Robust MCDA

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

ΘΑΛΗΣ – Πανεπιστήμιο Πειραιά
Μεθοδολογικές προσεγγίσεις για τη μελέτη της ευστάθειας σε προβλήματα λήψης αποφάσεων με πολλαπλά κριτήρια

Δ16 – Επιστημονικές δημοσιεύσεις

Π16 – Άρθρα σε επιστημονικά περιοδικά και πρακτικά συνεδρίων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Στοιχεία παραδοτέου

Δράση: Δ16 – Επιστημονικές δημοσιεύσεις

Τίτλος παραδοτέου: Π16 – Άρθρα σε επιστημονικά περιοδικά και πρακτικά συνεδρίων

Έκδοση: 04

Ημερομηνία: 30 Αυγούστου 2015

Υπεύθυνος σύνταξης: Καθηγητής Ιωάννης Σίσκος

Ομάδας σύνταξης: Όλες οι ερευνητικές ομάδες

Περιεχόμενα

1. Γενικά

1.1 Γενικά στοιχεία δράσης

Η δράση Δ16 αφορά τις επιστημονικές δημοσιεύσεις των ερευνητικών ομάδων που έχουν πραγματοποιηθεί σε όλη τη διάρκεια υλοποίησης του έργου.

Οι επιστημονικές δημοσιεύσεις βασίζονται στα ερευνητικά αποτελέσματα που έχουν παραχθεί σε διάφορες δράσεις (πακέτα εργασίας) του έργου και αφορούν:

- Δημοσιεύσεις σε διεθνή επιστημονικά περιοδικά με σύστημα κριτών
- Δημοσιεύσεις σε συλλογικούς τόμους (κεφάλαια βιβλίων) διεθνών εκδοτικών οίκων σε σύστημα κριτών
- Δημοσιεύσεις σε διεθνή/εθνικά επιστημονικά συνέδρια με σύστημα κριτών.

Η συνολική διαχείριση της συγκεκριμένης δράσης έχει πραγματοποιηθεί από τον επιστημονικό συντονιστή του έργου, Καθηγητή Ιωάννη Σίσκο, καθώς και από την Επιτροπή Παρακολούθησης, σύμφωνα με το πλάνο ποιότητας του έργου. Η Επιτροπή Παρακολούθησης του έργου αποτελείται από τους:

1. Καθηγητή Ιωάννη Σίσκο (συντονιστή έργου και υπεύθυνου της ερευνητικής ομάδας του ΠΑΠΕΙ)
2. Καθηγητή Κωνσταντίνο Ζοπουνίδη (υπεύθυνου της ερευνητικής ομάδας του Πολ. Κρήτης)
3. Καθηγητή Ιωάννη Ψαρρά (υπεύθυνου της ερευνητικής ομάδας του ΕΜΠ)
4. Professor Alexis Tsoukias, Research Director of CNRS-France (μετακαλούμενος εξωτερικός ερευνητής)

Η συμμετοχή της συγκεκριμένης επιτροπής στη δράση Δ16 είναι κυρίως έμμεση και επικουρική, παρέχοντας κυρίως υποστήριξη στην επιλογή του καταλληλότερου τρόπου δημοσιοποίησης των ερευνητικών αποτελεσμάτων. Θα πρέπει να σημειωθεί ότι η Επιτροπή Παρακολούθησης, στο πλαίσιο άλλων δράσεων του έργου δίνει τις γενικές κατευθυντήριες γραμμές της έρευνας και υποστηρίζει επικουρικά τις ερευνητικές ομάδες.

Η συγκεκριμένη δράση ανήκει στο γενικότερο πακέτο εργασίας που αφορά τη δημοσιότητα του έργου (συμπεριλαμβάνει τη δράση Δ15 αναφορικά με τη συμμετοχή των ερευνητικών ομάδων σε επιστημονικά συνέδρια, τη δράση Δ17 που αφορά την ιστοσελίδα του έργου, τη δράση Δ18 που αφορά τη διοργάνωση επιστημονικών συναντήσεων εργασιών/workshops και τη δράση Δ19 που αφορά την έκδοση μονογραφίας με τα βασικά επιστημονικά αποτελέσματα του έργου).

1.2 Γενικά στοιχεία παραδοτέου

Σύμφωνα με το πλάνο υλοποίησης, στα πλαίσια του συγκεκριμένου έργου είχαν προβλεφθεί 6 δημοσιεύσεις σε διεθνή επιστημονικά περιοδικά (2 δημοσιεύσεις ανά έτος ή 2 δημοσιεύσεις ανά ομάδα εργασίας).

Όπως παρουσιάζεται αναλυτικά και στην επόμενη ενότητα, ο συγκεκριμένος στόχος έχει υπερκαλυφθεί δεδομένου ότι οι δημοσιεύσεις των ερευνητικών ομάδων, κατά το χρόνο συγγραφής της συγκεκριμένης αναφοράς, περιλαμβάνουν:

- 10 δημοσιεύσεις σε διεθνή επιστημονικά περιοδικά με σύστημα κριτών
- 4 δημοσιεύσεις σε συλλογικούς τόμους (κεφάλαια βιβλίων) διεθνών εκδοτικών οίκων σε σύστημα κριτών
- 13 δημοσιεύσεις σε διεθνή/εθνικά επιστημονικά συνέδρια με σύστημα κριτών.

Θα πρέπει να σημειωθεί ιδιαίτερα ότι τα επιστημονικά περιοδικά που έχουν πραγματοποιηθεί οι δημοσιεύσεις του έργου είναι ιδιαίτερα υψηλού επιστημονικού κύρους. Για το λόγο αυτό στην επόμενη ενότητα δίδεται και ο συντελεστής επιρροής του εκάστοτε περιοδικού (εφόσον υπάρχει). Επίσης, πρέπει να τονισθεί ιδιαίτερα ότι η έρευνα που έχει πραγματοποιηθεί στα πλαίσια του έργου έχει αποσπάσει διακρίσεις σε διεθνή επίπεδο. Πιο συγκεκριμένα, ο υποψήφιος διδάκτορας Ελευθέριος Σίσκος, μέλος της Ομάδας Εξωτερικών Συνεργατών του Εθνικού Μετσόβιου Πολυτεχνείου, έχει λάβει υποτροφία για τη συμμετοχή του στο ELAVIO 2013 Summer School και την παρουσίαση της εργασίας του “Robust e-government evaluation based on multiple criteria decision analysis", μετά από αξιολόγησης από την ΕΕΕΕ (Ελληνική Εταιρεία Επιχειρησιακών Ερευνών), την EURO (Association of European Operational Research Societies) και την IFORS (International Federation of Operational Research Societies). Αναλυτικές πληροφορίες δίνονται στο http://www.euro-online.org/media_site/reports/ELAVIO_13_Siskos.pdf)

2. Δημοσιεύσεις

2.1 Δημοσιεύσεις σε επιστημονικά περιοδικά

Οι δημοσιεύσεις που έχουν πραγματοποιηθεί σε επιστημονικά περιοδικά με σύστημα κριτών στα πλαίσια του έργου είναι:

1. Delias, P., P. Manitsa, E. Grigoroudis, N. Matsatsinis, and A. Karasavvoglou (2013). Robustness-oriented group decision support: A case from ecology economics, *Procedia Technology*, 8, 285-291.
2. Doumpos, M., C. Zopounidis, and E. Galariotis (2014), Inferring robust decision models in multicriteria classification problems: An experimental analysis, *European Journal of Operational Research*, 236 (2), 601-611.
3. Mavrotas, G., O. Pechak, E. Siskos, H. Doukas, and J. Psarras (2015). Robustness analysis in multi-objective mathematical programming using Monte Carlo simulation, *European Journal of Operational Research*, 240 (1), 193-201.
4. Delias, P., M. Doumpos, E. Grigoroudis, P. Manolitzas, and N. Matsatsinis (2015). Supporting healthcare management decisions via robust clustering of event logs, *Knowledge-Based Systems*, 84, 203-213.
5. Mavrotas, G., J.R. Figueira, and E. Siskos (2015). Robustness analysis methodology for multi-objective combinatorial optimization problems and application to project selection, *Omega*, 52, 142-155.
6. Siskos, E. and N. Tsotsolas (2015). Elicitation of criteria importance weights through the Simos method: A robustness concern, *European Journal of Operational Research*, 246 (2), 543-553.
7. Xidonas, P., H. Doukas, G. Mavrotas, and O. Pechak (.). Environmental corporate responsibility for investments evaluation: An alternative multiobjective programming model, *Annals of Operations Research* (αποδεκτό).
8. Doumpos, M., P. Xidonas, S. Xidonas, and Y. Siskos (.). Development of a robust multicriteria classification model for monitoring the postoperative behaviour of heart patients, *Journal of Multi-Criteria Decision Analysis* (αποδεκτό).
9. Delias, P., M. Doumpos, and N. Matsatsinis (.). Business process analytics: a dedicated methodology through a case study, *EURO Journal on Decision Processes* (αποδεκτό).
10. Grigoroudis, E. and Y. Politis (.). Robust extensions of the MUSA method based on additional properties and preferences, *International Journal of Decision Support Systems* (αποδεκτό).

Η συντριπτική πλειοψηφία των συγκεκριμένων περιοδικών έχει υψηλούς δείκτες επιρροής. Πιο συγκεκριμένα, το Impact Factor (2014) των περιοδικών, όπου έχουν πραγματοποιηθεί οι δημοσιεύσεις των ερευνητικών ομάδων είναι:

- Omega (IF: 4.376)
- Knowledge-Based Systems (IF: 2.947)
- European Journal of Operational Research (IF: 2.358)
- Annals of Operations Research (IF: 1.217)

2.2 Δημοσιεύσεις σε συλλογικούς τόμους

Οι δημοσιεύσεις που έχουν πραγματοποιηθεί ως κεφάλαια βιβλίων σε συλλογικούς τόμους με σύστημα κριτών στα πλαίσια του έργου είναι:

1. Siskos, E., M. Malafekas, D. Askounis, and J. Psarras (2013). E-government benchmarking in European Union: A multicriteria extreme ranking approach, in: C. Douligeris, N. Polemi, A. Karantjias, and W. Lamersdorf (eds.), *Collaborative, trusted and privacy-aware e/m-services*, IFIP Advances in Information and Communication Technology Volume 399, Springer, New York, 338-348.
2. Doumpos, M. and C. Zopounidis (2014). The robustness concern in preference disaggregation approaches for decision aiding: An overview, in: T.M. Rassias, C.A. Floudas, and S. Butenko (eds.), *Optimization in science and engineering*, Springer, New York, 157-177.
3. Yannacopoulos, D., A. Spyridakos, and N. Tsotsolas (2014). Robustness analysis in multicriteria disaggregation-aggregation approaches for group decision making, in: F. Dargam, J.E. Hernández, P. Zaraté, S. Liu, R. Ribeiro, B. Delibašić, and J. Papathanasiou (eds.), *Decision Support Systems III: Impact of Decision Support Systems for global environments*, Lecture Notes in Business Information Processing Volume 184, Springer, New York, 167-180.
4. Delias, P., M. Doumpos, and N. Matsatsinis (2014). Robust discovery of coordinated patterns in a multi-actor business process, in: P. Zaraté, G. Camilleri, D. Kamissoko, and F. Amblard (eds.), *Group Decision and Negotiation 2014 (GDN 2014): Proceedings of the Joint International Conference of the INFORMS GDN Section and the EURO Working Group on DSS*, Toulouse University, Toulouse, 77-86.

Μέρος των δημοσιεύσεων αυτών αφορά πρακτικά συνεδρίων, τα οποία όμως έχουν δημοσιευτεί από διεθνείς εκδοτικούς οίκους.

2.3 Δημοσιεύσεις σε πρακτικά συνεδρίων

Οι δημοσιεύσεις που έχουν πραγματοποιηθεί σε πρακτικά διεθνών και εθνικών επιστημονικών συνεδρίων με σύστημα κριτών στα πλαίσια του έργου είναι:

1. Μαυρωτάς, Γ., Ο. Πετσάκ και Β. Παππά (2013). Επιλογή εύρωστου χαρτοφυλακίου επενδυτικών σχεδίων με μαθηματικό προγραμματισμό, Πρακτικά 9ου Επιστημονικού Συνεδρίου Χημικών Μηχανικών (http://9pesxm.chemeng.ntua.gr/fullpapers/EC0076.pdf).
2. Mavrotas, G. O. Pechak, D. Siatras, E. Siskos, and J. Psarras (2014). Project portfolio selection in a group decision making environment: Aiming at convergence with the iterative trichotomic approach, *Proceedings of the 2nd International Symposium & 24th National Conference on Operational Research*, HELORS, Athens, 30-36.
3. Grigoroudis, E. and Y. Politis (2014). A robust extension of the MUSA method based on desired properties of the collective preference system, *Proceedings of the 2nd International Symposium & 24th National Conference on Operational Research*, HELORS, Athens, 171-177.

4. Politis, Y. and E. Grigoroudis (2014). Combining performance and importance judgment in the MUSA method, *Proceedings of the 2nd International Symposium & 24th National Conference on Operational Research*, HELORS, Athens, 59-65.
5. Mavrotas, G., P. Xidonas, H. Doukas, and J. Psarras (2014). Constructing robust efficient frontiers for portfolio selection under various future returns scenarios, *Proceedings of the 2nd International Symposium & 24th National Conference on Operational Research*, HELORS, Athens, 178-184.
6. Mastorakis, K., E. Siskos, and Y. Siskos (2014). Value focused pharmaceutical strategy determination with multicriteria decision analysis techniques, *Proceedings of the 2nd International Symposium & 24th National Conference on Operational Research*, HELORS, Athens, 207-216.
7. Tsotsolas, N. and S. Alexopoulos (2014). Dealing with robustness in government decision-making using facilitated modelling, *Proceedings of the 2nd International Symposium & 24th National Conference on Operational Research*, HELORS, Athens, 363-372.
8. Pologiorgi, I., E. Grigoroudis, S. Tsafarakis, G. Baltas (2015). Customer satisfaction performance and importance judgments: An application of the MUSA+ model, *Proceedings of the 3rd International Symposium & 25th National Conference on Operational Research*, HELORS, Chania, 29-35.
9. Stavrou, D.I., N.P. Ventikos, and Y. Siskos (2015). Ranking risks of maritime activities with multicriteria decision aid: Application to a ship-to ship transfer operation, *Proceedings of the 3rd International Symposium & 25th National Conference on Operational Research*, HELORS, Chania, 174-180.
10. Tsotsolas, N. and S. Alexopoulos (2015). Robustness analysis approaches in political decision making, *Proceedings of the 3rd International Symposium & 25th National Conference on Operational Research*, HELORS, Chania, 214-218.
11. Politis Y. and E. Grigoroudis (2015). Analyzing robustness of the MUSA method through a simulation model, *Proceedings of the 3rd International Symposium & 25th National Conference on Operational Research*, HELORS, Chania, 219-224.
12. Spyridakos A., N. Tsotsolas, E. Siskos, and D. Yannacopoulos (2015). Estimating criteria weights exploiting priorities of the criteria and techniques of robustness analysis, *Proceedings of the 3rd International Symposium & 25th National Conference on Operational Research*, HELORS, Chania, 225-229.
13. Doumpos, M., C. Zopounidis, and P. Fragiadakis (2015). Assessing the financial performance of European banks under stress testing scenarios: A multicriteria approach, *Proceedings of the 3rd International Symposium & 25th National Conference on Operational Research*, HELORS, Chania, 272-278.

Παραρτήματα

Available online at www.sciencedirect.com
ScienceDirect
Procedia Technology 8 (2013) 285 – 291
Procedia Technology

6th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2013)

Robustness-oriented Group Decision Support. A Case from Ecology Economics

Pavlos Delias[a,*], Paschalia Manitsa[a], Evangelos Grigoroudis[b], Nikolaos Matsatsinis[b], Anastasios Karasavvoglou[a]

[a]*Kavala Institute of Technology, Agios Loukas, Kavala 65404, Greece*
[b]*Technical University of Crete, Kounoupidiana, Chania 73100, Greece*

Abstract

Problems of the field of Ecological Economics are inherently complex and by definition involve trade-offs among multiple criteria. Moreover, the decisions made involve multiple parties, often with conflicting interests. For these reasons, the multiple criteria decision aid (MCDA) paradigm appears as a valuable tool for the field of Ecological Economics and indeed as an indispensable tool in the cases where participatory decisions must be made. In this work we apply a robustness-oriented MCDA approach to reach a solution for a land use problem in Northern Greece. The mathematical modeling as well as the case study results are presented.

Selection and peer-review under responsibility of The Hellenic Association for Information and Communication Technologies in Agriculture Food and Environment (HAICTA)

Keywords: Ecological Economics, Multiple Criteria Decision Aid

1. Introduction

Problems of the field of Ecological Economics are inherently complex and by definition involve trade-offs among multiple criteria [1]. There are a number of reasons to avoid a single criteria approach [2] like neglecting certain aspects of realism and presenting features of one particular value-system as objective, just to name a few. Moreover, often the decisions made affect bigger sets than single humans (towns, cities or even larger geographical

* Corresponding author. Tel.: +30 510 462368.
E-mail address: pdelias@teikav.edu.gr

2212-0173 © 2013 The Authors. Published by Elsevier Ltd.
Selection and peer-review under responsibility of The Hellenic Association for Information and Communication Technologies in Agriculture Food and Environment (HAICTA)
doi:10.1016/j.protcy.2013.11.038

territories, local or national populations, societies etc.). Therefore, it is expected that multiple parties are involved in the decision process. For these reasons, the multiple criteria decision aid (MCDA) paradigm appears as a valuable tool for the field of Ecological Economics [3] and indeed as an indispensable tool in the cases where participatory decisions must be made [4].

Considering the evaluation of the natural capital and the ecosystem services, perhaps the most visible work is the work of Costanza et al. [5]. Several approaches using multiple evaluation factors have been presented [6], however the vast majority of works considers the ecological and the financial factor, underestimating the significance of the social factors. Neglecting or underestimating these factors leads to a misjudgment about the real value (or demand) of the ecosystem services for stakeholders. In [7], authors try to integrate social factors into the ecosystem service appraisal with a social welfare weight using the Ruoergai Plateau Marshes as a case study. However, the Analytic Hierarchy Process which is used as the multiple criteria tool, has been systematically criticized [8], [9].

In this work we apply a novel MCDA algorithm to support the decision about the land use in the area of Paggaio, Kavala, Greece. In particular, a convenience sample if six local stakeholders was interviewed to express its preferences about some cultivation alternatives (land use for photovoltaic systems was also included). The proposed method can be characterized as an attempt to combine preference relations with a UTA approach, which is actually a new trend aggregation – disaggregation approaches [10]. The idea of considering the whole set of compatible value functions to deal with ranking and choice problems was originally introduced in the UTA^{GMS} method [11], and further generalized in GRIP[12].

The family of the UTA methods has been also used in several studies of conflict resolution in multi-actor decision situations [13]. These studies refer to the development and application of group decision or negotiation support systems [14], [15], [16]. Beside UTA methods, Matsatsinis and Samaras [17] review several other aggregation- disaggregation approaches incorporated in group decision support systems. While group decision approaches aim to achieve consensus among the group of DMs or at least attempt to reduce the amount of conflict by compensation, collective decision methods focus on the aggregation of the DMs' preferences. Therefore, in the latter case, the collective results are able to determine preferential inconsistencies among the DMs, and to define potential interactions (trade-off process) that may achieve a higher group and/or individual consistency level.

The problem formulation and the model of the constructed linear problem are presented in the next Section, while a special section is dedicated to the robustness point of view of the method. Finally, preliminary results of the case study as well as some general conclusions are presented in the following sections.

2. Problem Modeling

Let m be the number of the decision makers involved in the problem under discussion. These decision makers (DMs) act as autonomous, self-interest agents. The notation $D = \{D_1, D_2, ..., D_m\}$ shall be used to symbolize them. Every agent (DM) has a weight of significance for the decision ruler (who is usually the responsible authority, as appointed by the government). This weight could represent the relative value that every agent has for the local society, its expertise level or it could be a parameter defined by formal statements. In any case, there should always be $\sum_{t=1}^{m} w_t = 1$.

Let n be the number of criteria $G = \{g_1, g_2, ..., g_n\}$, which will be used to evaluate the alternative solutions. The alternative solutions set can be of any finite size and it shall be notated as $A = \{a, b, ...\}$. Alternative solutions in this paper are nothing else than land usage, i.e., alternative ways to exploit land. Besides the existing solutions, the methodology suggested in this work introduces a set of reference alternatives A_R. According to [18] this set could be: a set of past decision alternatives past actions; a subset of decision actions, especially when A is large; or a set of fictitious actions, consisting of performances on the criteria, which can be easily judged by agents to perform global comparisons.

The concept of reference alternatives is common in the aggregation-disaggregation paradigm of the MCDA, however, the novelty of this method consists in non demanding a complete comparisons table. In particular, every agent (DM) is asked to express his/her preferences over just a subset of these reference alternatives. Representing by

A_{Rt} the set of the reference alternatives used for comparisons by the t[th] agent, the following should hold: $A_R = A_{R1} \cup A_{R2} \cup \ldots \cup A_{Rm}$. In order to compare alternatives, let us denote a preference relation S on $A \times A$, in a way that a means "alternative a is at least as good as b".

The ultimate goal of the methodology is to model the collective preferences of agents (DMs). To this end an additive value function u is introduced as following: $u(g) = \sum_{j=1}^{n} u_j(g_j)$. Each $u_j(g_j)$ is piecewise linear on $u_j(g_j), G_j = \{g_j^1, g_j^2, \ldots, g_j^{aj}\}$ being the number of level of performance of the jth criterion. In addition, the worst and the best performance have standard values as: $u_j(g_{j*}) = 0 \forall j, \sum_{j=1}^{n} u_j(g_j^{aj}) = 1$. Finally, the preferences relation is expressed on a value function basis as: $aSb \Leftrightarrow u[g(a)] - u[g(b)] \geq 0$.

3. A Robustness-oriented Algorithm

Each agent provides just two basic pieces of information: The first consists of a set of pairwise comparisons of some reference alternatives. These comparisons are made in terms of the preference relation defined in the previous section. This way, the t[th] decision maker provides a comparisons set $R_t \subset A_R \times A_R$, which could be of any size and include any reference alternatives. A comparison in that set (a row of the matrix) would indicate two alternatives (e.g. a and b) for which the preference relation a holds. The second piece of information needed is a set of intensities about the preference relations between couples of alternatives of A_{Rt}. Again, this comparisons' set does not have to be complete. More specifically, let I_t be the set of the "intensities" of the t[th] decision maker. An element of I_t would declare if a comparison (an element in R_t) is more "intense" than any other element in R_t. For example, a is more intensive than c.

The collective value function will be calculated through a linear regression problem. To this end, two variables z_{tk} and y_{tp} are introduced. The former refers to the k[th] preference relationship of the t[th] agent and the latter to the p[th] intensity declared by the t[th] agent. The linear problem is formulated as follows:

$$[\min] z = \sum_{t=1}^{m} \left(\sum_{k=1}^{|R_t|} z_{tk} + \sum_{p=1}^{|I_t|} y_{tp} \right)$$

subject to

$$u[\underline{g}(a)] - u[\underline{g}(b)] + z_{tk} \geq 0 \forall t = 1,2,\ldots,m; k = 1,2,\ldots,|R_t|$$

$$(u[\underline{g}(a)] - u[\underline{g}(b)]) - (u[\underline{g}(c)] - u[\underline{g}(d)]) + y_{tp} \geq 0, t = 1,2,\ldots m; p = 1,2,\ldots,|I_t|$$

$$u_j(g_j^{l+1}) - u_j(g_j^l) \geq 0, j = 1,2,\ldots,n; l = 1,2,\ldots,a_j - 1$$

$$u_j(g_j^1) = 0$$

$$\sum_{j=1}^{n} u_j(g_j^{a_j}) = 1$$

$$u_j(g_j^l) \geq 0, j = 1,2,\ldots,n; l = 1,2,\ldots,a_j; w_{tk} \geq 0; y_{tp} \geq 0; t = 1,2,\ldots,m; k = 1,2,..,|R_t|; p = 1,2,\ldots,|I_t|$$

Robustness analysis of the results provided by the Linear Problem is considered as a post-optimality analysis problem. What is actually applied is a slight alteration of the polyhedron defined by the constraints of the initial linear problem. The polyhedron is augmented by the additional constraint $z \leq z^* + \varepsilon, z^*, \varepsilon$ being the minimal error of the initial LP, and ε a very small positive number. A number of $T = \sum_{j=1}^{n}(a_j - 1)$ new linear problems are constructed and T value functions are calculated by maximizing and minimizing each value $u_j(g_j^l), j = 1,2,...,n; l = 2,...,a_j$, on the augmented polyhedron.

As a measure for the robustness of the marginal value functions the average stability indices $ASI(i)$ are used. An average stability index $ASI(i)$ is the mean value of the normalized standard deviation of the estimated marginal values on i[th] criterion and is calculated as

$$ASI(i) = 1 - \frac{1}{a_i - 1} \frac{\sum_{k=1}^{a_i-1} \sqrt{\left(T\left(\sum_{j=1}^{T} \left(u_k^j\right)^2\right) - \left(\sum_{j=1}^{T} u_k^j\right)^2 \right)}}{\frac{T}{a_i - 1}\sqrt{(a_i - 2)}}$$

Where u_j^k is the estimated value of the k[th] parameter in the j[th] additive value function.

The global robustness measure will be the average of $ASI(i)$ over all the criteria. If robustness measures are judged satisfactory, i.e. ASI indices are close to 1, then the final solution is calculated as the barrycentral value

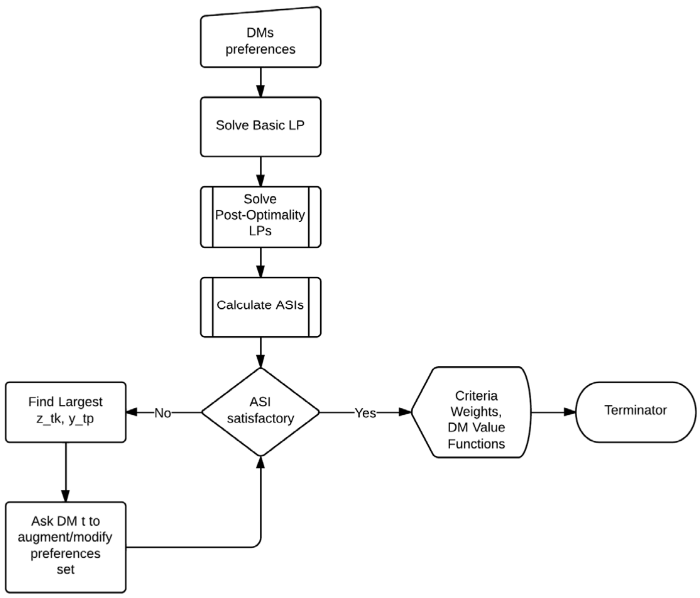


Fig. 1 The Flow chart of the Robustness-oriented algorithm

function. Else, the sets R_t and I_t should be enriched for one or more agents. The way to guide the R_t and I_t redefinition process is by checking the magnitude of the variables z_{tk} and y_{tp}. In particular, the larger these variables are, the greater the inconsistency they will prompt. So, Decision Maker*t* (who is related with z_{tk} and y_{tp}) shall be contacted by priority and thus the whole process (depicted inFig. 1) is guided by the robustness of the final solution.

4. The Case Study

The land of Paggaio, Kavala although very rich (after reclaiming a dried lake in 1930) has been cultivated in ways that affected both local environment and economies in a disadvantageous manner. Six local stakeholders – experts who represent different points of view were interviewed (two farmers, an agronomist, the president of local agricultural cooperatives, a resident and a complete feed mill owner) and expressed their preferences for 11 different land uses. The eleven alternatives for land use are a) Cultivation of colza (to extract oil and exploit the cake left), b) Cultivation of white poplar (Populus alba – Salicaceae) for the paper industry and biofuels, c) Sugar beets (cultivated Beta vulgaris) for biofuels and the food industry, d) Helianthus (sunflower) to mainly be used as a biofuel, e) Stevia for pharmaceutical or food industry, f) Photovoltaic parks, g) Barley for mash production, h) Wheat for the same purpose, i) Soybean also mash production, j) Maize and k) Pomegranate for the food industry as well as for pharmaceuticals.

Each alternative is evaluated against every criterion using a textual, ordinal 5-level scale. This multicriteria evaluation table is the same for all stakeholders. All stakeholders will express their preferences, however, not all are the same influential, namely, the "importance weight" of each might differ. In this case, stakeholders were selected based on a convenience basis, according to their profession – position and the following weights were assigned: 0.4 for the agronomist (C_6), 0.2 for the president of local cooperatives (C_1), 0.13 for the feed mill owner (C_4), 0.1 for each farmer (C_3& C_5) and 0.07 for the resident (C_2). These weights indicate the trade-off between the "expertise" of two stakeholders, while it is required to sum up to 1. Stakeholders are provided with the multicriteria evaluation table, and they express their preferences with statements like the ones described in the Problem Modeling section. In our case the stakeholders' preferences are presented in Table 1 Table 1:

Table 1. Stakeholders' preferences

Stakeholder	Preferences	Intensities
C_1	*{fSk, fSi, jSk, jSi, kSi}*	[1,3;2,5]
C_2	*{fSc, fSd, fSk, cSd, cSk}*	[2,1; 2,3;3,5]
C_3	*{jSc, jSf, jSk, jSd, cSk, cSd, fSk, fSd}*	[1,8;1,7;4,5]
C_4	*{fSj, fSi, fSh, jSh, jSi, hSi}*	[1,2;2,4;4,6]
C_5	*{fSd, fSi, dSk, fSc, dSc, iSc, kSc}*	[1,2;2,7;5,7]
C_6	*{fSi, fSd, fSk, jSi, jSd, jSk, jSi, iSk, dSk}*	[1,2;1,8;5,1]

Table 1 represents stakeholders' preferences in terms of pairwise comparisons (when such a comparison makes sense for the stakeholder) and in terms of intensities between those pairwise comparisons. The preferences set for each stakeholder contains the preference relations he declares (for instance C_1has declared that "alternative f is at least as good as k", "alternative f is at least as good as i", "alternative j is at least as good as k" etc. The intensities matrix contains as many rows as the number of the intensities declared (rows are separated by columns).Each row contains two numerical values, which correspond to the indices of the preferences relations involved.

For instance, row 1 can be interpreted like the following: The president of local cooperatives prefers the implementation of photovoltaic parks to the cultivation of pomegranate and to the cultivation of soybean, as well as he prefers the cultivation of maize to pomegranate and to soybean. He also prefers the cultivation of pomegranate to soybean. However, he considers his preference of photovoltaic system to pomegranate to be greater (more intense) than his preference of maize to soybean. The interesting part is that stakeholders do not need to express their preferences over the entire set of alternatives nor they need to declare intensities for every pair of relations. This is an important advantage of the proposed method that provides great flexibility to both the decision analysts and stakeholders.

Having solved the LP, results are presented in Table 2, however the overall ASI index is quite low. This means that additional input data (further clarifications on DMs' preferences) are needed. In particular, the need is for the

DMs with the largest z_{tk} and y_{tp} values (i.e., the president of cooperatives, the feed mill owner and the resident in descending order) to complement their preferences data. Moreover, additional intensities could be requested to make input information richer. The new data are presented in Table 3. Then a new iteration (re-solve the LP) follows and the robustness of the new results is re-evaluated.

Table 2. Preliminary Results (Iteration 1& 2)

Criterion	Weight (iter 1 / 2)	ASI (iter 1 / 2)
Environment friendliness	17%/ 18%	0.44 / 0.44
Exploitation of Natural Resources	21% / 19%	0.49 / 0.46
Land reuse potential	18% / 17%	0.52 / 0.51
Economical Performance	11% / 10%	0.44 / 0.44
Available Information	15% / 16%	0.45 / 0.45
Investment Attractiveness	18% / 18%	1 / 1

Table 3. Stakeholders' preferences update

Stakeholder	Preferences	Intensities
C_1	*{fSk, fSi, jSk, jSi, kSi, fSh, fSg, jSg, kSh, kSg, hSg}*	[1,3;2,5; 4,12]
C_2	*{fSc, fSd, fSk, cSd, cSk, fSj, fSi, cSj, cSi, dSj, dSi, kSj, kSi, jSi}*	[2,1; 2,3;3,5; 10,9]
C_3	*{jSc, jSf, jSk, jSd, cSk, cSd, fSk, fSd}*	[1,8;1,7;4,5]
C_4	*{fSj, fSi, fSh, jSh, jSi, his, fSc, fSd, fSg, jSc, jSd, jSg, hSc, hSd, hSg, iSc, iSd, iSg, dSc, gSc}*	[1,2;2,4;4,6; 19,20]
C_5	*{fSd, fSi, dSk, fSc, dSc, iSc, kSc}*	[1,2;2,7;5,7]
C_6	*{fSi, fSd, fSk, jSi, jSd, jSk, jSi, iSk, dSk}*	[1,2;1,8;5,1]

As it can be seen from Table 2 (iteration 2 elements), the ASI index is even lower after the new data. This is of course not a fortunate event since it signifies that the assessed collective model is not robust. This can be explained by the rigid attitude of the stakeholders who instead of adjusting their preferences with the rest ones, they prefer to intensify their personal opinion with additional declaration. The results demonstrate that this is a hard negotiation problem. Potential conflict resolution strategies would be to include more stakeholders into the process, to modify the stakeholders' weights, to eliminate certain decision criteria or certain land use alternatives.

5. Conclusions

In this work a multi-criteria methodology to support the land use decision was presented. What guide the reasoning component are the collective preferences of all stakeholders. Therefore, the final solution depends in a very direct way on the stakeholder's rationality. This infuses the system with an impressive flexibility but also with a disagreeable subjectivity. More specifically, modelling stakeholders as rational optimizers based on the suggested multiple criteria approach there emerge the same limitations with those of classical decision aid: There is a fuzzy borderline between what is and what is not feasible in real decision making contexts; the Decision makers' have seldom well shaped preferences. "In and among areas of firm convictions lie hazy zones of uncertainty, half held belief, or indeed conflicts and contradictions”[19]; many data are imprecise, uncertain, or ill-defined. In addition, sometimes, data may not be reflected appropriately into linear utility functions. Even more, in a real-world context, we shall not neglect complexity and time-issues: decisions have to be made in real time.

Despite the above limitations, the multiple criteria paradigm emerges as an endeavour to make an objective place for agents’ decisions. It provides a way to formalize pro-activeness guiding stakeholders to rational and transparent decisions.

Acknowledgements

This research has been co□financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) □ Research Funding Program: **THALES**. Investing in knowledge society through the European Social Fund.

References

[1] Daly H. E. Ecological economics principles and applications. Washington: Island Press; 2011.
[2] Roy B. Mulcritéria Methodology for Decision Aiding. Dordrecht: Kluwer; 1996.
[3] Shmelev S. E. Economic Valuation and Decision Making: MCDA as a Tool for the Future. Ecological Economics, Dordrecht: Springer Netherlands 2012;57–74.
[4] Garmendia E., Gamboa G. Weighting social preferences in participatory multi-criteria evaluations: A case study on sustainable natural resource management. Ecological Economics 2012;84:110–120.
[5] R. Costanza R., De Groot R., Grasso M., Hannon B., Limburg K., Naeem S., Paruelo J., and others. The value of the world’s ecosystem services and natural capital. Nature 1997;387(6630):253–260.
[6] Zhang B., Li W., and Xie G. Ecosystem services research in China: Progress and perspective. Ecological Economics 2010;69(7):1389–1395.
[7] Zhang X., Lu X.. Multiple criteria evaluation of ecosystem services for the Ruoergai Plateau Marshes in southwest China. Ecological Economics 2010;69(7):1463–1470.
[8] Whitaker R. Criticisms of the Analytic Hierarchy Process: Why they often make no sense. Mathematical and Computer Modelling 2007;46(7–8):948–961.
[9] Bana e Costa C. A., Vansnick J.-C. A critical analysis of the eigenvalue method used to derive priorities in AHP. European Journal of Operational Research 2008;187(3):1422–1428.
[10] Siskos Y. and Grigoroudis E. New Trends in Aggregation-Disaggregation Approaches. Handbook of Multicriteria Analysis 2010;103:189–214.
[11] Greco S., Mousseau V., and Słowiński R. Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions. European Journal of Operational Research 2008;191(2):416–436.
[12] Figueira J. R., Greco S., and Słowiński R. Building a set of additive value functions representing a reference preorder and intensities of preference: GRIP method. European Journal of Operational Research 2009;195(2):460–486.
[13] Jacquet-Lagrèze E. and Siskos Y. Preference disaggregation: 20 years of MCDA experience. European Journal of Operational Research 2001;130(2):233–245.
[14] Shakun M. F. Group decision and negotiation support in evolving, nonshared information contexts. Theory and Decision 1990;28(3):275–288.
[15] Shakun M. F.. Airline Buyout: Evolutionary Systems Design and Problem Restructuring in Group Decision and Negotiation. Management Science 1991;37(10):1291–1303.
[16] Matsatsinis N. and Delias P. A Multi-criteria Protocol for Multi-agent Negotiations. Methods and Applications of Artificial Intelligence 2004;3025:103–111.
[17] Matsatsinis N. F., Samaras A. P. MCDA and preference disaggregation in group decision support systems. European Journal of Operational Research 2001;130(2):414–429.
[18] Siskos Y., Grigoroudis E., and Matsatsinis N. UTA methods. Multiple criteria decision analysis: state of the art surveys, Springer New York ;2005. p. 297–344.
[19] Roy B. Paradigms and Challenges. Multiple Criteria Decision Analysis: State of the Art Surveys 2005;78:3-24.

European Journal of Operational Research 236 (2014) 601–611

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Inferring robust decision models in multicriteria classification problems: An experimental analysis

Michael Doumpos [a,*], Constantin Zopounidis [a,b], Emilios Galariotis [b]

[a] Technical University of Crete, Dept. of Production Engineering and Management, Financial Engineering Laboratory, University Campus, 73100 Chania, Greece
[b] Audencia Nantes School of Management – PRES LUNAM, Centre for Financial and Risk Management, 8 route de la Jonelière, 44312 Nantes, France

ARTICLE INFO

Article history:
Received 18 July 2013
Accepted 23 December 2013
Available online 3 January 2014

Keywords:
Multiple criteria analysis
Robustness
Disaggregation analysis
Monte Carlo simulation

ABSTRACT

Recent research on robust decision aiding has focused on identifying a range of recommendations from preferential information and the selection of representative models compatible with preferential constraints. This study presents an experimental analysis on the relationship between the results of a single decision model (additive value function) and the ones from the full set of compatible models in classification problems. Different optimization formulations for selecting a representative model are tested on artificially generated data sets with varying characteristics.

1. Introduction

The elicitation, modeling, and representation of preferential information are crucial steps in providing decision-makers (DMs) with sound decision analysis and aiding tools. Multiple criteria decision aid (MCDA) provides a wide arsenal of techniques and approaches to address such issues in the context of decision problems involving multiple (conflicting) criteria. Among others, MCDA techniques employ information on the preferential system of the DM to build criteria aggregation models for evaluating a set of alternative ways of action.

Information on the DM's preferential system and judgment policy can be obtained either directly or indirectly. In this paper we concentrate on the latter approach, referred to as "preference disaggregation analysis" (PDA, Jacquet-Lagrèze & Siskos, 2001). The disaggregation framework does not require the DM to provide the analyst with specific details on the parameters that define the criteria aggregation model. Instead, the model building process is based on the analysis of a small set of representative decision instances (reference set), using non-parametric regression techniques.

The quality of models resulting from disaggregation techniques depends not only on the information embodied in the sample of decision instances but also on the properties of the model fitting process. In this context, the issue of robustness has recently received much attention (Roy, 2010). The research in the area of building robust multicriteria decision models and obtaining robust recommendations with disaggregation techniques has adopted two main approaches. The first is based on the use of analytic methodologies for: (a) formulating preference relations and recommendations based on characterizations of the range of decision models compatible with the DM's judgments on the reference set (Greco, Mousseau, & Słowiński, 2010; Kadziński, Greco, & łowiński, 2012) and (b) building robust decision models that best represent the information embodied in the reference data (Bous, Fortemps, Glineur, & Pirlot, 2010; Doumpos & Zopounidis, 2007; Greco, Kadziński, & Słowiński, 2011). The second line of research has focused on using simulation techniques to sample different decision models compatible with the DM's preferences in order to form robust recommendations (Kadziński & Tervonen, 2013), thus enriching analytic procedures with a more detailed/explicit view of the outputs that can be obtained from the universe of compatible models.

Vetschera, Chen, Hipel, and Kilgour (2010) conducted an experimental investigation of the robustness of the information embodied in a reference set in the context of multicriteria classification problems. In this study we extend this analysis by focusing on the robustness and performance of representative decision models fitted on a set of reference alternatives using different optimization formulations. Using a good decision model that best represents the information provided by the DM on the reference data and provides robust results is of major importance in the context of decision aiding. Having an analytic or simulation-based characterization of all compatible models provides the DM with a comprehensive view of the range of possible recommendations that can be formed. On the other hand, a single representative

* Corresponding author. Tel.: +30 2821037318; fax: +30 2821069410.
E-mail addresses: mdoumpos@dpem.tuc.gr (M. Doumpos), kostas@dpem.tuc.gr (C. Zopounidis), egalariotis@audencia.com (E. Galariotis).

0377-2217/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2013.12.034

model is easier to use as it only requires the DM to “plug-in” the data for any alternative into a functional, relational, or symbolic model. Furthermore, the aggregation of all evaluation criteria in a single decision model enables the DM to get insight into the role of the criteria and their effect on the recommendations formulated through the model (Greco et al., 2011).

Traditional disaggregation techniques such as the family of the UTA methods (Siskos, Grigoroudis, & Matsatsinis, 2005) use post-optimality techniques based on linear programming in order to build a representative additive value function (AVF) defined as an average solution of some characteristic models compatible with the DM’s judgments. Recently, a number of other approaches have been proposed. For example, Greco et al. (2011) proposed a procedure (which implements max–min optimization models) for building a representative AVF that provides recommendations on possible assignments corresponding to the most stable results of a robust ordinal regression analysis. The proposed procedure is iterative allowing the DM to specify (interactively) at each iteration different targets that a representative model should achieve. Similar processes can also be used for the construction of representative AVFs in a group decision making context (Kadziński, Greco, & Słowiński, 2013). Kadziński and Tervonen (2013) extended this approach through its combination with a simulation process, which enhances the results of robust ordinal regression with assessments on the acceptability (i.e., confidence) of the assignments and proposed an optimization model to construct a model that best represents the simulation results. Instead of interactive and iterative model building procedures, other studies have focused on the introduction of optimization formulations based on new model fitting criteria. For instance, Doumpos and Zopounidis (2007) proposed a formulation based on the regularization principle of statistical learning, whereas Bous et al. (2010) presented a model based on the concept of the analytic center.

In this study we analyze such approaches (also introducing a new linear programming model) in order to examine the way in which their results represent the information provided by the DM’s reference judgments and their relationship with the robust recommendations that can be formulated on the basis of this information. Among others, the objectives of the analysis include the investigation of: (a) the association between robustness and the selection of representative decision models defined by parameters that lie near the “center” of the set that consists of all models compatible with the DM’s preferences, (b) the connection between the complexity of a decision model and its robustness and accuracy, and (c) the ability of different model inference procedures to cope with models of increasing complexity and the impact that the characteristics of the data have on the robustness of the inference process. The analysis is based on simulated data generated with different characteristics, in the context of multicriteria classification problems, which have recently received much attention among MCDA researchers (Zopounidis & Doumpos, 2002). We focus on decision models expressed in the form of linear and piecewise AVFs, which are widely used in MCDA. The results of the analysis contribute in improving the understanding of the features of disaggregation approaches that aim towards identifying representative decision models, as well as clarifying the relationship between the results of such approaches with the concept of robustness in decision aid.

The rest of the paper is organized as follows. Section 2 presents different optimization-based approaches for constructing AVFs in classification problems that best represent the set of models compatible with the DM’s judgments on some decision examples. Section 3 discusses the experimental setting used for the comparison of the selected approaches, whereas Section 4 presents and analyzes the obtained results. Finally, Section 5 concludes the paper and outlines some future research directions.

2. Inferring a representative additive value function in multicriteria classification problems

2.1. General framework

AVFs constitute a simple and easy to use modeling approach to decision aiding problems. They are based on a sound theoretical framework (multiattribute value theory), and despite their reliance on specific preferential independence conditions (Keeney & Raiffa, 1993), they are widely used in decision aiding and modeling.

Assuming that K criteria are used in a multicriteria evaluation context, an AVF introduces a criteria aggregation model, under which the global value (performance) of an alternative i is obtained as follows:

$$V(\mathbf{x}_i) = \sum_{k=1}^{K} w_k v_k(x_{ik}) \tag{1}$$

where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})$ is the vector with the data for alternative i on the evaluation criteria, $w_k \geqslant 0$ is the trade-off coefficient for criterion k (the normalization $w_1 + w_2 + \ldots + w_K = 1$ is often used) and $v_k(\cdot)$ is the marginal value function of criterion k. The marginal value functions define the partial performance of the alternative on each criterion, usually in a scale between 0 and 1.

Under the decision model (1) an alternative i is preferred over an alternative j if and only if $V(\mathbf{x}_i) > V(\mathbf{x}_j)$, whereas the alternatives are indifferent if $V(\mathbf{x}_i) = V(\mathbf{x}_j)$. In a multicriteria classification setting, each alternative should be classified in a set of N pre-defined categories $\{C_1, \ldots C_N\}$ ordered such that category C_1 includes the best alternatives and category C_N the worst ones. An AVF model can be easily used to classify any alternative i as follows:

$$t_\ell < V(\mathbf{x}_i) < t_{\ell-1} \iff \text{Alternative } i \text{ belongs to class } C_\ell \tag{2}$$

where $t_0 = 1 > t_1 > t_2 \cdots > t_{N-1} > t_N = 0$ is a set of thresholds that distinguish the categories. Cases where $V(\mathbf{x}_i) = t_\ell$ clearly lead to some ambiguity in the assignment of alternative i to one of the pre-defined categories (i.e., it can be assigned to C_ℓ or $C_{\ell+1}$). In the context of this study we assume that any test alternative i with $V(\mathbf{x}_i) = t_\ell$ is assigned to category C_ℓ.

The construction of the AVF can be simplified by setting $u_k(x_k) = w_k v_k(x_k)$, which leads to a rescaled set of marginal value functions $u_1, \ldots, u_K$ normalized in $[0, w_k]$. With this transformation, the AVF model (1) is expressed in the following equivalent form:

$$V(\mathbf{x}_i) = \sum_{k=1}^{K} u_k(x_{ik}) \tag{3}$$

The AVF model can be linear or nonlinear depending on the form of the marginal value functions. A convenient and flexible way to take into consideration a wide class of monotone marginal value functions, is to assume that they are piecewise linear. Under this scheme the scale of each criterion k is split into $s_k + 1$ subintervals defined by s_k break-points $\beta_0^k < \beta_1^k < \cdots < \beta_{s_k+1}^k$, between the least and the most preferred levels of the criterion (denoted by β_0^k and $\beta_{s_k+1}^k$, respectively). Thus, the marginal value of any alternative i on criterion k can be expressed as:

$$u_k(x_{ik}) = \sum_{r=1}^{s_k} p_{ik}^r d_{kr} \tag{4}$$

where $d_{kr} = u_k(\beta_r^k) - u_k(\beta_{r-1}^k) \geqslant 0$ is the difference between the marginal values at two consecutive break-points of criterion k and

$$p_{ik}^r = \begin{cases} 0 & \text{if } x_{ik} < \beta_{r-1}^k \\ \frac{x_{ik} - \beta_{r-1}^k}{\beta_r^k - \beta_{r-1}^k} & \text{if } x_{ik} \in \left[\beta_{r-1}^k, \beta_r^k\right] \\ 1 & \text{if } x_{ik} > \beta_r^k \end{cases} \tag{5}$$

Therefore, the AVF (3) can be expressed as a linear function of the step differences in the marginal values between consecutive break-points in the criteria's scale:

$$V(\mathbf{x}_i) = \sum_{k=1}^{K} \mathbf{p}_{ik}^{\top}\mathbf{d}_k \tag{6}$$

where $\mathbf{p}_{ik} = (p_{ik}^1, p_{ik}^2, \ldots, p_{ik}^{s_k})$ and $\mathbf{d}_k = (d_{k1}, d_{k2}, \ldots, d_{ks_k})$.

In a preference disaggregation framework for classification problems, the DM provides a reference set consisting of decision examples for M alternatives. The reference alternatives are classified into the pre-defined categories, and the objective is to infer the parameters of the AVF model (i.e., the vectors $\mathbf{d}_1, \ldots, \mathbf{d}_K$ and the classification thresholds) that are consistent with the classification of the alternatives. Thus, the inferred model of the form (6) should satisfy the following set of linear constraints:

$$V(\mathbf{x}_i) \geqslant t_\ell + \delta \qquad \forall \text{ alternative } i \text{ from category } C_\ell \ (1 \leqslant \ell \leqslant N-1) \tag{7}$$

$$V(\mathbf{x}_i) \leqslant t_{\ell-1} - \delta \qquad \forall \text{ alternative } i \text{ from category } C_\ell \ (2 \leqslant \ell \leqslant N) \tag{8}$$

$$\sum_{k=1}^{K} \mathbf{1}^{\top}\mathbf{d}_k = 1 \tag{9}$$

$$\mathbf{d}_k \geqslant \mathbf{0} \quad k = 1, \ldots, K \tag{10}$$

where $\mathbf{1} = (1, 1, \ldots, 1)$ is a vector of ones. Constraints (7) and (8) ensure that the model is consistent with the classification of the reference alternatives on the basis of the classification rule (2). In these constraints δ is a small positive constant used to avoid arbitrary results that arise when the global value of an alternative equals a classification threshold. Constraint (9) normalizes the AVF such that an ideal alternative (i.e., with the most preferred levels in each criterion) receives a global value equal to one, whereas the non-negative constraints (10) on the parameters of the AVF model ensure that the marginal value functions are non-decreasing (assuming that all criteria are expressed in maximization form).

If the DM's classifications of the reference alternatives are consistent with an AVF evaluation model, then the polyhedron defined from the above constraints will be non-empty, thus implying that there is an infinite number of alternative AVFs (each corresponding to a feasible solution) consistent with the DM's judgments of the reference set. This raises the issue of how can a single representative AVF be chosen from the set of feasible solutions of the above constraints. This issue is even relevant when inconsistencies exist in the decision examples of the reference set, as these inconsistencies can be resolved (algorithmically or interactively with the DM; Mousseau, Figueira, Dias, Gomes da Silva, & Clímaco, 2003), thus making the robustness concern still relevant in this case too.

In the following subsections we present the alternative approaches considered in this study for selecting a single AVF representing the DM's classifications of the reference alternatives. The selected approaches, include: (a) a post-optimality procedure that was the first to be introduced in order to explore some characteristic feasible solutions to the polyhedron (7)–(10) and obtain a "central" decision model, (b) a max–min formulation that has been used in several studies in a robust PDA context, (c) a recently proposed analytic center formulation that operationalizes the "centrality" concept in a more rigorous manner (compared to ad hoc post-optimality procedures), and (d) a new model based on the concept of the Chebyshev center of a polyhedron, which can be identified with a linear programming formulation.

2.2. Post-optimality analysis

To cope with the existence of multiple decision models compatible with the DM's evaluations of the reference alternatives, Jacquet-Lagrèze and Siskos (1982) introduced a heuristic post-optimality procedure, which involves the solution of K pairs of linear programs, corresponding to the maximization and the minimization of the trade-off constant for each criterion k, i.e.:

$$\max/\min \ \{\mathbf{1}^{\top}\mathbf{d}_k | \text{s.t.}: \ (7){-}(10)\} \tag{11}$$

The $2K$ solutions obtained from this post-optimality process are some characteristic extreme solutions of (7)–(10), and their average can be used to form a "representative" AVF model as an approximation of the polyhedron's centroid solution. Such a centroid solution can be considered as representative of the feasible polyhedron of compatible models as it is less likely to be affected by changes in the DM's judgments on the reference alternatives (i.e., thus being more robust).

2.3. A max–min formulation

Max–min optimization formulations are often used in PDA in order to infer the parameters of decision models from assignment examples. For instance, in the context of multicriteria classification problems, such formulations have been used by Zopounidis and Doumpos (2000) in the MHDIS method, Dias, Mousseau, Figueira, and Clímaco (2002) in the ELECTRE method, whereas Greco et al. (2011) used max–min formulations to infer a representative value function in robust multiple criteria classification procedure. Similar models, in the context of ranking problems where also considered by Beuthe and Scannella (2001).

In the PDA setting considered in this study, a max–min formulation to infer a model compatible with the DM's judgments on a set of reference decision instance can be expressed as follows:

$$\max\{\delta \ | \text{s.t.}: (7){-}(10)\} \tag{12}$$

This formulation seeks to maximize the minimum separating gap between two consecutive classes. Bous et al. (2010) note that such a formulation shrinks the original polyhedron, thus forming a more "central" set and yielding solutions that are away from the boundaries of the original polyhedron (i.e., the obtained decision model satisfies the DM's preferences in a clearer and more robust manner).

The above max–min approach can also be explained on the grounds of the regularization principle, which is a popular approach in statistical and machine learning for improving the robustness of prediction models with respect to changes in the reference set (Hastie, Tibshirani, & Friedman, 2001). Based on this approach, Doumpos and Zopounidis (2007) introduced a formulation, which in the case of a consistent reference set, can be expressed as follows (a regularization approach to construct an additive preference model in the context of the dominance-based rough set approach has also been presented by Dembczyński, Kotłowski, & Słowiński (2006)):

$$\min \left\{ \sum_{k=1}^{K} \mathbf{1}^{\top}\mathbf{d}_k \ | \text{s.t.}: \ (7), \ (8), \ (10) \right\} \tag{13}$$

The main feature of the model is that the normalization constraint (9) is no longer taken into consideration. Instead, the AVF model is normalized after the solution of the above problem is obtained. In particular, denoting by F^* the optimal objective function value of (13), the normalized AVF model is simply obtained by dividing the optimal solution of (13) with F^* (Doumpos & Zopounidis, 2007 described the conditions under which it is possible to have $F^* = 0$; nevertheless, this is not possible when the reference set is consistent). The following theorem shows the connection between formulations (12) and (13).

Theorem 1. *The solutions of problems (12) and (13) are equivalent.*

Proof 1. Suppose that (13) is solved for some user-defined $\delta = \delta_0 > 0$ and let $F^* > 0$ be the optimal objective function value. The optimal solution of (13) normalized with the procedure described above, is feasible to (12) and yields an objective function value for (12) equal to δ_0/F^*. If there was a solution to (12) with $\delta > \delta_0/F^*$, then rescaling it (i.e., multiplying) by δ_0/δ leads to a feasible solution for (13) with objective function value $\delta_0/\delta < F^*$, which contradicts the initial hypothesis that F^* is the minimum value for the objective function of problem (13).

Similarly, suppose that (12) is solved and let $\delta^* > 0$ denote its optimal objective function value. This solution is feasible to (13) for $\delta = \delta^*$ and the corresponding objective function value is equal to one. If there was another solution to (13) with objective function value F such that $0 < F < 1$, then dividing it by F leads to a solution that is feasible to (12) with objective function value $\delta^*/F > \delta^*$, which contradicts the initial hypothesis that δ^* is the maximum objective function value for problem (12).

Thus, the optimal solutions of the two problems only differ by a scaling factor. In that regard they are equivalent. □

2.4. The analytic center approach

The third modeling approach used in this study is based on the analytic center formulation introduced by Bous et al. (2010). The analytic center of a polyhedron is defined by a feasible solution that maximizes the logarithmic barrier function of the constraints' slacks. In the context of this study we adapt the optimization model of Bous et al. (2010) to find the analytic center of the polyhendron defined by (7)–(10). This is performed through the solution of the following convex nonlinear program, which is easily solvable with existing algorithms (e.g., Newton's method).

$$\begin{array}{lll}
\max & \sum_{i=1}^{M}(\ln s_i^+ + \ln s_i^-) + \sum_{k=1}^{K}\mathbf{1}^\top \ln \mathbf{y}_k & \\
\text{subject to:} & V(\mathbf{x}_i) - t_\ell - s_i^+ = \delta & \forall\, i \in C_\ell,\ 1 \leqslant \ell \leqslant N-1 \\
& V(\mathbf{x}_i) - t_{\ell-1} + s_i^- = -\delta & \forall\, i \in C_\ell,\ 2 \leqslant \ell \leqslant N \\
& \mathbf{d}_k - \mathbf{y}_k = \mathbf{0} & k = 1,\ldots,K \\
& \sum_{k=1}^{K}\mathbf{1}^\top \mathbf{d}_k = 1 & \\
& s_i^+,\ s_i^-,\ t_\ell,\ \mathbf{y}_k,\ \mathbf{d}_k \geqslant 0 & \forall\, i,\ell,k
\end{array} \tag{14}$$

Compared to the previous approaches, this formulation is based on a more rigorous definition of "centrality" for the resulting decision model. Furthermore, from an optimization perspective the solution to the above problem is unique (Bous et al., 2010), thus minimizing the ambiguity that often arises due to the existence of multiple optimal solutions in linear programming formulations for inferring the parameters of decision models.

2.5. A new formulation based on the Chebyshev center

The last model that we test in this study is a new variant-extension of model (12). Effectively, (12) constructs an AVF such that the minimum "satisfaction" of the constraints (7) and (8) is maximized (i.e., the minimum separating gap between the categories). However, there is no rigorous association between this optimality objective with the characteristics of the polyhedron (7)–(10) and its robustness. The analytic center model described earlier seeks to address this issue, by focusing on identifying the analytic center of the polyhedron.

Alternatively, it is possible to construct an AVF model from the Chebyshev center of the polyhedron. The Chebyshev center corresponds to a feasible solution from which the largest possible ball of radius r can be inscribed within the polyhedron (Boyd & Vandenberghe, 2004). In this study we employ this approach to find the Chebyshev center of the polyhedron (7)–(10). The following linear programming model is used for this purpose (for details see Boyd & Vandenberghe, 2004):

$$\begin{array}{lll}
\max & r & \\
\text{subject to:} & V(\mathbf{x}_i) - t_\ell - a_i r \geqslant 0 & \forall\, i \in C_\ell,\ 1 \leqslant \ell \leqslant N-1 \\
& V(\mathbf{x}_i) - t_{\ell-1} + b_i r \leqslant 0 & \forall\, i \in C_\ell,\ 2 \leqslant \ell \leqslant N \\
& \mathbf{d}_k - \mathbf{1}r \geqslant \mathbf{0} & k = 1,\ldots,K \\
& \sum_{k=1}^{K}\mathbf{1}^\top \mathbf{d}_k = 1 & \\
& \mathbf{d}_k,\ t_\ell,\ r \geqslant 0 & \forall\, \ell,k
\end{array} \tag{15}$$

where a_i and b_i are the Euclidean norms of the decision variables' (the vectors $\mathbf{d}_1,\ldots,\mathbf{d}_K$ and the classification thresholds) coefficients in each of the constraints (7) and (8), e.g. $a_i = \|(\mathbf{p}_{i1}, \mathbf{p}_{i2}, \ldots, \mathbf{p}_{iK}, -1)\|_2$.

3. Experimental setting

The models presented in the previous section are tested and compared through a Monte Carlo simulation study based on artificially generated data, adopting an approach similar to the one used by Vetschera et al. (2010).

All data used in the experimental analysis are generated from the multivariate normal distribution with zero mean, unit variance and correlations uniformly distributed in [0,0.2]. Similarly to Vetschera et al. (2010) we take into consideration different settings for the dimensionality of the data, as defined by the number of alternatives in the reference set, the number of criteria and classes, as follows:

- Number of classes: $N = 2, 3, 4$.
- Number of reference alternatives per class: $M/N = 3, 5, 10, 15$.
- Number of criteria: $K = 3, 5, 7$.

With these specifications, the reference sets used in the analysis involve both low dimensionality and complexity data (e.g., six alternatives from two categories with three criteria), up to larger and more complex ones (up to 60 alternatives in four classes with seven criteria). In all cases, a secondary test sample is also used consisting of 50 alternatives from each category.

For each combination of the above three factors, 100 simulation runs are performed.[1] To generate the data in each run, two data pools are first generated, each consisting of 1000 alternatives. The first pool is used to select (at random) the alternatives of the reference set, whereas the test alternatives are drawn from the second pool.

The classification of the alternatives is performed with the following procedure. First, all alternatives in the two data pools are evaluated with a random AVF and their global values (scores) are

[1] As a robustness check, the analysis was repeated with an additional set of 100 simulations. The differences between the two tests were found to be statistically insignificant even at the 10% level according to the Mann–Whitney non-parametric test.

obtained. Then, appropriate classification thresholds $1 > t_1 > t_2 > \cdots > t_{N-1} > 0$ are specified at predefined percentiles of the global values of the alternatives in the data pool used to formulate the reference set (i.e., the definition of the thresholds is done independently of the data pool from which the test data are derived). In particular, for two-class problems the threshold t_1 that distinguishes between the two categories is set equal to the median of the global values. For the three-class problems we use the 30% and 70% percentiles to set t_2 and t_1, respectively, whereas for the four-class problems the 20%, 50%, and 80% percentiles are used to define the thresholds $t_3,\ t_2,\ t_1$. Thus, in the multi-class instances, more alternatives are distributed in intermediate categories than the extreme ones, which is a realistic assumption. With these thresholds, all alternatives in the two data pools are assigned to the predefined number of categories. Finally, from each category a random selection is performed to formulate the reference and test sets with the composition (number of alternatives per category) noted above.

For simplicity, it is assumed that the DM's preferences are compatible with a linear AVF. Thus, a randomly generated linear AVF is used in each run of the simulation experiment to classify the alternatives in the two data sets. It should be noted, however, that employing a linear AVF model is not a restrictive setting, as piecewise linear additive models are also linear with respect to their parameters (i.e., they are expressed in the linear form (6)). Nevertheless, given that, in realistic cases, the actual preferential structure of the DM is not really known, an analyst may decide to employ a more general modeling form (e.g., piecewise linear AVF) than the one implied by the reference data in order to be able to get more general conclusions and gain insights that a simpler model (e.g., linear AVF) may fail to capture. When working with additive value models, this is based on the fact that a piecewise linear AVF completely covers a linear AVF, and consequently whatever result is derived by a linear AVF can also be obtained with a piecewise linear model, while the opposite is generally not true. However, as piecewise linear AVFs have more degrees of freedom, their robust inference from small reference sets is more involved and a poor PDA formulation may fail to provide good results. Thus, the robustness properties of the feasible polyhedron (7)–(10) are not only affected by the characteristics of the reference data, but also by the form of the decision model (e.g., as the AVF model becomes more complex, the polyhedron widens and the choice of a representative solution becomes more challenging). From this perspective, we also consider the inference of piecewise linear AVFs (with three subintervals for all marginal value functions of the criteria) from the reference data described above, in order to analyze how the above issue affects the robustness and quality of the results obtained with different PDA formulations (i.e., how the results of PDA formulations are affected when the set of alternative compatible models becomes larger).

All computational experiments were performed in MATLAB® R2012b using a PC with a quad-core Intel i7-2600 K/3.4 GHz processor and 16 GB of RAM.

4. Results

4.1. Analyzing the robustness of compatible decision models

In order to investigate the robustness features of the selected formulations, first we analyze the polyhedron induced by the constraints (7)–(10), for each of the artificially generated reference sets. In particular, for each reference set a hit-and-run sampling approach (Kroese, Taimre, & Botev, 2011; Tervonen, van Valkenhoef, Baştürk, & Postmus, 2013) is used to generate 5000 AVFs uniformly distributed in the polyhedron of all AVFs compatible with the classification of the reference alternatives. As previously mentioned, the sampling of compatible AVFs is repeated twice, first with a linear AVF and then with a piecewise linear model.

With each of the 5000 sampled compatible models, two assignment rules are employed to classify the alternatives in the test samples:

- *Robust assignment rule*: Each alternative i is classified into one of the predefined categories using all sampled AVFs and a class acceptability index ($CAI_{i\ell}$) is calculated as the frequency of the assignment of alternative i in a category C_ℓ (Kadziński & Tervonen, 2013; Tervonen, Figueira, Lahdelma, Dias, & Salminen, 2009). Thus, the CAI represents the likelihood that an alternative is classified into a specific category, on the basis of the information provided by the DM's judgments in the reference set. The aggregate assignment is then defined by the majority rule (i.e., alternative i is assigned to the most likely category with the maximum CAI).[2]
- *Centroid assignment rule*: The sampled AVFs are averaged to produce a single AVF corresponding to the centroid of the feasible polyhedron. This centroid AVF is then employed to classify the alternatives from the test set.

The robust assignment rule provides a benchmark for analyzing the robustness of the results obtained from the optimization formulations presented in Section 2. In general, the assignments of the robust assignment rule cannot be reproduced (exactly) by a single decision model, as they result from the combination of multiple models. On the other hand, the use of multiple decision models makes it very difficult for the DM to get straightforward insights on how the final recommendations are obtained from the available data. The centroid assignment rule overcomes this shortcoming as it is based on a single AVF constructed by averaging all compatible decision models. The results of the average (centroid) AVF are expected to approximate the robust assignments, but discrepancies between the two averaging procedures may occur.

