013) Implementing analytics: A blueprint for design, development, and adoption. Newnes
26. Data Mining Group. Predictive model markup language. See www.dmg.org
27. Guazzelli A, Lin W-C, Jena T (2012) PMML in action: unleashing the power of open standards for data mining and predictive analytics. CreateSpace
28. Yu E, Strohmaier M, Deng X (2006) Exploring intentional modeling and analysis for enterprise architecture. In: Enterprise Distributed Object Computing Conference Workshops, 2006. EDOCW'06. 10th IEEE International. IEEE, pp 32–32.

Vulnerability Analysis of a Signal-Controlled Road Network for Equilibrium Flow

Suh-Wen Chiou

1 Introduction

For an urban road network, most of travel delay is directly dependent on correct and continuous operation of signal settings at junctions. The reliability of a signal-controlled road network thus heavily depends on its vulnerability to a dangerous mix of probabilistic threats such as system random failure and signal-setting breakdown. The growing complexity of signal-controlled road junctions has increased opportunity of a formidable risk of sudden disruption such as loss of capacity on links due to signal setting failure at some road junction. Losing capacity in associated links when some signal settings fail to perform its intended function will have a negative wide impact on the performance of road network and give rise to substantial increase in total travel time on all road network users. The effects of signal-controlled road network-based disruption due to some given link capacity loss at signal-controlled junction will not only result in severe adverse consequence on traffic congestion but also increase substantial travel delay to all road users. The purpose of this chapter is to present a computationally tractable scheme to effectively evaluate the vulnerability of a signal-controlled road network to link capacity loss for user equilibrium flow. The performance of a signal-controlled road network can be conveniently assessed by a well-known TRANSYT model [59] in terms of a linear combination of delay rate and number of stops for all traffic streams. In order to model signal-controlled road network-based effect of link capacity loss on road user equilibrium flow, in this paper, a min-max complementarity problem is formulated and solved as a variational inequality where a robust equilibrium flow in the presence of a worst-case disruption of link capacity loss can be characterized. A modified projection method is presented to efficiently determine user equilibrium flow and identify important links whose disruption could cause substantial increase

S.-W. Chiou (✉)
Department of Information Management, National Dong Hwa University, Hualien, Taiwan
e-mail: chiou@mail.ndhu.edu.tw

© Springer International Publishing Switzerland 2015
F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_6

in travel delay to all road users. Numerical computations are conducted in a medium-sized road network to demonstrate the feasibility of the proposed solution scheme. The results indicate that the most critical signal-controlled traffic streams can be conveniently identified, when failed to perform its normal functions, would give rise to maximum travel delay to all road users.

Evaluation of network-based system for potential vulnerabilities is an important component of strategic planning, in particular in the context of reliability engineering and mitigating service disruptions. There has been a growing interest on the topic of vulnerability analysis for networked system over the past years [12, 24, 31, 33, 34, 37, 40]. Many methods have been proposed to facilitate vulnerability analysis, providing different measures employed for performance reliability of network-based system [3, 13, 15, 19–22, 26, 39, 45, 47, 48, 52, 55]. These methods differ primarily in how disruptions are evaluated and investigated. A disruption event, in the context of reliability engineering and system safety, is the set of networked components that are impacted, the degree to which they are disabled, and the performance and efficiency of networked system prior to the disruption. In some conditions, an affected component may be rendered completely inoperable by a disruption, e.g. removal of links in a road network. In other instances, a disruption may impact system performance to a lesser degree given that only some of the functionality of a component may be lost, as might be the condition with a signal setting failure causing associated link capacity loss and affecting traffic flow. Once disruption events are identified, the impact evaluation on the overall network effects is plausible. Impacts of disruption can range from those directly associated with networked system performance, such as connectivity, flow, or capacity reduction, to more complex associations like capacity loss impacts affecting users behaviour response and re-routing effects.

In vulnerability analysis, the ability to both characterize a range of disruption events and evaluate the impact of a specific set of disruption events perceived to be important is essential for managers to mitigate vulnerability of networked system to potential threats. However, in many situations, the event resulting in the greatest potential impact to networked system performance is of most interest. This is the scenario that managers and policy makers are often interested in when making decisions on where to direct funds and allocate defense resource for protecting or fortifying a networked system to potential threats [5, 23, 30, 32, 35, 36, 38, 43, 53]. If the scenario posing the greatest threat to a networked system can be identified, then mitigating vulnerability is possible. While a complete enumeration of disruptive scenarios can offer insight, the mathematical programming approaches have been widely employed to devise effective schemes to characterize critical networked system components and evaluate the underlying impacts to disruptions [46–51]. One mathematical programming approach seeking to identify those disruptive scenarios with the potential to most affect networked system performance with respect to the loss or fortification of components is interdiction modelling. The interdiction model is useful in the search for potentially important scenarios, especially since the topologies of networked system are often complex and relationships between components and network performance measure can be difficult to reconcile. In order to identify disruptive scenarios that pose the greatest threat to networked system vulnerability, a number of interdiction models have been devised to search for critical components in networked system or facilities in location science [14, 19, 25, 54, 61, 63]. Effective management of critical networked system requires the assessment of potential interdiction scenarios. Optimization based methods have been essential for identification and assessment of disruptive scenarios with potential to disable networked system operation most. Although a primary function of any networked system is the movement of flow between origins and destinations, the complexity and difficulty of mathematically abstracting interdiction impacts on networked system performance measure has been a challenge for researchers in the field of transportation science.

In order to understand overall reliability on transportation system performance in the presence of component degradation, Nicholson and Du [44] presented an integrated equilibrium model analyzing the performance and reliability of a degradable transportation system. The reliability of a degradable transportation system can be defined as the probability that the reduced flow of the system is less than a certain acceptable level of service. Asakura [4] used a Stochastic User Equilibrium model to assess the connectivity reliability of a small-sized network with the knowledge of probability distribution for link capacity degradations. The connectivity reliability of road network can be defined as a probability that there exists at least one path remain connected without disruption to a given destination within a given time period, that is, there continues to exist a connected path between a pair of nodes in the network when one or more links in the network have been impaired. For a transportation road network subject to traffic congestion, the network remains physically intact but performance of some part of it would be so severely affected by congestion and their use by traffic is curtailed. Most work in road network reliability accounts for behavioral change of road network users in response to occurrence of incidents [9, 27, 29, 41, 42, 56]. It is of interest to build various measures of performance reliability of road networks with certain occurrence of incidents. For instance, Chen at al. [17] investigated the reliability of capacity of road networks. On the other hand, vulnerability analysis of general road networks initially introduced by Berdica [11] is an alternative concept for identifying critical components in a transportation road network and evaluating the impacts of network degradation or disruptions. Vulnerability analysis of general road networks principally focuses on identifying the vulnerable facilities that could cause the most adverse impact on network performance as they are degraded. In the presence of uncertainty about the state of prevailing network performance, the road network user will intend to take a risk-averse position by spreading the risk across a number of routes with respect to major failure of components of transportation road networks. This is equivalently to supposing that road network user attempts to play through all possible eventualities before he seeks the least-cost route between a specific pair of origin and destination. More recently, Bell [7] and Bell and Cassir [8] applied a game-theoretic approach to measuring the expected performance of interest of a general road network and proposed a risk-averse model when road user faces unknown probabilities of incidents. Szeto et al. [57] and Szeto [58] gave a non-linear complementarity problem (NCP) formulation for a risk-averse traffic assignment with elastic travel demands. For a signal-controlled road network, the reliability of the network is heavily reliant on robust operation of signal settings and considerable traffic delay and queues could be incurred to all road users in the presence of signal setting failure. At the best knowledge of the author, there exists very limited research about road network reliability when signal setting failure is considered with exception of Lam et al. [28]. In Lam et al. [28], a simple user equilibrium model is proposed to account for route choice in a signal-controlled network where travel time and accident risk rate are evaluated by simple functions.

Traditionally, a measure of road vulnerability is a computationally intensive operation. Each network link is partially degraded or completely closed in turn and the performance of the degradable network is evaluated by solving a network equilibrium problem. Bell et al. [10] proposed a mixed route strategy to determine a risk-averse route choice for road users and minimize the maximum expected loss in the event of service disruption by conducting a simple heuristic using a simulation tool. Chen et al. [16] presented a network-based accessibility measure for assessing vulnerability of degradable road networks in terms of network travel time increase. A combined travel demand model was employed in Chen et al. [16] to model users' behavioral response and numerical computations were illustrated in a small-sized road network. While several studies [6, 10, 16, 26, 55] have investigated the measurement of transportation road network vulnerability and accounted for users behavioral response to service disruption, computational intractability in assessment of serviceability reduction has long been recognized as one of the most challenging issues facing researchers. In this paper, it is intended to fill this gap as far as evaluation of vulnerability of signal-controlled road network is concerned. A normative assessment of a signal-controlled road network reliability focusing on what could go wrong rather than what would go wrong is considered an alternative measure for vulnerability analysis of network performance to service disruptions. A well-known TRANSYT model [59] is considered to evaluate signal-controlled road network performance in terms of a combined measure of delay rates and number of stops. A min-max complementarity problem (CP) for vulnerability analysis of signal-controlled road network with equilibrium flow is established. An expected performance measure is employed in terms of a linear combination of expected delay rates and traffic queues in the presence of a worst-case of disruption of link capacity degradation a result of signal setting failure. Contrasting with conventional approaches in measuring the vulnerability of signal-controlled road network to link capacity degradation for user equilibrium flow, a min-max mathematical model in the sense of users' route choice is proposed for identification a set of disruptive traffic streams as a result of signal setting failure and evaluation of overall performance reliability of a signal-controlled road network with equilibrium flow.

The new insights of this chapter are stated in the following way. Firstly, a min-max mathematical model is established to characterize vulnerable links to service disruption of link capacity degradation due to signal setting failure in a signal-controlled road network and evaluate associated impacts affecting traffic flow on expected performance index of a signal-controlled road network.

According to Wardrop's first principle [60], the proposed min-max mathematical model can be equivalently formulated as a min-max variational inequality and effectively solved by a projection algorithm. Secondly, a min-max Nash game is envisaged between road network users who seek minimum travel cost route between specific pair of origin and destination, and a virtual attacker attempting to maximize expected trip cost of road users by disabling signal settings and decreasing associated link capacity at downstream which would cause the greatest adversity on signal-controlled road network performance. Thirdly, a modified projection method is developed to effectively solve the proposed min-max mathematical model with efficiently computational efforts. Numerical implementations are conducted on a medium-sized signal-controlled road network with equilibrium flows. For various scenarios of link capacity degradation, the proposed min-max model is demonstrated to be successfully solved with less computational efforts when compared to conventional approach. The numerical results indicate the feasibility of the proposed mathematical model for vulnerability analysis of signal-controlled road network with equilibrium flow. Moreover, the most important links can be successfully identified whose disruption could give rise to severity to overall networked system performance in terms of considerable travel delay increase. The remaining of this chapter is organized as follows. In next section, a min-max mathematical model expressed as a complementarity problem for vulnerability of signal-controlled road network with equilibrium flow is proposed. In order to measure signal-controlled road network performance under uncertainty, an expected performance measure in terms of linear combination of expected delay rates and number of stops for traffic streams (links) is employed. A robust user equilibrium flow is established following Wardrop's first principle in the presence of a worst-case of disruption of link capacity degradation. A min-max variational inequality is formulated for a robust user equilibrium flow in the presence of a worst-case of disruption of link capacity degradation. In Sect. 3, a modified projection method is presented to effectively solve a min-max variational inequality for robust user equilibrium flow in the presence of a worst-case of disruption of link capacity degradation. In Sect. 4, numerical computations are conducted in a medium-sized signal-controlled road network for a variety of disruptive scenarios of link capacity can be successfully identified. Comparative computations are also performed with conventional approach used in [6, 10], e.g. the method of successive averages (MSA) for vulnerability assessment of signal-controlled road network with equilibrium flow. Conclusion for this chapter and discussions for issues of interest of vulnerability analysis of signal-controlled road network are given in Sect. 5.

2 Problem Formulation

In order to characterize a range of disruptive scenarios of link capacity can be successfully identified with the potential to most affect the performance reliability of signal-controlled road network, in this section a min-max mathematical program is proposed to evaluate the expected performance of a signal-controlled road network for robust equilibrium flow. The expected performance of a signal-controlled road network for traffic flow can be approximately measured using TRANSYT model [59] in terms of a linear combination of weighted delay rates and number of stops. TRANSYT is a signal timing optimization program including a numerical model of traffic movement in which platoons of vehicular movements between adjacent junctions is a central feature. Regarding detailed calculations of indicator of traffic conditions at signal-controlled junctions, it can be referred to earlier results established in [18]. Furthermore, the expected travel cost for a specific pair of origin and destination in a signal-controlled road network can be measured by expected travel time along chosen routes. The expected route travel time is calculated as a sum of expected travel time on the links composing the route provided that the route travel time is additive. With regard to the notation used in the paper, we first introduce the notation as follows.

2.1 Notation

Let $G(N, L)$ denote a directed road network, where N represents a set of fixed time signal controlled junctions and L represents a set of links. Each traffic stream approaching any junction is represented by its own link. Within a particular time period, a given timing plan is implemented and the corresponding signal setting variables are assumed fixed.

W: a set of origin and destination (OD) pairs.

R_w: a set of routes connecting OD pair w, $\forall w \in W$

ϕ: a signal setting vector.

q: a travel demand matrix.

h: the average traffic flow on route in a vector form.

f: the average traffic flow on link in a vector form.

p_j: conditional probability of capacity degradation of link j due to signal setting failure.

Δs_j: a given capacity degradation on link j due to signal setting failure.

D_{ij}: conditional rate of delay incurred by road users on link i in the presence of a disruptive scenario of link capacity degradation on link j.

L_{ij}: conditional number of queuing vehicles incurred by road users for link i in the presence of a disruptive scenario of link capacity degradation on link j.

S_{ij}: conditional number of stops incurred by road users on link i in the presence of a disruptive scenario of link capacity degradation on link j.

d_{ij}: conditional average delay to road users arriving on link i in the presence of a disruptive scenario of link capacity degradation on link j.

t_i: the cruise travel time on link i.

c_i: the journey time on link i.

C_k: the cruise travel time of route k.

g_{jk}: the travel delay of route k in the presence of a disruptive scenario of link capacity degradation on link j.

λ: a link-route incident matrix where the entry $\lambda_{ik} = 1$ if link i lies on route k and 0 otherwise.

Λ: a OD-route incidence matrix where the entry $\Lambda_{kw} = 1$ if the route k connects the pair w of origin and destination and 0 otherwise, $\forall k \in R_w$, $\forall w \in W$.

2.2 The Expected Performance Measure of a Signal-Controlled Road Network

In a signal-controlled road network with disruptive scenarios of link capacity degradation perceived to be important, it is essential to mitigate vulnerability of a signal-controlled road network to a worst-case of link capacity degradation disruption that pose the greatest threat to road users in terms of substantial increase in total travel time. In this section, a maximized complementarity problem followed by equilibrium conditions is proposed to identify a worst-case of disruptive scenario of link capacity degradation due to signal setting failure. It aims to evaluate impacts of link capacity degradation on the performance reliability of a signal-controlled road network in which traffic streams would be severely affected and there would be substantial queue and traffic delay consequently incurred to all road users. For a signal-controlled road network with traffic flow, the measure of performance reliability of network in terms of traffic delay and queue can be effectively evaluated using a well-known model TRANSYT. The performance measure in TRANSYT is represented as a sum for signal-controlled traffic streams (links) of a weighted linear combination of estimated rate of delay and number of stops per unit time. Every traffic stream (link) in a signal-controlled road network has two states of indicator of traffic condition: normal state with certain traffic condition and disruptive state of link capacity degradation due to signal setting failure. For a normal state with certain traffic conditions, calculations of indicator of traffic condition can be referred to earlier results given in [18] and [62]. For a disruptive event of link capacity degradation due to signal setting failure, considerable travel delay could be incurred to all road users on associated degradable links.

Let D_i and S_i respectively represent estimate of rate of delay and number of stops on link i, the performance measure $P(\phi, f)$ with equilibrium flow f for a signal-controlled road network can be expressed as a sum of linear combination of weighted delay rates and number of stops in the following manner. Let W_{iD} and

W_{iS} be weighting factors for rate of delay and number of stops per unit time and M_D and M_S be the corresponding monetary factors, we therefore have

$$P(\phi,f) = \sum_{i \in L} D_i(\phi,f) W_{iD} M_D + S_i(\phi,f) W_{iS} M_S \tag{1}$$

In the presence of link capacity loss Δs_j on a degradable link j as a result of signal setting failure, the performance reliability measure $P_j(f, \Delta s_j)$ with equilibrium flow f for a signal-controlled road network can be represented in the following manner:

$$P_j(f, \Delta s_j) = \sum_{i \in L} D_{ij}(f, \Delta s_j) W_{iD} M_D + S_{ij}(f, \Delta s_j) W_{iS} M_S \tag{2}$$

The expected performance measure $AP(f, \Delta s)$ with equilibrium flow f for a signal-controlled road network in the presence of link capacity degradation Δs can be represented as follows.

$$\begin{aligned} AP(p,f, \Delta s) &= \sum_{i \in L} \sum_{j \in L} p_j D_{ij}(f, \Delta s_j) W_{iD} M_D + p_j S_{ij}(f, \Delta s_j) W_{iS} M_S \\ &= \sum_{j} P_j(f, \Delta s_j) p_j \end{aligned} \tag{3}$$

For a signal-controlled road network, a vulnerability analysis of signal-controlled road network subject to a worst-case of disruption of link capacity degradation can be considered establishing an upper bound of expected performance measure of a signal-controlled road network. In vulnerability analysis of signal-controlled road network, a maximized complementarity problem followed by equilibrium conditions is proposed. The purpose of establishing a maximized complementarity problem for a signal-controlled road network is stated in a twofold way. Firstly, it is to identify a set of disruptive scenarios of link capacity degradation that pose the greatest risk to all road users by causing substantial increase in total travel time. Secondly, it also aims to characterize impacts of link capacity degradation disruption on expected performance measure of a signal-controlled road network with equilibrium flow subject to traffic congestion. A maximized complementarity problem with performance reliability measure for a signal-controlled road network with respect to probabilities of occurrence of worst-case of link capacity degradation disruption due to signal setting failure is proposed first.

$$p_j \left(\underset{j}{Max} P_j(f, \Delta s_j) - P_j(f, \Delta s_j) \right) = 0, \quad \forall j \in L \tag{4}$$

In a maximized complementarity problem (4), the probability p_j of a worst-case of disruptive scenario of link capacity loss Δs_j on a degradable link j that pose the greatest risk to all road users by causing considerable increase in total travel time can be identified if and only if the following equilibrium conditions (5)–(6) are held.

If the performance reliability measure $P_j(f, \Delta s_j)$ of a disruptive scenario j of link capacity degradation is less than the maximum performance measure then the corresponding probability p_j of link capacity degradation disruption is zero. Or if the probability p_j of a disruptive scenario of link capacity degradation is greater than zero, then associated performance reliability measure $P_j(f, \Delta s_j)$ is equal to the maximum performance reliability measure. Therefore, the probability p_j is an occurrence of a worst-case of disruptive scenario of link capacity degradation and associated performance reliability measure $P_j(f, \Delta s_j)$ in a signal-controlled road network is a maximum performance measure in the presence of a worst-case of link capacity degradation disruption due to signal setting failure. The equilibrium conditions for probability p_j of occurrence of a disruptive scenario of link capacity degradation due to signal setting failure can be mathematically expressed in the following way. For every probability p_j of occurrence of a disruptive scenario of link capacity degradation together with a performance reliability measure $P_j(f, \Delta s_j)$ due to signal setting failure, it implies

$$p_j > 0 \qquad \Rightarrow \qquad P_j(f, \Delta s_j) = \underset{j}{Max}\, P_j(f, \Delta s_j) \tag{5}$$

and

$$p_j = 0 \qquad \Rightarrow \qquad P_j(f, \Delta s_j) \leq \underset{j}{Max}\, P_j(f, \Delta s_j) \tag{6}$$

2.3 The Expected Travel Time for Road Users

As it mentioned in Sect. 2.2, the expected performance measure of a signal-controlled road network can be evaluated when probability of a worst-case of link capacity degradation disruption is identified by a maximized complementarity problem (4) followed by equilibrium conditions addressed in (5)–(6). In this section, in the presence of a worst-case of disruptive scenario of link capacity degradation, the expected travel time for users in a signal-controlled road network can be characterized by a minimized complementarity problem followed by equilibrium conditions similarly. For a signal-controlled road network in the presence of normal traffic conditions, the average delay with signal settings ϕ and traffic flow f to road users can be denoted by a link delay function $d_i(\phi, f)$. According to traffic model used in TRANSYT, average rate of delay incurred by road users on link i can be characterize by

$$D_i(\phi, f) = f_i d_i(\phi, f), \qquad \forall i \in L \tag{7}$$

In the presence of a worst-case of disruptive scenario of link capacity loss Δs_j on a degradable link j due to signal setting failure, similarly, conditional average delay rate on link i can be in turn characterized by $d_{ij}(f, \Delta s_j)$. That is,

$$D_{ij}(f, \Delta s_j) = f_i d_{ij}(f, \Delta s_j), \qquad \forall i \in L \tag{8}$$

The journey time c_i on link i is represented as a sum of cruise time t_i on the link and conditional delay incurred by traffic at downstream junction and can be expressed below.

$$c_i = t_i + d_{ij}(f, \Delta s_j), \qquad \forall i \in L \tag{9}$$

In the presence of a worst-case of disruptive scenario of link capacity degradation Δs as a result of signal setting failure, let $sc_i(f, \Delta s)$ denote the expected travel time on link i, which can be expressed as a sum of cruise time t_i and expected delay $\sum_{j \in L} p_j d_{ij}(f, \Delta s_j)$. Since the probability p_j of a worst-case of disruptive scenario of link capacity loss Δs_j on a degradable link j can be identified by a maximized complementarity problem (4) followed by equilibrium conditions (5)–(6), it implies

$$sc_i(f, \Delta s) = t_i + \sum_{j \in L} p_j d_{ij}(f, \Delta s_j), \qquad \forall i \in L \tag{10}$$

In a signal-controlled road network for a pair w of origin and destination, $\forall\, w \in W$, there exists a route k connecting the origin and destination pair, $\forall\, k \in R_w$, $\forall\, w \in W$. The expected travel time SC_k along route k using the definition of (10) can be represented in the following manner: let $SC_k(h, \Delta s)$ and $g_{jk}(h, \Delta s_j)$ respectively denote expected travel time and conditional delay on route k in the presence of a worst-case of link capacity degradation disruption Δs, we have

$$\begin{aligned} SC_k(h, \Delta s) &= \sum_{i \in L} \lambda_{ik} sc_i(f, \Delta s) \\ &= \sum_{i \in L} \lambda_{ik} t_i + \sum_{i \in L} \lambda_{ik} \sum_{j \in L} p_j d_{ij}(f, \Delta s_j) \\ &= C_k + \sum_{j \in L} \sum_{i \in L} \lambda_{ik} p_j d_{ij}(f, \Delta s_j) \\ &= C_k + \sum_{j \in L} p_j g_{jk}(h, \Delta s_j) \end{aligned} \tag{11}$$

As it has been enunciated by Wardrop [60], the journey time on all routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route. In a signal-controlled road network the resulting traffic flow for user equilibrium is such that the travel time on all used routes connecting any given origin-destination pair will be equal. The travel time on all used routes will also be less than or equal to those on any of unused routes. At this point, no road user can experience a lower travel time by unilaterally changing routes and hence the signal-controlled road network is in user equilibrium. In accordance with

Wardrop's first principle, in the presence of a worst-case of disruptive scenario of link capacity degradation due to signal setting failure the resulting traffic flow for a robust route-choice game is such that the travel time on all used routes connecting any given origin-destination pair will be equal. The travel time in the presence of a worst-case of disruption of link capacity degradation on all used routes will also be less than or equal to those on any of unused routes. At this point, no road user in the presence of risk of link capacity degradation can experience a lower travel time by unilaterally changing routes and hence the signal-controlled road network in the presence of a worst-case of disruption of link capacity degradation is in robust user equilibrium.

In (11) the resulting traffic flow f is such that the expected travel time on all used routes k connecting any given origin-destination pair will be equal. The expected travel time SC_k on all used routes will also be less than or equal to those on any of unused routes. At equilibrium, no road user will gain an advantage by unilaterally changing his choice of route given choices of route made by other users in the presence of threat of link capacity degradation according to Wardrop's first principle. A robust user equilibrium model can be thus modeled as a minimized complementarity problem followed by equilibrium conditions. As it shown in (11), there exists a route k connecting a pair w of origin and destination in a signal-controlled road network. According to Aashtiani and Magnanti [1], a robust equilibrium flow h connecting a pair w of origin and destination along route k, $\forall k \in R_w$, $\forall w \in W$ in the presence of a worst-case of link capacity degradation disruption can be expressed as a minimized complementarity problem in the following way:

$$h_k\left(SC_k(h, \Delta s) - \underset{k}{Min} SC_k(h, \Delta s)\right) = 0 \tag{12}$$

and $q_w = \sum_{k \in R_w} \lambda_{kw} h_k$, $\forall w \in W$. Let π_w denote a minimum expected travel time for a pair w of origin and destination, $\forall w \in W$. The equilibrium conditions followed by a minimized complementarity problem (12) for a robust equilibrium flow h along route k connecting a pair w of origin and destination, $\forall k \in R_w$, $\forall w \in W$ are such that

$$h_k(SC_k(h, \Delta s) - \pi_w) = 0 \tag{13}$$

$$SC_k(h, \Delta s) - \pi_w \geq 0 \tag{14}$$

$$h_k \geq 0 \tag{15}$$

In (13)–(15), in the presence of a worst-case of disruption of link capacity degradation Δs due to signal setting failure, it shows that for all used routes k connecting a pair w of origin and destination carrying traffic flow the expected travel time SC_k, $\forall k \in R_w$, $\forall w \in W$ equals the minimum expected travel time π_w. While for any unused route carrying no traffic flow, the corresponding travel time is not less than the expected minimum travel time. Therefore the equivalent conditions in (13)–(15) for a robust user equilibrium flow h on route k, $\forall k \in R_w$, $\forall w \in W$

in a signal-controlled road network can be mathematically expressed in the following way. For any route k connecting a pair w of origin and destination, $\forall k \in R_w$, $\forall w \in W$, it implies

$$h_k > 0 \quad \Rightarrow \quad SC_k(h, \Delta s) = \pi_w \tag{16}$$

and

$$h_k = 0 \quad \Rightarrow \quad SC_k(h, \Delta s) \geq \pi_w \tag{17}$$

2.4 A Min-Max Mathematical Program

In order to simultaneously characterize a set of vulnerable links of signal-controlled road network in the presence of link capacity degradation and evaluate impacts on user equilibrium flow, a min-max mathematical program is proposed for complementarity problems respectively given in (4) and (12). As it expressed in a maximized complementarity problem (4), probabilities of occurrence of worst-case of link capacity degradation disruption together with expected performance measure of a signal-controlled road network can be mathematically characterized by equilibrium conditions stated in (5)–(6). Moreover, as it expressed in a minimized complementarity problem (12), a robust user equilibrium flow following Wardrop's principle can be determined by equilibrium conditions (13)–(15) on which the expected performance measure of a signal-controlled road network is based in the presence of a worst-case of link capacity degradation disruption due to signal setting failure. In other words, for a maximized complementarity problem (4) followed by equilibrium conditions (5)–(6), it aims to identify a probability of occurrence of worst-case of link capacity degradation disruption in a signal-controlled road network given traffic flow distribution characterized by (12). On the other hand, for a minimized complementarity problem (12) followed by equilibrium conditions (13)–(15), it focuses on evaluating impacts of a worst-case of link capacity degradation identified by equilibrium conditions (5)–(6) on expected performance measure in a signal-controlled road network such that a robust equilibrium flow can be determined following Wardrop's first principle.

In this section, a robust route-choice Nash game is envisaged between road users in a signal-controlled road network seeking minimum expected travel time route for a specific pair of origin and destination, and a virtual attacker attempting to maximize expected performance measure of a signal-controlled road network by disrupting link capacity and in turn causing considerable delay to all road users. A min-max variational inequality is proposed to characterize robust user equilibrium flow and associated expected performance measure of a signal-controlled road network in the presence of a worst-case of link capacity degradation disruption. A robust route choice Nash equilibrium is achieved if and only if a user equilibrium flow following conditions in (13)–(15) solves a min-max variational inequality on which a worst-case of link capacity degradation disruption satisfying conditions

(5)–(6) is based. At a route choice Nash equilibrium, a robust equilibrium flow satisfying equilibrium conditions (13)–(15) is mutually consistent with the probabilities of occurrence characterized by conditions (5)–(6) of a worst-case of link capacity degradation disruption in a signal-controlled road network. The critical traffic streams (links) that pose the greatest threat to expected travel time to all road users are also identified for robust user equilibrium flow on which a worst-case expected performance measure is based subject to disruptive scenario of link capacity degradation. In order to characterize a robust user equilibrium flow together with expected performance measure of a signal-controlled road network in the presence of link capacity degradation, a min-max variational inequality is introduced first.

In a maximized complementarity problem (4), the probability p with performance reliability measure given in (2) of characterizing a worst-case of disruptive scenario of link capacity degradation Δs due to signal setting failure with traffic flow f can be optimally determined by a maximization of expected performance measure given in (3). That is,

$$\underset{p}{Max} \quad AP(p,f,\Delta s) \tag{18}$$
$$= \sum_{i\in L}\sum_{j\in L} p_j D_{ij}(f,\Delta s_j) W_{iD} M_D + p_j S_{ij}(f,\Delta s_j) W_{iS} M_S$$

subject to $\sum_{j\in L} p_j = 1$

and $p_j \geq 0, \forall j \in L$

The expected performance measure of a signal-controlled road network obtained from (18) can be considered an upper bound for a robust user equilibrium flow with respect to probabilities of occurrence of a worst-case of link capacity degradation disruption. Such a maximization of expected performance measure can be generalized as a following variational inequality (19). For every probability p_j of occurrence of disruptive scenario of capacity loss Δs_j on a degradable link j with a corresponding performance reliability measure $P_j(f,\Delta s_j)$ introduced in (2), let $\Omega_p = \left\{\sum_{j\in L} p_j = 1, p_j \geq 0\right\}$. Find a probability $p \in \Omega_p$ of a worst-case of disruptive scenario of link capacity degradation with a corresponding performance reliability measure $P(f,\Delta s)$ which would cause the greatest travel delay to all road users such that

$$P(f,\Delta s)(p - p') \geq 0, \quad \forall p' \in \Omega_p \tag{19}$$

On the other hand, consider equilibrium conditions in (13)–(15) for a robust user equilibrium flow with a probability of occurrence of a worst-case of link capacity degradation disruption identified by (19). A general formulation for a robust user equilibrium flow following Wardrop's principle in the presence of a worst-case of link capacity degradation disruption can be expressed in terms of a following variational inequality. Let $\Omega = \{q = \Lambda h, h \geq 0\}$. Given a set of probabilities of occurrence of a worst-case of link capacity degradation disruption characterized by (19) together with performance reliability measure of a signal-controlled road network, a route flow h, $\forall h \in \Omega$ connecting pairs of origin and destination with expected travel time $SC(h, \Delta s)$ can be determined such that

$$SC(h, \Delta s)(h' - h) \geq 0, \quad \forall h' \in \Omega \tag{20}$$

Or in an equivalent form following (11):

$$(C + pg(h, \Delta s))(h' - h) \geq 0, \quad \forall h' \in \Omega \tag{21}$$

Let Ω^* denote a solution set for route flow h connecting pairs of origin and destination, which solving robust user equilibrium (20) with expected performance measure of a worst-case of link capacity degradation disruption.

Analogously, a robust user equilibrium flow in terms of link flow can be determined in a similar way: let $\Omega_f = \{f; f = \lambda h, q = \Lambda h, h \geq 0\}$. Given a set of probabilities of occurrence of a worst-case of link capacity degradation disruption characterized by (19) together with performance reliability measure of a signal-controlled road network, a link flow f, $\forall f \in \Omega_f$ with expected travel time $sc(f, \Delta s)$ can be determined such that

$$sc(f, \Delta s)(f' - f) \geq 0, \quad \forall f' \in \Omega_f \tag{22}$$

Or in an equivalent form following (10):

$$(t + pd(f, \Delta s))(f' - f) \geq 0, \quad \forall f' \in \Omega_f \tag{23}$$

Again, let Ω_f^* denote a solution set for link flow f in a signal-controlled road network solving robust user equilibrium (22) with expected performance measure of a worst-case of link capacity degradation disruption. A mutually consistent solution can be found below for a min-max variational inequality (19) and (22) such that a route choice Nash game equilibrium exists in a signal-controlled road network with robust equilibrium flow.

Proposition 1 (**Existence of a robust route choice Nash game**) In the presence of a worst-case of disruptive scenario of link capacity degradation due to signal setting failure, the performance measure in (19) and (20) is continuous and the feasible demand set Ω for route flow is a closed convex set. There exists a robust route choice Nash game (p, h) in a signal-controlled road network for (19) and (20) such that the following solutions are mutually consistent.

For a performance reliability measure $P(f, \Delta s)$ in the presence of disruptive scenarios of link capacity degradation Δs with flow f determined by (25), it is to characterize a probability $p \in \Omega_p$ of occurrence of link capacity degradation disruption causing the greatest increase in expected performance measure such that

$$P(f, \Delta s)(p - p') \geq 0, \quad \forall p' \in \Omega_p \tag{24}$$

And to identify a route flow h, $\forall h \in \Omega$ connecting pairs of origin and destination in a signal-controlled road network with expected travel time $SC(h, \Delta s)$ in the presence of probabilities of occurrence of worst-case of link capacity degradation disruption Δs characterized by (24) such that

$$SC(h, \Delta s)(h' - h) \geq 0, \quad \forall h' \in \Omega \tag{25}$$

It implies that a pair of mutually consistent points (p, h) exists in a signal-controlled road network, which satisfies a robust Wardrop's first principle that no user will take a route having a perceived cost higher than the lowest of the perceived travel time that prevail on the available routes in the light of the choices of all the other users. □

Proposition 2 (**Existence of a robust Nash game in terms of link flow**) In the presence of a worst-case of disruptive scenario of link capacity degradation due to signal setting failure, the performance measure in (19) and (22) are continuous and the feasible demand set Ω_f for link flow is a closed convex set. There exists a robust route choice Nash game (p,f) in a signal-controlled road network for (19) and (22) such that the following solutions are mutually consistent. For a performance reliability measure $P(f, \Delta s)$ in the presence of link capacity degradation Δs with flow f determined by (27), it is to characterize a probability $p \in \Omega_p$ of occurrence of link capacity degradation disruption causing the greatest increase in expected performance measure such that

$$P(f, \Delta s)(p - p') \geq 0, \quad \forall p' \in \Omega_p \tag{26}$$

And to identify link flow f, $\forall f \in \Omega_f$ with expected travel time $sc(f, \Delta s)$ in the presence of probabilities of occurrence of worst-case of link capacity degradation disruption Δs characterized by (26) such that

$$sc(f, \Delta s)(f' - f) \geq 0, \quad \forall f' \in \Omega_f \tag{27}$$

It implies that a pair of mutually consistent solutions (p,f) exists in a signal-controlled road network, which satisfies a robust Wardrop's first principle that no user will take a route having a perceived cost higher than the lowest of the perceived travel time that prevail on the available routes in the light of the choices of all the other users. □

3 Solution Approach

In order to determine a robust user equilibrium flow in the presence of a worst-case of disruptive scenario of link capacity degradation in a signal-controlled road network, in this section a solution approach is proposed to effectively characterize a robust Nash equilibrium as introduced in Proposition 2 for a min-max variational inequality (26) and (27). A variational inequality (26) identifying probabilities of occurrence of worst-case of link capacity degradation disruption can be equivalently expressed as a maximization problem as shown in (18) and conveniently solved as a linear programming problem. Thus in this section a modified projection method aiming to solve a variational equality (27) is proposed to characterize a robust user equilibrium flow in the presence of a worst-case of link capacity degradation disruption due to signal setting failure. In the following, a projection operator is introduced first for a robust user equilibrium flow in a signal-controlled road network in the presence of a worst-case of disruptive scenario of link capacity degradation. A distance measure of projection of traffic flow on the feasible set is defined such that a fixed point of the variation inequality (27) can be characterized. Moreover, a fixed point of a robust user equilibrium flow solves the variational inequality (27) if and only if a distance measure between iterates of solution to variational inequality (27) equals zero.

3.1 Projection Operator

In the presence of a worst-case of link capacity degradation disruption due to signal setting failure in a signal-controlled road network, a robust user equilibrium flow introduced in Proposition 2 can be effectively characterized using a projection-based approach for a robust Nash game in terms of link flow. Firstly, a projection operator is defined for a traffic flow f, $\forall f \in \Omega_f$ in the presence of probability p, $\forall p \in \Omega_p$ identified by (18) of occurrence of worst-case of link capacity degradation disruption in the following way.