With the sampled compatible AVFs, the following measures are used to analyze the robustness of the preferential information that the reference data provide:

- *Mean class acceptability index (MCAI)*. As defined above, the class acceptability index $CAI_{i\ell}$ indicates the percentage of compatible AVFs that assign alternative i in category C_ℓ. The MCAI is then defined by averaging the acceptability indices over all test alternatives, under a particular classification rule (e.g., robust or centroid).[3] In particular, let $y_1, y_2, \ldots, y_{M_{test}}$ denote the class assignments for M_{test} test alternatives, obtained with a given decision model (i.e., classification rule), such that $y_i \in \{C_1, \ldots, C_N\}$, for all $i = 1, 2, \ldots, M_{test}$. Then, the MCAI is defined as follows:

$$MCAI = 100\frac{1}{M_{test}}\sum_{i=1}^{M_{test}} CAI_{iy_i}$$

A MCAI close to 100 indicates that the assignments obtained with the considered classification rule for the test alternatives are robust as they are verified by all AVFs compatible with the information provided in the reference set. It should be noted that given a set of AVFs sampled uniformly from the feasible polyhedron, the MCAI can be computed for any classification rule (even

[2] Under this majority rule it is possible that the maximum CAI is attained for two or more different classes, which would lead to an ambiguous assignment. Such cases were not observed in the simulation experiment. Nevertheless, avoiding such ambiguous situations can be easily done by using an odd number of models in the majority rule.

[3] Vetschera et al. (2010) referred to this index as "overall robustness index".

with a single additive value function), as it only requires the comparison of some specific class assignments (e.g., the ones of a classification model) to the ones of the randomly generated AVFs.

- *Mean entropy of the assignments obtained from all sample AVFs.* While MCAI focuses on a specific assignment of the alternatives, entropy is used to consider the variability (randomness) in the results of all sampled models. In this study we employ the following entropy measure for the classifications of an alternative i by all AVFs:

$$E_i = 100\left(1 + \frac{1}{\ln N}\sum_{\ell=1}^{N} CAI_{i\ell} \ln CAI_{i\ell}\right)$$

Alternatives for which this entropy measure is close to 100 are classified in a single category by all AVFs compatible with the reference set, whereas the ambiguity is maximum for alternatives with entropy close to zero (i.e., in such cases $CAI_{i\ell} \approx 1/N$, for all $\ell = 1, \ldots, N$). The mean entropy is then employed by averaging the above entropy measure over all alternatives in a test sample.

- *Mean coefficient of variation (CV) of the criteria trade-offs in the sampled AVFs*. The previous two measures focus on the classification assignments of the alternatives. However, it may happen that robust assignments are obtained from models with very different specifications of their parameters. Naturally this leads to some ambiguity on how a single decision model can be specified that will best represent the different compatible sets of the model's parameters. In that sense, the robustness concern is not solely restricted to the outputs of alternative decision models, but it also involves the structural form of these models and their parameters. In this experimental analysis we analyze this issue by measuring the variability of all compatible criteria trade-offs through the following CV measure:

$$\mathrm{CV} = \frac{1}{K}\sum_{k=1}^{K} \frac{\sigma_k}{\overline{w}_k}$$

where σ_k is the standard deviation of the trade-off constant of criterion k in the sampled AVFs and $\overline{w}_k$ is the corresponding mean value. The CV is close to zero in cases where the criteria trade-offs are almost the same in all AVFs compatible with the information provided by a reference set (i.e., $\sigma_k \approx 0$), whereas CV becomes higher in cases where the variability of the trade-offs increases, thus indicating that the DM's judgments on the reference alternatives can be represented by a set of very different AVFs.

Table 1 summarizes the results from an ANOVA on ranks nonparametric full factorial analysis (Conover & Iman, 1981; Sawilowsky, 1990) for the above three robustness indicators as well the classification accuracy (CA) of the robust assignment rule for the test alternatives. The factors considered in the analysis include the characteristics of the reference set (number of criteria, classes, and alternatives from each class) and the AVF modeling form. Under the entropy and the CV indexes all main effects and interactions are found significant at the 1% level. The same applies to the four main effects and all two-way interactions for the CA and the MCAI. As far as the higher-order interactions are concerned, the combination of the AVF modeling form, with the number of criteria, and the number of categories is insignificant under both the CA and the MCAI, whereas the three way interaction of the three factors that describe the data (criteria, classes, alternatives per case) is significant only at the 5% level.

More detailed summary results for the robustness indicators are presented in Table 2. The entries in the table are averages computed over all data sets with the number of criteria (K), categories (N), and reference alternatives per category (M/N) indicated in the first column. In accordance with the results of Vetschera et al. (2010), all three measures clearly indicate that robustness improves significantly as more information is embodied in the reference set (i.e., when the number of alternatives in the reference set from each category (M/N) increases). On the other hand, as the number of criteria (K) increases, robustness gets lower. This is explained by the increase in the variability of the decision models, which is evident in the CV for the criteria's trade-offs.

As far as the effect of the number of categories (N) is concerned, the entropy measure and the CV of the criteria trade-offs indicate that robustness increases in problems with more than two categories. The MCAI, on the other hand provides mixed indications with minor differences in the case where a linear AVF is inferred (according to the Kruskal–Wallis test, the differences between the different settings for the number of categories are not significant at the 1% level), whereas with a piecewise linear AVF the MCAI decreases as the number of categories increases. The observed discrepancies for the three robustness indicators imply that in classification problems, robustness comparisons between problems with different number of categories should be made with caution when based on such measures of robustness.

The variability of the classification results as measured with the entropy index and the MCAI is consistently higher when a piecewise AVF decision model is employed. On the other hand, the coefficient of variation for the criteria trade-offs is lower for piecewise linear models compared to the case where a linear AVF is used. However, it should be noted that for a piecewise linear AVF, the criteria trade-offs are not the only parameters that define the decision model (the form of the marginal value functions is an additional important parameter). Thus, even though the trade-offs may exhibit lower variability in this case, the implications of this result are not directly comparable to the case of a linear model.

Overall, it is worth noting that except for the data characteristics of the reference data and the problem, the results confirm that the specification of an appropriate modeling form is an important factor related to the robustness of the results (this issue has also been highlighted by Stewart (1993, 1996)). In particular, using a more complex model than the one that actually expresses the DM's judgments in the reference set has a significant negative effect on the robustness of the information that the reference set provides.

As shown in Table 3 this has further implications for the performance (classification accuracy) of the models when applied to evaluate alternatives outside the reference set (i.e., test sample). The reported results for the robust assignment rule clearly indicate that increasing the degrees of freedom of the inferred decision model has a negative effect (the significance of the differences between the linear and the piecewise linear models was confirmed with the Wilcoxon signed-rank test at the 1% level). The effects of the number of criteria, the number of classes, and the number of reference alternatives in each category are very similar to the findings discussed above for the three robustness indicators. This is in line with the results reported by Vetschera et al. (2010) on the positive association between classification accuracy and robustness. Nevertheless, similarly to the remark made earlier on the interpretation of the robustness indicators for problems with different number of categories, it should again be noted that establishing a robustness–accuracy connection when referring to problems with such different characteristics, seems to be troublesome and deserves further analysis.

As far as the discrepancies between the robust and centroid assignment rules are concerned, they were found to be very limited, as the percentage of test alternatives for which the two rules provided different results was limited to 1–1.5% (on average) for reference sets with three alternatives per class and less than 1% for larger reference sets. This finding confirms that recommenda-

Table 1
ANOVA on ranks (F ratios and p values).

	CA	MCAI	Entropy	CV
AVF	1501.3 (0.000)	2495.6 (0.000)	2506.7 (0.000)	3632.0 (0.000)
Criteria	661.1 (0.000)	1566.2 (0.000)	1919.0 (0.000)	3512.4 (0.000)
Altern.	2221.5 (0.000)	7105.9 (0.000)	6876.3 (0.000)	3066.1 (0.000)
Classes	26.6 (0.000)	375.3 (0.000)	2559.2 (0.000)	4116.0 (0.000)
AVF × Criteria	6.0 (0.003)	7.7 (0.000)	7.8 (0.000)	99.3 (0.000)
AVF × Altern.	25.5 (0.000)	114.9 (0.000)	53.3 (0.000)	83.9 (0.000)
AVF × Classes	169.2 (0.000)	358.0 (0.000)	155.1 (0.000)	65.7 (0.000)
Criteria × Altern.	3.8 (0.001)	10.6 (0.000)	7.8 (0.000)	12.5 (0.000)
Criteria × Classes	4.2 (0.002)	4.1 (0.002)	4.2 (0.002)	11.6 (0.000)
Altern. × Classes	27.0 (0.000)	59.0 (0.000)	17.0 (0.000)	34.5 (0.000)
AVF × Criteria × Altern.	6.4 (0.000)	20.7 (0.000)	12.6 (0.000)	5.0 (0.000)
AVF × Criteria × Classes	0.9 (0.476)	1.4 (0.235)	4.8 (0.001)	7.5 (0.000)
AVF × Altern. × Classes	4.5 (0.000)	7.1 (0.000)	20.6 (0.000)	5.6 (0.000)
Criteria × Altern. × Classes	1.8 (0.039)	1.9 (0.029)	15.8 (0.000)	46.0 (0.000)
AVF × Criteria × Altern. × Classes	1.7 (0.067)	2.5 (0.003)	2.6 (0.002)	8.8 (0.000)

Table 2
Averages of robustness indicators.

		Linear AVF			Piecewise linear AVF		
		MCAI	Entropy	CV	MCAI	Entropy	CV
K	3	92.35	82.78	0.32	89.40	76.93	0.28
	5	89.77	76.97	0.44	86.73	71.13	0.34
	7	88.08	73.21	0.51	85.50	68.59	0.37
N	2	90.34	70.02	0.53	89.73	67.66	0.39
	3	89.70	79.12	0.40	86.56	72.77	0.33
	4	90.16	83.82	0.33	85.35	76.22	0.28
M/N	3	83.47	64.30	0.54	81.43	60.87	0.39
	5	88.26	73.67	0.47	85.43	68.55	0.36
	10	93.27	84.22	0.36	89.89	77.55	0.30
	15	95.27	88.42	0.32	92.09	81.90	0.28
	Overall	90.07	77.65	0.42	87.21	72.22	0.33

Table 3
Classification accuracies for the robust assignment rule.

		K			N			
		3	5	7	2	3	4	Overall
Linear AVF								
M/N	3	90.81	86.90	84.54	87.40	86.65	88.21	*87.42*
	5	94.14	91.76	90.38	91.30	91.61	93.37	*92.09*
	10	97.59	96.14	94.84	94.67	96.46	97.45	*96.19*
	15	98.59	97.72	97.10	96.40	98.06	98.95	*97.80*
	Overall	*95.28*	*93.13*	*91.72*	*92.44*	*93.19*	*94.50*	*93.38*
Piecewise linear AVF								
M/N	3	87.52	83.21	81.14	87.02	83.33	81.52	*83.96*
	5	91.38	87.76	85.77	90.40	87.69	86.81	*88.30*
	10	94.96	92.39	90.54	93.04	92.02	92.82	*92.63*
	15	96.74	94.80	93.29	94.97	94.30	95.56	*94.94*
	Overall	*92.65*	*89.54*	*87.68*	*91.36*	*89.34*	*89.18*	*89.96*

tions obtained from a decision model defined by the centroid of the set of solutions which are compatible with the DM's judgments, are robust in the sense that the likelihood of obtaining different recommendations with other compatible models is minimized. This justifies the attempts made in past studies to development formulations and approaches that aim to build decision models corresponding to some central solutions (e.g., the post-optimality of Section 2.2 or the formulations in Sections 2.4 and 2.5). The results in the next subsection focus on the comparison of the results obtained from such approaches, in light of the robustness results presented above.

4.2. Comparative analysis of the selected PDA formulations

The analysis in the previous section focused on providing some basic results on the characteristics of the polyhedron defined by the set of decision models compatible with DM's judgments on the reference alternatives. These results constitute the basis for comparing the four approaches described in Section 2 for building a decision model that best represents the reference data, namely:

- the basic post-optimality approach (11),
- the max–min model (12),
- the analytic center model (14), and
- the Chebyshev center model (15).

The comparative results presented in this section will be discussed in relation to: (a) the class acceptabilities (confidence) of the assignments of the models constructed with each approach, (b) the relationship between the parameters of the models and the ones corresponding to the actual and centroid models, and (c) the classification performance of the models when applied to the test samples.

4.2.1. The assignments' acceptability

Tables 4–6 summarize the results for the MCAI obtained from the four approaches, under the three main design factors (criteria, alternatives, classes). Each table presents the relative percentage differences between the MCAI obtained with the robust assignment rule and the ones from each of the four tested approaches. Obviously this difference is by definition non-positive as the robust rule assigns the alternatives into the most likely category. Having a single decision model providing similar results to the robust rule would be convenient in the context of robust decision aid.

The obtained results indicate that the decision models constructed with the analytic center formulation are the best performers overall, followed by the models obtained with the Chebyshev model, whereas the basic post-optimality approach and the max–min model provide worse results. This holds for both linear and piecewise linear AVFs, with the only difference being that the max–min model outperforms (overall) the post-optimality approach in the latter case, whereas in the former case the relative performance of these two approaches is reversed. Overall, and in accordance with the results reported in the previous section, the divergences between the models obtained with the four approaches and the robust rule become much larger as the degrees of freedom of the decision model increase. Nevertheless, this effect is weaker for the Chebyshev and the analytic center models.

As far as the number of criteria is involved (Table 4), the divergences between the four approaches and the robust rule, become larger as the number of criteria increases. This negative effect is higher for the max–min model compared to the other approaches (e.g., the divergence with seven criteria is more than double the one with three criteria).

The increase in the number of reference alternatives has a strong positive effect on the acceptability of the assignments pro-

Table 4
Relative percentage differences in MCAI compared to the robust assignment rule, by the number of criteria.

	3	5	7	Overall
Linear AVF				
Post-optimality	−1.71	−2.04	−3.10	−2.27
Max–min	−1.43	−2.55	−3.62	−2.52
Chebyshev cntr.	−1.28	−1.86	−2.37	−1.83
Analytic cntr.	−0.58	−0.88	−1.06	−0.84
Piecewise linear AVF				
Post-optimality	−7.15	−8.08	−8.86	−8.02
Max–min	−3.91	−6.55	−8.94	−6.43
Chebyshev cntr.	−2.31	−3.23	−3.99	−3.16
Analytic cntr.	−1.76	−2.12	−2.19	−2.02

Table 5
Relative percentage differences in MCAI compared to the robust assignment rule, by the number of reference alternatives.

	3	5	10	15	Overall
Linear AVF					
Post-optimality	−5.30	−2.66	−0.97	−0.54	−2.27
Max–min	−5.54	−3.03	−1.18	−0.71	−2.52
Chebyshev cntr.	−3.74	−2.21	−1.00	−0.61	−1.83
Analytic cntr.	−1.10	−0.97	−0.71	−0.60	−0.84
Piecewise linear AVF					
Post-optimality	−14.82	−10.17	−4.94	−3.01	−8.02
Max–min	−11.80	−7.84	−4.13	−2.62	−6.43
Chebyshev cntr.	−5.09	−3.81	−2.38	−1.62	−3.16
Analytic cntr.	−1.73	−2.10	−2.15	−2.06	−2.02

Table 6
Relative percentage differences in MCAI compared to the robust assignment rule, by the number of reference classes.

	2	3	4	Overall
Linear AVF				
Post-optimality	−2.53	−2.20	−2.09	−2.27
Max–min	−3.57	−2.31	−1.68	−2.52
Chebyshev cntr.	−2.09	−1.78	−1.62	−1.83
Analytic cntr.	−0.95	−0.84	−0.71	−0.84
Piecewise linear AVF				
Post-optimality	−7.56	−8.78	−7.73	−8.02
Max–min	−7.01	−6.56	−5.68	−6.43
Chebyshev cntr.	−2.75	−3.38	−3.37	−3.16
Analytic cntr.	−1.73	−2.16	−2.18	−2.02

duced by the four approaches (Table 5). Under a linear AVF with 15 references alternatives from each category, the MCAI for the results of the four approaches is very close to the MCAI of the robust assignment rule and the differences between the alternative approaches are limited (with the post-optimality approach producing slightly better results). In cases where a piecewise linear AVF is inferred, the post-optimality approach together with the max–min model and the Chebyshev center formulation improve the most with the use of more reference alternatives, whereas the results for the analytic center model appear to slightly worsen.

Finally, with respect to the number of categories (Table 6), all models provide better results in cases with four categories when a linear AVF is inferred, whereas with a piecewise linear model the effect of the number of categories appears mixed and less clear.

Table 7 provides a summary comparison of the four approaches in terms of their performance on MCAI. For each combination of the design factors (criteria, alternatives, classes; 36 combinations overall), the differences between each pair of approaches were assessed in terms of their statistical significance with a one-tailed Wilcoxon signed-rank test (at the 1% significance level with the Bonferroni–Holm correction to account for multiple comparisons). The table reports the number of factor combinations in which an approach (row) performed significantly better than another (column). As shown in the obtained results, the analytic center approach was never significantly outperformed by the other approaches. On the other hand, under a linear AVF setting it performed significantly better than the other approaches in a considerable number of factor combinations (19–21), mostly in cases with a small number of reference alternatives. With the piecewise linear setting, the number of cases where the analytic center model performed significantly better than the rest of the approaches is higher, again involving mostly cases with small references sets (for instance, no significant differences were observed in comparison to the Chebyshev model with references consisting of 10–15 alternatives from each category).

The above results on the relationship between the most robust assignments and the ones obtained with the models constructed with the four approaches, were also confirmed through the examination of the percentage of test alternatives for which the robust assignment was different from the results of the inferred AVFs. Under the linear AVF setting, this was found to be 3.88% for the analytic center model (on average), as opposed to 6.05% for the Chebyshev model, 6.78% for the post-optimality approach, and 7.08% for the max–min model. On the other hand, with the piecewise linear AVF, the frequency with which differences were observed from the robust assignment was high consistently higher for all approaches (6.94% for the analytic center, 9.06% for the Chebyshev model, 13.54% for the max–min model, and 15.46% for the post-optimality approach).

4.2.2. Criteria trade-offs

Except for the analysis of the assignments of the models developed with the four considered approaches we also examine the estimations obtained with regard to the criteria trade-offs in the constructed AVFs. It is worth noting that three of the approaches used in the comparison (i.e., the post-optimality approach, as well as the Chebyshev and analytic center models) are based on the identification of central solutions within the feasible set of a model's parameters. In that regard, we examine the relationship between the trade-offs in the models obtained with the considered formulation in comparison to the centroid model. Furthermore, comparisons are also performed with the trade-offs in the actual decision model used to classify the data. The mean absolute deviations (MAD) in the trade-off vectors for these comparisons are summarized in Tables 8–10.

With a linear AVF the centroid solution obtained by averaging the simulated compatible decision models, is the one that is closer to the actual trade-offs in the decision model used to classify the data (with an overall MAD equal to 4.29%), followed by the analytic center model. The Chebyshev model performs slightly better than the post-optimality approach, mainly in more complex cases (i.e., problems with seven criteria, four classes, and 10–15 reference alternatives from each category). Overall, the differences between the two methods were found significant at the 1% level according to the Wilcoxon signed-rank test. The Chebyshev model even outperforms the analytic center approach in problems with four categories (the difference being significant at the 1% level), as well as when larger reference sets are employed (i.e., 10–15 reference alternatives in each category; differences significant at the 1% level). On the other hand, the trade-offs in the models obtained with the max–min model are the ones that are most different from the actual trade-offs (with an overall MAD equal to 6.1%), even though its results improve significantly in multi-class instances (e.g., with four classes) as well as with larger reference sets.

As the degrees of freedom of the decision model increase (piecewise linear AVF), the trade-offs obtained with the analytic center model are the ones that best match the actual trade-offs

Table 7
Pairwise comparison of the methods in terms of MCAI.

	Post-optimality	Max–min	Chebyshev cntr.	Analytic cntr.
Post-optimality	–	3 (3)	0 (0)	0 (2)
Max–min	4 (13)	–	0 (0)	1 (1)
Chebyshev cntr.	5 (30)	13 (32)	–	0 (3)
Analytic cntr.	20 (30)	21 (28)	19 (16)	–

Number of factor combinations in which the approach in row performed significantly better than the approach in column (one-tailed Wilcoxon signed-rank test at the 1% significance level with Bonferroni–Holm correction); linear AVF outside the parentheses and piecewise linear in the parentheses.

Table 8
Mean absolute deviations (in %) from the criteria trade-offs of the actual model (deviations from the centroid model in parentheses), by the number of criteria.

	3	5	7	Overall
Linear AVF				
Centroid	4.53	4.42	3.94	4.29
Post-optimality	5.26	5.35	4.95	5.19
	(2.79)	(3.11)	(3.37)	(3.09)
Max–min	6.39	6.23	5.69	6.10
	(4.66)	(4.88)	(4.72)	(4.76)
Chebyshev cntr.	5.67	5.23	4.67	5.19
	(4.00)	(3.61)	(3.27)	(3.63)
Analytic cntr.	5.55	5.02	4.34	4.97
	(3.11)	(2.55)	(2.08)	(2.58)
Piecewise linear AVF				
Centroid	6.94	6.12	5.22	6.10
Post-optimality	10.21	8.62	7.66	8.83
	(8.66)	(4.91)	(3.61)	(5.73)
Max–min	8.89	7.77	6.92	7.86
	(7.40)	(5.38)	(4.44)	(5.74)
Chebyshev cntr.	7.07	6.06	5.14	6.09
	(4.41)	(2.80)	(2.10)	(3.10)
Analytic cntr.	6.19	5.56	4.76	5.50
	(4.19)	(2.27)	(1.48)	(2.65)

Table 9
Mean absolute deviations (in %) from the criteria trade-offs of the actual model (deviations from the centroid model in parentheses), by the number of alternatives.

	3	5	10	15	Overall
Linear AVF					
Centroid	6.73	4.97	3.14	2.33	4.29
Post-optimality	8.30	5.99	3.73	2.73	5.19
	(5.24)	(3.60)	(2.04)	(1.47)	(3.09)
Max–min	10.58	7.12	3.83	2.89	6.10
	(8.61)	(5.47)	(2.93)	(2.02)	(4.76)
Chebyshev cntr.	8.53	6.04	3.52	2.67	5.19
	(5.92)	(4.14)	(2.56)	(1.90)	(3.63)
Analytic cntr.	7.32	5.66	3.75	3.15	4.97
	(3.02)	(2.87)	(2.35)	(2.08)	(2.58)
Piecewise linear AVF					
Centroid	8.45	6.80	5.10	4.03	6.10
Post-optimality	13.23	10.34	6.71	5.04	8.83
	(7.92)	(6.63)	(4.66)	(3.69)	(5.73)
Max–min	12.16	8.86	5.76	4.66	7.86
	(8.25)	(6.56)	(4.56)	(3.60)	(5.74)
Chebyshev cntr.	8.21	6.67	5.13	4.34	6.09
	(3.55)	(3.38)	(2.93)	(2.54)	(3.10)
Analytic cntr.	7.89	6.08	4.36	3.69	5.50
	(2.00)	(2.53)	(2.98)	(3.07)	(2.65)

(overall MAD equal to 5.5%), followed by the Chebyshev model and the centroid solution, which both produce similar results. In this case, the post-optimality approach provides the worst results (overall MAD equal to 8.83%).

Table 10
Mean absolute deviations (in %) from the criteria trade-offs of the actual model (deviations from the centroid model in parentheses), by the number of classes.

	2	3	4	Overall
Linear AVF				
Centroid	6.48	3.87	2.53	4.29
Post-optimality	7.93	4.51	3.11	5.19
	(4.89)	(2.64)	(1.74)	(3.09)
Max–min	10.11	5.10	3.11	6.10
	(8.45)	(3.63)	(2.19)	(4.76)
Chebyshev cntr.	7.97	4.63	2.97	5.19
	(5.66)	(3.10)	(2.12)	(3.63)
Analytic cntr.	7.45	4.43	3.03	4.97
	(3.86)	(2.29)	(1.59)	(2.58)
Piecewise linear AVF				
Centroid	8.64	5.66	3.99	6.10
Post-optimality	12.22	8.33	5.94	8.83
	(7.88)	(5.50)	(3.80)	(5.73)
Max–min	12.07	6.83	4.68	7.86
	(8.59)	(5.23)	(3.40)	(5.74)
Chebyshev cntr.	8.42	5.70	4.14	6.09
	(3.64)	(3.24)	(2.42)	(3.10)
Analytic cntr.	7.49	5.17	3.85	5.50
	(2.95)	(2.84)	(2.15)	(2.65)

The trade-offs estimated through the analytic center model are also the ones that are most similar to the centroid solution, under both a linear and piecewise linear modeling setting (overall MAD equal to 2.58% and 2.65%, respectively). With a linear AVF the trade-offs produced with the post-optimality approach are closer to the centroid than the ones of the Chebyshev and max–min models, whereas under the piecewise linear case the Chebyshev model outperforms the two other approaches. These results indicate that, generally, the two approaches that operationalize the centroid concept directly into the model inference process (i.e., analytic and Chebyshev centers) do produce results that are indeed closer to the actual centroid of the feasible polyhedron, particularly in more constrained cases (i.e., with large reference sets, more criteria, and categories). However, with larger polyhedra derived from small-size reference data, such models may still provide poor proxies of centroid solutions.

4.2.3. Classification accuracy

At the final stage of the analysis, the classification accuracy of the models is examined for the test alternatives (out of sample accuracy). In this context, we define classification accuracy as the ratio between the number of correct classifications produced by a model (for the test alternatives) to the number of alternatives in a test sample. Detailed results are presented in Tables 11–13.

The overall results indicate that the analytic center formulation provides the highest accuracies, under both the linear and piecewise linear setting for the form of the AVF models, followed by the Chebyshev and max–min models, whereas the post-optimality approach provides the worst results. Compared to the robust assignment rule (cf. Table 3) the accuracies of the four approaches are consistently lower. With a linear AVF, the overall differences compared to the robust rule range between 0.98% for the analytic center model and 2.75% for the post-optimality approach, whereas for the piecewise linear setting they even higher (2.14% for the analytic center model up to 7.65% for the post-optimality approach). These results further confirm the association between robustness and classification performance which was found in Section 4.1 and also identified by Vetschera et al. (2010).

The classification performance of the analytic center model shows lower variability (compared to the other approaches) across the different settings for the number of criteria and alternatives. On the other hand, the improvement obtained with larger refer-

Table 11
Classification accuracies (in %) by the number of criteria.

	3	5	7	Overall
Linear AVF				
Post-optimality	93.01	90.70	88.17	90.63
Max–min	93.50	90.46	88.04	90.66
Chebyshev cntr.	93.67	91.28	89.40	91.45
Analytic cntr.	94.57	92.01	90.61	92.39
Piecewise linear AVF				
Post-optimality	84.74	82.15	80.03	82.31
Max–min	88.30	83.59	80.66	84.18
Chebyshev cntr.	89.86	86.44	84.69	86.99
Analytic cntr.	90.27	87.27	85.90	87.82

Table 12
Classification accuracies (in %) by the number of reference alternatives.

	3	5	10	15	Overall
Linear AVF					
Post-optimality	81.74	88.97	94.88	96.92	90.63
Max–min	81.77	88.86	95.03	97.00	90.66
Chebyshev cntr.	83.56	89.82	95.29	97.13	91.45
Analytic cntr.	86.21	91.05	95.49	96.82	92.39
Piecewise linear AVF					
Post-optimality	71.08	78.84	87.71	91.59	82.31
Max–min	74.55	81.48	88.56	92.14	84.18
Chebyshev cntr.	80.14	84.73	90.07	93.04	86.99
Analytic cntr.	82.86	86.49	89.94	91.97	87.82

Table 13
Classification accuracies (in %) by the number of classes.

Alternatives	2	3	4	Overall
Linear AVF				
Post-optimality	89.66	90.64	91.58	90.63
Max–min	88.81	90.51	92.68	90.66
Chebyshev cntr.	90.47	91.15	92.73	91.45
Analytic cntr.	91.49	92.21	93.48	92.39
Piecewise linear AVF				
Post-optimality	84.07	81.10	81.75	82.31
Max–min	85.55	83.00	84.00	84.18
Chebyshev cntr.	89.36	85.71	85.91	86.99
Analytic cntr.	90.46	86.74	86.25	87.82

Table 14
Pairwise comparison of the methods in terms of their classification accuracy.

	Post-optimality	Max–min	Chebyshev cntr.	Analytic cntr.
Post-optimality	–	0 (0)	0 (0)	0 (2)
Max–min	0 (7)	–	0 (0)	0 (3)
Chebyshev cntr.	0 (26)	6 (25)	–	0 (5)
Analytic cntr.	11 (24)	12 (22)	3 (8)	–

Number of factor combinations in which the approach in row performed significantly better than the approach in column (one-tailed Wilcoxon signed-rank test at the 1% significance level with Bonferroni–Holm correction); linear AVF outside the parentheses and piecewise linear in the parentheses.

ence sets is stronger for the other approaches. In fact, the max–min and the Chebyshev models outperform the analytic center approach in problems with 15 reference alternatives under both the linear and the piecewise linear modeling settings (the differences being significant in favor of the max–min and Chebyshev models under the piecewise linear setting in multi-category problem instances, according to a one-tailed Wilcoxon signed-rank test at the 1% significance level with the Bonferroni–Holm correction to account for multiple comparisons). Finally, similarly to the results discussed previously, the effect of the number of categories seems to be mixed, as higher accuracies are obtained in multi-class problems when a linear AVF is employed, whereas under the piecewise linear AVF setting the accuracies are higher in two-class problems.

Table 14 presents a summary comparison of the four approaches in terms of the statistical significance of the differences in their classification accuracies (at the 1% level). With the linear modeling setting there were 11 factor combinations where the analytic center model outperformed the post-optimality approach, 12 cases where it outperformed the max–min model, and three cases where it performed significantly better than the Chebyshev center model (all three cases involved small data sets with three alternatives from each category and two or three categories). With a piecewise linear AVF, there is an increase in the number of factor combinations in which the analytic center model outperformed the other approaches, but the same is also observed in the opposite direction. In particular, the post-optimality approach performed significantly better than the analytic center model in two instances, the max–min model in three cases, and the Chebyshev model in five cases. Again, all these instances involved multi-class data with 15 reference alternatives from each category.

5. Conclusions

In this study we presented an experimental investigation of some typical and recently proposed approaches for building a single AVF decision model representing the DM's judgments on a set of reference examples in a PDA framework for classification problems. A new approach based on the Chebyshev center of the feasible polyhedron for the decision model's parameters was also introduced.

The obtained results lead to conclusions and suggestions, which analysts, researchers, and DMs should consider when using PDA approaches for inferring preferential information and constructing decision models from data. Among others, the following main points can be highlighted:

- There is a strong positive association between the robustness of the recommendations obtained from a multicriteria decision model with central solutions of the polyhedron that describes the model's parameters. This was confirmed by the similarity of the results obtained under the robust and centroid classification rules as well as the good results that the analytic and Chebyshev center formulations provided compared to other model inference approaches.
- The differences between alternative model inference formulations become larger in cases where the polyhedron of the model's parameters is wide.
- Among the characteristics of the reference data, the number of reference alternatives from each category seems to be the most decisive factor, whereas on the modeling side, the number of free parameters of a model (i.e., its degrees of freedom) is also critical issue. On the other hand, robustness comparisons between problems with different number of categories can be troublesome (as alternative robustness measures may lead to conflicting indications).

These findings suggest that the use of a good model inference formulation can indeed make a significant difference in a PDA context, particularly when working with small reference sets and complex models. Approaches that operationalize the search for central solutions seem to be the best options in such situations. On the other hand, the aggregation of a limited set of (rather arbitrary selected) extreme feasible solutions generated with post-optimality techniques may yield poor results. In any case, larger reference sets should be sought (whenever possible), which will not only improve the formulation of more robust recommendations and accurate models, but also reduce the impact of the model inference approach employed.

However, the results of this study indicate that there is still room for developing new improved model inference formulations. Despite the good performance (relative to the other approaches considered in this study) of the models based on the concepts of the Chebyshev and the analytic center, their results were found to be inferior compared to the ones obtained with the robust and centroid rules. Thus, it is worthwhile to investigate the possibility of introducing new approaches (including interactive procedures, e.g., Greco et al. (2011)) that will facilitate the inference of better and more robust models, based on improved considerations of the concept of ‘‘centrality’’ for the feasible polyhedron.

Except for the above issue, the robustness concern should also be explored in a more general context of model selection, considering not only model complexity (that was considered in this study), but also the effect of using different modeling forms (i.e., other functional, relational or symbolic models such as outranking methods and decision rule approaches), including cases where the selected type of model has incompatibilities with the DM's system of preferences (e.g., when a model is inadequate to represent a complex preference structure). Other problem settings such as ordinal regression and choice problems can also be considered, together with further experimentation on real-world data. Finally, emphasis should be given to the construction of well-founded and meaningful indicators for measuring robustness with PDA approaches in a unified context applicable to different instances.

Acknowledgment

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program ‘‘Education and Lifelong Learning’’.

References

Beuthe, M., & Scannella, G. (2001). Comparative analysis of UTA multicriteria methods. *European Journal of Operational Research, 130*(2), 246–262.

Bous, G., Fortemps, P., Glineur, F., & Pirlot, M. (2010). ACUTA: A novel method for eliciting additive value functions on the basis of holistic preference statements. *European Journal of Operational Research, 206*(2), 435–444.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge: Cambridge University Press.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*(3), 124.

Dembczyński, K., Kotłowski, W., & Słowiński, R. (2006). Additive preference model with piecewise linear components resulting from dominance-based rough set approximations. In L. Rutkowski, R. Tadeusiewicz, Lotfi A. Zadeh, & J. M. Zurada (Eds.). *Artificial intelligence and soft computing – ICAISC 2006. Lecture notes in computer science* (4029, pp. 499–508). Berlin Heidelberg: Springer.

Dias, L., Mousseau, V., Figueira, F., & Clímaco, J. (2002). An aggregation/disaggregation approach to obtain robust conclusions with ELECTRE TRI. *European Journal of Operational Research, 138*(2), 332–348.

Doumpos, M., & Zopounidis, C. (2007). Regularized estimation for preference disaggregation in multiple criteria decision making. *Computational Optimization and Applications, 38*(1), 61–80.

Greco, S., Kadziński, M., & Słowiński, R. (2011). Selection of a representative value function in robust multiple criteria sorting. *Computers & Operations Research, 38*(11), 1620–1637.

Greco, S., Mousseau, V., & Słowiński, R. (2010). Multiple criteria sorting with a set of additive value functions. *European Journal of Operational Research, 207*(3), 1455–1470.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Jacquet-Lagrèze, E., & Siskos, J. (1982). Assessing a set of additive utility functions for multicriteria decision making: The UTA method. *European Journal of Operational Research, 10*(2), 151–164.

Jacquet-Lagrèze, E., & Siskos, Y. (2001). Preference disaggregation: 20 Years of MCDA experience. *European Journal of Operational Research, 130*(2), 233–245.

Kadziński, M., Greco, S., & Słowiński, R. (2012). Extreme ranking analysis in robust ordinal regression. *Omega, 40*(4), 488–501.

Kadziński, M., Greco, S., & Słowiński, R. (2013). Selection of a representative value function for robust ordinal regression in group decision making. *Group Decision and Negotiation, 22*(3), 429–462.

Kadziński, M., & Tervonen, T. (2013). Stochastic ordinal regression for multiple criteria sorting problems. *Decision Support Systems, 55*(1), 55–66.

Keeney, R., & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value trade-offs*. Cambridge: Cambridge University Press.

Kroese, D. P., Taimre, T., & Botev, Z. I. (2011). *Handbook of Monte Carlo methods*. New York: Wiley.

Mousseau, V., Figueira, J., Dias, L., Gomes da Silva, C., & Clímaco, J. (2003). Resolving inconsistencies among constraints on the parameters of an MCDA model. *European Journal of Operational Research, 147*(1), 72–93.

Roy, B. (2010). Robustness in operational research and decision aiding: A multi-faceted issue. *European Journal of Operational Research, 200*(3), 629–638.

Sawilowsky, S. (1990). Nonparametric tests of interaction in experimental design. *Review of Educational Research, 60*(1), 91–126.

Siskos, Y., Grigoroudis, E., & Matsatsinis, N. F. (2005). UTA methods. In J. Figueira, S. Greco, & M. Ehrogott (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 297–334). New York: Springer.

Stewart, T. J. (1993). Use of piecewise linear value functions in interactive multicriteria decision support: A Monte Carlo study. *Management Science, 39*(11), 1369–1381.

Stewart, T. J. (1996). Robustness of additive value function methods in MCDM. *Journal of Multi-Criteria Decision Analysis, 5*(4), 301–309.

Tervonen, T., Figueira, J., Lahdelma, R., Dias, J., & Salminen, P. (2009). A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research, 192*(1), 236–242.

Tervonen, T., van Valkenhoef, G., Baştürk, N., & Postmus, D. (2013). Hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research, 224*(3), 552–559.

Vetschera, R., Chen, Y., Hipel, K. W., & Kilgour, M. D. (2010). Robustness and information levels in case-based multiple criteria sorting. *European Journal of Operational Research, 202*(3), 841–852.

Zopounidis, C., & Doumpos, M. (2000). Building additive utilities for multi-group hierarchical discrimination: The M.H.Dis method. *Optimization Methods and Software, 14*(3), 219–240.

Zopounidis, C., & Doumpos, M. (2002). Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research, 138*(2), 229–246.

European Journal of Operational Research 240 (2015) 193–201

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Robustness analysis in Multi-Objective Mathematical Programming using Monte Carlo simulation

George Mavrotas [a,*], Olena Pechak [a], Eleftherios Siskos [b], Haris Doukas [b], John Psarras [b]

[a] Laboratory of Industrial and Energy Economics, School of Chemical Engineering, National Technical University of Athens, Zografou Campus, 15780 Athens, Greece
[b] Decision Support Systems Laboratory (EPU NTUA), School of Electrical and Computer Engineering, National Technical University of Athens, Zografou Campus, 15780 Athens, Greece

ARTICLE INFO

Article history:
Received 31 October 2013
Accepted 25 June 2014
Available online 3 July 2014

Keywords:
Multi-objective programming
Robustness analysis
Monte Carlo simulation

ABSTRACT

In most multi-objective optimization problems we aim at selecting the most preferred among the generated Pareto optimal solutions (a subjective selection among objectively determined solutions). In this paper we consider the robustness of the selected Pareto optimal solution in relation to perturbations within weights of the objective functions. For this task we design an integrated approach that can be used in multi-objective discrete and continuous problems using a combination of Monte Carlo simulation and optimization. In the proposed method we introduce measures of robustness for Pareto optimal solutions. In this way we can compare them according to their robustness, introducing one more characteristic for the Pareto optimal solution quality. In addition, especially in multi-objective discrete problems, we can detect the most robust Pareto optimal solution among neighboring ones. A computational experiment is designed in order to illustrate the method and its advantages. It is noteworthy that the Augmented Weighted Tchebycheff proved to be much more reliable than the conventional weighted sum method in discrete problems, due to the existence of unsupported Pareto optimal solutions.

1. Introduction

Optimization affected by parameter imprecision has been a focus of the mathematical programming community during the last twenty years. Solutions to optimization problems can exhibit remarkable sensitivity to perturbations in parameters of the problem, thus often rendering a computed solution infeasible, or significantly suboptimal, or both (Bertsimas, Brown, & Caramanis, 2011). Therefore the concept of robustness in mathematical programming has drawn attention of the scientific community in this field and it is usually under the umbrella of "robust optimization". By using the term "robustness" we actually mean that there is some kind of uncertainty (or imprecision) in the model and we want to be "at the safe side". Uncertainty can be present in various forms (uncertain data linked with future outcomes, imprecise model parameters, etc.).

Robustness can be defined as a degree to which a solution is insensitive to underlying assumptions within a model. Key elements of robust optimization are volatility and flexibility. The former asks for a solution that is relatively stable to data variations and hedges against bad outcomes while the latter is concerned with keeping options open in a sequential decision process having recourses for the effects of earlier decisions (Greenberg & Morisson, 2008).

The concept of "robust optimization" in operational research was introduced by Soyster in 1973 but it received significant attention after the work of Mulvey, Vanderbei, and Zenios (1995) introducing the robust model ("almost feasible") and robust solution ("close to optimal"). Ben-Tal and Nemirovski (1998, 2000) proposed a robust optimization approach to formulate continuous uncertain parameters and Bertsimas and Sim (2003, 2004) proposed robust optimization models preserving their linear structure and make them more tractable. See also Bertsimas et al. (2011) for a recent review.

Despite the vast work that has been done in last two decades on single objective mathematical programming problems the concept of robustness is not so extensively examined in multi-objective programming. Kouvelis and Yu in their seminal textbook devote a section to robustness and efficiency (Kouvelis & Yu, 1997; p. 59). Deb and Gupta (2006) introduce the concept of robustness in multi-objective optimization using meta-heuristics. Liesiö, Mild, and Salo (2007, 2008) focus on robustness in project selection using mathematical programming. Wang and Zionts (2006) provide robustness analysis for the Aspiration-level Interactive Method (AIM). Some recent works also deal with robustness and multi-objective optimization like Zhen and Chang (2012) where

* Corresponding author. Tel.: +30 2107723202.
E-mail address: mavrotas@chemeng.ntua.gr (G. Mavrotas).

http://dx.doi.org/10.1016/j.ejor.2014.06.039
0377-2217/

robustness is quantified and is used as a second objective function in a berth allocation problem which is heuristically solved. Roland, De Smet, and Figueira (2012) provide a stability radius for the efficient solutions in multi-objective combinatorial problems. In a recent paper Roy (2010) discusses the "multi-faceted" issue of robustness in the general context of operational research and not only in optimization. He claims that the "robustness concern" should be present regarding the choice of model parameters as they are imperfectly defined (p. 630). A concept of robustness in multi-objective optimization was also introduced by Figueira, Greco, Mousseau, and Słowiński (2008) especially regarding interactive multi-objective optimization. Finally, Lahdelma, Hokkanen, and Salminen (1998) introduced method SMAA (Stochastic Multiobjective Acceptability Analysis) in order to deal with uncertainties in multi-criteria problems with discrete alternatives. SMAA can be considered as a robustness (or stability) analysis in multi-criteria decision making methods (Tervonen, Figueira, Lahdelma, Almeida-Dias, & Salminen, 2009).

In the present paper we study the concept of robustness in multi-objective programming. In such problems we aim at selection of the most preferred among Pareto optimal solutions. Multi-objective programming combines two aspects: optimization and decision support as we have to deal with a subjective selection among objectively determined solutions. The question that we try to answer in this paper is "how sensitive is our choice on the preference parameters?". This means that the concept of robustness has to do only with the decision maker's preference and not on other model's data. We assume that the decision maker's preference corresponds to a set of weights for objective functions and we want to examine how sensitive is the obtained solution to the imposed weights. In Mulvey et al's. (1995) terminology we deal with a "solution robust" situation. We want to measure the robustness of Pareto optimal solutions having as source of uncertainty the precise definition of weights. For this task we design an integrated method that can be used in multi-objective discrete and continuous problems using a combination of Monte Carlo simulation and optimization (Vose, 1996).

In the proposed method we gradually increase the sampling space for the objective function weights and we observe how many times we obtain the selected most preferred solution. In order to quantify the "persistence" of the selected solution we introduce appropriate measures of robustness exploiting the results of the Monte Carlo simulation and optimization. In this way, we can compare Pareto optimal solutions according to their robustness, introducing one more characteristic for the solution quality. In addition, especially in multi-objective discrete problems, we can go one step further detecting new, hopefully most robust Pareto optimal solutions in the selected solution neighborhood. In this case our approach is not only descriptive but also prescriptive as it does not only measure the robustness of the Pareto optimal solutions but also proposes new Pareto optimal solutions in the neighborhood that may be more robust. The aim of the present paper is to examine the robustness of one Pareto optimal solution (the selected or most preferred) and not the robustness of the entire Pareto set.

It must be noted that under the term "sensitivity analysis" we change one parameter at a time. On the contrary, using Monte Carlo simulation (like e.g. in SMAA) we can simultaneously change the required parameters (the weights in our case) in a systematic way. The proposed approach is also different from the weight stability intervals used in multi-criteria methods (see e.g. Mareschal, 1988 for the PROMETHEE method) where the weights vary one at a time.

In order to examine the applicability of the proposed method, a computational experiment is designed using three cases: (1) a Multi-Objective Integer Linear Programming (MOILP) model, (2) a Multi-Objective Mixed Integer Linear Programming (MOMILP) model and (3) a Multi-Objective Linear Programming (MOLP) model. The particular characteristics of each kind of problems and the differences in their behavior are also discussed.

The remainder of the paper is as follows: In Section 2 we describe the methodological approach, in Section 3 we describe the computational experiment. In Section 4 we discuss the results. Finally, in Section 5 we present the main concluding remarks.

2. Methodological part

2.1. Description of the method

The starting point of our methodology is that there is a Pareto optimal solution (POS) that emerged as "the most preferred" using a multi-objective programming method. We refer to this solution as the reference POS and we denote it using an asterisk (POS*). Usually, the parameters that express the preferences of a decision maker are the weights that are attributed to the objective function. In most multi-objective interactive approaches these weights are provided directly by the decision maker or they can be indirectly determined. It must be mentioned that the objective functions should be first brought in the same scale (normalization) in order the weights to be meaningful. In the present paper we use the following normalization scheme:

$$z_k(x) = \frac{f_k(x) - f_k^{\min}}{f_k^{\max} - f_k^{\min}} \quad \text{for maximization criteria} \tag{1}$$

$$z_k(x) = \frac{f_k^{\max} - f_k(x)}{f_k^{\max} - f_k^{\min}} \quad \text{for minimization criteria} \tag{2}$$

where $f_k^{\min}$ and $f_k^{\max}$ are the minimum and the maximum of the kth objective function $f_k(x)$ as obtained from the payoff table. We must also keep in mind that although the meaning of the weights may be different from method to method, they always represent the preference parameters imposed by the decision maker (Steuer, 1986). Eventually, the reference POS* corresponds to a specific combination of weights for objective functions (either using a weighted sum approach, or weighted Tchebycheff, or goal programming, or aspiration reservation methods), (see e.g. Ehrgott & Gandibleux, 2002; Miettinen, 1999; Steuer, 1986; Wierzbicki & Granat, 1999).

The basic idea behind the proposed method is the following: we "relax" the weights that correspond to POS* by allowing them to take values in their "neighborhood". The "neighborhood" is determined by the decision maker and it is expressed by an interval around the so called reference weights and they are symbolized with w_p^* (p is the index of objective functions). These intervals are defined by the neighborhood parameter α which is actually a percentage of the initial weight. For example, a 10%-neighborhood means that corresponding weights (w_p^*) take values in the interval $[w \times (1 - 0.1), w \times (1 + 0.1)]$. For example, if we have three objective functions and the selected Pareto optimal solution corresponds to the weights 0.1, 0.3 and 0.6 then in Fig. 1 the 10%, 25% and 50% neighborhood is illustrated.

In order to better examine the weight neighborhood, the decision maker divides these weight intervals using a number G of grid points (usually G is set between 5 and 10). The higher the number G of grid points, the greater the accuracy in evaluation of robustness, but computation time also increases. Starting from the reference weights w_p^* grid points help to gradually expand in the weight neighborhood. If there are G grid points the whole process is completed in G steps. When we reach gth grid point the weight interval becomes:

$$\left[w_p^* \times (1 - g/G \times a), \quad w_p^* \times (1 + g/G \times a)\right] \tag{3}$$

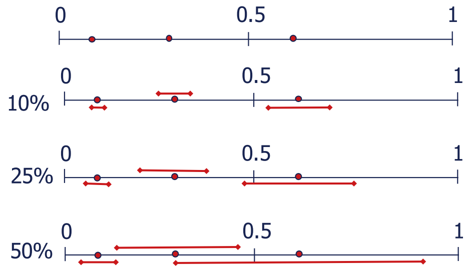


Fig. 1. Illustration of the weight neighborhood for α = 10%, α = 25% and α = 50%.

where w_p^* is the reference weight for objective function p and a is the neighborhood parameter i.e. the percentage of reference weight that defines the border of neighborhood around that weight. Therefore the maximum weight interval for w_p^* is:

$$\left[w_p^* \times (1-a), \quad w_p^* \times (1+a)\right] \tag{4}$$

Subsequently, for each grid point g a Monte Carlo simulation is performed in order to sample weights from the corresponding intervals using a joint uniform distribution (i.e. not one weight at a time, but all the weights simultaneously). The random sampling within weight intervals is performed using the sample and rejection technique. After a vector of uniformly distributed normalized weights has been generated, the weights are tested against their bounds. If any of the constraints is not satisfied, the entire set is rejected and the weight generation is repeated (see e.g. Tervonen & Lahdelma, 2007; p. 507). This simple technique proved sufficient for our case as the rejection rate is rather small (from 2% to 12% which means that for 1000 valid samples we need from 1020 to 1136 Monte Carlo iterations). Probably this is due to the nature of the weight space in our case. Specifically, the initial weights sum to 1 and subsequently we symmetrically extend their intervals around the initial value. Therefore, the resulting weight space can be easily represented by uniformly distributed vectors. However, in more complicated cases of weight space, more sophisticated algorithms for random weight generation need to be considered like e.g. the hit and run algorithm from Tervonen, van Valkenhoef, Basturk, and Postmus (2013).

After weight sampling, we use these weights in the scalarization function of the multi-objective programming model. It must be noted that the scalarization function may be of any kind that uses weights e.g. weighted sum, augmented Tchebycheff (or Chebyshev), goal programming, etc. (we shall return to this later on providing some conclusion regarding their sufficiency in specific kind of models). The results of optimization (i.e. the values of objective functions and the values of decision variables) are stored and we continue with the next Monte Carlo iteration. When we finish the N Monte Carlo iterations we obtain a set of N Pareto optimal solutions designated as S_g. We then count how many times in S_g we have produced the reference POS* which corresponds to the weights w_p^*. This frequency is an indication of the robustness of the most preferred solution due to small perturbations within weights. As we move to the next grid point we actually expand the weight interval of random sampling. Therefore we expect that as we move to wider intervals the frequency of POS* to drop. The degree of resistance to this drop is considered as a measure of robustness for the specific POS*.

In order to measure the degree of robustness according to this drop we introduce the Robustness Index (RI) which is calculated as follows: We calculate the frequency of POS* among the solutions obtained from Monte Carlo simulation for the specific grid point g. We draw the chart of frequency as a function of the width of sampling interval as indicated by the grid points. The ordinate of each point λ_g is the frequency of POS* in the set of solutions S_g and the abscissa is the range around w_p^* of the sampling interval. This is definitely a decreasing function and the so called Robustness Chart depicts the robustness of POS* as sampling intervals around w_p^* are getting wider. An illustrative Robustness Chart is shown in Fig. 2.

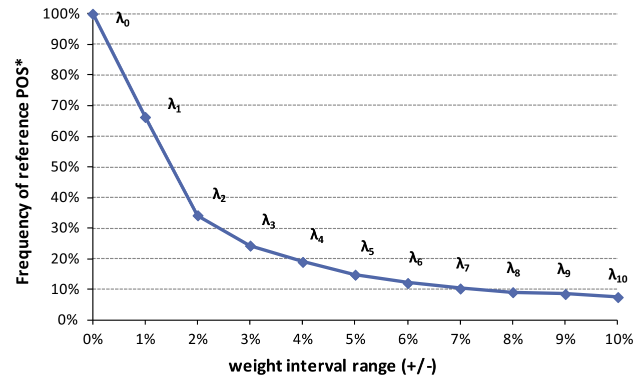


Fig. 2. Example of Robustness Chart.

From the Robustness Chart we can calculate the Robustness Index as a measure of robustness for a specific POS*. The Robustness Index is calculated as the area under Robustness Curve created by λ_g points divided by the full robustness area which is actually the case where for the whole interval the frequency remains at 100% (i.e. a rectangular area 100% $\times$ 10% in Fig. 2). The formula for calculating Robustness Index is the following:

$$\begin{aligned}
RI &= \frac{\left(\frac{\lambda_0+\lambda_1}{2} + \frac{\lambda_1+\lambda_2}{2} + \cdots + \frac{\lambda_{G-1}+\lambda_G}{2}\right) \times \frac{a}{G}}{a} \\
RI &= \frac{\frac{\lambda_0}{2} + \sum_{g=1}^{G-1} \lambda_g + \frac{\lambda_G}{2}}{G} \\
RI &= \frac{\frac{1}{2} + \sum_{g=1}^{G-1} \lambda_g + \frac{\lambda_G}{2}}{G}
\end{aligned} \tag{5}$$

From a mathematical point of view, the Robustness Index can be seen as the integral of the function that expresses the frequency of POS* in relation to the width of the sampling interval x (x is real and $x \leqslant a$) divided by the maximum robustness (= $a \times 100\%$). In other words, the function $fq(x)$ that expresses the frequency of POS* as a function of x, is integrated from 0 to a in order to provide the Robustness Index of POS*, as follows:

$$RI = \frac{\int_0^a fq(x)dx}{\alpha} \tag{5a}$$

Eq. (5) is actually the arithmetic calculation of the integral of Eq. (5a) using the corresponding grid points for discretization.

RI is a measure of robustness that can be used to compare different POS* on how much robust they are on small perturbations of weights. The higher the RI, the more robust the corresponding POS*. RI quantifies the concept of robustness and can provide useful information to the decision maker enriching his perception about candidate solutions before reaching his final choice. It must be noted that RIs are relevant for POS* of the same problem. The integrated flowchart of the robustness analysis algorithm is shown in Fig. 3.

It must be noted that the computational effort is almost insensitive to the involved number of objective functions. This can be attributed to the fact that we work with a scalarization of the objective functions which always leads to a solution of a single objective problem. Regarding the representation of the weight space and how it may explode with more objective functions (and therefore weights), we must keep in mind that we talk about

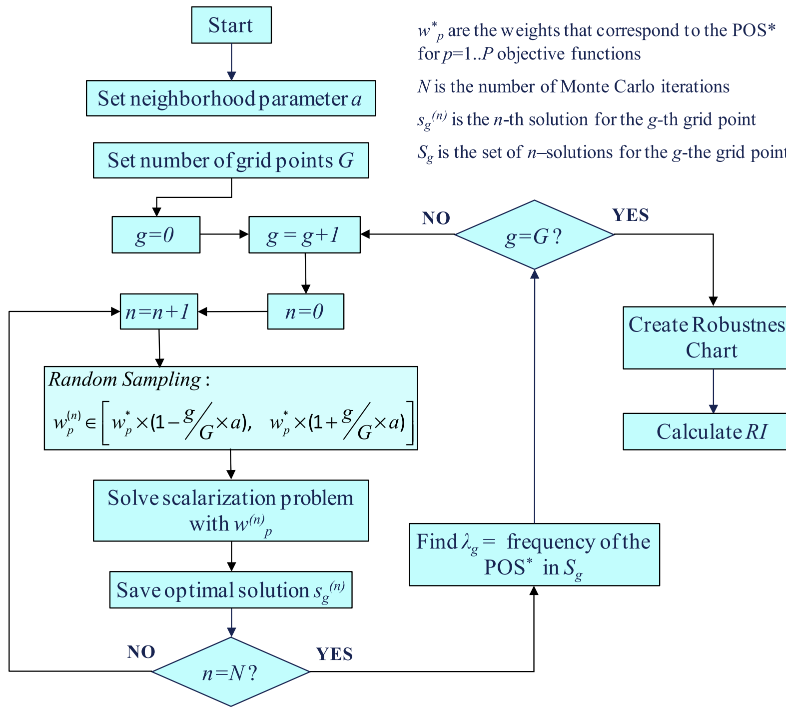


Fig. 3. Flowchart of the algorithm for robustness analysis in multi-objective programming.

a neighborhood around the original weights. This neighborhood is fairly restricted in comparison to the full weight space (the neighborhood is defined as a range of maximum ±20% around the original weights). Therefore, even for 5 or 6 objective functions 1000 Monte Carlo iterations can produce an adequate representation of the weight space. Moreover, if we want to increase the number of Monte Carlo iterations we can always reduce the number of grid points of the formulation in order to keep the computational effort to tractable limits (more iterations per round, but less rounds). Conclusively, the proposed approach escalates smoothly with the number of objective functions.

2.2. Comparison of Pareto optimal solutions

Some clarification is required on how we can conclude that one solution from S_g is the same with POS*. For this comparison we rely on the value of decision variables and not on the value of objective functions as there may be alternative optima. In Multi-Objective Integer Programming (MOILP) problems this is an easy task as the decision variables are integer (and very often binary) and therefore the comparison of two solutions is straightforward. In Multi-Objective Linear Programming (MOLP) and Multi-Objective Mixed Integer Linear Programming (MOMILP) problems we use the following approach: For every solution in S_g we calculate the percentage absolute deviation (relative deviation) of every decision variable's value from the value of the same decision variable in POS* according to the following equation:

$$rd_i = \frac{|x_i - x_i^{POS^*}|}{x_i^{POS^*}} \tag{6a}$$

where x_i is the value of ith variable and $x_i^{POS^*}$ the value of ith variable in POS*. It must be noted that if $x_i^{POS^*} = 0$ and $x_i = 0$ then rd_i is defined to be 0 otherwise if $x_i^{POS^*} = 0$ and $x_i \neq 0$ then rd_i is defined to be 1.

The percentage absolute deviation is calculated also for the objective function values according to:

$$rz_p = \frac{|z_p - z_p^{POS^*}|}{z_p^{POS^*}} \tag{6b}$$

where z_p is the value of pth objective function and $z_p^{POS^*}$ the value of pth objective function in POS*. Again we define that if $z_p^{POS^*} = 0$ and $z_p = 0$ then rz_p is defined to be 0 otherwise if $z_p^{POS^*} = 0$ and $z_p \neq 0$ then rz_p is defined to be 1. In this way we obtain the deviations in the decision variable space as well as in the objective space.

Subsequently we find the greatest relative deviation across decision variables and objective functions. If the greatest deviation is less than a tolerance (designated as *tol*) then the two solutions are considered to be equal. It must be noted that tolerances for the decision space and the objective space may be different. The method is illustrated with the following example: Assume that we have a bi-objective problem with 6 decision variables and 5 solutions to compare with POS* as shown in Table 1.

If we assume a tolerance limit of 10% for decision variables (the maximum deviation among all variables should be less than 10% in order to be considered "the same as POS*") and 3% for objective functions, we see that only solution S2 can be considered as "same as POS*"). However if we raise the tolerance limit to 20% and 6% for decision variables and objective functions respectively, then solution S5 is also added. Usually a tolerance limit of 5% or 10% is appropriate, but it depends on the problem. The greater the number of decision variables the lower the tolerance limit should be.

3. Computational experiments

We test our method for robustness analysis in multi-objective programming using three cases: (1) A MOILP problem dealing with project selection (2) a MOMILP problem dealing with stock portfolio optimization with cardinality constraints and (3) A MOLP problem dealing with stock portfolio optimization. In all three types of problems we tested two models, namely, (1) the simple weighted sum of objective functions and (2) the Augmented Weighted Tchebycheff (AWT) method (Steuer, 1986). The reason for using AWT is its capability to generate unsupported POS in MOILP problems (Steuer, 1986; p. 420). As we will see this is of crucial importance in the robustness analysis of MOILP problems. Although the meaning of the weights varies from method to method, we assume that

Table 1
Example with 6 decision variables.

	X1	X2	X3	X4	X5	X6	Z1	Z2
POS*	0	10	8.3	2.1	3.1	0	78.4	76.1
S1	0	10	8.2	2.6	3.3	0	79.7	79
S2	0	10	8	2.2	3.4	0	78.6	76.6
S3	2.1	10	0	2.3	3	0	57.8	49.8
S4	3.2	10	0	1.8	3.2	0	59.6	51.4
S5	0	10	7.7	1.7	3.2	0	76.5	73.1

Deviation from POS* (%)

							Max dev X			Max dev Z
S1	0.0%	0.0%	1.2%	23.8%	6.4%	0.0%	23.8%	1.7%	3.8%	3.8%
S2	0.0%	0.0%	3.6%	4.8%	9.7%	0.0%	9.7%	0.3%	0.7%	0.7%
S3	100%	0.0%	100%	9.5%	3.2%	0.0%	100%	26.3%	34.6%	34.6%
S4	100%	0.0%	100%	14.3%	3.2%	0.0%	100%	24.0%	32.5%	32.5%
S5	0.0%	0.0%	7.2%	19%	3.2%	0.0%	19%	2.9%	5.3%	5.3%

in both methods these preference parameters capture criteria importance. Objective functions are normalized to a common scale in order to have meaningful weights. As we will see, the results from two methods differ significantly according to the type of problem. For each case three reference POS* are generated (corresponding to a specific combination of weights) in order to calculate and compare their robustness measures.

3.1. Project portfolio selection problem (MOILP)

The first model is a capital budgeting problem concerning 133 renewable energy source (RES) projects. The objective is to find the most preferred project portfolio. There are 5 objective functions namely:

(1) regional development, (2) CO_2 emissions reduction, (3) economic performance measured with the internal rate of return (IRR) of the investment, (4) employment positions, (5) land use. The complete data are available in Makrivelios (2011).

The decision variables of the model are binary, indicating acceptance ($X_i = 1$) or rejection ($X_i = 0$) of the ith project in the final portfolio. The constraints of the model are:

- Available budget is 200 M€ (total cost of the 133 projects is 659 M€).
- Cost of projects in Central Greece should be less than 30% of the total cost.
- Cost of projects in Peloponnese should be less than 15% of the total cost.
- Cost of projects in East & West Macedonia, Northern & Southern Aegean, Epirus should be greater than 10% of the total cost.
- Number of projects from each technology should be between 20% and 60% of selected projects.
- Total capacity of selected projects should be greater than 300 megawatt.

The objective function of the weighted sum model is the following:

$$\max Z = \sum_{k=1}^{5} w_k \times z_k \quad (7)$$

where z_k are the normalized expressions calculated from Eqs. (1) and (2) for the five objective functions with:

$$f_1(x) = \sum_{i=1}^{133} reg_i \times X_i, \quad f_2(x) = \sum_{i=1}^{133} co2_i \times X_i, \quad f_3(x) = \sum_{i=1}^{133} irr_i \times X_i,$$
$$f_4(x) = \sum_{i=1}^{133} emp_i \times X_i, \quad f_5(x) = \sum_{i=1}^{133} lu_i \times X_i \quad (8)$$

with reg_i, $co2_i$, irr_i, emp_i, lu_i being the performance of the ith project in regional development, CO_2 emission reduction, IRR of investment, employment positions and land use respectively.

The AWT model differs from the weighted sum model in the objective function. It has also some additional constraints, decision variables and parameters. The objective function is:

$$\min Z^{AWT} = a + \rho \times \sum_{k=1}^{5} (f_k^{\max^*} - f_k(x)) \quad (9)$$

and the additional constraints are:

$$a \geqslant w_k \times \frac{f_k^{\max*} - f_k(x)}{f_k^{\max*} - f_k^{\min}} \quad \text{for} \quad k = 1, \ldots, 5 \quad (10)$$

where α is an auxiliary variable of minimax type, ρ a small parameter (=0.001) and $f_k^{\max^*}$ is the maximum obtained from the payoff table increased by one (corresponding to z^{**} used in Steuer, 1986; p. 420). The $f_k^{\min}$ is the constant definition point for objective k (Steuer, 1986, p. 426). We use a normalized distance from the ideal in order to make the weights of the two approaches (weighted sum and AWT) having similar meaning.

We consider three combinations of weights in order to examine the robustness of corresponding reference POS* as shown in Table 2.

The neighborhood parameter α is set to 20% which means that the weights will eventually vary to $[w_p^* \times (1 - 0.2), w_p^* \times (1 + 0.2)]$. For example, the weight intervals presented in Table 3 correspond to first weight combination from Table 2. The number of grid points G is set to 10 and the number of Monte Carlo iterations is set to 1000.

3.2. Portfolio selection problem (MOMILP)

In the second case we have a portfolio selection problem with 50 stocks from Eurostoxx 50 market. Objective functions are (1) the portfolio's return (2) the portfolio's risk as quantified by Mean Absolute Deviation (MAD) and (3) the portfolio's Dividend's Yield.

Table 2
Three combinations of weights corresponding to the reference POS*.

	Combination 1 (code: 12214)	Combination 2 (code: 34111)	Combination 3 (code: 41212)
Regional development	0.1	0.3	0.4
CO_2 emission reduction	0.2	0.4	0.1
Economic efficiency (IRR)	0.2	0.1	0.2
Employment	0.1	0.1	0.1
Land use	0.4	0.1	0.2

Table 3
Weight sampling intervals.

$g = 0$	$g = 1$ (±2%)	$g = 2$ (±4%)	...	$g = 10$ (±20%)
0.1	[0.098, 0.102]	[0.096, 0.104]	...	[0.080, 0.120]
0.2	[0.196, 0.204]	[0.192, 0.208]	...	[0.160, 0.240]
0.2	[0.196, 0.204]	[0.192, 0.208]	...	[0.160, 0.240]
0.1	[0.098, 0.102]	[0.096, 0.104]	...	[0.080, 0.120]
0.4	[0.392, 0.408]	[0.384, 0.416]	...	[0.320, 0.480]

Table 4
Three combinations of weights.

	Combination 1 (code: 433)	Combination 2 (code: 325)	Combination 3 (code: 523)
Expected return	0.4	0.3	0.5
Risk (measured as MAD)	0.3	0.2	0.2
Dividend's yield	0.3	0.5	0.3

It is a MOMILP model which it is described in detail in Xidonas and Mavrotas (2012). The resulting model is a MOMILP with 3 objective functions, 100 binary variables, 410 continuous variables and 520 constraints. The formulation of the weighted sum and the AWT model follows the guidelines of Section 3.1.

We test three combinations of weights in order to examine the robustness of corresponding POS* (see Table 4).

The neighborhood parameter α is set to 10% which means that weights will eventually vary within $[w_p^* \times (1-0.1), w_p^* \times (1+0.1)]$. The number of grid points G is set to 10 and the number of Monte Carlo iterations is set to 1000. Given that we have also continuous variables along with 0–1 variables, we follow the approach of Section 2.2 for comparison of solutions in S_g with POS*. In this case we consider a 10% tolerance limit for decision variables and 3% tolerance limit for objective functions (i.e. for one POS in S_g in order to be considered as "the same" with POS* the maximum relative deviation across the values of decision variables should not be greater than 10% and across the objective function values should not be greater than 3%).

3.3. Portfolio selection problem with only continuous variables (MOLP)

This is the same problem as in Section 3.2 but we have removed cardinality constraints, regulatory constraints and entrance thresholds that need binary variables for their formulation. Therefore it is eventually a MOLP problem with 3 objective functions, 260 continuous variables and 317 constraints. The same three combinations of weights as in Section 3.2 are used also in the MOLP case as well as the same neighborhood parameters and tolerance limit for the comparison of solutions.

4. Results and discussion

All the subsequent models were coded in GAMS (Brooke, Kendrick, Meeraus, & Raman, 1998) and solved using CPLEX 12.2. The computation time at each grid point for 1000 Monte Carlo iterations – optimizations was about 3 minutes for MOILP problems, 6 minutes for MOMILP problems and 1.5 minutes for MOLP problems in a core i5 64bit at 2.5 gigahertz. It must be noted that in the following problems we test multiple seeds for the random number generator in Monte Carlo simulation and the results do not differ significantly due to the great number of iterations.

4.1. MOILP model: Project portfolio selection problem

4.1.1. Comparison of weighted sum and Augmented Weighted Tchebycheff method

We use the weighted sum method and the Augmented Weighted Tchebycheff method with weights from Table 2. The terms of weighted sum are normalized in order to make the weights of objective functions more meaningful. Normalization is performed by dividing with the range from payoff table for each objective function. The results of all three weight cases are shown in Robustness Charts of Fig. 4 for both methods.

Regarding the weighted sum method we observe that the Pareto optimal solution that corresponds to first weight combination is the most robust among all three. Robustness Indices as calculated from Eq. (3) are 0.87, 0.73 and 0.82 for the three Pareto optimal solutions respectively.

However, things are much more different when we use Augmented Weighted Tchebycheff (AWT) method with the same sets of weights as shown in Fig. 4b. As we can see the drop is much steeper than in the weighted sum case. The reason is that the weighted sum method finds only supported Pareto optimal solutions in MOILP problems. On the contrary, as it was mentioned, AWT method is capable of finding also the unsupported Pareto optimal solutions. Therefore there are much more Pareto optimal solutions generated in the latter case and this is the reason why the frequency of POS* drops so quickly. For example, in problem 3 (code 41212) the maximum number of generated POS for 1000 Monte Carlo iterations is 5 using the weighted sum method while it increases up to 71 with AWT method. As a consequence, the Robustness Indices with AWT are 0.08, 0.26 and 0.15 for the three reference POS*, which are much lower than in weighted sum case. In addition, ranking of solutions according to their robustness has changed (the second solution – 34111 – is now the more robust while 12214 from first becomes last).

These results confirm previous studies where the number of supported Pareto optimal solutions was found to be only a small portion of total number of Pareto optimal solutions in MOILP

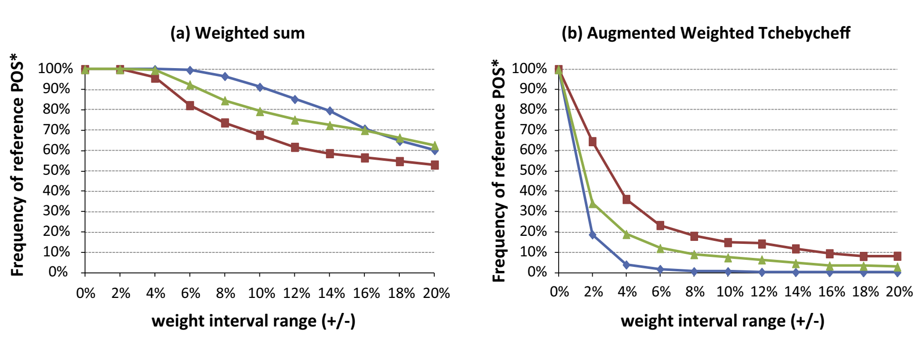


Fig. 4. Robustness Chart for (a) weighted sum method and (b) AWT method in the MOILP problem.

problems (see e.g. Mavrotas & Florios, 2013). Therefore the use of weighted sum method in MOILP problems should be avoided as it can leave many Pareto optimal solutions undiscovered.

4.1.2. Detecting the most robust Pareto optimal solutions in POS* neighborhood

The production of a great number of Pareto optimal solutions with Monte Carlo simulation provide us with an additional capability: We can examine more solutions and not just the reference solution POS*. In other words we can see which of Pareto optimal solutions, besides POS*, are most prominent in the neighborhood. This capability is applicable mainly in MOILP problems where each one solution is undoubtedly identified by the values of its decision variables. It is advisable to use the AWT method that provides both supported and unsupported Pareto optimal solutions. In practice, we implement the method as described in the following steps:

1. The top F in frequency among produced Pareto optimal solutions at the midrange grid point (i.e. the one that corresponds to $\alpha/2$), are isolated. F is usually from 5 to 10. The midrange grid point is chosen because we assume that most prominent POS in the neighborhood should have been appeared by then (when reaching the half range of weight intervals).
2. After processing all grid points, we count the frequency of each one of the F Pareto optimal solutions in every grid point. By monitoring these frequencies during processing of grid points, we can identify those that prevail across the neighborhood.
3. A joint Robustness Chart is created to reveal solutions that are more prominent in the weight neighborhood. An example of joint Robustness Chart is shown in Fig. 5. It is created from problem tch-34111 (second set of weights with AWT method) and we isolated the top 8 solutions form the midrange grid point (in this case, midrange grid point corresponds to ±10% around original weights).

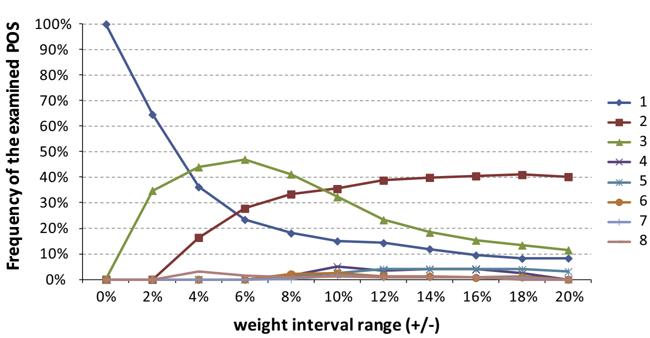


Fig. 5. Example of a joint Robustness Chart with $F = 8$ Pareto optimal solutions.

In chart of Fig. 5 solution 1 is the reference POS*. We can observe that in the neighborhood of ±5% the most prominent Pareto optimal solution is not the POS* but solution 3, while in a more extended neighborhood of ±10% solution 2 is almost as prominent as solution 3. However, when we move to ±20% solution 2 dominates by far with frequency almost 40% (i.e. using the weights from the interval $[w_p^* \times (1-0.2), w_p^* \times (1+0.2)]$ the 400 out of the 1000 produced solutions were equal to solution 2). Hence, the decision maker can select one of solutions 2 or 3 as more robust depending on the level of uncertainty he/she considers around criteria weights. It must be noted that this is a great aid for the decision maker as he discovers, apart from POS*, more candidate solutions in the weight neighborhood enriching his options. From our experience, it is not rare to have a case where the reference POS* is not the most robust solution among Pareto optimal solutions of its neighborhood.

4.2. MOMILP model: Stock portfolio selection problem

Again we use the weighted sum and the AWT method with weights of Table 4. As in Section 4.1 the terms of weighted sum are normalized in order to make the weights of objective functions more meaningful. The normalization is performed by dividing with the range from payoff table for each objective function. The results from the two methods and for the three set of weights are shown in Robustness Charts of Fig. 6.

For the weighted method we observe that the POS* which corresponds to the third weight combination is by far the most robust among all three POS*. Robustness Indices as calculated from Eq. (3) are 0.64, 0.43 and 0.98 for the three Pareto optimal solutions respectively.

In the case of AWT the results are different. As was the case in the MOILP problem (Section 4.1) we can see from Fig. 6b that the drop is steeper than in weighted sum method (although not so steep as in the MOILP case). As we have explained, the reason is the discontinuities of the feasible region in the presence of integer variables and the fact that the weighted sum method finds only supported Pareto optimal solutions in such problems. Therefore, much more different POS are generated with AWT method and the frequency of POS* is accordingly lower. In addition, unlike the weighted method, with AWT the second combination seems to outperform the first one. Robustness Indices as calculated from Eq. (3) are 0.12, 0.17 and 0.33 for the three reference POS* respectively.

4.3. MOLP model: Stock portfolio selection problem with only continuous variables

The results from the two methods and the 3 weight cases are shown in Robustness Chart of Fig. 7.

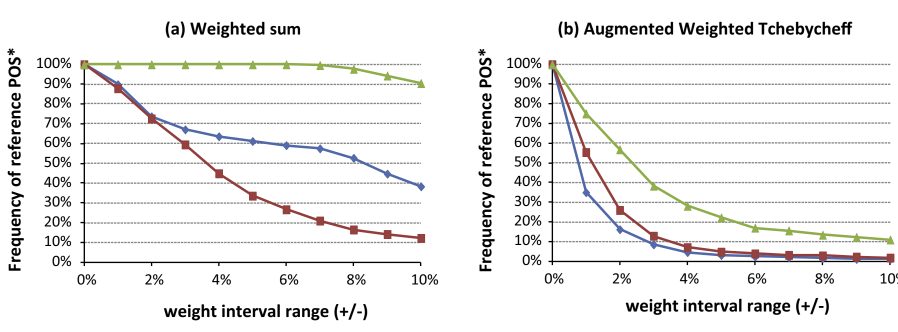


Fig. 6. Robustness Chart for (a) weighted sum method and (b) AWT method in the MOMILP problem.

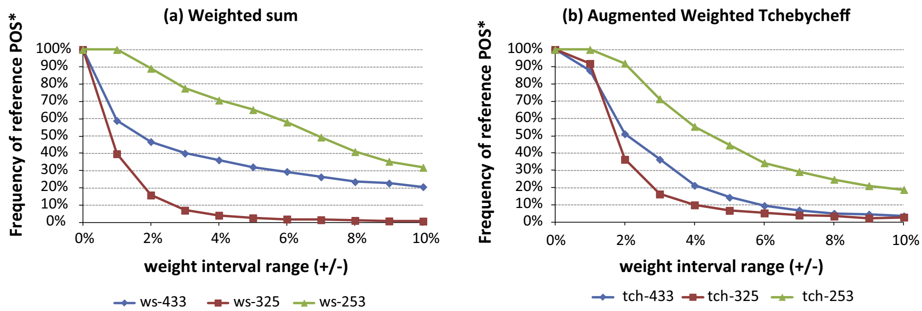


Fig. 7. Robustness Chart for (a) weighted sum method and (b) AWT method in the MOLP problem.

With the weighted sum method (Fig. 7a), we observe that the Pareto optimal solution that corresponds to the third weight combination is the most robust among the three in the specific neighborhood, as it was the case in the MOMILP problem. The Robustness Indices as calculated from Eq. (3) are 0.38, 0.12 and 0.65 for the three Pareto optimal solutions respectively.

The results from application of AWT method are shown in Fig. 7b. We can observe that in MOLP problems both methods (weighted sum and AWT) provide more or less similar results, especially in comparison with two previous cases. This is attributed to the lack of discrete variables that is mainly the reason of divergence between two approaches (i.e. existence of unsupported POS that remain undetected by weighted sum method). Robustness Indices as calculated from Eq. (3) are now closer to the weighted sum method, namely, 0.29, 0.23 and 0.53 for the three reference POS* respectively.

5. Conclusions

Robustness analysis in Multi-Objective Mathematical Programming can provide useful insight to the decision makers. In the present paper robustness analysis deals with weights of objective functions which are usually the most important preference parameters in this decision making context. The question is how small perturbations on weights may affect the final decision. We developed a methodology for measuring this kind of robustness based on Monte Carlo simulation. In contrast to sensitivity analysis where we change one parameter at a time, by using Monte Carlo simulation we can simultaneously alter the required parameters (in our case – the weights) and directly examine their effect on obtained solutions. Two new concepts of that are introduced with this paper, namely, the Robustness Chart and the Robustness Index may convey very useful information to the decision maker regarding the robustness of Pareto optimal solutions. Measuring robustness of candidate solution can lead to more robust decisions.

The comparison of reference solutions (POS*) regarding their robustness can be considered as a post optimality phase. However we can use robustness analysis in order to find new Pareto optimal solutions that may be more robust than the initial reference solution. For this reason, a systematic way of examining the neighborhood of reference solution is also proposed. In this way, we can discover additional Pareto optimal solutions in reference solution's neighborhood that may be more robust.

From the computational experiment it was confirmed that the weighted sum method in multi-objective problems with discrete variables leave several Pareto optimal solutions undetected. In this kind of problems the weighted sum method is insufficient as it leads to a significant underestimation of the size of Pareto set and provides misleading results concerning the robustness analysis of reference solutions. On the other hand, AWT proved to be much more appropriate for this kind of problems.

There are a lot of new features that can be considered for future research. One thing is the integration of all these characteristics of robustness analysis in multi-objective programming in an integrated platform where the decision maker provides just robustness analysis parameters and receives robustness analysis results. A second thing is to test the proposed method with more multi-objective programming methods that use weights. In addition, we can move one step further and adjust the method to Goal Programming where beyond the weights also the value of goal can be subject to small perturbations, following the same methodology (Monte Carlo simulation–optimization, comparison of solutions with POS*, etc.). A challenging task is to produce an approximation of the Pareto front using an adequate number of weight vectors and then apply the proposed method for each one of them, in order to identify robust regions of the Pareto front. Finally we can test the incorporation of more probability distributions in addition to the currently used uniform distribution following the guidelines of Steuer (1986) that uses a 50–50% Uniform – Weibull distribution for similar purposes.

Acknowledgments

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning". The authors would like to thank the two anonymous referees for their constructive comments.

Appendix A. The MOILP model of project portfolio selection

$$\max Z = \sum_{k=1}^{5} w_k \times z_k$$

$$\text{s.t.} \sum_{i=1}^{133} \cos t_i X_i = C$$

$$C \leqslant 200$$

$$\sum_{i \in STE} \cos t_i X_i \leqslant 0.3 \times C$$

$$\sum_{i \in PEL} \cos t_i X_i \leqslant 0.15 \times C$$

$$\sum_{i \in EMD,NAG,WMD,EPR,SAG} \cos t_i X_i \geqslant 0.1 \times C$$

$$0.2 \times \sum_{i=1}^{133} X_i \leqslant \sum_{i \in W} X_i \leqslant 0.6 \times \sum_{i=1}^{133} X_i$$

$$0.2 \times \sum_{i=1}^{133} X_i \leqslant \sum_{i \in SH} X_i \leqslant 0.6 \times \sum_{i=1}^{133} X_i$$

$$0.2 \times \sum_{i=1}^{133} X_i \leqslant \sum_{i \in PV} X_i \leqslant 0.6 \times \sum_{i=1}^{133} X_i$$

$$\sum_{i=1}^{133} mw_i X_i \geqslant 300$$

The variable C is the total cost of the project portfolio and the parameter mw_i is the capacity in MW of ith project.

References

Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research, 23*, 769–805.

Ben-Tal, A., & Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming, 88*, 411–424.

Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review, 53*(3), 464–501.

Bertsimas, D., & Sim, M. (2003). Robust discrete optimization and network flows. *Mathematical Programming (Series B), 98*, 49–71.

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations Research, 52*(l), 35–53.

Brooke, A., Kendrick, D., Meeraus, A., & Raman, R. (1998). *GAMS. A user's guide*. Washington: GAMS Development Corporation.

Deb, K., & Gupta, H. (2006). Introducing robustness in multi-objective optimization. *Evolutionary Computation, 14*(4), 463–494.

Ehrgott, M., & Gandibleux, X. (2002). *Multiple criteria optimization. State of the art annotated surveys*. Dordrecht: Kluwer Academic Publishers.

Figueira, J. R., Greco, S., Mousseau, V., & Słowiński, R. (2008). Interactive multiobjective optimization using a set of additive value functions. In J. Branke et al. (Eds.), *Multiobjective optimization, LNCS* (Vol. 5252, pp. 97–119).

Greenberg, J. H., & Morisson, T. (2008). In A. R. Ravindran (Ed.), *Robust optimization in operations research and management science handbook*. Taylor & Francis Group.

Kouvelis, P., & Yu, G. (1997). *Robust discrete optimization and its applications. Nonconvex-optimization and its applications* (Vol. 14). Dordrecht: Kluwer Academic Press.

Lahdelma, R., Hokkanen, J., & Salminen, P. (1998). SMAA – Stochastic multiobjective acceptability analysis. *European Journal of Operational Research, 106*(1998), 137–143.

Liesiö, J., Mild, P., & Salo, A. (2007). Preference programming for robust portfolio modeling and project selection. *European Journal of Operational Research, 181*, 1488–1505.

Liesiö, J., Mild, P., & Salo, A. (2008). Robust portfolio modeling with incomplete cost information and project interdependencies. *European Journal of Operational Research, 190*, 679–695.

Makrivelios, E. (2011). Multi-criteria evaluation of projects based on renewable energy sources in Greece. Diploma thesis in Master Program "Management of Industrial Systems". Joint program of National Technical University of Athens and University of Piraeus.

Mareschal, B. (1988). Weight stability intervals in multicriteria decision aid. *European Journal of Operational Research, 33*, 54–64.

Mavrotas, G., & Florios, K. (2013). An improved version of the augmented ε-constraint method (AUGMECON2) for finding the exact Pareto set in multi-objective integer programming problems. *Applied Mathematics and Computation, 219*, 9652–9669.

Miettinen, K. M. (1999). *Nonlinear multiobjective optimization*. Boston: Kluwer Academic Publishers.

Mulvey, J. M., Vanderbei, R. J., & Zenios, S. A. (1995). Robust optimization of large-scale systems. *Operations Research, 43*(2), 264–281.

Roland, J., De Smet, Y., & Figueira, J. R. (2012). On the calculation of stability radius for multi-objective combinatorial optimization problems by inverse optimization. *4OR, 10*, 379–389.

Roy, B. (2010). Robustness in operational research and decision aiding: A multi-faceted issue. *European Journal of Operational Research, 200*, 629–638.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research, 21*(5), 1154–1157.

Steuer, R. E. (1986). *Multiple criteria optimization. Theory, computation and application*. Malabar, FL: Krieger.

Tervonen, T., Figueira, J. R., Lahdelma, R., Almeida-Dias, J., & Salminen, P. (2009). A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research, 192*, 236–242.

Tervonen, T., & Lahdelma, R. (2007). Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research, 178*, 500–513.

Tervonen, T., van Valkenhoef, G., Basturk, N., & Postmus, D. (2013). Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research, 224*, 552–559.

Vose, D. (1996). *Quantitative risk analysis: A guide to Monte Carlo simulation modelling*. Wiley.

Wang, J., & Zionts, S. (2006). The aspiration level interactive method (AIM) reconsidered: Robustness of solutions. *European Journal of Operational Research, 175*, 948–958.

Wierzbicki, A. P., & Granat, J. (1999). Multi-objective modeling for engineering applications: DIDASN++ system. *European Journal of Operational Research, 113*(2), 374–389.

Xidonas, P., & Mavrotas, G. (2012). Multiobjective portfolio optimization with non-convex policy constraints: Evidence from the Eurostoxx 50. *The European Journal of Finance*. http://dx.doi.org/10.1080/1351847X.2012.733718.

Zhen, L., & Chang, D.-F. (2012). A bi-objective model for robust berth allocation scheduling. *Computers and Industrial Engineering, 63*, 263–272.

Knowledge-Based Systems 84 (2015) 203–213

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Supporting healthcare management decisions via robust clustering of event logs

Pavlos Delias [a], Michael Doumpos [b], Evangelos Grigoroudis [b,*], Panagiotis Manolitzas [b], Nikolaos Matsatsinis [b]

[a] Eastern Macedonia and Thrace Institute of Technology, Agios Loukas, 65404 Kavala, Greece
[b] Technical University of Crete, University Campus, Kounoupidiana, 73100 Chania, Greece

ARTICLE INFO

Article history:
Received 8 August 2014
Received in revised form 22 February 2015
Accepted 13 April 2015
Available online 18 April 2015

Keywords:
Process mining
Clustering
Robustness
Knowledge discovery
Healthcare management

ABSTRACT

Business processes constitute an essential asset of organizations while the related process models help to better comprehend the process and therefore to enable effective process analysis or redesign. However, there are several working environments where flows are particularly flexible (e.g., healthcare, customer service) and process models are either very hard to get created, or they fail to reflect reality. The aim of this paper is to support decision-making by providing comprehensible process models in the case of such flexible environments. Following a process mining approach, we propose a methodology to cluster customers' flows and produce effective summarizations. We propose a novel method to create a similarity metric that is efficient in downgrading the effect of noise and outliers. We use a spectral technique that emphasizes the robustness of the estimated groups, therefore it provides process analysts with clearer process maps. The proposed method is applied to a real case of a healthcare institution delivering valuable insights and showing compelling performance in terms of process models' complexity and density.

1. Introduction

Business processes are valuable assets of every organization. They control the revenue potential as much as they shape the cost profile of an organization. Processes directly affect the attractiveness of products and services as perceived by the market and they define the ability of organizations to adapt to new circumstances [10]. Therefore, it is no surprise that organizations strive to model, revise, and optimize their internal business processes, as well as the processes shared with other organizations.

In working environments with strong behavioural diversity (i.e., environments where deviations in the process control flows are common), business models are usually ambiguous [13]. In such environments, the problem concerning business process awareness can be defined as follows: are there any dominant patterns of process behaviour? Is it possible to identify groups of cases with similar behaviours? The objective of this paper is to propose a method that delivers compact and comprehensive synopses of flexible behaviours, keeping in mind the end goal to best support their analysis and improvement.

As an example, in this paper we consider a case study involving the clinical pathways of patients in a hospital, where there is a diverse set of paths followed depending on the peculiarities of each patient. The resulting complex behaviour of the business processes in such an environment can be observed through the trace that every patient leaves. On that account, a process mining perspective is followed in this study. The idea of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs, which are readily available in business information systems [37]. Event logs may store additional information about events (like the timestamp and the resource performing the activity). In other words, each case is leaving a trace, which corresponds to the observed behaviour.

When it comes to clinical pathways analysis, process mining techniques face a critical challenge: Patients routes vary significantly and in order to deliver comprehensive models, the event log should be someway summarized [34]. The authors in [17] propose a horizontal summarization, by partitioning the event log into time intervals. In [16], the authors exploit a rich dataset of patients' traces to summarize the clinical pathways based on a behavioural topic analysis. Indeed, as the authors in [19] discuss, the integration of medical knowledge can significantly improve the comprehensibility of the results.

* Corresponding author. Tel.: +30 28210 37346; fax: +30 282106940.
E-mail address: vangelis@ergasya.tuc.gr (E. Grigoroudis).

http://dx.doi.org/10.1016/j.knosys.2015.04.012
0950-7051/

Often, however, such medical knowledge is not available (e.g., relevant data are not recorded, or data are considered too sensitive to be provided, or even medical experts are not available for the process analysis project). In this paper, we focus on such cases; therefore, we follow a trace clustering approach that mostly relies on the control-flow features of the cases.

Trace clustering aims at discovering clusters with related behaviours. However, considering the set of traces in the event log all at once often leads to ambiguities because the event log contains traces of cases that may refer to very different behaviours (i.e., potentially unique or infrequent cases). By identifying clusters of diverse traces, process discovery techniques could be connected to subsets of behaviours and subsequently deliver more clear, coherent and comprehensible process models.

In this framework, this study contributes by proposing a method that relegates the effect of infrequent behaviours (without ignoring them) and eventually provides effective summarizations of the event log. This is achieved through clustering the traces using a more stable similarity metric. The stability of the metric is reached by introducing the concept of neighbourhood. This addition allows promoting any prevalent patterns, while it reduces the impact of isolated cases to the clusters' formation. In this way the proposed methodology provides compact information and meaningful insights to managers as it facilitates the derivation of a simple interpretation of a complex business process, thus allowing process stakeholders to communicate on an evidence-based ground.

The rest of the paper is organized as follows. The next section provides a brief overview of related works. In Section 3, we describe the case study of a public hospital and the proposed methodology. The approach is analytically presented in Section 4, while the obtained results are discussed in Section 5. Finally, a short discussion concludes the paper.

2. Related works

Flow variability in healthcare processes arises due to the highly customized medical guidelines that describe how patients are treated. Furthermore, it is possible that process analysts and stakeholders do not actually need a complete process model, but just an understanding of a dominant behavioural pattern. In cases where a process is expected to be realized over instances with very different behaviour, discovering a single model would seldom provide clear answers, since the generated models would be complex and confusing (i.e., "spaghetti" models as in [35], p. 301). Clustering different behaviours and discovering a process model per cluster has been identified as an effective solution [11].

An initial and influential approach, presented in [33], proposes the creation of feature vectors for each trace followed by the application of common clustering techniques. Features could be bag-of-activities, transitions, resources, case attributes, etc., while clustering techniques include *k*-means, agglomerative hierarchical clustering, and self-organizing maps. That work introduces the concept of "profiles" for traces, which allows for context information to be considered. However, the stability of the results is not discussed. An ordinary clustering technique (e.g., agglomerative hierarchical clustering) is also used in [18]. In this case, traces are evaluated for their similarities by the activities and the transitions vector. This similarity metric is simple, yet quite straightforward to infer control-flow similarities. While hierarchical clustering is effective in showing how different traces differ from each other, this form of clustering has its disadvantages. The primary disadvantage is that hierarchical clustering is only effective at splitting small amounts of data. When the event log is small, patterns and relationships between clusters are easily discernable. As the event log grows, so does the dendrogram, and this usually results in the loss of information. Besides, all determinations are strictly based on local decisions and a single pass of analysis.

The authors in [41] try to resolve spaghetti models through sequence clustering, i.e., identify frequent sequences of activities through a Markov chain representation. The proposed method could support the post-processing of cluster models (e.g., by discarding infrequent elements). However, the applied algorithm could result in multiple cluster solutions. For instance, the applied migrating-drifting means approach makes the final cluster solution dependent, to some extent, on the order in which the traces are considered for relocation.

Another approach is to use syntactic techniques which operate on the whole sequence "as-is" by applying string distance metrics such as the Levenshtein distance and the generic edit distance, in conjunction with standard clustering techniques [2,3]. A distinctive feature of this approach is that instead of assuming the causes that could explain the variation in process instances (e.g., due to different time periods as noted in [23]) – a task that requires intensive domain knowledge – clusters are created based on a simple similarity metric and variability causes are induced a posteriori (that is, we gain knowledge about the domain). An additional contribution is that the whole method is centred on the robustness of the final solution.

Many approaches, from the area of management and information technology, can be adopted by a healthcare organization in order to optimize its efficiency and effectiveness and to be competitive [22]. The authors in [8] provide a brief overview of business intelligence techniques applied to healthcare services. Moreover, data mining approaches can uncover new biomedical and healthcare knowledge for clinical and administrative decision making as well as generate scientific hypotheses from large experimental data [46]. Should the focus of the research is in discovering rules for temporal patterns (and not process models like in this work), several methods based on local patterns mining can be employed. Such rules are extracted as sequence patterns [7,12,29]. Another approach is to exploit temporal probabilistic models to model healthcare problems. In this category, Bayesian networks are the most visible technique [42]. An additional potential is to exploit temporal data of healthcare services to build predictive models. To this end, different learning algorithms have been applied. The focus of these works is on building predictive data mining models with temporal data (see [1]) using supervised or semi-supervised techniques, like positive-unlabelled learning [15]. However, data mining approaches are data-centric and not process centric [36]. Thus, their output is not directly related to the process mining approach proposed in this work.

Concentrating on process mining techniques, a visible work is that in [30], where the authors developed a methodology for the application of process mining techniques that leads to the identification of regular behaviour, process variants and exceptional medical cases. An additional use of process mining is to check for conformance (process stakeholders can match the assumed process model with the real one – derived from discovery in the event logs) and check if medical standards or administrative guidelines are followed.

3. Problem description

3.1. Introduction to case study

The hospital under consideration in this study is situated in the city of Chania, Greece. It was established in 2000 and has a capacity of 465 beds and 36 operating departments. The hospital is a general public health unit, providing first and secondary degree health care to the residents of the prefecture of Chania. On an annual basis, the hospital has more than 100,000 emergency patient visits and 120,000 external patient visits for medical exams, while the total number of inpatient visits is 25,000 and the number of surgeries is about 5500. This study focuses on the emergency department process. A rough (verbal) description of the process which was provided through interviewing the manager of the emergency department follows.

The emergency department has two subunits (ED1 and ED2). The first one (ED1), runs 16 hours per day and the second one (ED2), 24 hours per day. Generally, patients that arrive between 08:00 and 23:00 have to pass through registration (during the night shift there is no secretary available, due to cost cutting). Depending on the triage (extremely important cases are labelled with red triage, not severe cases are labelled with green, and the rest with yellow), patients can skip registration. Patients have to provide the necessary information (e.g., name, age) and pay a fee for medical examination. Afterwards, they have to wait at the waiting room. A nurse asks patients about their problems and characterizes the level of the triage. Patients that arrive by the ambulance are sent directly to ED2. Furthermore, patients with urgent problems (e.g., cardiological incidences or serious accidents) receive the highest priority level (red) and are sent directly to ED2. When a patient enters the room of diagnosis, the nurse checks his/her temperature, blood pressure, and heartbeat. Then the physician provides an initial examination. Depending on the level of triage, a patient waits for the lab results at the waiting room or in bed. When the physician delivers the results of the examination, there are three possible next steps. If the case is serious, the patient is sent to the appropriate department of the hospital. Alternatively, the patient may receive a prescription and is sent back home. The third choice is to decide that the patient will stay at the wards of the emergency department in order to make more lab tests.

The data used in this study were collected manually (by the nurses of the emergency department) during some random days of the first half 2013. Every second patient that visited the emergency department during those days was recorded. In particular, the triage and the type of incident were recorded for every patient, as well as the timestamp for each event of that patient. These events correspond to 21 event classes (the actual steps of the process, like for instance "arrival", "blood test", "X-ray test", etc.). Overall, 1867 events were recorded corresponding to 240 different patients.

3.2. Methodology

The process of the emergency department in our case study is governed by some rules and a general operational plan. However, there is no typical process model, not to mention a process model capable to describe every possible path. The lack of existence of such a model results in three major drawbacks. First, the management does not have a view of the flows inside the process (or it has an idealized view of them). Second, it is very hard to check for the compliance of operations to any guidelines. Finally, the performance-wise optimization of the process is not possible. Therefore, it is essential to come up with a process model to set up any management support activities. For this purpose, we propose the methodology illustrated in Fig. 1.

The first step is to discover the overall process model for the emergency department. Any discovery technique can be used at this step, for example the alpha miner [38] or the heuristic miner [45]. However, since we are particularly interested in assessing the variability of the flows within the process, we shall employ a discovery technique that is able to reveal all the paths instead of just the most frequent ones, for example the fuzzy miner without any abstractions [14] or the ILP miner [40].

The next step is to assess variability, which refers to the comprehensibility of the discovered process model. At this stage we suggest to evaluate variability empirically, through the visualization of the process map (a "spaghetti" or a neat diagram) or through a histogram of the variants' frequencies. If the process analyst evaluates the process flow variability as low, he/she can move on by showing the process map to the management board. If the analyst considers the variability as high, he/she has to define a similarity metric between different flows and to proceed by comparing the variants' similarities. Such a pairwise comparison may expose extraordinary, infrequent paths. If such paths are faint or of minor importance, the analyst can directly proceed by clustering the variants. However, if such paths are a critical part of the variants population, then an extra step is required. This step is about modifying the similarity metric towards a more robust version, which can deal more effectively with the presence of many infrequent (or even unique) items.

Eventually, either with the simple similarity metric, or with a more elaborated and robust one, a clustering algorithm may partition the variants into a finite number of coherent clusters. Finally, a process model can be discovered per cluster. The new categorized process models are expected to contain more effective visualizations and to guide more pointed interpretations.

4. Application

4.1. The variability of the flow

Following the proposed methodology, the first step is to discover the overall model. To this end, we applied the Fuzzy Miner [14] as implemented by Disco (http://fluxicon.com/disco). The discovered model is used as an input for the second step (variability assessment). The discovered process map is illustrated in the spaghetti diagram of Fig. 2, wherein the nodes represent the activities of the process (e.g., "arrival", "blood test", etc.) and the arcs represent the corresponding transitions. Due to its complexity (since the process flow does not follow a single specific pattern), the diagram of Fig. 2 is of little usefulness for the management committee.

Moreover, it is not possible to concentrate on the most frequent variants in the process, since, due to the high variation in the process, the most frequent variants would cover only a small percentage of the patients. As Table 1 shows, the three most frequent variants correspond to less than one third of the total traces. Moreover, to reach 75% of the traces population we need to consider 25 variants. This fact is illustrated in Fig. 3, where we plot the frequency of each variant. We see that for the 240 patients we observed 84 process variants. Only 31 patients followed the most frequent variant, while there are 68 variants that were followed by just one or two patients.

4.2. Defining a similarity metric

First of all, we shall note that the proposed method could work with any similarity metric. In [9] different similarity metrics are compared and evaluated. Some of the most popular metrics are the graph-edit distance, the cosine similarity, and the Euclidean, or Hamming distances that can be used for vector space models. However, our view is that there is no single optimal similarity metric for all domains and all kind of applications. In this work we propose a metric based on the cosine similarity, because its range is normalized. Nevertheless, the proposed methodology could accept virtually any other similarity metric, without needing to change anything.

The proposed metric eventually captures the similarities between two traces in terms of both their activities, as well as their

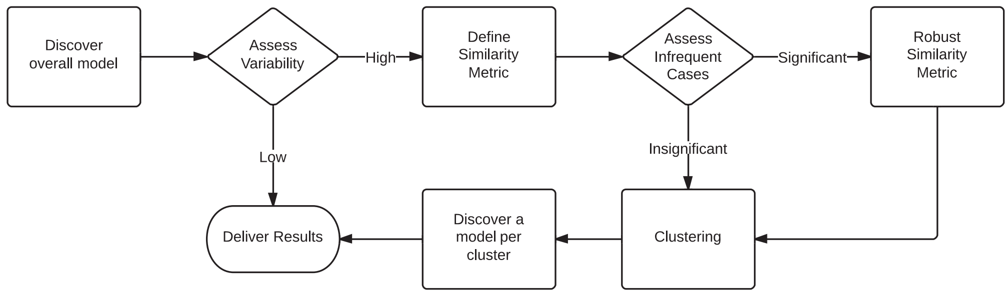


Fig. 1. The proposed methodology to support management decisions.

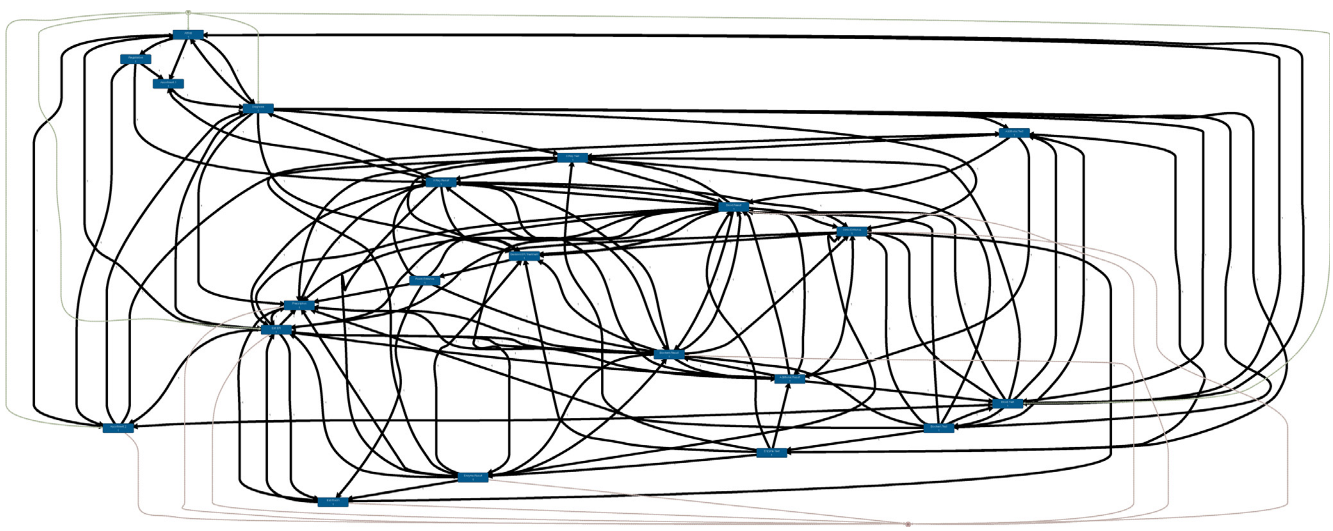


Fig. 2. The discovered process (real flows in the Emergency Department) – A spaghetti model.

Table 1
The three most frequent pathways of the emergency department process.

Pathway	Frequency (%)
Arrival → Assortment.2[a] → Diagnosis → Exit.ER	12.92
Arrival → Registration → Assortment.1 → Diagnosis → Prescription → Exit.ER	9.58
Arrival → Registration → Assortment.1 → Diagnosis → Exit.ER	8.75

[a] Assortment.2 is the activity of "assortment" that takes place in the ED2 Room, in contrast with Assortment.1 that takes place in the ED1 Room. This differentiation follows the different style of assortment that is performed, due to different staffing levels between the day and the night shifts.

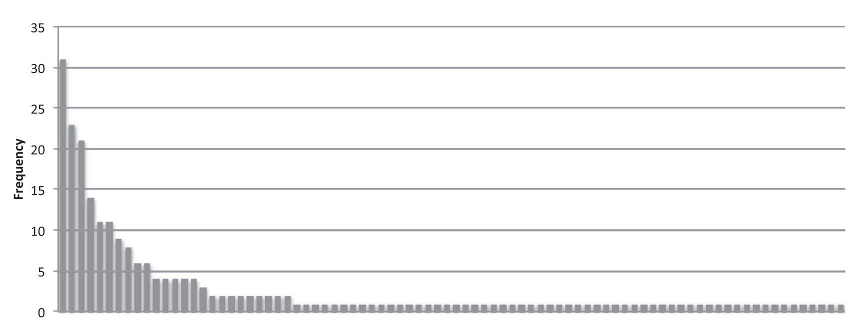


Fig. 3. Frequency of occurrence for every trace variant.

sequencing. This approach considers the dependencies among activities simultaneously with the resemblances of the activities in a trace. A similar approach is also followed in [18]. Nevertheless our approach is different in the components used, since in [18] similarities are calculated between process models while in this work, similarities refer to traces' variants.

To that end, two vectors are created for each trace: The first one, $a_k(i)$ is an ordered binary vector $(0$ and $1)$ whose k^{th} element is set equal to 1 if and only if activity k has been observed in trace i. The second one, $\mathbf{t}(i)$ is the vector format of a square matrix $\mathbf{M}(i)$ whose rows k and columns l are both equal to the number of activities $(k = l = 1, 2, \ldots, K)$. The elements of $\mathbf{M}(i)$ are calculated as follows:

$$M_{k,l}(i) = \frac{1}{d_{k,l}(i)} \text{ for every trace } i \quad (1)$$

where $d_{k,l}(i)$ is the distance of the transition between activities k and l in the i^{th} trace. That is, if k is directly followed by l, then $d_{k,l}(i) = 1$, whereas if l follows k after let's say 5 activities, then $d_{k,l}(i) = 5$. Assuming that $t_n(i) = M_{k,l}(i)$ with $n = (k-1)K + l$, it is easy to observe that $\mathbf{t}(i)$ is an ordered real vector defined in $[0, 1]$.

The cosine similarity for the activities vector will return higher values for pairs of traces that have more common elements (i.e., contain similar activities), while the cosine similarity for the transitions vector will return higher values for pairs of traces that involve similar precedences between two shared activities. The corresponding formulas are:

$$sim_{activities}(T_i, T_j) = \frac{\sum_k a_k(i) \times a_k(j)}{\sqrt{\sum_k a_k(i)^2 \times \sum_k a_k(j)^2}} \quad (2)$$

$$sim_{transitions}(T_i, T_j) = \frac{\sum_n t_n(i) \times t_n(j)}{\sqrt{\sum_n t_n(i)^2 \times \sum_n t_n(j)^2}} \quad (3)$$

Both formulas will return non-negative values in $[0, 1]$, since all elements of vectors $\mathbf{a}$ and $\mathbf{t}$ are non-negative. Finally, in order to attain a single metric, a weighted sum of the two coefficients can be calculated, so that the similarity s_{ij} between two traces i and j can be expressed by the following formula:

$$s(T_i, T_j) = s_{ij} = w_a sim_{activities}(T_i, T_j) + w_t sim_{transitions}(T_i, T_j) \quad (4)$$

where w_a and w_t are the weights for activities' cosine similarity and transitions' cosine similarity, respectively. There are many ways to assign values to these weights, e.g., direct assessment, goal programming or disaggregation methods [32]. In any case, it is important to note that w_a and w_t in (4) have the role of trade-offs between the different types of similarity (see a detailed discussion in [20]). However, since this is an issue beyond the scope of this work, we choose to work with the relative importance of w_a and w_t (i.e., $w_a + w_t = 1$), weighting the activities' similarity factor with 0.3 and the transitions' similarity factor with 0.7. We decided to promote the transitions similarity because the activities set is relatively small (just 21 distinct activities exist in the process), and many of them are common for most patients, albeit with a different sequence.

4.3. Reaching a robust metric

In real-world applications and especially in working environments with bending workflows (like healthcare), one should expect noise and outliers in the data. In particular, in the healthcare domain, outliers will often signify patients with special needs (e.g., to re-take some laboratory tests). Labelling these special cases as outliers is not an option (e.g., in our case such an approach would discard more than 2/3 of the data). Thus, such cases should be retained while controlling for their low frequency of occurrence.

Therefore, in this work we propose an adjustment of the similarity metric (which could be applied in general to any similarity metric) in order to handle infrequent data and noise, and therefore reach more robust results. We prefer not to refer to the infrequent cases as outliers since: (*i*) unique/infrequent routes constitute (overall) a significant percentage of the total variants and (*ii*) unique/infrequent routes are both accepted and expected for patients flow in an emergency department. The rationale of the proposed method is explained below.

The intuition of all clustering methods is to create coherent clusters, i.e., similarities among intra-cluster objects should be high, while similarities among objects of different clusters should be small. However, the existence of isolated traces would bias the distances (similarities) both to intra-cluster objects (they would appear less connected) and to inter-cluster objects (they would appear more connected). Therefore, in order to identify if an object is an isolated trace (i.e., an infrequent case – a kind of outlier as discussed above), one could take into account the neighbourhood of the object. The more crowded it is (many objects exist in its neighbourhood), the more likely would be for that particular object to describe a frequent behaviour. Thus, our efforts focus on finding a way to weight similarities by local densities.

In order to define the neighbourhood of each object, we adopt the ε-neighbourhood concept, according to which two objects are considered to be in the same neighbourhood when their distance is smaller than a specified threshold ε (i.e., when their similarity is greater than a specified threshold). The selection of the threshold value is an essential step of the process. In this work, we tested values from 0.6 to 0.9. Empirically, by examining the density distribution of the resulting similarity matrix, we chose to set $\varepsilon = 0.7$.

More formally, let N_i be the neighbourhood of an object i. Then we introduce a measure l_i to estimate the local density of an object i as follows:

$$l_i = \sum_{j \in N_i} s_{ij} \quad (5)$$

The new similarity metric, which will be better at detecting isolated cases through the amplification of the neighbourhood concept, is calculated as $s'_{ij} = s_{ij} l_i l_j$. The authors in [6] propose to assess the locality of the nodes in a graph through a weight function, i.e., to evaluate the vicinity by filtering out isolated cases (each case should have at least one neighbour). In this work, while we follow the neighbourhood concept, we avoid filtering out any cases by relegating the effect of isolated cases.

4.4. Clustering of traces

Having obtained a similarity matrix, the next step is to cluster the items (trace variants) into groups. Since the similarity matrix is a symmetric matrix whose rows and columns are the items, and cells are the values of the items' pairwise similarity, virtually any partitioning or agglomerative clustering technique can be used. Our selection is spectral clustering. Spectral clustering was selected because of its good results that have been demonstrated in the literature (see [43]), as well as because it provides a good recommendation about the number of clusters. In the next paragraphs, we concisely review spectral clustering and the algorithm proposed in [31].

4.4.1. Spectral clustering algorithm

Let $\mathbf{G} = (\mathbf{V}, \mathbf{S})$ be an undirected weighted graph with the set of vertices $\mathbf{V}$ consisting of the given points $\{x_i | i = 1, 2, \ldots, n\}$ and $\mathbf{S} = [s_{ij}]_{n \times n}$ a symmetric matrix with s_{ij} being the similarity (weight) of the edge connecting vertices i and j. It is very common for the s_{ij} to be calculated as the Gaussian kernel, however, in this work it is calculated by the procedure described in the previous section. We should note that by using the Gaussian kernel, it is not straightforward how multiple similarity dimensions (i.e., the activities and the transitions) can be calculated and weighted by their significance. The graph *Laplacian* $\mathbf{L}$ is defined as $\mathbf{L} = \mathbf{I} - \mathbf{S}$ where $\mathbf{I}$ is the identity matrix. Often the *Laplacian* is normalized, but in this work we follow [31] who propose to use the non-normalized matrix. In fact, in our case, the normalized version could yield unsatisfactory results, because after reducing the impact of infrequent cases, the graph is likely to contain many vertices with low degrees, thus leading to unstable cluster indicators [44].

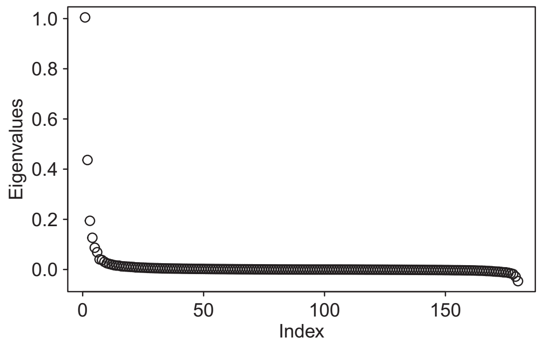


Fig. 4. Plotting the eigenvalues of the generalized eigenproblem $\mathbf{Lu} = \lambda\mathbf{Du}$. The first three are well separated.

The algorithm proceeds by considering the generalized eigenproblem $\mathbf{Lu} = \lambda\mathbf{Du}$, where $\mathbf{u}$ is the eigenvector and $\mathbf{D}$ is the degree matrix, defined as the diagonal matrix with the elements $d_i = \sum_j w_{ij}$ on the diagonal. Then, a matrix $\mathbf{U} \in \Re^{n\times k}$, containing the first k eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ as columns, is created, which introduces a (possibly non-convex) mapping of the original data to a k-dimensional subspace, which is defined by features corresponding to the derived eigenvectors. As noted in [28], if a partitional clustering algorithm (e.g., k-means) is applied directly to the original data, it may lead to unsatisfactory results, particularly when the clusters of interest correspond to non-convex regions. On other hand, applying a traditional clustering algorithm to cluster the data set $\mathbf{U}$ into k clusters, is equivalent to forming an arbitrary (possibly non-convex) cluster structure for the original data. In accordance with the algorithm of [31] and other spectral clustering algorithms (see for example [27,28]), in this study we employ the k-means algorithm to cluster the rows of $\mathbf{U}$ (i.e., to derive the final clustering of the trace variants in the sample) as described in the following paragraph.

Matrix $\mathbf{U}$ is a real-valued matrix, while a matrix with binary entries (1 if item i is assigned to cluster j, and 0 otherwise) is

Table 2
Complexity and coupling quality metrics of the clustering techniques. Bold values represent the best values for each column.

Technique	CFC	Structuredness	Density
Robust-spectral	**250**	642	**0.165**
Agglomerative	389	1267	0.169
k-means	307	831	0.167
EM	1258	**536**	0.206

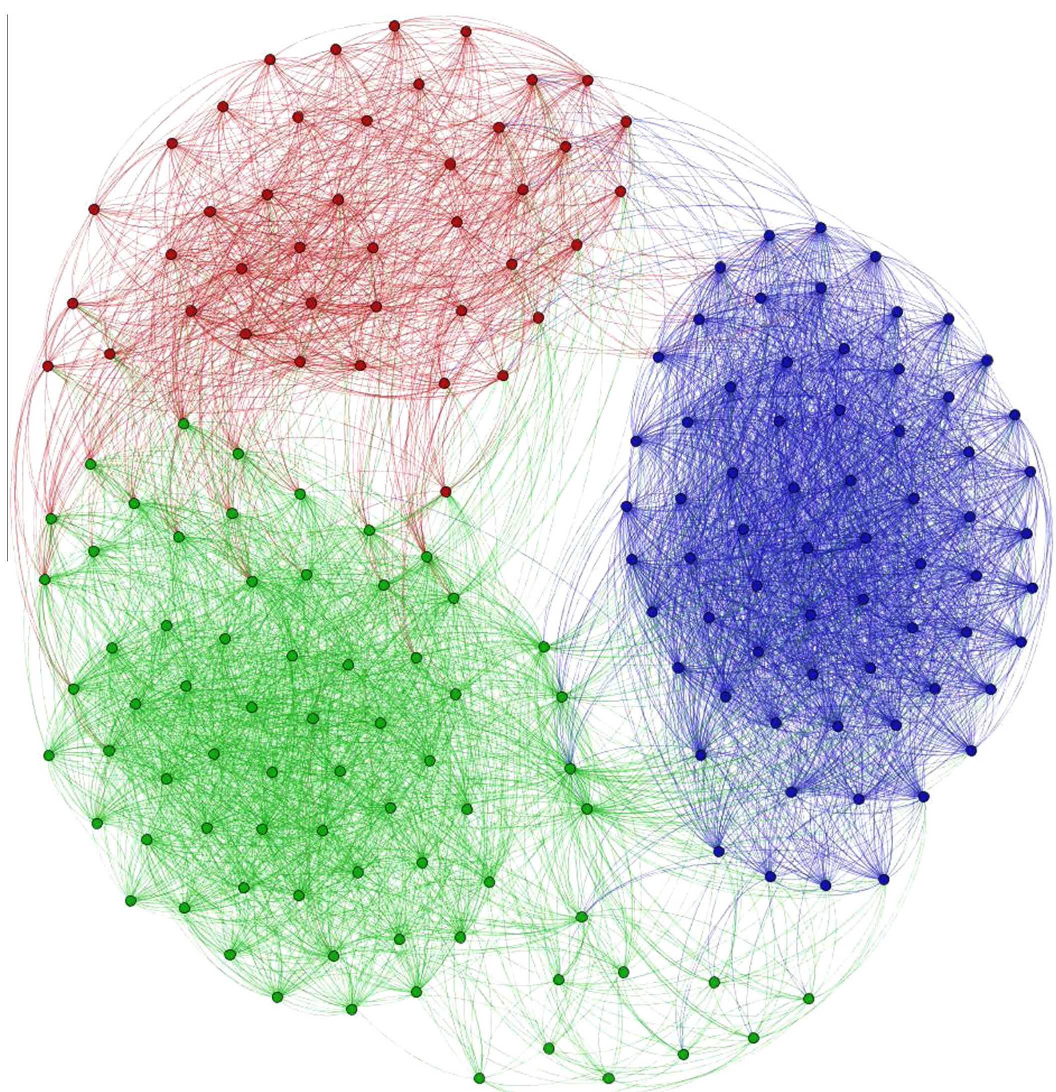


Fig. 5. Visualization of spectral clustering results using the Fruchterman–Reingold layout.

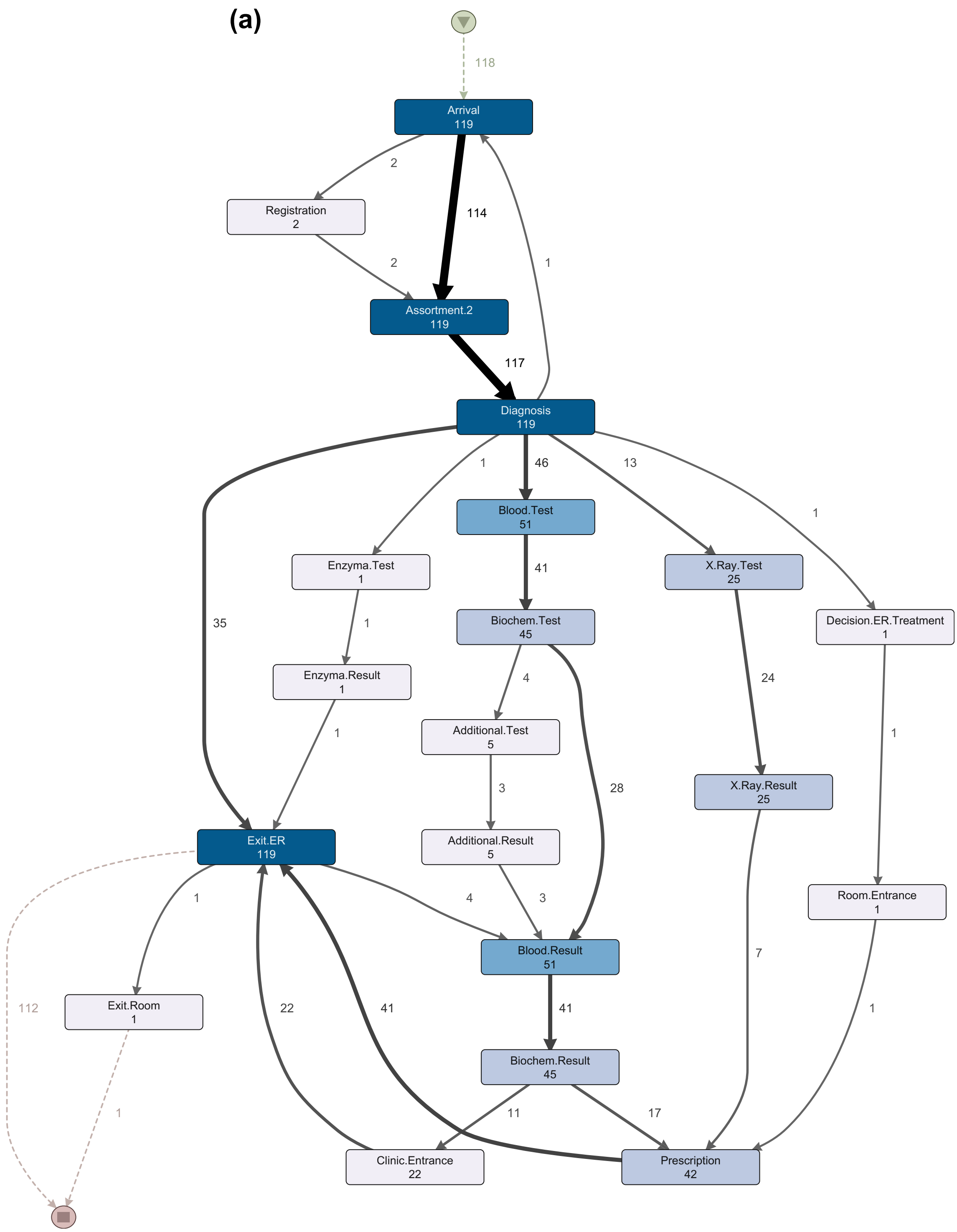


Fig. 6. Discovered process models for each clusters (visualizations are produced by Fluxicon Disco® – 100% for the "activities" and 66% for the "paths" parameters).

needed. Therefore, in order to accept **U** as a solution to our problem, it has to get transformed to a binary format. One simple solution, is to set the maximum value of each row equal to one and let the remaining values to be zeros. However, such an approach yields unsatisfactory results, particularly when there is no dominant maximum value at every row of **U**. An alternative approach, which is adopted in this paper, is to treat the n rows of matrix **U** as k-dimensional feature vectors. The next step is to apply the k-means clustering algorithm considering the rows of **U** as the population to be clustered in k classes. This way, items (traces variants) that have similar values to the eigenvectors (i.e., similar association degrees with the clusters) will eventually get clustered together.

Considering the number of clusters k, a good indication is to look for a sudden drop in the eigenvalues. Of course this number will largely depend on the special needs of each problem.

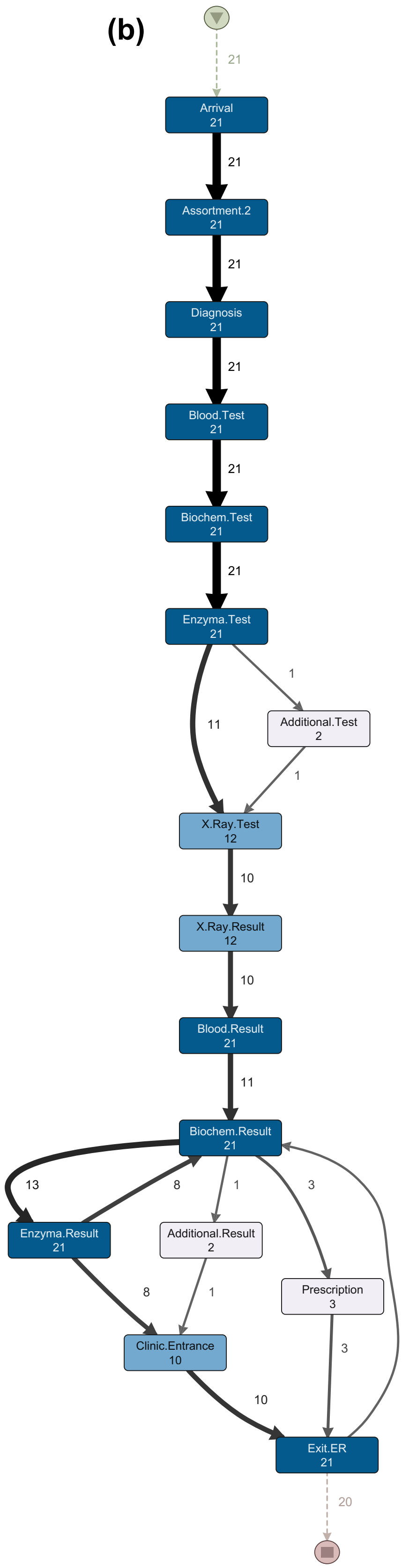


Fig. 6 (*continued*)

5. Results

5.1. Clustering results

An important decision during the clustering phase is to fix the number of clusters. As noted above, in spectral clustering the number of clusters can be specified by examining the eigenvalues obtained from the generalized eigenvalue problem. Fig. 4 illustrates all eigenvalues obtained for our data set. From such a plot, one can select the top k eigenvalues that are well separated from the rest. According to the results of Fig. 4 it is evident that choosing $k = 3$ satisfies the above rule. We also considered setting $k = 4$, but in this case it was found that the fourth cluster consisted of very few traces without a meaningful interpretation.

To get an impression of how spectral clustering assigns objects to clusters, we can consider the similarity matrix as an undirected graph, whose nodes are the clustered objects (trace variants) and edges connecting two nodes are their pairwise similarities (the greater the similarity, the greater the weight of the edge). Following this approach, Fig. 5 illustrates the clustering results. In particular, nodes are coloured according to their clustering membership, while edges are coloured according with their source node colour. The fact that there are few and thin lines connecting nodes from different clusters, while the intra-cluster edges are dense and thick, indicates that spectral clustering was able to provide coherent clusters.

5.2. Comparison with other techniques

In order to analyse the validity of the proposed methodology we compare the obtained results with some other trace clustering approaches. In particular, the proposed method is compared with:

(a) the New Agglomerative Clustering, as implemented in ProM 5.2 (http://promtools.org/prom5) using Euclidean distance metric and minimum variance method),
(b) the k-means approach,
(c) and The EM clustering algorithm, again as implemented in ProM 5.2 (with the default settings).

The three reference trace clustering approaches are described in [33]. We made comparisons with different number of clusters, but since the results are quite similar, we present just the results for the case of the three clusters.

After having obtained the clusters with every method, we discovered a Petri Net model for each cluster using the ILP algorithm [40], which is known to deliver well-fitting process models. Since the goal of this work is to deliver comprehensive process models, we evaluate the results with respect to their complexity and coupling, as it is measured by a set of objective metrics and discussed in [39]. More specifically, we use:

(i) The Control-Flow-Complexity (CFC) metric [4]. The CFC evaluates the complexity introduced in a process by the presence of XOR-split, OR-split, and AND-split constructs and it is highly correlated with the control-flow complexity of processes [5]. The desired situation is to have small complexity.
(ii) The structuredness metric, which recognizes different kinds of structures (basic patterns, such as sequence, choice, and iteration) and scores each structure by giving it some "penalty" value [21].

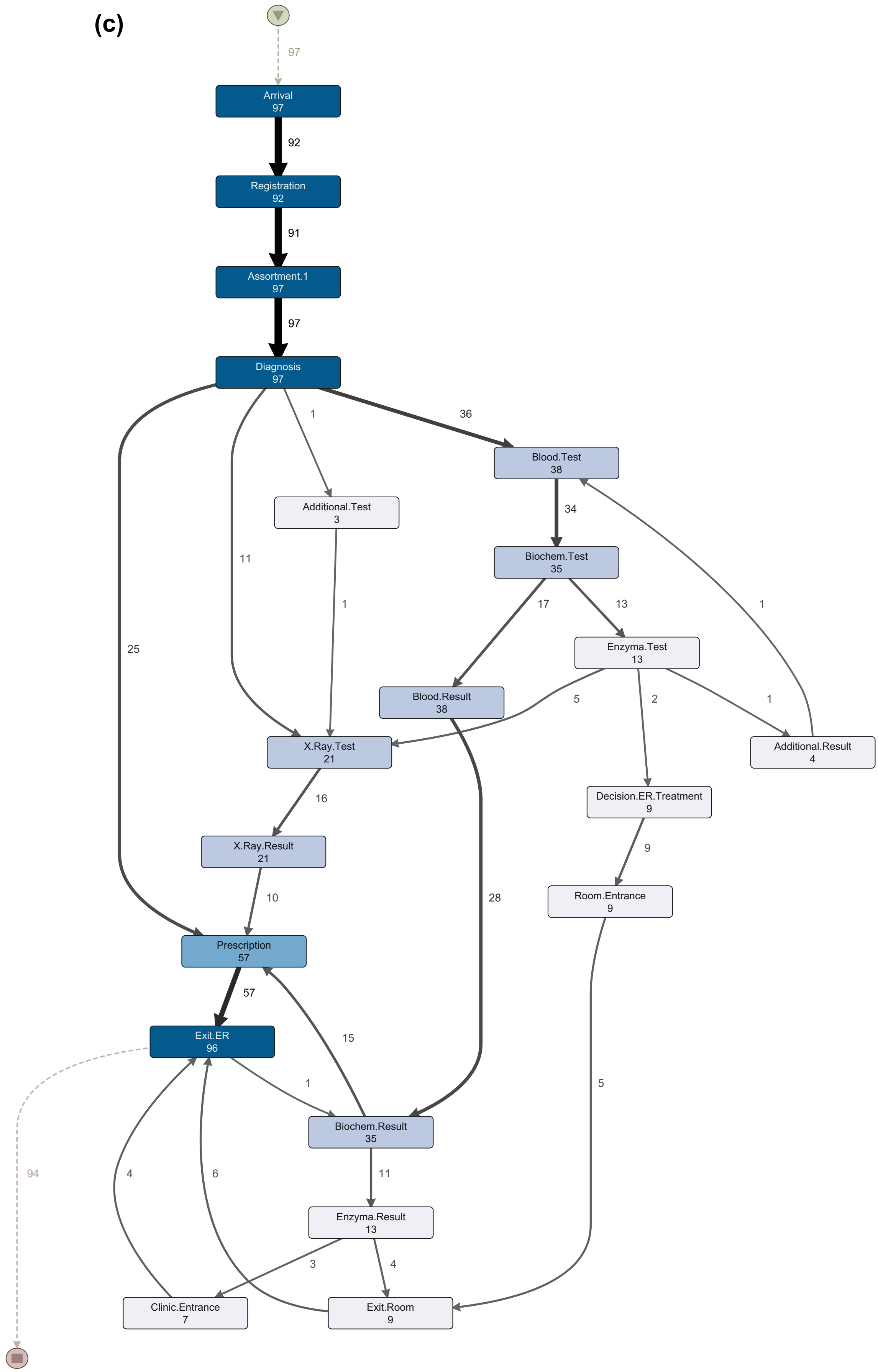


Fig. 6 (*continued*)

(iii) The density of the model as a coupling metric [24]. Density measures the number of interconnections among the activities of the model and has been found to be tightly connected with process models errors [25]. The smaller the density the better the quality of the process model.

The values presented in Table 2 refer to the average value for all clusters. Our approach yields better results with respect to the Control-Flow Complexity and the Density metric, and it is the second best one in the Structuredness metric. The technique that performs best in the Structuredness metric (EM) is however the worst performing one in the other two metrics.

5.3. Case study insights

The method described in the previous section allows the formulation of three, intuitively explained, process models that summarize all 84 variants. Although the set of all possible activities is relative small (21 activities), and the bags of activities are similar for every cluster (e.g., clusters 1 and 3 contain 20 out of the 21 activities; cluster 3 contains 16), there are significant differences among them.

In cluster 1 (Fig. 6a), ‘Assortment.1’ (assortment during day shifts) never happens, while in cluster 3 (Fig. 6c) ‘Assortment.2’ (assortment during night shifts) is not performed. Cases that belong to cluster 2 (Fig. 6b) do not perform ‘Registration’, ‘Assortment.1’, and the triplet ‘Decision.E.R.Treatment’, ‘Room.Entrance’, ‘Exit.Room’ (refer to the ED2 room).