Proposition 3 (**Projection operator for traffic pattern**) In a signal-controlled road network for any traffic pattern $x \in \Omega_f$ there exists a point

$$y = \mathrm{Pr}_{\Omega_f}(x) \tag{28}$$

known as the projection of x on the feasible set Ω_f such that

$$(z - y)(x - y) \leq 0, \qquad \forall z \in \Omega_f \tag{29}$$

The projection y with respect to Euclidean norm can be also represented

$$y = \Pr_{\Omega_f}(x) = Arg \underset{z \in \Omega_f}{Min} \|x - z\| \tag{30}$$

The projection y with respect to Euclidean norm has the following property:

$$\|x - y\| \le \|x - z\|, \qquad \forall z \in \Omega_f \tag{31}$$

□

According to Proposition 3, a fixed point of variational inequality (27) is characterized below for a robust user equilibrium flow introduced in Proposition 2 in the presence of a worst-case of link capacity degradation disruption due to signal setting failure in a signal-controlled road network.

Proposition 4 (**Fixed point of a robust user equilibrium flow**) In the presence of a worst-case of link capacity degradation disruption Δs due to signal setting failure in a signal-controlled road network, as introduced in Proposition 2, there exists a robust user equilibrium flow f for variational inequality (27) with expected travel time $sc(f, \Delta s)$ defined in (10) if and only if the traffic flow f in a signal-controlled road network is a fixed point of the following projection function:

$$\Pr_{\Omega_f}(I - sc(\cdot, \Delta s)) : \Omega_f \mapsto \Omega_f \tag{32}$$

That is, for every traffic flow $f \in \Omega_f^*$ in a signal-controlled road network

$$f = \Pr_{\Omega_f}(f - sc(f, \Delta s)) \tag{33}$$

In (33), equivalently, according to (10) f is a fixed point of the following projection function:

$$\Pr_{\Omega_f}(I - t - pd(\cdot, \Delta s)) : \Omega_f \mapsto \Omega_f \tag{34}$$

That is

$$f = \Pr_{\Omega_f}(f - t - pd(f, \Delta s)) \tag{35}$$

□

Proposition 5 (**Distance measure of a robust user equilibrium flow**) Following Proposition 4, let $e(f, \Delta s)$ denote a distance measure between iterates of solution to variational inequality (27), i.e.

$$e(f, \Delta s) = f - \Pr_{\Omega_f}(f - sc(f, \Delta s)) \tag{36}$$

For every fixed point, $f \in \Omega_f^*$ is a fixed point of the projection function (32) if and only if

$$e(f, \Delta s) = f - \mathrm{Pr}_{\Omega_f}(f - sc(f, \Delta s)) = 0 \tag{37}$$

□

According to Propositions 4 and 5, a distance measure defined in (37) can be regarded as a descent direction in the search for a robust user equilibrium flow of (27). Moreover, following Proposition 5, a traffic flow f, $f \in \Omega_f^*$ is a robust user equilibrium flow of (27) if and only if the distance measure characterized in (37) vanishes and it concludes that f is a fixed point of the projection function (32).

3.2 A Projection Approach

In order to effectively characterize a robust user equilibrium flow for a variational inequality (27), in this section a modified projection method using a distance measure defined in (36) is proposed. A distance measure introduced in Proposition 5 is employed as a descent direction in the effective search for a robust user equilibrium flow in (27). Following Proposition 3, a projection of traffic flow f in a signal-controlled road network in the presence of a worst-case of link capacity degradation disruption is defined in the following manner such that a fixed point of a projection function (32) can be characterized for a robust user equilibrium flow in a signal-controlled road network:

$$f^{(n+1)} = \mathrm{Pr}_{\Omega_f}\left(f^{(n)} - \rho e\left(f^{(n)}, \Delta s\right)\right) \tag{38}$$

and $\rho = \tau \frac{AP\left(p, f^{(n)}, \Delta s\right) - AP(p, f^*, \Delta s)}{\left\|e\left(f^{(n)}, \Delta s\right)\right\|^2}, \quad 0 < a \leq \tau \leq 2 - b, b > 0$

where a robust user equilibrium flow f^* and expected performance measure $AP(p, f^*, \Delta s)$ defined in (18) in the presence of a worst-case of link capacity degradation disruption Δs are supposed to be known and $\tau = \frac{1}{n}$. In (38), an effective search for a robust user equilibrium flow f^* in the presence of a worst-case of link capacity degradation disruption can be computationally tractably performed by consecutive projections of traffic flow on feasible set until a fixed point of a robust user equilibrium flow is characterized. By appropriately choosing parameters in determining a step length along search direction, a robust traffic flow is continuously updated and corresponding expected performance measure defined in (18) is successively improved from iteration to iteration. A robust user equilibrium flow in the presence of a worst-case of link capacity degradation disruption can be determined if and only if the distance measure defined in (36) vanishes such that a fixed point of projection function (32) for variational inequality (27) can be characterized.

3.3 An Efficient Computation Scheme

A robust user equilibrium flow f for a route choice Nash game can be effectively determined in this section using a modified projection method proposed in (38). A consonant solution pair (p,f) of a signal-controlled road network can be characterized for (26) and (27) in the presence of worst-case of link capacity degradation disruption. For a worst-case of disruption of link capacity degradation identified with occurrence of probabilities p due to signal setting failure, a mutually consistent robust user equilibrium flow f in a signal-controlled road network can be characterized in the following way.

Step 1. Start with initial traffic flow $f^{(k)}$ and probabilities $p^{(k)}$ of occurrence of a worst-case of link capacity degradation disruption. Set index $k=0$ and stopping threshold ε.

Step 2. Solve a maximization problem (18) with respect to probabilities $p^{(k+1)}$ of occurrence of a worst-case of link capacity degradation disruption as a result of signal setting failure for user equilibrium traffic flow $f^{(k)}$. Update the expected performance measure of a signal-controlled road network on which the occurrence of probabilities $p^{(k+1)}$ and traffic flow $f^{(k)}$ in the presence of link capacity degradation disruption are based. Set an upper bound estimate for a performance measure of a signal-controlled road network as $\overline{P} = AP\left(p^{(k+1)}, f^{(k)}, \Delta s\right)$ with respect to occurrence of probabilities $p^{(k+1)}$ of a worst-case of link capacity degradation disruption due to signal setting failure. Details about setting an upper bound estimate $\overline{P}$ with respect to occurrence of probability p are described in the following sub-steps.

Step 2-1. Evaluate a range of occurrence of performance reliability measure $P_j(f^{(k)}, \Delta s_j)$ for traffic flow $f^{(k)}$ in the presence of link capacity degradation Δs_j due to signal setting failure.

Step 2-2. Solve a maximization problem (18) with respect to occurrence of probability $p^{(k+1)}$ such that the expected performance measure for traffic flow $f^{(k)}$ is maximized, that is

$$p^{(k+1)} = \underset{p}{Arg\max}\, AP\left(p, f^{(k)}, \Delta s\right) \tag{39}$$

Step 2-3. Update the expected performance measure (18) with respect to occurrence of probabilities $p^{(k+1)}$ of a worst-case of link capacity degradation disruption Δs for traffic flow $f^{(k)}$. An upper bound estimate for expected performance of a signal-controlled road network can be characterized on which the occurrence of probabilities $p^{(k+1)}$ and traffic flow $f^{(k)}$ are based, that is

$$\overline{P} = AP\left(p^{(k+1)}, f^{(k)}, \Delta s\right) \tag{40}$$

Step 3. Solve a variational inequality (27) using a modified projection method (38) for a robust user equilibrium flow $f^{(k+1)}$ in the presence of occurrence of probabilities $p^{(k+1)}$ of a worst-case of link capacity degradation disruption Δs. The robust equilibrium flow $f^{(k+1)}$ is determined such that a fixed point of projection function (32) for variational inequality (27) can be characterized. Update the expected performance measure on which the occurrence of probabilities $p^{(k+1)}$ and traffic flow $f^{(k+1)}$ are based. Set a lower bound estimate for a performance measure of a signal-controlled road network as $\underline{P} = AP\left(p^{(k+1)}, f^{(k+1)}, \Delta s\right)$ for a robust user equilibrium flow $f^{(k+1)}$ in the presence of probabilities $p^{(k+1)}$ of occurrence of a worst-case of link capacity degradation disruption due to signal setting failure. Details about setting a lower bound estimate $\underline{P}$ with respect to traffic flow f are described in the following sub-steps.

Step 3-1. Evaluate the expected travel time $sc(f^{(k)}, \Delta s)$ in (27) for road users in the presence of occurrence of probabilities $p^{(k+1)}$ on which a worst-case of link capacity degradation disruption is based.

Step 3-2. Determine a robust user equilibrium flow $f^{(k+1)}$ for variational inequality (27) using a modified projection method (38) in the presence of occurrence of probabilities $p^{(k+1)}$ on which a worst-case of link capacity degradation disruption is based. At a robust user equilibrium flow $f^{(k+1)}$ a fixed point of projection function (32) is characterized for (27) and the following condition holds

$$f^{(k+1)} = \mathrm{Pr}_{\Omega_f}\left(f^{(k+1)} - sc\left(f^{(k+1)}, \Delta s\right)\right) \tag{41}$$

Step 3-3. Update the expected performance measure (18) with respect to a robust equilibrium flow $f^{(k+1)}$ in the presence of occurrence of probabilities $p^{(k+1)}$ on which a worst-case of link capacity degradation disruption is based. A lower bound estimate for the expected performance measure of robust equilibrium flow $f^{(k+1)}$ can be characterized on which the occurrence of probabilities $p^{(k+1)}$ under a worst-case of link capacity degradation is based. That is,

$$\underline{P} = AP\left(p^{(k+1)}, f^{(k+1)}, \Delta s\right) \tag{42}$$

Step 4. If the distance of expected performance measure between upper bound estimate $\overline{P}$ and lower bound estimate $\underline{P}$ is within a threshold ε, i.e. $\left|\overline{P} - \underline{P}\right| \leq \varepsilon$ of signal-controlled road network in the presence of a worst-case of link capacity degradation disruption, then stop. According to Proposition

2, we have a robust user equilibrium flow $f^{(k+1)}$ for a min-max variational inequality (26) and (27) in the presence of probabilities $p^{(k+1)}$ of occurrence of a worst-case of link capacity degradation disruption. Otherwise, increase index k by 1 and go to Step 2.

4 Numerical Computations

In order to investigate computational efficiency and solution effectiveness of proposed modified projection method for robust user equilibrium flow in the presence of a worst-case of link capacity degradation disruption, in this section, numerical computations are performed using a medium-scaled road signal-controlled network [2] with various levels of link capacity degradation disruption. Basic layout of the example road network is given in Fig. 1. This numerical test includes 22 pairs of trip-ends, 23 links at 6 signal-controlled junctions. Initial travel data for various pairs of origin and destination are given in Table 1. Using typical values found in practice, the minimum green time for each signal-controlled group is 7 s, and the clearance times are 5 s between incompatible signal groups. The maximum cycle time is set at 180 s. Three sets of various link capacity degradation disruption in terms of 7, 18 and 43 % are considered. For given signal settings, initial traffic flow on 23 links are provided in Table 2 together with corresponding expected performance measure defined in (3) in the presence of a worst-case of various levels of disruption of link capacity degradation. Computational results are summarized in Tables 3, 4, 5, 6, and 7. The resulting robust traffic flow on 23 links are summarized in Table 3 together with expected performance measure and improvement rates over initial signal settings in the presence of a worst-case of various sets of disruption of link capacity degradation. Also, the resulting consonant probabilities of occurrence of worst-case of link capacity degradation disruption are determined and summarized in Table 4 for robust Nash game introduced in (26) under various levels of disruption of link capacity degradation. Details for determining robust user equilibrium following computational steps given in Sect. 3.3 are respectively shown in Tables 5, 6, and 7 for various sets of link capacity degradation disruption. Moreover, in order to illustrate the effectiveness of modified projection method, numerical comparisons of expected performance measure are also made with conventional approach employed in [6, 10] such as the method of successive averages (MSA) when solving a robust user equilibrium flow in the presence of a worst-case of various sets of link capacity degradation disruption. Computational results are detailed in Tables 8, 9, and 10 for characterizing robust user equilibrium flow using MSA. Numerical comparisons are plotted in Figs. 2, 3, and 4 respectively for various sets of link capacity degradation disruption in the sense of gap of bound difference of expected performance measure defined in (40) and (42) between MSA and modified projection method. Implementations for carrying out the following computations were made on SUN SPARC SUNW, 900 MHZ processor with 4 GB RAM under Unix SunOS 5.8 using C++ compiler.

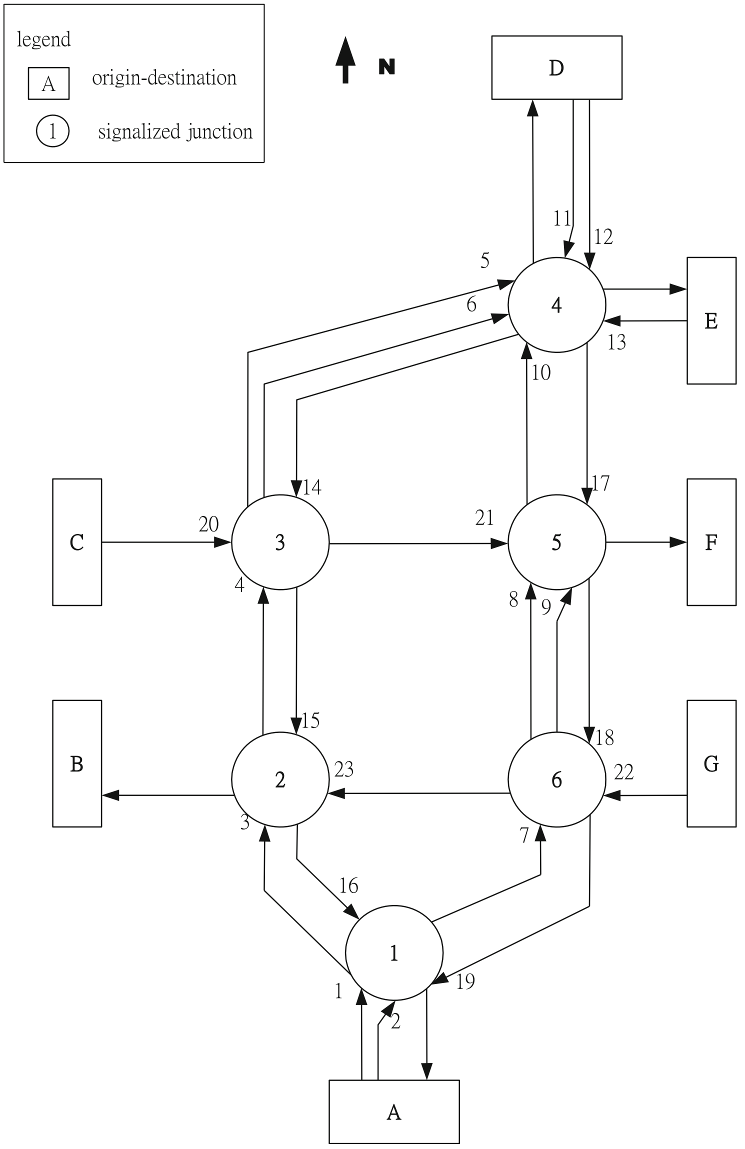


Fig. 1 Layout for Allsop and Charlesworth's network [2]

The stopping criterion is set when the relative difference in the objective function value is less than 0.15 %.

As it seen from Tables 3 and 4, in the presence of a worst-case of 7 % link capacity degradation disruption the robust traffic flow are centrally distributed on links 1, 3, 4, 5, 20, 21, 22 and 23 and moderately distributed on links 11, 14, 15, 16 and 19. The consonant probabilities of occurrence of a worst-case of 7 % link

Table 1 Base travel demands for Allsop and Charlesworth's road network [2] in veh/h

Origin/destination	A	B	D	E	F	Origin totals
A	–	250	700	30	200	1,180
C	40	20	200	130	900	1,290
D	400	250	–	50	100	800
E	300	130	30	–	20	480
G	550	450	170	60	20	1,250
Destination totals	1,290	1,100	1,100	270	1,240	5,000

Table 2 Initial traffic flow with various levels of link capacity degradation disruption

Link flow	7 %	18 %	43 %
f_1	950	950	950
f_2	230	230	230
f_3	950	950	950
f_4	840	840	840
f_5	880	880	880
f_6	150	150	150
f_7	230	230	230
f_8	250	250	250
f_9	90	90	90
f_{10}	260	260	260
f_{11}	460	460	460
f_{12}	290	290	290
f_{13}	450	450	450
f_{14}	740	740	740
f_{15}	790	790	790
f_{16}	610	610	610
f_{17}	460	460	460
f_{18}	350	350	350
f_{19}	680	680	680
f_{20}	1,290	1,290	1,290
f_{21}	1,050	1,050	1,050
f_{22}	1,250	1,250	1,250
f_{23}	810	810	810
PI (in $)	845	1,004	1,401

capacity degradation disruption are principally distributed at links 1, 3, 20, 21 and 22, which are mutually consistent with robust traffic flow distribution characterized by modified projection method. The vulnerability of signal-controlled road network can be identified by those links with moderate and heavy traffic flow in the presence of a worst-case of disruption of link capacity degradation. Moreover, in the presence of a worst-case of 18 % link capacity degradation disruption the robust traffic flow are centrally distributed rather on links 1, 3, 4, 5, 11, 14, 15, 16, 20, 21, 22 and 23 and moderately distributed on links 13 and 19. The consonant

Table 3 Robust equilibrium flow with various levels of link capacity degradation disruption

Link flow	7 %	18 %	43 %
f_1	1,075	1,087	1,061
f_2	105	93	119
f_3	1,075	1,087	1,061
f_4	935	940	902
f_5	974	971	926
f_6	148	153	158
f_7	105	93	119
f_8	157	155	195
f_9	88	86	83
f_{10}	168	167	206
f_{11}	461	597	562
f_{12}	290	153	188
f_{13}	450	450	450
f_{14}	794	968	942
f_{15}	842	995	970
f_{16}	661	821	948
f_{17}	406	232	258
f_{18}	298	146	170
f_{19}	629	470	342
f_{20}	1,290	1,290	1,290
f_{21}	1,055	1,080	1,080
f_{22}	1,250	1,250	1,250
f_{23}	778	779	918
PI (in $)	787	866	1,017
Improvement rates (%)	6.86	13.75	27.41

probabilities of occurrence of a worst-case of 18 % link capacity degradation disruption are rather distributed at links 1, 3, 14, 15, 20, 21 and 22, which are mutually consistent with robust traffic flow distribution characterized by modified projection method. Furthermore, in the presence of a worst-case of 43 % link capacity degradation disruption the traffic flow are diverted toward less congested links such as links 7, 8, 10 and 11 and the consonant probabilities of occurrence of a worst-case of link capacity degradation disruption are more evenly distributed at links 1, 3, 4, 11, 14, 15, 20, 21 and 22.

Computational efficiency of the modified projection method is detailed in Tables 5, 6, and 7 for characterizing robust user equilibrium flow in the presence of three sets of link capacity degradation disruption. As it indicated in Tables 5, 6, and 7, the modified projection method solves robust user equilibrium flow in the presence of various levels of disruption of link capacity degradation with success as the gap of bound difference of expected performance measure progressively decreases and gradually vanishes at robust traffic flow. For example, in the presence of a worst-case of 7 % link capacity degradation disruption, as it seen in Table 5, a robust user equilibrium flow is characterized using modified projection method

Table 4 Probabilities for robust equilibrium flow with various levels of link capacity degradation disruption

Link flow	7 %	18 %	43 %
p_1	0.13	0.09	0.09
p_2	0.00	0.00	0.03
p_3	0.11	0.07	0.12
p_4	0.06	0.04	0.07
p_5	0.05	0.04	0.05
p_6	0.01	0.01	0.00
p_7	0.00	0.01	0.00
p_8	0.00	0.01	0.00
p_9	0.00	0.00	0.00
p_{10}	0.00	0.00	0.00
p_{11}	0.02	0.03	0.05
p_{12}	0.00	0.01	0.00
p_{13}	0.02	0.04	0.05
p_{14}	0.04	0.06	0.07
p_{15}	0.05	0.08	0.08
p_{16}	0.04	0.04	0.06
p_{17}	0.02	0.02	0.00
p_{18}	0.00	0.03	0.00
p_{19}	0.06	0.04	0.00
p_{20}	0.13	0.12	0.11
p_{21}	0.10	0.11	0.09
p_{22}	0.12	0.10	0.07
p_{23}	0.04	0.05	0.06

Table 5 Performance measure under 7 % capacity degradation disruption using proposed method

(p,f)	$AP(p^{(k)},f^{(k)},\Delta s)$	$\overline{P}$	$\underline{P}$	Gap
$(p^{(0)},f^{(0)})$	845			
$(p^{(1)},f^{(0)})$	912	912		215
$(p^{(1)},f^{(1)})$	697		697	
$(p^{(2)},f^{(1)})$	875	875		153
$(p^{(2)},f^{(2)})$	722		722	
$(p^{(3)},f^{(2)})$	839	839		90
$(p^{(3)},f^{(3)})$	749		749	
$(p^{(4)},f^{(3)})$	798	798		27
$(p^{(4)},f^{(4)})$	771		771	
$(p^{(5)},f^{(4)})$	787	787		0
$(p^{(5)},f^{(5)})$	787		787	

following computation scheme given in Sect. 3.3. The expected performance measure with initial value of $845 is progressively degraded to the value of $787 with improvement rate of 6.86 %. The gap of bound difference of expected performance measure as defined in (40) and (42) is progressively decreasing and vanishes at which a robust user equilibrium flow is characterized. Numerical

Table 6 Performance measure under 18 % capacity degradation disruption using proposed method

(p,f)	$AP(p^{(k)},f^{(k)},\Delta s)$	$\overline{P}$	$\underline{P}$	Gap
$(p^{(0)},f^{(0)})$	1,004			
$(p^{(1)},f^{(0)})$	1,164	1,164		412
$(p^{(1)},f^{(1)})$	752		752	
$(p^{(2)},f^{(1)})$	1,098	1,098		317
$(p^{(2)},f^{(2)})$	781		781	
$(p^{(3)},f^{(2)})$	912	912		97
$(p^{(3)},f^{(3)})$	815		815	
$(p^{(4)},f^{(3)})$	875	875		31
$(p^{(4)},f^{(4)})$	844		844	
$(p^{(5)},f^{(4)})$	866	866		0
$(p^{(5)},f^{(5)})$	866		866	

Table 7 Performance measure under 43 % capacity degradation disruption using proposed method

(p,f)	$AP(p^{(k)},f^{(k)},\Delta s)$	$\overline{P}$	$\underline{P}$	Gap
$(p^{(0)},f^{(0)})$	1,401			
$(p^{(1)},f^{(0)})$	1,415	1,415		517
$(p^{(1)},f^{(1)})$	898		898	
$(p^{(2)},f^{(1)})$	1,370	1,370		429
$(p^{(2)},f^{(2)})$	941		941	
$(p^{(3)},f^{(2)})$	1,286	1,286		301
$(p^{(3)},f^{(3)})$	895		985	
$(p^{(4)},f^{(3)})$	1,083	1,083		89
$(p^{(4)},f^{(4)})$	994		994	
$(p^{(5)},f^{(4)})$	1,044	1,044		37
$(p^{(5)},f^{(5)})$	1,007		1,007	
$(p^{(6)},f^{(5)})$	1,017	1,017		0
$(p^{(6)},f^{(6)})$	1,017		1,017	

comparisons are also made with MSA in the sense of computational efficiency when solving a variational inequality (27) for characterizing robust user equilibrium flow. As it shown in Table 8, in the presence of a worst-case of 7 % link capacity degradation disruption, a robust user equilibrium flow is characterized using MSA instead following computation scheme given in Sect. 3.3. The expected performance measure with initial value of $845 is progressively improved after 9 iterations to the value of $787 and the gap of bound difference of expected performance measure is progressively decreasing and vanishes at which a robust user equilibrium flow is characterized.

Similarly, in the presence of a worst-case of 18 % link capacity degradation disruption, as it seen in Table 6, a robust user equilibrium flow is characterized using modified projection method following computation scheme given in Sect. 3.3. The expected performance measure with initial value of $1,004 is progressively degraded to the value of $866 with improvement rate of 13.75 %. The gap of bound

Table 8 Performance measure under 7 % capacity degradation disruption using MSA

(p,f)	$AP(p^{(k)},f^{(k)},\Delta s)$	$\overline{P}$	$\underline{P}$	Gap
$(p^{(0)},f^{(0)})$	845			
$(p^{(1)},f^{(0)})$	912	912		293
$(p^{(1)},f^{(1)})$	619		619	
$(p^{(2)},f^{(1)})$	875	875		240
$(p^{(2)},f^{(2)})$	635		635	
$(p^{(3)},f^{(2)})$	841	841		164
$(p^{(3)},f^{(3)})$	677		677	
$(p^{(4)},f^{(3)})$	822	822		120
$(p^{(4)},f^{(4)})$	702		702	
$(p^{(5)},f^{(4)})$	819	819		88
$(p^{(5)},f^{(5)})$	731		731	
$(p^{(6)},f^{(5)})$	803	803		56
$(p^{(6)},f^{(6)})$	747		747	
$(p^{(7)},f^{(6)})$	796	796		27
$(p^{(7)},f^{(7)})$	769		769	
$(p^{(8)},f^{(7)})$	789	789		9
$(p^{(8)},f^{(8)})$	780		780	
$(p^{(9)},f^{(8)})$	787	787		0
$(p^{(9)},f^{(9)})$	787		787	

difference of expected performance measure with initial value of \$412 is progressively decreasing and vanishes at which a robust user equilibrium flow is characterized. Numerical comparisons are also made with MSA in the sense of computational efficiency when solving a variational inequality (27) for characterizing robust user equilibrium flow. As it shown in Table 9, in the presence of a worst-case of 18 % link capacity degradation disruption, a robust user equilibrium flow is characterized using MSA instead following computation scheme given in Sect. 3.3. The expected performance measure with initial value of \$1,004 is progressively improved after 11 iterations to the value of \$866 and the gap of bound difference of expected performance measure with initial value of \$479 is progressively decreasing and vanishes at which a robust user equilibrium flow is characterized. Finally, in the presence of a worst-case of 43 % link capacity degradation disruption, as it seen in Table 7, a robust user equilibrium flow is characterized using modified projection method following computation scheme given in Sect. 3.3. The expected performance measure with initial value of \$1,401 is progressively degraded to the value of \$1,017 with improvement rate of 27.41 %. The gap of bound difference of expected performance measure with initial value of \$517 is progressively decreasing and vanishes at which a robust user equilibrium flow is characterized. Numerical comparisons are also made with MSA in the sense of computational efficiency when solving a variational inequality (27) for characterizing robust user equilibrium flow. As it shown in Table 10, in the presence of a worst-case of 43 % link capacity degradation disruption, a robust user equilibrium flow is characterized using MSA instead following computation

Table 9 Performance measure under 18 % capacity degradation disruption using MSA

(p,f)	$AP(p^{(k)},f^{(k)},\Delta s)$	$\overline{P}$	$\underline{P}$	Gap
$(p^{(0)},f^{(0)})$	1,004			
$(p^{(1)},f^{(0)})$	1,164	1,164		479
$(p^{(1)},f^{(1)})$	685		685	
$(p^{(2)},f^{(1)})$	1,139	1,139		436
$(p^{(2)},f^{(2)})$	703		703	
$(p^{(3)},f^{(2)})$	1,098	1,098		357
$(p^{(3)},f^{(3)})$	741		741	
$(p^{(4)},f^{(3)})$	1,043	1,043		274
$(p^{(4)},f^{(4)})$	769		769	
$(p^{(5)},f^{(4)})$	990	990		199
$(p^{(5)},f^{(5)})$	791		791	
$(p^{(6)},f^{(5)})$	945	945		140
$(p^{(6)},f^{(6)})$	805		805	
$(p^{(7)},f^{(6)})$	901	901		79
$(p^{(7)},f^{(7)})$	822		822	
$(p^{(8)},f^{(7)})$	879	879		30
$(p^{(8)},f^{(8)})$	849		849	
$(p^{(9)},f^{(8)})$	871	871		14
$(p^{(9)},f^{(9)})$	857		857	
$(p^{(10)},f^{(9)})$	869	869		6
$(p^{(10)},f^{(10)})$	863		863	
$(p^{(11)},f^{(10)})$	866	866		0
$(p^{(11)},f^{(11)})$	866		866	

scheme given in Sect. 3.3. The expected performance measure with initial value of \$1,401 is progressively improved after 14 iterations to the value of \$1,017 and the gap of bound difference of expected performance measure with initial value of \$710 is progressively decreasing and vanishes at which a robust user equilibrium flow is characterized. Numerical results for comparing the gap of bound difference of expected performance measure between MSA and modified projection method are also plotted in Figs. 2, 3, and 4 respectively for three sets of a worst-case of various levels of link capacity degradation disruption. As it seen from Figs. 2, 3, and 4, the modified projection method took less iteration to characterize a robust flow in the presence of a worst-case of various levels of link capacity degradation disruption than did MSA at which the gap of bound difference of expected performance measure vanishes.

Table 10 Performance measure under 43 % capacity degradation disruption using MSA

(p,f)	$AP(p^{(k)},f^{(k)},\Delta s)$	$\overline{P}$	$\underline{P}$	Gap
$(p^{(0)},f^{(0)})$	1,401			
$(p^{(1)},f^{(0)})$	1,415	1,415		710
$(p^{(1)},f^{(1)})$	705		705	
$(p^{(2)},f^{(1)})$	1,388	1,388		644
$(p^{(2)},f^{(2)})$	744		744	
$(p^{(3)},f^{(2)})$	1,349	1,349		558
$(p^{(3)},f^{(3)})$	791		791	
$(p^{(4)},f^{(3)})$	1,302	1,302		479
$(p^{(4)},f^{(4)})$	823		823	
$(p^{(5)},f^{(4)})$	1,276	1,276		422
$(p^{(5)},f^{(5)})$	854		854	
$(p^{(6)},f^{(5)})$	1,230	1,230		349
$(p^{(6)},f^{(6)})$	881		881	
$(p^{(7)},f^{(6)})$	1,187	1,187		269
$(p^{(7)},f^{(7)})$	918		918	
$(p^{(8)},f^{(7)})$	1,165	1,165		223
$(p^{(8)},f^{(8)})$	942		942	
$(p^{(9)},f^{(8)})$	1,140	1,140		173
$(p^{(9)},f^{(9)})$	967		967	
$(p^{(10)},f^{(9)})$	1,125	1,125		136
$(p^{(10)},f^{(10)})$	989		989	
$(p^{(11)},f^{(10)})$	1,092	1,092		87
$(p^{(11)},f^{(11)})$	1,005		1,005	
$(p^{(12)},f^{(11)})$	1,055	1,055		43
$(p^{(12)},f^{(12)})$	1,012		1,012	
$(p^{(13)},f^{(12)})$	1,034	1,034		19
$(p^{(13)},f^{(13)})$	1,015		1,015	
$(p^{(14)},f^{(13)})$	1,017	1,017		0
$(p^{(14)},f^{(14)})$	1,017		1,017	

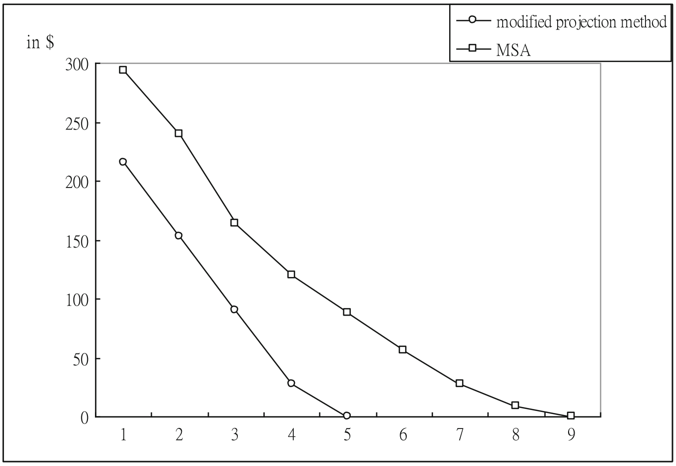


Fig. 2 Performance measure gap under 7 % capacity degradation disruption

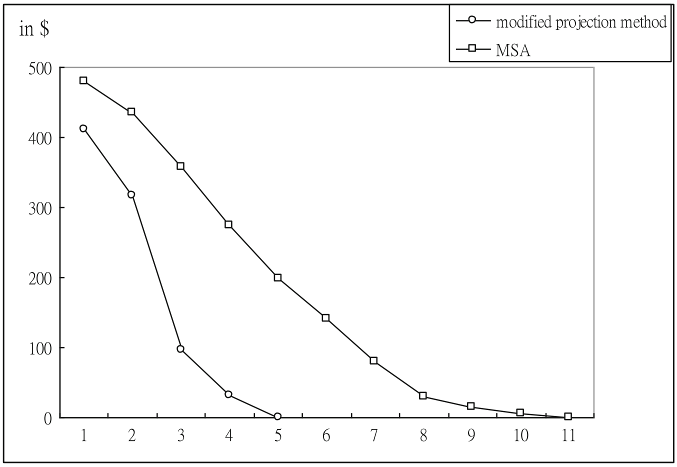


Fig. 3 Performance measure gap under 18 % capacity degradation disruption

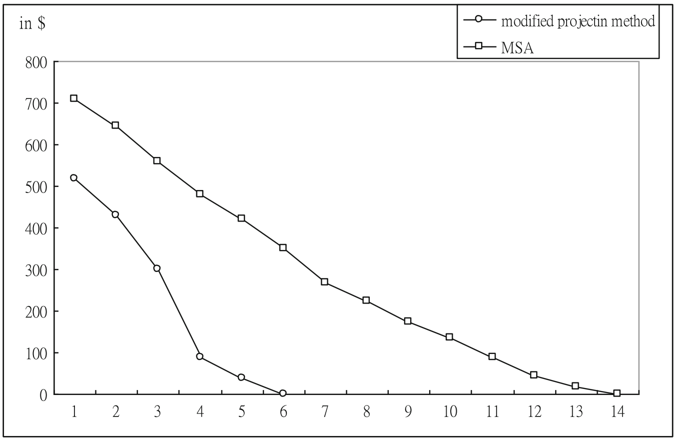


Fig. 4 Performance measure gap under 43 % capacity degradation disruption

5 Conclusion and Discussions

In this paper we presented a vulnerability analysis of signal-controlled road network in the presence of a worst-case of link capacity degradation for robust traffic flow following Wardrop's principle. A min-max mathematical model was proposed to formulate a robust user equilibrium flow problem and a modified projection method was presented to successfully solve a robust traffic flow problem in the presence of a worst-case of link capacity degradation disruption. Numerical computations were illustrated using a medium-sized road network for three sets of link capacity degradation disruption for proposed modified projection method in search for robust user equilibrium flow. Numerical comparisons were also made with conventional approach such as MSA in search for robust user equilibrium flow in the presence of a worst-case of various levels of link capacity degradation disruption. The results reported in this paper has demonstrated the feasibility and effectiveness of the proposed solution method when solving a robust user equilibrium flow in a signal-controlled road network under a worst-case of link capacity degradation disruption due to signal setting failure.

The method proposed in this paper can be extensively applied to a general road network design problem with continuous decisions such as link improvement program to mitigate potential consequence caused by uncertainty in travel demand. The present work on robust user equilibrium traffic flow can be considered a constraint for a robust area traffic control and a bilevel programming approach is attractive to deal with instances of robust network design problem. Extensions to the area of robust network design and traffic control in the presence of uncertainty will be another challenge facing researchers in the field of business analytics.

Acknowledgments The author is grateful to Editors for every aspect of arrangement about this book. The work reported in this paper has been financially supported via grants 98-2410-H-259-009-MY3 & 101-2628-H-259-001-MY2 from Taiwan National Science Council.