Considering the whole sample, the distribution with respect to the triage is green ∼34%, yellow ∼57%, and red ∼8%. This distribution is repeated only for cluster 1, which concerns approximately 50% of the patients. Cluster 2 (approximately 10% of patients) contains much more red triage cases than the normal percentage, while in cluster 3 (approx. 40% of cases) the yellow triage cases are over-represented.

In cluster 1, ‘Registration’ is often skipped while the unexpected phenomenon of green triage cases visiting the hospital during the night shift appears. This can be attributed to the economic crisis. Before the crisis, the emergency department in public hospitals in Greece was free). After the crisis and during the data collection period, there is a registration cost of 5 euros. However, in the hospital under consideration, there is no registration during the night because there is no secretary available. So, in cases of minor incidents, people prefer to visit the hospital at night in order to avoid paying the registration fee.

Considering cluster 2, skipping ‘Registration’ can be attributed to the higher emergency of cases. Moreover, this cluster has a higher frequency of lab tests (all patients have blood, biochemical or enzyme tests). Finally, this cluster has a high percentage of patients that enter a clinic rather than just leaving the hospital.

Cluster 3 is closest to the expected flow. People get registered, assorted, have some tests, and are forwarded towards the exit via the expected way (after a prescription or a treatment in the ED room).

6. Conclusions and future perspectives

In complex environments, process execution may significantly differ from an ideal process model. Process mining discovery techniques can facilitate the description and understanding of real-world process behaviours. However, there exist certain environments where processes are inherently complex. In these cases, direct process discovery usually deliver complex models, which are of limited practical usefulness.

The proposed methodology can address this problem effectively, by delivering a small number of simpler process maps. Our method contributes to trace clustering approaches by recommending a way to ameliorate the effect of unique (i.e., exceptional) cases. Moreover, by integrating a spectral approach to cluster the traces, we were able to obtain compelling results. The process models obtained through the proposed approach provide significant information of practical usefulness as they facilitate the understanding of an actual complex process and provide operational support on how it can be improved.

To illustrate the usefulness of the methodology, a case study from the healthcare sector was employed, involving the analysis of the diverse processes in the emergency department of a hospital. By applying the proposed methodology, we were able to expose three groups of patients that are homogeneous with respect to their pathways, and therefore we can comprehend the process flows more intuitively. Additionally, by correlating the patients’ clusters with their triage and since there exist estimations of the statistical distribution for the triage, it is possible to predict more accurately the workload per activity, or even to create a better resource allocation plan. Moreover, through the effective visualizations, there are credible chances to communicate the parameters of operations management to doctors, who usually claim that medical guidelines cannot allow for operational optimization.

Two critical points in the proposed method are the similarity metric and the choice of the neighbourhood threshold. For this work, the selected similarity metric focused just on control-flow attributes (bag-of-activities and transitions) while for the neighbourhood threshold, the selection was made through several experiments. Future work could concentrate on a further analysis of these two issues. In particular, regarding the similarity metric, additional criteria (case attributes) could be considered through a multi-attribute approach.

Regarding the selection of the neighbourhood threshold, exploiting concepts of preference modelling could contribute to more effective solutions [26]. In addition, adding domain knowledge to the clustering procedure could also be particularly useful. More specifically, domain experts could provide valuable information on the relationships between items (e.g., traces that must be clustered together or be separated). Incorporating such domain knowledge into the proposed computational process could potentially lead to improved results and provide insights derived from the discrepancies between the experts’ knowledge and the algorithmic outcomes.

Acknowledgements

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) – Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

References

[1] R. Bellazzi, F. Ferrazzi, L. Sacchi, Predictive data mining in clinical medicine: a focus on selected methods and applications, WIREs Data Min. Knowl. Discov. 1 (5) (2011) 416–430.

[2] R.P.J.C. Bose, W.M.P. Van der Aalst, Context aware trace clustering: towards improving process mining results, in: Proceedings of the SIAM International Conference on Data Mining, SDM 2009, Sparks, Nevada, 2009, pp. 401–412.

[3] R.P.J.C. Bose, W.M.P. Van der Aalst, Trace clustering based on conserved patterns: towards achieving better process models, in: S. Rinderle-Ma, S. Sadiq, F. Leymann (Eds.), Business Process Management Workshops, Springer, Berlin, Heidelberg, 2010, pp. 170–181.

[4] J. Cardoso, Evaluating the process control-flow complexity measure, in: 2005 IEEE International Conference on Web Services, IEEE Computer Society Press, Los Alamitos, 2005, pp. 803–804.

[5] J. Cardoso, Process control-flow complexity metric: an empirical validation, in: IEEE International Conference on Services Computing, 2006 (SCC 2006), IEEE Computing Society Press, Los Alamitos, 2006, pp. 167–173.
[6] H. Chang, D.-Y. Yeung, Robust path-based spectral clustering, Pattern Recogn. 41 (1) (2008) 191–203.
[7] J.H. Chang, Mining weighted sequential patterns in a sequence database with a time-interval weight, Knowl.-Based Syst. 24 (1) (2011) 1–9.
[8] P. Delias, P. Manolitzas, E. Grigoroudis, N. Matsatsinis, Applying process mining to the emergency department, in: J. Wang (Ed.), Encyclopedia of Business Analytics and Optimization, IGI Global, Hershey, PA, 2014, pp. 168–178.
[9] R. Dijkman, M. Dumas, B. Van Dongen, R. Käärik, J. Mendling, Similarity of business process models: metrics and evaluation, Inform. Syst. 36 (2) (2011) 498–516.
[10] M. Dumas, M. La Rosa, J. Mendling, H.A. Reijers, Fundamentals of Business Process Management, Springer, Berlin, Heidelberg, 2013.
[11] G. Greco, A. Guzzo, L. Pontieri, D. Saccà, Mining expressive process models by clustering workflow traces, in: H. Dai, R. Srikant, C. Zhang (Eds.), Advances in Knowledge Discovery and Data Mining, Springer, Berlin, Heidelberg, 2004, pp. 52–62.
[12] F. Guil, R. Marín, A tree structure for event-based sequence mining, Knowl.-Based Syst. 35 (2012) 186–200.
[13] C.W. Günther, Process Mining in Flexible Environments, PhD Dissertation, Technische Universiteit Eindhoven, Eindhoven, 2009.
[14] C.W. Günther, W.M.P. Van der Aalst, Fuzzy mining: adaptive process simplification based on multi-perspective metrics, in: G. Alonso, P. Dadam, M. Rosemann (Eds.), Business Process Management, Springer, Berlin, Heidelberg, 2007, pp. 328–343.
[15] R. Henriques, C. Antunes, Learning predictive models from integrated healthcare data: extending pattern-based and generative models to capture temporal and cross-attribute dependencies, in: 47th Hawaii International Conference on System Sciences (HICSS 2014), IEEE, Hawaii, 2014, pp. 2562–2569.
[16] Z. Huang, W. Dong, H. Duan, H. Li, Similarity measure between patient traces for clinical pathway analysis: problem, method, and applications, IEEE J. Biomed. Health Inform. 18 (1) (2014) 4–14.
[17] Z. Huang, X. Lu, H. Duan, W. Fan, Summarizing clinical pathways from event logs, J. Biomed. Inform. 46 (1) (2013) 111–127.
[18] J.-Y. Jung, J. Bae, L. Liu, Hierarchical clustering of business process models, Int. J. Innov. Comput. Inform. Control 12 (5) (2009) 1349–4198.
[19] U. Kaymak, R. Mans, T. Van de Steeg, M. Dierks, On process mining in health care, in: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Seoul, 2012, pp. 1859–1864.
[20] R. Keeney, H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Trade-Offs, Wiley, New York, 1976.
[21] K.B. Lassen, W.M.P. Van der Aalst, Complexity metrics for workflow nets, Inf. Softw. Technol. 51 (3) (2009) 610–626.
[22] R. Lenz, S. Miksch, M. Peleg, M. Reichert, D. Riaño, A. ten Teije (Eds.), Process Support and Knowledge Representation in Health Care, Lecture Notes in Artificial Intelligence, first ed., Springer, New York, 2013.
[23] D. Luengo, M. Sepúlveda, Applying clustering in process mining to find different versions of a business process that changes over time, in: F. Daniel, K. Barkaoui, S. Dustdar (Eds.), Business Process Management Workshops, Springer, Berlin, Heidelberg, 2012, pp. 153–158.
[24] J. Mendling, Testing Density as a Complexity Metric for EPCs, Technical Report No. JM-2006-11-15, Vienna University of Economics and Business Administration, Vienna, 2006.
[25] J. Mendling, M. Moser, G. Neumann, H.M.W. Verbeek, B.F. Van Dongen, W.M.P. Van der Aalst, Faulty EPCs in the SAP reference model, in: S. Dustdar, J. Fiadeiro, A. Sheth (Eds.), Business Process Management, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2006, pp. 451–457.
[26] P. Meyer, A.-L. Olteanu, Formalizing and solving the problem of clustering in MCDA, Eur. J. Oper. Res. 227 (3) (2013) 494–502.
[27] B. Nadler, S. Lafon, R. Coifman, I. Kevrekidis, Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators, in: Y. Weiss, B. Schölkopf, J. Platt (Eds.), Advances in Neural Information Processing Systems, vol. 18, MIT Press, Cambridge, MA, 2006, pp. 955–962.
[28] A.Y. Ng, M.I. Jordan, Y. Weiss, On spectral clustering: analysis and an algorithm, in: T.G. Dietterich, S. Becker, Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems, vol. 14, MIT Press, Cambridge, MA, 2001, pp. 849–856.
[29] D. Patnaik, P. Butler, N. Ramakrishnan, L. Parida, B.J. Keller, D.A. Hanauer, Experiences with mining temporal event sequences from electronic medical records: initial successes and some challenges, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2011, pp. 360–368.
[30] Á. Rebuge, D.R. Ferreira, Business process analysis in healthcare environments: a methodology based on process mining, Inform. Syst. 37 (2) (2012) 99–116.
[31] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (2000) 888–905.
[32] Y. Siskos, E. Grigoroudis, N. Matsatsinis, UTA methods, in: J. Figueira, S. Greco, M. Ehrgott (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer, New York, 2005, pp. 297–344.
[33] M. Song, C.W. Günther, W.M.P. Van der Aalst, Trace clustering in process mining, in: D. Ardagna, M. Mecella, J. Yang (Eds.), Business Process Management Workshops, Springer, Berlin, Heidelberg, 2009, pp. 109–120.
[34] S. Suriadi, R. Mans, M. Wynn, A. Partington, J. Karnon, Measuring patient flow variations: a cross-organisational process mining approach, in: C. Ouyang, J.-Y. Jung (Eds.), Asia Pacific Business Process Management, Lecture Notes in Business Information Processing, Springer, New York, 2014, pp. 43–58.
[35] W.M.P. Van der Aalst, Process Mining: Discovery, Conformance, and Enhancement of Business Processes, Springer, Berlin, Heidelberg, 2011.
[36] W.M.P. Van der Aalst, Process mining: overview and opportunities, ACM Trans. Manage. Inform. Syst. 3 (2) (2012) 1–17.
[37] W.M.P. Van der Aalst, A. Adriansyah, A.K.A. Medeiros, F. Arcieri, T. Baier, et al., Process mining manifesto, in: F. Daniel, K. Barkaoui, S. Dustdar (Eds.), Business Process Management Workshops, Springer, Berlin, Heidelberg, 2012, pp. 169–194.
[38] W.M.P. Van der Aalst, T. Weijters, L. Maruster, Workflow mining: discovering process models from event logs, IEEE Trans. Knowl. Data Eng. 16 (9) (2004) 1128–1142.
[39] I. Van der Feesten, J. Cardoso, J. Mendling, H.A. Reijers, W.M.P. Van der Aalst, Quality metrics for business process models, in: L. Fischer (Ed.), BPM and Workflow Handbook, vol. 144, Future Strategies Inc., Lighthouse Point, FL, 2007, pp. 179–190.
[40] J.M.E.M. Van der Werf, B.F. van Dongen, C.A.J. Hurkens, A. Serebrenik, Process discovery using integer linear programming, in: K.M. Van Hee, R. Valk (Eds.), Applications and Theory of Petri Nets, Springer, Berlin, Heidelberg, 2008, pp. 368–387.
[41] G.M. Veiga, D.R. Ferreira, Understanding spaghetti models with sequence clustering for ProM, in: S. Rinderle-Ma, S. Sadiq, F. Leymann (Eds.), Business Process Management Workshops, Springer, Berlin, Heidelberg, 2010, pp. 92–103.
[42] M. Velikova, J.T. Van Scheltinga, P.J.F. Lucas, M. Spaanderman, Exploiting causal functional relationships in Bayesian network modelling for personalised healthcare, Int. J. Approx. Reason. 55 (1) (2014) 59–73.
[43] D. Verma, M. Meila, A Comparison of Spectral Clustering Algorithms, Technical Report No. 98195-2350, Department of CSE, University of Washington, Seattle WA, 2005.
[44] U. Von Luxburg, A tutorial on spectral clustering, Stat. Comput. 17 (4) (2007) 395–416.
[45] A.J.M.M. Weijters, W.M.P. Van der Aalst, Rediscovering workflow models from event-based data using little thumb, Integr. Comput.-Aided Eng. 10 (2) (2003) 151–162.
[46] I. Yoo, P. Alafaireet, M. Marinov, K. Pena-Hernandez, R. Gopidi, J.-F. Chang, L. Hua, Data mining in healthcare and biomedicine: a survey of the literature, J. Med. Syst. 36 (4) (2012) 2431–2448.

Omega 52 (2015) 142–155

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Robustness analysis methodology for multi-objective combinatorial optimization problems and application to project selection ☆

George Mavrotas [a,*], José Rui Figueira [b], Eleftherios Siskos [c]

[a] Laboratory of Industrial and Energy Economics, School of Chemical Engineering, National Technical University of Athens, Zografou Campus, Athens 15780, Greece
[b] CEG-IST, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
[c] Decision Support Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Zografou Campus, Athens 15780, Greece

ARTICLE INFO

Article history:
Received 31 July 2014
Accepted 10 November 2014
Processed by Doumpos
Available online 20 November 2014

Keywords:
Robustness analysis
Multi-criteria
Combinatorial optimization
Multi-objective programming
Monte Carlo simulation
Visualization

ABSTRACT

Multi-objective combinatorial optimization (MOCO) problems, apart from being notoriously difficult and complex to solve in reasonable computational time, they also exhibit high levels of instability in their results in case of uncertainty, which often deviate far from optimality. In this work we propose an integrated methodology to measure and analyze the robustness of MOCO problems, and more specifically multi-objective integer programming ones, given the imperfect knowledge of their parameters. We propose measures to assess the robustness of each specific Pareto optimal solution (POS), as well as the robustness of the entire Pareto set (PS) as a whole. The approach builds upon a synergy of Monte Carlo simulation and multi-objective optimization, using the augmented ε-constraint method to generate the exact PS for the MOCO problems under examination. The usability of the proposed framework is justified through the identification of the most robust areas of the Pareto front, and the characterization of every POS with a robustness index. This index indicates a degree of certainty that a specific POS sustains its efficiency. The proposed methodology communicates in an illustrative way the robustness information to managers/decision makers and provides them with an additional supplement/tool to guide and support their final decision. Numerical examples focusing on a multi-objective knapsack problem and an application to academic capital budgeting problem for project selection, are provided to verify the efficacy and added value of the methodology.

1. Introduction

Imperfect knowledge of the exact value of parameters, which comprises imprecision, ill-determination, and uncertainty (see [25]), is currently a major issue in mathematical programming. The obtained optimal solutions can exhibit remarkable instability and high vulnerability/volatility to changes in the values of the parameters of the problem, often rendering therefore a computed solution significantly suboptimal or not adequate for further implementation, (Bertsimas et al., [6]; Roy [24]; Ben-Tal et al., [3], etc). Therefore, the concept of robustness in mathematical programming has drawn the attention of the scientific community in this field and is usually set under the umbrella of "robust optimization" [3]. In a more or less informal way, by using the term "robustness" we actually mean that there exists some kind of imperfect knowledge in the model parameters and we examine ways and tools to stay "at the safe side" and safeguard decision making.

The degree to which a solution is stable to the underlying uncertainties within a model is usually defined as robustness. The concept of "robust optimization" in Operational Research was introduced by Soyster [27] but it was not until the last 20 years that it flourished and gathered the attention of the scientific community, mainly with the works by Mulvey et al. [21], Ben-Tal and Nemirovski [1,2] and Bertsimas and Sim [4,5]. Recently, Soyster and Murphy [28] also studied the concept of duality in robust optimization using linear programming. The concept of robustness in integer programming applications, such as product design has been studied by Wang and Curry [31], while Sawik [26] proposed a robustness approach for the supply chain problem. The reader is prompted to see Bertsimas et al. [6] and Gabrel et al. [12] for recent reviews on the subject.

Although robustness has been extensively studied in single objective mathematical programming problems, in the case of

☆ This manuscript was processed by Associate Editor Doumpos.

* Corresponding author.
E-mail addresses: mavrotas@chemeng.ntua.gr (G. Mavrotas), figueira@tecnico.ulisboa.pt (J.R. Figueira), lsiskos@epu.ntua.gr (E. Siskos).

http://dx.doi.org/10.1016/j.omega.2014.11.005
0305-0483/

multi-objective optimization (MOO) several aspects remain to be explored. Kouvelis and Yu [15] in their seminal textbook devote a section to robustness and efficiency. A decade later, Deb and Gupta [7] introduced the concept of robustness in MOO using meta-heuristics. Some recent works also deal with robustness and MOO, for instance Zhen and Chang [32], where robustness is quantified and used as an additional objective function in a berth allocation problem. Roland et al. [22] provide a stability radius for the Pareto optimal solutions in multi-objective combinatorial problems. Roy [24] discusses the "multi-faceted" issue of robustness in the general context of operational research and not only in optimization. He initiates "robustness concern" and proposes a whole set of processes and actions when model parameters are imperfectly defined (p. 630). A concept of robustness in MOO was also introduced by Figueira et al. [10], especially in the case of interactive multi-objective optimization. Mavrotas et al. [20] have examined and analyzed the robustness of the most preferred efficient solution in MOO problems. Two recent works associated with robustness in multi-objective optimization are also worth mentioning: Ehrgott et al. [8] in a recent paper deal with the minmax approach of robustness, and Fliege and Werner [11] focused on robustness in a multi-objective context of portfolio selection.

In this paper we study the concept of robustness in multi-objective programming and especially in the generation (a posteriori) methods. These methods result in the generation of the whole set of efficient solutions (Pareto set) that includes all the Pareto optimal solutions (POS). The question that we attempt to answer in this paper is "How robust is the obtained Pareto set and the individual Pareto optimal solutions in the occurrence of changes or perturbations in model parameters?". We restrict our study to multi-objective combinatorial optimization (MOCO) problems, which, in their majority, concern multi-objective integer programming (MOIP) problems.

In our work we attempt to measure the robustness of POS, when uncertainty occurs by imprecision of the model's parameters. For this task we design an integrated methodology that can be applied in multi-objective discrete and combinatorial problems, using a combination of Monte Carlo simulation and optimization [30]. It must be noted that our approach does not constitute a sensitivity analysis over the results, where the instability of a single parameter is examined at a time. On the contrary, the use of Monte Carlo simulation, in SMAA method for instance [14], enables us to simultaneously alter the values of a number of parameters in a systematic way and extract holistic conclusions with respect to the robustness of the obtained solutions.

It is worth mentioning that Monte Carlo simulation has been also used for robustness analysis in multi-objective programming in the work of Mavrotas et al. [20]. However, in that work they studied the robustness of one specific Pareto optimal solution (most preferred solution) in relation to the preference parameters (weights) and not the robustness of the entire Pareto set in relation to the whole entity of the model's parameters. Since we are not practicing an exact method, but a simulation instead, we can refer to a pseudo-robustness analysis. Hereafter, we shall use the term robustness analysis, but we refer to a pseudo robustness analysis according to the terminology by Roy [23].

We denote as "reference Pareto set" the initial set of efficient solutions, the robustness of which we want to measure. In the proposed methodology we use Monte Carlo simulation in combination with the enhanced version of the ε-constraint generation method (AUGMECON2) that produces the exact Pareto set for MOIP problems [19]. Subsequently, we measure how many times a specific POS of the reference Pareto set is produced across the n Monte Carlo iterations. The higher the frequency, the more robust is the specific POS, since it exhibits a higher tendency to sustain its optimality. Consequently, besides the information regarding the performance of a POS to the criteria (objectives functions), we can provide the decision maker (DM) with an additional piece of post optimality information, namely the robustness measurements associated with perturbations in the model's parameters. A non-robust POS (i.e., it displays small appearance frequency in the Monte Carlo simulation–optimization process), signifies that it can be easily dominated by other solutions, when small perturbations in the model's parameters occur. Illustrative charts for problems with two and three objective functions are constructed, in order to depict the robustness of every POS in the reference Pareto front. Robustness indices for the POS as well as for the whole Pareto set are also calculated. In the end, we test the efficacy of the approach over two numerical examples and a case study regarding a capital budgeting problem for project selection with 108 binary decision variables.

The structure of the paper is as follows: In Section 2 we provide some basic concepts and definitions. In Section 3 we describe the methodology to measure robustness in MOCO problems. Section 4 illustrates two numerical examples in order to test the method, while Section 5 applies the method to an academic research proposal selection problem. Finally, in Section 6 we present the basic conclusions and discuss on some potential future perspectives of the work.

2. Concepts, definitions and notation

This section is devoted to some fundamental concepts on multi-objective combinatorial optimization, dominance and some of its other related concepts, robustness analysis or concerns, and some aspects on simulation.

2.1. MOCO problems

A multi-objective combinatorial optimization (MOCO) problem can be defined as follows:

Definition 1. (multi-objective combinatorial optimization problem). Let $I=\{1, 2,\dots, I,\dots,N\}$ denote a finite set of N objects or items, also called the ground set, and 2^I denote its power set (i.e., the set of all subsets of I), where $|2^I|=2^N$. Consider the subset $S\subseteq 2^I$ as the set of feasible solutions. Define the outcome/objective functions z_k: $I\to R$, such that the outcome vector of each solution $s\in S$ is as follows:

$$z(s)=(z_1(s),z_2(s),\dots,z_k(s),\dots,z_K(s)),\quad \text{where}\quad z_k(s)=\sum_{i\in S} c_{ik}$$

with c_{ik} being the value/outcome associate with each object $i\in S$, for kth objective function ($k=1, 2,\dots,K$). The MOCO problem consists of finding a subset of feasible solutions, $F\in S$, when "maximizing" all the functions z_k, for $k=1, 2,\dots,K$ (the sense of "maximizing" signifies that we search for a particular set of solutions called efficient solutions and defined in Section 2.2).

Any subset s of S is uniquely determined by its characteristic function Xs: $S\to\{0,1\}$ where $Xs(x)=1$ if $x\in S$ and $Xs(x)=1$ if $x\notin S$. With the help of this function, the problem of the above Definition 1 can be stated as a multi-objective optimization (MOO) problem:

$$\text{"maximize"}\{z_1(x),z_2(x),\dots,z_k(x),\dots,z_K(x)\}$$
$$\text{subject to}: x\in X\subseteq\{0,1\}^N$$

where $x=(x_1,\dots,x_i,\dots,x_N)$ is the vector of binary decision variables and X is the feasible region in the decision space. If the decision space is further described by the proper equalities/inequalities, the above MOO problem can be expressed as the following multiple objective integer programming (MOIP) problem with $i=1,\ldots,N$ binary {0–1} variables, $j=1,\ldots,M$ constraints and $k=1,\ldots,K$ objective functions:

$$\max\ z_1(x)=\sum_{i=1}^{N} c_{i1}x_i$$
$$\ldots$$
$$\max\ z_K(x)=\sum_{i=1}^{N} c_{iK}x_i$$
$$st$$
$$\sum_{i=1}^{N} d_{ij}x_i\{\leq,=,\geq\}\ b_j \qquad j=1,\ldots,M$$
$$x_i\in\{0,1\}. \tag{1}$$

where c_{ik} is the objective function coefficient of i-th decision variable in kth objective function, d_{ij} the technological coefficients of the ith decision variable to the j-th constraint, b_j the right hand side of the j-th constraint (RHS), x_i denotes the i-th binary decision variable. The image of X under all the objective functions leads to a feasible region Z (composed of finite set of outcome vectors z) in the objective space R^K.

2.2. Dominance and related concepts

A fundamental concept in MOO is the following one:

Definition 2. (dominance). Let $z', z''\in R^K$ denote two outcome vectors. Then, z' dominates z'', denoted by $z'\Delta z''$, if and only if $z'\geq z''$ and $z'\neq z''$ (i.e., $z_k'\geq z_k''$, for all $k=1,\ldots,K$, and $z_k'>z_k''$ for at least one k).

When applying the above concept to the feasible outcome vectors in Z we may distinguish between dominated and non-dominated outcome vectors. The set of non-dominated vectors in the criteria space is also known in the literature as the Pareto front (PF). The inverse image (in the decision space) of each vector in PF is called efficient solution. The whole set of efficient solutions is designated as Pareto set (PS). The produced PS from the above MOCO model is defined as the reference PS and is denoted as PS*. The corresponding Pareto front (i.e., the image of the Pareto set in the criteria space) is denoted as PF*. The PS* includes the vectors of values of the integer (usually binary) decision variables, while PF* includes the corresponding vectors of values of the objective functions. The Pareto optimal solutions in general and those accruing from the reference problem are denoted as POS and POS* respectively.

2.3. The imperfect knowledge of data

In our MOCO model there are only technical parameters, in the sense that there is no DM who expresses preference information, which should be further modeled using preference parameters. The technical parameters we dispose are thus the following ones:

- The coefficients of decision variables in the objective functions, c_{ik} for $i=1,\ldots,N$ and $k=1,\ldots,K$.
- The coefficients of decision variables in the constraints, d_{ij} for $i=1,\ldots,N$ and $j=1,\ldots,M$.
- The right hand side (*RHS*) of the constraints, b_j for $j=1,\ldots,M$.

The nature of the imperfect knowledge associated with these parameters may ensue from different sources (see [25]):

- The *arbitrariness* linked to the construction of the objective functions and constraints;
- The *uncertainty* which may derive from the application of a forecasting method for estimating the future vales of certain parameters;
- The *imprecision* that can be related to the tools used for measuring certain objects or events, and;
- The *ill-determination* sometimes attached to the definition of certain concepts.

All these sources of imperfect knowledge may lead to inaccurate and inadequate results. A robustness analysis and measurement of the results is therefore of utmost importance.

2.4. Robustness concerns

Robustness concepts can be applied to the method, algorithm, or results (say solutions). Roy [24] initiates the term of robustness concern and explains the main reasons for his choice. The author argues that "it incorporates concerns that must be taken into account *a priori*, at the time that the problem is formulated", what he characterized as frailty points, mainly related to the imperfect knowledge of data. Robustness concern leads to the construction of robust conclusions to be presented to the decision-makers in the form of additional recommendations. The concept of robust conclusions (which can be related to the solutions themselves) can be stated as follows (see Chapter 11, [9]):

Definition 3. (robust conclusions). A conclusion C^r is to be robust with respect to a domain Ω of possible values for the preference and technical parameters, if there is no particular set of parameters $\omega\in\Omega$ which clearly invalidates the conclusion C^r.

A robustness concern consists of all the possible ways that contribute to build synthetic recommendations, based on the robust conclusions. We are particularly interested in providing robust conclusions to the DM, regarding the solutions we generate. In the specific paper we delve into the robustness of the Pareto optimal solutions and specifically, we measure the robustness of the Pareto optimal solutions of the reference Pareto set (PS*), when small perturbations to the parameters of the model occur. If a reference Pareto optimal solution sustains its efficiency in the occurrence of perturbations to the parameters of the model, then it is considered to be robust. In order to study the robustness of the Pareto optimal solutions a simulation study is designed and analyzed.

2.5. Some aspects of a simulation study

Based on Monte Carlo simulation our simulation study contains two aspects:

1. Sampling: Random generation of model's parameters from specific predefined distributions
2. Optimization: Generation of the exact Pareto set of the obtained MOIP problem, after implementation of the augmented ε-constraints method.

Monte Carlo simulation (see, for example, [30]) is a popular algorithm–method that can be used when we want to study stochastic uncertainty. The sole prerequisite is the knowledge of the probability distributions of the uncertain data (a uniform distribution is used when uncertainty is described by just a range of values). In our case we combine Monte Carlo simulation with mathematical programming optimization, as explained in the next section.

The method is repetitive, based on predefined number of iterations T, determined in the beginning of the modeling. In each iteration t ($t=1,\ldots,T$) a cycle of (parameter sampling–optimization–storing of the results) is executed. In our specific case, in the form of a MOCO problem, the optimization process for the determination of the exact Pareto set is highly demanding. At the end of the simulation, we can draw conclusions on the robustness of the obtained Pareto optimal solutions of the reference set. This new index constitutes an additional piece of fruitful information for the decision support experience towards the selection of the most preferred among the Pareto optimal solutions. The next section describes in detail the proposed method.

3. Robustness methodology for MOCO problems

Section 3 sets the bases for robustness analysis of MOCO problems appropriately transformed to multi-objective integer programming problems, describing the methodological frame to be implemented. Then, an appropriate index to measure the robustness of the Pareto optimal solutions and the Pareto set, as a whole, is proposed.

3.1. Robustness analysis

The aim of the proposed method is to assess the robustness of the obtained Pareto set (PS) in a MOCO problem, in relation to small perturbations on the model's parameters. The small perturbations over the model's parameters (c_{ik}, d_{ij}, b_j) are expressed with a perturbation parameter α, which is usually a percentage of the original parameter's value. Subsequently using a Monte Carlo simulation process, we sample from the sampling intervals that are defined as:

$$[c_{ik}(1-\alpha), c_{ik}(1+\alpha)]$$
$$[d_{ij}(1-\alpha), d_{ij}(1+\alpha)]$$
$$[b_j(1-\alpha), b_j(1+\alpha)] \tag{2}$$

where α is the perturbation parameter that varies in [1%, 10%]. Its value depends on the magnitude of the parameters as well as the number of decision variables. In large problems, with multiple decision variables, the perturbation parameter a must be kept low. The reason is that a large a corresponds to an extended neighborhood around the original values of the coefficients. Consequently, the different combinations of alternative values for the sampled coefficients in the Monte Carlo iterations, increase dramatically. This might result to a vast number of different POS resulting from the Monte Carlo process and it would be difficult to draw conclusions about the robustness of the reference Pareto set (PS*). It must be noted that perturbation parameter α may be different across model's parameters (i.e., not a uniform a but a_{ik}, a_{ij}, a_j in Eq. (2)), in order to increase the degrees of freedom in uncertainty modeling.

It must be noted that the method can be applied for any type of distribution. The relevant distribution is appropriately modeled and the sampling with Monte Carlo process can be implemented as well. In addition, we can use different distributions for different parameters of the problem (objective function coefficients, right hand sides, technological coefficients), if we have this type of information. In this study we use as an example the uniform distribution as the more general case, when we have no information about the shape of the distribution (i.e., maximum uncertainty). However, if we want to analyze robustness in a less uncertain environment, the normal distribution may be used with parameters e.g. for the objective function coefficients $(\mu,\sigma)=(c_{ik}, \alpha\times c_{ik})$, where α is the perturbation parameter.

The combination of Monte Carlo simulation and optimization, using mathematical programming techniques, has become popular during the last years, due to recent algorithmic developments and the significant improvement in computer power ([18,20]). Although the computational complexity of these problems is usually high, the analysis of their robustness is still worthwhile, as it provides fruitful information regarding the stability of the final solutions.

The combination of Monte Carlo simulation–PS generation is performed as follows: in the t-th iteration of the Monte Carlo algorithm, we perform a sampling for all the parameters of the model from a uniform distribution, which is defined in a respective sampling interval. We then solve the multi-objective programming problem that arises, using the AUGMECON method [17]. Specifically, we use an improved version, AUGMECON2, which is more suitable for MOIP problems [19].

AUGMECON2, as all the variations of the ε-constraint method, varies parametrically the right hand side of the objective functions' constraints. Initially, it creates a grid on the range of the objective functions, each corresponding to a vector of the right hand side of the objective function constraints. In the case of MOIP problems, if the grid is dense enough, AUGMECON2 can produce the exact set of Pareto optimal solutions by effective scanning all grid points. The method can bypass redundant optimizations exploiting information from previous runs.

The key point here is that, by using AUGMECON2, we can produce the exact PS, i.e., all the POS. In this way, we are sure that no POS is left undiscovered throughout the simulation process. In each Monte Carlo iteration the following MOIP problem with $k=1$, …,K objective functions, $i=1$,…,N decision variables and $j=1$,…,M constraints is solved exactly:

$$\begin{aligned}
&\max\ z_1(x)^{(t)} = \sum_{i=1}^{N} c_{i1}^{(t)} x_i \\
&\quad \ldots \\
&\quad \max\ z_K(x)^{(t)} = \sum_{i=1}^{N} c_{iK}^{(t)} x_i \\
&\quad st \\
&\quad \sum_{i=1}^{N} d_{ij}^{(t)} x_i \{\le, =, \ge\}\ b_j^{(t)} \qquad j=1,\ldots,M \\
&\quad x_i \in \{0,1\}
\end{aligned} \tag{3}$$

where $c_{ik}^{(t)}$ is the objective function coefficient of i-th decision variable in k-th objective function during t-th Monte Carlo iteration and accordingly for the other parameters. In order to produce the exact PS, the coefficients of the objective functions are considered to be integer. If not, they are transformed to integer by multiplying them with an appropriate power of ten. X_i is a binary decision variable and therefore a POS is represented by a vector of "0" and "1" of size N. We then record the PS, of t-th iteration and we move to the next iteration ($t+1$), until we complete the required number of Monte Carlo iterations. It must be noted that the size of the produced PS, may vary across the T Monte Carlo iterations. Fig. 1 depicts an illustrative graph of the method.

After the solution of T MOIP, problems, we calculate the frequency of each one of the original POS of the reference Pareto set (PS*), across the T Pareto sets that have been generated. In order to be able to compare the binary solution vectors, we uniquely code the solution vectors by transforming them to large integers (see, e.g. [29]). Specifically, for the pth POS we use the following coding:

$$code_p = \sum_{i=1}^{N} 2^{(i-1)} \times \overline{X}_i \tag{4}$$

where $\overline{X}_i$ is the value of the X_i variable for the specific POS. In Table 1 we have 5 POS for a problem with 10 binary variables. The rightmost column assigns a unique code to each one of them.

This coding is effective for up to 50 variables. If we have more variables, we apply the aforementioned coding in parts of the solution with maximum 50 variables and we obtain the respective

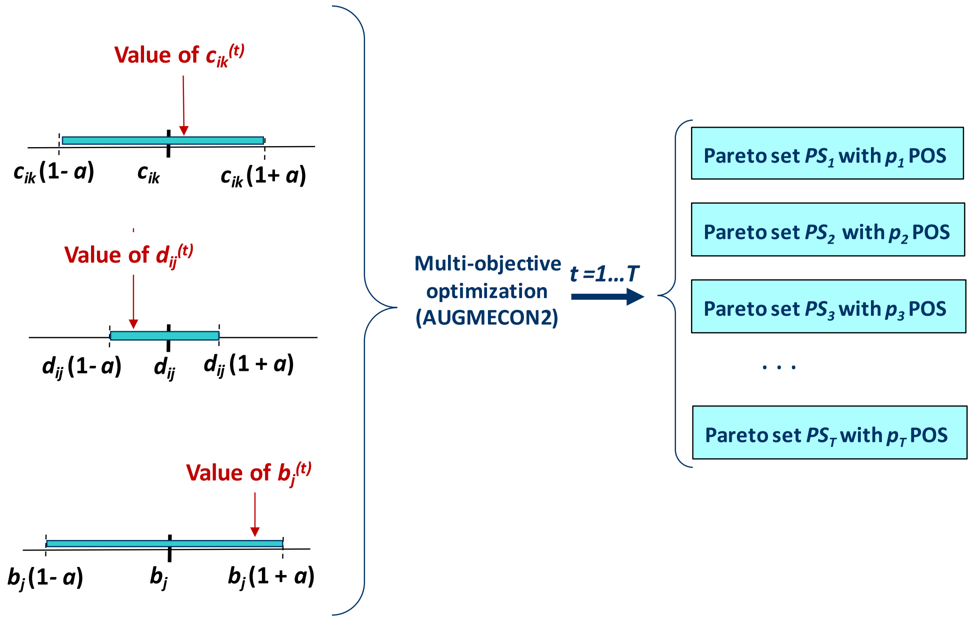


Fig. 1. Illustration of the Monte Carlo simulation–Pareto set generation approach.

Table 1
Example of POS coding.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Code
POS 1	0	0	1	0	1	1	1	0	0	1	628
POS 2	0	1	0	1	1	1	1	0	0	0	122
POS 3	0	1	1	1	0	0	1	0	0	1	590
POS 4	0	0	1	0	1	1	1	0	0	0	116
POS 5	1	1	1	0	1	0	1	0	1	1	855

number of long integer codes. For example, if we have 120 binary variables, we code the first decision variables 1–50 as part 1, then 51–100 as part 2 and finally 101–120 as part 3. Eventually, for the specific case, the code for each relevant Pareto optimal solution is composed of 3-digit integers.

3.2. Robustness assessment

The next step is to quantify the robustness of each reference POS. This task is accomplished by counting how many times across the T iterations we obtain a specific POS as part of the Pareto set. Assume that a reference POS, denoted as POS_p^*, is found in S_p out of T Pareto sets. Then, its "frequency" of appearance in the Pareto set is defined as the Robustness Index (RI_p) of POS_p^*, and it is expressed as:

$$RI_p = \frac{S_p}{T} \tag{5}$$

We can also extract a robustness index representative of the whole Pareto set, the Total Robustness Index (*TRI*), which is simply the average of all robustness indices (RI_p) of the Pareto optimal solutions.

$$TRI = \frac{\sum_{p=1}^{|PS^*|} RI_p}{|PS^*|} \tag{6}$$

where $|PS^*|$ is the size of the reference Pareto set. The flowchart of the proposed algorithm is shown in Fig. 2. The whole process can be implemented in GAMS Development Corporation [13].

3.3. Robustness information and decision support

Robustness information for the Pareto front is essential for the decision maker. As is well known, multi-objective optimization combines two major aspects: optimization and decision support. The decision maker explores the candidate solutions (i.e., the POS) and selects his/her most preferred. The robustness of the POS provides an additional piece of information to the DM, apart from their performance over the criteria. When robustness is neglected, the decision maker may result to decisions that are sensitive to small changes in the initial assumptions (parameters). In other words, the robustness of the results, expressed explicitly through the *RI* we propose, may be perceived as an additional criterion, which is considered and assessed after the end of the optimization procedure.

An illustrative way to present the robustness information is to express explicitly this information on the Pareto front. This can be achieved with the aid of a bubble chart, in which the coordinates of the POS are their objective functions' values and the size of the bubble is proportional to the corresponding robustness index. In this way the DM can easily perceive which areas of the Pareto front are robust and which are not. The following chart in Fig. 3 presents the idea for the visualization of robustness.

As it is obvious, visualization of the robustness information is only feasible for problems with up to three objective functions (i.e., three dimensions in the objective space, see an example in Section 4.2).

In addition, the DM can impose thresholds on the robustness of the Pareto optimal solutions. For example, he/she may prefer to select only the POS with RI $>=$50%. In this way, a more sparse Pareto front is created. Fig. 4 illustrates the instances when RI$>=$25%, 50% and 75%.

Alternatively, the DM, instead of setting thresholds to the Robustness Index, may construct Pareto fronts, based on the percentiles of the most robust POS (e.g. top 50% POS according to their robustness index).

3.4. Applicability of the approach

The proposed method is recommended for medium size MOCO problems, where the exact Pareto set can be effectively calculated. Nowadays, new effective optimization algorithms, accompanied by the increase in computational power, enable us

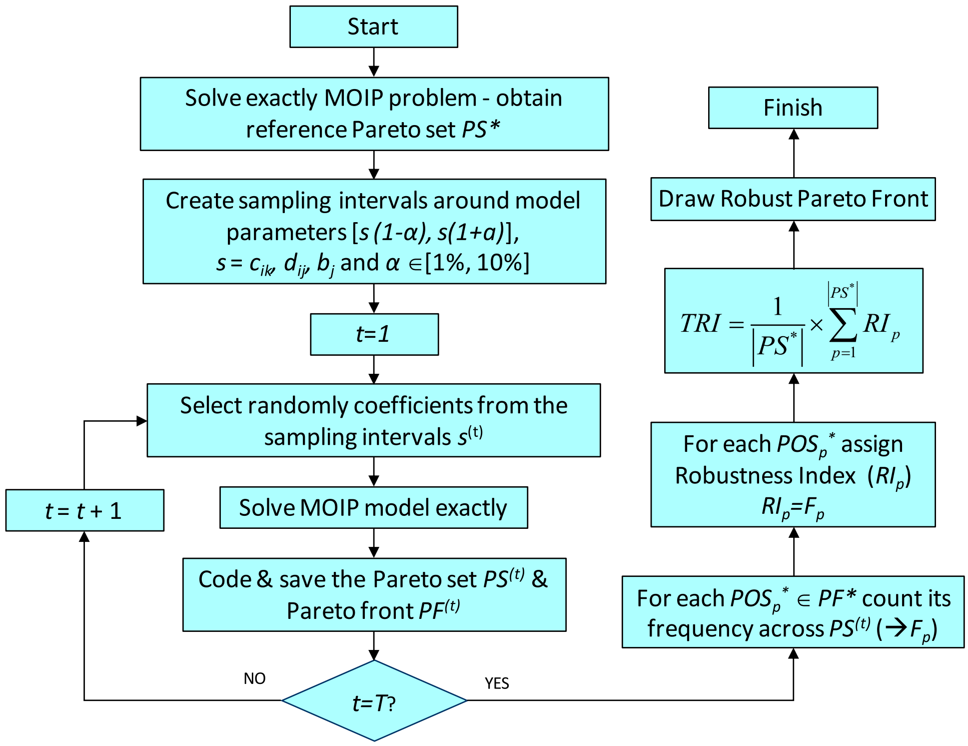


Fig. 2. The algorithm for estimating robustness in MOIP problems.

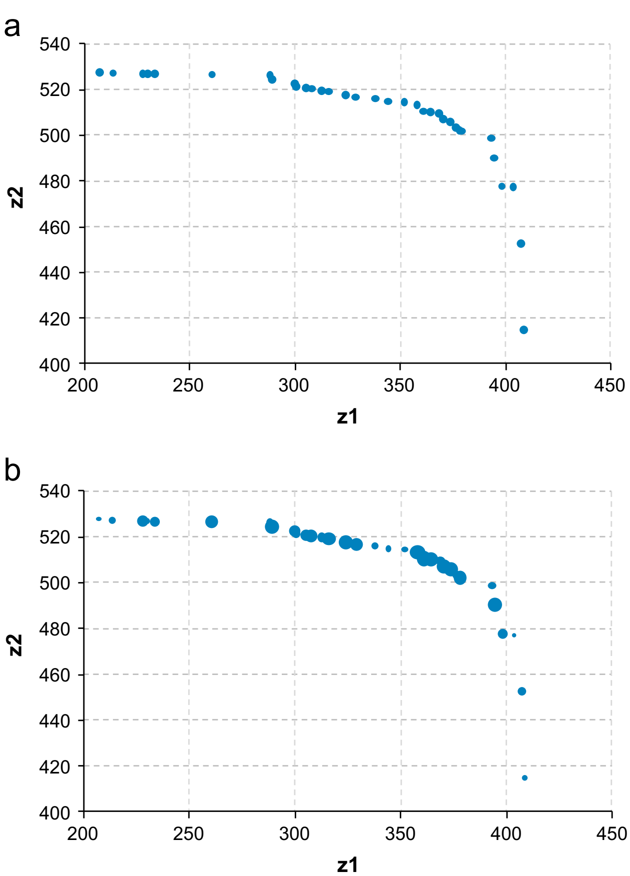


Fig. 3. Visualization of robustness information on the Pareto front. (a) Pareto front, (b) Pareto front with robustness information.

to solve exactly – by identifying all POS – more complicated MOCO problems. For the visualization of the robustness of the Pareto front we are confined to two or three objective functions (which is actually the vast majority of multi-objective optimization problems). The number of binary decision variables is limited to a few hundreds, in order to obtain the results from the Monte Carlo simulation–optimization process in reasonable computational time. For example if a MOIP problem, which corresponds to the reference problem, is solved exactly in 2 min, then the 1000 iterations will demand approximately 33 h.

It must be noted that as the number of objective functions and decision variables increases, the conclusions about the robustness of the Pareto front become vaguer. The explanation is that a huge number of POS is produced, during the Monte Carlo simulation process, and the robustness index of the reference POS converges to zero. For this reason it is unwise to use the proposed method in big problems because implications on the robustness of the Pareto front are ambiguous.

4. Numerical examples

In this section, we present two numerical examples. The first one is described by two objective functions, uncertainty over all model's parameters and a constant perturbation parameter. The second example bears three objective functions, with uncertainty only on the objective function coefficients. The second example is modeled using a non-constant perturbation parameter α.

4.1. Example with two objective functions

We first illustrate the approach with a knapsack problem with two objective functions, 50 items and a single family of constraints.

$$\max z_1 = \sum_{i=1}^{50} c_{i1} X_i$$

$$\max z_2 = \sum_{i=1}^{50} c_{i2} X_i$$

st

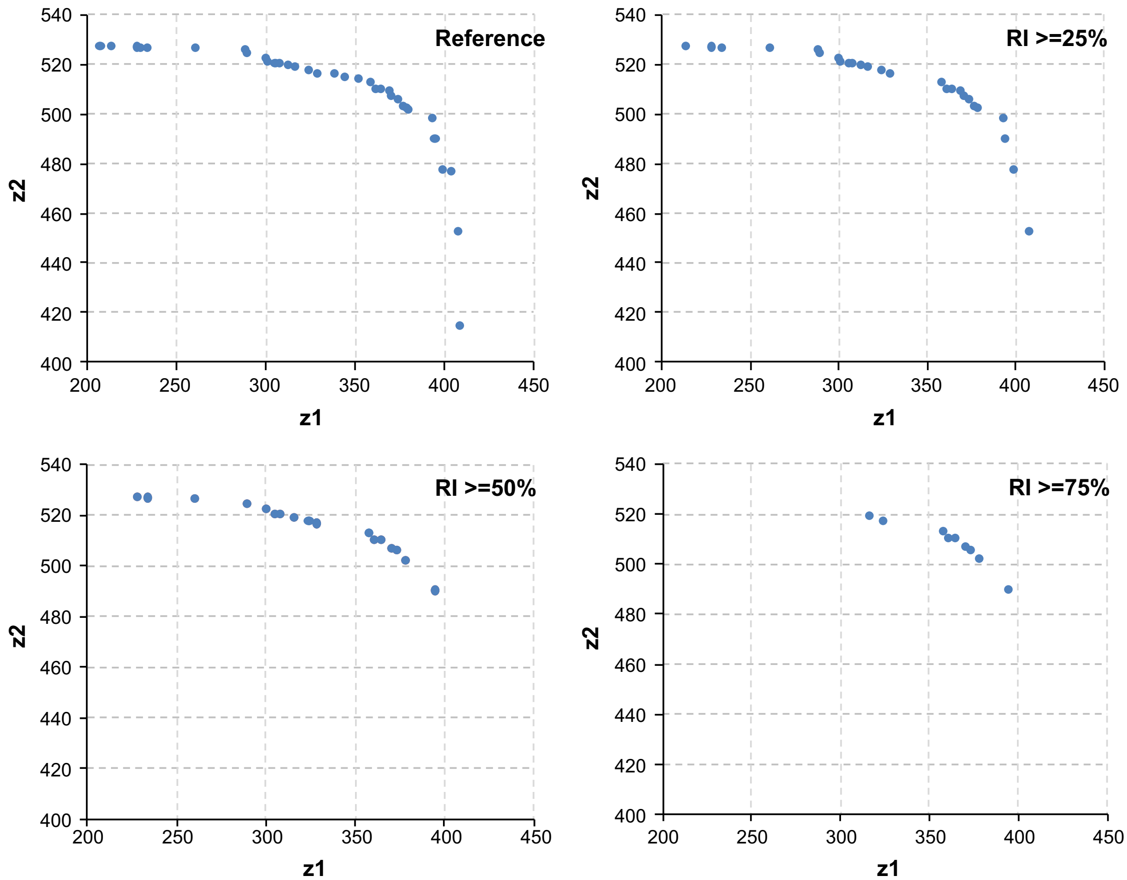


Fig. 4. Visualization of robustness information on the Pareto front.

$$\sum_{i=1}^{50} w_i X_i \le b \quad (7)$$

The objective functions coefficients are randomly generated within the range [100, 1000], using uniform distribution, and they are uncorrelated ($R=0.022$). The weights (w_i) are also randomly generated in the same way and b signalizes half the sum of the weights.

$$b = 0.5 \times \sum_{i=1}^{50} w_i \quad (8)$$

The detailed data for $\mathbf{c_1}$, $\mathbf{c_2}$, $\mathbf{w}$ and b are provided in Table A1 of Appendix A.

Initially, we solve the reference problem exactly with AUGMECON2 and we obtain 54 POS*. The solution time, using GAMS with CPLEX 12.2 in a core i5 2.5 GHz, is 9.56 s. Each POS is a vector of "0" and "1" and is appropriately coded using the process described in Section 3.1. Subsequently, we perform the Monte Carlo simulation–multi-objective optimization phase. We set the sampling intervals for all the model's parameters using a disturbance parameter $\alpha=5\%$. We set $T=1000$ Monte Carlo iterations, so we solve the problem 1000 times, after performing the random sampling of the model's parameters. The modeling and solution of the combination of Monte Carlo simulation and AUGMECON2 method are performed in the GAMS platform. All 1000 iterations last 8505 s (2 h 21 min 45 s). In total we produce 2693 different POS that are allocated to the 1000 PFs.

Then, we measure the frequency of each reference POS* across the 1000 PS. The higher the frequency, the more robust is the relative POS* of the reference problem. The frequency for the p-th reference POS*, as explained in Section 2.3, quantifies the robustness of the specific POS* and is denoted as the robustness index (RI_p). The Total Robustness Index that characterizes the reference Pareto set is calculated by formula (7) and is $TRI=0.243$.

In Fig. 5 we can see the conventional Pareto front along with the enhanced Pareto front, exploiting the additional information of the robustness of POS*. In Fig. 5a, all POS* are depicted with points of the same size, while in Fig. 5b the size of each POS* is proportional to its RI. In this way, the DM has the opportunity to isolate at a glance the areas of the Pareto front that are more robust than other. Hence, he/she can use robustness as an additional criterion in the pursuit of the most preferred among the POS*.

The effect of the number of Monte Carlo iterations on the results was also examined. Namely, in addition to the 1000 Monte Carlo iterations, we also performed the robustness assessment test with $T=100$ $T=500$ and $T=1000$ iterations, in order to examine the accuracy of the results. The results for the robustness indices of the 54 POS* are shown in Fig. 6.

We also tested the stochastic process with three different seeds for the random number generator in order to delve further into the stability of the method. We used three different seeds for the Monte Carlo simulation and 1000 iterations. The total robustness index TRI for the three different seeds was calculated 0.243, 0.241 and 0.244 respectively, which are considered to be similar to each other. In addition, the maximum difference among the robustness indices for the POS* from the three runs was 8.1% and the average difference was only 1.8%. Therefore, the stability of the method for 1000 runs is considered as adequate.

It must be noted that in case uncertainty is only existent in the objective function coefficients (using again $\alpha=5\%$), sampling is performed only for these specific parameters. As it is expected, the produced POS are fewer (287 instead of 2693) and the total robustness index is TRI = 0.688, which is significantly higher than in the previous case, where all model's parameters were considered as uncertain. In general, the more parameters we consider as uncertain, the lower the robustness indices for the reference Pareto set.

4.2. Example with three objective functions

Additionally, we test the method in a multi-objective multi-dimensional knapsack problem with three objective functions, 30 items and 3 constraints. It must be noted that the tri-objective problems are much more challenging than the bi-objective ones, because the exact calculation of the Pareto set is significantly more computationally demanding (see [19]). The model of the numerical example is described by the following mathematical formulae.

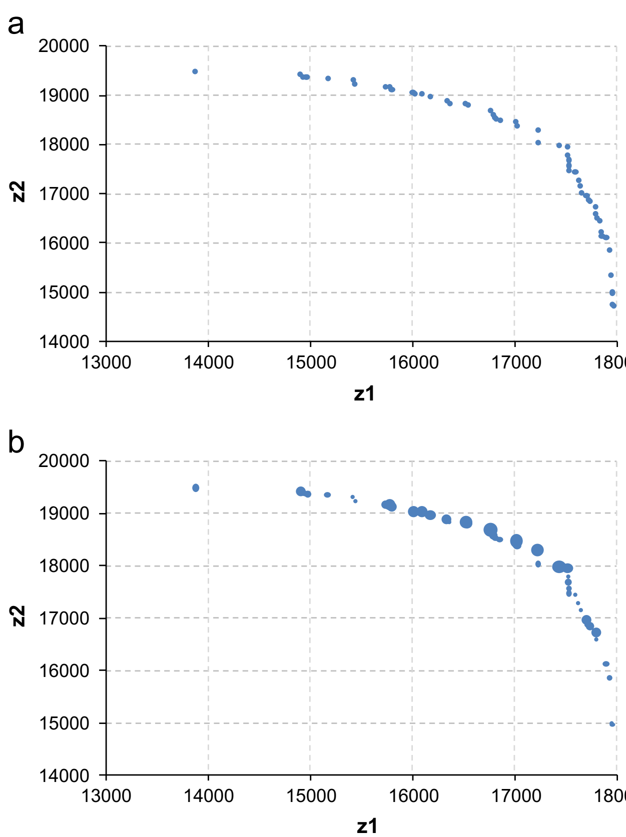


Fig. 5. Pareto fronts without (a) and with (b) robustness information embedded. (a) Pareto front, (b) Pareto front with robustness information.

The problem's input data are provided in Appendix A (Table A2).

$$\max z_k = \sum_{i=1}^{30} c_{ik} X_i \quad k = 1, \ldots, 3$$

$$st$$

$$\sum_{i=1}^{30} w_{ij} X_i \le b_j \qquad j = 1, \ldots, 3 \tag{9}$$

The objective functions coefficients are randomly generated within the range [1, 10] using uniform distribution and they are uncorrelated (greatest correlation coefficient R among the three is 0.12). The weights (w_{ij}) are similarly randomly generated in [10, 100]. b_j is half the sum of the weights as in the previous example. First, we solve the reference problem exactly with AUGMECON2 and we obtain 91 POS*. The solution time using GAMS with CPLEX 12.2, in a core i5 2.5 GHz is 78.43 s.

Then, we proceed to the robustness analysis of the PS*. In this example we consider only the objective function coefficients as the uncertain parameters. We also consider constant sampling intervals for each coefficient and we assume therefore a coefficient dependent perturbation parameter (a_{ik} instead of a). The sampling interval is defined as −1 and +1 around the original objective function coefficient. We perform 1000 iterations, in order to apply the combination of Monte Carlo simulation and AUGMECON2 method. The 1000 iterations lasted 73,098 s (20 h 18 min 18 s) and we produced 2221 different POS in total, which shape the 1000 exact Pareto sets. Subsequently, we measure the frequency of each one of the reference POS* across the 1000 PS. The Total Robustness Index that characterizes the reference Pareto set is calculated by Eq. (7) and is $TRI = 0.579$.

In Fig. 7 we can see the conventional Pareto front, along with the enhanced Pareto front, exploiting the additional information of the robustness of POS*. In this way, the decision maker can isolate at a glance the areas of the Pareto front that are more robust than other.

The values of the objective functions of the 91 POS* are presented in Appendix A (Table A4). The robustness indices for the 91 POS* are shown in Fig. 8. We observe that 22 out of 91 POS* have a Robustnes index greater than 80%. Similarly, 55 out of 91 POS* exhibit RI greater than 50% and 83 out of 91 POS* greater than 20%.

A small robustness index, such as that of POS* 34 (Fig. 8), indicates that we should probably avoid this solution, given that with a small perturbation of the objective function coefficients it will be probably get dominated by others. In other words, it will cease to be optimal. This conclusion emerges from the fact that POS* 34 is present as a non dominated solution in only 8% of the produced Pareto sets, while there are other POS* that are present in more than 90% of the Pareto sets as clearly shown in Fig. 8.

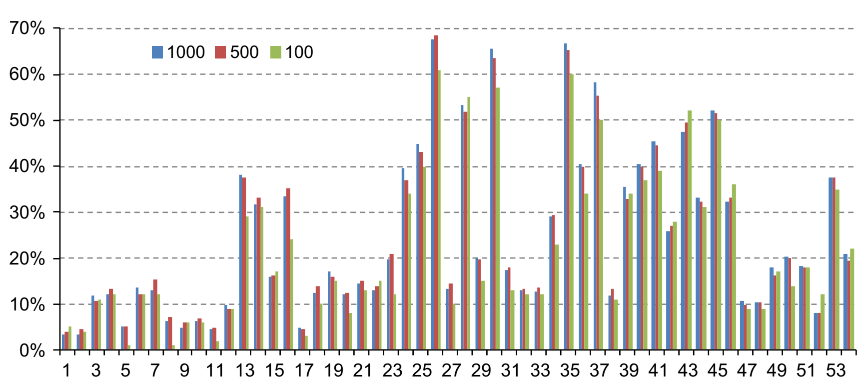


Fig. 6. The Robustness indices for different number of Monte Carlo iterations.

5. Application

In this section we illustrate the method through the use of an application dealing with project portfolio selection, taking into account two objective functions. The evaluation of the projects in the objective functions is characterized by imprecision and therefore robust POS*, which correspond to project portfolios, need to be identified. In Section 5.1 we describe the problem, then we present the corresponding MOIP model (Section 5.2) and eventually in Section 5.3 we show the robustness analysis results and discuss on them.

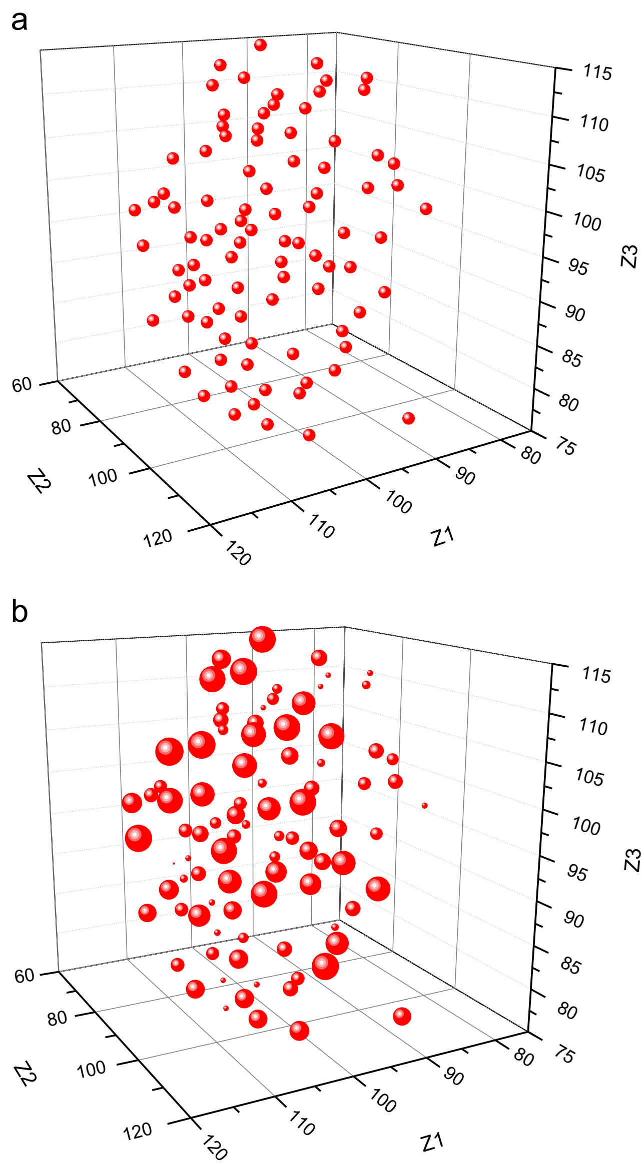


Fig. 7. Pareto fronts without (a) and with (b) robustness information embedded: (a) Pareto front, (b) Pareto front with robustness information.

5.1. Problem description

The application refers to an academic project portfolio selection problem. Similar to Mavrotas et al. [16], the case concerns a capital budgeting problem that examines the possibility of funding different research proposals. There are 108 research proposals from 6 departments. Each proposal is defined as either of basic or applied research. The cumulative cost of the 108 proposals is 19,780,000 Euros, while the available budget is 4000,000 Euros. The proposals are evaluated by a group of experts based on two criteria: (1) the project quality (denoted as *PQ*) that has to do with the project's characteristics (usefulness, innovation, quality of proposal, etc.) and (2) the human quality (denoted as *HQ*) that corresponds to the ability of the proposal's researchers to undertake the project (relevant experience of faculty, previous works, etc.). The scores for these two criteria are integers, namely a quantitative discrete, linear scoring scale in the range from 1 to 20. There are two types of segmentation constraints that need to be respected:

(1) *Department representation.* Instead of allocating evenly the total budget to the six departments and evaluate their own proposals, a comprehensive approach is proposed in order to have a more objective and effective project selection procedure. All 108 projects compete with each other but there exist upper and lower bounds for the representation of each department, which are the same for all six of them. More specifically, the share of each department must vary between 10% and 30% of the total number of projects that are eventually selected.
(2) *Type of research.* The research proposals are characterized either as applied research, or basic research in a proportion almost 2:1. Given that it is easier for the applied research projects to get funded from other sources, we limit their share, promoting therefore the basic research ones. Hence, an explicit constraint is formulated, stating that the budget allocated to basic research proposals should be at least 40% of the total budget.

The aim of the current analysis is to produce the PS* and explore the robustness of the POS* (proposals portfolios) with respect to small perturbations on the scoring of the proposals in the two criteria. In other words, we examine how sure are we that a specific portfolio remains Pareto optimal, given that the evaluation scores of the projects in the two criteria may vary ± 1 grade. For example, assume that the initial evaluation of project quality

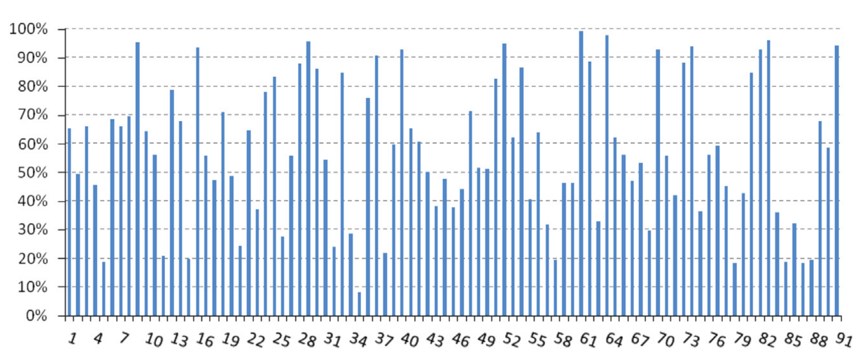


Fig. 8. The Robustness indices for the 91 POS*.

for the nth project is 17. During the robustness analysis this evaluation varies as 16, 17, 18 with equal probability for each grade. This is practiced for all 108 projects through the Monte Carlo simulation.

It must be noted that providing the whole Pareto set, we avoid the explicit use of weights in the criteria. The decision maker gets a clear picture of all Pareto optimal portfolios of projects, along with their characteristics (number of projects, allocated budget, representation of departments and type of research) before moving to his decision. Additionally, in the proposed approach, he may exploit the provided robustness information in the evaluation of the Pareto optimal portfolios when selecting his most preferred. Given that the evaluations of each project from the experts are more or less subjective, the decision maker can reduce this subjectivity by practicing robustness analysis. The robust portfolios indicate that they are more or less insensitive to small perturbations on the projects' evaluation grades.

5.2. Model building

5.2.1. Decision variables

The decision variables of the model are binary, indicating acceptance ($X_i=1$) or rejection ($X_i=0$) of the i-th research proposal in the final portfolio.

Table 2
The characteristics of the 20 POS*.

	PQ	HQ	Cost	Projects	dep1	dep2	dep3	dep4	dep5	dep6
1	490	470	3990	31	5	7	4	6	4	5
2	488	478	4000	31	5	7	4	7	4	4
3	486	483	3995	32	4	8	4	8	4	4
4	485	486	4000	32	4	8	5	7	4	4
5	484	488	4000	32	5	9	4	6	4	4
6	482	489	4000	32	5	9	4	6	4	4
7	480	492	4000	32	5	9	4	6	4	4
8	478	493	4000	32	5	9	4	6	4	4
9	477	494	4000	32	4	8	5	7	4	4
10	475	496	4000	32	4	9	4	7	4	4
11	473	498	4000	32	4	9	4	7	4	4
12	470	500	4000	32	4	9	4	7	4	4
13	468	501	4000	32	4	9	4	7	4	4
14	466	502	4000	32	4	8	4	8	4	4
15	459	504	4000	31	4	7	4	8	4	4
16	456	505	3990	31	4	7	4	8	4	4
17	455	506	4000	31	5	7	4	7	4	4
18	451	508	4000	31	4	8	4	7	4	4
19	437	509	4000	30	4	8	4	8	3	3
20	435	510	4000	30	4	7	4	9	3	3

5.2.2. Parameters

The parameters of the model are the score of each project for project quality (PQ) and human quality (HQ), the cost of each project, the total available budget (4 million Euros), the lower and upper bounds for the representation of each department (10% and 30% respectively) and the lower bound for the budget share dedicated to basic research (40%). The data of the 108 projects are given in Table A3 of Appendix A.