References

1. Aashtiani HZ, Magnanti TL (1981) Equilibria on a congested transportation network. SIAM J Algebr Discrete Meth 3(2):213–226
2. Allsop RE, Charlesworth JA (1977) Traffic in a signal-controlled road network: an example of different signal timings inducing different routeings. Traffic Eng Contr 18:262–264
3. Akgun I, Tansel B, Wood K (2011) The multi-terminal maximum-flow network-interdiction problem. Eur J Oper Res 211:241–251
4. Asakura Y (1999) Evaluation of network reliability using stochastic user equilibrium. J Adv Transp 33(2):147–158
5. Azaiez N, Bier VM (2007) Optimal resource allocation for security in reliability systems. Eur J Oper Res 181:773–786
6. Bell MGH (2003) The use of game theory to measure the vulnerability of stochastic networks. IEEE Trans Reliab 52(1):63–68
7. Bell MGH (2000) A game theory approach to measuring the performance reliability of transport networks. Transp Res B 34(6):533–546
8. Bell MGH, Cassir C (2002) Risk-averse user equilibrium traffic assignment: an application of game theory. Transp Res B 36(8):671–681
9. Bell MGH, Iida Y (1997) Transportation network analysis. Wiley, Chichester
10. Bell MGH, Kanturska U, Schmocker J-D, Fonzone A (2008) Attacker–defender models and road network vulnerability. Phil Trans R Soc A 366(1872):1893–1906
11. Berdica K (2002) An introduction to road vulnerability: what has been done, is done and should be done. Transp Policy 9:117–127
12. Bier VM, Hausken K (2013) Defending and attacking a network of two arcs subject to traffic congestion. Reliab Eng Syst Saf 112:214–224
13. Bier VM, Nagaraj A, Abhichandani V (2005) Protection of simple series and parallel systems with components of different values. Reliab Eng Syst Saf 87:315–323
14. Brown G, Carlyle M, Salmeron J, Wood K (2006) Defending critical infrastructure. Interfaces 36:530–544
15. Cappanera P, Scaparra MP (2011) Optimal allocation of protective resources in shortest-path networks. Transp Sci 45(1):64–80
16. Chen A, Yang C, Kongsomsaksakul S, Lee M (2007) Network-based accessibility measures for vulnerability analysis of degradable transportation networks. Netw Spat Econ 7:241–256
17. Chen A, Yang H, Lo HK, Tang WH (2002) Capacity reliability of a road network: an assessment methodology and numerical results. Transp Res B 36(3):225–252
18. Chiou S-W (2003) TRANSYT derivatives for area traffic control optimisation with network equilibrium flows. Transp Res B 37(3):263–290
19. Church RL, Scaparra MP (2007) Analyzis of facility systems' reliability when subject to attack or a natural disaster. In: Murray AT, Grubesic TH (eds) Critical infrastructure: reliability and vulnerability, Advances in spatial science. Springer, Berlin, pp 221–242
20. Church RL, Scaparra MP, Middleton RS (2004) Identifying critical infrastructure: the median and covering facility interdiction problems. Ann Assoc Am Geogr 94(3):491–502
21. Cormican KJ, Morton DP, Wood RK (1998) Stochastic network interdiction. Oper Res 46(2): 184–197
22. Grubesic TH, Matisziw TC, Murray AT, Snediker D (2008) Comparative approaches for assessing network vulnerability. Int Reg Sci Rev 31(1):88–112

23. Hausken K (2008) Strategic defense and attack for reliability Systems. Reliab Eng Syst Saf 93 (11):1740–1750
24. Hausken K, Levitin G (2009) Minmax defense strategy for complex multi-state systems. Reliab Eng Syst Saf 94:577–587
25. Israeli E, Wood K (2002) Shortest path network interdiction. Networks 40:97–111
26. Jenelius E, Mattsson L (2012) Road network vulnerability analysis of area-covering disruptions: a grid-based approach with case study. Transp Res A 46:746–760
27. Jenelius E, Petersen T, Mattsson L (2006) Importance and exposure in road network vulnerability analysis. Transp Res A 40:537–560
28. Lam WHK, Chan KS, Li Z-C, Bell MGH (2010) A risk-averse user equilibrium model for route choice problem in signal-controlled networks. J Adv Transp 44:219–230
29. Lam WHK, Shao H, Sumalee A (2008) Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. Transp Res B 42:890–910
30. Levitin G (2004) Maximizing survivability of vulnerable weighted voting systems. Reliab Eng Syst Saf 83:17–26
31. Levitin G (2007) Optimal defense strategy against intentional attacks. IEEE Trans Reliab 56: 148–156
32. Levitin G (2009) Redundancy vs. protection vs. false targets for systems under attack. Reliability. IEEE Trans Reliab 58(1):58–68
33. Levitin G, Ben-Haim H (2008) Importance of protections against intentional attacks. Reliab Eng Syst Saf 93:639–646
34. Levitin G, Gertsbakh I, Shpungin Y (2011) Evaluating the damage associated with intentional network disintegration. Reliab Eng Syst Saf 96:433–439
35. Levitin G, Hausken K (2013) Is it wise to leave some false targets unprotected? Reliab Eng Syst Saf 112:176–186
36. Levitin G, Hausken K, Ben-Haim H (2013) Defending majority voting systems against a strategic attacker. Reliab Eng Syst Saf 111:37–44
37. Levitin G, Lisnianski A (2003) Optimizing survivability of vulnerable series-parallel multi-state systems. Reliab Eng Syst Saf 79:319–331
38. Liberatore F, Scaparra MP, Daskin MS (2011) Analyzis of facility protection strategies against an uncertain number of attacks: the stochastic R-interdiction median problem with fortification. Comput Oper Res 38:357–366
39. Lim C, Smith J (2007) Algorithms for discrete and continuous multicommodity flow network interdiction problems. IIE Trans 39((Special Issue on Homeland Security)):15–26
40. Lins ID, Rego LC, Moura MC, Droguett EL (2013) Selection of security system design via games of imperfect information and multi-objective genetic algorithm. Reliab Eng Syst Saf 112:59–66
41. Lo HK, Luo XW, Siu BWY (2006) Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. Transp Res B 40(9):792–806
42. Lo HK, Tung YK (2003) Network with degradable links: capacity analysis and design. Transp Res B 37(4):345–363
43. Losada C, Scaparra MP, O'Hanley JR (2012) Optimizing system resilience: a facility protection model with recovery time. Eur J Oper Res 217:519–530
44. Nicholson A, Du Z-P (1997) Degradable transportation systems: an integrated equilibrium model. Transp Res B 31(3):209–223
45. O'hanley JR, Church RL (2011) Designing robust coverage networks to hedge against worst-case facility losses. Eur J Oper Res 209:23–36
46. Ramirez-Marquez JE, Rocco C (2009) Stochastic network interdiction optimization via capacitated network reliability modeling and probabilistic solution discovery. Reliab Eng Syst Saf 94:913–921
47. Ramirez-Marquez JE, Rocco C, Levitin G (2009) Optimal protection of general source–sink networks via evolutionary techniques. Reliab Eng Syst Saf 94(10):1676–1684

48. Ramirez-Marquez JE, Rocco C, Levitin G (2011) Optimal network protection against diverse interdictor strategies. Reliabil Eng Syst Saf 96(3):374–382
49. Rocco CM, Ramirez-Marquez JE (2009) Deterministic network interdiction optimization via an evolutionary approach. Reliab Eng Syst Saf 94:568–576
50. Rocco CM, Ramirez-Marquez JE (2010) A bi-objective approach for shortest-path network interdiction. Reliab Eng Syst Saf 59:232–240
51. Rocco CM, Ramirez-Marquez JE, Salazar DE (2010) Bi and tri-objective optimization in the deterministic network interdiction problem. Reliab Eng Syst Saf 95(8):887–896
52. Royset JO, Wood RK (2007) Solving the bi-objective maximum-flow network-interdiction problem. INFORMS J Comput 19(2):175–184
53. Salmeron J, Wood K, Baldick R (2004) Analyzis of electric grid security under terrorist threat. IEEE Trans Power Syst 19:905–912
54. Scaparra MP, Church RL (2008) A Bilevel mixed-integer program for critical infrastructure protection planning. Comput Oper Res 35:1905–1923
55. Sullivan JL, Novak DC, Aultman-Hall L, Scott DM (2010) Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: a link-based capacity-reduction approach. Transp Res A 44:323–336
56. Sumalee A, Kurauchi F (2006) Network capacity reliability analysis considering traffic regulation after a major disaster. Netw Spat Econ 6(3–4):205–219
57. Szeto WY, O'Brien L, O'Mahony M (2006) Risk-averse traffic assignment with elastic demands: NCP formulation and solution method for assessing performance reliability. Netw Spat Econ 6(3):313–332
58. Szeto WY (2011) Cooperative game approaches to measuring network reliability considering paradoxes. Transp Res C 19:229–241
59. Vincent RA, Mitchell AI, Robertson DI (1980) User guide to TRANSYT. TRRL report, LR888. Transport and Road Research Laboratory, Crowthorne
60. Wardrop JG (1952) Some theoretical aspects of road traffic approach. Proc Inst Civil Eng II 1(325–378):1952
61. Wollmer R (1964) Removing arcs from a network. J Oper Res Soc Am 12:934–940
62. Wong SC (1995) Derivatives of the performance index for the traffic model from TRANSYT. Transp Res B 29(5):303–327
63. Wood K (1993) Deterministic network interdiction. Math Comput Model 17:1–18

Analytical Model Implementing Objectives in EVM for Advanced Project Control

José Ignacio Muñoz-Hernández, José Ramón Otegi-Olaso, and Julen Rubio-Gómez

1 Introduction

In such a global industrial market as the current one, competitiveness is a key driver for success. This competitiveness requires the application of the latest technological advances in several fields as, for example, market analysis, manufacturing, management and new information systems. Science and technology go forward generating knowledge by means of the formulation of analytical models that could explain the real behavior of a system and could forecast its future behavior.

Project management is one of the Science disciplines which have experienced a bigger development generating methods and models to increase the efficiency of the definition and launching into the market of products and services, making project management indispensable for business results.

Aligned with this line of thinking, the top technology industries as, for example, aerospace and aeronautics, and government agencies such as NASA, have been exponentially empowering the project management discipline and have been generating guidelines and methods for its development. In parallel, international associations of project managers such as the Project Management Institute (PMI) or the International Project Management Association (IPMA) have been created, and are developing project management competencies interacting with thousands of practitioners and developing relationships with corporations, government

J.I. Muñoz-Hernández (✉)
University of Castilla-La Mancha (INEI-UCLM), Ciudad Real, Spain
e-mail: JoseIgnacio.Munoz@uclm.es

J.R. Otegi-Olaso
University of Basque Country, Bilbao, Spain
e-mail: joserra.otegi@ehu.es

J. Rubio-Gómez
Dresser-Rand Inc., Milano, Spain
e-mail: jrubio@dresser-rand.com

© Springer International Publishing Switzerland 2015
F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_7

agencies, universities and colleges, as well as training organizations and consulting companies.

In project management, project control has been developed to define the most efficient indexes showing the current project status and predicting the future scenario. Project control has experienced a quantum leap forward with the irruption of the analytic method of Earned Value Management.

The EVM is one of the project management techniques more used in the world nowadays. In 1965, the United States Air force acquisition managers defined 35 criteria which capture the essence of earned value management. In 1998, it was termed as the ANSI/EIA-748 Standard, and afterwards it has been developed exponentially. Advanced progressing upon EVM standards last research lines concepts have included objectives accomplishment measurement or the probabilistic analysis. Also in recent years, the business value is clearly oriented to the objectives accomplishment; correct definition, control and assessment of the objectives could mean the difference between the project success or failure. As it is explained in the technical universities, a project consisting in developing a new smartphone can be finished on schedule and cost according to the plan but if when the client has the phone on his hands the battery only lasts 30 s, probably the technical objectives are not fulfilled and the product acceptance in the market cannot be the expected!

The contribution of this work is to present a formulation for integrating the project objectives follow up using EVM and giving the managers a practical tool to control the projects in all aspects.

2 Enhancing Business Value with Advanced Project Management

2.1 Advanced Earned Value Management

EVM is a project management powerful technique used to measure and communicate the real progress of a project integrating the scope, schedule and cost, in monetary terms. EVM is a tool used by program managers to: (1) quantify and measure program/contract performance, (2) provide an early warning system for deviation from a baseline, (3) mitigate risks associated with cost and schedule overruns, (4) provide a means to forecast final cost and schedule outcomes. The EVM is well explained in literature, and it is possible to have a very good approach in the references [1].

Historically, it started by industrial engineers on the factory floor in the early 1900s [2], who for years have employed a three-dimensional approach to assess true "cost-performance" efficiencies. To assess their cost performance, they compared their earned standards (the physical factory output) against the actual cost incurred. Then, they compared their earned standards to the original planned standards (the physical work they planned to accomplish) to assess the schedule results. These efforts provided earned value in its most basic form. Most important, the industrial engineers defined a cost variance as the difference between the actual costs spent and the earned standards in the factory. This definition of a cost variance is perhaps the indication to determine whether one uses the earned-value concept.

In 1965 the United States Air force acquisition managers defined 35 criteria which capture the essence of earned value management. Two years later the U.S. Department of Defense (DoD) adopted these same criteria as part of their Cost/Schedule Control Systems Criteria (C/SCSC). Then, in 1996, after a rewrite of the C/SCSC 35 criteria by private industry, the DoD accepted the rewording of this criteria under a new title called Earned Value Management System (EVMS), and the total number of criteria was reduced to 32.

In 1998, National Defense Industrial Association (NDIA) obtained acceptance of the Earned Value Management System in the form of the American National Standards Institute, termed the ANSI/EIA-748 Standard, see [3].

In 2000, The Project Management Institute's (PMI's) A Guide to the Project Management Body of Knowledge (PMBOK Guide) [1] provided the basic terminology and formulas of EVM, and more details were provided in subsequent editions of the PMBOK Guide.

The private sector and the academic investigation have also shown great interest in applying EVM in recent years thanks to numerous publications promoting EVM principles and advanced project management software packages that incorporate EVM methods and analysis [4–6].

The development of the EVM in the last 10 years leads to an enhancement of the EVM from the standard defined in the ANSI/EIA-748 [21, 22]. Nowadays, the main EVM research lines are the following: the reliability assessment of the EVM indexes and forecasting methods of project final duration and cost [7–9, 23, 24]; the EVM extension called Earned Schedule [10, 11]; the consideration of project risks [12, 13]; the integration of the technical performance in the EVM [14–17]; and the use of fuzzy techniques and analytical curves for the EV determination [18, 25].

In order to explain the approach of EVM implementing objectives, first, EVM fundamentals overview will be given. The EVM is based on three basic metrics, Planned Value (PV), Actual Cost (AC) and Earned Value (EV) from which performance indexes and forecasting formulas are constructed.

The EVM variances are shown in Eqs. (1) and (2).

$$\mathbf{CV} = \mathbf{EV} - \mathbf{AC} \tag{1}$$

$$\mathbf{SV} = \mathbf{EV} - \mathbf{PV} \tag{2}$$

The EVM performance indexes are shown in Eqs. (3) and (4).

$$\mathbf{CPI} = \mathbf{EV}/\mathbf{AC} \tag{3}$$

$$\mathbf{SPI} = \mathbf{EV}/\mathbf{PV} \tag{4}$$

The EVM forecasting formulas provide calculations for the Cost Estimate at Completion (EAC) and the Time Estimate at completion (TEAC), based on three different assumptions:

– The assumptions underlying the original estimation are wrong, or no longer applicable due to changed conditions affecting the activity, work package, or project, a new Cost Estimate To Complete (ETC) and Time Estimate To Complete (TETC) need to be developed.

$$\mathbf{EAC} = \mathbf{AC} + \mathbf{ETC} \tag{5}$$

$$\mathbf{TEAC} = \mathbf{AT} + \mathbf{TETC} \tag{6}$$

– Past performance is not a good predictor of future performance, which problems or opportunities which affected performance in the past will not occur in the future, and that future performance will parallel the original plan. Bugdet At Completion (BAC), Baseline Schedule At Completion (SAC).

$$\mathbf{EAC} = \mathbf{AC} + \mathbf{BAC} - \mathbf{EV} \tag{7}$$

$$\mathbf{TEAC} = \mathbf{SAC} - \mathbf{TV} \tag{8}$$

– Past performance is a good predictor of future performance, that performance to date will continue into the future.

$$\mathbf{EAC} = \mathbf{AC} + \mathbf{BAC} - \mathbf{EV}/\mathbf{CPI} \tag{9}$$

$$\mathbf{TEAC} = \mathbf{SAC}/\mathbf{SPI} \tag{10}$$

The EVM basic parameters can be shown in Fig. 1.

According to the recent research lines, in this work an "Advanced EVM" is considered for project management which includes the objectives control to evaluate the risks and different project scenarios.

2.2 Project Driver: Objectives Accomplishment

Project management should be applied to every project where the owners of the final product wish to ensure that the expended resources are used efficiently. On major projects the application of good project management tools will aid in the selection of the right course when managers need to make financial and time allocation decisions.

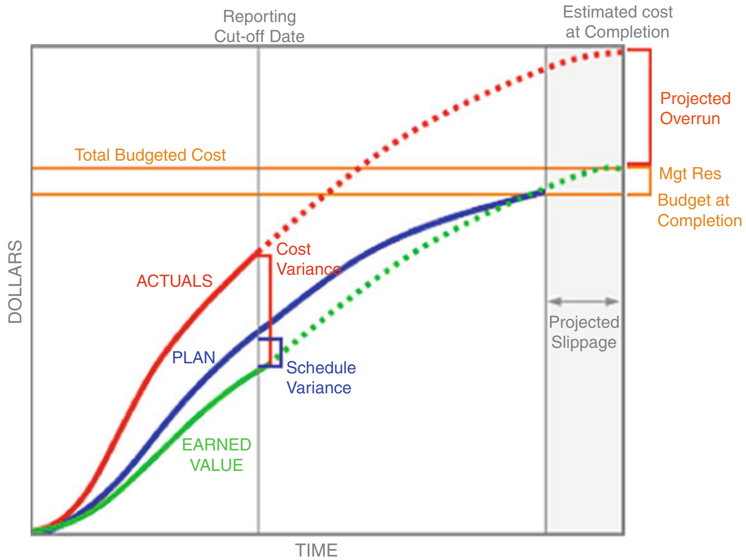


Fig. 1 EVM key parameters. Source: http://evm.nasa.gov/images/key-data.gif

According to PMBOK [1], a project is a temporary endeavor undertaken to create a product, service or result. The temporary nature of projects indicates a definite beginning and an end. The end is reached when the project's objectives are achieved or when the project is terminated because its objectives will not be met or cannot be met, or when there is no longer the need that gave rise to the project.

Business and product development are objectives oriented. One of the drivers of project control must be the accomplishment of objectives. In Fig. 2 the typical product development flow chart is presented and it can be seen that in project management for the launching of new products, it is important to highlight the fulfilling of technical objectives.

At the beginning of a project, it is usual to create a document called Statement of Work (SOW) where the tasks to be done for the achievement of the targets of the project are defined. From the SOW document on, the specifications for all the subsystems of the product or service of the project are defined.

Therefore, there is a definition of the project objectives in a top-down structure from the top management level to the task breakdown structure at the final level. And those objectives are summarized in the project documentation; this is, in the genesis of the project.

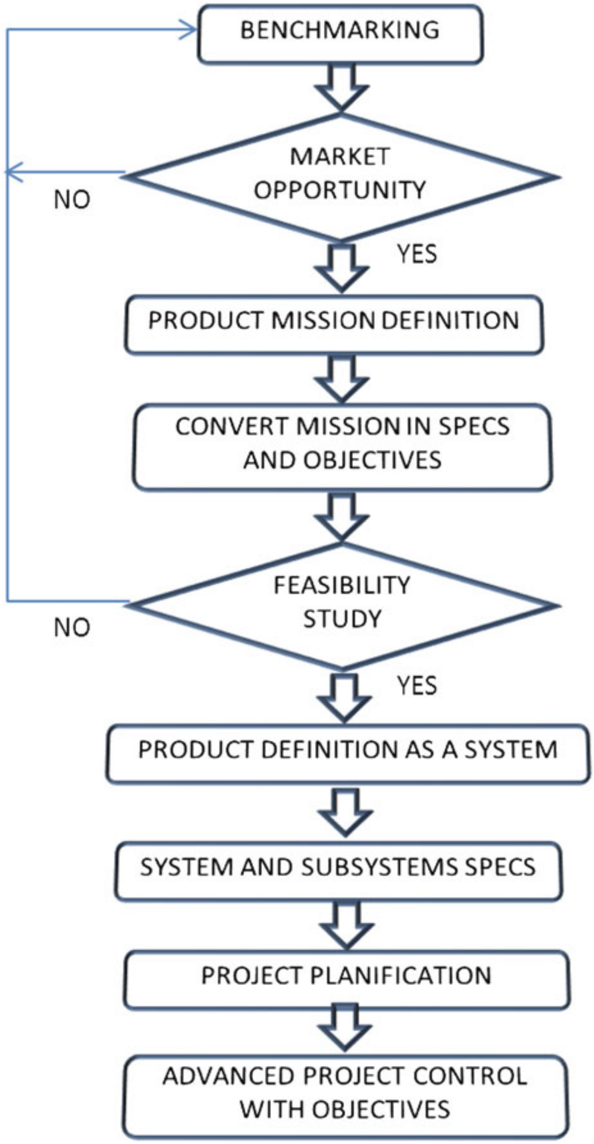


Fig. 2 Product development process oriented to objectives

3 An Analytical Model Implementing Objectives in EVM

In the projects control with EVM, more accurate analytical models are being developed including new control parameters in order to obtain more reliable information about the current and future situation of the project.

In this section, an analytical model implementing the accomplishment of the project objectives in the earned value management system is presented. It can also provide output about forecasting of objectives deviations impact at the project completion.

This formulation is a new contribution to develop the EVM especially useful for product development projects where the objectives related with product features are essential for success.

The model is based on two main concepts. The first one is the evaluation of the objectives accomplishment weighted by the objectives criticism, and the second one, the objectives deviations impact estimation at the end of the project.

3.1 Evaluation of Project Objectives Accomplishment

As it is commented in the previous section, one of the keys to decide about the project success is the objectives accomplishment.

Normally, in product development projects, objectives show different levels of criticism between them. This means that, there are objectives which are a must in the product and whose non-fulfillment implies that the product could be rejected for market launching. Mean-while, there are other objectives that are preferable and searched for but there is a bigger margin for their acceptance.

The weighting of objectives on the light of their criticism is done by the chief engineer, the product technology responsibles and/or the project managers. And this fact is usually the *BIG-BANG* of the project definition phase.

3.2 Impact of Initial Phases in the Objectives Accomplishment

Product development projects are structured following well established phases based on experience for the correct product definition. Typically these are the following ones: *conceptual design, detail design, prototypes manufacturing, testing and series launching*. These phases might differ from one project to another, be subdivided or be a part of other phases, but they commonly appear in the practice of product development projects.

In each phase, the objectives accomplishment deviation impact is different, and specifically, the impact is bigger in initial phases and it decreases as long as the last phases are reached. This means, that one deviation in the product definition at the concept phase produces a multiplier effect in the project end which is difficult to rectify.

For example, an error in the concept design phase while defining the number of engines needed to propel a commercial aircraft could lead to the project cancelation or a complete project re-thinking if it is detected in testing phase that this number of engines is not correct. On the other hand, an error in the final phase of serial launching, as a wrong dimensioning of engines production line could be corrected more easily and it implies less risk for the project.

This idea, that gives more weight to the initial phases, is captured in the model presented in this work, remarking the impact of the initial deviations in the objectives at the end of the project.

3.3 Analytical Model Description

The analytical model for implementing the project objectives in the EVM is built in for steps.

First, a weight to each project objective in function of its criticism is assigned.

Second, in each project phase, each objective accomplishment is evaluated and the weighted average of objectives accomplishment is calculated.

Third, an impact factor is assigned to each project phase in function of the effect that a deviation of the objectives in that phase has at the end of the project. Earlier the project phases bigger the impact factor. In this model the inverse of project progress is considered as the impact factor, so the impact factor can be calculated as the Budget at Completion (BAC) divided by the Planned Value (PV).

Four, the objectives weighted average powered to the inverse of project progress is applied to the EV, in order to implement both the objectives accomplishment and the initial phases impact into the EV.

The earned value implementing the objectives control and the impact of the initial phases is defined in this work as the Earned Value and Objectives (EVO).

Figure 3 shows the four steps to obtain the EVO from the project data.

Thus, the EV is minored by the objectives if they are not all fulfilled, and the model puts on the table the fact that it is not enough with the planned task completion, but also it is checked that the tasks have led to product specifications accomplishment.

The EV formulation implementing project objectives is formulated by Eq. (11).

$$\mathbf{EVO} = \left(\frac{\sum_{\mathbf{i}}^{\mathbf{n}} \mathbf{Xi} * \mathbf{Wi}}{\sum_{\mathbf{i}}^{\mathbf{n}} \mathbf{Wi}} \right)^{\mathbf{IF}} * \mathbf{EVi} \tag{11}$$

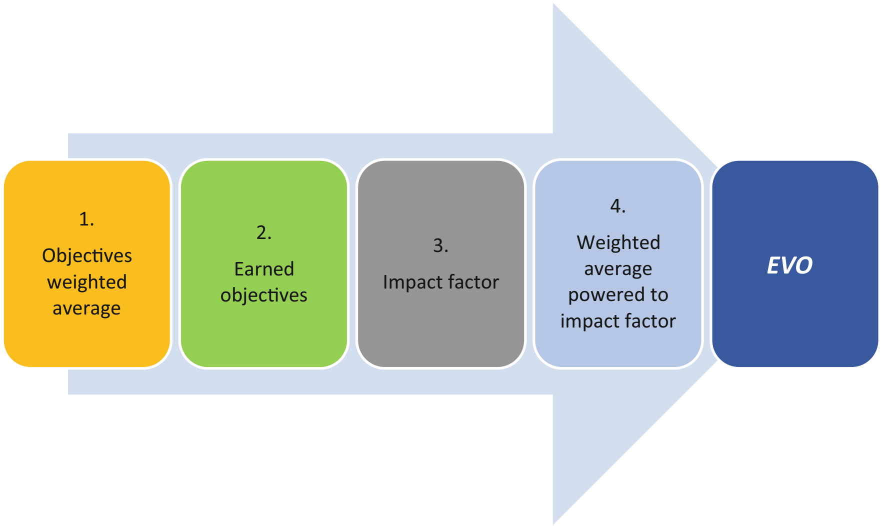


Fig. 3 Four steps for building the analytical model

Where,

EVO = Earned Value and Objectives
EV = Earned Value in each project phase
Xi = objective achievement in each project phase
Wi = objective weight
IF = phases impact factor, calculated as the inverse of project progress
IF = BAC/PVi
n = number of project phases

Besides, the model presented in this work incorporates a formulation to make a forecasting of the Earned Value and Objectives (EVO) at the project completion.

As it was commented in Sect. 2 about the EVM overview, there could be different forecasting formulas depending on the considerations made about the future project progress.

In this model, it is considered that past performance is a good predictor of future performance. Thus, the Earned Value and Objectives At Completion (EVOAC) is formulated by Eq. (12):

$$\mathbf{EVOAC} = \mathbf{EVO} + (\mathbf{BAC} - \mathbf{EVO})/\mathbf{SPIevo} \tag{12}$$

Where,

EVOAC = Earned Value and Objectives At Completion
EVO = Earned Value and Objectives
BAC = Budget At Completion
SPIevo = Schedule Performance Index with EVO. Calculated by Eq. (13).

$$\mathbf{SPIevo} = \mathbf{EVO/PV} \tag{13}$$

At project completion, it is necessary to earn all the planned value; however, the EV reduction when the objectives fulfillment is implemented implies that more actions have to be taken to successfully terminate the project.

3.4 *Example of Application of the Model*

In this section, an example of the analytical model implementing the objectives in the EV is shown.

This example is based on the general features of a typical commercial aircraft development project. The project has five phases and seven objectives. It is considered that the project progress follows a PERT function.

The PERT function can be used to model events which are constrained to take place within an interval defined by a minimum and maximum value. For this reason, the Pert function is used extensively in PERT, Critical Path Method (CPM), Joint Cost Schedule Modeling (JCSM) and other project management/control systems [19] to describe the time to completion and the cost of a task.

The PERT function is a family of continuous functions defined on the interval [0, 1] parameterized by two positive shape parameters, denoted by α and β, that appear as exponents of the random variable and control the shape of the function.

In project management, shorthand computations are widely used to estimate the mean and standard deviation of the beta distribution like the ones collected in Eqs. (14) and (15).

$$\mu(\mathbf{x}) = \frac{\mathbf{a} + 4\mathbf{b} + \mathbf{c}}{\mathbf{c}} \tag{14}$$

$$\boldsymbol{\sigma}(\mathbf{x}) = \frac{\mathbf{c} - \mathbf{a}}{6} \tag{15}$$

where a is the minimum, c is the maximum, and b is the most likely value (the mode for $\alpha > 1$ and $\beta > 1$).

The above estimate for the mean μ(x) is known as the PERT three-point estimation and it is exact for either of the following values of β (for arbitrary α within these ranges):

- $\beta = \alpha > 1$
- $\beta = 6 - \alpha$ for $5 > \alpha > 1$

The PERT function used in the model has $\beta = \alpha = 3$, and it is shown in the following graph.

Figure 4 shows the PERT function and then the PERT cumulative function modeling the typical project progress behavior with smooth begin and end. Also, in Fig. 4, the Impact Factor function is calculated as the ratio of the Budget At

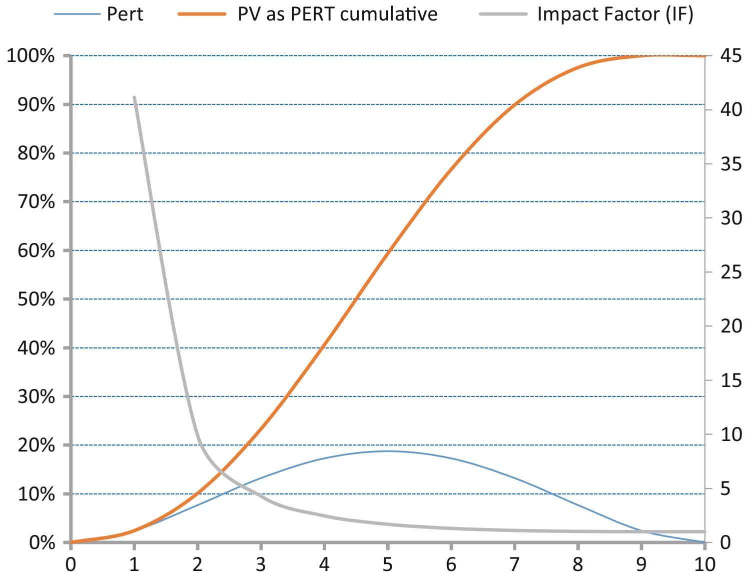


Fig. 4 PERT, PERT cumulative and impact factor functions

Table 1 Model application example

EVM parameters		
Phase	PV	EV
0	0,00	0,00
1	0,10	0,07
2	0,41	0,30
3	0,77	0,63
4	0,97	0,85
5	1,00	0,90

PV and EV values

Completion and the Planned Value, where the Planned Value is the PERT cumulative function. The first point of the Impact Factor function tends to infinite as the first point of the Planned Value tends to zero, thus, it is a singularity not considered. This fact also happens in the real projects where in the first point there is no significant planned value.

Once project data is known, it is possible to calculate the EVM parameters. The Planned Value is the project progress modeled with the PERT cumulative function. For EV calculation certain task completion in each project phase is considered. In Table 1 and Fig. 5, the PV and EV values for this example are shown.

The goal of this method is the objectives accomplishment. For this to happen monitoring of objectives is fundamental. Thus, the next step is to evaluate the accomplishment of objectives at every phase. Then, the weighted average is calculated (Table 2).

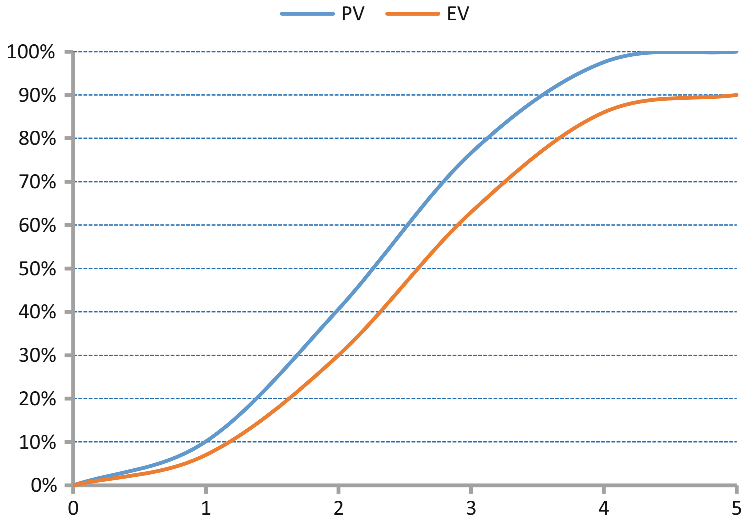


Fig. 5 Model application example. PV and EV curves

Table 2 Model application example

Objective	Weight	Project phases					
		0	1	2	3	4	5
		Earned	Earned	Earned	Earned	Earned	Earned
Thrust	0,3	0	0,95	0,90	0,90	0,85	0,85
Compresion	0,1	0	1,00	1,00	1,00	1,00	1,00
Passengers	0,05	0	1,00	1,00	0,90	0,90	0,90
Performance	0,2	0	1,00	1,00	0,90	0,95	0,97
Weigth	0,2	0	1,00	1,00	1,00	1,00	1,00
Consumption	0,05	0	0,95	0,90	0,90	0,85	0,85
Autonomy	0,1	0	0,90	0,85	0,95	0,95	0,98
Weighted average		0	0,973	0,950	0,935	0,928	0,936

Objectives weighted average

Earned Value and Objectives (EVO) is the calculated. This earned value implementing objectives takes into account not only the work finished, but also the degree of Project objectives accomplishment and the impact of the initial phases.

PV, EV and EVO values may be observed in Table 3 and Fig. 6.

Besides implementing objectives in the EVM, as previously commented, a forecasting of the impact of the deviation of the objectives at the end of the project may be done. This impact is calculated by the power of the weighted average to the inverse of project progress, i.e., the Impact Factor. This Impact Factor considers that a deviation in the initial phases has a larger effect at the end of the project than a deviation in the final phases.

Table 3 EVO calculation

PV	EV	Objectives weighted average	Impact factor	EVO
0,000	0,00	0,000	–	–
0,101	0,07	0,973	9,891	0,053
0,406	0,30	0,950	2,461	0,264
0,766	0,63	0,935	1,305	0,577
0,976	0,86	0,928	1,025	0,796
1,000	0,90	0,936	1,000	0,842

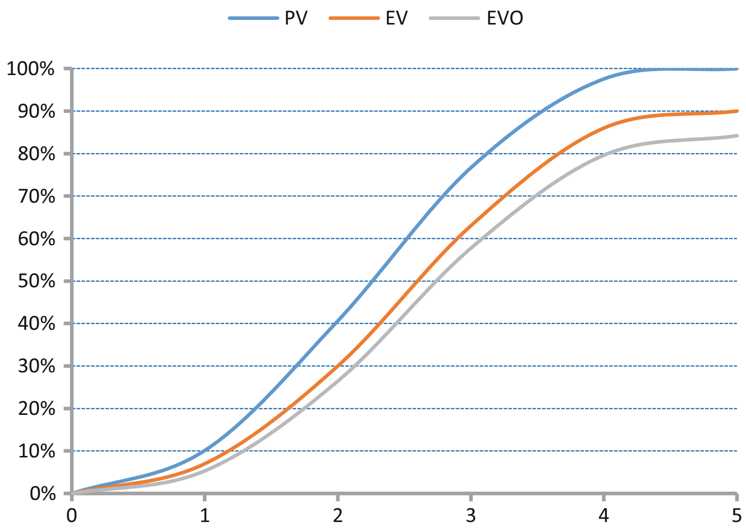


Fig. 6 EVO representation against PV and EV

Figure 7 shows EVO and EVOAC linked by lines which express the projection of the EVO at the end of the project to obtain the EVOAC.

4 Case Study

The analytical model has been applied to a real engineering product development project where the drivers are the technical objectives for the operating of a combustion engine for power generation.