5.2.3. Constraints

The constraints of the model are the following:

(a) The total cost cannot exceed 4 million Euros

$$\sum_{i=1}^{108} \cos t_i \times X_i = TC$$
$$TC \le 4000 \tag{10}$$

where cost_i is the cost of the i-th proposal in thousand Euros and TC is the decision variable that expresses the total cost of the portfolio of proposals.

(b) The representation of departments in the obtained solutions are constrained by lower and upper bounds:

$$\sum_{i \in D(j)} X_i \le 0.3 \times \sum_{i=1}^{108} X_i \quad \text{for } j=1,\ldots,6$$
$$\sum_{i \in D(j)} X_i \ge 0.1 \times \sum_{i=1}^{108} X_i \quad \text{for } j=1,\ldots,6 \tag{11}$$

where $D(j)$ denotes the set of projects proposed from the j-th department.

(c) The lower bound for the share of basic research proposals is 40% of the total funding.

$$\sum_{i \in Basic} \cos t_i \times X_i \ge 0.4 \times TC \tag{12}$$

5.2.4. Objective functions

The two objective functions are: (1) maximization of total usefulness of the research portfolios and (2) maximization of total faculty sufficiency. They are modeled as

$$\max z_1 = \sum_{i=1}^{108} PQ_i \times X_i$$
$$\max z_2 = \sum_{i=1}^{108} HQ_i \times X_i \tag{13}$$

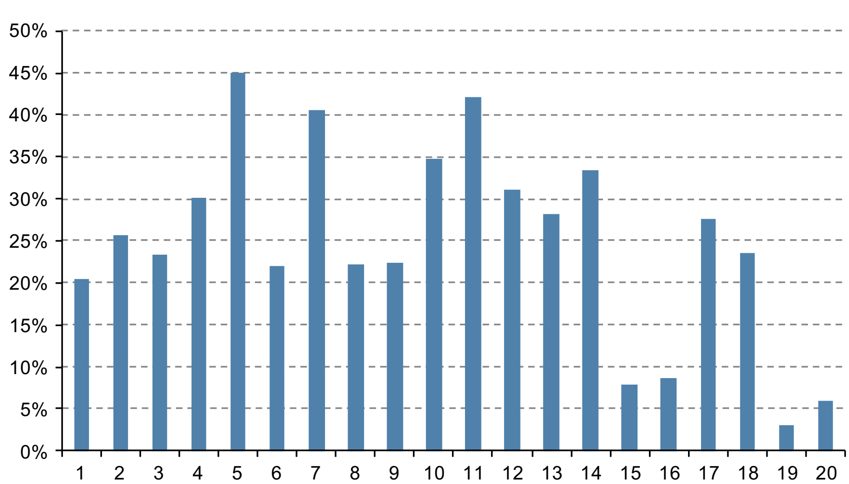


Fig. 9. The Robustness indices for the 20 POS*.

PQ_i and HQ_i are the scores for Project quality and human quality respectively for ith proposal (receiving values from 1 to 20).

5.3. Results and discussion

First, we solve the problem with the original objective function coefficients in order to obtain the reference Pareto set. Using

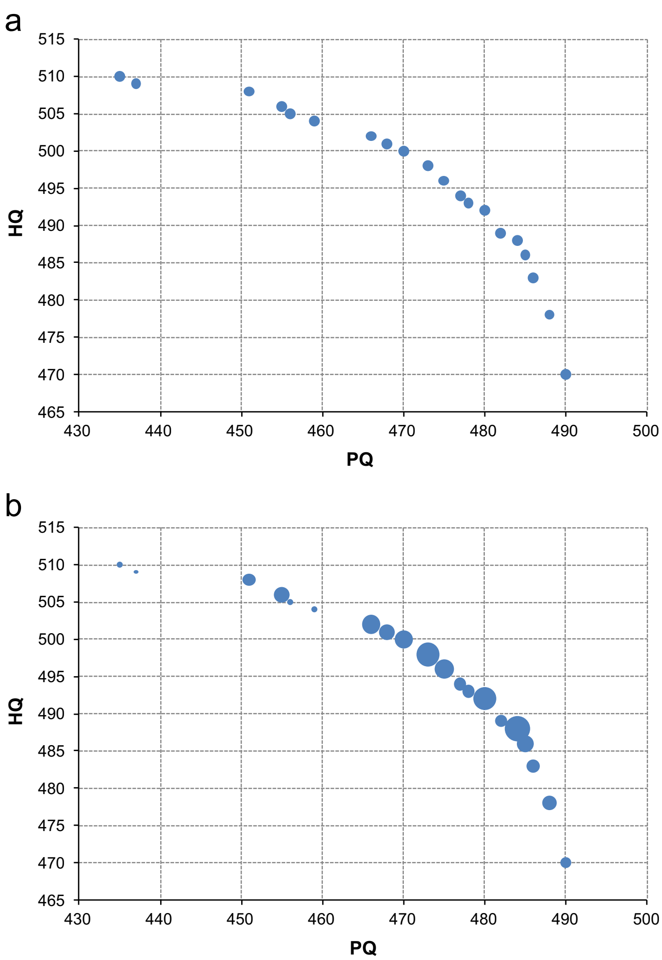


Fig. 10. Pareto fronts of the case study without (a) and with (b) robustness information embedded: (a) Pareto front, (b) Pareto front with robustness information.

AUGMECON2 (coded in GAMS and solved by CPLEX 12.2), we produce the exact Pareto set (PS*) that includes 20 POS*, which means 20 combinations (portfolios) of research proposals. The computational time in an Intel Q9650 core2 at 3 GHz with 4 GB RAM running Windows 7 at 64 b is 7.2 s.

Subsequently, analysis of robustness in the specific capital budgeting problems is performed as follows: We assume that a source of uncertainty is the evaluation score of the projects in the two criteria, as extracted by the experts. We create sampling intervals around the objective function coefficients that express the evaluation uncertainty and actually define the perturbation values. Given the values of the objective function coefficients, which range from 1 to 20, the sampling coefficients are in the neighborhood of ± 1, i.e., the sampling intervals are defined as $[PQ_i - 1, PQ_i + 1]$ and $[HQ_i - 1, HQ_i + 1]$. For example, if the score for Project Quality for a proposal is 14, in the Monte Carlo simulation approach may obtain the values 13, 14 and 15, all three with equal probabilities.

We set the number of Monte Carlo iterations to $T = 1000$ and we model the whole process in GAMS (Monte Carlo simulation with random number generation, generation of exact Pareto set and reporting). In each Monte Carlo iteration the exact Pareto set is obtained using AUGMECON2. The solution time after 1000 iterations is 6887 s (1 h 54 min 47 s). The number of different POS obtained in the 1000 Pareto sets varies from 16 to 31 across. In these 1000 Pareto sets there are 2531 different POS that at least appear in a single iteration (almost half of them appear just once).

Then, we code the obtained Pareto optimal solutions and we measure the frequency of the 20 POS* across the 1000 iterations. The values of the objective functions of the 20 POS* along with their cost and department distribution are shown in Table 2.

The robustness index RI_p for each POS* is shown in Fig. 9.

The total robustness index TRI that characterizes the Pareto set is obtained as the average of RI_p and is TRI = 24.9%. The Pareto front in the criteria space is shown in Fig. 10. Fig. 10a shows the conventional Pareto front (all Pareto points are of the same size), while Fig. 10b shows the Pareto front with the incorporated robustness information (the size of the Pareto points grows proportionally to the RI).

The visualization of robustness aids the DM to easily detect the areas of the Pareto front that are more robust than others. For instance, in our specific case the most robust area of the Pareto front is spotted at the middle. On the other hand, the area close to the maximization of HQ is the less robust. The visualization of the information is also useful with regard to specific POS*. For example, if the DM decides to focus on portfolios 4, 5 and 6

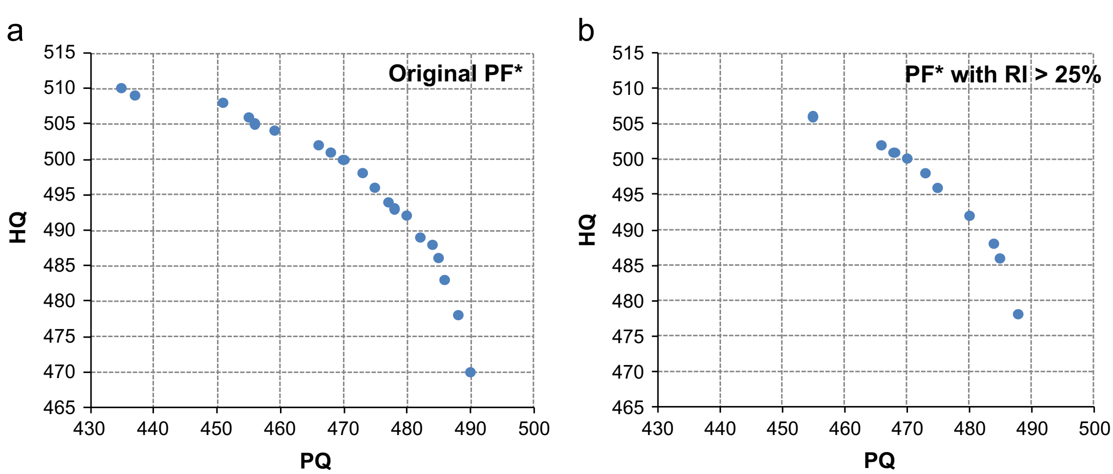


Fig. 11. Pareto fronts for different levels of Robustness (full and with RI > 25%). (a) Original Pareto front, (b) Pareto front with RI > 25%.

(4th, 5th, 6th from the bottom-right), which have close performance in the two objectives, he/she can immediately realize that solution 5 is the most robust.

Another informative chart concerns the isolation of the POS* that exhibit RI greater than a predefined threshold (see Section 3.3). In this way, we limit the information overload to the DM, reducing the number of POS* to be compared and considered for selection. In our case we have set the Robustness Index threshold to 25% and consequently only 10 out of the 20 project portfolios qualified and are presented to the DM, as shown in Fig. 11.

Conclusively, the added value of the chart embedding robustness information is obvious and significant during decision support. The DM can spot at a glance the Pareto optimal solutions as well as the regions of the Pareto front that are more robust. Therefore, robustness can be envisaged as an extra criterion to be considered in the decision making process, towards the selection of the most preferred solution. In other words, the DM, apart from the values of the objective functions, can also take into account the robustness of the Pareto optimal solutions.

6. Conclusions

In this paper we propose a methodology that can be used to assess the robustness of the Pareto set, as well as the individual Pareto optimal solutions in MOCO problems. The MOCO problems are translated into appropriate MOIP problems and are solved exactly. The key of the proposed method lies in the capability to produce the exact Pareto set in MOIP problems by using the AUGMECON2 method. We combine AUGMECON2 with Monte Carlo simulation process in order to perform a thorough examination of the parameters space. The computational complexity of the procedure is significant, because in each Monte Carlo iteration we solve a MOIP problem exactly, which means that we produce all the POS.

The most meaningful information and added value that emerges from the proposed approach is the robustness measurement of the whole Pareto front and the specific Pareto optimal solutions. We also propose the visualization of the robustness information using appropriate 2-D and 3-D charts. These charts are very helpful for the DM as he can easily detect regions of high and low robustness and focus appropriately. In addition, such charts reduce the information overload, by presenting to him/her only the “robust” Pareto optimal solutions. This information can be exploited in the decision making procedure, when it comes to selecting the most preferred Pareto optimal solution.

This study has set the bases for an elaborative analysis and measurement of the robustness of integer multi-objective problems, implying also some very interesting perspectives for future research. For instance, we made use of uniform distribution for the parameters in the Monte Carlo simulation but when relevant information is possessed, the DM can use an alternative distribution. In this case the perturbation parameter can be quantified by appropriate distribution parameters like e.g. the standard deviation of a normal distribution, and it is an instance which is interesting to study. Then, its results can be compared to the ones of the uniform distributions and calculate to what extent the *RI* may ameliorate.

Moreover, in the current study we deal only with discrete problems. As part of our future research we are also planning to extend the robustness analysis, including also continuous variables (i.e., mixed integer and linear multi-objective problems). In this case, a different approach should be followed in order to incorporate effectively the continuous variables and visualize appropriately the robustness information. Regarding the specific application to academic research projects, a future work may incorporate the uncertainty of the projects' cost to the robustness analysis. In this case we will have uncertain parameters not only in the objective functions but also in the model's constraints.

Acknowledgements

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program “Education and Lifelong Learning”. The authors would like to thank the three anonymous referees for their constructive comments.

Appendix A

See Tables A1–A4.

Table A1
The data for the first knapsack problem of the numerical example.

Item	c_1	c_2	W
1	576	235	402
2	184	186	120
3	607	657	710
4	418	990	513
5	833	574	391
6	875	286	845
7	210	178	316
8	242	391	296
9	201	418	282
10	247	529	935
11	332	183	150
12	875	940	962
13	290	224	320
14	822	282	377
15	207	998	862
16	467	687	859
17	648	815	931
18	379	121	119
19	201	478	938
20	169	561	513
21	736	277	426
22	929	203	609
23	137	177	902
24	148	551	573
25	945	709	114
26	225	793	469
27	222	903	365
28	655	547	893
29	981	426	158
30	593	225	866
31	378	387	587
32	106	927	112
33	190	545	598
34	873	590	180
35	790	850	726
36	611	582	512
37	160	915	419
38	458	185	640
39	209	501	639
40	437	189	165
41	194	240	805
42	764	972	461
43	415	426	567
44	593	560	990
45	516	865	812
46	749	316	936
47	507	804	698
48	815	531	663
49	457	447	856
50	407	935	335
RHS ($=\boldsymbol{b}$)	13958.5		

Table A2
The data for the second knapsack problem of the numerical example.

Item	c_1	c_2	c_3	w_1	w_2	w_3
1	2	2	8	84	19	89
2	7	9	1	49	96	57
3	3	5	5	68	93	82
4	9	3	5	20	64	63
5	8	1	10	97	72	75
6	3	3	2	74	91	67
7	10	8	1	60	32	99
8	8	3	4	30	96	97
9	6	6	10	13	44	95
10	6	7	5	95	76	28
11	5	10	5	19	69	83
12	9	4	10	41	82	51
13	2	6	8	17	51	99
14	1	9	2	95	38	43
15	6	1	9	73	52	38
16	3	9	3	12	22	93
17	10	5	2	66	83	76
18	6	5	8	55	27	31
19	1	6	3	75	70	75
20	1	5	2	20	56	90
21	4	8	5	56	29	32
22	10	3	7	80	89	76
23	7	9	3	59	86	84
24	2	3	6	66	48	42
25	6	3	7	25	13	33
26	7	6	9	70	95	65
27	5	5	3	95	66	44
28	7	4	2	96	94	77
29	5	4	3	62	16	11
30	6	6	1	74	44	38
	b_1	b_2	b_3			
RHS	873	906.5	966.5			

Table A4
The coordinates of the 91 Pareto optimal solutions of the 3-objective problem of Section 4.2.

#	Z_1	Z_2	Z_3	Nos.	Z_1	Z_2	Z_3
1	114	85	87	*47*	89	102	96
2	112	91	82	*48*	113	76	98
3	111	95	80	*49*	111	78	99
4	109	96	84	*50*	105	93	96
5	108	98	78	*51*	101	95	99
6	107	102	80	*52*	98	99	100
7	106	104	78	*53*	94	101	97
8	102	108	77	*54*	107	87	100
9	99	109	84	*55*	106	89	97
10	89	111	77	*56*	104	91	98
11	102	105	81	*57*	103	92	97
12	108	97	81	*58*	81	100	98
13	110	94	88	*59*	103	90	99
14	107	100	84	*60*	110	79	100
15	105	101	81	*61*	109	80	104
16	104	101	91	*62*	102	89	103
17	102	103	85	*63*	100	90	101
18	101	105	82	*64*	106	84	105
19	112	88	90	*65*	97	92	104
20	111	90	88	*66*	96	97	101
21	108	95	86	*67*	89	98	101
22	107	98	89	*68*	85	99	101
23	106	99	86	*69*	93	93	103
24	99	104	92	*70*	92	94	106
25	97	108	86	*71*	86	95	104
26	96	105	87	*72*	84	96	103
27	94	106	89	*73*	100	87	106
28	91	107	91	*74*	97	91	107
29	114	82	95	*75*	102	81	106
30	95	105	94	*76*	102	80	107
31	109	91	92	*77*	99	85	107
32	108	93	89	*78*	101	78	108

Table A3
The data for the 108 projects of the application.

#	PQ	HQ	Cost (k€)	Dept	Res. type[a]	#	PQ	HQ	Cost (k€)	Dept	Res. type[a]	#	PQ	HQ	Cost (k€)	Dept	Res. type[a]
1	18	13	320	1	A	37	10	13	250	2	A	73	10	6	165	4	A
2	15	14	160	1	A	38	18	16	150	2	A	74	14	16	260	5	A
3	16	19	185	1	A	39	6	11	130	2	A	75	9	8	160	5	A
4	18	12	330	1	A	40	19	12	365	3	A	76	6	6	155	5	B
5	12	15	110	1	B	41	14	16	120	3	B	77	20	19	280	5	A
6	11	16	100	1	B	42	20	16	280	3	A	78	14	12	205	5	A
7	20	16	110	1	B	43	10	14	120	3	A	79	10	9	215	5	A
8	9	6	150	1	A	44	18	13	300	3	A	80	20	18	330	5	A
9	20	14	210	1	A	45	8	9	115	3	B	81	11	10	255	5	A
10	18	15	130	1	B	46	14	10	120	3	B	82	14	18	105	5	B
11	11	14	190	1	A	47	17	18	190	3	A	83	8	12	105	5	B
12	18	15	240	1	A	48	16	12	335	3	A	84	12	9	150	5	A
13	12	10	240	1	A	49	15	20	160	3	A	85	12	10	100	5	B
14	15	10	190	1	A	50	20	16	240	3	A	86	13	18	150	5	B
15	18	18	180	1	A	51	10	5	110	3	A	87	12	10	130	5	B
16	14	11	170	2	A	52	14	16	190	3	A	88	15	14	210	6	A
17	16	15	185	2	A	53	17	20	125	3	A	89	12	20	270	6	A
18	12	14	215	2	A	54	14	9	305	3	A	90	7	9	145	6	B
19	11	17	105	2	B	55	14	18	135	4	A	91	14	13	285	6	A
20	14	10	110	2	B	56	12	16	190	4	A	92	19	11	145	6	B
21	11	6	160	2	B	57	18	18	310	4	A	93	19	15	255	6	A
22	15	12	155	2	B	58	13	12	215	4	A	94	11	14	165	6	A
23	6	10	155	2	B	59	10	6	270	4	A	95	14	18	180	6	A
24	16	17	190	2	B	60	20	16	315	4	A	96	13	9	140	6	A
25	14	16	125	2	A	61	16	14	240	4	A	97	18	13	130	6	A
26	6	7	135	2	A	62	16	17	150	4	B	98	19	12	185	6	A
27	17	16	105	2	B	63	18	14	110	4	B	99	8	6	125	6	B
28	10	9	205	2	A	64	16	14	135	4	B	100	12	14	260	6	A
29	20	13	220	2	A	65	14	18	150	4	B	101	9	12	165	6	A
30	14	20	180	2	A	66	18	17	115	4	B	102	10	20	165	6	A
31	16	15	190	2	A	67	12	11	130	4	B	103	20	13	250	6	A
32	10	6	190	2	A	68	14	20	180	4	B	104	17	12	150	6	A
33	20	16	115	2	A	69	15	18	165	4	A	105	13	17	175	6	A
34	6	16	105	2	A	70	14	18	130	4	B	106	17	15	110	6	B
35	20	16	145	2	A	71	17	20	195	4	A	107	20	14	260	6	A
36	11	12	115	2	A	72	5	9	105	4	A	108	18	17	280	6	A

[a] Research type: A = applied research, B = basic research.

Table A4 (*continued*)

#	Z_1	Z_2	Z_3	Nos.	Z_1	Z_2	Z_3
33	107	97	92	*79*	96	80	108
34	110	88	91	*80*	95	81	109
35	110	84	92	*81*	93	87	109
36	102	100	93	*82*	101	74	111
37	110	83	99	*83*	98	78	112
38	109	87	93	*84*	94	80	110
39	108	89	96	*85*	88	81	110
40	107	95	95	*86*	82	83	110
41	98	101	95	*87*	86	79	111
42	97	103	94	*88*	80	80	111
43	109	86	96	*89*	99	72	113
44	101	97	94	*90*	87	78	113
45	99	98	96	*91*	93	72	115
46	100	96	96				

References

[1] Ben-Tal A, Nemirovski A. Robust convex optimization. Mathematics of Operations Research 1998;23:769–805.
[2] Ben-Tal A, Nemirovski A. Robust solutions of linear programming problems contaminated with uncertain data. Mathematical Programming 2000;88:411–24.
[3] Ben-Tal A, El Ghaoui L, Nemirovski A. Robust optimization. Princeton University Press; 2009.
[4] Bertsimas D, Sim M. Robust discrete optimization and network flows. Mathematical Programming Series B 2003;98:49–71.
[5] Bertsimas D, Sim M. The price of robustness. Operations Research 2004;52(l):35–53.
[6] Bertsimas D, Brown DB, Caramanis C. Theory and applications of robust optimization. SIAM Review 2011;53(3):464–501.
[7] Deb K, Gupta H. Introducing robustness in multi-objective optimization. Evolutionary Computation 2006;14(4):463–94.
[8] Ehrgott M, Ide J, Schöbel A. Minmax robustness for multi-objective optimization problems. European Journal of Operational Research 2014;239(1):17–31.
[9] Figueira JR, Greco S, Ehrgott M. Multiple criteria decision analysis: state of the art surveys. New York: Springer; 2005.
[10] Figueira JR, Greco S, Mousseau V, Słowiński R. Interactive multiobjective optimization using a set of additive value functions. In: Branke J, et al., editors. Multiobjective optimization, LNCS, 5252; 2008. p. 97–119.
[11] Fliege J, Werner R. Robust multiobjective optimization & applications in portfolio optimization. European Journal of Operational Research 2014;234(2):422–33.
[12] Gabrel V, Murat C, Thiele A. Recent advances in robust optimization: an overview. European Journal of Operational Research 2014;235:471–83.
[13] GAMS Development Corporation. General Algebraic Modeling System (GAMS) Release 23.5.2 Washington, DC, USA; 2010.
[14] Lahdelma R, Hokkanen J, Salminen P. SMAA—stochastic multiobjective acceptability analysis. European Journal of Operational Research 1998;106:137–43.
[15] Kouvelis P, Yu G. Robust discrete optimization and its applications. Nonconvex-optimization and its applications. . Dordrecht: Kluwer Academic Press; 1997.
[16] Mavrotas G, Diakoulaki D, Kourentzis A. Selection among ranked projects under segmentation, policy and logical constraints. European Journal of Operational Research 2008;187:177–92.
[17] Mavrotas G. Effective implementation of the e-constraint method in multi-objective mathematical programming problems. Applied Mathematics and Computation 2009;213:455–65.
[18] Mavrotas G, Florios K, Vlachou D. Energy planning of a hospital using mathematical programming and Monte Carlo simulation for dealing with uncertainty in the economic parameters. Energy Conversion and Management 2010;51:722–31.
[19] Mavrotas G, Florios K. An improved version of the augmented ε-constraint method (AUGMECON2) for finding the exact Pareto set in multi-objective integer programming problems. Applied Mathematics and Computation 2013;219:9652–69.
[20] Mavrotas G, Pechak O, Siskos E, Doukas H, Psarras J. Robustness analysis in multi-objective mathematical programming using Monte Carlo simulation. European Journal of Operational Research 2015;240:193–201.
[21] Mulvey JM, Vanderbei RJ, Zenios SA. Robust optimization of large-scale systems. Operations Research 1995;43(2):264–81.
[22] Roland, J., De Smet, Y., Figueira, J.R.. On the calculation of stability radius for multi-objective combinatorial optimization problems by inverse optimization, 4OR 10; 2012. p. 379–89.
[23] Roy B. A missing link in OR-DA: robustness analysis. Foundations of Computing and Decision Sciences 1998;23:141–60.
[24] Roy B. Robustness in operational research and decision aiding: a multi-faceted issue. European Journal of Operational Research 2010;200:629–38.
[25] Roy B, Figueira JR, Almeida-Dias J. Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: theoretical results and practical issues. Omega 2014;43:9–20.
[26] Sawik T. On the robust decision-making in a supply chain under disruption risks. International Journal of Production Research 2014. http://dx.doi.org/10.1080/00207543.2014.916829.
[27] Soyster AL. Convex programming with set-inclusive constraints and applications to inexact linear programming. Operations Research 1973;21(5):1154–7.
[28] Soyster AL, Murphy FH. A unifying framework for duality and modeling in robust linear programs. Omega 2013;41(6):984–97.
[29] Steuer, R.E. (1986) Multiple criteria optimization. theory, computation and application, Krieger, Malabar, FL.
[30] Vose D. Quantitative risk analysis: a guide to Monte Carlo simulation modelling. Wiley; 1996.
[31] Wang X, Curry DJ. A robust approach to the share-of-choice product design problem. Omega 2012;40(6):818–26.
[32] Zhen L, Chang DF. A bi-objective model for robust berth allocation scheduling. Computers and Industrial Engineering 2012;63:263–72.

European Journal of Operational Research 246 (2015) 543–553

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Elicitation of criteria importance weights through the Simos method: A robustness concern

Eleftherios Siskos [a,*], Nikos Tsotsolas [b]

[a] National Technical University of Athens, 9, Iroon Polytechniou Str., 15780 Zografou, Athens, Greece
[b] University of Piraeus, 80, Karaoli & Dimitriou Str., 18534 Piraeus, Greece

ARTICLE INFO

Article history:
Received 30 January 2014
Accepted 21 April 2015
Available online 25 April 2015

Keywords:
Multiple criteria
Decision analysis
Criteria weights
Robustness analysis
Simos method

ABSTRACT

In the field of multicriteria decision aid, the Simos method is considered as an effective tool to assess the criteria importance weights. Nevertheless, the method's input data do not lead to a single weighting vector, but infinite ones, which often exhibit great diversification and threaten the stability and acceptability of the results. This paper proves that the feasible weighting solutions, of both the original and the revised Simos procedures, are vectors of a non-empty convex polyhedral set, hence the reason it proposes a set of complementary robustness analysis rules and measures, integrated in a Robust Simos Method. This framework supports analysts and decision makers in gaining insight into the degree of variation of the multiple acceptable sets of weights, and their impact on the stability of the final results. In addition, the proposed measures determine if, and what actions should be implemented, prior to reaching an acceptable set of criteria weights and forming a final decision. Two numerical examples are provided, to illustrate the paper's evidence, and demonstrate the significance of consistently analyzing the robustness of the Simos method results, in both the original and the revised method's versions.

1. Introduction

A significant factor pertaining to the non-compensatory multicriteria decision aiding models (MCDA), such as the outranking methods (i.e. ELECTRE and PROMETHEE), is the criteria weighting, or the importance of the criteria. Generally, these parameters imprint the preferences of a single decision maker (DM) to the model. The existing methods, which are widely used to assess the criteria importance weights, could be classified into two categories: (i) direct assessment procedures, where the DM is asked to explicitly express the criteria weights in terms of percentages, and (ii) indirect methods, inferring the weights from pairwise comparisons of the criteria or reference alternatives. Most of these procedures use mathematical programming formulations (see the reference by Pekelman & Sen, 1974 or the MCDA survey by Figueira, Greco, & Ehrgott, 2005).

The second category of methods includes among others:

- the method of cards proposed by Simos (1990a, 1990b) that will be described in the following section;
- the method of centralized weights (Solymosi & Dombi, 1986), which requests from the DM a number of ordinal comparisons of criteria that are formulated as linear inequalities, in order to obtain the centroid of the vertices of a polyhedron;
- the TACTIC method (Vansnick, 1986) in which the relative importance of the criteria is depicted and assessed as a system of functional representations of relations;
- DIVAPIME (Mousseau, 1995), which has been adapted to the ELECTRE methods and is implemented by making pairwise comparisons of fictitious alternatives, in order to support the elicitation of importance variation intervals;
- the analytic hierarchy process (AHP), proposed by Saaty (1994), where the DM is asked to provide pairwise comparisons over the priority of criteria on a prespecified numerical scale; and
- MACBETH (Bana e Costa, De Corte, & Vansnick, 2012) which infers the weights as values of attractiveness from pairwise comparisons of the criteria on a qualitative scale, measuring thus the magnitude of attractiveness.

Recently, Bisdorff, Meyer, and Veneziano (2014) proposed a mixed integer linear programming model to infer the criteria importance weights from overall outranking statements, by maximizing the stability of the induced median-cut outranking digraph. The outranking statements are acknowledged by the DM during an MCDA procedure.

The method proposed by Jean Simos in 1990 has gained popularity and has been applied to different types of problems, due to its simplicity, and the convenience it provides to a DM to express

* Corresponding author. Tel.: +302107723609; fax: +302107723550.
E-mail addresses: lsiskos@epu.ntua.gr (E. Siskos), ntsotsol@unipi.gr (N. Tsotsolas).

http://dx.doi.org/10.1016/j.ejor.2015.04.037
0377-2217/

her/his preferences. Specifically, it requires the construction of a hierarchy on the evaluation criteria, by involving the DM to a "playing cards" procedure, in order to attribute numerical values to them. Nevertheless, the process recommended by Simos and its revised version proposed by Figueira and Roy (2002) have some robustness issues. In particular, they arbitrarily calculate a unique weighting vector, even though there exist infinitely more weight vectors, also satisfying the preferential statements, which have been defined by the DM during the initial arrangement of the cards.

According to Figueira et al. (2005), a robustness concern consists of all possible ways that contribute in building synthetic recommendations based on robust conclusions. In the case of Simos method, a well-structured framework will be developed and used, in order to further facilitate the study of robustness concerns in outranking methods. Furthermore, the framework will address several other issues, affecting the quality of the outcome, in terms of robustness, such as the level of ratio z, introduced in the revised version of Simos (Shanian, Milani, Carson, & Abeyarante, 2008). In addition, the framework has to be appropriately adapted to support the implementation of the ELECTRE and PROMETHEE methods when interaction between criteria is taken into account (Figueira, Greco, & Słowiński, 2009) and when a multiple criteria hierarchy process is applied.

The aim of this paper is to expose the arbitrariness of the estimations made through the Simos method (robustness problem) and to propose amendment measures, in order to support DMs in identifying the preferable importance weights themselves. From this point of view, the methodological recommendations propounded in this paper should be considered as complementary and indispensable, when choosing to practice the method. All these rules and measures form a methodological framework, which, if adapted to the original or revised method of cards, can now be referred to as the Robust Simos Method.

A brief presentation of the Simos method in its original and revised versions, accompanied by an extended literature review of its implementation, is provided in Section 2. The robustness issues associated with the method are outlined in Section 3, while Section 4 proposes some robustness rules and measures. These formulate the Robust Simos Method, which supports the elicitation of a representative and "acceptable" set of weights. In Section 5, two numerical examples illustrate the paper's evidence and propositions. These experiments come to prove the massive impact of the instability of the weights on the robustness of the final results. The conclusions of this paper are in Section 6.

2. A review of the Simos methods

This section describes the original Simos method, as well as its revision by Figueira and Roy (2002). Section 2.2 presents the state-of-the-art, of the use of the method, in the scientific literature.

2.1. Description of the Simos method

The original Simos method consists of the following three steps, concerning the interaction with the DM and the collection of information:

1. The DM is given a set of cards with the name of one criterion on each (n cards, each corresponding to a specific criterion of a family F). A number of white cards are also provided to the DM.
2. The DM is asked to rank the cards/criteria from the least to the most important, by arranging them in an ascending order. If multiple criteria have the same importance, she/he should build a subset by holding the corresponding cards together with a clip.
3. The DM is finally asked to introduce white cards between two successive cards (or subsets of ex aequo criteria) if she/he deems that the difference between them is more extensive. The greater the difference between the weights of the criteria (or the subsets of criteria), the greater the number of white cards that should be placed between them. Specifically, if u denotes the difference in the value between two successive criteria cards, then one white card means a difference of *two times u*, two white cards mean a difference of *three times u*, etc.

The information provided by the DM is utilized by the Simos method for the determination of the weights, according to the following algorithm:

i. ranking of the subsets of ex aequo from the least important to the most important, considering also the white cards,
ii. assignment of a position to each criterion/card and to each white card,
iii. calculation of the non-normalized weights, and
iv. determination of the normalized weights.

The least qualified card is given *Position 1*, while the most qualified one receives *Position n*. The non-normalized weight of each rank/subset is determined by dividing the sum of positions of a rank, by the total number of criteria belonging to it. The non-normalized weights are then divided by the total sum of positions of the criteria in each rank (excluding the white cards), in order to normalize them. The obtained values are rounded off to the lower or higher nearest integer value.

Following the criticism of Scharlig (1996) that the method processes the information unrealistically, Figueira and Roy (2002) expressed objections to the way the Simos procedure determines the weights. One of the main issues indicated is that it elicits only one set of weights that satisfies the model expressed by the DM. However, other sets of weights could probably better fit the DM's preferences on the relative importance of the criteria. Such sets of weights cannot be obtained by the Simos' procedure. A second point of criticism is that the procedure processes criteria with the same importance (i.e. the same weight), in a non-robust way. If one tries to re-order the cards between two subsets, she/he realizes that the distance (difference of weights) between the subsequent subsets has changed in an uncontrolled way. This phenomenon occurs because the difference of weights between two successive subsets of criteria is automatically influenced by the number of existing cards in these subsets. The user however "does not have a real or absolute perception of the way in which the numerical values are determined by the procedure". Finally, Figueira and Roy do not agree with the rounding of the normalized weights to 100, because they perceive this as a non-realistic process.

In their effort to address these issues, Figueira and Roy (2002) proposed a revised version of the Simos method. In addition to the three-step data collection process, the new procedure introduces a fourth step, which demands from the DM to state *"how many times the last criterion is more important than the first one in the ranking"* (ratio z). This ratio is used in order to define a fixed interval between the weights of criteria or their sub-sets. The variable u denotes this interval: $u = (z - 1)/e$, where e is the number of different weight classes (namely single card, subsets of cards, and white card).

2.2. State of the art

The Simos method, although exhibiting considerably easy, almost naïve, data collection and implementation, has been extensively used in the scientific literature. Several authors have made use of the method, mostly combined with ELECTRE type methods, in order to assess the importance of the criteria weights. A review of the literature unveiled a very wide area of application, from energy planning and environmental evaluation problems, to project selection and mechanical engineering problems. Forty such applications are depicted in Table 1. It has also been noticed that many business and market surveys make use of the Simos method, in order to assign weights to the evaluation criteria, but they are rarely published.

Table 1
A review of applications of the Simos method.

Research paper	MCDA method	Revised Simos	Stability/ Sensitivity	Application field, type of problem
Shanian et al. (2008)	ELECTRE III	Yes	Yes	Industrial Design, Material Selection
Özcan, Çelebi, and Esnaf (2011)	ELECTRE I	No	No	Warehouse location selection
Kodikara, Perera, and Kularathna (2010)	PROMETHEE	Yes	No	Urban water supply problem
Fontana, Morais, and Almeida (2011)	SMARTER (SMARTS)	Yes	No	Urban water conservation Strategies
Marzouk, Hamid, and El-Said (2014a)	Priority index model	No	No	Projects prioritization
Marzouk, Nouh, and El-Said (2014b)	Rating model	No	No	Projects rating system
Pictet and Bollinger (2008)	MAVT	No	No	Public procurement in Switzerland
Herssens, Jureta, and Faulkner (2008)	PROMETHEE I	Yes	No	Service selection process
Merad, Verdel, Roy, and Kouniali (2004)	ELECTRE TRI	Yes	No	Risk zoning and management
Cavallaro (2009)	PROMETHEE	No	No	Assessment of concentrated solar thermal technologies
Augusto, Lisboa, Yasin, and Figueira (2008)	ELECTRE III	Yes	No	Firms benchmarking
Jabeur and Martel (2007)	Ordinal sorting method	Yes	No	Group decision making
Dawson and Schlyter (2012)	MAVT	Yes	No	Site suitability for solar thermal power station
Ashari and Parsaei (2014)	ELECTRE III	No	No	Weapon selection
Cavallaro (2010)	ELECTRE III	No	No	Assessment of thin-film photovoltaic production processes
Shanian and Savadogo (2006)	TOPSIS	Yes	No	Material selection
Jabeur et al. (2004)	ELECTRE III, PROMETHEE I, II	Yes	No	Group decision making
Aretoulis et al. (2010)	Linear weighting model	No	No	Supplier evaluation and selection
Alexopoulos et al. (2012)	ELECTRE II	No	No	Evaluation of strategic publishing actions
Bojovic, Bonzanigo, and Giupponi (2012)	ELECTRE	Yes	No	DSS development
Kostoglou et al. (2014)	AHP	No	No	Comparative analysis of Greek Universities web sites
Huaylla et al. (2013)	ELECTRE III	No	No	Selection of residential photovoltaic systems
Wilkens and Schmuck (2012)	PROMETHEE	Yes	No	Evaluation of energy scenarios
Merad et al. (2013a)	ELECTRE III	Yes	No	Evaluation of sustainable development actions within an organization
Madlener, Kowalski, and Stagl (2007)	PROMETHEE I, II	Yes	No	Appraisal of renewable energy scenarios
Cai, Liao, and Wang (2012)	GDM method	Yes	No	Group decision making
Azadeh et al. (2011)	TOPSIS	Yes	No	Optimum operator assignment in cellular manufacturing systems
Merad, Dechy, Marcel, and, Linkov (2013b)	ELECTRE III	Yes	No	Assessment of the governance of sustainability
Merad, Dechy, and Marcel (2011)	ELECTRE III	Yes	No	Assessment of sustainable development actions
Xidonas et al. (2007)	PROMETHEE I, II	No	No	Appraisal of consumer credit banking products
Jamali and Nejati (2009)	TOPSIS	Yes	No	Material selection problem
Bahraminasab and Jahan (2011)	VIKOR	Yes	No	Material selection problem
Pictet and Bollinger (2005)	Various MCDA methods	Yes	No	Group decision making in public procurement
Shanian et al. (2012)	ELECTRE III	Yes	No	Material selection
Esseghir and Mellouli (2006)	ELECTRE TRI	Yes	No	DSS for the evaluation of scientific research structures
Brifcani et al. (2012)	TOPSIS, VIKOR, ELECTRE	Yes	No	Review of techniques for material selection
Figueira et al. (2011)	Electre Tri-C	Yes	No	Assisted reproduction sorting model
Ribas and da Silva Rocha (2014)	Fuzzy AHP	Yes	No	Prioritization of energy efficiency investments
Bollinger, Maystre, and Slowinski, (1997)	ELECTRE III	No	Yes	Location of an incineration plant for municipal waste
Wang and Zionts (2014)	-	Yes	No	Comparison of generalized Rank-Sum method with Simos

It is recognized that Figueira and Roy triggered a wide application of the Simos methodology, since most of the aforementioned studies make use of their revised version. However, none of them investigates the robustness of the method, or examines the stability of the weights obtained by Simos, except partially for Shanian et al. (2008), who examine the impact of the parameter "z", on the final results. The majority of papers use Simos for simplicity and convenience reasons, without examining or analyzing its risks, a fact which jeopardizes the quality and questions the validity of their results.

3. Some theoretical concerns

In the frame of the theoretical aspect of this study, two formal properties, related to the results of the Simos criteria weighting method(s), are considered:

Property 1. The weighting solution of the Simos method(s) is a vector of a non-empty convex polyhedral set.

Proof. Without loss of generality, suppose the family of n criteria $g_1, g_2, \ldots, g_n$ is rearranged in such a way that g_j is less or equally important than g_{j+1} for every $j = 1, 2, \ldots, n-1$, according to steps 1 and 2 of the Simos procedure. Suppose additionally that the DM has introduced into the ranking k white cards $wc_1, wc_2, \ldots, wc_k$ $(k < n)$ during step 3. Let us denote $p_1, p_2, \ldots, p_n$ the unknown weights of criteria, and $w_1, w_2, \ldots, w_k$ the unknown weights of the white cards.

Then, the following system of linear relations results:

For every $j = 1, 2, \ldots, n-1$, and $h = 1, 2, \ldots, k$:

- If g_j is followed by g_{j+1}, and g_{j+1} belongs to the same importance class as g_j, write:

$$p_j = p_{j+1} \quad (1)$$

- If g_j is followed by g_{j+1}, and g_{j+1} belongs to a higher importance class, write:

$$p_j < p_{j+1} \Leftrightarrow p_{j+1} - p_j \geq \delta \quad (2)$$

- If between g_j and g_{j+1} a white card wc_h has been inserted, write:

$$p_j < w_h \text{ and } w_h < p_{j+1} \Leftrightarrow w_h - p_j \geq \delta \text{ and } p_{j+1} - w_h \geq \delta \quad (3)$$

$$p_1 + p_2 + \ldots + p_n = 1 \quad (4)$$

$$p_1 \geq 0, p_2 \geq 0, \ldots, p_n \geq 0; w_1 \geq 0, w_2 \geq 0, \ldots, w_k \geq 0 \quad (5)$$

In the case of the revised Simos method, the following equation should be added, where the z value is actually articulated by the DM:

$$p_n = zp_1 \quad (6)$$

δ is set equal to a minimal value, say 0.01 (1 percent) for instance, to distinguish between two consecutive classes of the ranking. Then all the Simos weighting solutions belong to the polyhedral set:

$$\boldsymbol{P} = \{\, \boldsymbol{p} \in \boldsymbol{R}^n \,/\boldsymbol{p} \text{ satisfies the system of relations (1)–(6)} \} \quad (7)$$

It is still pending to prove that $\boldsymbol{P}$ is non-empty. Suppose, without loss of generality, that no criteria are of equal importance according to the DM's answers. This implies that there are no equations of type (1). Obviously, when transforming the inequalities (2)–(3) into equations, by adding $(n + k - 1)$ slack variables, the linear system (2)–(6) comprises $(n + k) + (n + k - 1) = 2(n + k) - 1$ variables and only $(n + k - 1) + 1 + 1 = (n + k) + 1$ equations. Consequently, $\boldsymbol{P}$ is non-empty because it always has a solution.

Property 2. The polyhedral set $\boldsymbol{P}$ either contains a single criteria weighting or an infinite number of weighting vectors that are all consistent with the DM's criteria ranking.

Proof. Following the findings of Property 1, it suffices to identify cases where only one weighting solution exists. Obviously, two such cases could be stated:

Case 1. All criteria are equally important (the DM's ranking consists of a single class of importance). It is thus obvious that $\boldsymbol{P} = \left\{\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)\right\}$.

Case 2. The DM classifies the first n_1 criteria in one weak class and the rest of n_2 criteria in a second strong class ($n_1 + n_2 = n$); no white card is inserted between the two classes. Let us denote p and q as the common weights of the criteria in the two classes, respectively. According to the relations (1), (4), and (6), the following two relations hold:

$$n_1 p + n_2 q = 1$$

$$q = zp, \text{ where } p = \frac{1}{(n_1 + n_2 z)}$$

It follows that $\boldsymbol{P}$ comprises again a single weighting solution in which the first n_1 components are equal to p and the remaining n_2 are equal to q.

4. Robust recommendations

It has previously been shown that the information extracted by the DM, as part of the original, as well as the revised Simos method, is not sufficient to ensure a single or robust criteria weighting. On the contrary, there exists an infinite number of weighting vectors that are consistent to the DM's ranking, which form a hyper-polyhedron $\boldsymbol{P}$. In other words, although these weighting vectors are compatible with the DM's ranking of the set of criteria, she/he is totally unaware of their existence and unjustifiable exclusion during the implementation of the method. Apart from this, it is possible that this plethora of different sets of weights causes major variation in the final results, when used in outranking or other MCDA approaches, e.g. rank reversals, when implementing the ELECTRE II–III methods.

Consequently, the set of robustness rules, which are developed in Section 4.1, should be accounted for and applied by decision analysts, when selecting to implement the Simos procedure for the assessment of the criteria weights.

4.1. Robustness rules

The pursuit of one or more of the following robustness rules is essential, if the analyst wishes to solidly implement the Simos method and obtain tangible and adequately supported results.

1. Compute the variation range of the weight of each separate criterion, by solving $2n$ linear programs of the following type (Max–Min approach):

$$\text{Min}\, p_j \;\&\; \text{Max}\, p_j, \text{ for } j = 1, 2, \ldots, n \tag{8}$$

s.t.

$$\boldsymbol{p} \in \boldsymbol{P} \tag{9}$$

2. Compute the average weighting vector (barycenter) of all different vectors (from the $2n$ solutions obtained in the former rule), as a more representative weighting solution in the hyper-polyhedron $\boldsymbol{P}$ (M-N Average); see also Greco, Kadziński, Mousseau, and Słowiński (2011) about the specification of representative parameter sets.
3. Find and record all the vertices of the polyhedron $\boldsymbol{P}$, by using the Manas and Nedoma (1968) analytical algorithm. This algorithm traverses the vertices of the Hamiltonian graph, listing them simultaneously, until all are identified. Then, calculate a new average weighting vector, which represents the barycenter of $\boldsymbol{P}$ (see also the related technique proposed by Solymosi & Dombi, 1986).
4. Implement a random weight sampling algorithm/technique to produce and statistically analyze a great number of weighting sets from the polyhedron. A relevant technique is the stochastic multiobjective acceptability analysis (SMAA) and has been devised by Lahdelma, Hokkanen, and Salminen (1998). A significant number of sampling algorithms have already been proposed (see Lovasz, 1999; Smith, 1984; Smith & Tromble, 2004; Tervonen, Valkenhoef, Baştürk, & Postmus, 2012). The analyst is then capable of computing an average weighting solution, which could be considered as a representative solution within $\boldsymbol{P}$.
5. Visualize the ranges of variation of the criteria weights and/or the polyhedron they define, in order to more comprehensively perceive the extent of the instability.
6. Following the implementation of recommendation 4, calculate the ratio of the volume of the criteria polyhedron and the unconstrained criteria area. When implementing a random sampling algorithm, such as the "Simplex Point Picking", this ratio is interpreted as the number of samples that hit the criteria polyhedron to the number of all samples.
7. For each of the recommendations from 2 to 4, compute the robustness measure *ASI* (Average Stability Index), which is the mean value of the normalized standard deviation of the estimated weights:

$$\text{ASI} = 1 - \frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{m\sum_{j=1}^{m} p_{ij}^2 - \left(\sum_{j=1}^{m} p_{ij}\right)^2}}{\frac{m}{n}\sqrt{n-1}} \tag{10}$$

where m is the number of weighting instances of the system, n is the number of criteria, and p_{ij} is the weight of ith criterion for the jth instance. ASI takes values in the interval [0,1], and usually the robustness of the weights is considered as adequate, when convergence to the maximum value of 1 is achieved (Grigoroudis & Siskos, 2010).

4.2. Robust decision aiding

Suppose now that a decision analyst is willing to use a multicriteria outranking or other decision aiding model, in order to choose an action, rank or sort a set of actions A, taking into account that the criteria weightings are confined in the hyper-polyhedron $\boldsymbol{P}$. In this case, the following activities are recommended:

- Build on A two distinct outranking relations, the necessary outranking ($aS^{N}b \Leftrightarrow aSb$, i.e. action a outranks action b, for every weighting vector $\boldsymbol{p} \in \boldsymbol{P}$), and the possible outranking ($aS^{P}b \Leftrightarrow$ there is at least one weighting vector $\boldsymbol{p} \in \boldsymbol{P}$ for which aSb; (see Figueira et al., 2009; Greco, Mousseau and Słowiński, 2008 for definitions and properties of these outranking relations).
- Define the maximum and minimum possible ranking positions for every action in A, by means of mixed integer linear programming techniques (see Kadzinski, Greco, & Slowinski, 2012).
- In the cases of ELECTRE I and ELECTRE IS choice methods, statistically compute the possibility/probability that an action a belongs to the kernel of the outranking graph.

Table 2
Some weighting sets calculated using different algorithms.

	p_a	p_b	p_c	p_d	p_e	p_f	p_g	p_h	p_i	p_j	p_k	p_l
Simos (no z)	13.00	9.00	2.00	5.00	12.00	9.00	2.00	13.00	9.00	9.00	15.00	2.00
Revised Simos ($z = 6.5$)	13.20	8.80	2.40	4.50	11.00	8.80	2.40	13.20	8.80	8.80	15.30	2.40
Manas–Nedoma average (no z)	14.84	7.25	0.80	2.86	10.65	7.25	0.80	14.84	7.25	7.25	25.41	0.80
Max–Min average (no z)	15.20	6.92	0.93	3.17	11.41	6.92	0.93	15.20	6.92	6.92	24.57	0.93
Manas–Nedoma average ($z = 6.5$)	13.44	8.23	2.64	4.45	10.69	8.23	2.64	13.44	8.23	8.23	17.16	2.64
Max–Min average ($z = 6.5$)	13.04	7.92	2.81	4.82	10.74	7.92	2.81	13.04	7.92	7.92	18.27	2.81

Table 3
ASI of the different algorithms.

	Manas–Nedoma (no z)	Max–Min (no z)	Manas–Nedoma ($z = 6.5$)	Max–Min ($z = 6.5$)
ASI	0.809	0.811	0.922	0.920

- Following the implementation of random sampling within ***P***, compute entropy measures associated with outranking relations between actions in A and ranking positions for each action separately (see Greco, Siskos, & Slowinski, 2013). Entropy is a statistical index indicating either the probability that an alternative *a* outranks an alternative *b*, or the probability that alternative *a* maintains its initial position in the ranking.

4.3. An outline of the Robust Simos Method

All the above ideas and set of actions should be integrated into a methodological framework that is called Robust Simos Method. The new algorithmic procedure comprises an initialization phase, which consists of transforming the DM's hierarchy of criteria into a convex n-dimension polyhedron ***P***, which is defined by all the linear constraints (1)–(6) cited in Section 3. This is followed by a holistic procedure, consisting of two robustness control stages.

The purpose of the first stage is the examination and analysis of the stability of the criteria weights, making use of the robustness rules proposed in Section 4.1. In the second stage of the Robust Simos Method, the focal point is the implementation of the MCDA method and the decision support procedure that follows. Specifically, the stability of the decision model's results is assessed, in close cooperation between the analyst and the DM, practicing certain robust decision aiding measures (cf. Section 4.2). In case the results of the dual robustness control procedure are judged as unacceptable and/or inadequate to support a decision, the analyst may request additional preference information by the DM in order to further shrink the polyhedron ***P***.

5. Numerical experiments

Section 5 builds upon the theoretical robustness concerns and recommendations, presented in the previous sections, and experiments on their efficacy. Several robustness measurements and recommendations are examined and implemented, in order to test their efficiency. Two numerical examples are presented: (i) the classic environmental planning problem originally proposed by Simos (1990a) and Maystre, Pictet, and Simos (1994), concerning the evaluation of environmental solutions, and (ii) a metro lines extension planning problem.

5.1. Environmental planning problem

Let us consider a family *F* of twelve criteria, used by Maystre et al. (1994) and later by Figueira and Roy (2002), to implement the Simos methods: $F = \{a, b, c, d, e, f, g, h, i, j, k, l\}$. Let us also suppose that the DM groups together the cards, associated with the criteria of the same importance, into six different subsets of *ex aequo,* and that she/he also uses a white card, as follows:

Ranking of the subset of *ex aequo* criteria: *{c, g, l}, {d}, {white card}, {b, f, i, j}, {e}, {a, h}, {k}*

For the use of revised Simos, based on the preferential information in formula (6), the value of *z* is set to *6.5,* which means: $p_k = 6.5p_c$.

The preferential information concerning the grouping of the criteria cards, the insertion of the white card and the setting of the *z* value to 6.5, has been predefined by Jean Simos and the authors of the revised Simos method. It is thus deliberately preserved for the case of the numerical example presented here.

For the complete set of requirements of the Simos method, the polyhedron ***P*** is defined by the following linear constraints (relations 1–6; $\delta = 0.01$):

Polyhedron P:

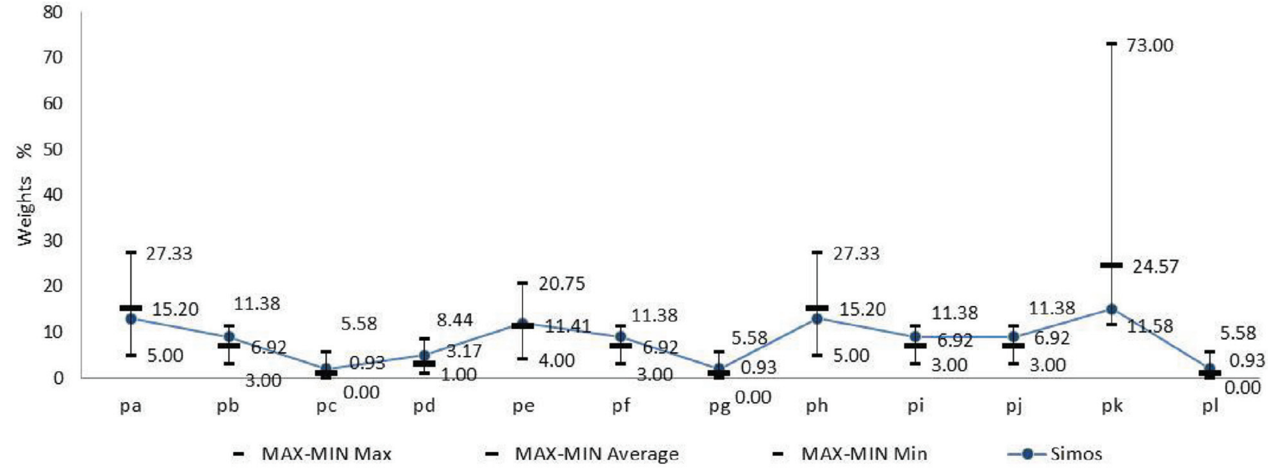
$$p_g - p_c = 0;\ p_l - p_g = 0;\ p_d - p_l \geq 0.01;\ w - p_d \geq 0.01;$$
$$p_b - w \geq 0.01;\ p_f - p_b = 0;\ p_i - p_f = 0;\ p_j - p_i = 0;$$
$$p_e - p_j \geq 0.01;\ p_a - p_e \geq 0.01;\ p_h - p_a = 0;$$
$$p_k - p_h \geq 0.01;\ p_k = 6.5p_c;$$
$$p_a \geq 0;\ p_b \geq 0;\ p_c \geq 0;\ p_d \geq 0;\ p_e \geq 0;\ p_f \geq 0;\ p_g \geq 0;$$
$$p_h \geq 0;\ p_i \geq 0;\ p_j \geq 0;\ p_k \geq 0;\ p_l \geq 0;\ w \geq 0.$$

In a first step, *Robustness rules 1* and *5* are practiced, which are related to the calculation and visualization of the variation ranges of the criteria weights. After solving the *2 x n* = 24 linear programs (8)–(9), having set $\delta = 0.01$, the range (max-min) of the weights is obtained. Fig. 1 compares these weight ranges with the values obtained by the Simos and the revised Simos procedures.

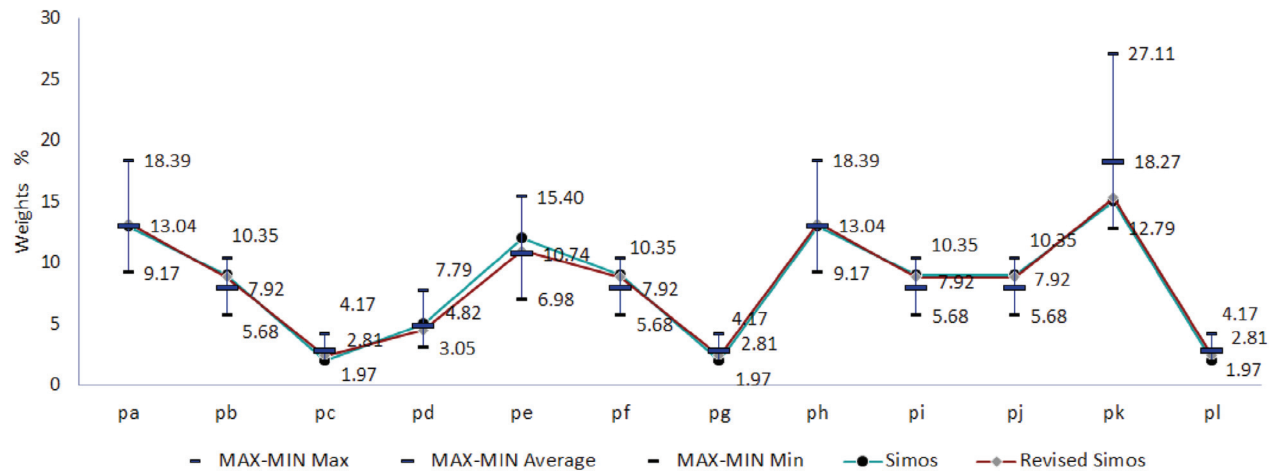
The ranges between the minimum and the maximum value of each weight are remarkably significant, and in the specific case, when the constraint $z = 6.5$ is missing (Fig. 1a), they are totally uncontrollable. For instance, the range for the weight of the *k-th* criterion, p_k, in that case is [11.58 – 73.00].

Table 2 (two first rows) shows the normalized weighting vectors as calculated by the procedure of the Simos and the revised Simos methodology. It also depicts the barycenters, as proposed by *Robustness rule 3*, of the sets of solutions calculated with the Manas-Nedoma (*Robustness rule 4*) and the Max-Min algorithms with and without the accounting of the *z* value. Table 3 presents the values of ASI (*Robustness rule 7*) corresponding to the last four cases in Table 2. Specifically, for the case of the Manas-Nedoma algorithm, ASI is equal to 0.922, while the exclusion of ratio *z* causes the decrement of ASI to 0.809.

It is obvious that the introduction of ratio *z* considerably improves the robustness of the weighting solutions, but the space in which the latter are constrained, as shaped by the DM's arrangement of the criteria cards, is still excessive. This fact jeopardizes the stability of the evaluation results and may lead to decisions that are not in tune with the DM's viewpoints. In particular, there still exists a multiplicity of different set of weights, arbitrarily not calculated by the revised Simos method, which are possible to cause variations in the results.



(a) Weight ranges when the preferential information z is not considered (Original Simos)



(b) Weight ranges when the preferential information z is considered (Revised Simos)

Fig. 1. The calculated range of the weights using different algorithms, (a) weight ranges when the preferential information z is not considered (original Simos), (b) weight ranges when the preferential information z is considered (revised Simos).

The effect of the weightings instability on the evaluation results is examined and analyzed in the following example.

5.2. Metro lines extension planning problem

The application example presented here is a hypothetical problem, concerning the prioritization of six possible metro lines extensions, based on the criteria modeling proposed by Hugonnard and Roy (1983); the data are given in Hurson and Siskos (2014). The following evaluation criteria are used in the evaluation and ranking of the six alternative actions:

- g_1: Number of habitants and workers served per km of line (scale: 0–1 millions)
- g_2: Number of travellers per km of line per day (scale: 0–75 thousands)
- g_3: Construction cost per km of line (scale: 0–125 millions €)
- g_4: Return rate of the investment (scale: 0–20 percent)
- g_5: Network organization index (scale: 0–1)
- g_6: Urban efficiency index (scale: 0–1)

All the separate evaluations of six metro extension lines over the six criteria are presented in Table 4.

Table 4
Multicriteria evaluation of the six metro extension lines.

Extension	g_1	g_2	g_3	g_4	g_5	g_6
A	0.30	40	50	10	0.8	0.5
B	0.18	35	35	15	0.8	0.8
C	0.10	20	25	12	0.5	0.6
D	0.15	30	30	15	0.5	0.6
E	0.80	60	100	8	0.6	0.4
F	0.60	50	80	10	0.5	0.5

The DM for this problem groups the cards associated with the criteria having the same importance, in four different subsets of *ex aequo*, as follows: {5,6}, {4}, {1,2}, {3}. She/he also makes use of white cards, which lead to the following subsets of *ex aequo* criteria: {5,6}, {white card}, {4}, {1,2}, {white card}, {3}.

5.2.1. Application of the original Simos methodology

For the first case, where no white cards are inserted, a set of weight vectors is produced by applying the Max–Min and the Manas–Nedoma algorithms. The different weight vectors are presented in Table 5, along with the weight vector produced by applying

Table 5
The weight vectors of $\boldsymbol{p}^i \in \boldsymbol{P}$ calculated (percent) using different algorithms (no white cards).

Weighting rule	p_1	p_2	p_3	p_4	p_5	p_6
Simos	21.00	21.00	29.00	15.00	7.00	7.00
Max–Min average	22.70	22.70	32.90	12.50	4.60	4.60
Manas–Nedoma average	19.57	19.57	43.57	10.90	3.19	3.19
1st vertex of the polyhedron	2.00	2.00	95.00	1.00	0.00	0.00
2nd vertex of the polyhedron	17.33	17.33	18.33	16.33	15.33	15.33
3rd vertex of the polyhedron	25.00	25.00	26.00	24.00	0.00	0.00
4th vertex of the polyhedron	32.67	32.67	33.67	1.00	0.00	0.00

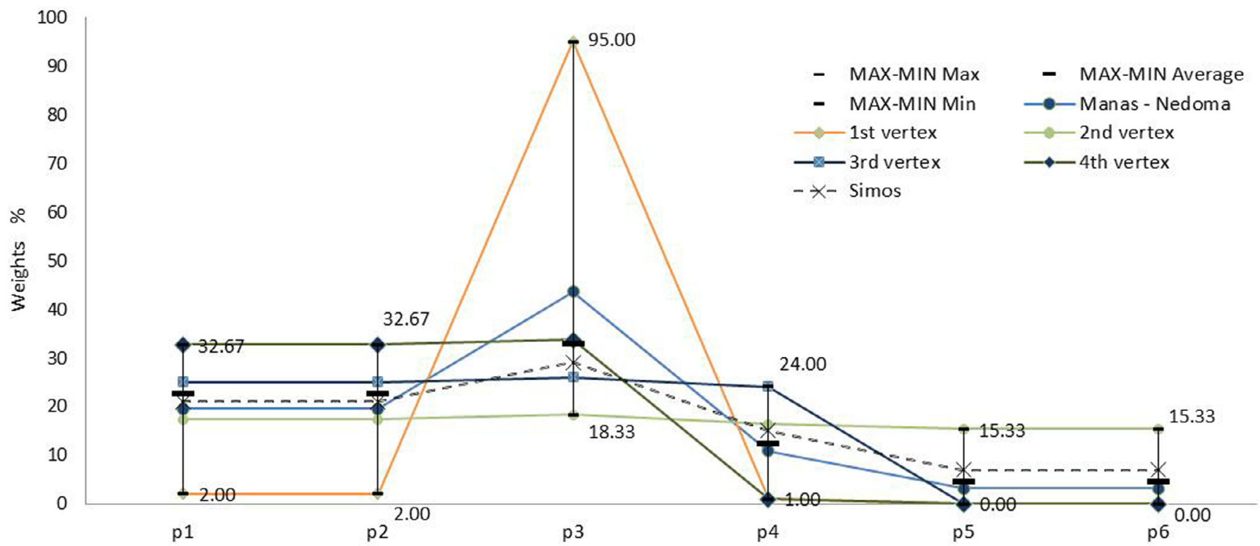


Fig. 2. The calculated range of the weights using different algorithms (no white cards).

Table 6
Ranking of the six alternatives (case with no white card).

Weighting rule	1st	2nd	3rd	4th	5th	6th
Simos	B,D	–	C	A,F	–	E
Max–Min average	B	A,D	–	C,E	–	F
Manas–Nedoma average	B,D	–	C	A,F	–	E
1st vertex of the polyhedron	C	D	B	A	F	E
2nd vertex of the polyhedron	B	A,D	–	C,E	–	F
3rd vertex of the polyhedron	B	F	D,E	–	A	C
4th vertex of the polyhedron	E	F	A	B	D	C

the Simos procedure. The four different weight vectors represent four distinct vertices of the hyper-polyhedron.

Each one of the weight vectors $\boldsymbol{p} \in \boldsymbol{P}$ of Table 4 represents a value system, which is compatible with the preferences expressed by the DM, during the initial arrangement of the criteria cards.

Fig. 2 shows a pictorial representation of the possible variation of each separate weight and the weight vectors calculated using the different algorithms and rules. The values of each specific vector are depicted with a line of the same color and bullets of the same style.

The next step concerns the implementation of a "naïve" MCDA ranking methodology, using the different sets of weights in Table 5, in order to identify possible alternations in the rankings. A combination of the ELECTRE outranking method with PROMETHEE II ranking procedure is implemented, once for each different weighting set. In particular, ELECTRE I is applied, to produce an outranking relation, while PROMETHEE II is applied on this relation, in order to obtain a complete ranking of the alternatives, on the basis of net outranking flows. Concordance thresholds are set to $s = 0.5 + \varepsilon$, where ε is a very small positive number, and simultaneously, no veto thresholds are applied. Table 6 presents the results of this implementation for each weight vector of Table 5, while Table 7 exhibits the extreme ranking positions for each alternative; $P^*(a)$ and $P_*(a)$ represent the highest and the lowest ranking position obtained by each alternative, respectively.

Table 7
Extreme ranking positions of metro lines (case with no white card).

	A	B	C	D	E	F
$P^*(a)$	2nd	1st	1st	1st	1st	2nd
$P_*(a)$	5th	4th	6th	5th	6th	6th

In the second case, with the insertion of two white cards, a set of weight vectors was produced by applying the Max–Min and the Manas-Nedoma algorithms again. The different weight vectors are presented in Table 8, and depicted in Figure 3, along with the weight vector obtained by applying the Simos procedure.

ELECTRE combined with PROMETHEE II ranking procedure is again implemented, once for each weight vector of Table 8. Concordance level is again set to $s = 0.5 + \varepsilon$ without making use of any vetoes. The results of the multiple implementations are presented in Table 9. Table 10 illustrates the extreme ranking positions for each alternative.

5.2.2. Application of the revised Simos methodology

In this paragraph, the revised Simos methodology is implemented, in order to test and measure the robustness of its results. In this case, the most stable instance including the two white cards is examined. The parameter of the revised Simos method has been considered as $z = 6$. Table 11 and Figure 4 present the Max–Min and Manas–Nedoma average weighting sets, 3 vertices of the hyper-polyhedron and the weighting set produced after the application of the Revised Simos procedure.

By applying the ELECTRE method in combination with the PROMETHEE II ranking procedure again, and by setting the concordance level to $s = 0.5 + \varepsilon$ without any vetoes, the different rankings

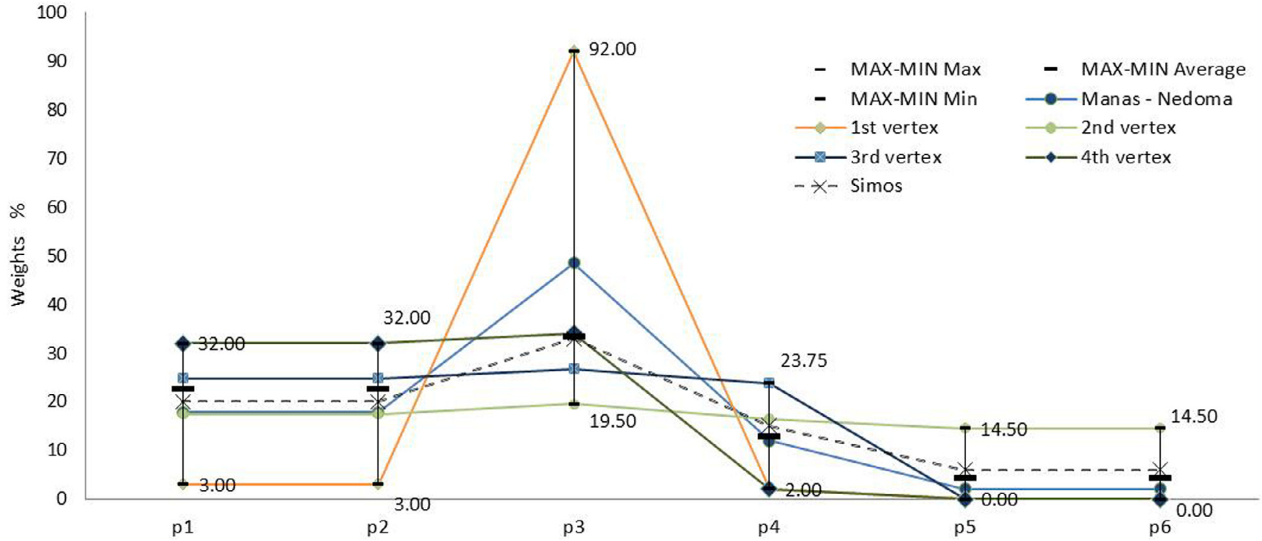


Fig. 3. The calculated range of the weights using different algorithms (2 white cards and original Simos).

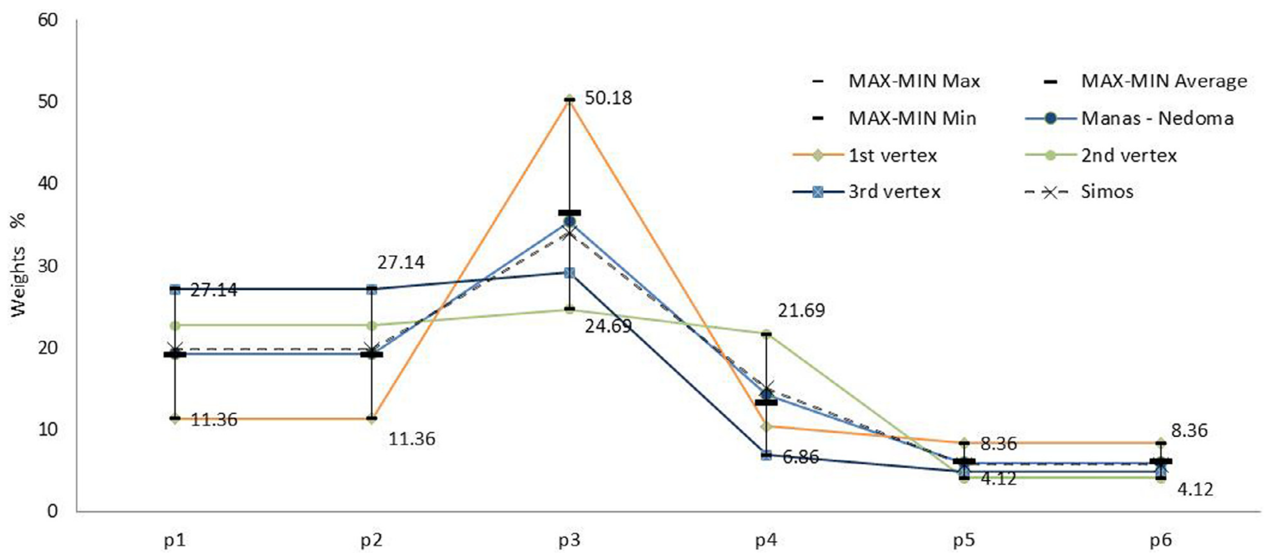


Fig. 4. The calculated range of the weights using different algorithms (2 white cards and revised Simos).

Table 8
The weight vectors of $p^i \in P$ calculated (percent) using different algorithms (2 white cards).

Weighting rule	p_1	p_2	p_3	p_4	p_5	p_6
Simos	20.00	20.00	33.00	15.00	6.00	6.00
Max–Min average	22.58	22.58	33.28	12.88	4.35	4.35
Manas–Nedoma average	17.76	17.76	48.41	11.98	2.05	2.05
1st vertex of the polyhedron	3.00	3.00	92.00	2.00	0.00	0.00
2nd vertex of the polyhedron	17.50	17.50	19.50	16.50	14.50	14.50
3rd vertex of the polyhedron	24.75	24.75	26.75	23.75	0.00	0.00
4th vertex of the polyhedron	32.00	32.00	34.00	2.00	0.00	0.00

Table 9
Ranking of 6 alternatives (case with 2 white cards).

	1st	2nd	3rd	4th	5th	6th
Simos	B,D	–	C	A,F	–	E
Max–Min average	B	D	C	A,F	–	E
Manas–Nedoma average	B,C,D	–	–	A	F	E
1st vertex of the polyhedron	C	D	B	A	F	E
2nd vertex of the polyhedron	B,D	–	C	A,F	–	E
3rd vertex of the polyhedron	B,D	–	C	A,F	–	E
4th vertex of the polyhedron	E	F	A	B	D	C

for each weight vector $\boldsymbol{p} \in \boldsymbol{P}$ of Table 11 are calculated. The results, surfacing after the consecutive implementations of the synergy of the MCDA methods, are presented in Table 12. Table 13 depicts the extreme ranking positions for each alternative, as calculated in these multiple implementations.

Table 14 presents the calculation of the ASI for all the different instances of the metro case study. It is obvious that the ASI improves in the presence of white cards, and especially when the revised Simos procedure is used, instead of the original. The shrinking of the range of weights, when z is introduced, is clearly depicted in Fig. 5.

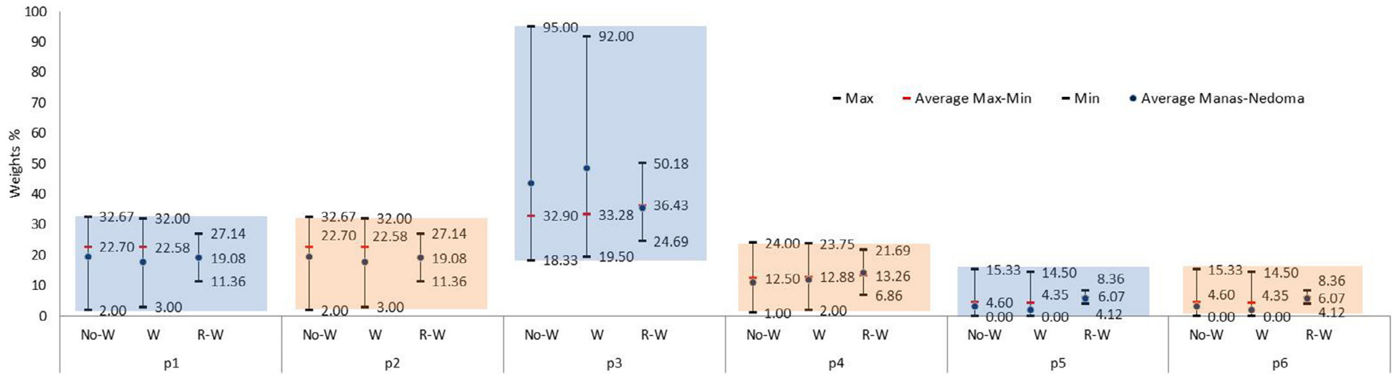


Fig. 5. Range of weights and the Max–Min and Manas–Nedoma representative weighting sets for the Metro case study and its three different cases.

However, again, under no circumstances, the ASI or the visualized criteria weight ranges can be considered as acceptable.

Fig. 5 summarizes and visually compares the possible weight ranges for the three different cases of the Metro case study: original Simos without and with white cards (No-W and W) and revised Simos with white cards (R-W). It also depicts the representative weight vectors, calculated using the Manas–Nedoma and the Max–Min algorithms.

Table 10
Extreme ranking positions of metro lines (case with 2 white cards).

	A	B	C	D	E	F
$P^*(a)$	3rd	1st	1st	1st	1st	2nd
$P_*(a)$	4th	4th	6th	5th	6th	5th

Table 12
Ranking of 6 alternatives using ELECTRE II with PROMETHEE II sorting procedure (case with 2 white cards and revised Simos).

	1st	2nd	3rd	4th	5th	6th
Revised Simos	B,D	–	C	A,F	–	E
Max–Min average	B,D	–	C	A,F	–	E
Manas–Nedoma average	B,D	–	C	A,F	–	E
1st vertex of the polyhedron	C,D	–	B	A,F	–	E
2nd vertex of the polyhedron	B,D	–	C	A,F	–	E
3rd vertex of the polyhedron	E	F	A	B	D	C

Table 13
Extreme ranks (case with 2 white cards and revised Simos).

	A	B	C	D	E	F
$P^*(a)$	3rd	1st	1st	1st	1st	2nd
$P_*(a)$	4th	4th	6th	5th	6th	4th

6. Conclusions

The contribution of this paper lies in the exposure of the robustness issues of both the original and the revised Simos methods, in the elicitation of the criteria importance weights. In fact, the DM is completely unaware of the existence of several sets of weights, which are also compatible with her/his preferences, and have been unjustifiably excluded or neglected during the implementation of the method.

On the other hand, the analyst cannot depend exclusively on the Simos method when it comes to eliciting the preferential information from a DM, in order to calculate the importance weights. When one decides to use the convenience of the Simos procedure, she/he has to apply one or more of the complementary robustness rules and recommendations proposed in this paper. The robustness analysis techniques, proposed in this paper, aid the analysts and DMs to: (i) gain insight on the whole set of weighting solutions, (ii) select a single set of criteria weights, and/or (iii) apply robust rules based on multiple sets of acceptable weights. The visualization of the variation of the weights in particular is an indispensable measure that significantly aids the DMs in limiting the initial feasible weighting space, and eventually, in fixing the criteria weights by themselves.

In case the results are unstable, the decision analyst is required to ask for further preferential information from the DM, through an iterative process. This process ends when the robustness rules produce high levels of stability and the results of the evaluation exhibit no or negligible deviations. The additional preference information may concern information extracted, for instance, through pairwise comparisons between hypothetical alternatives or criteria weights. Alternatively, the decision analyst may prompt the DM to pinpoint the criteria weights, refraining therefore from any further robustness controlling procedures.

The numerical examination of the original Simos method, and the possible weighting results it bears, showed that the weights are almost uncontrollable. On the other hand, the revised Simos procedure, characterized by the additional piece of information requested from the DM, confines to some extent the feasible weighting space, leading therefore to results of higher robustness. However, this amendment is

Table 11
The weight vectors of $p^i \in P$ calculated (percent) using different algorithms (2 white cards and Revised Simos, $z = 6$).

Weighting rule	p_1	p_2	p_3	p_4	p_5	p_6
Revised Simos	19.80	19.80	34.00	15.00	5.70	5.70
Max–Min average	19.08	19.08	36.43	13.26	6.07	6.07
Manas–Nedoma average	19.28	19.28	35.35	14.31	5.89	5.89
1st vertex of the polyhedron	11.36	11.36	50.18	10.36	8.36	8.36
2nd vertex of the polyhedron	22.69	22.69	24.69	21.69	4.12	4.12
3rd vertex of the polyhedron	27.14	27.14	29.14	6.86	4.86	4.86

Table 14
ASI of the different algorithms for the metro case study.

	No white card		2 white cards			
	Manas–Nedoma	Max–Min	Manas–Nedoma (no z)	Max–Min (no z)	Manas–Nedoma ($z = 6$)	Max–Min ($z = 6$)
ASI	0.662	0.659	0.672	0.677	0.844	0.842

not adequate, since the feasible weighting vectors are still unstable to an extent that they still cause significant ranking variations. Specifically, for the case of the Metro lines case study, no conclusion/decision can be drawn, since almost every alternative may get ranked best or worst.

Acknowledgments

The authors wish to acknowledge financial support from the Ministry of Education, Religious Affairs, Culture and Sports of Greece and the European Social Fund, under grant THALIS, MIS 377350: Methodological approaches for studying robustness in multiple criteria decision making projects.

References

Alexopoulos, S., Siskos, Y., Tsotsolas, N., & Hristodoulakis, N. (2012). Evaluating strategic actions for a Greek publishing company. *Operational Research, 12*(2), 253–269.

Aretoulis, G., Kalfakakou, G., & Striagka, F. (2010). Construction material supplier selection under multiple criteria. *Operational Research, 10*(2), 209–230.

Ashari, H. E., & Parsaei, M. (2014). Application of the multi-criteria decision method ELECTRE III for the weapon selection. *Decision Science Letters, 3*(4), 511–522.

Augusto, M., Lisboa, J., Yasin, M., & Figueira, J. R. (2008). Benchmarking in a multiple criteria performance context: An application and a conceptual framework. *European Journal of Operational Research, 184*(1), 244–254.

Azadeh, A., Kor, H., & Hatefi, S. M. (2011). A hybrid genetic algorithm-TOPSIS-computer simulation approach for optimum operator assignment in cellular manufacturing systems. *Journal of the Chinese Institute of Engineers, 34*(1), 57–74.

Bahraminasab, M., & Jahan, A. (2011). Material selection for femoral component of total knee replacement using comprehensive VIKOR. *Materials and Design, 32*(8–9), 4471–4477.

Bana e Costa, C. A., De Corte, J.-M., & Vansnick, J.-C. (2012). MACBETH. *International Journal of Information Technology and Decision Making, 11*(2), 359–387.

Bisdorff, R., Meyer, P., & Veneziano, T. (2014). Elicitation of criteria weights maximising the stability of pairwise outranking statements. *Journal of Multi-Criteria Decision Analysis, 21*(1–2), 113–124.

Bojovic, D., Bonzanigo, L., & Giupponi, C. (2012). Drivers of change in Southern European agriculture: Online participatory approaches for the analysis of planned and autonomous adaptation strategies. In R. Seppelt, A. A. V. S. Lange, & D. Bankamp (Eds.), *2012 International congress on environmental modelling and software. Managing resources of a limited planet* (pp. 684–691).

Bollinger, D., Maystre, L. Y., & Slowinski, R. (1997). Un cas concret de la contribution de l'aide multicritère à la décision à la coordination inter-cantonale pour l'incinération des déchets urbains. In *45e Journées du Groupe de Travail Européen Aide Multicritère à la Décision, Celàkovice, République Tchèque* 20–21 Mars.

Brifcani, N., Day, R., Walker, D., Hughes, S., Ball, K., & Price, D. (2012). A review of cutting-edge techniques for material selection. In *Proceedings of the 2nd international conference on advanced composite materials and technologies for aerospace applications, June 11–13, 2012* http://collections.crest.ac.uk/3975/1/fulltext.pdf Accessed online 16.11.14.

Cai, F. L., Liao, X., & Wang, K. L. (2012). An interactive sorting approach based on the assignment examples of multiple decision makers with different priorities. *Annals of Operations Research., 197*(1), 87–108.

Cavallaro, F. (2009). Multi-criteria decision aid to assess concentrated solar thermal technologies. *Renewable Energy, 34*(7), 1678–1685.

Cavallaro, F. (2010). A comparative assessment of thin-film photovoltaic production processes using the ELECTRE III method. *Energy Policy, 38*(1), 463–474.

Dawson, L., & Schlyter, P. (2012). Less is more: strategic scale site suitability for concentrated solar thermal power in Western Australia. *Energy Policy, 47*, 91–101.

Esseghir, H., & Mellouli, K. (2006). A decision support system for the evaluation of the scientific research structures in Tunisia. In *Proceedings of the ICSSSM'06: 2006 international conference on service systems and service management: 2* (pp. 872–877).

Figueira, J., & Roy, R. (2002). Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. *European Journal of Operational Research, 139*, 317–326.

Figueira, J., Greco, S., & Ehrgott, M. (Eds.). (2005). *State-of-Art of multiple criteria decision analysis*. Dortrecht: Kluwer Academic Publishers.

Figueira, J., Greco, S., & Słowiński, R. (2009). Building a set of additive value functions representing a reference preorder and intensities of preference: GRIP method. *European Journal of Operational Research, 195*(2), 460–486.

Figueira, J. R., Almeida-Dias, J., Matias, S., Roy, B., Carvalho, M. J., & Plancha, C. E. (2011). Electre Tri-C, a multiple criteria decision aiding sorting model applied to assisted reproduction. *International Journal of Medical Informatics, 80*(4), 262–273.

Fontana, M. E., Morais, D. C., & Almeida, A. T. (2011). A MCDM model for urban water conservation strategies adapting Simos procedure for evaluating alternatives intra-criteria. *Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science, 6576*, 564–578.

Greco, S., Mousseau, V., & Słowiński, R. (2008). Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research, 191*(2), 416–436.

Greco, S., Kadziński, M., Mousseau, V., & Słowiński, R. (2011). ELECTRE^{GKMS}: Robust ordinal regression for outranking methods. *European Journal of Operational Research, 214*(10), 118–135.

Greco, S., Siskos, Y., & Slowinski, R. (2013). Entropy measures to control robustness in ordinal regression models. In *Proceedings of the 77th EURO working group on MCDA, Rouen, France*.

Grigoroudis, E., & Siskos, Y. (2010). *Customer satisfaction evaluation*. New York: Springer.

Herssens, C., Jureta, I. J., & Faulkner, S. (2008). Dealing with quality tradeoffs during service selection. In *5th International conference on autonomic computing, ICAC 2008* (pp. 77–86). art. no. 4550829.

Huaylla, F., Berrah, L., & Cliville, V. (2013). A decision-aiding approach for residential photovoltaic system choice: An application to the French context. *Advances in Production Management Systems. Competitive Manufacturing for Innovative Products and Services, 397*, 232–239.

Hugonnard, J. C., & Roy, B. (1983). Le plan d'extension du métro en banlieue Parisienne. In E. Jacquet-Lagrèze, & J. Siskos (Eds.), *Méthode de decision multicritère* (pp. 39–65). Paris: Hommes et Techniques.

Hurson, Ch., & Siskos, Y. (2014). A synergy of multicriteria techniques to assess additive value models. *European Journal of Operational Research, 238*(2), 540–551.

Jabeur, K., Martel, J. M., & Khélifa, S. B. (2004). A distance-based collective preorder integrating the relative importance of the group's members. *Group Decision and Negotiation, 13*(4), 327–349.

Jabeur, K., & Martel, J. M. (2007). An ordinal sorting method for group decision-making. *European Journal of Operational Research, 180*(3), 1272–1289.

Jamali, R., & Nejati, M. (2009). Women's career progression barriers and organisational justice: A study of Iranian society. *Business Strategy Series, 10*(5), 311–328.

Kadzinski, M., Greco, S., & Slowinski, R. (2012). Extreme ranking in robust ordinal regression. *Omega, 40*, 488–501.

Kodikara, P. N., Perera, B. J. C., & Kularathna, M. D. U. P. (2010). Stakeholder preference elicitation and modelling in multi-criteria decision analysis—a case study on urban water supply. *European Journal of Operational Research, 206*(1), 209–220.

Kostoglou, V., Papathanasiou, J., Moschidis, O., & Ladopoulou, P. (2014). A comparative analysis of Greek universities' presence on the World Wide Web using an analytical MCDM methodology. *International Journal of Multicriteria Decision Making, 4*(3), 279–298.

Lahdelma, R., Hokkanen, J., & Salminen, P. (1998). SMAA – Stochastic multiobjective acceptability analysis. *European Journal of Operational Research, 106*(1), 137–143.

Lovasz, L. (1999). Hit-and-run mixes fast. *Mathematical Programming, 86*, 443–461.

Madlener, R., Kowalski, K., & Stagl, S. (2007). New ways for the integrated appraisal of national energy scenarios: The case of renewable energy use in Austria. *Energy Policy, 35*(12), 6060–6074.

Manas, M., & Nedoma, J. (1968). Finding all vertices of a convex polyhedron. *Numerische Mathematik, 14*, 226–229.

Marzouk, M., Hamid, S. A., & El-Said, M. (2014a). A methodology for prioritizing water mains rehabilitation in Egypt. *HBRC Journal* (in press). Available online 13 May 2014, http://dx.doi.org/10.1016/j.hbrcj.2014.03.002.

Marzouk, M., Nouh, A., & El-Said, M. (2014b). Developing green bridge rating system using Simos' procedure. *HBRC Journal, 10*(2), 176–182.

Maystre, L., Pictet, J., & Simos, J. (1994). *Methodes multicriteres ELECTRE – Description, conseils pratiques et cas d'application a la gestion environnementale*. Lausanne: Presses Polytechniques et Universitaires Romandes.

Merad, M. M., Verdel, T., Roy, B., & Kouniali, S. (2004). Use of multi-criteria decision-aids for risk zoning and management of large area subjected to mining-induced hazards. *Tunnelling and Underground Space Technology, 19*(2), 125–138.

Merad, M., Dechy, N., & Marcel, F. (2011). Sustainable development and climate change challenges: Case of a public organization. *Climate: Global change and local adaptation*, 193–208 Chapter 11.

Merad, M., Dechy, N., Serir, L., Grabisch, M., & Marcel, F. (2013a). Using a multi-criteria decision aid methodology to implement sustainable development principles within an organization. *European Journal of Operational Research, 224*(3), 603–613.

Merad, M., Dechy, N., Marcel, F., & Linkov, I. (2013b). Multiple-criteria decision-aiding framework to analyze and assess the governance of sustainability. *Environment Systems and Decisions, 33*(2), 305–321.

Mousseau, V. (1995). Eliciting information concerning the relative importance of criteria. In P. M. Pardalos, Y. Siskos, & C. Zopounidis (Eds.), *Advances in multicriteria analysis* (pp. 17–43). Dortrecht: Kluwer Academic Publishers.

Özcan, T., Çelebi, N., & Esnaf, Ş. (2011). Comparative analysis of multi-criteria decision making methodologies and implementation of a warehouse location selection problem. *Expert Systems with Applications, 38*(8), 9773–9779.

Pekelman, D., & Sen, S. K. (1974). Mathematical programming models for the determination of attribute weights. *Management Science, 20*(8)), 1217–1229.

Pictet, J., & Bollinger, D. (2005). The silent negotiation or how to elicit collective information for group MCDA without excessive discussion. *Journal of Multi-Criteria Decision Analysis, 13*(5–6), 199–211.

Pictet, J., & Bollinger, D. (2008). Extended use of the cards procedure as a simple elicitation technique for MAVT. Application to public procurement in Switzerland. *European Journal of Operational Research, 185*(3), 1300–1307.

Ribas, J. R., & da Silva Rocha, M. (2014). A decision support system for prioritizing investments in an energy efficiency program in favelas in the city of Rio de Janeiro. *Journal of Multi-Criteria Decision Analysis* (In press).

Saaty, T. L. (1994). *The analytic hierarchy process*. New York: McGraw Hill.

Scharlig, A. (1996). *Pratiquer Electre et PROMETHEE – Un complément à décider sur plusieurs critères*. Collection Diriger L'Entreprise, Lausanne: Presses Polytechniques et Universitaires Romandes.

Shanian, A., & Savadogo, O. (2006). TOPSIS multiple-criteria decision support analysis for material selection of metallic bipolar plates for polymer electrolyte fuel cell. *Journal of Power Sources, 159*(2), 1095–1104.

Shanian, A., Milani, A. S., Carson, C., & Abeyarante, R. C. (2008). A new application of ELECTRE III and revised Simos' procedure for group material selection under weighting uncertainty. *Knowledge-Based Systems, 21*(7), 709–720.

Shanian, A., Milani, A. S., Vermaak, N., Bertoldi, K., Scarinci, T., & Gerendas, M. (2012). A combined finite element-multiple criteria optimization approach for materials selection of gas turbine components. *Journal of Applied Mechanics, 79*(6), 061019 Article number.

Simos, J. (1990a). *L'évaluation environnementale: Un processus cognitif négocié*. Lausanne: Thèse de doctorat, DGF EPFL.

Simos, J. (1990b). *Evaluer l'impact sur l'environnement: Une approche originale par l'analyse multicritère et la négociation*. Lausanne: Presses Polytechniques et Universitaires Romandes.

Smith, R. L. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research, 32*, 1296–1308.

Smith, N. A., & Tromble, R. W. (2004). *Sampling uniformly from the unit simplex*. Johns Hopkins University: Department of computer science. Center for Language and Speech processing.

Solymosi, T., & Dombi, J. (1986). A method for determining the weights of criteria: the centralized weights. *European Journal of Operational Research, 22*, 35–41.

Tervonen, T., Valkenhoef, G. V., Baştürk, N., & Postmus, D. (2012). Hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research, 224*, 552–559.

Vansnick, J.-C. (1986). On the problem of weights in multiple criteria decision making (the noncompensatory approach). *European Journal of Operational Research, 24*, 288–294.

Wang, J., & Zionts, S. (2014). Using ordinal data to estimate cardinal values. *Journal of Multi-Criteria Decision Analysis* (in press). doi:10.1002/mcda.1528.

Wilkens, I., & Schmuck, P. (2012). Transdisciplinary evaluation of energy scenarios for a German village using multi-criteria decision analysis. *Sustainability, 4*(4), 604–629.

Xidonas, P., Flamos, A., Koussouris, S., Askounis, D., & Psarras, J. (2007). On the appraisal of consumer credit banking products within the asset quality frame: A multiple criteria application. *Operational Research, 7*(2), 255–283.

Ann Oper Res
DOI 10.1007/s10479-015-1820-x

Environmental corporate responsibility for investments evaluation: an alternative multi-objective programming model

Panos Xidonas · Haris Doukas · George Mavrotas · Olena Pechak

Abstract Current financial and economic crisis, as well as growing environmental pressures put seriously under question traditional development patterns. The need to develop alternative models able to address current economic situation through the exploitation of sustainable patterns is of crucial importance. The innovation of this current study is the incorporation of energy and environmental corporate responsibility (EECR) in decision making, supporting particularly the development of a new model for investment evaluation. A bi-objective programming model is introduced in order to provide the Pareto optimal portfolios (Pareto set) based on the net present value of projects and the EECR score of firms. A systematic decision making approach using Monte Carlo simulation and multi-objective programming is also developed in order to deal with the inherent uncertainty in the objective functions' coefficients. The robustness of the Pareto set as a whole, as well as the robustness of the individual Pareto optimal portfolios is also assessed. The proposed approach facilitates banking organizations and institutions to the selection of firms applying for financial support and credit granting, within the frame of their EECR. Finally, an illustrative real-world application of the proposed model is presented.

Keywords Multi-objective programming · Economic crisis · Corporate social responsibility · Energy and environment · Project portfolio selection · Uncertainty · Robustness

P. Xidonas
ESSCA École de Management, 55 quai Alphonse Le Gallo, 92513 Paris, France

H. Doukas (✉)
Decision Support Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Zografou Campus, 15780 Athens, Greece
e-mail: h_doukas@epu.ntua.gr

G. Mavrotas · O. Pechak
Laboratory of Industrial and Energy Economics, School of Chemical Engineering, National Technical University of Athens, Zografou Campus, 15780 Athens, Greece

Published online: 12 March 2015

1 Introduction

One of the major reasons for economic crises is the irrational distribution of resources. The problem of project selection deals with exactly this kind of problems. Project selection is one of the most common and oldest problems in operations research (OR). Financial organizations often face the problem of selection within a set of projects to fund. Several OR techniques are involved in this kind of problems like e.g. multiple criteria decision analysis (MCDA), mathematical programming (MP). These techniques are widely exploited in relevant decision problems, such as portfolio selection, choice among alternative projects or investment opportunities, student selection, military applications, capacity expansion (see e.g. Golabi et al. 1981; Mavrotas and Rozakis 2009; Salo et al. 2011; Martínez-Costa et al. 2014).

Project portfolio selection problem is defined as the problem of selecting a subset of projects usually based on one or more criteria that have to fulfill specific constraints. In the presence of the imposed constraints (e.g. policy, segmentation) a simple MCDA method does not suffice. Combinatorial character of the problem implies the use of optimization methods aiming at the portfolio of projects that satisfy constraints and achieves the "best" performance. A combination of projects is defined as project portfolio. Usually the "best" performance is expressed emphasizing on economic and financial criteria. Criteria related with the promotion of sustainable practices, fostering green growth, were not taken into consideration in traditional models (Hobbs and Meier 2000).

However, current financial and economic crisis, as well as growing socio-economic and environmental pressures, including climate change, put seriously under question traditional development patterns. The need to develop alternative models able to address current economic situation through the exploitation of sustainable patterns is of crucial importance (Hobbs and Meier 2000; Doukas et al. 2012). Companies are at the heart of the Europe 2020 Strategy, taking into consideration their vital role towards national prosperity and sustainable development (SD). Enterprises have to integrate social and environmental concerns in their business operations and in their interaction with stakeholders on a voluntary basis, within the framework of the corporate social responsibility (CSR) concept.

Companies, more than other stakeholders, have to address the problem in a long term plan, and become a driving force for adoption of relative initiatives towards "green" development and promotion of energy efficiency and environmentally friendly practices, within the CSR framework (Doukas et al. 2013). CSR has been incorporated recently in decision models using Data Envelopment Analysis (Lee and Farzipoor Saen 2012), inventory policy (Barcos et al. 2013) and supply chain (Hsueh 2014) among others. The penetration of energy and environmental policies, as an aspect of CSR is definitely small and CSR does not appear to be a systematic activity in new conditions of European market, a conclusion further confirmed by Apostolakou and Jackson (2009) and Gjølberg (2009a, b) studies. However, relevant works in various fields have been detected recently like e.g. in supplier selection (Hashemi et al. 2014). In this context, new tools and methods are required to foster green entrepreneurship and green energy growth.

The innovation of the current study is the incorporation of energy and environmental corporate responsibility (EECR) in decision making, supporting particularly the development of a new model for investment evaluation. This model can assist financial institutions (green loans) and governmental bodies funding energy—environmental friendly investments. The EECR performance of a firm is considered as an evaluation criterion of the submitted project. Therefore, in our study the drivers of optimization are two objective functions: (1) The net present value (NPV) that represents the economic dimension and characterizes each project and (2) the EECR index that represents the CSR and characterizes each firm that submits the project. In this way, firms with increased EECR are rewarded without ignoring the economic performance of relevant projects.

The resulting multi-objective model (specifically bi-objective) does not provide an optimal portfolio but a set of Pareto optimal portfolios among which the most preferred one is selected by the decision maker (DM). In general, multi-objective optimization increases degrees of freedom within decision making process providing not an optimal solution (as in single objective optimization) but a set of candidate solutions among which the DR chooses. Therefore, the set of Pareto optimal solutions (Pareto set) is essential information in an integrated decision making approach. It must be noted that we deal with multi-objective integer programming (MOIP) models and we can produce the exact Pareto set (i.e. all the Pareto optimal solutions). It is also important to note that, especially the last years, the multi-objective character of project portfolio selection is addressed with multi-objective metaheuristic methods that produce an approximation of the Pareto set (see e.g. Yu et al. 2012; Tavana et al. 2013; Hassanzadeh et al. 2014a).

This work is going one step further, considering also the uncertainty characterizing basic parameters of the model, which are the coefficients of objective functions, namely the NPV of each project and the EECR score of each firm. Given that these values are actually estimations, we follow a systematic approach to deal with the inherent uncertainty. The latter is considered to be of stochastic nature, i.e. we have a probability distribution instead of a crisp number for the values of objective functions' coefficients. It must be noted that a similar approach for project selection problems with multiple criteria that deals with stochastic uncertainty in projects' evaluation is stochastic multiobjective acceptability analysis (SMAA) introduced by Lahdelma et al. (1998). However, SMAA cannot handle the case of multiple constraints that are imposed to the constraints but is used only with independent alternatives in a MCDM context.

The current paper introduces an innovative approach that deals with parameters' uncertainty in a MOIP model and eventually converges to the final Pareto set. It uses the main idea of the iterative trichotomic approach (ITA) (Mavrotas and Pechak 2013a, b). ITA was originally designed for single objective problems of project portfolio selection. It gives information about the degree of certainty for the inclusion or rejection of a specific project in the final portfolio. The version of ITA described in the current paper deals with multi-objective problems of project portfolio selection and provides information about the degree of certainty for inclusion of a specific portfolio in the final Pareto set, expanding thus its application area from project level to portfolio level. This kind of information is essential for the DR to be more confident to select project portfolios that have high degree of certainty regarding their Pareto optimality. In this respect, the DR has a sufficient tool to measure the robustness of the final Pareto set as well as the robustness of specific portfolios that appear in the final Pareto set. Robustness in project portfolio selection has also been addressed in a different way in the works of (Liesiö et al. 2008; Hassanzadeh et al. 2014a, b).

The remainder of the paper is structured as follows: In Sect. 2 the methods, concepts and terminology that will be used in the proposed model are briefly presented, with the focus on adaptation of ITA for the multi-objective case. In Sect. 3 the development of the MOIP model is being elaborated, along with the way the EECR scores are calculated and the relevant constraints. In Sect. 4 the application of the proposed model is presented and the results of multi-objective ITA are discussed, giving emphasis to the kind of additional information that is available to the DR. Finally, in Sect. 5 the main concluding remarks are summarized.

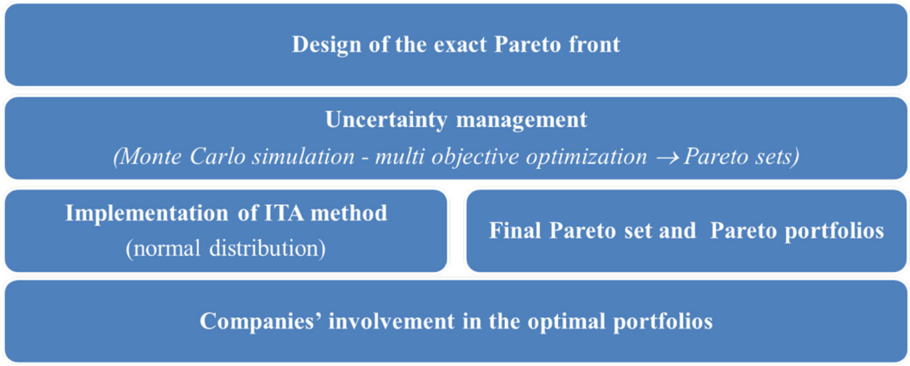


Fig. 1 The adopted procedure

2 Methods, concepts and terminology

The overall procedure that was adopted for the addressed multi-objective project portfolio selection problem is graphically illustrated in Fig. 1.

In the following sections, a more detailed description of the methods deployed will be presented.

2.1 Iterative trichotomic approach (ITA) to multi-objective project portfolio selection problems

The basic idea of current work is to extend the applicability of ITA to the case of multi-objective optimization. ITA was originally designed for project portfolio selection under the framework of MP and more specifically integer programming (IP). It was used with a single objective function reflecting the optimization criterion. The uncertainty associated with objective function coefficients has a stochastic nature (probability distributions instead of crisp numbers).

The term "trichotomy" refers to the separation of a set into three parts. In this context, the proposed decision making process ITA is based on the fact that projects are assigned to one of three groups based on their performance and current level of uncertainty. The latter is incorporated using probability distributions for coefficients of the objective function, which usually express projects' performance. Individual projects' performances are summed up in the objective function, which is the driver of optimization. Monte Carlo simulation is performed using sampling from these distributions. Subsequently with the sampled objective function's coefficients the IP model is solved leading to an optimal portfolio. This pair of sampling and optimization is the core of calculations. The number of Monte Carlo simulations is set to a large number T (e.g. $T = 1000$) which means that the sampling and optimization cycle is performed T times. The output of this process is 1000 optimal portfolios based on the sampling of model's parameters (in this case—projects' performance). Eventually, the set of projects is divided into three subsets (classes): green projects that are present in the final portfolio under all circumstances (i.e. in all Monte Carlo simulations), red projects that are absent from the final portfolio under all circumstances, and grey projects that are included in part of final portfolios. The classification in three subsets is not new in the literature. Liesiö et al. (2007, 2008) used a similar approach in the framework of robust programming. However, the way projects are assigned to each set is different. In addition, Mavrotas and

Rozakis (2009) applied similar concepts in a student selection problem for a post graduate program.

The term "iterative" indicates that the proposed process is developed in a series of computation rounds (or cycles). A predetermined number of computation rounds is defined from the beginning and every round feeds its subsequent until a convergence to the final portfolio is attained. From round to round the uncertainty is reduced for grey projects, and some of them are forced to become either green or red. The uncertainty reduction can be performed either by inclusion of more information or by an automatic uniform narrowing of grey projects' probability distributions.

The concept behind the trichotomic approach is that a DM can focus on projects at stake. The "sure" projects (either in or out of the portfolio) are determined and the DM can shift his attention to "ambiguous" projects (e.g. the grey set). The method provides quantitative and qualitative information that cannot be acquired using e.g. the expected values of distributions. In the latter case, the DM is provided with a unique optimal portfolio or, in other words, which are "go" and "no go" projects, without any discrimination about the degree of certainty for each one of them. On the contrary, in trichotomic approach, DM is provided with fruitful information about certainty degree of each project in the portfolio.

Project portfolio selection is by definition a multi-objective problem. Different points of view should be taken into account. One approach is to aggregate these points of view to a single metric through multicriteria analysis and subsequently optimize the resulting single objective problem where coefficients of objective function are multicriteria scores (Mavrotas et al. 2008). Alternatively, one can use a goal programming approach aggregating the objective functions based on their distance from individual goals (see e.g. Zanakis et al. 1995; Santhanam and Kyparisis 1996).

In the above mentioned approaches, the DR has to assign weights to criteria or goals in order to aggregate them to a single objective function (scalarization). Another approach is to keep individual criteria as separate objective functions and proceed to a multi-objective optimization generating the Pareto set of the problem (or the Pareto front in criteria space). The Pareto set comprises Pareto optimal solutions (or Pareto portfolios in our case). The DR then examines the Pareto front before reaching his final choice. These methods are called "a posteriori" or "generation" methods in the popular Hwang and Masud (1979) terminology for multi-objective optimization methods (first generate Pareto front, examine it, and then select the most preferred Pareto portfolio). Their aim is not just to provide the most preferred solution but also to generate the Pareto set (either exactly or its approximation).

In the current work, we adapt ITA to the multi-objective case. While in original ITA we provide the certainty degree of a specific project to be member of the optimal portfolio given underlying uncertainty, in multi-objective ITA we provide the degree of certainty of a specific portfolio of projects to be member of the Pareto set. A schematic representation of the multi-objective ITA is shown in Fig. 2.

Unlike original ITA, in multi-objective ITA the first iteration has no red set as we don't have any portfolios to be excluded. In the first iteration we have the maximum number of generated portfolios as candidate final Pareto optimal portfolios (POPs). In subsequent iterations some of these portfolios are not present anymore in any Pareto set so they join the red set. As we move from round to round, the uncertainty of parameters (objective functions' coefficients) is reduced (e.g. reduce the standard deviation of a normal probability distribution or shrink the interval of a uniform probability distribution). As we reduce the uncertainty, more portfolios from grey set move to green (appear in all Pareto sets). The red set is implied indirectly by the initially generated portfolios that are not present in any current Pareto set.

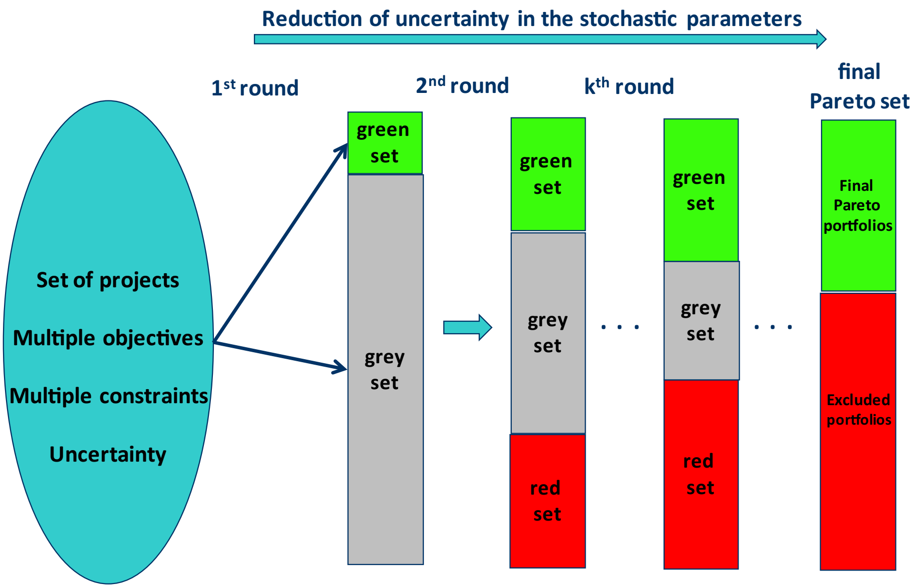


Fig. 2 Graphical representation of multi-objective ITA

In order to describe our model we first present the relevant concepts and terminology, then the mathematical definitions of the robustness measures and then the algorithm that can be used to compute these measures in the framework of ITA.

2.2 Concepts and terminology

We will start with some terminology. The POPs of projects are actually the Pareto optimal solutions of the multi-objective integer problem with binary variables:

$$\begin{aligned} &\max\ Z_1 = \sum_{i=1}^{N} c_{i1} X_i \\ &\ldots \\ &\max\ Z_K = \sum_{i=1}^{N} c_{iK} X_i \\ &st \\ &\mathbf{X} \in S \\ &X_i \in \{0, 1\} \end{aligned} \tag{1}$$

where N is the number of candidate projects, c_{ik} is the objective function coefficient of i-th project in k-th objective function, X_i is a binary decision variable indicating if the i-th project from initial set is selected ($X_i = 1$) or not ($X_i = 0$), and S represents the feasible region formulated by all the imposed constraints. Apart from the usual budget constraints, segmentation and policy constraints, interactions and interdependencies among projects can be also taken into account in the formulation of decision space S (Mavrotas et al. 2003; Liesiö et al. 2007). Eventually, a POP is represented by a vector of "0" and "1" of size N. According to the multi-objective version of ITA method each one of the initial POPs is eventually characterized as red or green as we gradually decrease the uncertainty in model's parameters. The reduction of uncertainty in the model's parameters is performed in *computation rounds*.

In each computation round we solve a great number ($t = 1, \ldots, T$ with e.g. $T = 1000$) of problems like model (1), with different model's parameters, specifically different objective function coefficients using a Monte Carlo simulation approach (see e.g. Vose 1996).

$$
\begin{aligned}
&\max \ Z_1^{(t)} = \sum_{i=1}^{N} c_{i1}^{(t)} X_i \\
&\ldots \\
&\max \ Z_K^{(t)} = \sum_{i=1}^{N} c_{iK}^{(t)} X_i \\
&st \\
&\mathbf{X} \in S \\
&X_i \in \{0, 1\}
\end{aligned}
\tag{2}
$$

where $c_{ik}^{(t)}$ is the objective function coefficient of i-th project in k-th objective function during t-th Monte Carlo iteration. The values of $c_{ik}^{(t)}$ are sampled from the selected probability distributions (uniform, normal, triangular etc). Therefore, in each computation round T Pareto sets ($PS_t, t = 1, \ldots, T$) are produced. The generation of each one of the T Pareto sets is performed using the AUGMECON2 method (Mavrotas and Florios 2013). AUGMECON2 is an improved version of the well known ε-constraint method, especially appropriate for MOIP problems like model (1). It must be noted that AUGMECON2 is capable of generating the exact Pareto set in MOIP problems which means that no Pareto optimal solution is left undiscovered.

Like in original ITA, in each computation round we have three sets where all the POPs p are allocated: The green set (G), the red set (R) and the grey set (Y). The membership relation for each portfolio p in G, R and Y are shown below.

$$
\begin{aligned}
&p \in G : \forall t \in \{1, \ldots, T\}, \quad p \in PS_t \\
&p \in R : \forall t \in \{1, \ldots, T\}, \quad p \notin PS_t \\
&p \in Y : \exists t \in \{1, \ldots, T\}, \quad p \in PS_t
\end{aligned}
\tag{3}
$$

In other words the green set includes the portfolios p that are present in all Pareto sets ($PS_1, \ldots, PS_T$) of the computation round, the red set includes the portfolios that were produced in the initial computational round but are not present in any of T Pareto sets in current computational round and the grey set includes portfolios that are present in some of T Pareto sets. In order to be more specific about the round r that a green, red and grey set refers to we use the notation G_r, R_r and Y_r. To facilitate the decision process, we can define membership thresholds for the green set by relaxing membership requirements. For example, we may set a "green" threshold of 95 % which means that a portfolio is considered to be a member of green set if it is present in the Pareto set for at least 95 % of iterations.

2.3 Robustness measures

Robustness of the POPs in multi-objective ITA is associated with how sure we are about the membership of a specific portfolio in the final (definitive) Pareto set, which is obtained in the last computation round. As uncertainty is reduced going from one computation round to the next, the sooner a POP enters the green set, the more sure we are about its participation

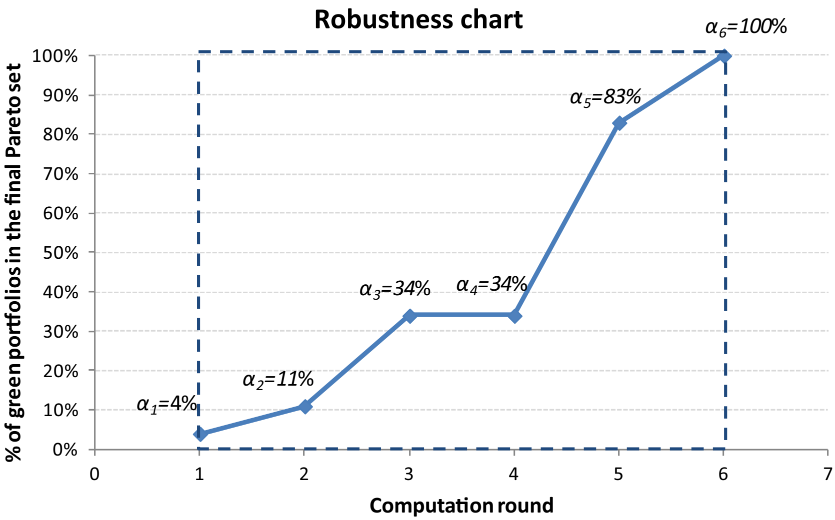


Fig. 3 Example of Robustness Chart with R = 6

in the final portfolio. Therefore, for the POPs, the measure of robustness can be quantified with the *Robustness Degree* for each POP (RD_p) which is defined as follows:

$$RD_p = \frac{R - r_p}{R} \tag{4}$$

where r_p is the computation round that p-th portfolio enters the green set (i.e. becomes member of the final Pareto set) and R the total number of computation rounds. As it is obvious from Eq. (4) Robustness Degree of p-th portfolio varies in $[0, (R-1)/R]$ and the closer it is to 1 the more robust is the specific portfolio.

We have also developed a measure of robustness for the final Pareto set according to how early in the decision process the final POPs are entering the green set. The more green portfolios we have from early rounds (i.e. when we have greater uncertainty), the more robust is the final Pareto set. On the contrary, if the majority of green portfolios is identified in last rounds, it means that the final Pareto set is not so stable.

For the robustness of the final Pareto set we introduce the *Robustness Index (RI)*. In order to calculate the RI we need to draw the so called *Robustness Chart* where the percentages of green portfolios that are available on r-th round (denoted as a_r) are plotted as a function of the computation round. The resulting curve is called *Robustness Curve*. In Fig. 3 we can see an example of a Robustness Chart with the corresponding Robustness Curve. We can observe that from round 2 to round 3 there are no new portfolios added in the green set. This may happen especially when the maximum number of rounds (R) is relatively high.

The RI of the final Pareto set is calculated as the area below the robustness curve, divided by the rectangle area denoted by the dashed rectangular in Fig. 2. The dashed rectangular actually expresses the maximum robustness ($RI = 1$) that occurs when already from the first computation round (i.e. when we have the maximum uncertainty), only one Pareto set is produced from all Monte Carlo iterations. The minimum robustness occurs when all green portfolios are added in the final Pareto set on the last round ($RI = 0$). RI takes values between 0 and 1 and it is calculated using the trapezoid rule for piecewise linear functions according to the following equations:

$$RI = \left(\frac{a_1 + a_2}{2} + \frac{a_2 + a_3}{2} + \cdots + \frac{a_{R-1} + a_R}{2}\right) / (R-1)$$
$$RI = \left[\frac{a_1}{2} + \sum_{r=2}^{R-1} a_r + \frac{a_R}{2}\right] / (R-1)$$
$$RI = \left[\frac{a_1}{2} + \sum_{r=2}^{R-1} a_r + \frac{1}{2}\right] / (R-1) \tag{5}$$

For example, from Fig. 2 we can calculate the corresponding *RI* as:

$$RI = \left[\frac{0.04}{2} + 0.11 + 0.34 + 0.34 + 0.83 + \frac{1}{2}\right] / 5 = 42.8\,\%$$

2.4 The algorithm

As it was mentioned, ITA proceeds with computation rounds (or cycles). The DR initially determines the number R of computation rounds. In the first round, the Monte Carlo sampling is performed using appropriate probability distributions for the uncertain parameters. The results define the green and grey set denoted as G_1 and Y_1. On second round the variance of Y_1 projects' parameters is reduced proportionally to the number of total rounds R. This reduction depends on the form of distribution. For example, for a normal distribution we reduce the standard deviation by $1/(R-1)$, or, for a uniform distribution, we cut $1/(2(R-1))$ from both edges of the range.

The variance reduction follows a uniform pattern across rounds. For example, in case of normal distribution, we reduce the standard deviation *sd* by $1/(R-1)$ after each round. This means that after round r, the reduction of standard deviation is $sd \times (r-1)/(R-1)$. Thus, in the final round projects' parameters (objective function coefficients) are considered as deterministic (have no variance at all). Therefore the final round produces only one Pareto set which is the final Pareto set that comprises the final Pareto portfolios. The flowchart of the decision making process is depicted in Fig. 4.

After the end of the multi-objective—ITA algorithm we have all the information for computing the Robustness Degree of each one of the POPs, for creating the Robustness Chart and computing the RI of the Pareto set. In addition we can provide the DR with informative charts that illustrate the Pareto front with additional information about the robustness of each POP. The latter is explicitly shown in the application in the next section.

3 Model building

This idea of incorporating energy and environmental issues in CSR is rather recent (Doukas and Psarras 2010; Doukas et al. 2012, 2014). In the present application a multi-criteria project portfolio selection problem is addressed taking into account both economic and environmental criteria. Given the uncertainty in quantifying the economic as well as the environmental performance of projects, multi-objective ITA method is an appropriate choice to extract results about the robustness of obtained project portfolios.

As it was mentioned before, the MP model that represents the optimization problem is a MOIP problem with the following characteristics:

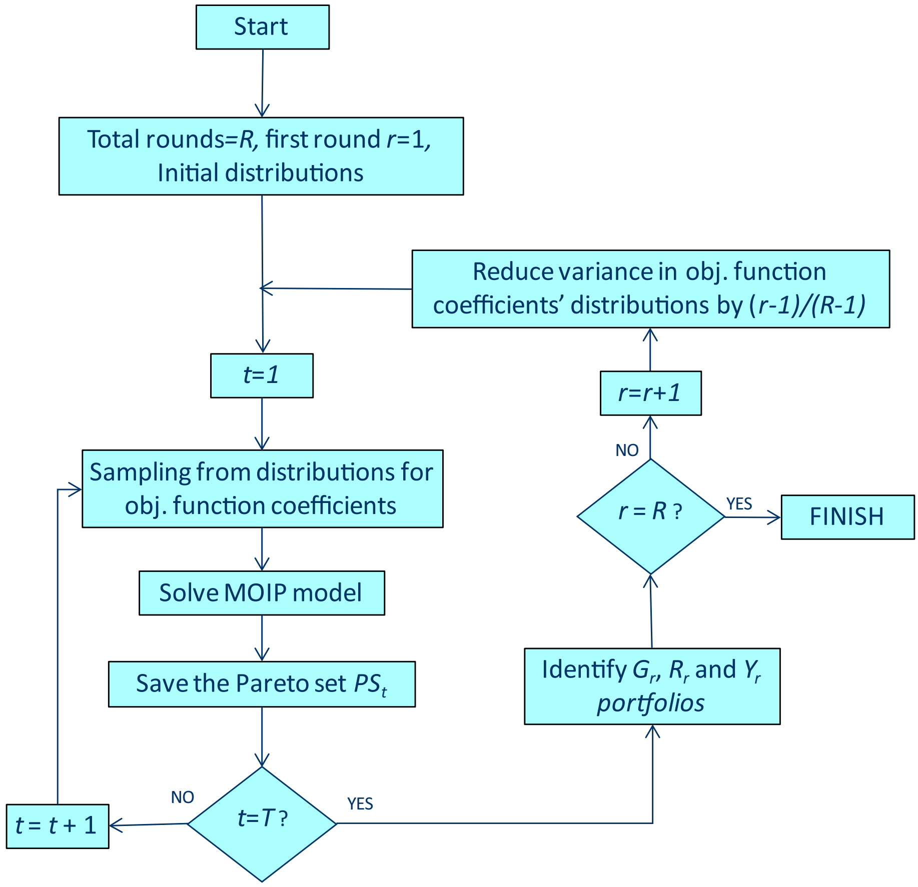


Fig. 4 Illustration of the multi-objective—ITA algorithm

3.1 Decision variables

In the specific case, firms' applications are expressed with 0–1 decision variables, with X_i denoting the i-th firm or application.

More specifically:

- If $X_i = 1$, then the corresponding application is approved.
- Otherwise, if $X_i = 0$, the corresponding application is rejected.

3.2 Objective functions

In the specific model we have two objective functions, namely the NPV of a portfolio and the EECR index of a portfolio. They are both additive functions of individual projects' relevant values.

$$\text{portfolio's } EECR\text{:}\quad max\ Z_1 = \sum_{i=1}^{N} eecr_i X_i$$

$$\text{portfolio's } NPV: \quad max\ Z_2 = \sum_{i=1}^{N} npv_i X_i \tag{6}$$

The parameters npv_i and $eecr_i$ are the NPV of the specific project application and the EECR score of the specific applied company.

3.3 EECR calculation

The adopted procedure used for calculation of the EECR scoring was based upon the ordered weighted average (OWA) operator. According to the literature, OWA operators were introduced by Yager (1988). An aggregation operator is a function $F : I^n \to J$ where I and J are real intervals. I denotes the set of values to be aggregated and J denotes the corresponding result of aggregation. The set of aggregation operators is denoted as $A_n(I, J)$.

An OWA operator is an aggregation operator from $A_n(I, J)$ with an associated vector of weights $w \in [0, 1]^n$, such that:

$$Fw(x) = \sum_{i-1}^{n} w_i \times b_i, \quad \text{where} : \sum_{i-1}^{n} w_i = 1 \tag{7}$$

and b_i denoting the performance of the alternative in the criteria $x_1, \ldots, x_n$.

The criteria to be selected have to be operational, exhaustive in terms of containing all points of view, monotonic and non-redundant since each criterion should be countered only once, as pointed out by Bouyssou (1990). With respect to this, the research focuses on the provision of a small but clearly understood set of evaluation criteria, which can form a sound basis for the comparison of the examined firms in terms of their systematic energy and environmental policy integration as a part of CSR and SD. Concisely, all six criteria are presented in Table 1. The data from these firms were mainly collected from the global reporting initiative disclosure database (GRI 2013).

3.4 Constraints

The model includes constraints, imposed by each banking institution's specific credit policy. First of all, a budget constraint is used in order to secure that the cumulative cost of approved applications does not exceed the overall budget.

$$\sum_{i=1}^{N} \cos t_i \mathrm{X}_i \le avb \tag{8}$$

where avb is the total available budget and $cost_i$ the cost of i-th project application. In the specific application the available budget is 3 M€ while the total cost of all 40 projects is 9.4 M€.

Specific bounds are imposed to control the distribution of projects according to their category, across various sectors. In particular, we don't want a specific project category to dominate in portfolio which is expressed as "no sector or region is allowed to have more than half of the total approved applications". This condition is expressed with the following constraints:

Table 1 The criteria

Criteria	Description
C1: Management commitment	The degree to which Management of a firm prioritizes actions related to the energy and environmental corporate policy, sets specific targets and corresponding time schedule for their accomplishment
C2: Monitoring progress and related impact	The degree to which a firm adopts procedures and protocols for monitoring the set of targets, specific progress made in each related activity and the corresponding impact in companies operation and activation in the market
C3: Participation in dissemination activities	Reflects firms' participation in dissemination activities in broader community, including among others, educational and information activities regarding environmental practices, organization of workshops, conferences and other events, and sponsorships
C4: Promotion of renewable energy	Refers to the firms' involvement for investment in projects and initiatives related to renewable energy sources—wind power, solar power (thermal, photovoltaic and concentrated), hydro-electric power, tidal power, geothermal energy and biomass
C5: Promotion of energy efficiency	The extent to which a firm incorporates initiatives to provide energy-efficient products and services, to reduce direct and indirect energy consumption and other energy conservation practices and technological improvements
C6: Waste and water management	This criterion demonstrates the effort of firms in reducing total water use or discharge and the adoption of waste management activities

$$\sum_{i \in S} X_i \leq 0.5 \times \sum_{i=1}^{N} X_i \quad \text{for } S = \textit{Sector}\ 1, 2, 3, 4 \tag{9}$$

$$\sum_{i \in R} X_i \leq 0.5 \times \sum_{i=1}^{N} X_i \quad \text{for } R = \textit{Region}\ 1, 2, 3, 4 \tag{10}$$

In order to assure that all sectors and regions will be present in final portfolios we also add the following condition: "all sectors and areas will be funded with at least 10 % of the total cost". This condition is expressed with the following constraints:

$$\sum_{i \in S} \cos t_i X_i \geq 0.1 \times \sum_{i=1}^{N} \cos t_i X_i \quad \text{for } S = \textit{Sector}\ 1, 2, 3, 4 \tag{11}$$

$$\sum_{i \in R} \cos t_i X_i \geq 0.1 \times \sum_{i=1}^{N} \cos t_i X_i \quad \text{for } R = \textit{Region}\ 1, 2, 3, 4 \tag{12}$$

In the framework of ITA, the uncertainty characterizing the estimation of projects' NPV as well as the calculation of firm's EECR score is expressed with normal distributions for relevant projects' values. Specifically we take as mean value for the normal distributions the estimated value presented in Table 4 of the appendix and as standard deviation of the initial round the 5 % of the mean. This is done for the NPV as well as the EECR values. From round to round we reduce the standard deviation of corresponding normal distributions to 4, 3, 2, 1 and 0 % in the final round. The whole process (model building, random sampling, Pareto set generation) is implemented within GAMS platform (GAMS 2010).

4 Application and results discussion

For the application we have 40 projects from 40 different firms, with a geographical, sectoral distribution as follows in Table 2:

The parameters' values of the model as well as the membership of projects in various sets (sectoral and geographical) are shown in Table 4 of the Appendix. It must be noted that more types of constraints may be considered in the MP framework like e.g. the specific number (or range) of accepted applications (projects to be finally funded), or constraints for mutually exclusive projects etc.

We performed 1000 Monte Carlo iterations in each computation round and the computation time varied between 7181 and 9150 s from round to round in a core i-5 running at 2.5 GHz. It must be noted that in the specific application, we set a 99 % acceptance threshold for the green set (if a portfolio is present in 99 % of Pareto sets i.e. in 990 out of 1000).

The results of multi-objective ITA are shown in Table 3. There are in total 398 POPs that participate in 1000 Pareto sets of the initial round. Among them only four were present in all Pareto sets. At subsequent iterations we reduce the standard deviation of sampling distributions as shown in the first column of Table 3. Eventually, on the last round we obtain the final Pareto set that comprises 31 POPs of projects. These portfolios contain from 18 to 28 projects.

The additional information that we have from ITA is that we are aware which of these 31 portfolios can be considered more certain than others. The degree of certainty for each portfolio is directly related to the corresponding round that it enters the green set. In Fig. 5 we can see this picture very clear. The darker the portfolio's background the more certain we are about its Pareto optimality. From Fig. 4 we have at a glance which portfolios are more robust given the uncertainty in the model's parameters. The DR can exploit this information in his final selection.

A challenging task is to incorporate the robustness information in the Pareto front. As it is well known, Pareto front of a multi-objective problem is a graph of the Pareto set in criteria space. When we have 2 or 3 objective functions the Pareto front can be easily visualized. The

Table 2 Characteristics of the 40 projects

Geographical regions	Sectors
11 southern European enterprises	11 energy enterprises
10 northern European enterprises	9 industrial enterprises
13 central European enterprises	7 electrical equipment enterprises
6 Greek enterprises	13 enterprises from other sectors

Table 3 The results of multi-objective ITA from round to round

		Computation time (sec)	Green	Red	Grey
σ = 5 %	Round 1	9178	4	0	394
σ = 4 %	Round 2	8247	4	109	285
σ = 3 %	Round 3	8592	5	215	178
σ = 2 %	Round 4	7811	9	275	114
σ = 1 %	Round 5	8685	16	324	54
σ = 0 %	Round 6	7.3*	31	367	0

* For just one iteration as there is no uncertainty quantified by standard deviation

Fig. 5 Coloring code for the 31 portfolios

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	

robustness of each portfolio can be expressed with a bubble chart, where the size of bubble being the portfolio's Robustness Degree (see Sect. 2).

The upper chart in Fig. 5 is the conventional Pareto front with 31 Pareto optimal solutions (different portfolios). The lower chart embodies also robustness information. The robustness information is visualized with the size of the bubble. The greater the Robustness Degree of a POP (i.e. the earlier it enters the green set), the greater the size of the bubble. This kind of information is essential for the DR as he can recognize regions of the Pareto front with higher or lower robustness.

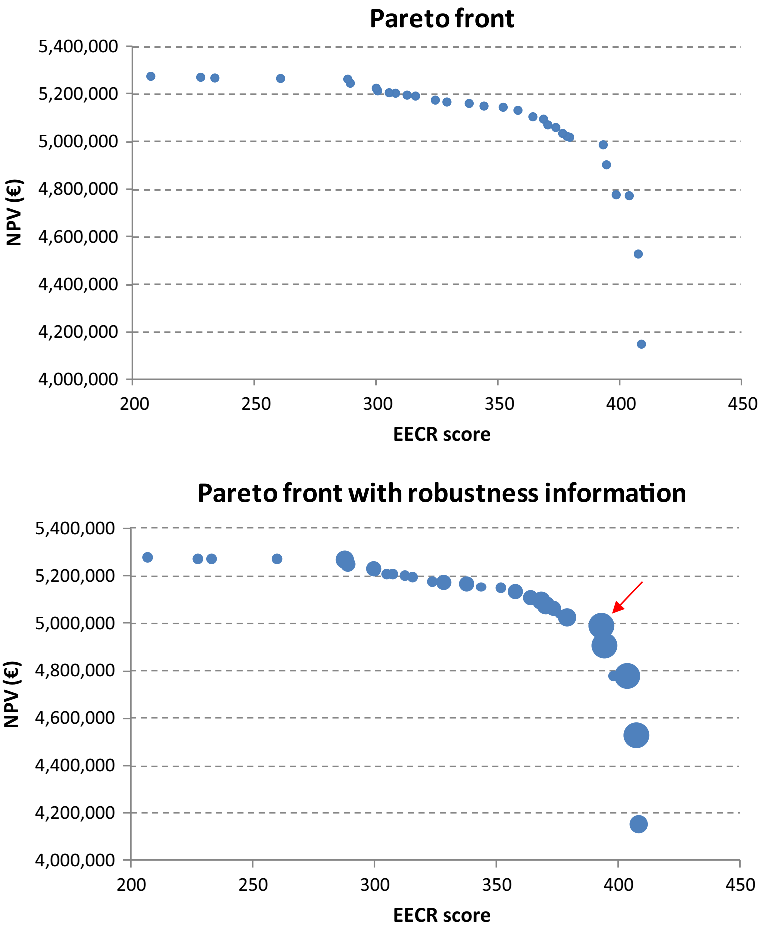


Fig. 6 Visualizing robustness with bubble charts

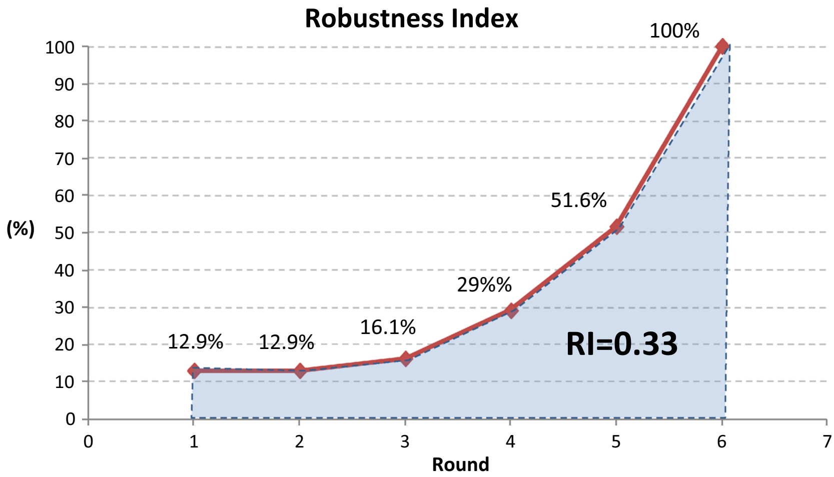


Fig. 7 The Robustness Chart

From this chart the DR can draw conclusions about criteria values of each solution (and therefore assess the trade-off) as well as about the robustness of solutions.