Project began in January 2004 and was delayed 1 year by material procurement problems. But once new deadlines were negotiated with suppliers, it was launched again in 2005 and an overall tracking of the technical targets, cost and schedule began. In this project, the engineering consultancy delivered a design that was slightly below the technical targets of engine performance and efficiency, which was penalized in their fees. The engine efficiency target was 47 % but an engine

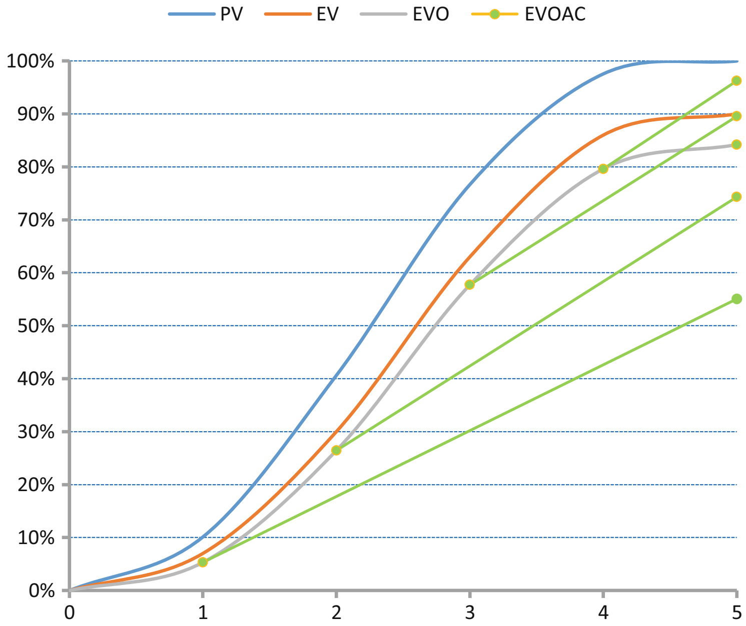


Fig. 7 EVOAC representation

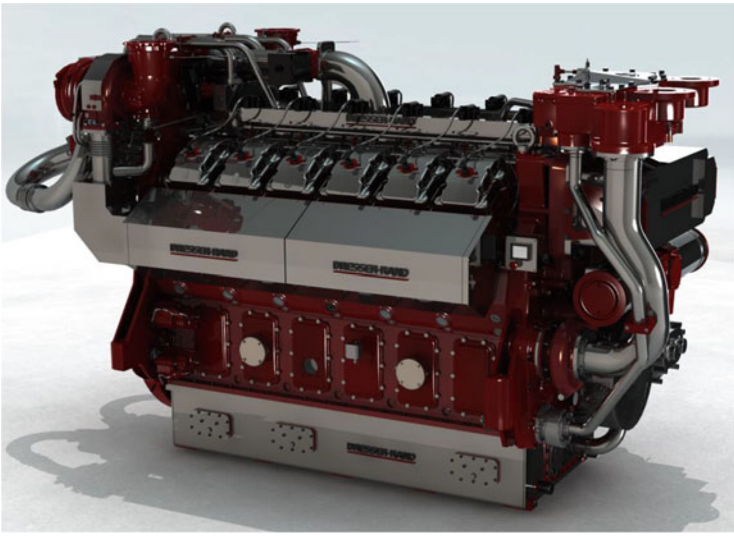


Fig. 8 V12 combustion engine. Dresser-Rand Inc. courtesy

with 45 % efficiency was obtained. Finally, the project was completed in January 2008, bringing to the market the new engine with a very good acceptance and overall rating of satisfactory (Fig. 8).

Table 4 Project objectives

Objectives	Weight	Values	Max	Min
Performance	0,3	47 %	50 %	44,5 %
Power	0,05	2.140	2.250	1.995
Pressure	0,05	15	15,75	14,25
Cost (€/kW)	0,3	120	140	100
Durability	0,2	60.000	63.000	57.000
Emissions	0,05	500	525	475
Consumption	0,05	0,2	0,21	0,19
Weighted average			1,088 %	0,913 %

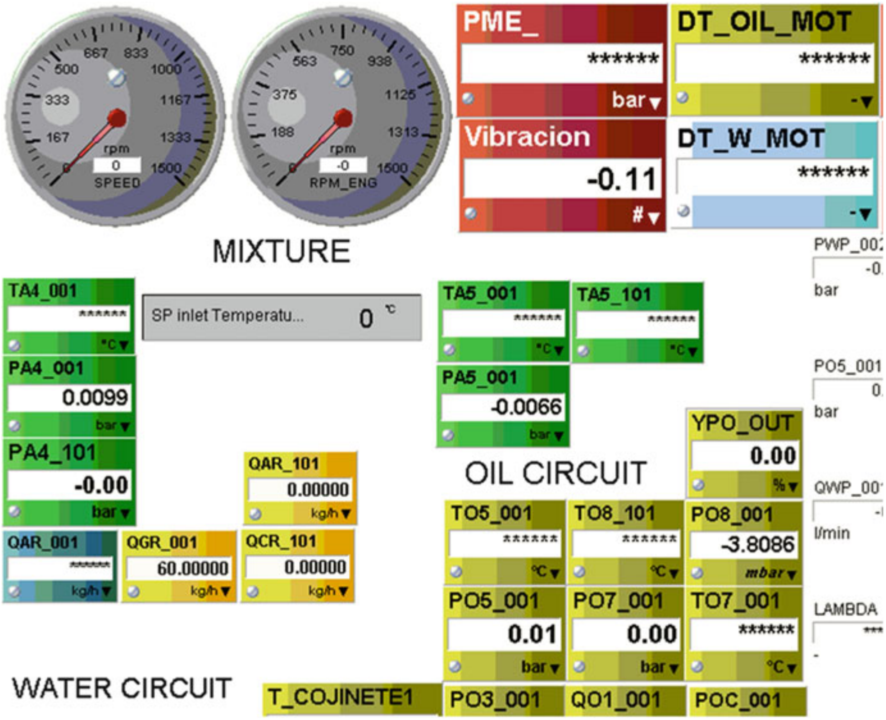


Fig. 9 Engine control unit monitoring engine performance

The project was scheduled in six phases of product development, such as, the concept design, detail design, prototypes construction, performance test, endurance test and series launching. The main objectives and their maximum and minimum acceptable values were defined in the project in order to decide the margins of acceptance for engine launching to the market. In Table 4 are presented the values for the objectives and the margins, as well as, the weighted average of the margins.

In Fig. 9, it is shown the tracking of the engine main running parameters, which were the objectives during the project.

Table 5 Project cumulative cost data

Project cumulative cost data (Thousands of Euros)							
Date	Materials	Investment	Labour	Testing	Engineering outsourcing	Trips	Total
Dec-05	178	206	55	14	425	10	889
Jan-06	148	206	74	22	678	10	1.139
Apr-06	174	219	131	50	678	10	1.263
Jul-06	248	259	172	75	679	11	1.444
Oct-06	308	260	197	93	679	12	1.549
Nov-06	316	260	208	101	679	12	1.576
Dec-06	383	262	211	104	995	12	1.967
Oct-07	417	279	309	176	1.443	13	2.637
Jun-08	444	279	401	242	1.522	16	2.905
Aug-08	445	279	425	261	1.523	16	2.950
Nov-08	472	279	475	281	1.526	23	3.056
Dec-08	473	279	488	281	1.526	23	3.070

In the analyzed project the available information for tracking along all the project life cycle was the following documents:

– The initial baseline schedule in MS Project software. It is detailed with multiple tasks levels and including the starting and finishing dates, as well as the dependences between them.
– The total budget of the project made at the beginning of the project and its breakdown in six general cost issues as follows: material costs, tooling investments, engineering hours, outsourcing expert consultancy support hours, testing costs and trips.
– A monthly report with the technical, economical and scheduling tracing. The technical part of the report includes a list of all manufacturing drawings and the testing results. The economical part of the report included cost monthly figures of the general spending issues collected from the invoices. The scheduling part of the report includes approximate deadlines for the critical tasks but not a detailed scheduling tracking.
– The general accounting of the project with the invoiced costs per month. There are available monthly figures of six general cost issues as it is shown in Table 5.

EVO method is here below applied, with real data from this project to analyze if it is possible to report status of the project in time, cost and technical compliance, to predict future project outcomes.

First of all, EVM has been applied to the data of the case study. The basic metrics for this methodology have been successfully obtained, such as AC, PV and EV, and their cumulative values which have been plotted throughout the project.

The total budget is equally distributed in the project periods to obtain the PV as used by some authors [20] (Fig. 10).

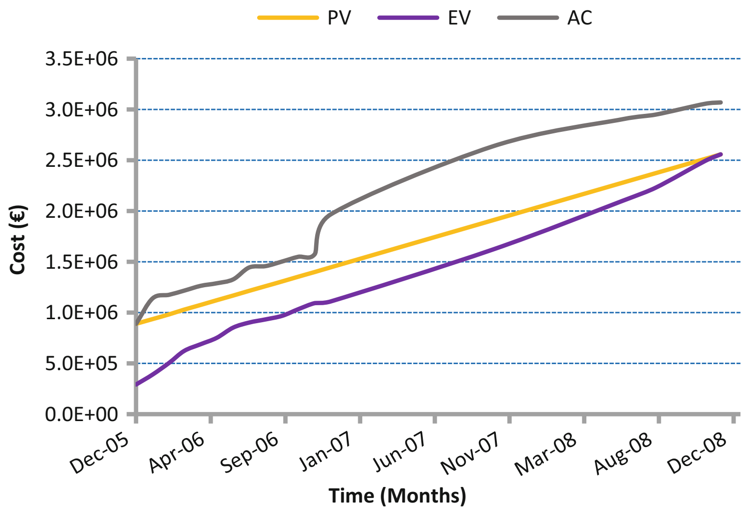


Fig. 10 EVM parameters PV, EV and AC

Table 6 Objectives accomplishment measurement and weighted average

Objective	Weight	Project phases					
		1	2	3	4	5	6
		Earned	Earned	Earned	Earned	Earned	Earned
Performance	0,3	1,00	0,96	0,90	1,00	0,85	0,87
Power	0,05	1,00	1,00	1,00	1,00	1,00	1,00
Pressure	0,05	1,00	1,00	1,00	1,00	1,00	1,00
Cost €/Kw	0,3	1,00	1,00	1,00	0,90	0,95	0,97
Durability	0,2	1,00	1,00	1,00	1,00	1,00	1,00
Emissions	0,05	1,00	0,95	0,90	0,9	0,85	0,87
Consumption	0,05	0,9	0,90	0,80	0,99	0,98	0,98
Weighted average		0,995	0,981	0,955	0,965	0,931	0,947

In parallel, it is possible to evaluate the objectives accomplishment measurement for each project phase. In this case study, the earned objectives in the first three phases are based on previous experience on five similar projects.

In Table 6 the objectives accomplishment measurement is shown, i.e., the "earned objectives", at the different project phases. From this measurement, the earned objectives weighted average is calculated (Table 7).

The new parameter EVO compared with EV and PV is presented in Fig. 11.

In Fig. 12, EVO values are shown including the maximum and minimum for acceptance.

Figure 13 shows the EVO maximum and minimum and their projection at the end of the project, this is, EVOACmin and EVOACmax, connected by lines. The areas formed by the EVOmin-EVOACmin-EVOACmax-EVOmax establish the limits for the EVO for each point of control. In this picture the EV is also plotted in order to locate the EVO margins in the EVM system.

Table 7 EVO calculation

PV	EV	Objectives weighted average	Impact factor	EVO	EVOmax	EVOmin
981.841	504.183	0,995	2,605	497.641	628.348	397.910
1.120.934	750.057	0,981	2,282	717.093	909.583	609.603
1.260.026	932.495	0,955	2,030	849.273	1.107.005	775.425
1.399.119	1.092.581	0,965	1,828	1.022.706	1.275.116	925.359
2.280.037	2.102.206	0,931	1,120	1.941.324	2.311.257	1.898.478
2.558.222	2.558.222	0,947	1,000	2.424.554	2.783.774	2.336.052

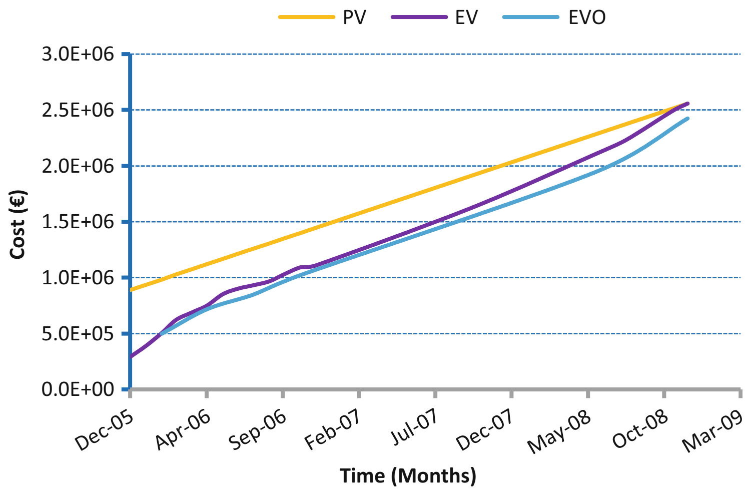


Fig. 11 PV, EV and EVO curves

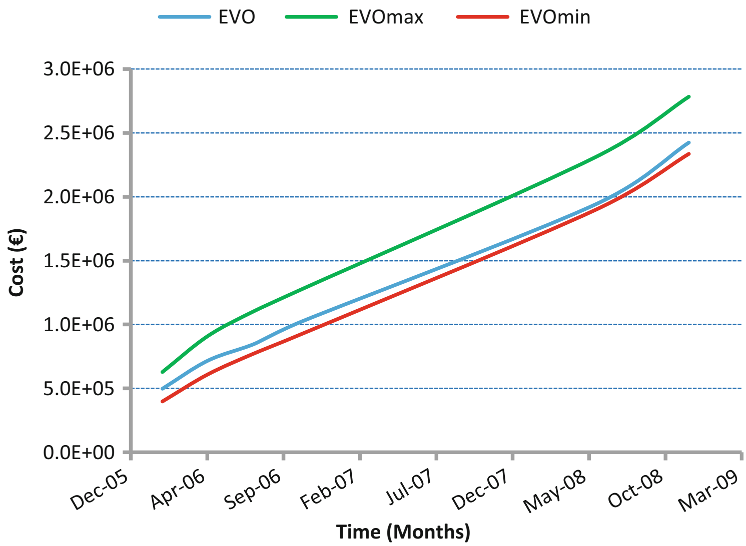


Fig. 12 EVO and its margins, EVOmax and EVOmin

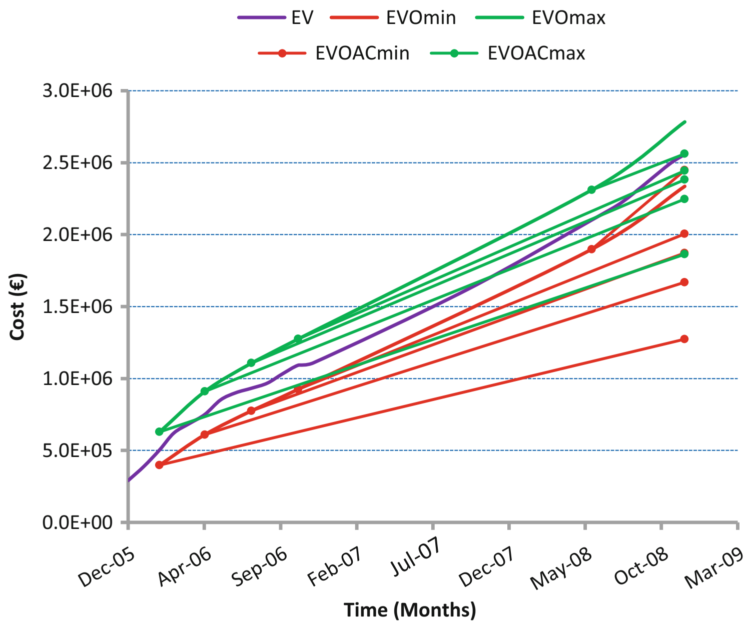


Fig. 13 EVO and EVOAC margins

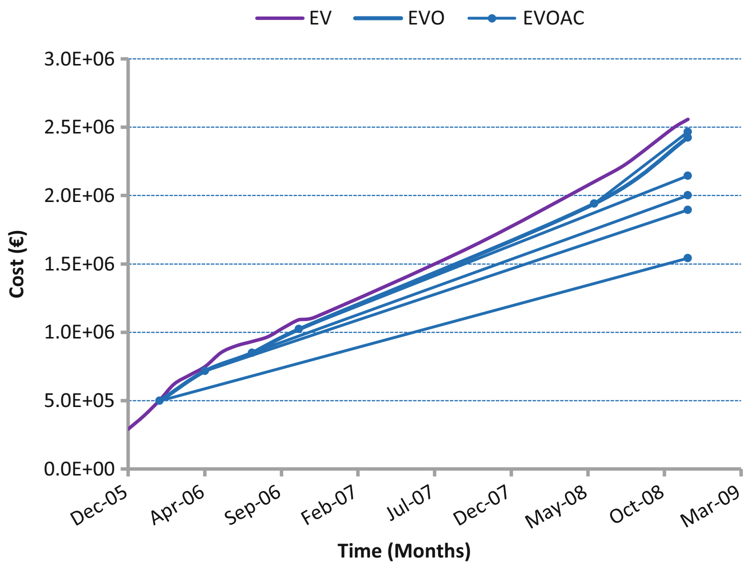


Fig. 14 EVO and EVOAC

Finally, Fig. 14 shows the EVO and the EVOAC obtained in this case study. These values must be inside the limits, and any deviation could be analyzed to take corrective actions.

5 Conclusions

Once the described model has been applied to real engineering project data, as it was in the case study, the following conclusions can be extracted:

– The evaluation of the objectives accomplishment measurement requires technical point of view, but it is possible to calculate it and their estimation at the end of the project using available project data and the formula presented in this work.
– EVO values are consistent during the complete project life. At the beginning of the project the objectives are closer to the planned ones but the impact of a deviation is bigger. This is the opposite at the end of the project.
– EVOAC demonstrates the impact of the initial phases in the objectives accomplishment at the end of the project. In this sense, EVOAC can quantify the effects of the deviations at crucial initial phases, regarding technical objectives.
– EVOAC must be between EVOACmax and EVOACmin at the end of the project, so that EVOAC measurement during the project allows correcting the product performance if needed.
– This formulation permits project managers to take decisions not only about cost and schedule but also about product definition and its embedded specifications.

References

1. PMBOK (2004) A guide to the Project Management Body of knowledge. Project Management Institute
2. Fleming Q, Koppelman J (1998) Earned value project management. J Def Softw Eng
3. American National Standards Institute (1998) Earned value management systems. ANSI/EIA-748-A-1998
4. DOD Earned Value Management Implementation Guide. U.S. Department of Defense. http://www.acq.osd.mil/evm/resources/guidance-references.shtml. Accessed 10 February 2012
5. Humphreys and Associates. http://www.humphreys-assoc.com/evms/evms-implementation-ta-a-18.html. Accessed 17 February 2012
6. National Aeronautics and Space Administration. NASA. http://evm.nasa.gov. Accessed 3 February 2012
7. Noori S, Bagherpour M, Zorriasatine F (2008) Designing a control mechanism for production planning problems by means of earned value analysis. J Appl Sci 8(18):3221–3227, ISSN 1812-5654
8. Vanchoucke M, Vandevoorde S (2008) Earned value forecast accuracy and activity critically. Measurable News Summer (3):13–16
9. Fleming Q, Koppelman J (2009) The two most useful earned value metrics: the CPI and the TCPI. J Def Softw Eng
10. Lipke W (2003) Shedule is different. The Measurable News, 31–34
11. Lipke W, Zwikael O, Henderson K, Anbari F (2009) Prediction of project outcome: the application of statistical methods to earned value management and earned schedule performance indexes. Int J Project Manag 27(4):400–407
12. Pajares J, López-Paredes A (2011) An extension of the EVM analysis for project monitoring: the Cost Control Index and the Schedule Control Index. Int J Proj Manag 29(5):615–621

13. Acebes F, Pajares J. A new approach of project management under uncertainty. Int J Proj Manag. Accepted 13 August 2013
14. Solomon P (2001) Practical software measurement, performance-based earned value. J Def Softw Eng
15. Solomon P (2005) Performance-based earned value. J Def Softw Eng
16. Solomon P (2006) Practical performance-based earned value. J Def Softw Eng
17. Solomon P (2008) Integrating systems engineering with earned value management. Meas News 4.
18. Moslemi L, Shadrokh S, Salehipour A (2011) A fuzzy approach for the earned value management. Int J Proj Manag 29(6):764–772
19. Malcolm DG, Roseboom JH, Clark CE, Fazar W (1958) Application of a technique for research and development program evaluation. Oper Res 7(5):646–669. doi:10.1287/opre.7.5.646. ISSN 0030-364X
20. Anbari F (2003) Earned value project management method and extensions. Project Manag J 34 (4):12–23
21. Project Management Institute (2005) The practice standard for earned value management. Project Management Institute
22. Kerzner H (2003) Project management. A systems approach to planning, scheduling and controlling. Wiley: New York
23. Kwak Y, Anbari F (2012) History, practices, and future of earned value management in government: perspectives from NASA. Proj Manag J 43(1):77–90
24. Fleming Q, Koppelman J (2006) Start with "Simple" Earned Value. . .on All Your Projects. Meas News
25. Xu J (2010) Project integrated management based on quality earned value. In: Proceedings of the 2nd International Conference on Information Science and Engineering, pp 432–435

Use of Excellence Models as a Management Maturity Model (3M)

Jose Ramón García Aranda and Fausto Pedro García Márquez

1 Introduction

The management and improvement of business processes are critical approaches for the organisational design from the 1990s, where it was introduced the advent of new business management strategies focused on the analysis and design of workflows and processes in an organisation. It can be found a large number of methodologies and tools employed to aligning an organisation's business processes with the requirements and needs of clients through an integrated management approach.

A maturity model usually includes a sequence of levels that creates an anticipated, desired, or logical path from an initial state to maturity. An organisation's current maturity level represents the progress on improving business effectiveness and efficiency while works for improving innovation, flexibility, and integration considering the technology available.

Usually, maturity models define different levels for increasing capability and maturity, establishing different criteria for every level, for example:

Level 1: Initial
Level 2: Managed
Level 3: Established
Level 4: Predictable/Quantitatively managed
Level 5: Optimised

J.R. García Aranda (✉)
University of Zaragoza, Zaragoza, Spain
e-mail: joser@unizar.es

F.P. García Márquez
University of Castilla-La Mancha, Castilla-La Mancha, Spain
e-mail: FaustoPedro.Garcia@uclm.es

© Springer International Publishing Switzerland 2015

F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_8

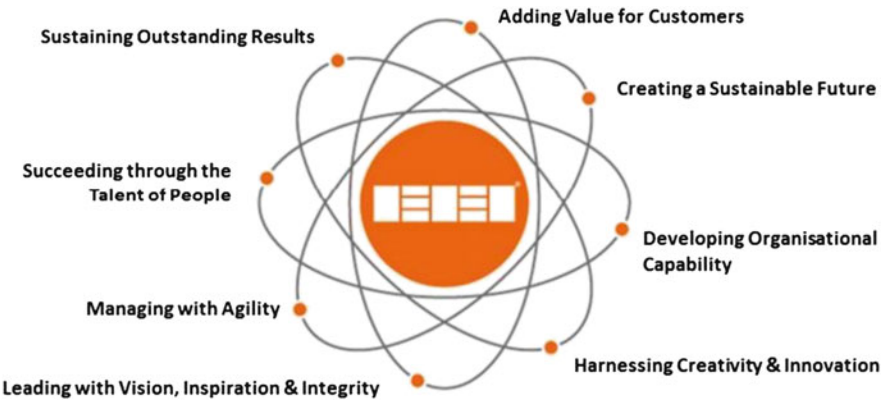


Fig. 1 The fundamental concepts of excellence [1]

The methodology of excellence models is similar to the maturity models, especially in the European Foundation for Quality Management (EFQM) excellence model [1] (see Fig. 1). It presents different levels of maturity depending on the evolution of an organisation through the implementation of best management practices:

Level 1: Unable to demonstrate.
Level 2: Limited ability to demonstrate.
Level 3: Able to demonstrate.
Level 4: Fully able to demonstrate.
Level 5: Recognized as a global role model.

In conclusion, the maturity model could be understood on how to measure the improvement of the organisation in any field establishing different levels of evolution.

2 Structure of an Excellence Model for EFQM

Excellent organisations try to get and improve the requirements of the stakeholders around to his organization. Excellence models provide the structure to analyse the level reached through a set of three integrated components:

Principles.
Criteria.
An assessment framework.

2.1 Principles

In model, the fundamental concepts of excellence outline the foundation for achieving sustainable excellence (Fig. 1).

According to EFQM:

> they can be used as the basis to describe the attributes of an excellent organisational culture. Each of the Concepts is important in its own right but maximum benefit is achieved when an organisation can integrate them all into its culture.

The fundamentals concepts are:

- Adding value for customers.
- Creating a sustainable future.
- Developing organisational capability.
- Harnessing creativity & innovation.
- Leading with vision, inspiration & integrity.
- Managing with agility.
- Succeeding through the talent of people.
- Sustaining outstanding results.

2.2 Criteria

Figure 2 shows the fundamental concepts of excellence form the basis for the criteria of the EFQM excellence model.

The EFQM excellence model is based on nine criteria, five of these are "Enablers", pictured on the left-hand side of the Model, and four are "Results", shown on the right-hand side of the Model. The "Enabler" criteria cover what an organisation does and how it does it. The "Results" criteria cover what an organisation achieves.

The "Enablers" criteria are:

- Leadership.
- Strategy.
- People.
- Partnerships & resources.
- Processes, products & services.

The "Results" criteria are:

- Customer results.
- People results.
- Society results.
- Business results.

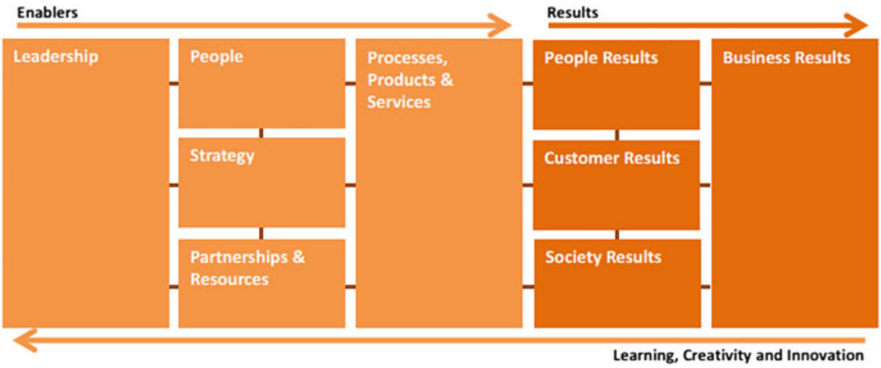


Fig. 2 EFQM criteria [1]

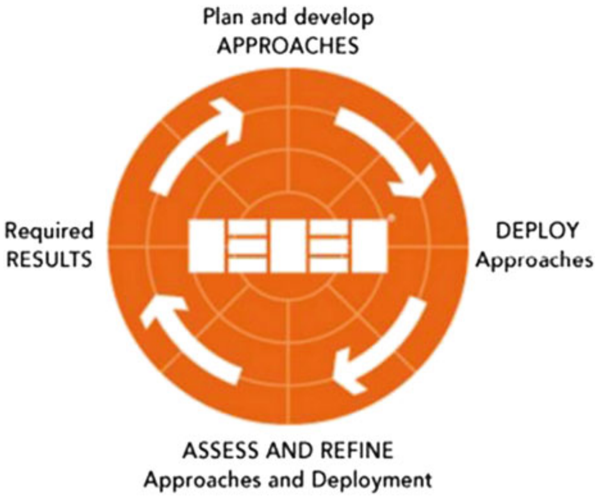


Fig. 3 Structure of the RADAR logic [1]

2.3 *Assessment Framework*

The RADAR logic is a dynamic assessment framework that provides a structured approach to questioning the performance of an organisation (Fig. 3).

The RADAR logic includes these five elements:

- R: To what extent are results used to set targets for process performance?
- A: To what extent is a clear approach defined and understood?
- D: To what extent is the approach deployed?
- A: To what extent is the process assessed?
- R: To what extent is the process reviewed?

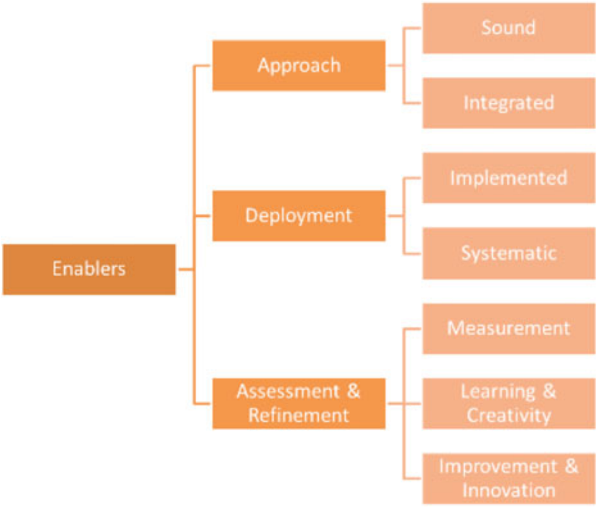


Fig. 4 Enablers (attributes) [1]

EFQM explain the RADAR logic, at the highest level, as a statement that an organisation needs in order to [1]:

- Determine the "results" it is aiming to achieve as part of its strategy
- Plan and develop an integrated set of sound "approaches" to deliver the required results both now and in the future
- Deploy the approaches in a systematic way to ensure implementation
- Assess and Refine the deployed approached based on monitoring and analysis of the results achieved and ongoing learning activities

This management approach leads to a continuous improvement, assessing the maturity of the approaches implemented and the excellence of the results achieved.

The RADAR elements can be broken down into a series of attributes (Figs. 4 and 5) which contain guidance on what we expect the organization to demonstrate. In the case of enablers, the main attributes are shown in Fig. 4.

In the case of results, the main attributes are presented in Fig. 5.

3 Business Process Maturity Models and Excellence Models

Based on the assumption of predictable patterns of organisational evolution and change, maturity models generally represent approaches about how an organisation's capabilities evolve in a stage-by-stage manner along an anticipated, desired, or logical path. The main objective of the maturity models is to define the stages of

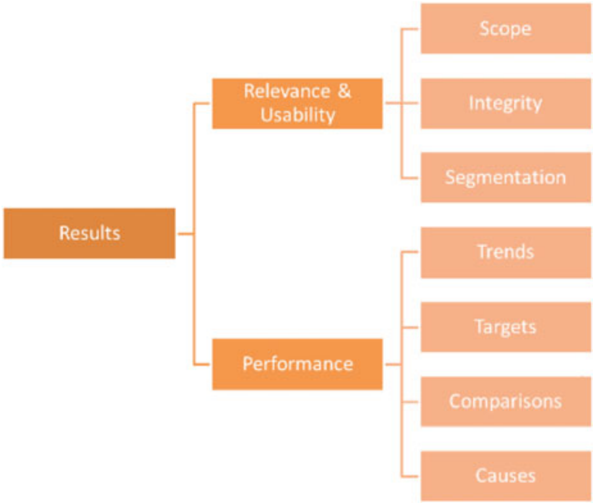


Fig. 5 Results (attributes) [1]

maturation paths [2]. This includes the characteristics of each stage and the logical relationship between them.

According to B. Curtis and J. Alden [3] the Business Process Maturity Model (BPMM):

> describes an evolutionary improvement path that guides organisations as they move from immature, inconsistent business activities to mature, disciplined processes. The BPMM orders these stages so that improvements at each stage provide a foundation on which to build improvements undertaken at the next stage. Thus, an improvement strategy drawn from the BPMM provides a roadmap for continuous process improvement. It helps identify process deficiencies in the organisation and guides the improvements in logical, incremental steps.

The RADAR logic is employed as a structured approach to support the performance of an organisation, assessing the maturity of the approaches that the organisation has been implemented. In other words, it is used to guide improvement initiatives and to control progress as BPMM.

4 Measuring Management Maturity: A Real Case Study

The proposed management maturity model (3M) has two levels:

- Maturity level related to the management values (qualitative stage).
- Maturity level related to the management criteria (quantitative stage).

4.1 Assessment of Management Values: Qualitative Evaluation

It is necessary to establish an evaluation scale for the management principles as an initial step. The evaluation of the management maturity determines the extent to which the organisation has succeeded in implementing holistic management values. Some management tools, like common assessment framework (CAF) [4], have integrated similar practices in their self-assessment processes.

The evaluation scale has five qualitative maturity levels:

- Level 1: Unable to demonstrate.
- Level 2: Limited ability to demonstrate.
- Level 3: Able to demonstrate.
- Level 4: Fully able to demonstrate.
- Level 5: Recognized as a global role model.

It can be compared that with the common levels in business process maturity models. It can be found similitudes between both methodologies, known as "pyramid of levels":

- Unable to demonstrate vs. Initial/Foundation.
- Limited ability to demonstrate vs. Emerging/Managed.
- Able to demonstrate vs. Defined/Standardized/Established.
- Fully able to demonstrate vs. Predictable/Quantitatively Managed.
- Recognized as a global role model vs. Optimised.

This evaluation scale is based on the fundamental concepts of excellence (Table 1):

The assessment of the management values relates to the maturity level achieved by the organisation as a result of the work with a self-assessment process, where the organisation determines the level reached.

4.2 Assessment of Management Criteria: Quantitative Evaluation

In a second step, after this first general appraisal, the organisation can use RADAR tool for enablers and results as an evaluation method used (beyond national or international awards) within organisations to obtain a score through a self-assessment process.

The evaluation scale has five quantitative maturity levels:

- Level 1: Unable to demonstrate (0–10 %)
- Level 2: Limited ability to demonstrate (15–35 %)

Table 1 Evaluation scale (management values)

Management maturity levels	Level 1 Unable to demonstrate	Level 2 Limited ability to demonstrate	Level 3 Able to demonstrate	Level 4 Fully able to demonstrate	Level 5 Recognized as a global role model
Principle of excellence 1. Sustaining outstanding results					
Definition	The initiation level has not been reached	The organisation identifies relevant stakeholders and result areas	The organisation defines a set of targets and results to be achieved in relation to the relevant stakeholders' needs	The organisation systematically monitors the results it achieves and uses it for continuous improvement	The organisation is recognised as a global role model in this area
Principle of excellence 2. Adding value for customers					
Definition	The initiation level has not been reached	The organisation focuses on the needs of existing and potential customers	The organisation involves customers in the evaluation and improvement of its performance	The organisation responds to the needs of customers by developing and delivering activities, products and services	The organisation is recognised as a global role model in this area
Principle of excellence 3. Leading with vision, inspiration & integrity					
Definition	Initiation level has not been reached	Leaders establish a clear mission statement	Leaders establish vision and values. They drive and inspire people towards excellence	Leaders demonstrate the capability to keep a constancy of purpose in a changing environment	The organisation is recognised as a global role model in this area
Principle of excellence 4. Managing with agility					
Definition	The initiation level has not been reached	Processes are identified and managed	The implementation of the strategy and planning of the organisation is enabled and assured through the processes	Processes are continuously improved for effectiveness on the basis of internal performance measurement, bench learning and/or benchmarking	The organisation is recognised as a global role model in this area

Principle of excellence 5. Succeeding through the talent of people					
Definition	The initiation level has not been reached	The organisation takes initiatives for developing and involving people	The organisation develops competencies and involves people in a structured way to improve products, services and processes	The organisation creates a working environment of shared values and a culture of trust, openness, empowerment and recognition	The organisation is recognised as a global role model in this area
Principle of excellence 6. Harnessing creativity & innovation					
Definition	The initiation level has not been reached	The organisation learns from its activities and performance and looks for opportunities for improvement	Continuous improvement is promoted in the organisation, through sharing knowledge and taking into account people's suggestions	The organisation systematically challenges the status quo, encourages, accepts and integrates innovation and regularly compares its performance to other organisations	The organisation is recognised as a global role model in this area
Principle of excellence 7. Developing organisational capability					
Definition	The initiation level has not been reached	The organisation identifies its partners	The organisation formalises partnerships to reach mutual advantages	The organisation manages partnerships in a win-win situation to enable delivery of enhanced value and to optimise the use of resources	The organisation is recognised as a global role model in this area
Principle of excellence 8. Creating a sustainable future					
Definition	The initiation level has not been reached	The organisation is aware of its impact on society (social and environmental)	The organisation is actively involved in activities related to social responsibility and ecological sustainability	The organisation meets or exceeds the major expectations and requirements of the local and—where appropriate—the global community	The organisation is recognised as a global role model in this area

- Level 3: Able to demonstrate (40–60 %)
- Level 4: Fully able to demonstrate (65–85 %)
- Level 5: Recognized as a global role model (90–100 %)

The first step to set the criteria to fix a percentage score to each criterion point. This is achieved by considering each of the elements and attributes of the RADAR tool for each criterion component. The scoring summary sheet is therefore used to combine the percentage scores awarded, corrected by a weight, and then derive a score on a scale of 0–1,000 points.

It is remarkable that the excellence models, i.e. the European Model of Excellence, have within them weighted scoring allocations. Despite the fact that these weights have been widely accepted and reviewed frequently, these are "generic" weights that cannot be adapted to the specific reality of the organisation.

This point is critical to step from the excellence model scoring tools to a management maturity model, in which an organisation establishes not exactly the criteria but some adjustments in weighs in its environment and social context.

Table 2 presents the RADAR tool for enablers, where this quantitative evaluation can be an excellent base to know five different levels of evolution (maturity) in any field: Processes, Projects, Systems, internal management methodologies and dynamics, etc.

Table 3 shows the RADAR tool for results, where RADAR allows identifying the maturity level of enablers and also the results:

5 Conclusions

It has been demonstrated that management maturity models can be excellent tools for measuring and benchmarking the quality of management performance.

In terms of structure, a Management Maturity Model can be built according to the main components of any Excellence Models: Principles, criteria and the assessment framework. These elements, specially the criteria, are basic for world-class management practices, beyond the excellence model chosen, and are all fairly similar to each other.

In the case of the EFQM Excellence Model, the systematic use of an adapted tool as RADAR, with a new perspective as a measurement of maturity, would allow to strongly improve management practices in a holistic way.