In the specific case, it seems that the robust Pareto optimal solutions are in the region of high EECR (horizontal axis). This also means that the values of EECR have less uncertainty, and this is true, into consideration the detailed and precise way of their calculations.

Promising solutions are on the knee of the Pareto curve where the slope changes sharply meaning that with a little sacrifice in one objective function we can achieve large improvement in the other. A promising solution (portfolio) in our case is the one pointed with an arrow. This means that a small compromise from the maximum EECR value leads to a great improvement in NPV. Besides, as it is evident from the size of the bubble, the specific solution is among the most robust. Conclusively, the robustness of the Pareto optimal solutions which is visualized in Fig. 6 can be regarded as an additional characteristic that helps the DR to evaluate the attractiveness of the obtained POPs.

The overall robustness of the final Pareto set can be measured using the RI. The Robustness Chart and the RI of specific case can be depicted in Fig. 7. Applying Eq. (2) we calculate the RI as the area underneath the Robustness Curve which is $RI = 0.33$.

Regarding all 40 projects, we can measure their presence in the Pareto front by counting how many times each one of them appears in 398 initial Pareto portfolios and how many in times in 31 final Pareto portfolios as shown in Fig. 8.

The initial Pareto portfolios correspond to maximum uncertainty. From Fig. 7 we can extract information about the robustness of the individual projects. The closer they are the two frequency rates (in the initial and in the final Pareto portfolios) for one project, the more robust are the conclusions for the participation frequency of the specific project. From Fig. 7 we can observe that there are projects included in more than 90 % of Pareto portfolios (even when maximum uncertainty is considered, i.e. in the initial round) like projects 7, 11, 13, 20, 21, 24, 35, 38, 39, 40) and other projects that never appear in Pareto portfolios (19, 23, 26, 29, 36, 37).

Moreover, based on the results, it can be noted that companies requesting for larger loans, while having a low EECR index, tend to be rejected. On the other hand, companies asking for smaller loans and having a high NPV index, tend to be approved.

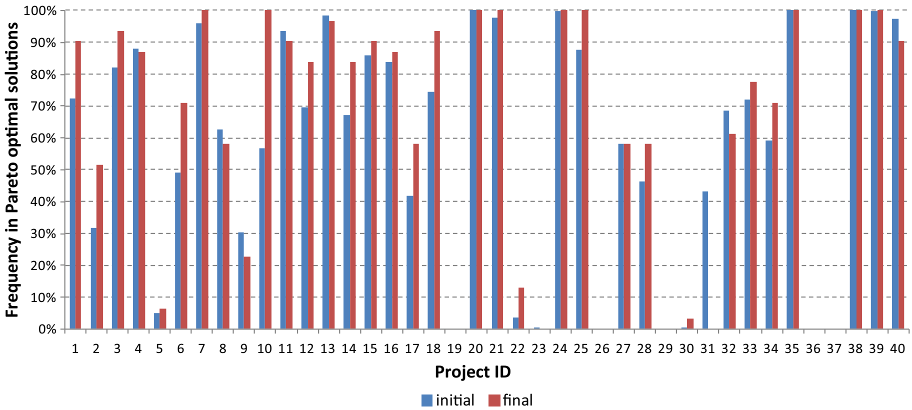


Fig. 8 Frequency of projects in the initial and final Pareto portfolios

5 Conclusions

Project portfolio selection is a challenging problem that sometimes involves multiple objectives and multiple constraints (budget, policy, allocation etc.) that should be satisfied. The combinatorial character of the problem implies the use of discrete optimization methods.

With the proposed methodology, banks and financial institutions do not take into consideration only usual and traditional economic performance in order to finance a project, but also additional ones, such as energy and environmental. The concept of this model can support fruitful decision making towards sustainable transition towards green growth, fostering green corporate responsibility. This is also in accordance with European Commission's objectives to foster firms to report related data in a transparent and explicit way. The proposed decision support model can also enhance the appropriate absorption of Structural and Cohesion Funds, assuring the energy and environmental responsibility of related firms.

In particular, in the presented case, two objective functions represent economic (NPV) and energy and environmental (EECR) dimensions of the submitted projects. A MOIP model is developed with these two objective functions and the exact Pareto set of project portfolios is generated. Moreover, we consider the underlying uncertainty of objective function coefficients (NPV of projects and EECR score of firms). For this reason, a multi-objective version of ITA is introduced so that it can convey useful information to the DR regarding the robustness of eventually obtained Pareto set.

The combination of Monte Carlo simulation and multi-objective programming via the systematic framework of ITA provides us with fruitful insights regarding the robustness of Pareto optimal solutions. The iterative approach gradually converges to the final Pareto set. Useful information emerged from this process is not just the Pareto optimality of project portfolios, but also their robustness in relation to perturbations in objective function coefficients (degree of robustness). Specific measures are developed in order to assess the robustness of the Pareto set as a whole as well as for each Pareto portfolio individually. We also obtain information regarding the specific projects and their frequency in POPs. The hybrid combination of two methodological tools (Monte Carlo simulation and multi-objective optimization) can effectively handle the specific green credit granting problem, where in addition to the consideration of multiple criteria, alternatives must obey to particular policy constraints.

Several issues can be considered for future research. Different probability distributions can be tested for the objective function coefficients. In addition, the underlying uncertainty may be extended to other model parameters beyond the objective function (i.e. to parameters associated with constraints). Moreover, the combination of Monte Carlo simulation and multi-objective optimization is a promising approach that may be used to address the robustness in multi-objective programming problems outside the ITA framework. For future research we can test the method in larger problems and with different probability distributions.

Acknowledgments The authors would like to thank the anonymous referees. This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning". Olena Pechak would like to thank the Hellenic State Scholarship Foundation (IKY) for financial support of her PhD studies.

Appendix

See Table 4.

Table 4 Projects' data

	EECR	NPV (€)	C ost (€)	Sector	Region
1	12.97	2500	5930	S1	R3
2	14.66	49,800	50,830	S1	R3
3	9.76	8300	5000	S1	R2
4	6.23	63,600	33,860	S1	R3
5	6.99	244,600	191,870	S2	R1
6	14.64	36,700	37,500	S2	R1
7	7.10	14,100	6070	S2	R1
8	11.92	22,500	23,030	S2	R4
9	11.81	261,300	190,000	S2	R1
10	21.59	455,000	422,670	S3	R2
11	13.64	696,800	415,000	S3	R1
12	13.59	53,900	39,330	S3	R1
13	3.86	238,900	95,330	S1	R4
14	9.62	3400	5630	S4	R1
15	40.00	600	7370	S4	R1
16	2.95	74,600	37,670	S4	R2
17	25.87	4900	30,100	S1	R4
18	5.25	12,500	5700	S4	R2
19	11.39	389,900	909,310	S4	R3
20	11.67	378,100	160,300	S4	R4
21	15.39	53,100	26,190	S4	R2
22	17.13	51,400	161,010	S4	R3
23	5.76	460,100	353,420	S3	R1
24	8.93	422,800	184,410	S1	R3
25	16.12	146,900	87,910	S4	R2

Table 4 continued

	EECR	NPV (€)	C ost (€)	Sector	Region
26	12.38	477,100	614,620	S1	R2
27	7.19	431,600	277,040	S1	R3
28	21.95	208,500	158,790	S3	R3
29	4.70	324,400	1,410,180	S2	R1
30	18.07	324,100	533,640	S3	R1
31	7.75	603,200	529,130	S4	R2
32	4.54	648,800	396,670	S2	R4
33	19.18	179,600	123,640	S1	R3
34	15.85	220,000	149,770	S1	R1
35	22.01	204,300	93,050	S4	R2
36	4.04	352,100	311,780	S4	R3
37	19.39	223,000	772,970	S3	R2
38	17.81	228,800	117,580	S2	R3
39	12.86	428,500	190,870	S4	R4
40	5.85	516,100	262,030	S2	R1

References

Apostolakou, A., & Jackson, G. (2009). *Corporate social responsibility in Western Europe: An institutional mirror or substitute?*. University of Bath School of Management Working Paper Series, University of Bath, Bath, UK.

Barcos, L., Barroso, A., Surroca, J., & Tribo, J. A. (2013). Corporate social responsibility and inventory policy. *International Journal of Production Economics*, *143*, 580–588.

Bouyssou, D. (1990). Building criteria: A prerequisite for MCDA. In C. A. Bana e Costa (Ed.), *Readings in multiple criteria decision aid* (pp. 58–80). Berlin: Springer.

Doukas, H., Marinakis, V., & Psarras, J. (2012). Greening the Hellenic corporate energy policy: An integrated decision support framework. *International Journal of Green Energy*, *9*(6), 487–502.

Doukas, H., Mavrotas, G., Xidonas, P., & Psarras, J. (2013). Incorporating energy and environmental corporate responsibility in capital budgeting: A multiobjective approach. In *Proceedings of EURO 2013*, Rome, Italy.

Doukas, H., Tsiousi, A., Marinakis, V., & Psarras, J. (2014). Linguistic multi-criteria decision making for energy and environmental corporate policy. *Information Sciences (Special Issue on "New Challenges of Computing with Words in Decision Making")*, *258*, 328–338.

Doukas, H., & Psarras, J. (2010). *Integrated framework for enterprises energy and environmental policies (ENTREE Policies)*. Brussels: Enterprise 2020, A strategic initiative by corporate social responsibility (CSR) Europe.

GAMS Development Corporation. (2010). General Algebraic Modeling System (GAMS) Release 23.5.2. Washington, DC, USA. www.gams.com

Gjølberg, M. (2009a). Measuring the immeasurable? Constructing an index of CSR practices and CSR performance in 20 countries. *Scandinavian Journal of Management*, *25*, 10–22.

Gjølberg, M. (2009b). The origin of corporate social responsibility: Global forces or national legacies? *Socio-Economic Review*, *7*(4), 605–637.

Global Reporting Initiative (GRI) Disclosure Database http://database.globalreporting.org/. Accessed 06 July 2013.

Golabi, K., Kirkwood, C. W., & Sicherman, A. (1981). Selecting a portfolio of solar energy projects using multiattribute preference theory. *Management Science*, *27*, 174–189.

Hashemi, S. H., Karimi, A., & Tavana, M. (2014). An integrated green supplier selection approach with analytic network process and improved Grey relational analysis. *International Journal of Production Economics*. doi:10.1016/j.ijpe.2014.09.027.

Hassanzadeh, F., Nemati, H., & Sun, M. (2014a). Robust optimization for interactive multiobjective programming with imprecise information applied to R&D project portfolio selection. *European Journal of Operational Research*, *238*, 41–53.

Hassanzadeh, F., Modarres, M., Nemati, H., & Amoako-Gyampah, K. (2014b). A robust R&D project portfolio optimization model for pharmaceutical contract research organizations. *International Journal of Production Economics*, *158*, 18–27.

Hobbs, B. F., & Meier, P. (2000). *Energy decisions and the environment: A guide to the use of multicriteria methods*. Dordrecht: Kluwer Academic Publishers.

Hsueh, C.-F. (2014). Improving corporate social responsibility in a supply chain through a new revenue sharing contract. *International Journal of Production Economics*, *151*, 214–222.

Hwang, C. L., & Masud, A. S. (1979). *Multiple objective decision making, methods and applications: A state-of-the-art survey*. Berlin: Springer. 351.

Lahdelma, R., Hokkanen, J., & Salminen, P. (1998). SMAA: Stochastic multiobjective acceptability analysis. *European Journal Operations Research*, *106*, 137–143.

Lee, K.-H., & Farzipoor Saen, R. (2012). Measuring corporate sustainability management: A data envelopment analysis approach. *International Journal of Production Economics*, *140*, 219–226.

Liesiö, J., Mild, P., & Salo, A. (2007). Preference programming for robust portfolio modeling and project selection. *European Journal Operations Research*, *181*, 1488–1505.

Liesiö, J., Mild, P., & Salo, A. (2008). Robust portfolio modeling with incomplete cost information and project interdependencies. *European Journal Operations Research*, *190*, 679–695.

Martínez-Costa, C., Mas-Machuca, M., Benedito, E., & Corominas, A. (2014). A review of mathematical programming models for strategic capacity planning in manufacturing. *International Journal of Production Economics*, *153*, 66–85.

Mavrotas, G., Diakoulaki, D., & Capros, P. (2003). Combined MCDA–IP approach for project selection in the electricity market. *Annals of Operations Research*, *120*, 159–170.

Mavrotas, G., Diakoulaki, D., & Kourentzis, A. (2008). Selection among ranked projects under segmentation, policy and logical constraints. *European Journal Operations Research*, *187*, 177–192.

Mavrotas, G., & Florios, K. (2013). An improved version of the augmented ε-constraint method (AUGMECON2) for finding the exact Pareto set in multi-objective integer programming problems. *Applied Mathematics and Computation*, *219*, 9652–9669.

Mavrotas, G., & Pechak, O. (2013a). The trichotomic approach for dealing with uncertainty in project portfolio selection: Combining MCDA, mathematical programming and Monte Carlo simulation. *International Journal of Multicriteria Decision Making*, *3*(1), 79–97.

Mavrotas, G., & Pechak, O. (2013b). Combining mathematical programming and Monte Carlo simulation to deal with uncertainty in energy project portfolio selection. In F. Cavallaro (Ed.), *Assessment and simulation tools for sustainable energy systems* (pp. 333–356). London: Springer.

Mavrotas, G., & Rozakis, S. (2009). Extensions of the PROMETHEE method to deal with segmentation constraints. *Journal of Decision Systems*, *18*, 203–229.

Salo, A., Keisler, J., & Morton, A. (2011). *Portfolio decision analysis: Improved methods for resource allocation. International series in operations research & management science*. New York: Springer.

Santhanam, R., & Kyparisis, G. J. (1996). A decision model for interdependent information system project selection. *European Journal of Operational Research*, *89*, 380–399.

Tavana, M., Khalili-Damghani, K., & Abtahi, A.-R. (2013). A fuzzy multidimensional multiple-choice knapsack model for project portfolio selection using an evolutionary algorithm. *Annals of Operations Research*, *206*, 449–483.

Vose, D. (1996). *Quantitative risk analysis: A guide to Monte Carlo simulation modelling*. New York: Wiley.

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man, and Cybernetics*, *18*, 183–190.

Yu, L., Wang, S., Wen, F., & Keung Lai, K. (2012). Genetic algorithm-based multi-criteria project portfolio selection. *Annals of Operations Research*, *197*, 71–86.

Zanakis, S. H., Mandakovic, T., Gupta, S. K., Sahay, S., & Hong, S. (1995). A review of program evaluation and fund allocation methods within the service and government sectors. *Socio Economic Planning Sciences*, *29*(1), 59–79.

JOURNAL OF MULTI-CRITERIA DECISION ANALYSIS
J. Multi-Crit. Decis. Anal., 1–n/a (2015)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/mcda.1547

Development of a Robust Multicriteria Classification Model for Monitoring the Postoperative Behaviour of Heart Patients

MICHAEL DOUMPOS[a*], PANAGIOTIS XIDONAS[b], SOTIRIOS XIDONAS[c] and YANNIS SISKOS[d]
[a] *Financial Engineering Laboratory, School of Production Engineering and Management, Technical University of Crete, Chania 73100, Greece*
[b] *ESSCA, École de Management, 55 quai Alphonse Le Gallo, Paris 18534, France*
[c] *Second Department of Cardiology, Division of Cardiac Electrophysiology, Evaggelismos General Hospital, Athens Greece*
[d] *Department of Informatics, University of Piraeus, 80, M. Karaoli & A. Dimitriou St., Piraeus 18534, Greece*

ABSTRACT

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia occurring in 2% of the general population, while the assuming projected incidence in 2050 will rise to 4.3%. This paper presents a multicriteria methodology for the development of a model for monitoring the post-operative behaviour of patients who have received treatment for AF. The model classifies the patients in seven categories according to their relapse risk, on the basis of seven criteria related to the AF type and pathology conditions, the treatment received by the patients and their medical history. The analysis is based on an extension of the UTilités Additives DIScriminantes (UTADIS) method, through the introduction of a two-stage model development procedure that minimizes the number and the magnitude of the misclassifications. The analysis is based on a sample of 116 patients who had pulmonary veins isolation in a Greek public hospital. The classification accuracy of the best fitted models scores between 71% and 84%. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS: multiple criteria analysis; health care; disaggregation analysis; multicriteria classification

1. INTRODUCTION

Atrial fibrillation (AF) is the most common arrhythmia and can be either symptomatic or not. Its prevalence increases with age, and it appears that one in four adults older than 40 years has a lifetime risk of developing AF of approximately 25%. The major mechanism that initiates and perpetuates AF relies on rapid electrical discharges from the pulmonary veins (PV) that return oxygenated blood from the lungs to the left atrium (LA) of the heart. Electrical isolation of the PV with the application of high frequency current across the ostia of the PV is particularly effective for elimination of AF and is widely used in cardiac electrophysiology departments of tertiary hospitals.

However, even after PV isolation (PVI), AF often recurs. Recurrence is classified as early when it takes place 48 h after the operation, late when it occurs within 30 days and very late for cases more than 30 days after the operation. The efficacy of PVI depends on several medical variables, and the assessment of the AF recurrence risk is of major importance in order to decide the most suitable treatment for a patient. Analytic decision models can be particularly useful for defining post-operative AF treatment.

Empirical evidence has shown that medical decision support systems often improve significantly the medical decision process (Garg *et al.*, 2005; Kawamoto *et al.*, 2005), in different activities (e.g. diagnosis, therapy, monitoring and prevention) and contexts such as acute care, primary care and patient advice (Ammenwerth *et al.*, 2013). Data mining, computational intelligence and statistical pattern recognition techniques have been widely used for diagnostic purposes (for an overview, see the work of Hardin and Chhieng,

*Correspondence to: Michael Doumpos, Financial Engineering Laboratory, School of Production Engineering and Management, Technical University of Crete, Chania 73100, Greece. E-mail: mdoumpos@dpem.tuc.gr

Received 19 February 2015
Accepted 25 July 2015

2007). Such methods have also been used in predicting and detecting AF and other forms of cardiac arrhythmia (Alonso-Atienza *et al*., 2012; Chesnokov, 2008; Mohebbi and Ghassemian, 2012), mostly using complex machine learning/data mining models that emphasize the accuracy of the results rather than their inter pretability. However, as noted by Berner and La Lande (2007), many physicians are hesitant to use such systems because the reasoning behind them is not transparent and they are not built on the grounds of knowledge derived from the medical literature, or rules and guidelines issued by clinical associations based on clinical trials' results, registries and experts' consensus (see for instance, the work of Camm *et al*., 2012).

Multicriteria decision aid (MCDA) is well suited in this context, providing a constructive approach for developing medical support systems that combine the physicians' expert judgments with evidence-based clinical practice, in a patient-centred clinical decision-making context (Dolan, 2010). Medical applications of MCDA methods cover, among others, generic computer-aided diagnostic systems (Du Bois *et al*., 1989; Rahimi *et al*., 2007), specialized diagnostic and screening models (Belacel, 2000; Dolan and Frisina, 2002; Goletsis *et al*., 2004), decision aiding in evidence-based medicine (O'Sullivan *et al*., 2014; van Valkenhoef *et al*., 2013), medication risk analysis and appraisal (Goetghebeur *et al*., 2012; Tervonen *et al*., 2011), therapy planning (Hamacher and Küfer, 2002; Schlaefer *et al*., 2013) and and the setting of medical practice guidelines and policy interventions (Angelucci *et al*., 2008; Baltussen *et al*., 2010; Postmus *et al*., 2014).

In this paper, we present a novel MCDA approach for the construction of a decision model that supports the analysis of AF recurrence risk. The model can be used both preoperatively and post-operatively to assess possible options for performing the PVI operation, consult with patients regarding the operation and assess the status of patients after the operation in order to prescribe a proper post-operative pharmacological therapy when needed. The model provides estimates on the AF recurrence risk as well as insights into the factors that contribute to AF recurrence. These factors relate both to the characteristics of patients and the way the PVI operation are performed. The model is expressed in the form of an additive value function, which allows the modelling of nonlinear relationships between the considered factors and the AF recurrence risk, while retaining the interpretability of simpler linear models. The additive form of the model provides both overall risk estimates and the marginal effects due to each separate factor.

The analysis is based on a sample of 116 patients who have undergone PVI operation in a major Greek hospital and have been classified into seven recurrence risk categories according to their post-operative condition. The model is developed through a multicriteria classification approach in the context of disaggregation analysis (Jacquet-Lagrèze and Siskos, 2001) on the basis of the available data. Given the multi-category nature of the problem, a new mixed-integer programming formulation is introduced that takes into account not only the number of misclassifications but also their magnitude. These two model fitting criteria are handled through a lexicographic process, and the robustness of the model is also analysed. The results demonstrate that the proposed MCDA modelling approach can provide not only a useful medical decision aid model but also guidelines and insights into the role of the AF recurrence risk assessment criteria.

The rest of the paper is organized as follows. Section 2 describes the problem context regarding the assessment of AF recurrence risk and the prognostic attributes used in the modelling process. Section 3 is devoted to the proposed multicriteria methodology for constructing the prognostic medical decision model, whereas section 4 presents the application of the methodology and discusses the obtained results. Finally, Section 5 concludes the paper and proposes some future research directions.

2. PROBLEM SETTING

Atrial fibrillation is the most common sustained cardiac arrhythmia occurring in 2% of the general population, while the projected incidence in 2050 will rise to 4.3%. The prevalence of AF increases with age, from <0.5% at 40–50 years, to 5%–15% at 80 years (Kirchhof *et al*., 2012; Wann *et al*., 2011). Men are more often affected than women. The lifetime risk of developing AF is approximately 25% in those who have reached the age of 40 years (You *et al*., 2012).

Atrial fibrillation is characterized electrocardio graphically by low-amplitude baseline oscillations (fibrillator, f-waves that lead to chaotic and irregular atrial rhythm) and an irregular ventricular rhythm, which leads to abnormal contraction and consequently

DOI: 10.1002/mcda

inadequate emptying of both atria of the heart in every cardiac cycle (heart beat). The abnormal and occasionally slow blood flow into the atria in AF patients leads to thrombus formation, thus increasing the risk of stroke, organ ischemia and other acute medical conditions that need hospitalization and have increased mortality risk.

The most common causes for developing AF are excessive alcohol intake, myocardial infarction, pericarditis, myocarditis pulmonary embolism and hyperthyroidism. Other risk factors include congestive heart failure, aortic and mitral valve disease, left atrial enlargement, obstructive sleep apnea and advanced age. On the other hand, the effects of AF in the cardiovascular system have been well studied; it doubles the risk of mortality, triples the risk for hospitalization and increases the risk of stroke nearly five times. Overall, AF promotes heart failure, and heart failure aggravates AF to worsen patients' overall prognosis.

Atrial fibrillation that terminates spontaneously within 7 days is termed paroxysmal, and AF of more than seven continuous days is called persistent. AF persistent for more than 1 year is termed longstanding, whereas longstanding AF refractory to electrical cardioversion is called permanent.

Depending on the characteristics of AF, its treatment can be based on pharmacological rate and rhythm control strategies. Left atrial catheter ablation is another option for long-term management involving patients who remain symptomatic despite other treatments. Catheter ablation is an electrophys iological operation during which multiple endocardial lesions are created by the multiple applications of high frequency current created by an external generator through ablation catheters (Figure 1). The aim of the operation is to isolate electrically the ostia (entrances) of the PV that return oxygenated blood from the lungs into the LA with the use of fluoroscopy and an electroanatomic mapping system for the navigation of ablation catheters in the heart.

The efficacy of AF ablation (PVI) varies widely depending mainly on medical variables like the type of AF, duration of AF, duration of the last AF episode, diameter and volume of the LA, the number of applications of high frequency current and the time of fluoroscopy. It is very important for electrophysiologists to choose the right patients, that is, with certain values of medical variables related to AF prior to PVI, who are more likely to benefit from the operation and remain free from arrhythmia for as long as possible, taking also into consideration the risk of adverse events due to the operation (which is estimated to by about 1%–3%).

In this context, this study employs a multicriteria methodology to determine the risk of AF recurrence. The analysis is based on a sample of 116 patients

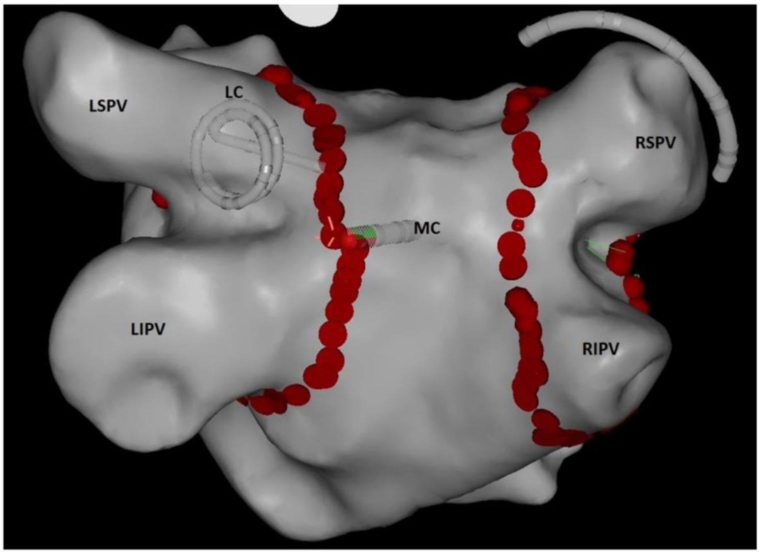


Figure 1. Postero-anterior view of left atrium (endocardial lesions—red dots—created by application of high frequency current through the irrigated tip ablation mapping catheter (MC); lasso catheter (LC) records endocardial potentials at the ostia of left superior pulmonary vein (LSPV), left inferior pulmonary vein (LIPV), right superior pulmonary vein (RSPV) and right inferior pulmonary vein (RIPV)).

DOI: 10.1002/mcda

who had PVI in a Greek public hospital. The condition of the patients was monitored after the operation, and AF recurrence was characterized as 'early' for cases in which it occurred during the first 48 h post-operatively, 'late' when in the first month and 'very late' if it occurred more than a month after the operation. Thus, the patients were classified into seven ordinal recurrence risk categories, ranging from high risk cases (class YYY), corresponding to patients for whom AF recurrence occurred at all three-time windows (early, late and very late), to patients for which PVI was successful as no recurrence occurred (class NNN).[1] The risk order of the classes was defined in cooperation with a cardiologist with experience on the treatment of AF and the PVI operation. Table I illustrates the definition of the recurrence risk categories and the number of sample patients in each class. It should be noted that early recurrence is often observed without any future complications whereas late or very late recurrence is more likely to be associated with cases that may require additional treatment. Thus, patients in category YNN are considered to be of lower risk than those in category NNY. Furthermore, patients with late recurrence are more likely to require additional treatment compared with patients with no late recurrence. This is why the top three risk categories (YYY, NYY and NYN) all correspond to patients with late recurrence.

The assessment of the AF recurrence risk is based on the seven prognostic criteria noted in the succeeding texts, which have been selected in cooperation with an expert medical decision-maker and existing medical guidelines on the risk factors of AF and its treatment (Camm *et al*., 2012; Kirchhof *et al*., 2012). In particular, the assessment criteria involve the following measures:

Table I. Definition of the atrial fibrillation recurrence risk categories

Recurrence period				
Early	Late	Very late	Class labels	No. of cases
Yes	Yes	Yes	YYY	8
No	Yes	Yes	NYY	8
No	Yes	No	NYN	9
Yes	No	Yes	YNY	2
No	No	Yes	NNY	12
Yes	No	No	YNN	3
No	No	No	NNN	74

- AF type (paroxysmal, persistent and permanent): ordinal criterion, such that a permanent type is associated with higher recurrence risk whereas a paroxysmal type is associated to lower risk.
- Duration of AF (number of years since first episode of AF): positively associated to recurrence risk (i.e. the larger the duration of AF, the higher the risk).
- Duration of the last AF episode (in days): positively associated to recurrence risk.
- LA diameter (measured in two-dimension echocardiogram in millimetre): positively associated to recurrence risk.
- LA volume (calculated automatically by echocardiograph software using longitudinal and transverse dimensions in cubic centimetre): positively associated to recurrence risk.
- Number of applications of high frequency current (each application lasts 60 s): negatively associated to recurrence risk (i.e. the risk of recurrence decreases with the number of applications of high frequency current).
- Time of fluoroscopy (duration of fluoroscopy used in order to visualize catheters and navigate them across cardiac chambers and PV in minutes): positively associated to recurrence risk.

The aforementioned assessment criteria combine attributes about the nature of the arrhythmia in each patient as well as attributes that are related to the PVI operation. The combination of such factors in an aggregate recurrence risk assessment model can support medical doctors in a number of ways, both preoperatively and post-operatively. First, it allows them to assess the risk of recurrence preoperatively based on the cardiovascular diagnostic characteristics of patients and the parameters that define how the PVI operation can be performed. In that respect, a model combining such factors can guide medical doctors to differentiate the ablation strategy during operation (i.e. increase the number of applications of high frequency current or make additional lesion lines in patients with longstanding AF and dilated LA) and provide patients with personalized estimated success rate during preoperative consultation. Furthermore, through such a model, cardiologists can change the type and duration of post-operative pharmacological therapy (i.e. more potent antiarrhythmic drugs in high risk patients).

First, it allows them to assess (post-operatively) additional treatments that may improve the condition of the patients and minimize the AF recurrence risk. Furthermore, through such a model, medical doctors

DOI: 10.1002/mcda

can analyse the trade-offs between the factors that define the nature of the PVI operation (applications of high frequency current and fluoroscopy time) while controlling for the characteristics of the AF for each patient. This allows doctors to decide on make informed decisions about the best way to perform the operation in order to minimize the recurrence risk.

3. MULTICRITERIA METHODOLOGY

In the context of the problem setting described in the previous section, the development of a decision model that facilitates the monitoring of the patients can be considered as a multicriteria classification problem. Multicriteria classification problems have received much interest among MCDA researchers over the past couple of decades, and several modelling approaches have been developed (Zopounidis and Doumpos, 2002). In this study, we employ an additive value function model. In particular, denoting by $\mathbf{x}_i = (x_{i1}, \ldots, x_{in})$ the vector with the available data for patient i on a set of n recurrence risk attributes, the patient's overall recurrence risk is assessed with the following additive function:

$$V(\mathbf{x}_i) = \sum_{j=1}^{n} w_j v_j(x_{ij}), \text{ with } \sum_{j=1}^{n} w_j = 1, \qquad (1)$$

where w_j is the (non-negative) weight for criterion j (the weights represent the trade-offs the decision-maker is willing to make among the criteria) and $v_j(\cdot)$ is the marginal value function for criterion j, normalized in [0, 1]. The additive model is well founded from a theoretical point of view (Keeney and Raiffa, 1993) and has been used in a wide range of multicriteria evaluation problems. The additive form of the model makes it easy to use and comprehend. The comprehensibility of the model is an important feature that greatly helps medical doctors to understand the model's logic, thus improving the practical usefulness of the model. More complex modelling forms (e.g. a multi-linear value function) take into account interactions between the decision criteria at the expense of yielding models, which are difficult to construct and understand.

In the modelling setting followed in this study, it is assumed that the higher the global value $V(\mathbf{x}_i)$ of patient i, the higher is his/her recurrence risk. Thus, with the additive model (1), a patient i is classified into risk group k if and only if $t_k < V(\mathbf{x}_i) < t_{k-1}$, where $t_0 > 1 > t_1 > t_2 > \cdots > t_{q-1} > t_q > 0$ is a set of thresholds that distinguish between the q recurrence risk categories $C_1, \ldots, C_q$ (e.g. $q = 7$ for the sample used in this study).[2] In accordance with the aforementioned interpretation of the additive value model, the categories are risk-ordered such that C_1 corresponds to high risk patients (i.e. category YYY in Table I) and C_q to low risk ones (category NNN in Table I).

The construction of the additive model and the estimation of the separating thresholds are performed using a preference disaggregation approach (Jacquet-Lagrèze and Siskos, 2001), namely, the UTADIS II method (Doumpos and Zopounidis, 2002), which adapts the framework of the UTilités Additives (UTA) method (Jacquet-Lagrèze and Siskos, 1982) to classification problems. In this context, the evaluation model (1) is fitted on a set of data (reference set) for m patients already classified in q recurrence risk categories. The objective of the model-fitting process is to construct a decision model that is as compatible as possible with the predefined classification of the patients in the reference set. The constructed model can then be calibrated (if needed) through an interactive process with the medical decision-maker and then used to evaluate the risk for patients in a real time setting. In the UTADIS II approach, the fitting of the model is based on the solution of the following mixed-integer programme (MIP):

$$\begin{aligned}
&\min \frac{1}{q}\sum_{k=1}^{q} \frac{1}{m_k} \sum_{\forall i \in C_k} \left(\sigma_i^+ + \sigma_i^-\right) && \\
&\text{s.t. } V(\mathbf{x}_i) - t_k + \sigma_i^+ \geq \delta && \forall\, i \in C_k\ (k = 1, \ldots, q-1) \\
&\quad V(\mathbf{x}_i) - t_{k-1} - \sigma_i^- \leq -\delta && \forall\, i \in C_k\ (k = 2, \ldots, q), \\
&\quad V(\mathbf{x}_*) = 0,\ V(\mathbf{x}^*) = 1 && \\
&\quad t_{k-1} - t_k \geq \varepsilon && k = 1, \ldots, q-1 \\
&\quad \sigma_i^+, \sigma_i^- \in \{0, 1\} && i = 1, \ldots, m,
\end{aligned} \qquad (2)$$

where m_k denotes the number of patients in the reference set from category C_k whereas σ_i^+ and σ_i^- are binary slack variables associated with patients misclassified by the additive model. In particular, σ_i^+ equals one if a patient is misclassified in a lower risk category compared to the one he/she actually belongs to (i.e. when the model underestimates the actual recurrence risk), whereas σ_i^- denotes the misclassification into higher risk classes (i.e. overestimation of risk). The first two constraints define these error variables on the basis of the threshold-based classification rule. In both constraints, δ is a small user-defined positive constant used to handle ambiguous classification results, which arise when the risk score of a patient equals one of the

DOI: 10.1002/mcda

classification thresholds (in the analysis, we set $\delta = 0.0001$). The third set of constraints normalizes the additive model in [0, 1], such that a low risk patient (denoted by $\mathbf{x}_*$) is assigned a risk score of 0, whereas a patient with the highest risk (denoted by $\mathbf{x}^*$) is assigned the maximum risk score of 1. These two extremes ($\mathbf{x}_*$ and $\mathbf{x}^*$) can either be defined through medical expertise or through the data used in the analysis. In this study, we followed the latter approach, defining $\mathbf{x}_*$ and $\mathbf{x}^*$ by the minimum and maximum levels, respectively, of the criteria described in the previous section (except for the number of applications of high frequency current, which is negatively related to recurrent risk; for this criterion, $\mathbf{x}^*$ was defined by the minimum level of the criterion and $\mathbf{x}_*$ by its maximum). Thus, a high risk patient ($\mathbf{x}^*$) has permanent AF, large durations, large LA diameter/volume, small number of applications of high frequency current during the PVI operation and large fluoroscopy time.

Finally, the fourth constraint of problem (2) defines the minimum difference between two consecutive classification thresholds, with ε being a user-defined positive constant (in this study, we used $\varepsilon = 0.02$). The objective function of problem (2) minimizes the total weighted classification error for the patients in the reference set. The weighting of the errors for each patient i from risk class C_k by $1/m_k$ imposes a balance among all risk categories, thus ensuring that the classifications of the fitted model will not be biased towards classes with a large number of patients. The aforementioned optimization problem can be formulated as a linear MIP, through the modelling of the marginal value functions of the additive model (1) as piecewise linear functions of the data (for details, see Doumpos and Zopounidis, 2002; Jacquet-Lagreze and Siskos, 1982).

Even though problem (2) is easy to solve for medium-size reference sets (with existing powerful MIP solvers), it fails to distinguish between the magnitude of the classification errors, which is an important issue in multi-category ordinal classification problems such as the one considered in this study. In such cases, instead of using the total number of misclassifications as the modelling fitting criterion, the mean absolute error is a more meaningful objective. Imposing weights to account for the imbalances in the number of patients in each risk category in the reference set, the mean-weighted absolute error (MWAE) is defined as follows:

$$\frac{1}{q}\sum_{k=1}^{q}\frac{1}{m_k}\sum_{\forall i \in C_k}|\hat{y}_i - y_i|, \tag{3}$$

where $y_i = \{1, 2, \ldots, q\}$ is the actual risk category for patient i and $\hat{y}_i$ is classification of the patient by the decision model. The construction of an additive value function model that optimizes this fitting measure for a given reference set can be performed with the following MIP formulation:

$$\begin{array}{lll}
\min & \displaystyle\sum_{k=1}^{q}\frac{1}{m_k}\sum_{\forall i \in C_k}\sum_{\ell=1}^{q}\left(\xi_{i\ell}^{+} + \xi_{i\ell}^{-}\right) & \\
\text{s.t.} & V(\mathbf{x}_i) - t_k + \xi_{ik}^{+} \geq \delta & \forall i \in \{C_1, \ldots, C_k\},\ k = 1, \ldots, q-1 \\
 & V(\mathbf{x}_i) - t_{k-1} - \xi_{ik}^{-} \leq -\delta & \forall i \in \{C_k, \ldots, C_q\},\ k = 2, \ldots, q \\
 & t_k - t_{k-1} \geq \varepsilon & k = 1, \ldots, q-1 \\
 & V(\mathbf{x}_*) = 0,\ V(\mathbf{x}^*) = 1 & \\
 & \xi_{ik}^{+}, \xi_{ik}^{-} \in \{0, 1\} & i = 1, \ldots, m,\ k = 1, \ldots, q
\end{array} \quad . \tag{4}$$

Compared to model (2), this formulation distinguishes between the possible misclassifications for a patient i from risk category C_k through the binary error variables ξ^{+} and ξ^{-}. More specifically, the first constraint compares the global value of every patient belonging in the set of risk categories $\{C_1, C_2, \ldots,$

DOI: 10.1002/mcda

$C_k\}$ (for each, $k=1, \ldots, q-1$), against the lower threshold t_k of category C_k. For instance, a patient from the high risk category C_1 (i.e. $k=1$) is compared (successively) against t_1 (the lower threshold of category C_1) and t_2 (the lower threshold of category C_2), up to t_{q-1} (the lower threshold of category C_{q-1}). Each of these comparisons is associated with a different error variable $\xi^+_{i\ell}$ ($\ell=k, \ldots, q-1$), which equals to 1 if and only if a patient i that actually belongs to the set of risk categories $\{C_1, C_2, \ldots, C_k\}$ is misclassified in any of the risk categories $\{C_{\ell+1}, C_{\ell+2}, \ldots, C_q\}$. For example, if a patient from the risk category C_1 is assigned into category C_3, then $\xi^+_{i1} = \xi^+_{i2} = 1$, thus indicating that there is a two-notch difference between the actual and the estimated classification.

In a similar manner, the second constraint compares the global value of every patient from the set of risk categories $\{C_k, C_{k+1}, \ldots, C_q\}$ (for each, $k=2, \ldots, q$), against the upper threshold t_{k-1} of category C_k. For instance, a patient from the low risk category C_q is compared (successively) against t_{q-1} (the upper threshold of category C_q) and t_{q-2} (the upper threshold of category C_{q-1}), up to t_2 (the upper threshold of category C_2). These comparisons are associated with error variables $\xi^-_{i\ell}$ ($\ell=2, \ldots, k$), which equal to 1 if and only if a patient i that actually belongs to the set of risk categories $\{C_k, C_{k+1}, \ldots, C_q\}$ is assigned (misclassified) into any of the risk categories $\{C_1, C_2, \ldots, C_{\ell-1}\}$.

Thus, the sum $\xi^+_{ik} + \xi^+_{i,k+1} + \cdots + \xi^+_{i,q-1}$ for a patient from risk category C_k equals the difference $\hat{y}_i - y_i$ (as in the example noted previously), when the patient is assigned into a lower risk category compared to its actual risk level (i.e. $\hat{y}_i > y_i$), whereas the sum $\xi^-_{i2} + \cdots + \xi^-_{ik}$ equals the difference $y_i - \hat{y}_i$, when the patient is assigned into a higher risk category compared to its actual risk level (i.e. $\hat{y}_i < y_i$). Obviously, the ordinal definition of the risk categories implies that $\xi^+_{i\ell} = 1$ whenever $\xi^+_{i,\ell+1} = 1$ and $\xi^-_{i\ell} = 1$ whenever $\xi^-_{i,\ell-1} = 1$.

4. RESULTS

4.1. Empirical setting

In this study, the two model fitting formulations described in the previous section are employed in a lexicography manner. In particular, model (2) is first used to obtain an additive evaluation model that minimizes the total weighted number of misclassifications while ignoring their magnitude. Then, at a second stage, problem (4) is solved to minimize the MWAE while controlling for the number of misclassified patients on the basis of the solution of model (2), that is, by adding the following constraint to problem (4):

$$\sum_{\forall i \in C_k} \left(\xi^{\pm}_{ik} + \xi^{\pm}_{i,k-1} \right) = E^*, \qquad (5)$$

where $E^* = \sum_i \left(\sigma^+_i + \sigma^-_i \right)$ is the total number of misclassifications corresponding to the solution of model (2).

All optimization problems are solved with a quad-core personal computer with an Intel i7-2600 K/3.4 GHz processor and 16 GB of RAM, using the Gurobi 6 solver. With this computational environment, the mixed integer linear programming formulation of UTADIS II was easily solved to optimality, whereas problem (4) was much more challenging due to its increased complexity. In that respect, a time limit of 1 h was imposed during the solution process.

Table II. Model fitting metrics

	UTADIS II	MWAE	MWAE-Lex	WOLR
Overall classification accuracy	0.724	0.569	0.724	0.345
Average classification accuracy	0.843	0.691	0.793	0.244
Mean-weighted absolute error	0.679	0.530	0.588	1.627

MWAE, mean-weighted absolute error; WOLR, weighted ordinal logistic regression model.

4.2. Analysis of results

Table II presents some main model fitting measures for three different additive evaluation models, including the model resulting from the solution of the UTADIS II problem (2), the one obtained from the MWAE problem (4), the MWAE-Lex model obtained from the combination of the previous two approaches through the aforementioned lexicographic scheme and a weighted ordinal logistic regression model (King and Zeng, 2001). For each evaluation model, three fitting indices are calculated, namely: (i) the overall classification accuracy, defined as the percentage of patients correctly classified by the

DOI: 10.1002/mcda

Table III. Classification matrices for the UTADIS II and MWAE-Lex models (all entries in %)

			Model's classification						
			YYY	NYY	NYN	YNY	NNY	YNN	NNN
Actual classification	UTADIS II	YYY	100.0	0.0	0.0	0.0	0.0	0.0	0.0
		NYY	0.0	100.0	0.0	0.0	0.0	0.0	0.0
		NYN	0.0	0.0	100.0	0.0	0.0	0.0	0.0
		YNY	0.0	0.0	0.0	100.0	0.0	0.0	0.0
		NNY	33.3	8.3	0.0	0.0	58.3	0.0	0.0
		YNN	33.3	0.0	0.0	0.0	0.0	66.7	0.0
		NNN	4.1	17.6	5.4	2.7	2.7	2.7	64.9
	MWAE-Lex	YYY	100.0	0.0	0.0	0.0	0.0	0.0	0.0
		NYY	12.5	50.0	12.5	0.0	12.5	12.5	0.0
		NYN	0.0	0.0	66.7	0.0	11.1	0.0	22.2
		YNY	0.0	0.0	0.0	100.0	0.0	0.0	0.0
		NNY	16.7	0.0	8.3	0.0	66.7	0.0	8.3
		YNN	0.0	0.0	0.0	0.0	0.0	100.0	0.0
		NNN	2.7	1.4	6.8	8.1	4.1	5.4	71.6

model; (ii) the average classification accuracy, defined by the objective function of problem (2) and (iii) the MWAE index (3).

The basic UTADIS II performs best in terms of the average classification accuracy, which corresponds to the objective function of the MIP (2), whereas the MWAE model minimizes the MWAE on the basis of the optimization problem (4). The performance of the MWAE model, however, on the two classification accuracy criteria is significantly lower compared to the results of UTADIS II. The MWAE-Lex approach provides a good balance between the two other models. In particular, compared to UTADIS II, MWAE-Lex has slightly lower average classification accuracy (by about 6% in relative terms) while improving the weighted absolute error by about 13% (again in relative terms). On the other hand, compared to MWAE, the MWAE-Lex model has a bit higher weighted absolute error but yields much higher classification accuracies. Finally, the ordinal logistic regression model performs consistently worst than all multicriteria models.

Table IV. Weights of the criteria in the decision models developed with UTADIS II and the lexicographic approach

	UTADIS II	MWAE-Lex
Type of AF	0.00	2.04
AF duration	14.29	10.70
AF episode duration	16.32	21.32
LA diameter	18.41	12.48
LA volume	12.33	13.03
Applications	12.19	19.21
Fluoroscopy time	26.46	21.21

MWAE, mean-weighted absolute error; AF, atrial fibrillation; LA, left atrium.

The detailed classification matrices for the results of the additive decision models constructed with the UTADIS II and the MWAE-Lex approaches are presented in Table III. It is evident that the UTADIS II model performs very well for patients in the high risk categories YYY–YNY but it leads to some significant misclassifications. For instance, about 33% of the patients from the low risk category YNN are classified as very risky cases (category YYY), whereas 17.6% of the patients with no recurrence indications (category NNN) are classified as high risk patients in category NYY. Overall, the UTADIS II model is clearly biased towards overestimating the recurrence risk as all classification errors involve cases misclassified into higher risk categories. On the other hand, the decision model constructed with the lexicographic scheme provides more balanced results with a considerable reduction of the major misclassifications noted previously.

Detailed results for the weights of the criteria in the UTADIS II and MWAE-Lex models are presented in Table IV. Both models indicate that the duration of AF episodes, the number of applications of high frequency current during AF ablation, the fluoroscopy time and the LA volume are major factors contributing to the decisions regarding the monitoring and evaluation of a patient's condition. The type of AF on the other hand, seems to be a less important factor.

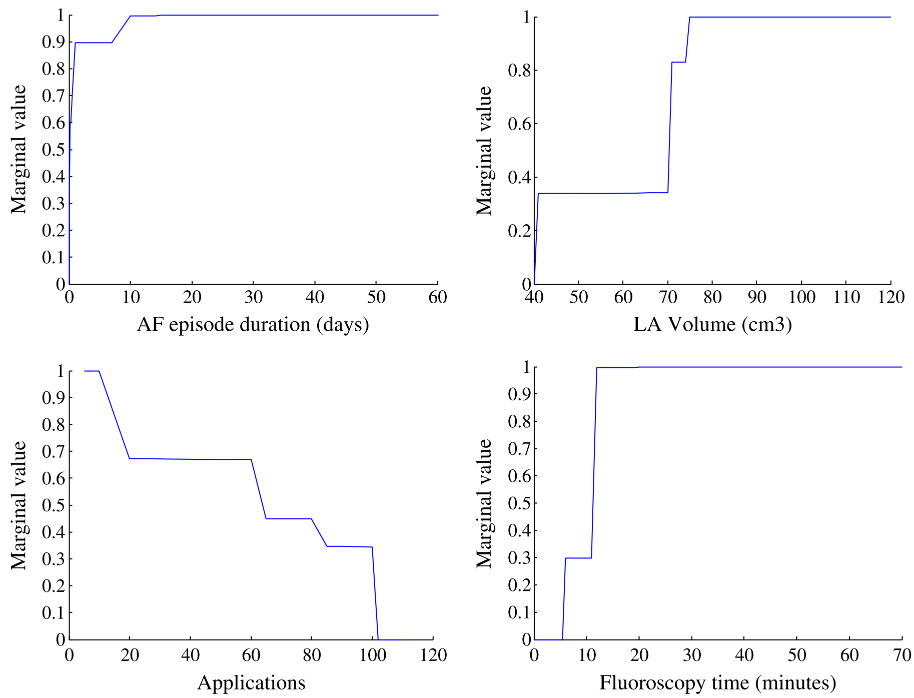


Figure 2. Marginal value functions for the four evaluation criteria with the highest weights. AF, atrial fibrillation; LA, left atrium.

The marginal value functions for the criteria with the highest weights in the MWAE-Lex decision model are illustrated in Figure 2. The function for the AF episode duration criterion has a concave form indicating that the recurrence risk increases rapidly even for cases with low AF episode duration and remains at high levels for cases with duration above 1 day. A similar concave form is also evident for the fluoroscopy time criterion, according to which the recurrence risk increases significantly in cases where the fluoroscopy time is more than 10 min. All high risk patients had fluoroscopy time greater than 10 min. This may be influenced by difficulties faced with navigating and positioning the catheters into the PV for patients with high LA volume. Additionally, patients with longstanding AF episodes have more intense and chaotic electric disorganization of the atria, demanding prolonged and repeated lesions for PVI, which are associated with longer operational times and therefore longer fluoroscopy times. On the other hand, the marginal value function for the LA volume criterion reveals that recurrence risk increases significantly for patients with LA volume above 70 cm^3. It is worth noting that under normal conditions, LA volume ranges between 25 and 58 cm^3. Therefore, the model does confirm that the ablation operation is likely to be unsuccessful for patients with LA volume much higher than normal levels. Finally, the function for the number of applications of high frequency current during AF ablation has a decreasing form, with the recurrence risk being much lower when there are more than 100 applications. These insights provide cardiologists with a disaggregated view of the global recurrence risk assessment result for each particular patient, in terms of his/her medical status on each one of the prognostic attributes. This is valuable information that strengthens the medical decision-maker's confidence on the model's reasoning and results, facilitates their qualitative analysis and supports the process for providing sound medical treatment to individual patients.

4.3. Robustness analysis

In a preference disaggregation context, such as the one adopted in this study for the inference of preferential information from a set of decision instances, the robustness of the obtained conclusions is a critical issue. The robustness concern (Roy, 2010) has recently received considerable attention among MCDA researchers. The MCDA literature related to the robustness concern in disaggregation techniques can be categorized into two main streams. The first

DOI: 10.1002/mcda

focuses on providing a range of recommendations (instead of single point results) based on the full set of models compatible with the information provided by the decision-maker (see, for instance, Greco *et al*., 2010). When inconsistencies exist in the data (i.e. classification errors), these are resolved (Mousseau *et al*., 2003) prior to the formulation of the recommendations. An alternative approach adopts a post-optimality perspective focusing on investigating the existence of multiple optimal or near-optimal models, after a decision model has been constructed using a set of reference examples (Siskos and Grigoroudis, 2010).

In this study, we adopt the latter approach in order to examine the existence of alternative decision models that describe the classification of the given patients in the available sample in the same way the obtained MWAE-Lex model does. If other very different models exist, that would raise concerns about the validity of the recommendations derived with the MWAE-Lex model for patients outside the reference sample.

Similarly to the post-optimality analysis often employed in the context of UTA-like methods (Jacquet-Lagrèze and Siskos, 1982; Siskos and Grigoroudis, 2010), in order to examine the existence of other optimal models, we first fix all the classification assignments obtained from the MWAE-Lex model and then check the variability of different models that provide the same assignments for the patients in the sample. More specifically, let $\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_7$ denote the sets of patients assigned by the MWAE-Lex model in each of the seven recurrence risk classes. Then, all additive value models that are compatible with the assignments of the MWAE-Lex model should satisfy the following constraints:

$$\begin{array}{ll} V(\mathbf{x}_i) \geq t_k + \delta & \forall i \in \hat{C}_k,\ k = 1, \ldots, 6 \\ V(\mathbf{x}_i) \leq t_{k-1} - \delta & \forall i \in \hat{C}_k,\ k = 2, \ldots, 7 \\ t_k - t_{k-1} \geq \varepsilon & k = 1, \ldots, 6 \\ V(\mathbf{x}_*) = 0,\ V(\mathbf{x}^*) = 1 & \end{array} . \quad (6)$$

In order to explore the robustness of the solutions in the polyhedron defined by these constraints, we follow two approaches. First, a post-optimality analysis (Jacquet-Lagrèze and Siskos, 1982) is employed to identify extreme solutions corresponding to the maximization and minimization of the weight for each criterion (separately). Additionally, we also examine the divergence between the weights of the criteria in the developed MWAE-Lex model and the ones that correspond to the analytic centre of the aforementioned polyhedron. As noted by Bous *et al*. (2010), decision models close to the centre of feasible polyhedron are more robust representations (compared to solutions near the boundaries) of the preferential information embodied in a set of reference examples. The identification of analytic centre can be easily performed through the solution of an optimization problem with linear constraints and logarithmic barrier objective function (Bous *et al*., 2010). A similar approach for the construction of a robust and representative sorting tool is outlined by Greco *et al*. (2011).

The criteria weights obtained from the aforementioned two approaches are shown in Table V (the post-optimality results include the minimum, maximum and the average of each criterion's weight). The results obtained from the post-optimality approach indicate that there are only very minor variations in the weights of the criteria between different models compatible with the assignments of the MWAE-Lex model. Furthermore, both the post-optimality results as well as those obtained from the analytic centre are extremely similar to the ones of the MWAE-Lex model (cf. Table IV). The robustness of decision model developed with the lexicographic approach was also verified with the average stability index (ASI) proposed by Grigoroudis and Siskos (2002), which provides a comprehensive measure of the robustness of an inferred additive value model taking into account not only the weights of the criteria but also variations with respect to the marginal value functions. By definition, ASI ranges in a 0%–100% scale, with higher values indicating more stable models. In the context of the data in this study, the ASI of the MWAE-Lex model was found to be 99.63%, slightly improved over the ASI for the UTADIS II model (99.18%).

Table V. Robustness analysis results for the weights of the criteria

	Post-optimality (min, mean and max)	Analytic centre
Type of AF	[2.00, 2.05, 2.07]	2.05
AF duration	[10.64, 10.75, 11.07]	10.74
AF episode duration	[21.24, 21.29, 21.59]	21.31
LA diameter	[10.48, 11.42, 12.82]	11.32
LA volume	[12.94, 14.12, 15.28]	14.20
Applications	[19.08, 19.14, 19.33]	19.18
Fluoroscopy time	[21.14, 21.23, 21.49]	21.21

AF, atrial fibrillation; LA, left atrium.

5. CONCLUSIONS AND FUTURE PERSPECTIVES

The development of medical decision aiding models is a challenging issue with important practical implications. In this study, we presented a real-world case study involving the development of such a model for monitoring the post-operative condition of AF patients. The model combines a number of medical factors that are potential predictors of AF recurrence and classify patients into risk categories.

A preference disaggregation approach was used to develop an appropriate model, combining two main fitting criteria through a lexicographic scheme. This lexicographic approach was found to lead to a good trade-off between the fitting criteria, resulting to a model with a small number and magnitude of misclassifications, with the overall accuracy rate ranging higher than 70%. The model performed very well in identifying high risk patients, whereas low-risk cases were found to be more difficult to be evaluated accurately, thus indicating the more detailed analysis is further needed for such cases. In that regard, it could be particularly beneficial to combine the model's results with the expertise and judgement of expert cardiologists as well as to examine the usefulness of additional prognostic criteria. Among the recurrence risk criteria used in the analysis, the duration of the most recent AF, the volume of the LA and the two criteria related to the PVI treatment (number of applications of high frequency current and fluoroscopy time) were found to contribute to the assessment of AF recurrence risk. The conducted robustness analysis verified the validity of these results. These findings are in accordance with complex nature of AF recurrence, which is due to a combination of factors regarding the nature of a patient's AF, his/her physical characteristics and the PVI operation. According to an expert medical doctor, the results of the model were found to be satisfactory, both in terms of their classification performance as well as in terms of their interpretation, their implications in practice and the insights that it provides.

Future research can focus on the consideration of a number of different variables of the aforementioned medical procedure or other interventional methods. On the methodological side, other model fitting criteria could be considered model, focusing, for instance, on eliminating/reducing important errors for specific risk categories or patient cases, which are explained poorly by the constructed model. The use of efficient optimization techniques (e.g. meta-heuristics) is also a point that could be considered in order to improve the computational efficiency of the model construction process. Comparisons with other multicriteria and data mining techniques could also be considered, focusing on the robustness of the results for patients outside the reference set (out of sample generalization ability).

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Programme 'Education and Lifelong Learning'.

ENDNOTES

1. The case YYN is missing from the analysis, as it is highly unlikely a patient with early and late atrial fibrillation recurrence to go asymptomatic at the very late time window (there was no such case in our data sample).
2. When $V(\mathbf{x}_i) = t_k$ for some $k = 1, \ldots, q-1$, then the classification of patient i is arbitrary. In such cases, we assume that patient i is assigned to risk group k (no such cases were observed in our application).

REFERENCES

Alonso-Atienza F, Rojo-Álvarez JL, Rosado-Muñoz A, Vinagre JJ, García-Alberola A, Camps-Valls G. 2012. Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection. *Expert Systems with Applications* **39**: 1956–1967.

Ammenwerth E, Nykänen P, Rigby M, de Keizer N. 2013. Clinical decision support systems: need for evidence, need for evaluation. *Artificial Intelligence in Medicine* **59**: 1–3.

Angelucci E, Barosi G, Camaschella C, Cappellini MD, Cazzola M, Galanello R, Marchetti M, Piga A, Tura S. 2008. Italian Society of Hematology practice guidelines for the management of iron overload in thal assemia major and related disorders. *Haematologica* **93**: 741–752.

Baltussen R, Youngkong S, Paolucci F, Niessen L. 2010. Multi-criteria decision analysis to prioritize health interventions: capitalizing on first experiences. *Health Policy* **96**: 262–264.

Belacel N. 2000. Multicriteria assignment method PROAFTN: methodology and medical application. *European Journal of Operational Research* **125**: 175–183.

Berner ES, La Lande TJ. 2007. Overview of clinical decision support systems. In *Clinical Decision Support Systems* Berner ES (ed). Springer: New York; 3–22.

DOI: 10.1002/mcda

Bous G, Fortemps P, Glineur F, Pirlot M. 2010. ACUTA: a novel method for eliciting additive value functions on the basis of holistic preference statements. *European Journal of Operational Research* **206**: 435–444.

Camm AJ, Lip GYH, De Caterina R, Savelieva I, Atar D, Hohnloser SH, Hindricks G, Kirchhof P. 2012. 2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association. *European Heart Journal* **33**: 2719–2747.

Chesnokov YV. 2008. Complexity and spectral analysis of the heart rate variability dynamics for distant prediction of paroxysmal atrial fibrillation with artificial intelligence methods. *Artificial Intelligence in Medicine* **43**: 151–165.

Dolan JG. 2010. Multi-criteria clinical decision support: a primer on the use of multiple criteria decision making methods to promote evidence-based, patient-centered healthcare. *Patient* **3**: 229–248.

Dolan JG, Frisina S. 2002. Randomized controlled trial of a patient decision aid for colorectal cancer screening. *Medical Decision Making* **22**: 125–139.

Doumpos M, Zopounidis C. 2002. *Multicriteria Decision Aid Classification Methods, Applied Optimization*, Springer: New York.

D Bois P, Brans JP, Cantraine F, Mareschal B. 1989. MEDICIS: an expert system for computer-aided diagnosis using the PROMETHEE multicriteria method. *European Journal of Operational Research* **39**: 284–292.

Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, Sam J, Haynes RB. 2005. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *Journal of the American Medical Association* **293**: 1223–1238.

Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. 2012. Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. *Medical Decision Making* **32**: 376–388.

Goletsis Y, Papaloukas C, Fotiadis DI, Likas A, Michalis LK. 2004. Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis. *IEEE Transactions on Biomedical Engineering* **51**: 1717–1725.

Greco S, Kadziński M, Słowiński R. 2011. Selection of a representative value function in robust multiple criteria sorting. *Computers and Operations Research* **38**: 1620–1637.

Greco S, Mousseau V, Słowiński R. 2010. Multiple criteria sorting with a set of additive value functions. *European Journal of Operational Research* **207**: 1455–1470.

Grigoroudis E, Siskos Y. 2002. Preference disaggregation for measuring and analysing customer satisfaction: the MUSA method. *European Journal of Operational Research* **143**: 148–170.

Hamacher HW, Küfer K-H. 2002. Inverse radiation therapy planning—a multiple objective optimization approach. *Discrete Applied Mathematics* **118**: 145–161.

Hardin MJ, Chhieng DC. 2007. Data mining and clinical decision support systems. In *Clinical Decision Support Systems*, Berner ES (ed). Springer: New York; 44–63.

Jacquet-Lagrèze E, Siskos J. 1982. Assessing a set of additive utility functions for multicriteria decision making: the UTA method. *European Journal of Operational Research* **10**: 151–164.

Jacquet-Lagrèze E, Siskos Y. 2001. Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research* **130**: 233–245.

Kawamoto K, Houlihan CA, Balas EA, Lobach DF. 2005. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* **330**: 765.

Keeney RL, Raiffa H. 1993. *Decisions with Multiple Objectives: Preferences and Value Trade-offs*, Cambridge University Press: Cambridge.

King G, Zeng L. 2001. Logistic regression in rare events data. *Political Analysis* **9**: 137–163.

Kirchhof P, Lip GYH, Van Gelder IC, Bax J, Hylek E, Kaab S, Schotten U, Wegscheider K, Boriani G, Brandes A, Ezekowitz M, Diener H, Haegeli L, Heidbuchel H, Lane D, Mont L, Willems S, Dorian P, Aunes-Jansson M, Blomstrom-Lundqvist C, Borentain M, Breitenstein S, Brueckmann M, Cater N, Clemens A, Dobrev D, Dubner S, Edvardsson NG, Friberg L, Goette A, Gulizia M, Hatala R, Horwood J, Szumowski L, Kappenberger L, Kautzner J, Leute A, Lobban T, Meyer R, Millerhagen J, Morgan J, Muenzel F, Nabauer M, Baertels C, Oeff M, Paar D, Polifka J, Ravens U, Rosin L, Stegink W, Steinbeck G, Vardas P, Vincent A, Walter M, Breithardt G, Camm AJ. 2012. Comprehensive risk reduction in patients with atrial fibrillation: emerging diagnostic and therapeutic options—a report from the 3rd Atrial Fibrillation Competence NETwork/European Heart Rhythm Association consensus conference. *Europace* **14**: 8–27.

Mohebbi M, Ghassemian H. 2012. Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal. *Computer Methods and Programs in Biomedicine* **105**: 40–49.

Mousseau V, Figueira J, Dias L, Gomes da Silva C, Clímaco J. 2003. Resolving inconsistencies among constraints on the parameters of an MCDA model. *European Journal of Operational Research* **147**: 72–93.

O'Sullivan D, Wilk S, Michalowski W, Słowiński R, Thomas R, Kadziński M, Farion K. 2014. Learning the preferences of physicians for the organization of result lists of medical evidence articles. *Methods of Information in Medicine* **53**: 344–356.

Postmus D, Tervonen T, van Valkenhoef G, Hillege HL, Buskens E. 2014. A multi-criteria decision analysis perspective on the health economic evaluation of medical interventions. *Eur. J. Heal. Econ.* **15**: 709–716.

DOI: 10.1002/mcda

Rahimi S, Gandy L, Mogharreban N. 2007. A web-based high-performance multicriteria decision support system for medical diagnosis. *International Journal of Intelligence Systems* **22**: 1083–1099.

Roy B. 2010. Robustness in operational research and decision aiding: a multi-faceted issue. *European Journal of Operational Research* **200**: 629–638.

Schlaefer A, Viulet T, Muacevic A, Fürweger C. 2013. Multicriteria optimization of the spatial dose distribution. *Medical Physics* **40**: 121720.

Siskos Y, Grigoroudis E. 2010. New trends in aggregation-disaggregation approaches. In *Handbook of Multicriteria Analysis* Zopounidis C, Pardalos PM (eds). Springer: Berlin Heidelberg; 189–214.

Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. 2011. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Statistics in Medicine* **30**: 1419–1428.

Van Valkenhoef G, Tervonen T, Zwinkels T, de Brock B, Hillege H. 2013. ADDIS: a decision support system for evidence-based medicine. *Decision Support Systems* **55**: 459–475.

Wann LS, Curtis AB, January CT, Ellenbogen KA, Lowe JE, Estes NAM, Page RL, Ezekowitz MD, Slotwiner DJ, Jackman WM, Stevenson WG, Tracy CM, Fuster V, Rydén LE, Cannom DS, Le Heuzey J-Y, Crijns HJ, Olsson SB, Prystowsky EN, Halperin JL, Tamargo JL, Kay GN, Jacobs AK, Anderson JL, Albert N, Hochman JS, Buller CE, Kushner FG, Creager MA, Ohman EM, Ettinger SM, Guyton RA, Tarkington LG, Yancy CW. 2011. 2011 ACCF/AHA/HRS focused update on the management of patients with atrial fibrillation (updating the 2006 guideline): a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* **123**: 104–123.