Table 2 Evaluation scale (management criteria—enablers) [1]

RADAR tool for enablers						
	Guidance	Unable to demonstrate	Limited ability to demonstrate	Able to demonstrate	Fully able to demonstrate	Recognised as a global role model
Approach						
Sound	The approaches have a clear rationale, based on the relevant stakeholder needs, and are process based	No evidence or anecdotal	Some evidence	Evidence	Clear evidence	Comprehensive evidence
Integrated	The approaches support strategy and are linked to other approaches as appropriate	No evidence or anecdotal	Some evidence	Evidence	Clear evidence	Comprehensive evidence
Deployment						
Implemented	The approaches are implemented in relevant areas, in a timely manner	No evidence or anecdotal	Implemented in about 1/4 of relevant areas	Implemented in about 1/2 of relevant areas	Implemented in about 3/4 of relevant areas	Implemented in all relevant areas
Structured	The execution is structured and enables flexibility and organisational agility	No evidence or anecdotal	Some evidence	Evidence	Clear evidence	Comprehensive evidence
Assessment and refinement						
Measurement	The effectiveness and efficiency of the approaches and their deployment are appropriate measured	No evidence or anecdotal	Some evidence	Evidence	Clear evidence	Comprehensive evidence
Learning & Creativity	Learning and creativity is used to generate opportunities for improvement or innovation	No evidence or anecdotal	Some evidence	Evidence	Clear evidence	Comprehensive evidence
Improvement & Innovation	Outputs from measurement, learning and creativity are used to evaluate, prioritise and implement improvements and innovations	No evidence or anecdotal	Some evidence	Evidence	Clear evidence	Comprehensive evidence

(continued)

Table 2 (continued)

RADAR tool for enablers						
	Guidance	Unable to demonstrate	Limited ability to demonstrate	Able to demonstrate	Fully able to demonstrate	Recognised as a global role model
Overall Score		0–10 %	15–35 %	40–60 %	65–85 %	90–100 %
Scale		0 %	25 %	50 %	75 %	100 %

Table 3 Evaluation scale (management criteria—results) [1]

RADAR tool for results						
	Guidance	Unable to demonstrate	Limited ability to demonstrate	Able to demonstrate	Fully able to demonstrate	Recognised as a global role model
Relevance & usability						
Scope & Relevance	A coherent set of results, including key results, are identified that demonstrate the performance of the organisation in terms of its strategy, objectives and the needs and expectations of the relevant stakeholders	No results or anecdotal information	Results address 1/4 of relevant areas and activities	Results address 1/2 of relevant areas and activities	Results address 3/4 of relevant areas and activities	Results address all of relevant areas and activities
Integrity	Results are timely, reliable and accurate	No results or anecdotal information	Adequate integrity in 1/4 of results	Adequate integrity in 1/2 of results	Adequate integrity in 3/4 of results	Adequate integrity in all results
Segmentation	Results are properly segmented to provide meaningful insights	No results or anecdotal information	Properly segmentation in 1/4 of results	Properly segmentation in 1/2 of results	Properly segmentation in 3/4 of results	Properly segmentation in all results
Performance						
Trends	Positive trends or sustained good performance over at least 3 years	No results or anecdotal information	Positive trends and/or satisfactory performance on some results	Positive trends and/or sustained good performance on many results over the last 3 years	Strongly positive trends and/or sustained excellent performance on most results over at least 3 years	Strongly positive trends and/or sustained excellent performance in all areas over at least 5 years

(continued)

Table 3 (continued)

RADAR tool for results

	Guidance	Unable to demonstrate	Limited ability to demonstrate	Able to demonstrate	Fully able to demonstrate	Recognised as a global role model
Targets	Relevant targets are set and consistently achieved for the key results, in line with the strategic goals	No Results or anecdotal information	Achieved and appropriate for about 1/4 of results	Achieved and appropriate for about 1/2 of results	Achieved and appropriate for about 3/4 of results	Achieved and appropriate for all results
Comparisons	Relevant external comparisons are made and are favourable for the key results, in line with the strategic goals	No results or anecdotal information	Favourable comparisons for about 1/4 of results	Favourable comparisons for about 1/2 of results	Favourable comparisons for about 3/4 of results	Favourable comparisons for all results
Confidence	There is confidence that performance levels will be sustained into the future, based on established cause & effect relationships	No results or anecdotal information	Performance can be sustained (with cause and effect visible) for about 1/4 of results	Performance can be sustained (with cause and effect visible) for about 1/2 of results	Performance can be sustained (with cause and effect visible) for about 3/4 of results	Performance can be sustained (with cause and effect visible) for all results
Overall Score		0–10 %	15–35 %	40–60 %	65–85 %	90–100 %
Scale		0 %	25 %	50 %	75 %	100 %

References

1. European Foundation for Quality Management (2013) The EFQM excellence model (framework)
2. Röglinger M, Pöppelbuß J, Becker J (2012) Maturity models in business process management. Bus Process Manag J 18(2):328–346
3. Curtis B, Alden J (2007) The business process maturity model (BPMM): what, why and how. BPTrends Column
4. European CAF Resource Centre (2013) CAF external feedback

AHP and Intuitionistic Fuzzy TOPSIS Methodology for SCM Selection

Babak Daneshvar Rouyendegh

1 Introduction

Multi-Criteria Decision-Making (MCDM) is a modeling and methodological tool for dealing with complex engineering problems [1]. Many mathematical programming models have been developed to address MCDM problems. However, in recent years, MCDM methods have gained considerable acceptance for judging different proposals. Intuitionistic fuzzy set (IFS) theory introduced by Atanassov [2] is an extension of the classical Fuzzy Set (FS), which is a suitable tool to deal with the vagueness and uncertainty decision information [2]. Recently, some researchers have shown interest in the IFS theory and applied it in the field of MCDM [3–10]. However, IFS has also been applied to many areas, such as medical diagnosis [11–13] decision-making problems [6–8, 14–31], pattern recognition [33–38], supplier sélection [39, 40], entreprise partners selection [41], personnel selection [42], evaluation of renewable energy [43], facility location selection [44], web service selection [45], printed circuit board assembly [46], management information system [47] and project selection [48].

The AHP proposed by Saaty [49] is one of the most popular methods in the based on the preference relation in the decision-making process [49]. The AHP is a well-known method for solving decision-making problems. In this method, the decision-maker (DM) performs pair-wise comparisons and, then, the pair-wise comparison matrix and the eigenvector are derived to specify the weights of each parameter in the problem. The weights guide the DM in choosing the superior alternative.

We study the AHP-IFT methodology here where all the values are expressed in Intuitionistic fuzzy numbers collected. To do that, we first present the concept of AHP and determine the weight of the criteria based on the opinions of decision-

B.D. Rouyendegh (✉)
Department of Industrial Engineering, Atilım University, P.O. Box 06836, İncek, Ankara, Turkey
e-mail: babekd@atilim.edu.tr

© Springer International Publishing Switzerland 2015
F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_9

makers. Then, we introduce the concept of IFT and develop a model based on such opinions. The rest of the paper is organized as follows: Sect. 2 provides the materials and methods—mainly AHP, Fuzzy Set Theory (FST) and Intuitionistic Fuzzy Set (IFS). The AHP-IFT methodology is introduced in Sect. 3. How the proposed model is used in a numerical example is explained in Sect. 4. The conclusions are provided in the final section.

2 Preliminaries

2.1 Basic Concept of AHP

The AHP is a general theory of measurement. It is used to derive relative priorities on absolute scales from both discrete and continuous paired comparisons in multilevel hierarchic structures. These comparisons may be taken from a fundamental scale that reflects the relative strength of preferences. The AHP has a special concern with deviation from consistency and the measurement of this deviation, and with dependence within and between the groups of elements of its structure. It has found its widest applications in MCDM. Generally, the AHP is a nonlinear framework for carrying out both deductive and inductive thinking without the use of syllogism [50].

The AHP proposed by Saaty [49] is a flexible method for selecting among alternatives based on their relative performance with respect to a given criteria [51, 52]. AHP resolves complex decisions by structuring alternatives into a hierarchical framework. Such hierarchy is constructed through pair-wise comparisons of individual judgments rather than attempting to prioritize the entire list of decisions and criteria. This process has been given as follows [53]:

Describe the unstructured problem; Detail the criteria and alternatives; Recruit pair-wise comparisons among decision elements; Use the Eigen-value method to predict the relative weights of the decision elements; Compute the consistency properties of the matrix, and Collect the weighted decision elements.

The AHP techniques form a framework of the decisions that use a one-way hierarchical relation with respect to decision layers. The hierarchy is constructed in the middle level(s), with decision alternatives at the bottom. The AHP method provides a structured framework for setting priorities at each level of the hierarchy using pair-wise comparisons that are quantified using a 1–9 scale as demonstrated in Table 1.

Table 1 The fundamental scale

Importance intensity	Definition
1	Equal importance
3	Moderate importance of one over another
5	Strong importance of one over another
7	Very strong importance of one over another
9	Extreme importance of one over another
2, 4, 6, 8	Intermediate values

2.2 FST

Zadeh [54] introduced the FST to deal with uncertainty and vagueness. A major contribution of FST is capability in representing uncertain data. FST also allows mathematical operators and programming to be performed in the fuzzy domain. An FS is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership function, which assigns to each object a grade of membership ranging "between" zero and one [55, 56].

A tilde '~' will be placed above a symbol if the symbol shows an FST. A Triangular Fuzzy Number (TFN) $\widetilde{M}$ is shown in Fig. 1. A TFN is denoted simply as (a,b,c). The parameters a, b and c $(a \leq b \leq c)$, respectively, denote the smallest possible value, the most promising value, and the largest possible value that describe a fuzzy event. The membership function of TFN is as follows:

Each TFN has linear representations on its left and right side, such that its membership function can be defined as

$$\mu\left(\frac{x}{\widetilde{M}}\right) = \begin{cases} 0, & x < a, \\ (x-a)/(b-a), & a \leq x \leq b, \\ (c-x)/(c-b), & b \leq x \leq c, \\ 0, & x > c. \end{cases} \tag{1}$$

The left and right representation of each degree of membership as in the following:

$$\widetilde{M} = M^{l(y)}, M^{r(y)} = (a + (b-a)y, c + (b-c)y), \quad y \in [0,1] \tag{2}$$

where $l(y)$ and $r(y)$ denote the left-side representation and the right-side representation of a fuzzy number(FN), respectively. Many ranking methods for FNs have

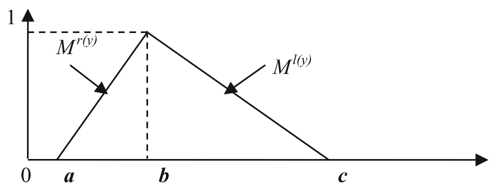


Fig. 1 A TFN $\widetilde{M}$

been developed in the literature. These methods may provide different ranking results [57].

While there are various operations on TFNs, only the important operations used in this study are illustrated. Two positive TFNs (a_1, b_1, c_1) and (a_2,b_2,c_2) have been given as follows:

$$\begin{aligned}(a_1,\ b_1,\ c_1)\ +\ (a_2,\ b_2,\ c_2)\ &= (a_1 +\ a_{2,}, b_1 +\ b_2, c_1 + c_2),\\ (a_1,\ b_1,\ c_1)\ -\ (a_2,\ b_2,\ c_2)\ &= (a_1 -\ a_{2,}, b_1 -\ b_2,\ c_1 -\ c_2),\\ (a_1,\ b_1,\ c_1) *\ (a_2,\ b_2,\ c_2)\ &= (a_1 *\ a_{2,},\ b_1 *\ b_2,\ c_1 *\ c_2)\ ,\\ (a_1,\ b_1,\ c_1)/(a_2,\ b_2,\ c_2)\ &= \left(a_{1/}c_{2,},\ b_{1/}b_2,\ c_1/a_2\right).\end{aligned} \tag{3}$$

2.3 Basic Concept of IFS

The following formulas briefly introduce some necessary introductory basic concepts of IFS. IFS A in a finite set R can be written as:

$$\chi_{\mathrm{ij}} = \left(\mu_{\mathrm{ij}}, v_{\mathrm{ij}}, \pi_{\mathrm{ij}}\right),$$

Where,

$\boldsymbol{\mu_{ij}}$: Degree of membership of the the i_{th} alternative with respect to j_{th} criteria

$\boldsymbol{v_{ij}}$: Degree of non-membership of i_{th} alternative with respect to j_{th} criteria

$\boldsymbol{\pi_{ij}}$: Degree of hesitation of the i_{th} alternative with respect to the j_{th} criteria

R is an intuitionistic fuzzy decision matrix.

$$\begin{aligned}&A = \{\langle r, \mu_A(r), v_A(r)\rangle | r \in R\}\\ &\mathit{where}\\ &\mu_A(r) : \mu_A(r) \in [0, 1], R \to \quad [0, 1]\\ &v_A(r) : v_A(r) \in [0, 1], R \to \quad [0, 1]\end{aligned} \tag{4}$$

μ_A and v_A are the membership function and non-membership function, respectively, such that

$$0 \le \mu_A(r) \oplus v_A(r) \le 1 \quad \forall r \in R \quad R \to \quad [0, 1] \tag{5}$$

A third parameter of IFS is $\pi_A(r)$, known as the intuitionistic fuzzy index or hesitation degree of whether r belongs to A or not

$$\pi_A(r) = 1 - \mu_A(r) - v_A(r) \tag{6}$$

$\pi_A(r)$ is called the degree of indeterminacy of r to A.

It is obviously seen that for every $r \in R$:

$$0 \leq \pi_A(r) \leq 1 \quad \pi_A(r) \tag{7}$$

If $\pi_A(r)$ is small the knowledge about r is more certain. However, if $\pi_A(r)$ is great, this knowledge is rather uncertain. Obviously, when

$$\mu_A(r) = 1 - v_A(r) \tag{8}$$

For all elements of the universe, the ordinary FST concept is recovered [46].

Let A and B are IFSs of the set R. Then, the multiplication operator is defined as follows (2).

$$A \otimes B = \{\mu_A(r).\mu_B(r), v_A(r) + v_B(r) - v_A(r).v_B(r) | r \in R\} \tag{9}$$

3 AHP-IFT Hybrid Method

To rank a set of alternatives, the AHP-IFT methodology as an outranking relation theory is used to analyze the data of a decision matrix. We assume m alternatives and n decision criteria. Each alternative is evaluated with respect to the n criteria. All the values assigned to the alternatives with respect to each criterion form a decision matrix.

In this study, our model integrates two well-known models, AHP and IFT. The evaluation of the study based on this hybrid methodology is given in Fig. 2. The procedure for AHP-IFT methodology ranking model has been given as follows:

Let $A = \{A_1, A_2, \ldots, A_m\}$ be a set of alternatives and $C = \{C_1, C_2, \ldots, C_n\}$ be a set of criteria. It should be mentioned here that the presented approach mainly utilizes the IFT method proposed in [39, 42–44, 48]. The procedure for AHP-IFT methodology is conducted in seven steps presented as follows:

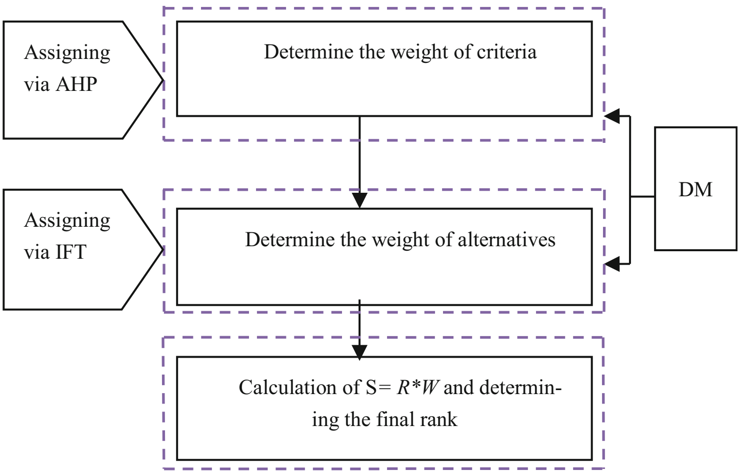


Fig. 2 Schematic diagram of the AHP–IFT

Table 2 Fundamental scale of absolute numbers

Importance intensity	Definition	Definition
1	Very Bad (VB)	Equal importance
3	Bad(B)	Moderate importance of one over another
5	Medium Best(MB)	Strong importance of one over another
7	Good(G)	Very strong importance of one over another
9	Very Good(VG)	Extreme importance of one over another

Step 1

Determine the weight of the criteria based on the opinion of decision-makers (W).

In the first step, we assume that the decision group contains $1 = \{1_1, 1_2, \ldots, 1_1\}$ DMs. The DMs is given the task of forming individual pair-wise comparisons by using standard scale as in Table 2.

Step 2

Determine the weights of importance of DMs:

In the second step, we assume that the decision group contains $l = \{l_1, l_2, \ldots, l_l\}$ DMs. The importance's of the DMs are considered as linguistic terms which are assigned to IFNs. Let $D_l = [\mu_l, v_l, \pi_l]$ be an intuitionistic fuzzy number for rating of k_{th} DM. Then, the weight of l_{th} DM can be calculated as:

$$\lambda_l = \frac{\left(\mu_l + \pi_l\left(\frac{\mu_l}{\mu_l + v_l}\right)\right)}{\sum_{l=1}^{k}\left(\mu_l + \pi_l\left(\frac{\mu_l}{\mu_l + v_l}\right)\right)} \quad where\, \lambda_l \in [0, 1] \quad and \quad \sum_{l=1}^{k} \lambda_l = 1. \tag{10}$$

Step 3

Determine the Intuitionistic Fuzzy Decision Matrix (IFDM).

Based on the weight of DMs, the aggregated intuitionistic fuzzy decision matrix (AIFDM) is calculated by applying the intuitionistic fuzzy weighted averaging (IFWA) operator Xu [58]. In a group decision-making process, all the individual decision opinions need to be fused into a group opinion to construct AIFDM [58].

Let $R^{(l)} = (r_{ij}^{(l)})_{m \times n}$ be an IFDM of each DM. $\lambda = \{\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k\}$ is the weight of DM as result, are equal.

$$R = (r_{ij})_{m \times n},$$

Where

$$r_{ij} = IFWA_\lambda\left(r_{ij}{}^{(1)}, r_{ij}{}^{(2)}, \ldots, r_{ij}{}^{(l)}\right)$$
$$= \lambda_1 r_{ij}{}^{(1)} \oplus \lambda_2 r_{ij}{}^{(2)} \oplus \lambda_3 r_{ij}{}^{(3)} \oplus \ldots \oplus \lambda_k r_{ij}{}^{(k)}$$
$$= \left[1 - \prod_{l=1}^{k}\left(1 - \mu_{ij}{}^{(l)}\right)^{\lambda_l}, \prod_{l=1}^{k}\left(v_{ij}{}^{(l)}\right)^{\lambda_l}, \prod_{l=1}^{k}\left(1 - \mu_{ij}{}^{(l)}\right)^{\lambda_l} - \prod_{l=1}^{k}\left(v_{ij}{}^{(l)}\right)^{\lambda_l}\right]. \quad (11)$$

Step 4
Calculate $S = R^*W$:

In the step 4, a weight of criteria (W) with respect to IFDM (R) is defined as follows:

$$S = R * W \quad (12)$$

Step 5
Determine the intuitionistic fuzzy positive and negative ideal solutions:

In this step, the intuitionistic fuzzy positive ideal solution (IFPIS) and intuitionistic fuzzy negative ideal solution (IFNIS) have to be determined. Let J_1 and J_2 be the benefit criteria and cost criteria, respectively. A^* is IFPIS and A^- is IFNIS. Then, A^* and A^- are equal to:

$$A^* = \left(r_1^{'*}, r_2^{'*}, \ldots, r_n^{'*}\right), r_j{}^{'*} = \left(\mu_j^{'*}, v_j^{'*}, \pi_j^{'*}\right), j = 1, 2, \ldots, n \quad (13)$$

and

$$A^- = \left(r_1^{'-}, r_2^{'-}, \ldots, r_n^{'-}\right), r_j{}^{'-} = \left(\mu_j^{'-}, v_j^{'-}, \pi_j^{'-}\right), j = 1, 2, \ldots, n \quad (14)$$

Where

$$\mu_j^{'*} = \left\{\left(\max_i \left\{\mu_{ij}^{'}\right\} j \in J_1\right), \left(\min_i \left\{\mu_{ij}^{'}\right\} j \in J_2\right\}, \quad (15)$$

$$v_j{}^* = \left\{\left(\min_i \left\{v_{ij}^{'}\right\} j \in J_1\right), \left(\max_i \left\{v_{ij}^{'}\right\} j \in J_2\right\}, \quad (16)$$

$$\pi_j^{'*} = \left\{\left(1 - \max_i \left\{\mu_{ij}^{'}\right\} - \min_i \left\{v_{ij}^{'}\right\} j \in J_1\right), \left(1 - \min_i \left\{\mu_{ij}^{'}\right\} - \max_i \left\{v_{ij}^{'}\right\} j \in J_2\right\}, \quad (17)$$

$$\mu_j^{'-} = \left\{\left(\min_i \left\{\mu_{ij}^{'}\right\} j \in J_1\right), \left(\max_i \left\{\mu_{ij}^{'}\right\} j \in J_2\right\}, \quad (18)$$

$$v_j^{'-} = \left\{ \left(\max_i \left\{ v_{ij}^{'} \right\} j \in J_1 \right), \left(\min_i \left\{ v_{ij}^{'} \right\} j \in J_2 \right\}, \tag{19}$$

$$\pi_j^{'-} = \left\{ \left(1 - \min_i \left\{ \mu_{ij}^{'} \right\} - \max_i \left\{ v_{ij}^{'} \right\} j \in J_1 \right), \left(1 - \max_i \left\{ \mu_{ij}^{'} \right\} - \min_i \left\{ v_{ij}^{'} \right\} j \in J_2 \right\}. \tag{20}$$

Step 6

Determine the separation measures between the alternative:

We can make use of the separation between alternatives on IFS, distance measures proposed by Atanassov [59], Szmidt and Kacprzyk [60], and Grzegorzewski [61] including the generalizations of Hamming distance, Euclidean distance and their normalized distance measures. After selecting the distance measure, the separation measures, S_i^* and S_i^-, of each alternative from IFPIS and IFNIS, are calculated:

$$S_i^* = \frac{1}{2} \sum_{j=1}^{n} \left[\left| \mu_{ij}^{'} - \mu_j^{'*} \right| + \left| v_{ij}^{'} - v_j^{'*} \right| + \left| \pi_{ij}^{'} - \pi_j^{'*} \right| \right] \tag{21}$$

$$S_i^- = \frac{1}{2} \sum_{j=1}^{n} \left[\left| \mu_{ij}^{'} - \mu_j^{'-} \right| + \left| v_{ij}^{'} - v_j^{'-} \right| + \left| \pi_{ij}^{'} - \pi_j^{'-} \right| \right] \tag{22}$$

Step 7

Make the final ranking

In the final step, the relative closeness coefficient of an alternative A_i with respect to the IFPIS A^* is defined as follows:

$$C_i^* = \frac{S_i^-}{S_i^* + S_i^-} where 0 \leq C_i^* \leq 1. \tag{23}$$

The alternatives are ranked according to the descending order of C_i^*'s score.

4 Numerical Examples

In this section, we describe how an AHP-IFT methodology is applied via an example. The criteria to be considered in the selection of projects are determined by an expert team from the decision group. In our study, we employ six evaluation criteria. The attributes which are considered here in the assessment of A_i (i = 1,2,. . .,6) are: (1) C_1 as benefit; (2) C_2,.., C_6 as cost. The committee evaluates the performance of alternatives A_i (i = 1,2,. . .,4) according to the attributes C_j (j = 1,2,. . .,6), respectively. Therefore, one cost criterion, C_1, and five benefit criteria, C_2,. . .,C_6 are considered. After preliminary screening, four alternatives

Table 3 The importance weight of the criteria

Criteria	DM_1	DM_2	DM_3	DM_4
C_1	*G*	*VG*	*VG*	*MB*
C_2	*VG*	*VG*	*VG*	*VG*
C_3	*MB*	*G*	*VG*	*MB*
C_4	*G*	*G*	*VG*	*G*
C_5	*G*	*VG*	*MB*	*G*
C_6	*MB*	*G*	*MB*	*VG*

Table 4 Linguistic term for rating DMs

Linguistic terms	IFNs
Very Important	(0.80, 0.10)
Important	(0.50, 0.20)
Medium	(0.50, 0.50)
Bad	(0.3, 0.50)
Very Bad	(0.20, 0.70)

A_1, A_2, A_3, and A_4, remain for further evaluation. A team of four DMs,—such as; DM_1, DM_2, DM_3, and DM_4—is formed to select the most suitable alternative.

The importance weight of the criteria given by the four DMS appear in Table 3.

The opinions of the DMs on the criteria are aggregated to determine the weight of each criterion.

$$W_{\{R_1,R_2,R_3,R_4,R_5,R_6\}} = \begin{bmatrix} 0,170 \\ 0,205 \\ 0,148 \\ 0,170 \\ 0,159 \\ 0,148 \end{bmatrix}^T$$

Also, the degree of the DMs on group decision, shown in Table 4, and the linguistic terms used for the ratings of the DMs, appear Table 5.

We construct the aggregated IFDM based on the opinions of DMs. The linguistic terms are shown in Table 6.

The ratings given by the DMs to six alternatives appear in Table 7.

The aggregated IFDM based on aggregation of DMs' opinions is constructed as follows:

Table 5 The importance of DMs and their weights

	DM_1	DM_2	DM_3	DM_4
Linguistic terms	Very important	Medium	Important	Important
Weight	0.342	0.274	0.192	0.192

Table 6 Linguistic terms for rating the alternatives

Linguistic terms	IFNs
Extremely good (EG)	[1.00; 0.00;0.00]
Very good (VG)	[0.85;0,05; 0.10]
Good (G)	[0.70; 0.20;0.10]
Medium bad (MB)	[0.50; 0.50;0.00]
Bad (B)	[0.40; 0.50;0.10]
Very bad (VB)	[0.25; 0.60;0.15]
Extremely bad (EB)	[0.00, 0.90,0.10]

$$R = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \end{array} \begin{array}{cccccc} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ (0.80,0.08,0.12) & (0.69,0.20,0.11) & (0.76,0.12,0.12) & (0.80,0.09,0.11) & (0.78,0.11,0.11) & (0.69,0.20,0.11) \\ (0.68,0.20,0.12) & (0.78,0.11,0.11) & (0.74,0.13,0.13) & (0.78,0.11,0.11) & (0.69,0.21,0.10) & (0.75,0.13,0.12) \\ (0.82,0.07,0.11) & (0.79,0.10,0.11) & (0.79,0.10,0.11) & (0.84,0.05,0.11) & (0.84,0.05,0.11) & (0.84,0.05,0.11) \\ (0.83,0.16,0.1) & (0.75,0.14,0.11) & (0.70,0.19,0.11) & (0.81,0.08,0.11) & (0.82,0.07,0.11) & (0.85,0.05,0.10) \\ (0.55,0.38,0.07) & (0.42,0.52,0.06) & (0.64,0.40,0.06) & (0.55,0.33,0.12) & (0.54,0.33,0.13) & (0.40,0.54,0.06) \\ (0.75,0.13,0.12) & (0.69,0.19,0.12) & (0.75,0.13,0.12) & (0.75,0.13,0.12) & (0.85,0.05,0.10) & (0.78,0.11,0.11) \end{array}$$

After the weights of the criteria and the rating of the projects were determined, the aggregated weighted IFDM was constructed as follows:

$$R = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \end{array} \begin{array}{cccccc} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ (0.136, 0.0136, 0.020) & (0.141, 0.041, 0.023) & (0.112, 0.018, 0.018) & (0.136, 0.015, 0.019) & (0.124, 0.017, 0.017) & (0.102, 0.030, 0.016) \\ (0.116, 0.034, 0.020) & (0.160, 0.023, 0.023) & (0.110, 0.019, 0.019) & (0.133, 0.019, 0.019) & (0.110, 0.033, 0.016) & (0.111, 0.019, 0.018) \\ (0.139, 0.012, 0.019) & (0.162, 0.021, 0.023) & (0.117, 0.015, 0.016) & (0.143, 0.009, 0.019) & (0.134, 0.008, 0.017) & (0.124, 0.007, 0.016) \\ (0.141, 0.027, 0.017) & (0.154, 0.029, 0.023) & (0.104, 0.028, 0.016) & (0.138, 0.014, 0.019) & (0.130, 0.011, 0.017) & (0.126, 0.007, 0.015) \\ (0.094, 0.065, 0.012) & (0.086, 0.107, 0.012) & (0.095, 0.059, 0.009) & (0.094, 0.056, 0.020) & (0.086, 0.052, 0.021) & (0.059, 0.080, 0.009) \\ (0.128, 0.022, 0.020) & (0.141, 0.039, 0.025) & (0.111, 0.019, 0.018) & (0.128, 0.022, 0.020) & (0.135, 0.008, 0.016) & (0.115, 0.016, 0.016) \end{array}$$

Then IFPIS and IFNIS are provided as follows:

$$A^* = \{(0.141, 0.012, 0.847), (0.162, 0.021, 0.817), (0.117, 0.015, 0.868), (0.143, 0.009, 0.848), (0.135, 0.008, 0.857), (0.126, 0.007, 0.867)\}$$

$$A^- = \{(0.094, 0.065, 0.841), (0.086, 0.107, 0.807), (0.095, 0.059, 0.846), (0.094, 0.056, 0.850), (0.086, 0.052, 0.862), (0.059, 0.080, 0.861)\}$$

The negative and positive separation measures based on normalized Euclidean distance for each alternative, and the relative closeness coefficient are calculated as Table 8.

Table 7 Ratings of the alternatives

Alternative	Criteria	DM_1	DM_2	DM_3	DM_4
A_1	C_1	*VG*	*VG*	*G*	*G*
	C_2	*G*	*VG*	*MB*	*MB*
	C_3	*VG*	*G*	*B*	*VG*
	C_4	*VG*	*VG*	*G*	*G*
	C_5	*VG*	*VG*	*MB*	*G*
	C_6	*G*	*VG*	*MB*	*MB*
A_2	C_1	*G*	*VG*	*MB*	*B*
	C_2	*VG*	*VG*	*G*	*MB*
	C_3	*VG*	*VG*	*B*	*B*
	C_4	*VG*	*VG*	*MB*	*G*
	C_5	*G*	*G*	*G*	*G*
	C_6	*VG*	*VG*	*MB*	*B*
A_3	C_1	*VG*	*VG*	*G*	*VG*
	C_2	*VG*	*G*	*G*	*VG*
	C_3	*VG*	*G*	*VG*	*G*
	C_4	*VG*	*VG*	*VG*	*VG*
	C_5	*VG*	*VG*	*VG*	*VG*
	C_6	*VG*	*VG*	*VG*	*VG*
A_4	C_1	*MB*	*G*	*MB*	*VG*
	C_2	*G*	*VG*	*G*	*G*
	C_3	*MB*	*VG*	*G*	*G*
	C_4	*VG*	*G*	*VG*	*VG*
	C_5	*VG*	*VG*	*G*	*VG*
	C_6	*VG*	*VG*	*VG*	*VG*

Table 8 Separation measures and relative closeness coefficient of each alternative

Alternatives	S^*	S^-	C_i^*
A_1	2,563	2,737	0.516
A_2	2,570	2,725	0.515
A_3	2,500	2,798	0.528
A_4	2,530	2,773	0.523

5 Conclusion

In this paper, AHP-IFT methodology is incorporated in selecting Supply Chain (SCM). The purpose of the study was to use an MCDM Method which combines AHP and IFT to evaluate a set of alternatives in order to reach the most suitable alternative. In the evaluation process, the ratings of each alternative, given by Intuitionistic fuzzy information, are represented as IFNs. AHP is used to assign weights to the criteria while IFT is employed to calculate the full-ranking of the alternatives. The AHP-IFT methodology was used to aggregate rating DMs.

Multiple DMs are often preferred rather than a single DM to avoid minimizes partiality in the decision process. Therefore, group decision making process for alternative selections considered effective. This is because it combines the idea of different DMs using a scientific MCDM method. In real life, information and performances regarding different settings are usually uncertain. Therefore, the DMs are unable to express their judgments on the best alternatives and/or criteria with crisp values, and such evaluation are very often expressed in linguistic terms, instead AHP and IFT are suitable ways to deal with MCDM because the contains a vague perception of DMs' opinions. A numerical example is illustrated and finally, the, results indicate that Among six alternatives with respect to six criteria, after using this methodology, the best ones are three, four, six, one, two and, five. The presented approach not only validates the methods, but also considers a more extensive list of benefit—and—cost oriented criteria suitable selecting the best. The AHP-IFT methodology has potential to deal with similar types of situations with uncertainty in MCDM problems.

References

1. Rouyendegh BD (2011) The DEA and intuitionistic fuzzy TOPSIS approach to departments' performances: a pilot study. J Appl Math 2011:1–16. doi:10.1155/2011/712194
2. Atanassov KT (1986) Intuitionistic fuzzy sets. Fuzzy Sets Syst 20:87–96
3. Gau WL, Buehrer DJ (1993) Vague sets. IEEE Trans Syst Man Cybern 23:610–614
4. Bustine H, Burillo P (1996) Vague sets are intuitionistic fuzzy sets. Fuzzy Sets Syst 79:403–405
5. Chen SM, Tan JM (1994) Handling multi criteria fuzzy decision-making problems based on vague set theory. Fuzzy Sets Syst 67:163–172
6. Hong DH, Choi CH (2000) Multi criteria fuzzy decision-making problems based on vague set theory. Fuzzy Sets Syst 114:103–113
7. Szmidt E, Kacprzyk J (2002) Using intuitionistic fuzzy sets in group decision making. Control Cybern 31:1037–1053
8. Atanassov KT, Pasi G, Yager RR (2005) Intuitionistic fuzzy interpretations of multi-criteria multi-person and multi-measurement tool decision making. Int J Syst Sci 36:859–868
9. Xu ZS, Yager RR (2006) Some geometric aggregation operators based on intuitionistic fuzzy sets. Int J Gen Syst 35:417–433
10. Liu HW, Wang GJ (2007) Multi-criteria decision-making methods based on intuitionistic fuzzy sets. Eur J Oper Res 179:220–233
11. De SK, Biswas R, Roy AR (2001) An application of intuitionistic fuzzy sets in medical diagnosis. Fuzzy Sets Syst 117:209–213
12. Szmidt E, Kacprzyk J (2001) Intuitionistic fuzzy sets in some medical applications. Lect Notes Comput Sci 2206:148–151
13. Szmidt E, Kacprzyk J (2004) A similarity measure for intuitionistic fuzzy sets and its application in supporting medical diagnostic reasoning. Lect Notes Comput Sci 3070:388–393
14. Li DF (2005) Multi attribute decision making models and methods using intuitionistic fuzzy sets. J Comput Syst Sci 70:73–85
15. Liu HW, Wang GJ (2007) Multi criteria fuzzy decision-making methods based on intuitionistic fuzzy sets. Eur J Oper Res 179:220–233
16. Xu ZS (2007) Intuitionistic preference relations and their application in group decision making. Inf Sci 177:2363–2379
17. Xu ZS (2007) Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making. Fuzzy Optim Decis Making 6:109–121

18. Xu ZS (2007) Models for multiple attribute decision making with intuitionistic fuzzy information. Int J Uncertainty Fuzziness Knowledge-Based Syst 15:285–297
19. Lin F, Ying H, MacArthur RD, Cohn JA, Barth-Jones D, Crane LR (2007) Decision making in fuzzy discrete event systems. Inf Sci 177:3749–3763
20. Xu ZS, Yager RR (2008) Dynamic intuitionistic fuzzy multi-attribute decision making. Int J Approximate Reasoning 48:246–262
21. Li DF (2008) Extension of the LINMAP for multi attributes decision making under Atanassov's intuitionistic fuzzy environment. Fuzzy Optim Decis Making 7:17–34
22. Wei GW (2009) Some geometric aggregation function and their application to dynamic multiple attribute decision making in the intuitionistic fuzzy setting. Int J Uncertainty Fuzziness Knowledge-Based Syst 17:179–196
23. Xu ZS, Cai XQ (2009) Incomplete interval-valued intuitionistic fuzzy preference relations. Int J Gen Syst 38: 871–886.
24. Li DF, Wang YC, Liu S, Shan F (2009) Fractional programming methodology for multi-attribute group decision–making using IFS. Appl Soft Comput J 9:219–225
25. Xia MM, Xu ZS (2010) Some new similarity measures for intuitionistic fuzzy value and their application in group decision making. J Syst Sci Syst Eng 19:430–452
26. Xu ZS, Hu H (2010) Projection models for intuitionistic fuzzy multiple attribute decision making. Int J Inf Technol Decis Making 9:267–280
27. Xu ZS, Cai X (2010) Nonlinear optimization models for multiple attribute group decision making with intuitionistic fuzzy information. Int J Intell Syst 25:489–513
28. Tan C, Chen X (2010) Intuitionistic fuzzy Choquet integral operator for multi-criteria decision-making. Expert Syst Appl 37:149–157
29. Xu ZS (2010) A deviation-based approach to intuitionistic fuzzy multiple attribute group decision making. Group Decis Negot 19:57–76
30. Park JH, Park IY, Kwun YC, Tan X (2011) Extension of the TOPSIS method for decision making problem under interval-valued intuitionistic fuzzy environment. Appl Math Modell 35:2544–2556
31. Chen TY, Wang HP, Lu HP (2011) A multi-criteria group decision-making approach based on interval–valued intuitionistic fuzzy sets: a comparative perspective. Expert Syst Appl 38:7647–7658
32. Xia MM, Xu ZS (2012) Entropy/cross-entropy based group decision making under intuitionistic fuzzy environment. Inf Fusion 13:31–47
33. Li DF, Cheng CT (2002) New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions. Pattern Recognit Lett 23:221–225
34. Liang ZZ, Shi PF (2003) Similarity measures on intuitionistic fuzzy sets. Pattern Recognit Lett 24:2687–2693
35. Hung WL, Yang MS (2004) Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance. Pattern Recognit Lett 25:1603–1611
36. Wang WQ, Xin XL (2005) Distance measure between intuitionistic fuzzy sets. Pattern Recognit Lett 26:2063–2069
37. Zhang CY, Fu HY (2006) Similarity measures on three kinds of fuzzy sets. Pattern Recognit Lett 27:1307–1317
38. Vlachos IK, Sergiadis GD (2007) Intuitionistic fuzzy information–applications to pattern recognition. Pattern Recognit Lett 28:197–206
39. Boran FE, Genç S, Kurt M, Akay D (2009) A multi-criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. Expert Syst Appl 36:11363–11368
40. Kavita SP, Kumar Y (2009) A multi-criteria interval-valued intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. Comput Sci 5908:303–312
41. Ye F (2010) An extended TOPSIS method with interval-valued intuitionistic fuzzy numbers for virtual enterprise partner selection. Expert Syst Appl 37(10):7050–7055. doi:10.1016/j.eswa

42. Boran FE, Genç S, Akay D (2011) Personnel selection based on intuitionistic fuzzy sets. Hum Factors Ergon Manuf Serv Ind 21:493–503
43. Boran FE, Boran K, Menlik T (2012) The evaluation of renewable energy technologies for electricity generation in Turkey using intuitionistic fuzzy TOPSIS. Ene Sou, Part B: Eco, Plan Pol 7:81–90
44. Boran FE (2011) An integrated intuitionistic fuzzy multi-criteria decision-making method for facility location selection. Math Comput Appl 16:487–496
45. Wang P (2009) QoS-aware web services selection with intuitionistic fuzzy set under consumer's vague perception. Expert Syst Appl 36:4460–4466
46. Shu MS, Cheng CH, Chang JR (2006) Using intuitionistic fuzzy set for fault-tree analysis on printed circuit board assembly. Microelectron Reliab 46:2139–2148
47. Gerogiannis VC, Fitsillis P, Kameas AD (2011) Using combined intuitionistic fuzzy set-TOPSIS method for evaluating project and portfolio management information system. IFIP Int Fed Inf Proc 364:67–81
48. Rouyendegh BD (2012) Evaluating projects based on intuitionistic fuzzy group decision making. J Appl Math 2012:1–16. doi:10.1155/2012/824265
49. Saaty TL (1980) The analytic hierarchy process. McGraw-Hill, New York
50. Saaty TL, Vargas LG (2006) Decision making with the analytic network process. Spring Science, LLC 1–23
51. Boroushaki S, Malczewski J (2008) Implementing an extension of the analytical hi*er*archy process using ordered weighted averaging operators with fuzzy quantifiers in ArcGIS. Comput Geosci 34:399–410
52. Lin L, Yuan XH, Xia ZQ (2007) Multicriteria fuzzy decision- making methods based on intuitionistic fuzzy sets. J Comput Syst Sci 73:84–88
53. Vahidnia MH, Alesheika AA, Alimohammadi A (2009) Hospital site selection using AHP and its derivatives. J Environ Manage 90:3048–3056
54. Zadeh LA (1969) Fuzzy sets. Inf Cont 8:338–353
55. Kahraman Ç, Ruan D, Doğan I (2003) Fuzzy group decision-making for facility location selection. Inf Sci 157:135–150
56. Rouyendegh BD, Erol S (2010) The DEA–FUZZY ANP department ranking model applied in Iran Amirkabir university. Acta Polytech Hungarica 7:103–114
57. Kahraman Ç, Ruan D, Ethem T (2002) Capital budgeting techniques using discounted fuzzy versus probabilistic cash flows. Inf Sci 42:57–76
58. Xu ZS (2007) Intuitionistic fuzzy aggregation operators. IEEE Trans Fuzzy Syst 15:1179–1187
59. Atanassov KT (1999) Intuitionistic fuzzy sets. Springer, Heidelberg
60. Szmidt E, Kacprzyk J (2003) A consensus-reaching process under intuitionistic fuzzy preference relations. Int J Intell Syst 18:837–852
61. Grzegorzewski P (2004) Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the Hausdorff metric. Fuzzy Sets Syst 148:319–328

The Assessment and Selection of Hedge Funds

Joaquin López Pascual

1 Introduction

Hedge funds are capturing the interest of institutional investors as well as high net worth individuals. Hedge funds have put in crisis traditional portfolio construction methods based on mean and variance analysis. The academic community has not hesitated to research this fascinating universe of principal importance searching for solutions and answers from investors.

Understanding hedge funds is not a very easy task. There are a number of complexities involved in investments related to hedge funds. Legal and compliance, operations, qualitative analysis and quantitative analysis, and technology related questions means that operational due diligence is a very important concept in the allocation to hedge funds. High level of interest has been concentrated on the hedge fund industry as a paradigm of alternative investments; however the hedge fund industry is a heterogeneous group. One way to classify hedge funds is according to the investment strategy used, each offering a different degree of return and risk. Their historical return distributions provide key information in order to understand the strategies behaviour.

In this Chapter, we briefly introduce the basic concepts about investment in hedge funds. We will provide some general definitions and introduction to hedge funds, their analysis and strategies. For a more extensive guide over hedge funds analysis please refer to López Pascual and Cuellar [1]: "Assessment and Selection of Hedge Funds and Funds of Hedge Funds" Working Paper N°5 2010, CUNEF. Finally we will focus on the benchmark hedge funds and the importance of hedge fund managers in the fund selection process.

In the front line of investment companies investing in emerging markets are hedge funds, introducing sophisticated investment mechanisms and instruments,

J.L. Pascual (✉)
Department of Finance, University Rey Juan Carlos Madrid, Madrid, Spain
e-mail: joaquinlopezpascual@gmail.com

© Springer International Publishing Switzerland 2015

F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_10

presenting the problems related to risk analysis in hedge funds and problems related to accounting or valuation of illiquid assets. Moreover, we introduce the various most common hedge fund strategies and their expected risks. Can we reach a market neutral investment strategy by investing in hedge funds? How can we build a diversified portfolio of hedge funds? How can we measure risk adjusted returns and total risk in hedge funds? These are some of the questions to which we aim to respond in this Chapter.

One of the most important attributes of a hedge fund is the ability to perform above a certain hurdle rate at all times no matter what market conditions prevail. This attribute has been called market neutral, which under no circumstances should be considered as neutral to the markets. As the LTCM (Long Term Capital Management) experience has demonstrated, there is no hedge fund that can be completely unaffected by a general adverse prevailing market condition. However, some managers are able to turn an adverse market condition into an opportunity, delivering extraordinary returns during market turmoil.

In general, it is considered that hedge funds have to be beta neutral or that the level of correlation with the performance benchmark of the market where the fund is involved should be as close to zero as possible. The principal function of the hedge fund in this conceptual frame would be of at least capital preservation in bear markets and capital appreciation in bullish markets.

This definition calls for reviewing the concept of absolute returns, which have been in the area of investment since the inception of hedge funds into the arena of investment vehicles. Recent research [2] explores the frontiers of alpha generation. It is considered that a portfolio manager is exposed to beta, but returns exceeding beta exposure can be attributed to manager's skills as measured by the alpha. However, as already mentioned, hedge fund managers are not always able to generate alpha and they are even sometimes not able to beat passively managed investment portfolios, such as index funds, which does not necessarily mean that they are not alpha generators given their non-directional investment style. As we know, beta can be obtained in the market to significantly cheaper prices than hedge fund fees, just by investing in an index replicating an investment portfolio or by using derivatives or, most recently ETFs, which are very liquid actively managed instruments and are able to provide a number of products for beta generation. Beyond the beta, the most important aspect in hedge fund selection is the manager's abilities to generate returns by his skills because, as demonstrated by the research mentioned above, there is no such thing as an absolute-return investor, but a relative return investor. It is correct to assess that a well managed hedge fund is one that has a zero or nearly close to zero beta coefficient, as we can observe in the Fig. 1, while enjoying a high degree of alpha in its portfolio returns.

The question is how an investor can be able to assess the level of alpha generation by a hedge fund manager. Analysing the track record of the fund is a possible answer. However, in doing so, investors should be aware that historical performance is not a guarantee of future returns. The consistencies between historical and future returns have to be carefully assessed considering a number of parameters that result in higher and consistent alpha creation. However, one should

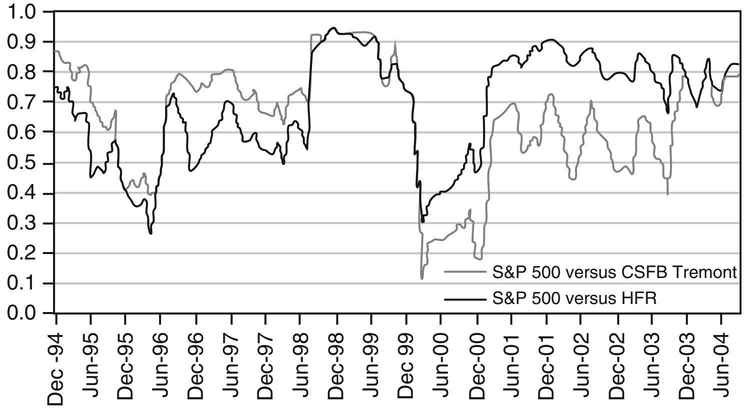


Fig. 1 Rolling 12 month correlation between the S&P 500 and the HFR equity hedge index and the CSFB tremont long/short equity index from the period 1993–2004 (*Source*: Bloomberg)

consider that linear factor models such as the ones developed by Markowitz or Sharpe are unable to capture hedge fund's nonlinear return features.

In line with this assessment, Fung and Hsieh [3] have developed a model based on asset-based style factors. These factors with statistical significance may not necessarily be associated to any strategy or specific investment style. The statistical clustering created by using principal component analysis (PCA) is able to group common risk and returns characteristics of the sample. This is very important because hedge funds are actively managed investment organizations, so timing and leverage are relevant influential factors of the investment style and strategy. The attractivity of the non-correlated returns generated by hedge funds bearing low or neutral beta and a high alpha should be assessed in the context of portfolio diversification. Kat [4] established that the undesired effects of hedge funds that are attributable to negative skewness and high kurtosis can be eliminated through the use of out-of-the-money put options or by investing in other hedging strategies. In this context, it is clear that hedge fund returns are not "superior", but different than returns generated by other asset classes. Needless to mention, a diversified hedge fund portfolio has for a retail investor a prohibitive cost, given the fact that the minimal investment in an average hedge fund is in the order of USD 1 million and a diversified portfolio should have about 10–15 underlying vehicles.

One way to classify hedge funds is according to the investment strategies employed. The strategies perform differently according to the economic cycle, each offering a different degree of return and risk. Therefore performance, generated in a specific part of an economic cycle, that seem to have achieved consistent high excess returns could underperform systematically once that the economic cycle changes. The returns generated by a hedge fund have to be understood in the context of the strategy used and the economic cycle. This implies a double problem for investors:

- The allocation strategy will depend on what investors are looking for. For instance, are they looking for a dynamic hedge to the equity market? In this case a negative beta strategy as "short bias" could be the right decision. It will be more complicated if the investors looks for absolute returns. In this case, the first decision is to decide between a passive or a dynamic approach. Another complicated choice would be a hedge fund that enhances portfolio efficiency.
- In the quest for the right hedge fund, a key factor to understand is the intrinsic behaviour of the strategy followed by the manager. Then, the investors have to look for a manager with an "edge" in the specific strategy.

2 Definitions

2.1 What Is a Hedge Fund?

Alternative investments usually refer to a number of investment vehicles that are not correlated with the financial markets. In general, whenever a reference is made to alternative investments, it is in reference to hedge funds. Although hedge funds are part of the alternative investment universe, they are not the only component. This universe of alternative investments is composed of a number of investment vehicles, predominantly products of financial innovation.

There is no an universally accepted definition of hedge fund. However, the common characteristics of the term hedge fund are; private investment fund that invest in a wide range of assets and employs a great variety of investment strategies. Due to their nature hedge funds have almost no restrictions in the use of derivatives, leverage or short-selling. This combination of capacity, instruments and flexibility in their investment decisions makes a significant difference relative to the more regulated, mutual funds.

Also, the combination of these resources has allowed hedge funds to exploit new market opportunities through investment strategies.

2.2 Investment Strategies

There is also no consensus regarding the number of investment strategies used by hedge funds. Financial technology evolves and the universe of investment assets is constantly growing. Therefore new investment strategies are continually developing to exploit market opportunities. Even hedge funds that invest in the same type of assets can try to make money taking exposure to different risk factors. For example a hedge fund investing in convertible bonds could be aiming to get equity, credit, volatility, liquidity, interest rate exposure, or a combination of several of them. The exposure to each of these factors could be exploited through different strategies.

Therefore, it is important to note that different strategies provide a different degree of return and risk.

2.3 Hedge Fund Indexes

Hedge funds have no formal obligation to disclose their results, however most of the funds release, at least monthly, their returns to attract new investors. With this information some data vendors have built performance hedge fund indexes, as well as sub indexes according to the fund strategy.

Some of these data providers are;

- Hedge Fund Research (HFR)
- Zurich Capital Markets
- CSFB Tremont
- Hennesse
- Tuna
- Barclays

2.4 Historical Return Analysis

The historical return analysis provides an important source of information for evaluating and understanding hedge funds investment styles.

Through explicit or implicit analysis we can try to explain the funds performances and to classify investment styles.

- *Explicit analysis*. The aim is to identify and measure the sensitivity of real factors that explain the historical returns. An example could be to model the returns as a linear function of various macro economic factors or indexes.
- *Implicit analysis*. The idea is to identify certain statistical factors that explain the historical returns. One the most used methods is the principal component analysis (PCA). The PCA ranks explanatory factors with the highest possible variance with the constraint that each one has to be orthogonal to the previous components.

In addition, comparing the time series returns of a hedge fund against the returns of its peer group will allow us to assess the investment skills of the manager.

From an investor's perspective, it is important to maintain a clear view of the risk exposure gained by a hedge fund investment in relation not only to the returns but also with the investment vehicle strategy. Different strategies yield not only different risk exposures but expose the investment to different risk classes. In this respect, it is important to conceptualize the risk. Some investors wrongly believe that by investing in bonds or in an investment fund, which invests in fixed income

Accounting risk	Fiduciary risk	Political risk
Bankruptcy risk	Hedging risk	Prepayment risk
Basis risk	Horizon risk	Publicity risk
Call risk	Iceberg risk	Regulatory risk
Capital risk	Interest rate risk	Reinvestment risk
Collateral risk	Knowledge risk	Rollover risk
Commodity risk	Legal risk	Spread risk
Concentration risk	Limit risk	Systemic risk
Contract risk	Liquidity risk	Taxation risk
Currency risk	Market risk	Technology risk
Curve construction risk	Maverick risk	Time lag risk
Daylight risk	Modeling risk	Volatility risk
Equity risk	Netting risk	Yield curve risk
Extrapolation risk	Optional risk	

Fig. 2 Partial listing of risk universe in relation to hedge funds (*Source*: Author based on CMRA)

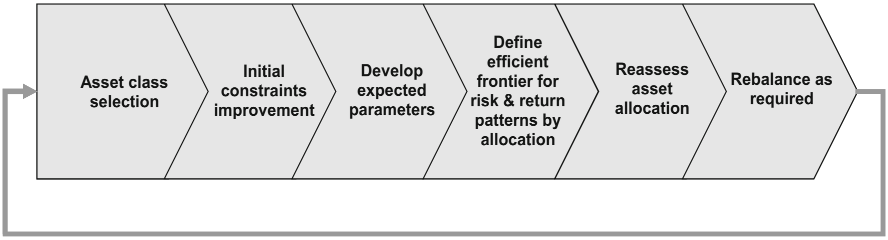


Fig. 3 Asset allocation process (*Source*: Author)

securities, they are only exposed to interest rate risk or credit risk. A brief list of possible risks that investors face in financial markets can be summarised in Fig. 2.

Institutional investors have traditionally used asset allocation as the core process in order to determine their investment strategy. The process of asset allocation is important; however, it does not take into account the dynamic changes in risk appetite and the changing dynamics of risk in the investment portfolio. Risk budget monitoring introduces a different dimension in the investment process as a function of volatility, correlation, and investment volume itself (Fig. 3).

Risk budgeting is a tool that should not be seen as an optimization process, because the optimization process in asset allocation uses a traditional mean-variance approach to efficiently allocate assets in a trade-off process of risk and returns. The objective of optimal investment risk management has to be such that it allows the investor to acquire less risk for a larger return or more return in exchange for the current risk exposure.

Other than the universe of possible risks mentioned in Fig. 2, hedge funds gain exposure through poor market liquidity, use of leverage, high turnover, heavy use of derivatives instruments, correlation to unrelated assets and transparency risk, to mention just a few. Risk measurement in traditional investment vehicles or asset classes seems to be a very straightforward exercise when compared with hedge funds.

Asset allocation is concerned with optimal asset combination, thus mathematically it is equivalent to a constrained optimization process. The process of asset allocation is much simpler than portfolio construction. Brinson et al. [5,6] established that more than 90 % of the variability of an investment portfolio is due to asset allocation. The advantage of the asset allocation process is that we resolve the optimization process at the asset class level instead of at the single security level. This is simpler because it is easier to estimate expected future returns at the asset class level than at the single security level and because the correlations are clearly established in order to build a diversified allocation. In this frame, we should consider investment in alternative funds as an asset class problem within the optimization process of asset allocation. Empirical research [7] has robustly established the virtues of including alternative assets in the allocation process given the low and even negative correlations with traditional asset classes.

One of the main challenges for investors is the poor transparency of hedge funds, which allows for very important risk misspecifications. The non-stationarity of risk due to the dynamic asset allocation of hedge fund managers is another challenge in risk measurement. Under these circumstances, it is very difficult to reduce measurement error to near zero. Identifying risk in a dynamic investment environment requires high frequency assessment and great accuracy. Factor analysis cannot only assist in identifying risk factors but also the rate of change of those factors. Factor analysis can determine the aggregate factors explaining investment returns. This analysis can be used either as forward risk modelling or as inverse modelling.

Forward risk modelling uses assumed pre-existing risk factors to assess the risk universe of the investment portfolio. If the investor has allocated investments to hedge funds using a convertible arbitrage strategy, we can assume risk factors correlated to fixed income securities as well as stocks, because such an investment strategy is exposed not only to risk factors related to the yield curve but also because when the hedge fund manager exercises his option in a convertible bond, he is automatically gaining exposure to stock market risks.

Static forward modelling (SFM) analyses the returns and finds the factors that can fit in the return's model. By definition, SFM is a replication strategy using future contracts or other trading assets. The modelling eliminates sequentially uncorrelated factors that assist in explaining the stream of returns. In practice, SFM is used as an early warning system for the fund of funds manager, because when the manager sees a new factor emerging which can affect the returns directly or indirectly, the manager should try to rebalance the portfolio eliminating the style drifting underlying position.

Inverse risk modelling uses principal component analysis (PCA) in order to analyse time series of returns and establishes all possible patterns with exposure to risk factors explaining the returns. Using the covariance matrix, the manager extracts the eigenvectors with maximum explanatory power in statistical terms, but because these eigenvectors are not the real economic variables such as actual gold price or the exchange parity of currencies, the manager must correlate the characteristics of those statistical factors to real factors. Interpretation is in this case absolutely critical but many times is not even possible.

Non-stationary or dynamic factor analysis takes into consideration relative changes of exposure along a time series of factors or combination of factors and their weights in explaining the returns of a portfolio. Managers have to take into account a sufficiently long horizon that explains the trade-off between risk and returns. When the factors and the returns converge in a time series, there is an alignment in the risk factors and the established strategy. Observation has to be maintained for a certain period of time because at a certain point the exposures could be subject to variations and diversions, letting the manager without knowledge of the new risk factors. The use of multi-scale correlation methods can assist portfolio managers in establishing the right time horizon for the analysis. Two significant risks in the analysis can be found. The first is that the time horizon of the assessment is too short and the point of divergence between the explaining factors and the portfolio return streams cannot be evaluated with a certain degree of accuracy, and the second is that the established time horizon is too long diluting the effects so much that the factors combination and the moment relation can hardly be visualized.

Detecting changes in correlations or non-stationarities across time is very useful for the investor because with assistance of this multi-scale correlation method, we can build an error map. If the error map becomes non-zero, it is because the correlation between the explaining factors and the returns has collapsed. Collapse in return attributions are a warning indicator that the fund manager has changed the strategy or is entering into a strategy shifting process that should trigger an immediate explanation by the fund manager to the investor about this change and the new risk factors implicated in such a strategy move. Another indicator of strategy shifting is sudden factor dispersion, which is given through introduction of new explanatory factors or alterations in the eigenvectors of the covariance matrix, which again we insist are statistical factors that have to be correlated to real economically relevant factors such as interest rate risk, volatility index (VIX), or gold price, to mention a few examples.

The practice of investment portfolio risk budgeting in the context of hedge fund management is to align risk budgeting with a coherent risk measurement methodology in order to obtain an appropriate risk amount. There are a number of variations of VaR methodologies of which the most utilized in the hedge fund industry is certainly CVaR [8, 9].

3 Investment Strategies and Indexes

We will use some of the HFR sub indexes as proxies for hedge fund strategies performance.

HFR have developed a series of benchmark indexes designed to reflect hedge fund industry performance by constructing equally weighted composites of constituent funds. The funds selected in each index are filtered through manager's due diligence and other qualitative requirements. The classification is done through statistical analysis, cluster analysis, correlation analysis, optimization and Monte Carlo simulations. This information is available on the HFR web site.

One of the main problems with the hedge fund indexes is the survivorship bias. Many hedge funds that were included at some point in the indexes might now not comply with the index requirements or might be defunct. HFR minimizes this problem by trying to receive a fund's performance until the point of the final liquidation of the fund.

HFR has created the following index classification.

Hedge Fund Strategy Classifications			
Equity Hedge	Event Driven	Macro	Relative Value
Equity Market Neutral	Activist	Active Trading	Fixed Income - Asset Backed
Fundamental Growth	Credit Arbitrage	Commodity: Agriculture	Fixed Income - Convertible Arbitrage
Fundamental Value	Distressed / Restructuring	Commodity: Energy	Fixed Income - Corporate
Quantitative Directional	Merger Arbitrage	Commodity: Metals	Fixed Income - Sovereign
Sector: Energy/Basic Materials	Private Issue/ Regulation D	Commodity: Multi	Volatility
Sector: Technology/Healthcare	Special Situations	Currency: Discretionary	Yield Alternatives: Energy Infrastructure
Short Bias	Multi-Strategy	Currency: Systematic	Yield Alternatives: Real Estate
Multi-Strategy		Discretionary Thematic	Multi-Strategy
		Systematic Diversified Multi-Strategy	

Source: www.hedgefundresearch.com

Based on these indexes we will show inherent characteristics of some of the most relevant hedge fund strategies and we will assimilate the historical returns of each investment strategy taking a long or short positions in plain vanilla options.

3.1 Ratios

The analysis of hedge funds performances through ratios is an easy and intuitive way to measure the efficiency of an investment. Although many researchers, Sharpe himself, study the deficiencies and limitations of the ratio, rating agencies and institutional investors include this ratio in their performance and risk measurements as appointed López Pascual and Cuellar [10].

The two most used ratios are; Sharpe and Sortino, both measure the excess returns of an investment per unit of risk. In the case of the Sharpe ratio, the unit of risk is calculated as the standard deviation of the investment returns. For the Sortino ratio, the unit of risk is measure as the standard deviation of the negative returns. In other words, a measure of excess return against downward price volatility.

The statistical characteristics of the hedge funds returns that we have described in the previous section result in overestimated Sharpe or Sortino ratios, therefore these ratios tend to overvalue the efficiency of hedge funds and drive to over allocation in this asset class. This technology has limitations therefore the results have to be understood in the context of the selected strategy and the inherent risks. In this direction, López Pascual and Cuellar [1] propose a complementary system for evaluating the inherent risks of each hedge fund through a radar visualization of strategy exposure.

4 How to Benchmark Hedge Funds. Should They Be Benchmarkable?

Over time, a number of database vendors have started to make their own index. The most important are those produced by: ABN Amro, Altvest (currently Morningstar), MAR (Managed Account Reports), Credit Suisse First Boston/ Tremont, Hedge Fund Research, Hennessee Group, Morgan Stanley Capital Indices, Standard & Poor, Greenwich Alternative Investments, and Barclay Hedge Fund Index. In the majority of cases, these indices are equally weighted with few exceptions. The most important exception is Credit Suisse/Tremont, which is capitalization weighted. The debate about which index is more appropriate continues in the hedge fund community, which considers that a capitalization weighted index does not necessarily reflect the hedge fund industry because hot money normally flows in higher intensity into big hedge funds, making the results of these funds more relevant than the rest. However, capitalization indices are the norm in the universe of equity and bonds.

The importance of a well designed and functioning index is fundamental for investors. A well designed index can be subject to elaboration by financial products, such as an index fund, and opens the possibilities in future for hedge fund related derivative investable instruments. However, since hedge funds are a result of an entrepreneurial process in the financial industry, it is very difficult to group funds under a certain style or strategy that could ideally fit them. Another problem is standardizing asset sizes given that AUM reporting is on a voluntary basis and it is in the interest of the fund manager to inflate the AUM by consolidating offshore and onshore AUM.

Reporting to an index is one of the few marketing tools that hedge funds have in order to reach retail investors. Moreover, the problem of the low transparency of the alternative investment vehicle is not very compatible with an investable index and if the index demands a high degree of transparency, it risks developing a non-representative benchmark. The financial community, retail and institutional investors, as well as financial institutions need to have a well designed and dependable benchmark. Credit Suisse/Tremont has imposed a number of conditions to its index's constituent, such as transparency, a minimum AUM of USD 10 million, audited reports, and various other conditions with the result that their index has become an investable index, which is the most important one in the industry.

Nevertheless, hedge fund indices as well as other indices suffer the same statistical imperfections. The most important deficiencies originate from the fact that reporting is voluntary and in most cases the results are unaudited. In this frame, logically the worst managers do not report to the index but also funds that are closed to new investors do not report because they are not marketing to new investors. Thus, the indices probably do not represent the returns of the best and the worst managers. Some data vendors do not consider certain investment styles as hedge funds and therefore those funds are also not included. This is the case of funds implementing the style of managed futures, which are not included by some data vendors in their indices.

Another important issue is the so-called survivorship bias. Since the average lifespan of a hedge fund is 5.5 years, there are a large number of hedge funds included in the indices that have been reporting and that are defunct either because they went out of business or for consolidation reasons. This situation can create inaccuracies in the rankings because it distorts both the positioning and the average results. Considering an example explained by Lhabitant [11], if we think about 10 marathon runners reaching the finish line out of the 100 that originally started the race, we would not know how to refer to runner number 10, since runner 10 could be the one placed last or one of the top 10 performers. Fortunately, most data vendors maintain defunct funds in their databases in a segregated manner. Moreover, the backfilled and instant history bias is related to hedge funds that now enter into the fund index and back report their historical results to the index, altering the past average returns of the index. Others deciding to not report historical returns also distort the overall picture.

There are a number of fundamental reasons why the financial community requires a purely dedicated and dependable hedge fund index. One of the most important is that non-dedicated indices are not a dependable benchmark that can be applied to the alternative investment universe, because hedge funds make extensive use of derivatives, short selling, and leverage while other asset classes do not present these characteristics. Moreover, we have seen that the universe of hedge funds is very heterogeneous and therefore styles are very mixed and hardly classifiable using a simplistic method. Because of routine trading in illiquid securities, many hedge funds present smoothing appraisals similar to real estate valuation indices and therefore non-dedicated indices are fully inappropriate to benchmark hedge fund managers. Institutional investors need to provide investors and trustees with reference bearings on the portfolio returns of their assets under management. An index also serves the purpose of monitoring behaviour of hedge funds in relation to other funds. Following Bailey [12], we can consider that the most important properties of an index are: simplicity and ease of understanding, replicability, comparability in terms of homogeneous open prices, taxes etc., and being representative of the investment vehicle's universe.

The problem with the benchmarks is that they normally operate with ratios that are more intuitive rather than being based on statistical or financial economic theory. Some of these statistical ratios are:

- Capture indicator: average of the captured performance
- Up capture indicator: the fund average return divided by the benchmark average, considering only the periods when the market is up. The ratio is best when at its highest value.
- Down capture indicator: the fund average return divided by the benchmark average, considering only the periods when the market is down. The ratio is best when at its lowest value.
- Up number ratio: measures the number of periods when the fund was up, during the time when the benchmark was up. The larger the ratio, the better.
- Down number ratio: measures the number of periods when the fund was down, during the time when the benchmark was down. The lowest the ratio, the better.
- Up percentage ratio: measures the number of periods that the fund outperformed the benchmark when the market was up.
- Down percentage ratio: measures the number of periods that the fund outperformed the benchmark when the market was down.
- Percent gain ratio: the number of periods that the fund was up.
- Ratio of number of negative months over total month: indicator of downside risk of the fund.

In order to overcome all the problems inherent to hedge funds indices, some investors have started to benchmark managers according to their peers and reaching excellent results. The dynamic trading strategies, leverage, and non-market assets make benchmarking certain investment styles very difficult to assess. In this respect, recent research by Gregoriou and Zhu [13] has implemented DEA (Data Envelopment Analysis) as a dependable statistical tool in benchmarking hedge funds and CTA to obtain ranking of funds according to different criteria. This system is ideal for benchmarking managers against their peers by visualising relative efficiency in the frame of the efficient frontier.

Hedge fund indices can be considered in general terms as valid approximation to benchmarks in the hedge fund industry despite the distortions induced by the lack of transparency and other factors. Moreover, investors should try to benchmark managers against peers under different criteria in order to obtain dependable rankings which could be used in efficient portfolio rebalancing or asset allocation.

5 Assessment and Selection of Managers

The importance of hedge fund managers in the fund selection process is the most relevant element in the hedge fund. Normally in the start-up fund, the manager, trader, and risk officer are the same person. Unless the manager has previous experience in hedge funds, he will be specialized solely in building long only portfolios. The investment organization relies on the talent of the manager for its success. As compensation, managers receive a management fee on the AUM and a participation on the returns of the fund, the so-called performance fee, which normally compensates the manager for the performance of the investment above a certain hurdle rate, which could be the returns of a risk free asset or Libor, for example. Managers usually offer to investors a high watermark clause. This clause allows investors to refrain from paying a performance fee when their investment does not outperform the returns obtained during the last compensation period. It also acts as an incentive to avoid losses, because no bonus would be paid to the manager until the investor has recovered the losses experienced during the last compensation period, which normally is quarterly. High watermark provisions could negatively influence a manager's portfolio architecture because of the interaction of convex compensation and a long horizon instead of the convexity of the compensation scheme.

Research has established that convex payoff structures incentivates risk shifting in investment organisations [14, 15]. Indeed, not penalizing the manager for his audacity or volatility acquired by the portfolio positions could be the result of the fee's structure in an investment company. However, empirical research conducted on a statistically significant sample of hedge funds [16] has proven that hedge funds invest a constant fraction of the assets in risk free assets with the rest invested in a mean variance asset portfolio.

The process of wealth transfer from the investor to the manager has to be justified by the risk adjusted returns being in excess of a certain rate, such as returns on risk free assets or by the level of alpha generated by the manager. Some hedge fund managers charge a performance fee only for returns exceeding a certain hurdle rate, such as 90 days T-Bills. Contrary to mutual funds, a hedge fund manager's compensation scheme depends deeply on the strategies implemented in order to generate alpha. The high watermark acts as both an incentive for the investor not only to invest in the fund initially but also to retain the investor in the fund. However, investors must consider the timeframe mismatch between performance fees accruing on a monthly basis and high watermarks being set at the beginning of the year on an annualized basis.

Many investors have criticised the fee structure of hedge funds using variations of the following example. When a fund charges 20 % performance fee and returns

USD 100 in the first year, then the fee charged is USD 20. In the second year, the fund manager produces minus returns of USD 100, but then does not charge any fees under the high watermark clause. The net performance of the fund over these 2 years was zero percent and the total performance fee paid to the fund manager was USD 20. The high watermark is the principal reason why hedge funds voluntarily close. When a hedge fund manager carries forward a negative balance and in the next period generates a good performance, due to the high watermark clause, the manager is unable to charge a performance fee and therefore the usual practise is to close the investment company and create a new one.

According to several hedge fund databases, hedge fund managers charge management fees between 0 and 6 % of the AUM and performance fees from 0 to 42.5 %. The factors deciding what fees may apply are uncertain, as neither volatility nor past performance nor AUM are explanatory variables. The role of the high watermark is to lock-in underperformance in the fund because once the assets decrease to a significant level below the high watermark, the manager cannot provide to the investor the perception of excess returns, and therefore he redeems his shares in the fund. Therefore, the decision to liquidate the fund under these circumstances can be considered as endogenous. This aspect constitutes one of the most important to have under consideration while assessing the massive collapse of hedge funds during the credit crisis that inflicted severe unexpected losses to different hedge fund strategies as explained by López Pascual and Cuellar [17].

An important difference between hedge fund managers and mutual fund managers is that in the mutual fund industry, the better performing the fund, the more funds this fund will manage to raise. Empirical research [18, 19] has demonstrated that with the success of mutual funds, the flow of investment is directed to both top performers and bad performers alike. On the other hand, research made on hedge funds [20] shows that top hedge fund performers are hesitant to accept new funds. This management behaviour among hedge funds is explained by the capacity constraints imposed by limited arbitraged-in expectations of opportunities where from a certain funding level, more investment increases the systematic risk given a certain investment strategy. The aforementioned research has regressed the net fund growth on the lagged return in cross section and established that hedge fund management technology is not linear and therefore does not accept more investment at any given time. Hedge fund managers managing large funds grow slower than smaller hedge funds.

Another important aspect to consider while assessing hedge fund managers in a fund is to determine if and to what extent the managers invest in the fund themselves. In the past, it has been considered an important aspect if the managers in a fund are investors as well. In general, there is the perception among investors that a fund with investing managers underperforms funds with no investing managers. The inferred conclusion from this statistic is that when managers start a fund and invest for others, they increase risk using leverage because they know that in the event of a fund failure, they will find employment in a financial institution (however, this principle may no longer apply in a severe bear market). When the manager is wealthy and invests his own money in the fund, he becomes risk averse and conservative in his investment decisions, taking less risk than other peers and therefore limiting the returns of the fund. On the other hand, investors can also be averse to funds where the members of the management do not invest, alleging that if the managers are not putting money into their own strategy, then why should someone else do so.

Managers can establish liquidity gates and lockup periods for the fund. This is an important aspect when a fund is investing in illiquid securities or during periods when large volumes of so-called hot money is flowing in the industry as well as in order to prevent the devastating effects of massive redemptions as experienced during the 2007–2008 credit crisis. The main problem related to hot money is that some strategies demand a certain time to produce returns, while hot money is only interested in short term performance while on the other side, panic scenarios are only interested in the flight to quality. Another of the principal mistakes made by money managers is to allocate short term liquid assets to long term strategies, which induces hedge fund managers to provision for eventual redemptions reducing the amount of total investment or increasing leverage and therefore risk in order to accommodate the demands of hot money.

Some strategies using derivative or fixed income instruments require certain maturity terms and permitting some investors to redeem before maturity of the investment could affect severely the NAV calculation, which could result in benefiting short term investors at the cost of long term investors. In the event of collective simultaneous redemptions, hedge fund managers impose liquidity gates where if certain quantities of investors wish to redeem, they can only obtain a certain percentage of their investment, having to wait for a certain time for the remainder, which is subject to the end calculation of the NAV or the audited balance sheet. Excepted from the reminder are investments in illiquid assets that the manager can allocate within his discretionary mandate and segregate from the portfolio in a sidepocket or private equity like structure, which allows for protection from massive redemptions during a period that can last up to 8 years and represents normally between 5 and 10 % of the total assets under management. However, reporting on such positions allocated to private equity like structures normally lacks the kind of transparency for the investor that private equity fund managers provide to their investors. In the same way, funds of hedge funds managers are forced to impose or pass along the same redemption conditions of their constituent underlying funds to their investors in order to be able to not dislocate investments. In the context of duration of the investment and lock-up periods, we have to consider recent research [10], where hedge funds establishing longer lock-up periods perform significantly better than funds offering none or shorter lock-up periods.

References

1. López Pascual J, Cuellar RD (2010) Assessment and selection of hedge funds and funds of hedge funds. Working Paper N 5 2010, Cunef

2. Waring MB, Siegel LB (2006) The myth of the absolute-return investor. Financ Anal J 62 (2):14–21
3. Fung W, Hsieh DA (2001) The risk in hedge fund strategies: theory and evidence of trend follower. Rev Financ Stud 41:313–41
4. Kat HM (2005) Integrating hedge funds into the traditional portfolio. J Wealth Manage 7 (4):51–57
5. Brinson GP, Hood LR, Beebower GL (1986) Determinants of portfolio performance II: an update. Financ Anal J 42(4):35–43
6. Brinson GP, Singer BD, Beebower GL (1991) Determinants of portfolio performance II: an update. Financ Anal J 47(3):40–48
7. Lintner J (1983) The potential role of managed commodity-financial futures accounts (and/or funds) in portfolios of stocks and bonds. Presentation at the Financial Analysts Federation
8. Rockefellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J Risk 2:21–41
9. Rockefellar RT, Uryasev S (2001) Conditional value-at-risk for general loss distributions. Research Report 2001–2005. ISE Department, University of Florida
10. López Pascual J, Cuellar RDR (2007) The challenges of launching, rating, and regulating funds of hedge funds. J Derivatives Hedge Funds 13(3):247–262
11. Lhabitant F-S (2004) Hedge funds: quantitative insights. John Willey & Sons, Hoboken, NJ
12. Bailey JV (1992) Are manager universes acceptable benchmarks? J Portfol Manage 18:9–13
13. Gregoriou G, Zhu J (2005) Evaluating hedge fund and CTA performance, data envelopment analysis approach. John Wiley & Sons, Hoboken, NJ
14. Carpenter J (2000) Does option compensation increase managerial risk appetite? J Financ 21:2311–2331
15. Ross SA (2004) Compensation, incentives, and the duality of risk aversion and riskiness. J Financ 59:207–225
16. Panageas S, Westerfield MM (2005) High-water marks: high risk appetites? Convex compensation, long horizons and portfolio choice. Working Paper 03–06, The Rodney L. White Center for Financial Research, The Wharton School, University of Pennsylvania
17. López Pascual J, Cuellar RD (2008) Can the hedge fund regulation limit the negative impact of a systemic crisis. Universia Bus Rev 20:42–53
18. Sirri E, Tuffano P (1992) The demand for mutual fund services by individual investors. Working Paper, Harvard Business School
19. Chevalier J, Ellison G (1995) Risk-taking by mutual funds as a response to incentives. NBER Working Paper No. 5234
20. Goetzmann W, Ingersoll J Jr, Ross S (2003) High-water marks and hedge fund management contracts. J Financ LVIII(4):1685–1718

Functional Data Analysis with an Application in the Capital Structure of California Hospitals

Feng Mai and Chaojiang Wu

1 Introduction

Functional Data Analysis or FDA is a set of statistical tools for analyzing data consisting of random functions (or surfaces for two-dimension functions), where each function is considered one sample element. FDA is a relatively new technique, which was first coined in Ramsay and Dalzell [1] and systematically introduced in Ramsay and Silverman [2]. The random functions are typically considered to come from an underlying smooth stochastic process and observed data are considered its realizations. FDA offers a set of tools for analyzing the data generated from such stochastic processes. One example is the growth study discussed in Ramsay and Silverman [2], where the heights of young girls are measured at 31 different ages unevenly, with an annual interval or from 2 to 8 years. The measurements on each child are observed in discrete values, but they reflect a smooth growing process that in principle can be measured at any frequency. Thus the first step in functional data analysis is to uncover the underlying function. This functional treatment offers the advantages of easily dealing with missing values, or uneven observations over time. It also smooths out the noise or measurement errors, which can be quite helpful in real data explorations.

Because of the difficulty to parameterize the processes, the uncovery of the underlying functions and models of such functions are usually nonparametric in nature, which allows for flexible modeling. In Sect. 2, we will introduce some nonparametric smoothing techniques and mathematical notations in approximating

F. Mai (✉)
Department of Operations, Business Analytics and Information Systems, University of Cincinnati, Cincinnati, OH, USA
e-mail: maifg@mail.uc.edu

C. Wu
Department of Decision Sciences, Drexel University, Philadelphia, PA, USA
e-mail: cw578@drexel.edu

© Springer International Publishing Switzerland 2015
F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_11

the functions. This nonparametric approach does not assume *a priori* parametric form of the function so the number of parameters could potentially go to infinity. This nonparametric feature makes FDA distinctly different from multivariate analysis, where finite-dimensional random vectors are delt with. The functional approach also has the advantage of studying the dynamics of the system by employing the functional derivatives and differential equations. The infinite dimension of functions will not pose any difficulty. This is due to the bases expansion approach which we will introduce later, where the infinite dimension function parameters are practically stored with the finite dimensional bases and their coefficients. The dimension of the bases may not depend on the observation frequency, which will increase the dimensions of analysis in traditional multivariate analysis. With the increasing volume of data from more frequent sampling and longer observations, and changing dynamics of the system, functional data analysis offers many advantages over traditional statistical tools.

The goals of functional data analysis lie in three major categories: exploratory, confirmatory and predictive. Functional data analysis is particularly suitable for exploratory analysis, where one is usually interested in finding new patterns as well as recovering known and obvious features. This is especially so when the measurements are subject to measurement errors, or the measurements are sparse but certain properties hold (such as monotonicity). Exploratory functional data analysis could potentially work well with Data Mining techniques, where both techniques focus on the data at hand, with less interest in inferential statement about the population characteristics. On the other hand, confirmatory analysis tends to make inferential statements about the population characteristics, and resembles a lot like the linear regressions used in many fields to confirm certain theories. Predictive functional data analysis tends to focus on the prediction of new observations, based on the data at hand. In Sect. 3 we will use an example to illustrate the techniques such as functional principal analysis and clustering, mainly as an exploratory tools, and functional linear models, often used in a confirmatory fashion. The functional prediction is less commonly seen but it can turn out to outperform traditional statistical models, see e.g., Sood et al. [3].

As described above, in modern business analytics, functional data analysis has several advantages that set it apart from more traditional techniques. First of all, because of the functional approach, missing values, observations of different frequencies, and observations with measurement errors can be handled with ease. Second, as Jank and Shmueli [4] point out, FDA enables us to study the trend, processes and dynamics of individual units over time. This is important as longitudinal data and panel data are becoming more common. Moreover, FDA can be coupled with more traditional statistics and econometrics models and thus offer researchers the familiar frameworks to work with. Therefore, in several recent academic papers from various business disciplines FDA has shown promising results.

The rest of the book chapter is organized as follows. We describe the basis expansions method of FDA in Sect. 2. In Sect. 3, we review some of recent studies.

In Sect. 4 we study the changes in the capital structures of California hospitals using FDA and illustrate FDA in a more detailed manner.

2 Function Representations Using Basis Expansions

Since functional data analysis works with functions, it is important to know how raw data can be turned into smooth functions. Consider the discretely recorded *n* observations of y_j at time t_j, where y_j could have measurement errors. Time t_j is the continuum over which the functional data are observed. In general, t_j can be time, space, or frequency and so forth. The observed y_j is assumed to come from the underlying smooth function *f(t)*, that is

$$\boldsymbol{y} = f(\boldsymbol{t}) + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$, $f(\boldsymbol{t})$ and $\boldsymbol{\varepsilon}$ are all column vector of length *n*, $\boldsymbol{\varepsilon}$ being the noise or measurement error term. Smooth functions are functions such that one or more derivatives exist. Thus the derivatives can be estimated from the function once the function *f(t)*, the trajectory of the quantity of interests, is uncovered from the data. This is particularly useful in the study of system dynamics, in which velocity and acceleration, corresponding to the first and second derivative respectively, are often subjects of interest. Functional data analysis shows clear advantages in this regard because the estimation of derivatives can be rather noisy by using discretized approach such as first differences.

In functional data analysis, the function *f(t)* is usually represented by the method of basis expansion. More specifically, any function can be approximated arbitrarily well by a weighted sum of a large number of basis functions, which are mathematically independent of each other. That is, for given basis series $B_k(t), k = 1,\ldots,K,$ the smooth function *f(t)* can be approximated by a linear combination of the bases

$$f(t) = \sum_{k=1}^{K} \beta_k B_k(t)$$

where β_k are the coefficients that define a function for the given basis $B_k(t)$. Thus the potentially infinite dimensional arbitrary function can be represented with finite dimensional parameters β_k. In functional data analysis, the number of bases or the dimension of the expansion *K* is left to be determined by data as to better reflect the characteristics of the data. This dimension may also depend on which basis functions to choose. In functional data analysis, there are two most commonly used bases: Fourier basis (for periodic data), and B-spline basis (for non-periodic data), while other bases are available.

For periodic data, Fourier bases or Fourier series are used to approximate the function *f(t)*

$$f(t) = \beta_0 + \beta_1 \sin(\omega t) + \beta_2 \cos(\omega t) + \beta_3 \sin(2\omega t) + \beta_4 \cos(2\omega t) + \ldots$$

where $2\pi/\omega$ is the period. For open-ended data, B-spline basis function is used. The B-spline basis function is a piecewise polynomial function of degree p, with the breakpoints at $t_1, t_2, \ldots, t_b$. The number of parameters to define a B-spline function is $p + b - 1$, i.e., order of the polynomials (one more than the degree p) plus number of interior breakpoints. Figure 2 shows the spline bases of order four with three interior breakpoints equally spaced in the time period 2001–2012. Other useful bases include wavelets bases, exponential bases, power bases, polynomial bases, constant basis, and monomial basis and so forth.

Once the functional form is expressed with the basis expansion, some key FDA methods can be applied. These methods include functional principal component analysis, functional clustering, and functional linear regressions, each of which needs additional mathematical representations. However, because these techniques are in principle similar to the multivariate statistics and linear regression, we take a practical approach in the Sect. 4 by illustrating the methods using existing software packages. The readers are referred to Ramsay and Silverman [2] for detailed mathematical treatments.

3 Recent Applications of FDA in Business Studies

This section is not an exhaustive review of FDA literature, rather we would like to highlight several typical applications of FDA in various business fields.

3.1 Production and Operations

Ramsay and Ramsey [5] use FDA to analyze the dynamics of nondurable goods production. They study the time series of U.S. monthly non-seasonally adjusted non-durable good production since January 1919. They propose a linear differential equation model with time varying coefficients to explain the seasonal variations. They show that seasonal components in the time series can be modeled by a third-order continuous time smooth deterministic differential equation. They reveal interesting variation in the dynamics of the system over major events such as the depression and war.

Abraham et al. [6] propose a functional clustering procedure that consists of two stages: fitting the functional data by B-splines and partitioning the estimated model coefficients using a k-means algorithm. Strong consistency of the clustering method is proven and a real-world example from cheese production industry is provided to demonstrate the procedure.

Gastón et al. [7] illustrate how FDA can be used in the simulation of arrival processes. In particular, the authors use FDA to estimate the cumulative mean function of a non-homogeneous Poisson Process. As an application, the time-varying arrival process for patient arrivals to a primary health center during 150 days is modeled using FDA.

3.2 Marketing and Sales

Bapna et al. [8] apply the functional linear regression modeling to study the price formation process in online auctions. A sample of 1009 auctions from Ebay.com is used in the study. Covariates include seller's choices such as starting price, reserve and duration, competition such as number of bidders and other market characteristics.

Wang et al. [9] further Bapna et al. [8]'s study and propose a dynamic auction forecasting model based on FDA. Price dynamics, price lags, and information related to sellers, bidders, and auction design are used in the model to forecast future price dynamics. The model allows forecasting of an ongoing auction using unevenly spaced bid data.

Sood et al. [3] propose an FDA model as an alternative to the Classic Bass model for market penetration of new products. The data is gathered from 21 products across 70 countries, for a total of 760 categories. The FDA model is able to integrate information of new products in other categories and countries in order to predict the market penetration process of a new product. Functional Principal Components, Regression and Clustering are used in the study. The FDA model is shown to have better prediction accuracy than the Classic Bass model.

Foutz and Jank [10] propose functional shape analysis (FSA) of virtual stock markets (VSMs) to address the pre-release demand forecasting in motion picture industry. FSA can identify small number of distinguishing shapes such as the last-moment velocity spurt. These functional shapes can carry information about movies' future demand and produce dynamic forecasts. The results show that the FSA framework leads to improved forecast accuracy compared with other methods.

4 Case Study: Capital Structure of California Hospitals

This section illustrates applications of the functional data analysis. In particular we use functional clustering and functional regression methods to analyze the California hospital data. We provide readers with the essential code needed to replicate the analysis. We use the *fda* library in R 3.0.2 to conduct our analysis. The library is developed by Ramsay and Silverman [2]. A detailed description the package can be found in (Graves et al. [11]). The following command can be used to install and load the library in R:

```
install.packages('fda')
library('fda')
```

4.1 Data and Preparation

Our dataset contains name, facility number, ownership type, total assets, and equity of over 300 California hospitals from 2001 to 2012. The original dataset is obtained from the Office of Statewide Health Planning and Development (OSHPD) of the State of California. Per regulation, all licensed hospitals in California must submit the data on utilization, finance and other metrics to the OSHPD. All data can be downloaded from the official website of OSHPD (www.oshpd.ca.gov). Excluding the hospitals that changed ownership, merged or acquired by other entities, and ones with too many missing records, we retain 296 hospitals in our sample.

4.2 Construction of Functional Curves

Our original data are the reported equity and total assets of hospitals during the 12 years. As the first step, we use functional curves to represent the dynamics of each hospital's equity ratio. The assumption is that there is an underlying continuous, smooth curve which describes how equity ratio changes over time. In certain hospitals, especially not-for-profit hospitals, the equity could be negative, which means the hospital owes more debts than its assets. We define the capital structure as the equity/asset ratio, or simply the equity ratio. A smaller number means higher portion of debt, and negative number means the debt has outnumbered the asset. Thus this equity ratio reflects the equity level in relation to its asset. Consider Fig. 1 which displays the equity/assets ratios for four selected hospitals. We can see that the equity/assets ratios in each of these four hospitals have heterogeneous trends which can be described by various continuous curves. In the top left panel, the hospital had a relative stable equity ratio until 2010 then experienced a sudden drop, indicating the recent deterioration of financial situations. The hospital in the top right panel had a stable declining trend. Notice the negative ratios reflect the troubled financial status of the hospital by having more debts than assets. This troubled situation has becoming worse over the past years. The bottom left panel describes a hospital with an increasing ratio starting from 2008, while in the bottom right panel the hospital has an inversed U-shape trend.

In order to discover the underlying functions that generate the equity ratios, we start by fitting the data with functional curves. We use the basis expansion method discussed in Sect. 2. Specifically, for hospital i, $i = 1, \ldots, 296$, we use the linear combination of basis functions with coefficients β_{ik} to represent a function of equity ratio of time $ER(t)_i$. That is

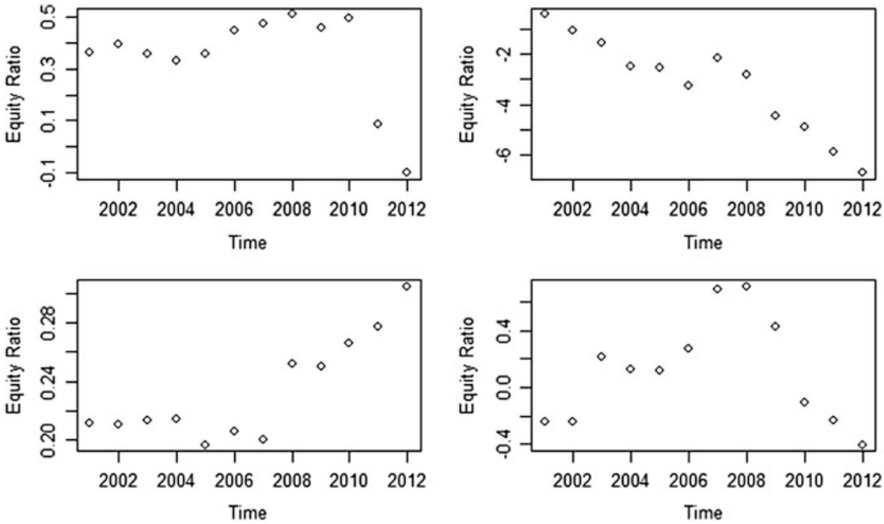


Fig. 1 Equity ratios of four hospitals

$$ER(t)_i = \sum_k \beta_{ik} B_k = \boldsymbol{\beta}_i^{'} \boldsymbol{B}$$

Note that the sample of 296 hospitals shares a common set of B_k, but each will have a different coefficients vector β_{ik}.

We can choose the appropriate basis systems according to the desired functional shapes. For example, if the data is periodic then we can consider the Fourier basis. The *fda* package includes several types of basis functions such as constant basis, monomial basis, Fourier basis, and B-spline basis. In our study we use B-spline basis because of the lack of periodic feature in the data and the numerical stability and flexibility of B-splines. We refer the readers to Ramsay and Silverman [2] for discussions in selection of bases in functional data analysis, and De Boor et al. [12] and Schumaker [13] for more detailed theory of splines.

Figure 2 plots a set of B-spline bases that we use to represent the smooth functions, where the bases of order four and three equally spaced interior breakpoints are used. With this B-spline basis, the time interval of interest from 2001 to 2012 is broken into four equally spaced sub-intervals and within each interval a polynomial of order four is fitted. The *fda* package uses term *order* (degree + 1) to describe the order of polynomials. The default order of the B-spline in *fda* is 4, meaning that piecewise cubic polynomials are used. Although it is possible to use different order of polynomials in each sub-interval, in our study we will keep the order constant. Figure 2 is generated using the following code:

```
basis=create.bspline.basis(rangeval = c(2001,2012), nbasis=7,
norder=4)
plot(basis)
```

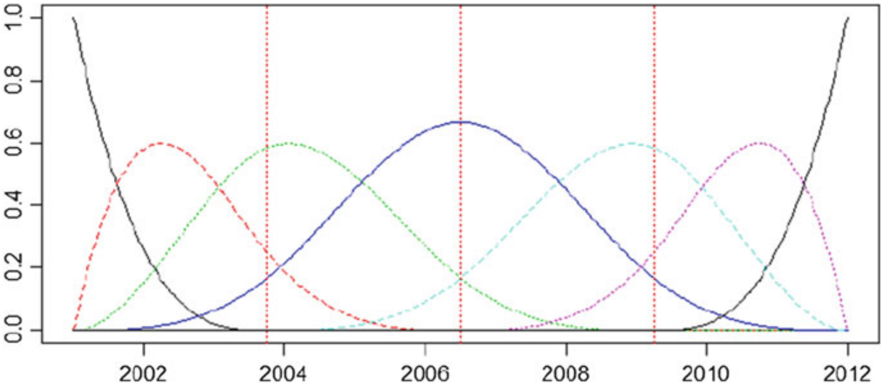


Fig. 2 B-spline bases used to represent smooth equity ratio function. Here B-splines of order four with three equally spaced interior breakpoints are used

The argument nbasis specifies number of basis functions, which is equal to the sum of order and number of interior breakpoints or knots. The argument rangeval defines the interval over which the functional data can be evaluated, in our case from 2001 to 2012.

Other than number of knots b and order of polynomial p, another parameter that can have an impact on the resulting curves is the smoothing penalty λ. With a non-zero λ, the coefficients $\boldsymbol{\beta_i}$ are called penalized spline estimators. These parameters combined can balance between the goodness of fit of the curves and over-fitting: order of the polynomials determines the accuracy in derivatives, number of knots decides number of breaks, and smoothing parameter λ specifies how much we wish to penalize curvatures in the model. Figure 3 shows the effect of different λ on how data is represented using a functional curve. When λ is small most of the curvatures are preserved in the fitted curve, while with very large λ can eliminate the curvatures. Using an extremely large λ will result in simple linear regression fit. In functional data analysis, although some automatic method of choosing the right λ is available, sometimes researchers still need to choose desired λ by some judgment on the level of desired smoothness.

As an example, the curve fitted using the penalized B-spline with $\lambda = 0.1$ can be generated using the code below. The fdPar function creates a functional parameter object that specifies the basis system and the smoothing parameter. Then a single observation y is transformed into a functional data object using smooth.basis according to the options defined in functional parameter object. Lastly the plot and lines commands are used to plot the fitted smoothed curve.

```
fdpar = fdPar(fdobj = basis,lambda = 0.1)
sb = smooth.basis(argvals = 2001:2012, y=
as.numeric(hostpital[,5]), fdParobj=fdpar)
plot(x=NULL,y=NULL,ylim=c(-0.2,0.5),xlim=c(2001,
2012),ylab="Equity Ratio", xlab="Year")
lines(sb)
```

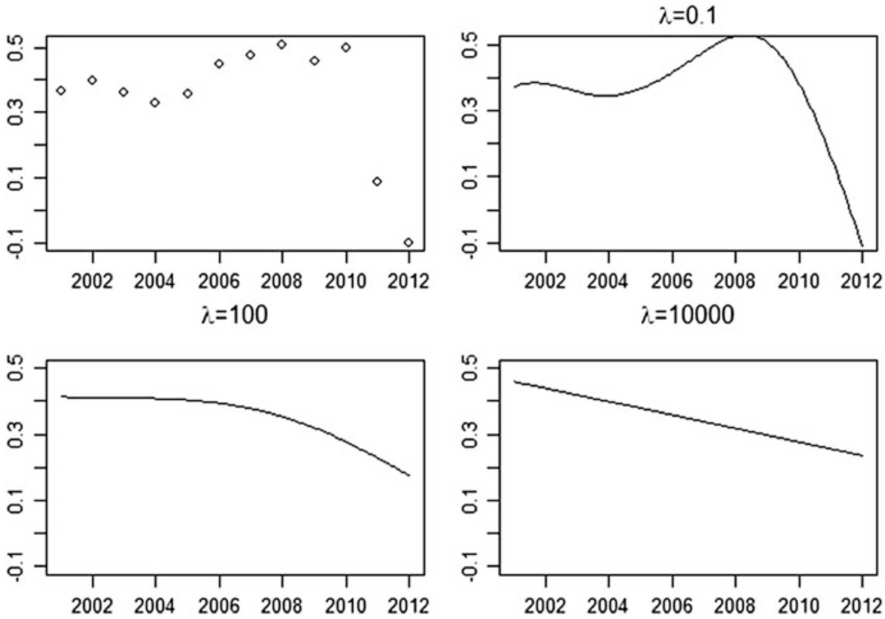


Fig. 3 Effect of smoothing penalty on fitted curves

One data-driven way of determining λ is to use the generalized cross-validation (GCV) developed by Craven and Wahba [14]. The GCV is defined as:

$$GCV(\lambda) = \left(\frac{n}{n - df(\lambda)}\right)\left(\frac{SSE}{n - df(\lambda)}\right)$$

A simple grid search procedure as follows can be used to identify the "optimum" λ that balances the smoothness and closeness to the raw data.

```
loglambda = seq(−5,5,0.1)
gcv = cbind(loglambda, NA)
for (i in 1:length(loglambda)){
  gcv[i,2]  =  sum(lambda2gcv(log10lambda=loglambda[i],argvals  =
2001:2012, y= hostpital, fdParobj=fdpar))
}
plot(gcv)
```

The dynamic of the estimated functional curve can now be further studied. For example, we can analyze its higher-order derivatives, as long as the polynomials

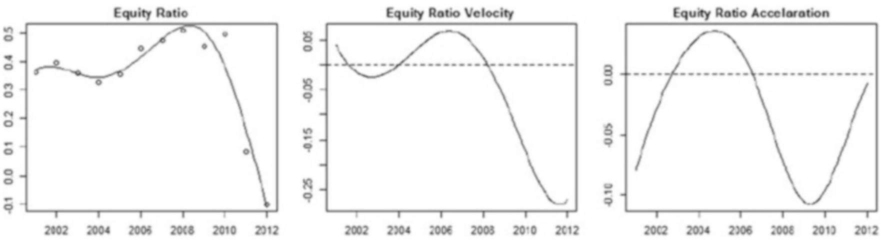


Fig. 4 Functional curve, velocity, and acceleration of equity ratio of a hospital

used to construct the basis system have sufficient order. The deriv.fd function can be used to calculate the first derivative (velocity) and second derivative (acceleration) of a functional curve as displayed in Fig. 4.

4.3 Functional Clustering

In this subsection, we demonstrate how we can group the hospitals according to how their equity ratios change during the time period. Cluster analysis is a method that can be used to find natural groupings in multivariate data sets. In our case, each hospital's equity ratio has a functional curve with infinite dimension. After summarizing each curve E_i using its decomposition in B-spline basis, each E_i can be represented at a lower dimension using their coefficients. Therefore, we can group the functions by partitioning the model coefficients using a clustering procedure on the $\boldsymbol{\beta}_i$ s.

Before we start, we standardize each hospital's equity ratio from year 2001 to 2012 since we would like to cluster the hospitals according to their trends rather than the numerical value of their equity ratios. After recovering the functional objects with 7 basis functions using the *fda* package, we proceed to work with the 296×7 coefficient matrix that can represent the functional curves.

Different clustering methods can be applied at this stage. We perform a *k*-means clustering on the coefficients because of the simplicity and popularity of the *k*-means algorithm. In *R* the analysis can be done using the kmeans function. We use the Duda index [15] to determine that the best number of clusters for this data set is $k = 4$. The cluster means and 95 % confidence intervals are computed by bootstrapping the coefficients within each cluster. Figure 5 illustrates the four cluster solutions for the 296 hospitals.

Examining Fig. 5 yields some interesting observations. The functional clustering results indicate that most changes in equity ratios occurred between 2006 and 2008. Hospitals in Cluster 2 had a U-shaped curve with bottom located around 2007. Meanwhile hospitals in Cluster 3 and Cluster 4 all had obvious inflection point on equity ratios during the same period. Without looking at the functional clustering results, it is difficult to use traditional econometrics models to detect these changes,

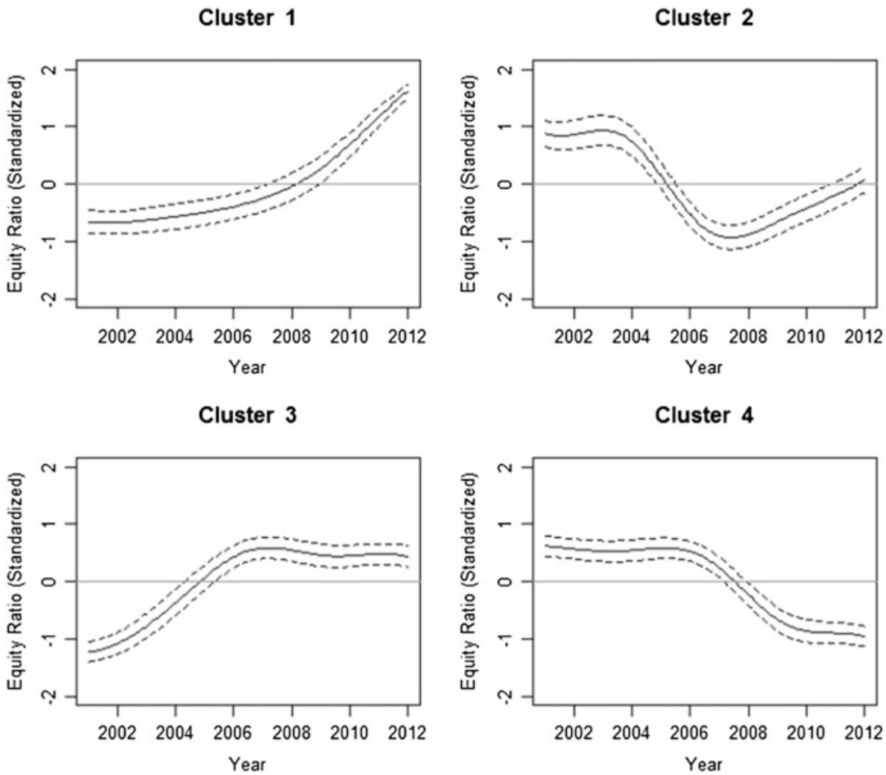


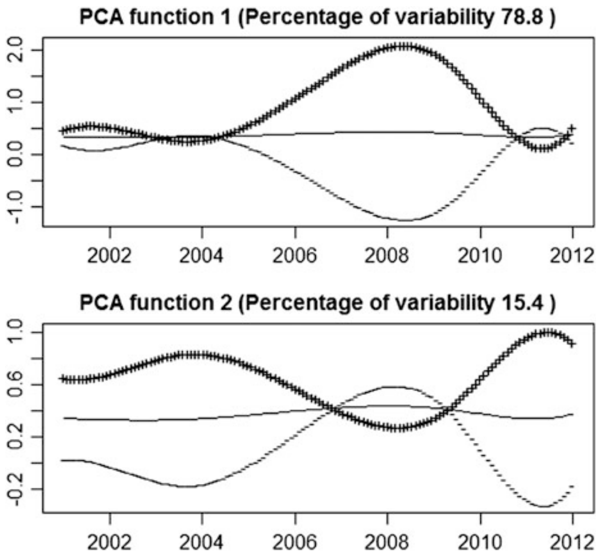
Fig. 5 Cluster means and 95 % confidence intervals

as these groups had various change directions and non-linear dynamics. Based on the exploratory analyses, we may frame follow-up research questions such as the policy and economic factors that lead to these changes, possible consequences in terms of healthcare outcomes, and intra-cluster characteristics differences of the hospitals in sample.

4.4 Functional Principle Component Analysis

As a method to conduct exploratory analysis on variability in multivariate data, Principal components analysis (PCA) is usually used to reduce dimensionality of data through eigenvalue decomposition of the covariance matrix and finding the few orthogonal directions with largest variance. PCA can be combined with functional data analysis to provide an overview of the typical shapes in a set of functional curves. With the 7-basis representation of the functional data, we can carry out a standard PCA on the resulting coefficient matrix in 7-dimensional space using the command below.

Fig. 6 The two principal component functions as perturbations of the mean




```
equity_ratio_pca = pca.fd(sb$fd, nharm = 2,centerfns = TRUE)
plot.pca.fd(equity_ratio_pca)
```

Figure 6 provides plots of the first two principle components of the equity ratios. These two principle components explain more than 94 % of the variability of how the equity ratios of hospitals change during the years. The solid curve is the overall mean while the ± curves are the mean plus and minus a small multiple of the principal component. The plots reveal that the majority of the variability in equity ratio occurred around 2008, a conclusion that echoes what we observed from functional clustering the standardized equity ratios.

4.5 Functional Regression

Functional data can also be combined with classical linear models. When either the response variable or predictor variable(s), or both are functional in nature, the extension of commonly-used linear models are called functional linear models. These models are especially useful in longitudinal data analysis because they allow coefficients to vary over time.

As an example, we investigate whether the dynamics of the equity ratios have any explanatory power on the quality outcome. One may argue that because the hospitals in financial distress may not be able to invest much on the newer diagnostic facilities, or attract high quality medical staff, that the quality outcome may become not as good as the hospitals in good financial status. We use the hospital level fatality rate for upper gastrointestinal (GI) bleeding in 2012 as the

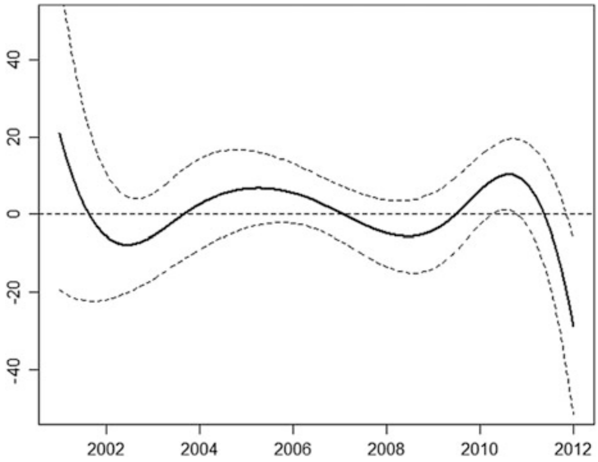


Fig. 7 Time-varying coefficient and confidence interval

medical quality outcome measure. According to Agency for Healthcare Research and Quality (National Quality Measures), in the United States more people are admitted to the hospital for GI bleeding than for congestive heart failure. The annual rate of hospitalization is estimated to be more than 300,000 hospitalizations per year, at a cost of \$2.5 billion, with a case-fatality rate of 7–10%. In this problem the response variable (GI bleeding fatality rate) is a scalar and the predictor variables are functional objects. We use the similar approach as in earlier analyses, that is, we use the 7-dimensional coefficients $\boldsymbol{\beta}_i$ to represent the functional objects of infinite dimension. Since only a fraction of the hospitals have both financial data and quality outcome measures, we have 153 observations in the model. We can use the function fRegress to estimate the parameters of the functional regression. Figure 7 displays the Time-varying coefficient and 95 % confident interval of the coefficient. The plot suggests that overall there is only weak predictive power when using equity ratio as the explanatory variable. This is so because the curve is largely flat except at the two boundaries. In fact, if we ignore the boundary effect, which is usually caused by scarcity of data, the confidence intervals cover the horizontal line of zero. This is similar to testing the hypothesis of regression coefficient being zero.

References

1. Ramsay JO, Dalzell C (1991) Some tools for functional data analysis. J R Stat Soc Series B (Methodological):539–572
2. Ramsay JO, Silverman BW (2005) Functional data analysis. Springer, New York
3. Sood A, James GM, Tellis GJ (2009) Functional regression: a new model for predicting market penetration of new products. Mark Sci 28(1):36–51
4. Jank W, Shmueli G (2006) Functional data analysis in electronic commerce research. Stat Sci 21(2):155–166

5. Ramsay JO, Ramsey JB (2002) Functional data analysis of the dynamics of the monthly index of nondurable goods production. J Econ 107(1):327–344
6. Abraham C, Cornillon P-A, Matzner-Løber E, Molinari N (2003) Unsupervised curve clustering using B-splines. Scand J Stat 30(3):581–595
7. Gastón M, León T, Mallor F (2008) Functional data analysis for non homogeneous poisson processes. In: Simulation conference WSC 2008, Winter. IEEE, Piscataway, NJ, pp 337–343
8. Bapna R, Jank W, Shmueli G (2008) Price formation and its dynamics in online auctions. Decis Support Syst 44(3):641–656
9. Wang S, Jank W, Shmueli G (2008) Explaining and forecasting online auction prices and their dynamics using functional data analysis. J Bus Econ Stat 26(2)
10. Foutz NZ, Jank W (2010) Research note-prerelease demand forecasting for motion pictures using functional shape analysis of virtual stock markets. Mark Sci 29(3):568–579
11. Graves S, Hooker G, Ramsay J (2009) Functional data analysis with R and MATLAB. Springer, New York
12. De Boor C, De Boor C, De Boor C, De Boor C (1978) A practical guide to splines, vol 27. Springer, New York
13. Schumaker LL (1981) Spline functions: basic theory, vol 1981. Wiley, New York
14. Craven P, Wahba G (1978) Smoothing noisy data with spline functions. Numer Math 31 (4):377–403
15. Duda RO, Hart PE (1973) Pattern classification and scene analysis, vol 3. Wiley, New York
16. National Quality Measures C Gastrointestinal (GI) hemorrhage: mortality rate. Agency for Healthcare Research and Quality (AHRQ). http://www.qualitymeasures.ahrq.gov/content.aspx?id=38495. Accessed 11 June 2014

A Software Application to Optimize the Visits of Sales/Marketing Agents to Their Customers in a Brewing Company

Marcos Colebrook, Ana González-Larsson, and Antonio Sedeño-Noda

1 Introduction

The biggest brewing company (the name is undisclosed due to confidentiality reasons) in the Canary Islands (Spain) is among the leading local business, generating wealth and development for the regional economy, with nearly 11,700 customers, as well as fostering nearly 900 direct jobs and 13,000 indirect jobs. The estimated revenue (in 2009) was €137 million (=US$188 million).

For such a big company in the brewing business, the relationship with its customer is vital. Salespersons are required to visit a set of assigned customers each week for marketing purposes. This information can be obtained from a database table which is consulted through a PDA (Personal Digital Assistant). Every time a salesperson visits a customer, the date and a description of the visit is written down in the PDA and transferred to the database. Each of the 53 salespersons has to accomplish approximately 2,000 visits a year.

In this sense, the company wanted to improve the visit planning application because it was not very efficient, and it did not consider the history of past visits. The problem also was justified by:

- The need to optimize company's resources.

M. Colebrook (✉)
Departamento de Ingeniería Informática, Universidad de La Laguna, Tenerife, Spain
e-mail: mcolesan@ull.edu.es

A. González-Larsson
NOVASOFT TIC, Tenerife, Spain
e-mail: anaglezlarsson@gmail.com

A. Sedeño-Noda
Departamento. Matemáticas, Estadística e Investigación Operativa, Universidad de La Laguna, Tenerife, Spain
e-mail: asedeno@ull.edu.es

© Springer International Publishing Switzerland 2015

F.P. García Márquez, B. Lev (eds.), *Advanced Business Analytics*,
DOI 10.1007/978-3-319-11415-6_12

- To balance the numbers of visits of each salesperson.
- The requirement that each customer should be visited regularly within a fixed schedule.

Unfortunately, there are few references in the literature regarding this problem. Ferrer et al. [2] presented a procedure to design the routes for the salespersons of a multinational entertainment company. The company's selling network had around 450 customers covered by 8 salespersons, whereas the rental network held 400 customers managed by 6 salespersons. Each salesperson was assigned most of the customers in the zone, and had to visit their customers an established number of times (1–5) within a month. The objective function was to minimize the weighted sum of the travel time plus overtime penalties.

In another related paper, Polacek et al. [7] dealt with determining daily routes for a traveling salesperson of a large and global food wholesaler, who had to visit up to 335 customers. Besides, each customer had to be visited at least once a year, with some customers requiring one visit per month. The primary objective was to minimize the total travel time of the salesperson. Another goal addressed was to minimize the number of days needed by the salesperson to visit all customers in a given month. They also enhanced the objective function by considering periodicity requirements for customer visits.

These papers are more concerned with routing than with planning. Besides, both the objective and the dimensions of our problem are quite different, mainly due to the requirements of the company.

The remainder of the chapter is organized as follows. In Sect. 2, the problem is described in detail. Section 3 presents the initial goals and the final scope of the requirements that was developed as a software application. The software application that the company had before tackling this project along with other software solutions are described in Sect. 4. In Sect. 5, we present and discuss the mathematical models for both the Static and the Dynamic Planning, whereas the solution algorithms are developed in Sect. 6. Finally, the computing results are provided in Sect. 7 before the conclusions of the chapter in Sect. 8.

2 Problem Description

The company has 53 salespersons (SP) that must accomplish a certain number of visits per year to their 11,700 customers. Each salesperson belongs to a fixed delegation and it is assigned to a set of customers. In turn, each customer belongs to a presales zone (PSZ) and has associated a category to indicate the number of visits (24, 18, 12, 4, 3 and 2) he/she should receive during 1 year time.

The main question for each salesperson is: *Which is the best planning for visiting my set of customers?*

Clearly, it would not be very logical that all the visits should be performed in the first month. Therefore, the optimal planning should try to space the visits in time as best as possible.

On the other hand, it is not suitable for a salesperson to have a set of weeks with lots of visits and another set with few or none, since this would mean that there is no regularity in their work. This can be achieved by evenly spreading all the visits of a salesperson to his/her customers, and taking into account holidays on each week.

Each customer belongs to a presale zone, so customers belonging to the same zone are geographically near. The company was also interested in bounding the number of presale zone that a salesperson must visit to a certain number (say for example 5), so each salesperson would not go on long trips every day, which in turn can save fuel (money) and time.

The optimal solution (planning) should consider the equitable distribution of visits all over the year (divided in working weeks), and that each salesperson should not pass through more than 5 PSZ each week.

Finally, the visits already performed in the past should also be considered, in order to plan the remaining visits in the future (dynamic planning). For instance, inserting promotional campaigns, new customers, etc.

3 Initial Goals and Final Scope

Given the above description of the problem, at a first stage, we planned the following goals according to the company's requirements:

Static Planning

For each salesperson, perform a Static Planning (12 months ahead, from January 1st to December 31st) of all visits to its customers, given the following hints:

- Visits should be adequately spaced on time over the final planning.
- For each salesperson, the visits should be equitably distributed over all the year taking into account the weeks with holidays.

Dynamic Planning

For each salesperson, accomplish a Dynamic Planning, with regards to the following:

- Take into account visits already made in the past months to carry out a new optimal planning: this requirement was the most innovative for the company.
- Make a balanced distribution of visits.
- Visits should be also adequately spaced on time (regular frequency of visits).

Promotional Campaigns

Given a planning, insert Promotional Campaigns, which are considered as special visits made to a selected set of customers in a determined range of dates.

Besides, the final scope involved all these initial goals plus the following requirements:

- Each salesperson should visit at most five presale zones each week.
- Chance to choose which salesperson planning could be re-scheduled at any time.
- Friendly user interface to build the planning.
- User interface to insert Promotional Campaigns in the database.

4 Software Applications

By the time the project was discussed, there were several applications that could be used for planning in large companies. Many of them had different types of interesting features, such as salespersons route computation, work planning, vehicle localization in real time, etc.

OPTI-TIME
Opti-Time SA is a company offering independent and complementary solutions to plan and optimize sales visits. One of the products they offer is *TourSolver* [6]. It can be used to optimally arrange the order of visits to the customers to increase efficiency. It also improves the visibility of the sales teams' visits and facilitate their orientation in the route.

LOGISPLAN
Logisplan Professional Appointments [3] is a leading optimization software oriented to transportation and logistics that calculates routes and optimal loads for each vehicle in the fleet, providing the roadmap for each of them.

Previous Company's Application
To perform each salesperson's static planning, the company had a previous software application, which was really a SQL Server Stored Procedure inside the data base. This method ran over all the customers' information to insert in the database the corresponding visit trying to distribute them in time as far as possible in order to be non-consecutive.

It was actually executed once a year, and calculated the static planning of the entire year (i.e., from January 1st to December 31st). This caused several problems:

- This planning did not take into account visits made in the past.
- The visits were not reasonably distributed: some weeks were overloaded, and some other very underused.
- It did not insert all the visits planned for a certain year, and the process was not fully automated.
- They could not insert Promotional Campaigns.
- It did not take into account the 5 Pre-Sale Zone per week limit.

In conclusion, every software application we have briefly described can be used for a specific planning task, with regard mainly to vehicle routing and not so explicitly to visit planning. Therefore, none of them fit exactly to the company's needs. This was mainly the reason of the new application.

The model presented in the following section along with the subsequent algorithm development completely solved the company's problem. Besides, being developed as a custom tailored application, the company avoided the cost of licensing other external products.

5 Mathematical Model

Tables 1 and 2 describe both the parameters and variables of the problem.

As an initial approach, and for each SP, the problem was modeled using a bipartite graph (see Fig. 1), where the set of *NC* customers are represented as the left nodes, and the set of *NW* weeks are denoted as the right nodes. Besides, nodes *A* and *B* represent the source (in this case, the salesperson) and the sink (dummy node), respectively.

Then, given the input parameters of Table 1, the problem can be expressed as finding the right assignment between the customers and the weeks, so that all the requirements are fulfilled.

After this initial graph, and depending on the type of planning (static or dynamic), we devised the following mathematical models.

Table 1 Parameters of the problem

Parameters	Description
NC	Number of customers, $c = 1,\ldots, NC$
NW	Number of weeks, $w = 1,\ldots, NW$. Depending on the year, $NW = 52$ or 53
$V[c]$	Number of visits to a certain customer c in a certain period of time (usually a year)
$MinVpW[w]$	Minimal number of visits that a SP could make in a week w
$MaxVpW[w]$	Maximal number of visits that a SP could make in a week w
$F[c]$	Frequency of visits (in weeks) to a customer c

The term *VpW* stands for "Visits per Week"

Table 2 Decision variables of the model

Variables	Description
$X[c,w]$	Binary variable equal to 1 if customer c is visited in week w, and 0 otherwise
$VpW[\mathrm{w}]$	Number of visits for a certain SP in week w
$NVpW$	Maximal number of visits per week

The term *VpW* stands for "Visits per Week"

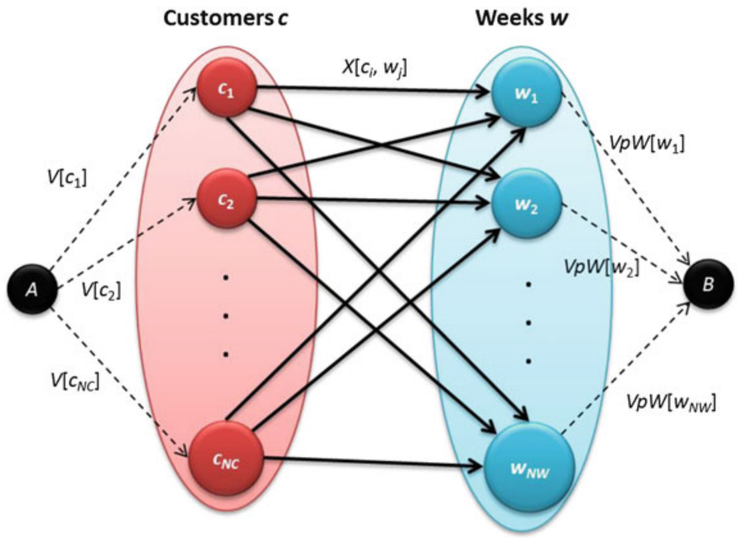


Fig. 1 Network model of the problem as a bipartite graph

5.1 Static Planning

The model for the Static Planning is as follows:

minimize $NVpW$

$$\text{(C1)} \sum_{c=1}^{NC} X[c,w] = VpW[w], \qquad w = 1,\ldots,NW$$

$$\text{(C2)} \sum_{w=1}^{NW} X[c,w] = V[c], \qquad c = 1,\ldots,NC$$

$$\text{(C3)}\ MinVpW[w] \le VpW[w] \le \min(NVpW, MaxVpW[w]) \qquad w = 1,\ldots,NW$$

$$\text{(C4)} \left\{ \begin{array}{r} X[c,1]+X[c,2]+\cdots+X[c,F[c]] \le 1 \\ X[c,F[c]+1]+X[c,F[c]+2]+\cdots+X[c,2\cdot F[c]] \le 1 \\ X[c,2\cdot F[c]+1]+X[c,2\cdot F[c]+2]+\cdots+X[c,3\cdot F[c]] \le 1 \\ \vdots \\ X[c,NW-F[c]+1]+X[c,NW-F[c]+2]+\cdots+X[c,NW] \le 1 \end{array} \right\} \ c = 1,\ldots,NC$$

$$\text{(C5)}\ 0 \le X[c,w] \le 1 \qquad \left\{ \begin{array}{l} c = 1,\ldots,NC \\ w = 1,\ldots,NW \end{array} \right.$$

$$\text{(C6)}\ VpW[w] \ge 0 \qquad w = 1,\ldots,NW$$

$$\text{(C7)}\ NVpW \in \mathbf{Z}^+$$

Objective Function

We wish to minimize the maximal value of visits in a week $NVpW$. This allows to distribute equitably the weekly task of each SP.

$$\text{minimize} \quad NVpW$$

Constraint (C1)
For each week $w = 1, \ldots, NW$, the total number of visits in such week must be equal to the sum of visits to all customers.

$$\sum_{c=1}^{NC} X[c,w] = VpW[w] \tag{1}$$

Constraint (C2)
For each customer $c = 1, \ldots, NC$, the total number of visits to that customer must be equal to the total sum of visits in all weeks to that customer.

$$\sum_{w=1}^{NW} X[c,w] = V[c] \tag{2}$$

Before describing constraint (C3), we must explain how the values of $MinVpW[\cdot]$ and $MaxVpW[\cdot]$ are computed. For each week $w = 1, \ldots, NW$, we define a $WeekWeight[w] = 1$. Then, for each non-working day (i.e. holidays), we subtract 0.2 to the $WeekWeight[w]$. For example, in a 5-day working week w with two holidays, we get $WeekWeight[w] = 1 - 0.2 - 0.2 = 0.6$. Thus, the values of $MinVpW[w]$ and $MaxVpW[w]$ are the following:

$$z[w] = \frac{\sum_{c=1}^{NC} V[c]}{\sum_{w=1}^{NW} WeekWeight[w]} \cdot WeekWeight[w]$$

$$MinVpW[w] = \lfloor z[w] \rfloor, \quad MaxVpW[w] = \lceil z[w] \rceil$$

This parameters are fine-tuned later in the algorithm to get a feasible solution.

Constraint (C3)
For each week $w = 1, \ldots, NW$, the total number of visits cannot exceed the maximal number of visits $\min(NVpW, MaxVpW[w])$ that a SP can make in a certain week, and also cannot be less than the minimal number of visits $MinVpW[w]$.

$$MinVpW[w] \leq VpW[w] \leq \min(NVpW, MaxVpW[w]) \tag{3}$$

In the right hand side expression, $NVpW$ is usually the smaller value, except for the weeks with $WeekWeight[\cdot] < 1$.

Constraints (C4)

For each customer $c = 1, \ldots, NC$, the frequency of visits $F[c]$ depends on the number of visits $V[c]$ per time unit (week):

$$F[c] = \left\lfloor \frac{NW}{V[c]} \right\rfloor$$

The frequency of visits should be adequately spaced and it should not be accumulated in adjacent weeks:

$$\begin{aligned} X[c,1] + X[c,2] + \cdots + X[c,F[c]] &\leq 1 \\ X[c,F[c]+1] + X[c,F[c]+2] + \cdots + X[c,2 \cdot F[c]] &\leq 1 \\ X[c,2 \cdot F[c]+1] + X[c,2 \cdot F[c]+2] + \cdots + X[c,3 \cdot F[c]] &\leq 1 \\ &\vdots \\ X[c,NW - F[c]+1] + X[c,NW - F[c]+2] + \cdots + X[c,NW] &\leq 1 \end{aligned} \tag{4}$$

In most cases the frequency $F[c]$ is not integer, so we have to add the remainder weeks to the last group. For example, given a customer c who is going to be visited three times in a year with 53 weeks, the ideal frequency of visits is:

$$F[c] = \left\lfloor \frac{53}{3} \right\rfloor = 17 \text{ weeks}$$

The three groups of weeks where the three visits should be accomplished are the following:

- 1 visit in the first group of weeks: 1, 2, 3, . . ., 15, 16, 17.
- 1 visit in the second group of weeks 18, 19, 20, . . ., 32, 33, 34.
- 1 visit in the third group of weeks: 35, 36, 37, . . ., 50, 51, **[52, 53]**.

Since there are two weeks remaining (52 and 53), we assign them to the last group. This might imply that the weeks on this last group will not be as balanced as the rest of weeks.

Besides, once the model is solved, the result might yield that some visits must be performed in consecutive weeks. For example, in week $F[c]$ (first group of constraints) and the next one in week $F[c]+1$ (second group). This is solved using a method discussed in Sect. 6.2.1.

To avoid an excessive load of constraints to the mathematical solver, we decided to include only these simple constraints, instead of the complete set of frequency constraints:

$$\sum_{j=k}^{k+F[c]} X[c, j+1] \leq 1, \quad k = 0, \ldots, NW - F[c]$$

This lack of constraints truly speeds up the first basic and feasible solution of the problem.

Constraints (C5)–(C7)
Finally, for each customer $c = 1, \ldots, NC$ and for each week $w = 1, \ldots, NW$, the decision variables range over the following values:

$$0 \leq X[c, w] \leq 1 \tag{5}$$

$$VpW[w] \geq 0 \tag{6}$$

$$NVpW \in \mathbf{Z}^{+} \tag{7}$$

To make the problem more tractable, we relaxed variables $X[c,w]$ to be real numbers between 0 and 1. In case any variable $X[c,w]$ gets a real value, we just round up to 1 the first $X[c,w] > 0$ for the weeks of the same group of constraints (C4). This allows obtaining an initial basic solution much faster, which is later enhanced with specific algorithms (this will be discussed in Sect. 6).

It is worth mentioning that constraint (C6) is superfluous due to (1) and (5), and hence, it can be removed from the model.

Finally, the model without the constraints in (C4) corresponds to a parametric maximum flow problem in a bipartite graph, and it can be solved in the same time complexity than the maximum flow problem (see [1]).

5.2 *Dynamic Planning*

Before presenting the model for the Dynamic Planning, we must introduce some additional parameters which are described in Table 3.

Taking into account the previous additional parameters, we present the mathematical model for the Dynamic Planning, which resembles the Static Planning model, but it has some slight modifications:

Table 3 Additional parameters for the Dynamic Planning model

Parameters	Description
PW	Number of weeks already planned (past weeks)
$V[c]$	Number of visits to a certain customer c that remain to be planned in the future, considering the visits already performed in the past
$R[c,w]$	Records the past visits. Variable equal to 1 if customer c was visited in week w, and 0 otherwise

$$\text{minimize} \; NVpW$$

$$\text{(C8)} \sum_{c=1}^{NC} X[c,w] = VpW[w], \qquad w = PW+1,\ldots,NW$$

$$\text{(C9)} \sum_{w=PW+1}^{NW} X[c,w] = V[c], \qquad c = 1,\ldots,NC$$

$$\text{(C10)} \; MinVpW[w] \le VpW[w] \le \min(NVpW, MaxVpW[w]) \qquad w = PW+1,\ldots,NW$$

$$\text{(C11)} \; X[c,w] = R[c,w] \qquad \begin{cases} c = 1,\ldots,NC \\ w = 1,\ldots,PW \end{cases}$$

$$\text{(C12)} \left\{ \begin{array}{r} X[c,PW+1]+X[c,PW+2]+\cdots+X[c,F[c]]=1 \\ X[c,F[c]+1]+X[c,F[c]+2]+\cdots+X[c,2\cdot F[c]]=1 \\ X[c,2\cdot F[c]+1]+X[c,2\cdot F[c]+2]+\cdots+X[c,3\cdot F[c]]=1 \\ \vdots \\ X[c,NW-F[c]+1]+X[c,NW-F[c]+2]+\cdots+X[c,NW]=1 \end{array} \right\} \quad c = 1,\ldots,NC$$

$$\text{(C13)} \; 0 \le X[c,w] \le 1 \qquad \begin{cases} c = 1,\ldots,NC \\ w = PW+1,\ldots,NW \end{cases}$$

$$\text{(C14)} \; VpW[w] \ge 0 \qquad w = PW+1,\ldots,NW$$

$$\text{(C15)} \; NVpW \in \mathbf{Z}^+$$

Since the main goal of the model has not changed, the objective function is the same as the one presented in the Static Planning.

Constraint (C8) and (C9)

These constraints are the same as constraints (C1) and (C2), but the week index w starts from $PW + 1$ to avoid the past weeks.

Constraint (C10)

This constraint is very similar to constraint (C3) of the Static Planning model. However, the weeks in constraint (C10) range in $w = PW + 1, \ldots, NW$ to avoid past weeks, and the values of $MinVpW[w]$ and $MaxVpW[w]$ are recomputed accordingly with the updated number of visits of each customer in the remaining weeks.

$$MinVpW[w] \le VpW[w] \le \min(NVpW, MaxVpW[w]) \tag{8}$$

Constraints (C11)

This is a new constraint that is used only in the Dynamic Planning model. For each customer $c = 1, \ldots, NC$, and for each week $w = 1, \ldots, PW$, parameter $R[c,w]$ records the visits already performed in the past weeks w to customer c.

$$X[c, w] = R[c, w] \tag{9}$$

Constraint (C12)
This constraint resemble constraint (C4) of the Static Planning model, though the weeks start at $PW + 1$.

$$\begin{aligned} X[c, PW+1] + X[c, PW+2] + \cdots + X[c, F[c]] &= 1 \\ X[c, F[c]+1] + X[c, F[c]+2] + \cdots + X[c, 2 \cdot F[c]] &= 1 \\ X[c, 2 \cdot F[c]+1] + X[c, 2 \cdot F[c]+2] + \cdots + X[c, 3 \cdot F[c]] &= 1 \\ &\vdots \\ X[c, NW - F[c]+1] + X[c, NW - F[c]+2] + \cdots + X[c, NW] &= 1 \end{aligned} \tag{10}$$

Constraints (C13)–(C15)
Same as constraints (C5)–(C7) but the weeks start in $PW + 1$.

6 Algorithm Development

Once the problem has been modeled in both its static and dynamic versions, we proceed to describe the algorithms that have solved these models. First, we introduce the software tools used to devise the solution, and subsequently, we describe the two algorithms developed.

6.1 Software Tools

For the development of this solution, we used the following tools:

Integrated Development Environment (IDE) and Programming Language
We used Microsoft Visual Studio 2008 (version 9.0.30729.1 SP) as the IDE, specifically, Microsoft Visual C# [5]. However, any other IDE supporting C# as the programming language (as open sourced SharpDevelop) should work fine.

Database Management
To get the data from the company's database, we used Microsoft SQL Server 2008.

Solver
Microsoft Solver Foundation (MSF) is a set of tools for mathematical modeling and optimization problems using OML (Optimization Modeling Language), and that integrates directly with Visual C#. Due to the large number of variables and constraints, we had to use MSF Enterprise Edition [4].

6.2 Solution Algorithms

Before describing the specific methods to solve the Static and Dynamic Planning, Fig. 2 shows a flow diagram with the different steps of the process to get the solutions.

6.2.1 Static Planning

The solution to the Static Planning returns a planning of visits from January 1st to December 31st of the current year. This version ignores the past visits, and it can be run at any time.

1. Basic feasible planning
Initially, a basic planning is performed, which consists of the planning for all SPs and their associated customers considering only those with more than two visits per year. The reason to postpone the planning of this type of customers lies in the fact that when customers with a high number of visits are inserted first, the frequency of visits of the final solution is highly improved. Besides, we also want to avoid the basic planning with two visits too close between them.

The solution provided by this stage of the process is show in Table 4.

2. Remove visits in adjacent weeks
After obtaining the basic planning, we apply a swap method to remove visits in two adjacent weeks. As we mentioned in constraint (C4), a visit might be inserted, for example, in week $F[c]$ and another in $F[c] + 1$.

Therefore, this method swaps visits with another different customer in previous or subsequent weeks. In the case of a customer with a high number of visits (for instance, 24), it is almost inevitable that two visits won't occur in consecutive weeks.

For example, customer A has two visits in a row in weeks 12 and 13. We look for customer B who has a visit in week 11. Besides, there is no problem for customer

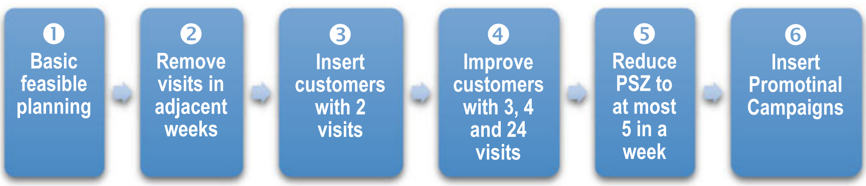


Fig. 2 Flow diagram of the process to get the optimal planning

Table 4 Data table returned by the first stage if the process

Name	Description
CUSTOMER	Code of the customer
WEEK	Week number
STARTING DATE	Date of the first day (Monday)
ENDING DATE	Date of the last day (Sunday)
SALESPERSON	Salesperson code
PRESALE ZONE	Presales Zone code

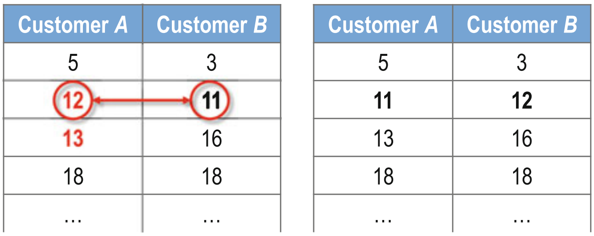


Fig. 3 An example of the swap method to solve the visits in adjacent weeks

B to have a visit in week 12. Then, we swap weeks 12 and 11 between customers *A* and *B*, as shown in Fig. 3.

3. Insert customers with two visits

With this new planning, and for each SP, we go through all the customers with two visits and we try to insert one visit in the emptiest week *w* belonging to the first part of the year, and always guaranteeing that $VpW[w] \leq MaxVpW[w]$.

Then, the second visit is inserted the emptiest week within the range $[w+f-2, w+f+2]$, where $f=26$ is the ideal frequency for customers with two visits.

Once this method is completed, the planning covers all the visits to all customers to be accomplished by all SPs.

4. Improve the frequency for customers with 3, 4 and 24 visits

First, we try to improve the frequency for customers with three and four visits by checking whether the distance between one visit and the next one is less than a half the ideal frequency. Then, we insert the visit in the emptiest week within the biggest gap of weeks.

For example, customer *A* will receive four visits this year (frequency = 13). As it is shown in Fig. 4, it has two visits in weeks 13 and 16, respectively, that are very near in time since the gap between them is less than $13/2=6$ weeks. Then, we remove week 16 from the planning, and we replace it for the best week in the range of possible solutions that, in this case, is week 21.

The method also tries to improve the frequency for customers with 24 visits that have two visits in consecutive weeks, and the swap method previously described in step 2 failed to improve this planning. To solve this problem, we look for the

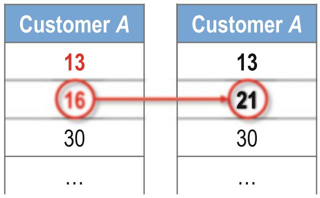


Fig. 4 Improving the frequency for a customer with four visits

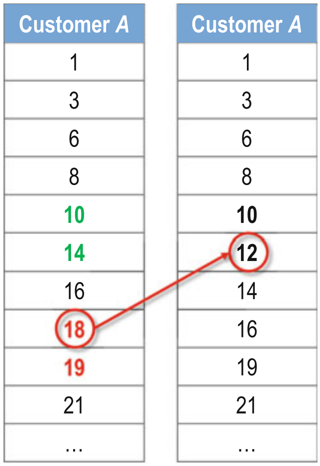


Fig. 5 Improving the frequency of a customer with 24 visits

position in the planning with a gap of at least 3 weeks, and then we insert one of the two consecutive visits in that position.

For example, in Fig. 5 we can see that the visit planning of customer *A* has two consecutive visits in weeks 18 and 19. Fortunately, there is a gap of size 3 (weeks 11, 12 and 13) between weeks 10 and 14. Therefore, we shift the visit of week 18 to week 12.

By virtue of this improvement we have solve two problems. On the one hand, there are no visits in successive weeks. On the other hand, the new planning is more balanced and hence, the weeks are not too separated from each other.

5. Reduce PSZ to at most 5 in a week
In this final step, a procedure will try to keep each SP's planning below 5 PSZ in the same week. To do this, the method proceeds on each week with more than five different PSZ and we try to swap other visits placed in previous or subsequent weeks. Actually, not only a single visit is swapped, but an entire group of visits from PSZ *X* to another group of visits in PSZ *Y*.

The result of this method dramatically improves the overall solution of the planning, reducing the PSZ to 5 or below on a large number of weeks that exceeded this limit.

At this stage, we have obtained the final optimized planning.

6. Insert Promotional Campaigns
As we mentioned in the initial goals, the Promotional Campaigns are special visits made to a selected set of customers in a particular range of dates. If a visit has been already set within that range of dates, the planning remains unchanged. Otherwise, we must insert a new visit in the emptiest week of that range of dates.

6.2.2 Dynamic Planning

The steps to obtain the Dynamic Planning are almost the same ones than the Static Planning, with the slight difference that only the (53 or 52)—*PW* future weeks will be considered, since the past *PW* weeks are already planned.

The planning that was already accomplished for the past weeks *PW* is obtained from a table in the database, and the resulting data is inserted in the model using parameter $R[c,w]$, for all the customers $c = 1,\ldots,NC$ and the corresponding weeks $w = 1,\ldots,PW$.

For each customer *c*, the frequency of visits $F[c]$ must be recomputed according to the remaining number of visits. Besides, we can choose which SP will be re-planned, leaving the rest of them unchanged.

The main similarities and differences with the Static Planning described in Sect. 6.2.1 are the following:

- The swap method is almost the same, but it will only check the planning for those weeks greater than *PW*.
- The method to insert the customers with two visits is different, since it has to verify if such visits were performed completely (2), partially (just 1) or none of them.
- The procedures to improve the frequency for customers with three and four visits, as well as the method to reduce the number of PSZ in a week, will only consider weeks from $PW + 1$ onwards.
- The step to improve the frequency of customers with 24 visits is not applied.

At this point, it is worth mentioning that, as the project progressed, the company realized and concluded that **re-planning every few months considering the past visits** would yield better results than re-planning once a year. Therefore, they decided to **run the Dynamic Planning every 3 months**, and every time there is a change in the data from the customers, presales zones or salespersons.

7 Computing Results

As we stated in the introduction, the company has 11,700 customers approximately and 53 salespersons. Some salespersons have few customers (<100) assigned, and some others have a lot of them (>400). Anyway, each salesperson must achieve 2,000 visits a year, approximately.

For each salesperson's planning model, we get around 13,000 variables and 26,000 constraints. Accordingly, we then run all 53 models, one for each salesperson. The running times on a regular laptop computer were 75 min for the Static Planning and 90 min for the Dynamic Planning.

The results obtained were very promising as a result of:

- For all customers and salespersons, all visits were fulfilled.
- The number of planned visits in a week remains below the maximum number of visits, and over the minimum number of visits.
- All visits are relatively spaced in time, although not perfectly spaced, since otherwise the equitable distribution of weeks would not hold.
- It was intended in the planning to leave no consecutive weeks, but sometimes this was impossible for customers with high number of visits (24).
- It was also intended for each salesperson to keep the number of presale zones in a week below 5, but sometimes this was not possible for salespersons with a large number of customers who are located in different presales zones.

Next, we comment out some of the results achieved for both the Static and the Dynamic Planning.

7.1 Static Planning Results

For the Static Planning, we obtained in the worst case that a salesperson with the greatest number of visits (3,938) had the following results:

- Number of Visits per Week (*NVpW*): 79 visits.
- Number of weeks with more than 5 PSZ: only 2 weeks with 6 PSZ, and the rest of them with 5 PSZ or less.
- Maximal number of visits per week ($MaxVpW[\cdot]$): 82 visits.
- Minimal number of visits per week ($MinVpW[\cdot]$): 49 visits.

In the case of a salesperson with an average number of visits (1,946), we obtained:

- Number of Visits per Week (*NVpW*): 36 visits.
- Number of weeks with more than 5 PSZ: no week has more than 5 PSZ.
- Maximal number of visits per week ($MaxVpW[\cdot]$): 40 visits.
- Minimal number of visits per week ($MinVpW[\cdot]$): 24 visits.

7.2 Dynamic Planning Results

The results of the Dynamic Planning considering 3 months in the past (12 weeks) for a salesperson with the higher number of visits (3,938) are the following:

- Number of Visits per Week (*NVpW*): 90 visits.
- Number of weeks with more than 5 PSZ: there is only one week with 7 different PSZ, and 10 weeks with 6 PSZ. The rest of them are less than or equal to 5 PSZ.
- Maximal number of visits per week ($MaxVpW[\cdot]$): 91 visits.
- Minimal number of visits per week ($MinVpW[\cdot]$): 54 visits.

We remark that, for this particular salesperson, the visits accomplished in the past were not available due to a synchronization problem with his PDA. Consequently, the planning considered a full year starting from the following week.

On the other hand, the results for a salesperson with an average number of visits (1,946) are as follows:

- Number of Visits per Week (*NVpW*): 47 visits.
- Number of weeks with more than 5 PSZ: there is only one week with 6 different PSZ, and the rest of them are less than or equal to 5 PSZ.
- Maximal number of visits per week ($MaxVpW[\cdot]$): 50 visits.
- Minimal number of visits per week ($MinVpW[\cdot]$): 30 visits.

Finally, Fig. 6 shows an example of a real planning for two customers with 12 visits each. The columns in the table (from left to right) denote the customer code, the planned week for the visit, starting date, ending date, salesperson code and presales zone code.

8 Conclusions

The initial goal of this project was to improve the planning of visits for the 53 salespersons of a brewing company to their 11,700 customers, in order to optimize the company's resources and improve efficiency.

In this sense, several goals were accomplished from the view of each salesperson:

- We obtained a planning with regular visits and equitably spaced in time.
- Balanced working weeks.
- No more than five Pre-Sales Zones was achieved in a remarkable number of cases.
- The Static Planning plans the annual visits for all salespersons to all the customers from January 1st to December 31st, regardless of the past visits.
- The Dynamic Planning takes into account the visits already made in the past.
- Promotional Campaigns can be inserted in the current planning.
- Besides, the user interface allows selecting which salespersons will be optimized.

All these requirements were fully completed by generating a basic feasible planning with a mathematical solver, and then running specialized methods to improve the initial solution.

	Customer	Week	Starting date	Ending date	SP code	PSZ code
	Cliente	Semana	FechaIni	FechaFin	Comercial	ZonaPV
1	PV200018	2	2010-12-13 00:00:00.000	2010-12-19 00:00:00.000	00001292	ZPR201
2	PV200018	5	2011-01-03 00:00:00.000	2011-01-09 00:00:00.000	00001292	ZPR201
3	PV200018	9	2011-01-31 00:00:00.000	2011-02-06 00:00:00.000	00001292	ZPR201
4	PV200018	14	2011-03-07 00:00:00.000	2011-03-13 00:00:00.000	00001292	ZPR201
5	PV200018	19	2011-04-11 00:00:00.000	2011-04-17 00:00:00.000	00001292	ZPR201
6	PV200018	21	2011-04-25 00:00:00.000	2011-05-01 00:00:00.000	00001292	ZPR201
7	PV200018	26	2011-05-30 00:00:00.000	2011-06-05 00:00:00.000	00001292	ZPR201
8	PV200018	29	2011-06-20 00:00:00.000	2011-06-26 00:00:00.000	00001292	ZPR201
9	PV200018	33	2011-07-18 00:00:00.000	2011-07-24 00:00:00.000	00001292	ZPR201
10	PV200018	39	2011-08-29 00:00:00.000	2011-09-04 00:00:00.000	00001292	ZPR201
11	PV200018	42	2011-09-19 00:00:00.000	2011-09-25 00:00:00.000	00001292	ZPR201
12	PV200018	51	2011-11-21 00:00:00.000	2011-11-27 00:00:00.000	00001292	ZPR201
13	PV200084	4	2010-12-27 00:00:00.000	2011-01-02 00:00:00.000	00001292	ZCAT02
14	PV200084	8	2011-01-24 00:00:00.000	2011-01-30 00:00:00.000	00001292	ZCAT02
15	PV200084	11	2011-02-14 00:00:00.000	2011-02-20 00:00:00.000	00001292	ZCAT02
16	PV200084	13	2011-02-28 00:00:00.000	2011-03-06 00:00:00.000	00001292	ZCAT02
17	PV200084	17	2011-03-28 00:00:00.000	2011-04-03 00:00:00.000	00001292	ZCAT02
18	PV200084	21	2011-04-25 00:00:00.000	2011-05-01 00:00:00.000	00001292	ZCAT02
19	PV200084	28	2011-06-13 00:00:00.000	2011-06-19 00:00:00.000	00001292	ZCAT02
20	PV200084	32	2011-07-11 00:00:00.000	2011-07-17 00:00:00.000	00001292	ZCAT02
21	PV200084	36	2011-08-08 00:00:00.000	2011-08-14 00:00:00.000	00001292	ZCAT02
22	PV200084	39	2011-08-29 00:00:00.000	2011-09-04 00:00:00.000	00001292	ZCAT02
23	PV200084	42	2011-09-19 00:00:00.000	2011-09-25 00:00:00.000	00001292	ZCAT02
24	PV200084	50	2011-11-14 00:00:00.000	2011-11-20 00:00:00.000	00001292	ZCAT02

Fig. 6 An example of a real planning for two customers (PV200018 and PV200084) with 12 visits each

There is no doubt that, although the planning has successfully enhanced the previous method, it could be further improved to better fit the frequency of customers with 12, 18 and 24 visits, particularly in the last case that happens quite often. Since our main goal is to hold the number of visits per week initially computed by the solver, some visits remain adjacent in consecutive weeks and there is no way to perform any swap in order to solve it.

Furthermore, the running time could be significantly reduced by changing the mathematical solver for a parametric maximum flow algorithm with additional constraints.

Finally, just to mention that the application has been running every three months since January 2011.

Acknowledgments This work is partly supported by the former Spanish Ministry of Science and Innovation through the Research Project with reference MTM2010-18591 and by the new Spanish Ministry of Economy and Competitiveness through the Research Project with reference MTM2013-43396-P.

References

1. Ahuja RK, Magnati TL, Orlin JB (1993) Network flows: theory, algorithms, and applications. Prentice-Hall, Upper Saddle River, NJ
2. Ferrer L, Pastor R, García-Villoria A (2009) Designing salespeople's routes with multiple visits of customers: a case study. Int J Prod Econ 119:46–54
3. LOGISPLAN (2014) Logisplan Professional Appointments. http://planificacionderutas.logisplan.com/en/description-logisplan-appointments.html. Accessed 10 Jun 2014
4. MSF (2014) MS Solver Foundation. http://msdn.microsoft.com/en-us/devlabs/hh145003.aspx. Accessed 10 Jun 2014
5. MVS (2014) MS Visual Studio. http://msdn.microsoft.com/en-us/vstudio/aa718325. Accessed 10 Jun 2014
6. OPTI-TIME (2014) TourSolver. http://www.opti-time.com/en/route-optimization-software. Accessed 10 Jun 2014
7. Polacek M, Doerner KF, Hartl RF et al (2007) Scheduling periodic customer visits for a traveling salesperson. Eur J Oper Res 179:823–837