You JJ, Singer DE, Howard PA, Lane DA, Eckman MH, Fang MC, Hylek EM, Schulman S, Go AS, Hughes M, Spencer FA, Manning WJ, Halperin JL, Lip GYH. 2012. Antithrombotic therapy for atrial fibrillation: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* **141**: e531S–75S.

Zopounidis C, Doumpos M. 2002. Multicriteria classi fication and sorting methods: a literature review. *European Journal of Operational Research* **138**: 229–246.

DOI: 10.1002/mcda

EURO J Decis Process
DOI 10.1007/s40070-015-0050-4

ORIGINAL ARTICLE

Business process analytics: a dedicated methodology through a case study

Pavlos Delias[1] · Michael Doumpos[2] · Nikolaos Matsatsinis[2]

Received: 29 September 2014 / Accepted: 16 July 2015

Abstract Business process analytics is a set of techniques that can be applied to event datasets created by logging the execution of business processes, and emerges as a promising decision aid field. In this work we propose a methodology based on the process mining approach to guide the implementation of process analytics projects. Following a conceptual analysis of existing methodologies, we extract the common methodological steps and present a practical synthesis. The proposed methodology reaches the business need of exposing more than just a static, marginal snapshot of performance by considering a process perspective. We present the methodology in tandem with a case study of a customer service request handling process. We analyze a real dataset containing events from an incident and a problem management information system, and deliver results that eventually can raise the capacity of the company to manage the process.

Keywords Process analytics · Process mining · Methodology

Mathematics Subject Classification 90B50 · 68T99

✉ Pavlos Delias
pdelias@teikav.edu.gr

Michael Doumpos
mdoumpos@dpem.tuc.gr

Nikolaos Matsatsinis
nikos@dpem.tuc.gr

[1] Department of Accounting and Finance, Eastern Macedonia and Thrace Institute of Technology, Kavala, Greece

[2] School of Production Engineering and Management, Technical University of Crete, Chania, Greece

Published online: 31 July 2015

1 Introduction

Customer service request handling is a reactive business process that is triggered when a customer submits a service request to the help desk of a company. It has been identified as a core function of modern organizations, due to its tight relationship with their marketing function (Wilson et al. 2012). Establishing a service response capability includes a number of actions (Grance et al. 2004), like creating a service response policy, setting guidelines for communicating with outside parties regarding customer requests, selecting a team structure and staffing model, establishing relationships between the help desk team and other groups, both internal (e.g., technical support teams) and external, determining what services the incident response team should provide, and staffing and training the incident response team.

There are multiple factors that affect the complexity of the process, such as the number of support teams involved, the organizational hierarchy, the number of products/product categories being served, special business rules etc. Due to the complexity of this process, special IT systems are often employed, which can significantly add to the business value (Lin and Kao 2014). A common practice reference model that introduces standard best practices for IT service management is the Information Technology Infrastructure Library (ITIL) (Hochstein et al. 2005). Nevertheless, the processes described in ITIL are deliberately non-prescriptive, therefore the process flow is not enforced [e.g., by a workflow engine (Tarantilis et al. 2008)]. In practice, the actual behavior can significantly vary, not just according to the organizational implementation, but also because of a plethora of other implementation parameters as well (e.g., the resource performing the activities). Process mining (Van der Aalst et al. 2012) is a promising approach to expose the real behavior of the process from IT systems' logs and conceals business process optimization potentials (Van der Aalst 1996).

The process mining approach has recently attracted researchers for the service request management process analysis (van Dongen et al. 2013). Since the respective process takes place in a highly flexible environment, multiple techniques are typically combined to deliver a solution. In De Weerdt et al. (2012), the authors propose a combination of trace clustering and text mining to enhance process discovery techniques with the purpose of retrieving more useful insights from process data, while in Ferreira and Mira da Silva (2008) process mining is used to assess whether a business process is implemented according to ITIL guidelines.

In this work we propose a methodology based on the process mining approach to discover coordinated patterns of behavior in a customer service request handling process. Our efforts are not centered in delivering a standard framework, but rather in guiding the implementation of a process analytics application. The goal of this paper is to demonstrate through a real-world case study a roadmap for evidence-based decision making. The case study concerns an IT system used by Volvo IT to support incidents reported by the IT service users. It is a reactive business process, and although there is an Organization structure and some general rules (Steeman 2013), the company actively looks for inefficiencies. The proposed methodology and its actions deliver effective analytics for such a business quest. Actions are described with a clear reference to the case study, however they are relevant and applicable to any case where event-based data sets are available. We should emphasize that besides guiding an application, additional motivation for delivering such a methodology is to support best practices reporting and sharing, and to endorse evidence-based approaches for decision making.

The next section is a brief presentation to scaffold readers into the proposed approach. Next, we present phases and actions in parallel with their concrete instantiations concerning the case study. Last, a short discussion on the limitations and future work concludes the paper.

2 Outline of the proposed approach

The explosion of generated data has lead to several proposals of methodologies for practitioners to extract useful insights from datasets, KDD (Fayyad et al. 1996), SEMMA (Matignon 2007), and CRISP-DM (Chapman et al. 2000) being the most popular among them. Concentrating on process mining, we regard the L* model (Van der Aalst 2011, pp. 283–286) that portrays the basic steps to improve a structured (Lasagna) process, context specific approaches [e.g., methodologies for healthcare (Rebuge and Ferreira 2012; Delias et al. 2015)], and the work of Bozkaya et al. (2009) to exploit different perspectives of process mining for a specific purpose (to gain an quick overview). Van der Heijden (2012) followed a System Engineering Process to identify the requirements and the main activities of a process mining project and to deliver (like CRISP-DM) “an industry-, tool-, and application neutral methodology”. There are evident overlaps between the above methodologies. Following a conceptual analysis, we can extract the common methodological steps, necessary to deliver a process analytics project. Thus, the proposed methodology *synthesizes* existing works in order to accelerate project delivery. It focuses and summarizes the bottom line of the cited works, leading to expedited knowledge discovery. The methodology is illustrated in Fig. 1. It consists

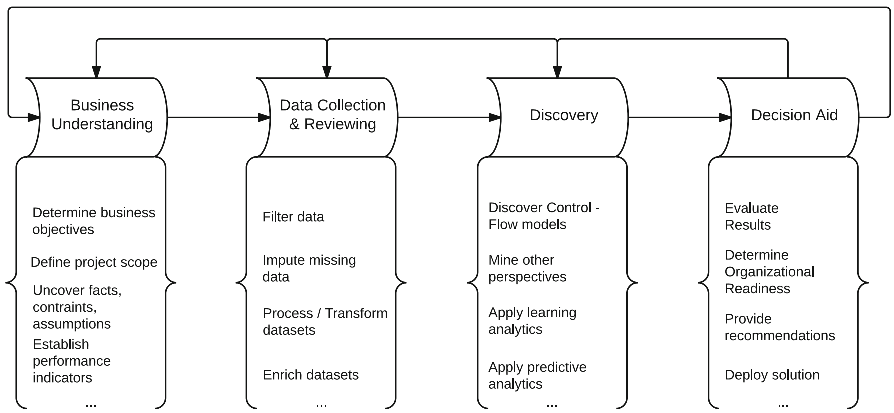


Fig. 1 The proposed methodology: phases and actions

of a set of actions classified into four phases, defined as a higher level of abstraction. Although a sequence is demonstrated (aiming to suggest a consistent and progressive development), it is not rigid. Switching between phases is an expected, as well as an essential part for any project implementation. In addition, the dots at the bottom of the list of actions for each phase are used for suspense, to indicate that the actions' lists are not complete checklists but coarse guides.The methodology assumes that a process notion is omnipresent (it exists in problem definition, data format, solutions' intuition, etc.). Therefore, a process mining approach is qualified, since ordinary data analysis or data mining techniques would fail to capture the sequencing of the related events. The basic phases are Business Understanding (figuring out the business context and developping the shape of solutions); Data Collection and Reviewing (acquiring and preparing the raw material); Discovery (extracting bits of knowledge); and Decision Aid (building a rapport between results and business goals). The methodology will be presented in parallel with the case study implementation, while we commit the next sections to the analytical description of the steps.

3 Business understanding

The rationale of this phase is to help the analyst arrive at a stage of reflection where she has a clear understanding of the business context, and where she can assess how alternative actions can contribute to the business objectives. The case study concerns Volvo IT Belgium. The company's support system comprises of three levels: The first line operates as a common help desk/service desk. The Second line comprises of specialized functional teams within a higher organizational line. The third line is a team of specific product or technical experts and is also within a higher organizational line. The company provided a dataset (Steeman 2013) from its information system that supports the incidents management for the 2013 edition of the BPI challenge. The suggested actions to reach business understanding are enumerated in the following subsections.

3.1 Determine business objectives and define process scope

Determining business objectives implies informally describing the problem to be solved, specifying all business questions as precisely as possible, and pointing out expected benefits in business terms (Chapman et al. 2000). In the case study, the primary goal of the incident management process is restoring a customer's normal service operation as quickly as possible when incidents arise, ensuring that the best possible levels of service quality and availability are maintained. So, the pertinent business question is if there are any particular patterns that delay the resolution of issues. Since the quick resolution of the issue is defined within the Service Level Agreement of the company, there are evident gains in settling tactics to avoid those patterns.

Defining process scope deals with acknowledging what parts of the process can be tracked through the logged data and which type of information is both available and useful (van der Heijden 2012). Concerning the case study, each record contains a number of variables such as the unique ticket number of the service request, the impact of the case (a measure of the business criticality of the incident), the case status (queued, accepted, completed or closed) and sub-status (assigned, awaiting assignment, cancelled, closed, in progress, wait or unmatched), the business area of the user reporting the incident, the technology-wise division of the organization, the support team that will try to respond to the service request, and the location that takes the ownership of the support team.

The process is roughly the following: a customer submits a service request. The process reactively triggers a "first line" response, in other words, the Service Desk or the Expert Help Desk tries to resolve the issue. When this is not possible, the case should be escalated to Second Line and/or Third Line teams. The dataset contains 65,533 timestamped events related to the incident management process.

3.2 Uncover facts, constraints, and assumptions

There is an announced policy of the company that most of the incidents need to be resolved by the first line support teams (mainly service desks). This is called "*Push to Front*" tactic and it is mostly a matter of efficiency. Pushing to front, allows the 2nd and 3rd line support teams to focus on their special, more demanding tasks (usually not related to customer service support). Unless this tactic is consistently applied a lot of 'easy', big volume cases will end up in those lines. As such, pushing to front is an important coordinated pattern that may arise during the process execution.

Besides pushing work towards the front, any team upon receiving a task can either try to resolve the issue by itself or hand over the task to another team (of the same or of another line). Handover of work is an ordinary action, however if this is excessively used, it may have an inadmissible effect on process efficiency. Namely, extensive handover may reveal dodging or deferring behavior. The opposite (extensive takeover) may also reveal some undesired elements, like lack of collaboration mentality or lack of knowledge transferring. Therefore, the inter-team handovers may also include coordinated patterns of (social) behavior.

A special case of handover of work is when support teams send the same case to each other again and again. We shall call this undesirable situation "*Ping Pong*". Ping Pong is also an undesirable coordinated behavior that may affect significantly the process performance.

3.3 Establish performance indicators

Every indicator should follow some basic requirements, like representativeness, simplicity and easiness of interpretation, feasible data collection, etc. (Franceschini et al. 2007). Generally, each indicator refers to a specific objective, that is to say a sort of reference point used as a basis of comparison. Indicators may originate from a global performance measurement system of the company, but it is recommended to define ad-hoc indicators for the specific process mining project. In this case study, there is a single performance indicator: resolution time. However, since the focus in on discovering patterns that affect the primary indicator, we shall define two secondary indicators: *Push to Front* and *Ping Pong*.

Push to Front is measured by a binary variable for each case (a case can either push to front or not) while for the overall performance measuring the percentage of cases that are pushing to front is enough. Since Push to Front is a desired behavior, the greater the percentage, the better for the enterprise. Ping Pong is measured by a numerical variable, because a case may have multiple Ping Pongs, and the amount of Ping Pongs undoubtedly affects the resolution time. We further discuss the need for a numerical scale in Sect. 5.2.

4 Data collection and reviewing

This phase consists of data manipulation actions. It is a time-consuming phase that demands for data filtering, dealing with missing values, transforming representations of the variables, and adding new information to the existing dataset. Luckily, the case study dataset has been preprocessed by its provider (Steeman 2013) in a way that very few data manipulation actions were required. In particular: we did not apply any filters to data, and we did not face any missing values problems.

The dataset in its original format contains a list of timestamped events. It is quite hard to elicit patterns of behavior from within this format, since the sequencing of events and their aggregation per case are not exploited. Therefore, the leading step is to reach a process perspective for the dataset. Therefore, we committed data to process format following two different perspectives (and thus creating two different datasets)

1. Control flow-wise (trajectories of status/sub-status changes)
2. Social-wise (transactions among support teams or lines)

Finally, to enrich the dataset, we created an additional variable for the service line where every event is performed. This information was embedded within the Support Team variable, so we extracted the pertinent values from the original variable. If no value for the line was logged for a support team (ST), we assumed it to be a 1st line ST. In case that a ST spread over more than one line, we used the most front one.

5 Discovery

5.1 Discover control-flow

The control-flow perspective focuses on the control-flow, i.e., the ordering of activities. The goal of mining this perspective is to find a good characterization of all possible paths expressed in some process notation (Van der Aalst 2011, p. 11), or in other words the goal is to answer the question “what does the actual process look like?” (Bozkaya et al. 2009). The intuition of a common automated process discovery algorithm is to scan the Event Log for sequencing patterns and then to try

Table 1 Status and sub-status alternatives

Status	Sub-status
Accepted	Assigned, in progress, wait, wait-user, wait-customer, wait-implementation, wait vendor
Queued	Awaiting assignment
Completed	In call, resolved, closed, cancelled
Unmatched	Unmatched

to aggregate them. However, many different concepts and techniques have been proposed. The interested reader is redirected to (Van der Aalst 2011, ch. 5–6) for a relevant discussion.

For the specific case study, control-flow refers to how the status/sub-status of a case changes during its lifecycle. There are 13 distinct alternatives for the status/sub-status of a case (presented in Table 1). Although the set of activities (status changes) is small, we noticed that there are 2278 different variants of the same process (for a dataset of 7554 cases). Out of these 2278 variants, just 88 have a frequency higher than 100, while the dominant variant represents just a 23 % of total cases, a fact that confirms that the process environment is highly flexible.

Since there is no strict sequencing rule, discovering an exact behavior would not reflect the real situation, and would probably be of little importance. In general terms, cases go from some *Accepted* sub-status to either a *Completed* sub-status or to *Queued*. In the latter option, the case returns to an *Accepted* sub-status. A process map is depicted in Fig. 2, where some labels for performance measures are printed. In particular, the heavier the weight of an edge, the worst its performance. The illustration has been created using Disco® (Fluxicon 2012) and it is a direct way to visualize the process' bottlenecks. The largest delays happen between Completed-Resolve and Completed-Closed (7.2 days), Accepted-Wait User and Completed-Resolve (5.3 days) and Accepted-Wait Implementation and Completed-Resolved (4.7 days). It is also interesting to note that there is a meantime of 4.3 days between the Completed-Closed status and the Accepted-In Progress status, a fact that indicates that some cases are closed only to be re-initiated after 4–5 days.

5.2 Mine other perspectives

Besides control-flow, other common perspectives are the social or organizational perspective (which focuses on what actors are involved and how they are related), and the case perspective (which focuses on properties of cases). Concerning the case study, the patterns described in Sect. 3.2 are social (organizational) patterns, therefore our focus is on mining that perspective.

First of all, we need to evaluate the "Ping Pong" and the "Push to Front" patterns for each case, based on the descriptions of Sect. 3.2. To this end, the algorithms illustrated in Fig. 3 were developed. The algorithms follow the definitions of the patterns. We recall that the definition of Push to Front in this paper refers to the case when the 1st line support teams can resolve the service

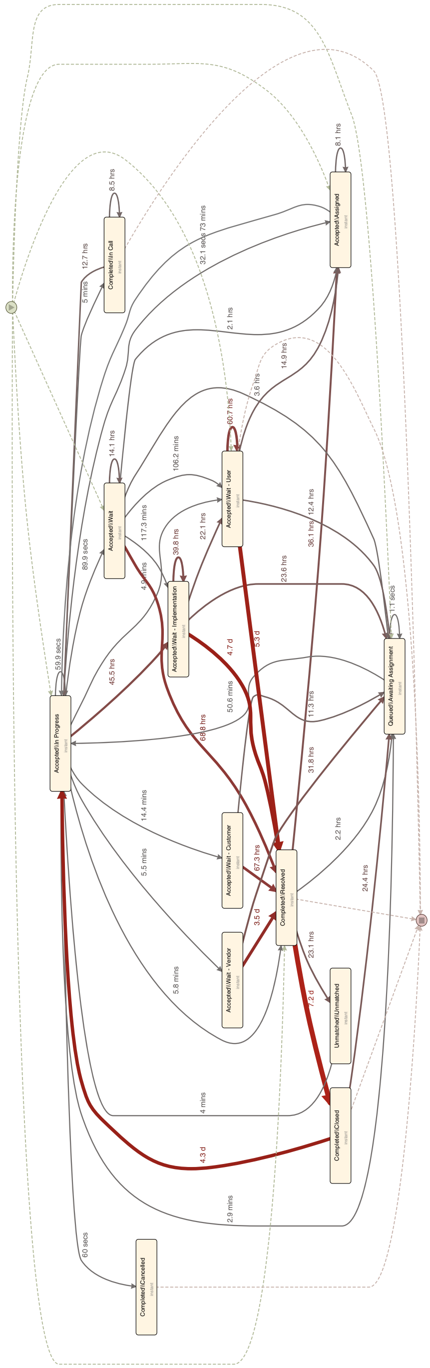
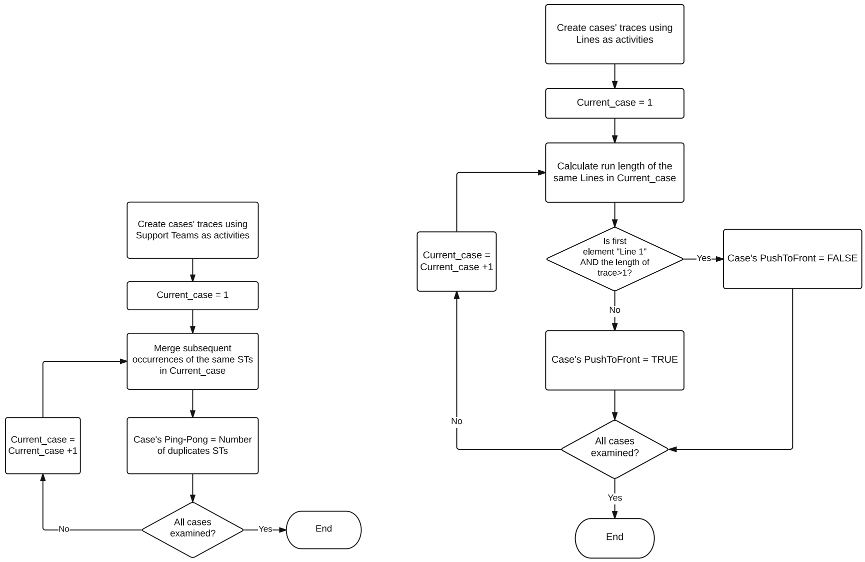


Fig. 2 Process performance map



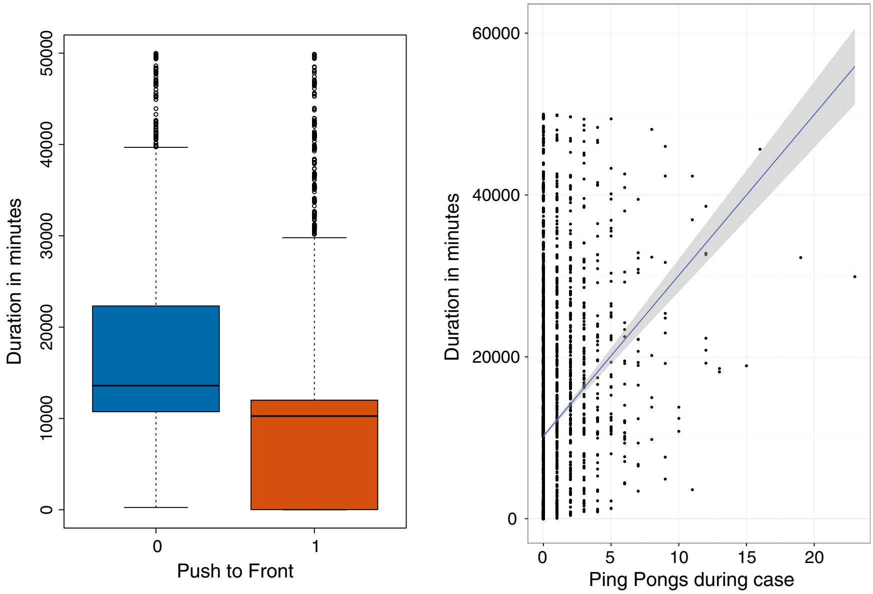
(a) Evaluating the Ping Pong behavior from the Event Log.

(b) Evaluating the Push to Front behavior from the Event Log

Fig. 3 Flow charts to evaluate the social patterns

request without interference of a 2nd or 3rd line support team. The definition of "Ping Pong" is that a Ping Pong occurs when a support team is revisited during the case, after it has passed the work to another team. However, we count a single Ping Pong per support team, even if this is revisited multiple times. This definition allows for a numeric representation of the Ping Pong behavior (a case may have multiple Ping Pongs, yet attributed to different teams).

Figure 4 illustrates these effects for the mainstream cases (outliers, i.e., cases that last more than 50,000 min are removed). In particular, Fig. 4a depicts what is the difference in duration between cases that *Push to Front* and cases that do not. The drawn boxes are rectangles with edges defined by the lower and upper quartiles (25 % and 75 % respectively). The line inside the box is located at the median while values greater than 1.5 times of the upper quartile are presented as dots. It is clear that cases that do *Push to Front* are resolved quicker than cases that don't. While for *Push to Front* a binary variable is sufficient, for *Ping Pong* a numerical scale is preferred. An illustrative argument for this choice is presented in Fig. 5, where we see that it would not be fair to evaluate Ping Pong with a binary variable, since the number of Ping Pongs has a strong effect on the process behavior. In this point we shall remind that a Ping Pong is assigned per team, i.e., even if a pair of teams handover their work multiple times during a case, that will still count for two (one for each team that is revisited). Figure 4b plots a simple regression line between the duration (in minutes) of cases and the number of Ping Pongs they contain. As expected, both behaviors (lacking *Push to Front* and *Ping Pong*) have a negative effect on the case duration.



(a) Duration of cases that Push to Front (1) or not (0)

(b) Duration of cases over the number of Ping Pongs they contain

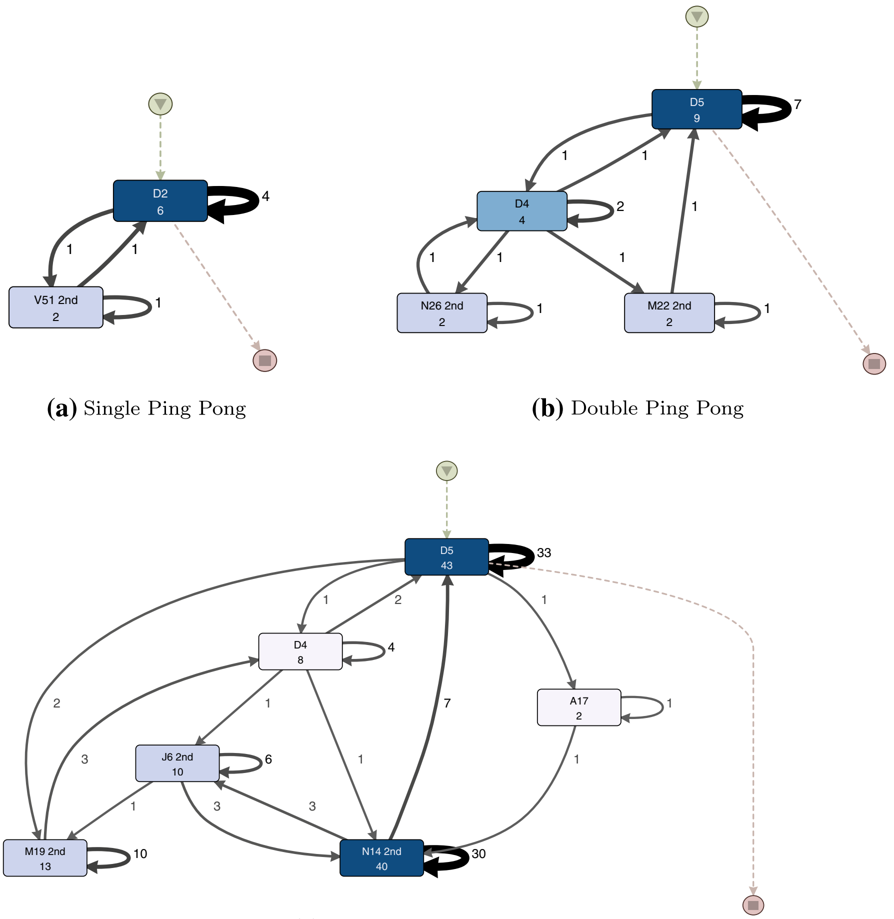
Fig. 4 The effect on case duration

5.3 Apply learning analytics

Learning analytics come in many shapes. Trying to profoundly epitomize, we can name as learning analytics techniques that estimate relationships among variables, determine which variables are important in predicting future values, as well as techniques that segment the data into homogeneous groups. Casting this description to the case study context, we observe that identifying a set of important factors is a highly relevant question in the customer service field (Tseng and Huang 2007). In this section, we propose techniques that would reach an answer when the focus is on a process view. The intuition of this task is to discover the features that have a great impact to the process flow and thus facilitate process improvement of reengineering by detecting, listing or classifying best practices (Reijers and Mansar 2005). In particular, we perform a discrepancy analysis for the observed behavioral variation, as well as we try to assess the importance of factors that cause cases to *Push to Front* or to *Ping Pong*.

5.3.1 Discrepancy analysis

Considering a case as an animating process, discrepancy measures the among-cases variability of the cases' life-cycle trajectories. Therefore, higher discrepancy, would reflect a greater level of uncertainty about the path followed by the cases. In this section, we integrate the sequence discrepancy analysis with the regression tree



(a) Single Ping Pong **(b)** Double Ping Pong

(c) Multiple Ping Pong

Fig. 5 Nodes are *support teams* and *arrows* show the handovers of work. A *numerical scale* for the Ping Pong behavior is preferable. The illustrations have been created using Disco® (Fluxicon 2012)

method introduced in Studer et al. (2011). The intuition of this regression tree method is the following: start with all cases grouped in an initial node. Then, recursively partition each node using values of another variable. At each node, the variable and the split are chosen in such a way that the resulting child nodes differ as much as possible from one another or have, more or less equivalently, lowest within-group discrepancy. The process is repeated on each new node until a certain stopping criterion is reached.

An apparent barrier to the application of the above method is that it is not straightforward to calculate the "mean" trace. Therefore, the discrepancy (variance) of the traces will be defined from their pairwise dissimilarities. Perhaps the most popular dissimilarity measure used for sequence analysis is the generalized Levenshtein distance. It is defined as the lowest cost of transforming one sequence

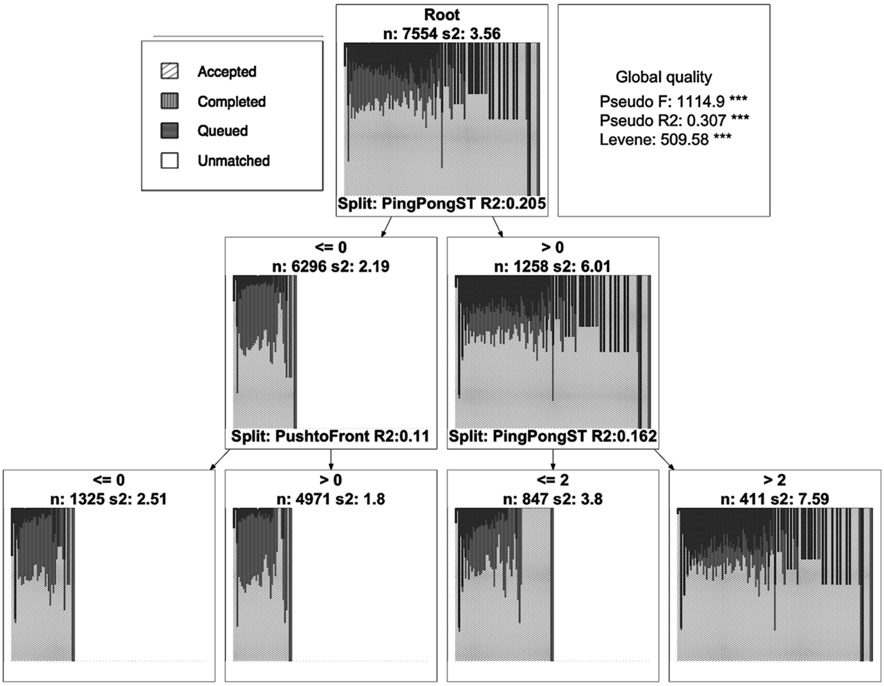


Fig. 6 Discrepancy analysis for cases lifecycle trajectories

into the other by means of state insertions–deletions and state substitutions. However, we still need to find a way to gauge the contribution of each instance to the overall variance. To this end, we exploit the generalization of the Ward criterion (Batagelj 1988). In particular, Batagelj (1988) introduced the notion of a *gravity center* of a set of sequences and proposes a formula to calculate the distance of any sequence from it. This proposition allows the calculation of metrics like the sum of squares of these distances and the residual within the sum of squares. Based on this fact, and following the ANOVA mindset, Studer et al. (2011) introduced a metric to measure the part of the discrepancy that is explained by differences in group positioning (and they call it pseudo-R^2) and a metric to compare the explained discrepancy to the residual discrepancy (they call it pseudo-F).

To build the regression tree, we use the pseudo-R^2 as a splitting criterion (we choose to split based on the variable that yields the highest R^2). As a stopping criterion, we trust the pseudo-F significance. In other words, we no longer split a branch as soon as we get a non-significant F (considering a *p-value* of 0.05) for the selected split. For the implementation of this method, we used the TraMineR (Gabadinho et al. 2011) package of R.

We examined the role of just two predictors (Push to Front and Ping Pong), and as illustrated in Fig. 6, these two social patterns alone explain approximately 30 % of the total discrepancy. Both of them result in clustered behaviors. In particular, the first split is among cases that *Ping Pong* or not (0 and greater than 0). Cases of the later category (no *Ping Pong*) last significantly less and visit a lot less frequently the "*Queued*" status. At the second level, the leftmost split is among cases that *Push to Front* (>0) and not (0). We regard that cases that Push to Front reach a "*Completed*" status earlier, and that their average duration is smaller. The rightmost split is again based on the *Ping Pong* behavior, but this time the critical value is two. Cases that Ping Pong more than twice spend an important percentage of their lifetime in a "*Queued*" status, and are naturally prolonged.

5.3.2 Detecting the factors' importance

In the previous section we elaborated on learning the role of *Ping Pong* and *Push to Front* to the variation of the process. In this paragraph, we try to discover what are the factors that affect these coordinated patterns. To this end, we propose to use a tree-based method. The basic reason that favored this choice is that tree-based methods are more easily interpreted by non-experts. However, a drawback of trees is their accuracy level, since they suffer from high variance. Therefore, in order to create a more powerful and robust model, we propose to use Random Forests (RF) (Breiman 2001). The basic idea of RF is that they grow a number of decision trees on bootstrapped training samples. During the creation of every tree, and every time a split is considered, a random set of characteristics (predictors) is used. There are a number of reasons why Random Forests are expected to deliver better results. First of all, since the new dataset is only a subset, it is likely that the number of records it contains is small. RF are more suitable for this kind of problems (small number of records with respect to the number of predictors). Then, by considering different characteristics for every split, RF can deal with high-order interactions and correlated characteristics (Strobl et al. 2008). Moreover, through RF, it is possible to obtain a summary of the importance of each characteristic (how significant it is for the branching decisions) using the Gini index (for classification trees) or the RSS (for regression trees).

More specifically, since we have a numerical scale for *Ping Pong*, we grow a regression tree, using seven factors as predictors: the existence of the *Push to Front* behavior, the Support Team (*ST*) where the case was initiated (we kept just the top 30 of STs, using "other" for the rest), the *Country* of the ST, the code of the *Product*

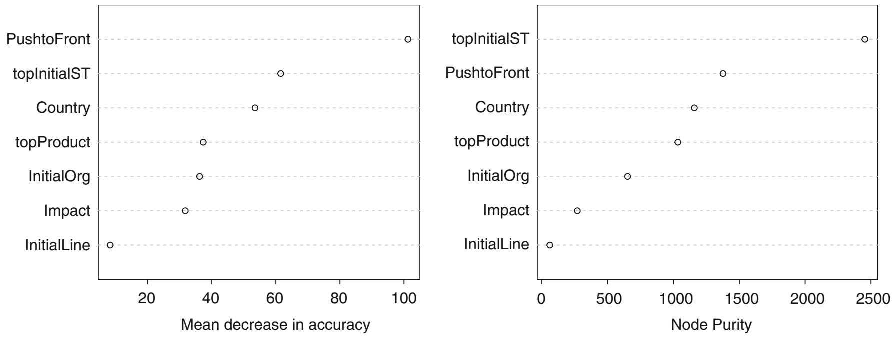


Fig. 7 Variable importance with respect to the *Ping Pong* behavior

(again we kept just the top 30 ones), the *Organization Line* where the case was initiated, the *Impact* of the incident, and the *Line* of the ST (1st, 2nd or 3rd). We measure the importance of factors with two ways: The first calculates the mean decrease of accuracy in predictions when the corresponding variable is left out of the model (leftmost case of Fig. 7), and the second considers the total decrease in node impurity that results from splits over the corresponding variables. As Fig. 7 shows, the top three factors that lead to extended *Ping Pong* is the *Push to Front* tactic, the ST where the incident handling was initiated and its Country.

As far as it concerns the *Push to Front* behavior, we grow a Random Forest of classification trees to assess the variables' importance. A typical way to evaluate the importance is the Gini index, which is actually a measure of total variance across all classes. Nevertheless, the Gini index has been criticized as a tool to asses the importance of characteristics because it is biased in favor of continuous variables and variables with many categories. Therefore, we follow a permutation scheme, as proposed in Strobl et al. (2007). The basic idea of employing permutation tests, is that if the variable is not important (the null hypothesis), then rearranging the values of that variable will not degrade prediction accuracy. Following this method (which yields an accuracy of 89.54 %) *Country* appears to be the most important factor, followed by the *Product* code.

In the previous Sects. 5.3.1, 5.3.2 we applied techniques to fathom the process behavior. The same techniques can be used to predict future outcomes and trends (Predictive Analytics). The most common situations are predicting group membership, predicting a future value, and predict relevant conditions. However, this kind of analysis is not included in this work, due to lack of relevant data.

6 Decision aid

Decision aid is about "providing decision makers with the most favorable conditions possible for the type of behavior which will increase coherence between the evolution of the process, on the one hand, and the goals and/or systems of values within which these actors operate on the other" (Roy 1994). There are some basic actions contributing to this mission. We shall dedicate the following paragraphs to briefly describe them.

6.1 Evaluate results

Evaluating results involves reviewing them to determine whether they are still tied to the original questions, whether they meet the business objectives, and assess the business value they deliver. In case that the business objectives are not met, the analysts should report the reasons of the decline.

Considering the primary objective (resolution time), results are clearly targeted, revealing either paths that slow down the process (Fig. 2), or correlations of resolution time with behavioral patterns (Fig. 4), or even a view of variation of the life-cycles (Fig. 6). Concerning the secondary objectives, with respect to *Ping Pong*, results suggest an important finding, since now the company can focus on specific factors (Push to Front, Support Team, and Country) and look for assignable causes. As long as for the *Push to Front* pattern, the knowledge gain is again important, since by focusing on the *Product* factor we can spot the products that display strong *Push to Front* behavior and the ones that don't. Based on these results, we can provide perceptive recommendations (described in the next paragraph).

6.2 Determine organizational readiness and provide recommendations

Effective assessment of the organization readiness should result a smooth transition and an increased user satisfaction with the proposed changes. Unfortunately, for this case study, we did not have access to relevant information (e.g., are process stakeholders willing to reinforce and reward positive teamwork behaviors? Are they willing to allow time for personnel to attend training?), therefore this step is skipped, while the following recommendations are not aware of any corporate particularities (probably existent, yet not available).

Following the evidence of the previous sections, and to deal with the *Ping Pong* effect, we shall recommend focusing on STs. This way we can detect that 80 % of the total Ping Pongs is due to less than 5 % of the STs. STs with extended Ping Pong behavior can thus be identified, enabling the company to take perceptive actions. By focusing on the *Country* factor, we regard that the largest average of Ping-Pongs per case belongs to the Netherlands (4.7 per incident) or that most Ping-Pongs happen within Belgium (with an average of 1.66 per incident). This knowledge facilitates the company to go deeper and look for the reasons that these specific countries are prone to Ping-Pong.

Moreover, to stimulate the Push to Front behavior, a possible response policy could be to assign Product codes that perform low on *Push to Front* directly to other lines, or to train the service desk (1st Line) specific for these products, or even to create a knowledge sharing mechanism that will capture solutions specific to those product codes. Focusing on the *Country* factor, we regard for instance that Poland and USA are countries that are Pushing to Front while India has the worst performance. This piece of information should make the company aware and drive it to search concretely for the reasons (e.g., is it a matter of poor training or cultural differences?). Assuming, that we are willing to trade-off accuracy performance for more direct interpretation, it is possible to grow a single classification tree and get a number of 'rules' that can classify/predict *Push to Front*. Such an output would allow for rule-based process monitoring and support the timely investigation of undesired patterns (Caron et al. 2013).

6.3 Deployment

Deployment for this methodology has the meaning of providing decision aid by participating in the final decision legitimization (Roy and Damart 2002). In particular, the analyst should be able to enlighten and scientifically accompany decision-making notably (Roy 1993):

- by making the objective stand out more clearly from the less objective (the entire methodology has an evidence-based mentality)
- by separating robust from fragile conclusions (we applied a robust classification technique to explain the factors affecting the behavior and to deliver a predictive model for undesired behaviors)
- by avoiding the pitfall of illusory reasoning and by emphasizing incontrovertible results (the effect of every variable can be pointedly exhibited).

7 Conclusions

In this work we presented a dedicated approach based on process mining to guide the implementation of process analytics projects. We explored a real case study with the goal to provide insights to this implicit business process and to raise the capability of the company to handle service requests. This work demonstrated that a process perspective generates knowledge gains since ordinary data analysis methods may miss salient information of event based data sets. Our methodology was capable to detect how some social-wise patterns (behaviors) are related with performance and provide insights about the factors that shape these behaviors. Ultimately, the proposed methodology exemplifies how business decisions and process analysis can benefit from the analytical capabilities of a process mining approach.

We avoided to convey the methodology as a standard framework, since the following limitations are acknowledged: The phases include sets of generic actions. Such actions do not necessarily suggest specific techniques. For example, we do not make any particular recommendations about the diagnostic techniques for determining business objectives or defining the project's scope. Likewise, we do not provide recommendations about selecting specific process mining techniques. Yet, that would be a very interesting step for future improvement, since the usefulness of a mining technique decidedly depends on the available data, domain knowledge, expertise, business culture, and the objectives of the project. Moreover, the methodology does not include any mapping technique to match the deliverables with the business objectives, namely to assure that the delivered analytics best support the project's goals. It is out of the scope of this work, but very relevant for future enhancements to add any monitor and control over time functions for the effectiveness and efficiency of the solutions. Finally, the methodology evolved and has been validated through a case study. Future work would target additional validation methods. One way is to collect evidence and to check for realizations of the indicated actions in other applications. A different way is to rely on expert judgement for a more qualitative evaluation to competently meet the challenges of process analytics projects.

Acknowledgments This research has been co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)-Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

References

Batagelj V (1988) Generalized ward and related clustering problems. In: Classification and related methods of data analysis. North-Holland, Amsterdam, pp 67–74

Bozkaya M, Gabriels J, Werf JMvd (2009) Process diagnostics: a method based on process mining. IEEE, pp 22–27. doi:10.1109/eKNOW.2009.29. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4782560

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Caron F, Vanthienen J, Baesens B (2013) Comprehensive rule-based compliance checking and risk management with process mining. Decis Support Syst 54(3):1357–1369

Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R (2000) CRISP-DM 1.0 step-by-step data mining guide. Technical report. The CRISP-DM consortium. http://www.crisp-dm.org/CRISPWP-0800.pdf

De Weerdt J, vanden Broucke S, Vanthienen J, Baesens B (2012) Leveraging process discovery with trace clustering and text mining for intelligent analysis of incident management processes. In: 2012 IEEE congress on evolutionary computation (CEC), pp 1–8. doi:10.1109/CEC.2012.6256459

Delias P, Doumpos M, Grigoroudis E, Manolitzas P, Matsatsinis N (2015) Supporting healthcare management decisions via robust clustering of event logs. Knowledge-based systems. doi:10.1016/j.knosys.2015.04.012. http://linkinghub.elsevier.com/retrieve/pii/S0950705115001501

van Dongen B, Weber B, Ferreira D, De Weerdt J (eds) (2013) Business Process intelligence challenge 2013, vol 1052, CEUR

Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI Mag 17(3):37–54. doi:10.1609/aimag.v17i3.1230

Ferreira DR, Mira da Silva M (2008) Using process mining for ITIL assessment: a case study with incident management. Bournemouth University

Fluxicon (2012) Disco. Fluxicon. http://www.fluxicon.com/disco/

Franceschini F, Galetto M, Maisano D (2007) Management by measurement: designing key indicators and performance measurement systems. Springer, Berlin, New York

Gabadinho A, Ritschard G, Müller NS, Studer M (2011) Analyzing and visualizing state sequences in r with traminer. J Stat Softw 40(4):1–37

Grance T, Kent K, Kim B (2004) Computer security incident handling guide. NIST Special Publication, pp 800–61

Hochstein A, Zarnekow R, Brenner W (2005) Itil as common practice reference model for it service management: formal assessment and implications for practice. In: Proceedings of IEEE international conference on the e-Technology, e-Commerce and e-Service, 2005. EEE ’05, pp 704–710. doi:10.1109/EEE.2005.86

Lin WT, Kao TWD (2014) The partial adjustment valuation approach with dynamic and variable speeds of adjustment to evaluating and measuring the business value of information technology. Eur J Oper Res 238(1):208–220

Matignon R (2007) Data mining using SAS enterprise miner. Wiley-Interscience, Hoboken, NJ

Rebuge Á, Ferreira DR (2012) Business process analysis in healthcare environments: a methodology based on process mining. Inf Syst 37(2):99–116. doi:10.1016/j.is.2011.01.003. http://linkinghub.elsevier.com/retrieve/pii/S0306437911000044

Reijers H, Mansar SL (2005) Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. Omega 33(4):283–306

Roy B (1993) Decision science or decision-aid science? Eur J Oper Res 66(2):184–203. doi:10.1016/0377-2217(93)90312-B. http://linkinghub.elsevier.com/retrieve/pii/037722179390312B

Roy B (1994) On operational research and decision aid. Eur J Oper Res 73(1):23–26. doi:10.1016/0377-2217(94)90136-8. http://linkinghub.elsevier.com/retrieve/pii/0377221794901368

Roy B, Damart S (2002) L’ analyse coûts-avantages, outil de concertation et de légitimation? Metropolis 108(109):7–16

Steeman W (2013) Bpi challenge 2013, incidents. doi:10.4121/uuid:500573e6-accc-4b0c-9576-aa5468b10cee

Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinform 8(1):25

Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. BMC Bioinform 9(1):307

Studer M, Ritschard G, Gabadinho A, Müller NS (2011) Discrepancy analysis of state sequences. Sociol Methods Res 40(3):471–510. doi:10.1177/0049124111415372

Tarantilis C, Kiranoudis C, Theodorakopoulos N (2008) A web-based ERP system for business services and supply chain management: application to real-world process scheduling. Eur J Oper Res 187(3):1310–1326

Tseng TLB, Huang CC (2007) Rough set-based approach to feature selection in customer relationship management. Omega 35(4):365–383. doi:10.1016/j.omega.2005.07.006. http://linkinghub.elsevier.com/retrieve/pii/S0305048305000964

Van der Aalst WMP (1996) Petri net based scheduling. Oper-Res-Spektrum 18(4):219–229

Van der Aalst WMP (2011) Process mining: discovery, conformance and enhancement of business processes, 1st edn. Springer, New York

Van der Aalst WMP, Adriansyah A, Medeiros AKA, Arcieri F, Baier T, Blickle T, Bose JC, Brand P, Brandtjen R, Buijs J, Burattin A, Carmona J, Castellanos M, Claes J, Cook J, Costantini N, Curbera F, Damiani E, Leoni M, Delias P, Dongen BF, Dumas M, Dustdar S, Fahland D, Ferreira DR, Gaaloul W, Geffen F, Goel S, Günther C, Guzzo A, Harmon P, Hofstede A, Hoogland J, Ingvaldsen JE, Kato K, Kuhn R, Kumar A, Rosa M, Maggi F, Malerba D, Mans RS, Manuel A, McCreesh M, Mello P, Mendling J, Montali M, Motahari-Nezhad HR, Muehlen M, Munoz-Gama J, Pontieri L, Ribeiro J, Rozinat A, Seguel Pérez H, Seguel Pérez R, Sepúlveda M, Sinur J, Soffer P, Song M, Sperduti A, Stilo G, Stoel C, Swenson K, Talamo M, Tan W, Turner C, Vanthienen J, Varvaressos G, Verbeek E, Verdonk M, Vigo R, Wang J, Weber B, Weidlich M, Weijters T, Wen L, Westergaard M, Wynn M (2012) Process mining manifesto. In: Van der Aalst WMP, Mylopoulos J, Rosemann M, Shaw MJ, Szyperski C, Daniel F, Barkaoui K, Dustdar S (eds) Business process management workshops, vol 99. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 169–194. http://www.springerlink.com/index/10.1007/978-3-642-28108-2_19

van der Heijden T (2012) Process mining project methodology: developing a general approach to apply process mining in practice. Master Thesis, Technische Universiteit Eindhoven, Eindhoven. http://alexandria.tue.nl/extra2/afstversl/tm/Van_der_Heijden_2012.pdf

Wilson A, Zeithaml VA, Bitner MJ, Gremler DD (2012) Services marketing: integrating customer focus across the firm, 2nd Eu. McGraw Hill, UK

Robust extensions of the MUSA method based on additional properties and preferences

Evangelos Grigoroudis
School of Production Engineering and Management, Technical University of Crete
University Campus, Kounoupidiana, GR73100 Chania, Greece
e-mail: vangelis@ergasya.tuc.gr

Yannis Politis
School of Science and Technology, Hellenic Open University
Parodos Aristotelous 18, GR26 335, Patra, Greece
e-mail: politis@ergasya.tuc.gr

Abstract: The MUSA method is a collective preference disaggregation approach following the main principles of ordinal regression analysis under constraints using linear programming techniques. The method has been developed in order to measure and analyse customer satisfaction and it is used for the assessment of a set of marginal satisfaction functions in such a way that the global satisfaction criterion becomes as consistent as possible with customer's judgments. The main objective of the method is to assess collective global and marginal value functions by aggregating individual judgments. This study evaluates different extensions of the MUSA method with the introduction of additional constraints in the basic linear programming formulation of the method. These constraints concern special properties for the assessed average indices and additional customer preferences about the importance of the criteria and they may be modelled as multiobjective linear programming problems. The main aim of the study is to show how the introduction of these additional constraints and information may improve the stability of the estimated results. An illustrative example is presented in order to show the applicability of this approach.

Keywords: MUSA method; robustness analysis; importance judgements; customer satisfaction

Author

Biographical notes:

Evangelos Grigoroudis is Associate Professor on management of quality processes in the School of Production Engineering and Management of the Technical University of Crete, Greece (2002-). He followed postgraduate studies in Technical University of Crete, Greece from where he received his Ph.D. degree in 1999. He has received distinctions from the Hellenic Operational Research Society, the Academy of Business and Administrative Sciences, the World Automation Congress, the Foundation of Ioannis and Vasileia Karayianni, the Technical University of Crete, and the State Scholarships Foundation of Greece. He acts as reviewer for more than 50 scientific journals, and he is associate editor of the Operational Research: An International Journal, International Journal of Decision Support Systems, Journal of Knowledge Economy, International Journal of Social Ecology and Sustainable Development, Journal of Innovation and Entrepreneurship, Journal of Technology, Innovation and Education, and International Journal of Food and Beverage Manufacturing and Business Models and member of the Editorial Board of the scientific journals: International Journal of Information and Decision Sciences, International Journal of Information Systems in the Service Sector, International Journal of Multicriteria Decision Making. He is author/editor of 10 books on the measurement of service quality, the business strategy and management, and the multicriteria decision aid approaches, as well as of a significant number of research reports and papers in scientific journals and conference proceedings.

Yannis Politis was born in 1974 in Piraeus. He has a Bachelor in Production Engineering and Management and a M.Sc. in Operational Research from the Department of Production Engineering and Management of the Technical University of Crete. In 2005 he was declared a Doctor from the Technical University of Crete. He served for several years as a business consultant for quality issues. He is an expert and he has designed, implement and audit a large number of quality management systems for various business sectors. He works since 2009 in the Internal Audit Departments of the Prefecture of Athens and the Region of Attica as a chief auditor. He is a member of the board of the Hellenic Operational Research Society. He has taught in the Department of Production Engineering and Management of the Technical University of Crete and in the Business Administration Departments of the Technological Educational Institutes of Athens and Crete. He serves as a Collaborative Academic Staff of the Hellenic Open University for the postgraduate course of Quality Assurance, since 2012. He is an author of papers and studies in scientific journals and conference proceedings. His research interests include quality management, business excellence, customer satisfaction measurement, internal audit, multicriteria decision analysis, operational research, organization assessment.

1 Introduction

The MUSA (MUlticriteria Satisfaction Analysis) method is a preference disaggregation model for measuring and analysing customer satisfaction (Grigoroudis and Siskos, 2002, 2010). It follows the principles of ordinal regression analysis and aims at evaluating the satisfaction level of a set of individuals (customers, employees, etc.) based on their values and expressed preferences.

The philosophy of preference disaggregation in multicriteria analysis is to assess/infer preference models from given preferential structures and to address decision-aiding activities through operational models (Siskos et al., 2005). In this context, the MUSA method uses linear programming techniques in order to assess a set of marginal satisfaction functions in such a way that the global satisfaction criterion becomes as consistent as possible with customers' judgments.

Considering that the MUSA method is based on a linear programming (LP) modelling, the problem of multiple or near optimal solutions appears in several cases. This has an impact on the robustness of the provided results. The quality of collected data and the incapability to interact with customers complicates the task of finding stable solutions. In addition, there might be a gap between the decision-maker's "true" model and the model resulting from the disaggregation computational mechanism as noted by Siskos et al. (2005).

The problem of robustness analysis in multicriteria decision aid (MCDA) models has gained significant attention during recent years. As emphasised by Roy (2010), robustness may be considered as an enabling tool for decision analysts to resist the phenomena of approximations and ignorance zones. However, robustness should also refer to the results and the decision support activities (e.g. conclusions, argumentation). In the particular area of ordinal regression analysis, several approaches have been proposed, using LP as the main inference mechanism, such as UTA-GMS (Greco et al., 2008), GRIP (Figueira et al., 2009), and RUTA (Kadzinski et al., 2013). A detailed discussion of robustness analysis in the context of ordinal regression may be also found in Greco et al. (2010), Kadzinski et al. (2012), and Corrente et al. (2013).

In the MUSA method, robustness is considered as a post/near-optimality analysis problem, especially on the form and the extent of the polyhedron of the feasible solutions, where the observed variance in the post-optimality step indicates the degree of instability of the final results. In this study we follow the general methodological framework for managing robustness proposed by Siskos and Grigoroudis (2010), which consists of the following steps: (a) infer a representative additive value model, (b) develop a robustness measure, (c) examine alternative rules of robustness analysis (e.g., addition preference judgements), if the robustness measure form step (b) is not satisfactory.

In this context, there are several ways to overcome potential stability problems and increase the robustness of the MUSA results. For example, customers may be asked to give additional information (e.g., information about the importance of the criteria along with the usual satisfaction questions) or additional constraints may be introduced in the basic LP of the method taking into account desired properties of the estimated collective preference system. All these approaches are able to reduce the polyhedron of the feasible solutions. Several studies have shown that the introduction of additional information or constraints in ordinal regression models increases the stability of the final results (Greco et al, 2008, 2010; Corrente et al., 2012, 2013). A different approach based on a set of compatible preference models is presented by Angilella et al. (2014). They propose the MUSA-INT method as a generalised approach that takes also into account positive and negative interactions among criteria, and considers a set of utility functions representing customers' satisfaction. In this context, different customer satisfaction profiles may be evaluated using the set of compatible value functions.

This study evaluates different extensions of the MUSA method with the introduction of additional constraints in the basic LP of the method. In particular, a customer satisfaction survey may include, besides the usual performance questions, preferences about the importance of the criteria. Using such questions, customers are asked either to judge the importance of a satisfaction criterion based on a predefined ordinal scale, or rank the set of satisfaction criteria according to their importance. All these performance and importance preferences are modelled using LP techniques in order to assess a set of marginal satisfaction functions in such a way that the global satisfaction criterion and the importance preferences become as consistent as possible with customer's judgments.

Furthermore, the LP formulation of the method gives the ability to consider additional constraints regarding special properties of the assessed model variables. An interesting type of such constraints concerns additional properties for the assessed average indices, which include:

a. Average satisfaction indices (mean value of the global and marginal value functions) that can be considered as the basic performance norms.
b. Average demanding indices, which indicate customers' demanding level and represent the average deviation of the estimated value functions from a "normal" (linear) function.

The main aim of the study is to show how, by incorporating these additional constraints in the LP of the original MUSA method, the stability of the estimated results may be improved. The proposed approach is modelled as a multiobjective linear programming (MOLP) problem, while different stability and fitting measures have been used in order to analyse and compare the provided results. It should be noted that given the LP formulation, the MUSA method is rather flexible, allowing to examine several extensions.

The rest of this paper is organised into five sections. Section 2 presents briefly the basic principles of the MUSA method, as well as the post-optimality analysis and the robustness measures. Section 3 is devoted to the modelling of additional constraints in the basic LP of the MUSA method (special properties for the assessed average indices and additional customer preferences about the importance of the criteria). Section 4 proposes an extension of the MUSA method and presents a heuristic approach for solving the formulated MOLP problem. A simple illustrative example in order to evaluate the reliability of the results and demonstrate the applicability of the proposed approach is given in section 5. Finally, section 6 discusses some concluding remarks, as well as future research in studying the robustness analysis of the MUSA method.

2 MUSA method

2.1 Mathematical development

The MUSA method is a multicriteria preference disaggregation approach that provides quantitative measures of customer satisfaction. It considers the qualitative form of customers' judgments (Siskos et al., 1998; Grigoroudis and Siskos, 2002) and its main objective is the aggregation of individual judgments into a collective value function assuming that customer's global satisfaction depends on a set of n criteria or variables representing service characteristic dimensions.

MUSA assesses global and partial value (satisfaction) functions Y^* and X_i^* given customers' judgments about their global satisfaction Y and their satisfaction with regard to the set of discrete criteria X_i. The main objective of the method is to achieve the maximum consistency between the value function Y^* and the customers' judgments Y.

The method follows the principles of ordinal regression analysis under constraints using LP techniques (Jacquet-Lagrèze and Siskos, 1982; Siskos and Yannacopoulos, 1985; Siskos, 1985). The ordinal regression analysis equation with the introduction of a double-error variable has the following form:

$$\begin{cases} Y^* = \sum_{i=1}^{n} b_i X_i^* - \sigma^+ + \sigma^- \\ \sum_{i=1}^{n} b_i \end{cases} \tag{1}$$

where the value functions Y^* and X_i^* are normalised in the interval $[0,100]$, b_i is the weight of the *i*-th criterion, and σ^+ and σ^- are the overestimation and the underestimation errors, respectively.

According to the aforementioned definitions and assumptions, the customers' satisfaction evaluation problem may be formulated as a LP in which the goal is the minimisation of the sum of errors under the constraints:

a. ordinal regression equation (1) for each customer,
b. normalisation constraints for Y^* and X_i^* in the interval $[0,100]$, and
c. monotonicity constraints for Y^* and X_i^*.

In order to decrease the computational effort required for optimal solution search, a set of transformation variables can be introduced in the model. These transformation variables represent the successive steps of the value functions Y^* and X_i^*, and their introduction reduces the size of the previous LP (Siskos and

Yannacopoulos, 1985; Siskos, 1985). The transformation equation can be written as follows (see also Figure 1):

$$\begin{cases} z_m = y^{*m+1} - y^{*m} & \text{for } m = 1, 2, \ldots, \alpha - 1 \\ w_{ik} = b_i x_i^{*k+1} - b_i x_i^{*k} & \text{for } k = 1, 2, \ldots, \alpha_i - 1 \text{ and } i = 1, 2, \ldots, n \end{cases} \tag{2}$$

where y^{*m} is the value of the y^m satisfaction level, x_i^{*k} is the value of the x_i^k satisfaction level, and α and α_i are the number of global and partial satisfaction levels.

According to the aforementioned definitions and assumptions, and by introducing the z_m and w_{ik} variables, the basic LP of the MUSA method can be written as follows:

$$\begin{cases} [\min] F = \sum_{j=1}^{M} \sigma_j^+ + \sigma_j^- \\ \text{subject to} \\ \sum_{i=1}^{n} \sum_{k=1}^{x_i^j - 1} w_{ik} - \sum_{m=1}^{y^j - 1} z_m - \sigma_j^+ + \sigma_j^- \quad \text{for } j = 1, 2, \ldots, M \\ \sum_{m=1}^{\alpha - 1} z_m = 100 \\ \sum_{i=1}^{n} \sum_{k=1}^{\alpha_i - 1} w_{ik} = 100 \\ z_m, w_{ik}, \sigma_j^+, \sigma_j^- \geq 0 \quad \forall m, i, k, j \end{cases} \tag{3}$$

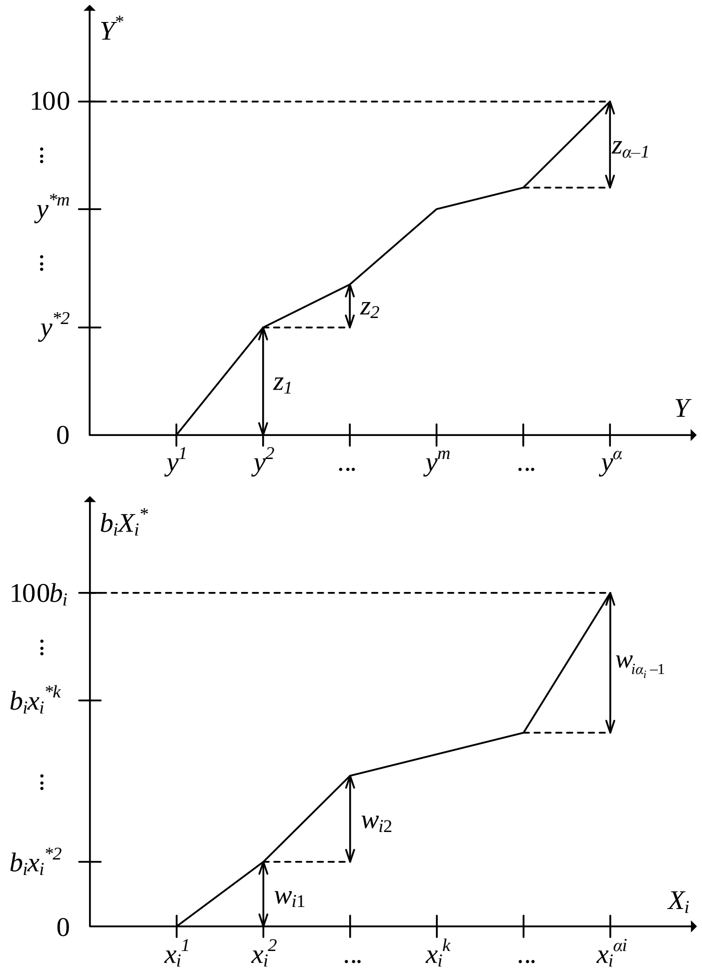


Figure 1. Transformation variables z_m and w_{ik} in global and partial value functions (Grigoroudis and Siskos, 2010)

where M is the number of customers, and y^j and x_i^j are the j -th level on which variables Y and X_i are estimated (i.e., global and partial satisfaction judgments of the j-th customer).

The principles and the main methodological framework of the MUSA method have been developed by Siskos et al. (1998) and Grigoroudis and Siskos (2002), while a discussion and a more detailed presentation of the method may also be found in Grigoroudis and Siskos (2010).

2.2 Post-optimality analysis

A post-optimality analysis stage is also included in the MUSA method in order to face the problem of multiple or near optimal solutions. Considering that the method is based on LP modelling, post-optimality analysis can give insight about the stability of the provided results. The MUSA method applies a heuristic method for near optimal solutions search, which is based on the following (Siskos, 1984; Siskos and Grigoroudis, 2010):

a. In several cases, the optimal solutions are not the most interesting, given the uncertainty of the model parameters.
b. The number of the optimal or near optimal solutions is often huge, and therefore an exhaustive search method (reverse simplex, Manas-Nedoma algorithms) requires a lot of computational effort.

The final solution is obtained by exploring the polyhedron of near optimal solutions, which is generated by the constraints of the previous LP (see Figure 2). During the post-optimality analysis stage of the MUSA method, n LPs (equal to the number of criteria) are formulated and solved. Each LP program maximises the weight of a criterion and thus the solutions give the internal variation of the weight of all criteria, and consequently give an idea of the importance of these criteria in the decision-maker’s preference system (Siskos et al., 2005).

The post-optimality analysis LP has the following form:

$$\begin{cases} [\min] F' = \sum_{k=1}^{\alpha_i - 1} w_{ik} \quad \text{for } i = 1, 2, \ldots, n \\ \text{subject to} \\ F \leq F^* + \varepsilon \\ \text{all the constraints of LP(3)} \end{cases} \tag{4}$$

where F^* is the optimal value of the objective function of LP(3) and ε is a small percentage of F^*.

The average of the optimal solutions given by the n LPs(4) may be considered as the final solution of the problem. In case of instability, a large variation of the provided solutions appears and the final average solution is less representative.

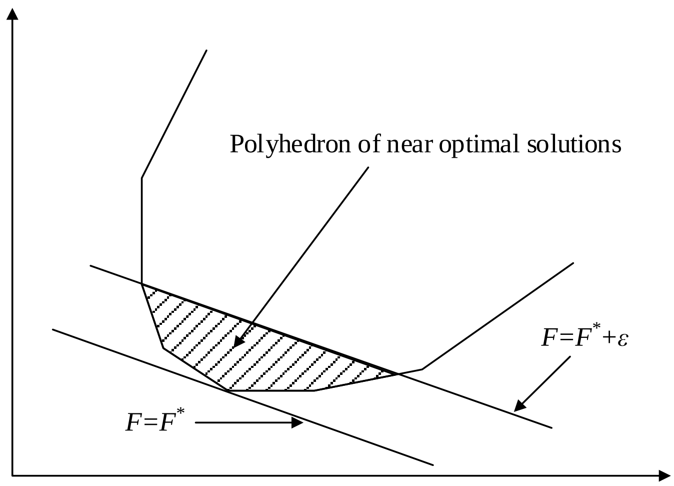


Figure 2. Post-optimality analysis (Jacquet-Lagrèze and Siskos, 1982)

2.3 Average fitting and stability indices

The fitting level of the MUSA method refers to the assessment of a preference collective value system (value functions, weights, etc.) for the set of customers with the minimum possible errors. Therefore, the optimal values of the error variables indicate the reliability of the value system that is evaluated.

Several fitting measures may be used depending on the optimum error level and the number of customers. Grigoroudis and Siskos (2002) propose the following simple average fitting index AFI_1:

$$AFI_1 = 1 - \frac{F^*}{100M} \tag{5}$$

where F^* is the minimum sum of errors of the initial LP(3).

AFI_1 is normalised in the interval $[0,1]$, and it is equal to 1 if $F^* = 0$, that is when the method is able to create a preference value system with zero errors. On the other hand, AFI_1 takes its worst value only when the pairs of the error variables σ^+ and σ^- take the maximum possible values.

An alternative fitting indicator is based on the percentage of customers with zero error variables. This means that, for these customers, the estimated preference value systems fits perfectly with their expressed satisfaction judgments. This average fitting index AFI_2 can be assessed as follows:

$$AFI_2 = \frac{M_0}{M} \tag{6}$$

where M_0 is the number of customers with $\sigma^+ = \sigma^- = 0$.

The previous fitting indicators are rather simple and can be easily calculated. However, they present several disadvantages. For example, AFI_1 may rarely take small values, since usually $F^* \ll 100M$ because it is unreasonable for all the error variables in a regression-type model to have their maximum possible values, i.e., $\sigma^+ + \sigma^- = 100 \quad \forall j$. For this reason, AFI_1 usually overestimates the fitting ability of the MUSA method. On the other hand, AFI_2 examines only the existence of non-zero errors, without taking into account the values of these error variables. Therefore, in several cases AFI_2 underestimates MUSA's fitting level. Additionally, the values of AFI_2 may not give a reliable indication for the overall fitting ability of the MUSA method, since a small (or high) value of AFI_2 does not imply a respective small (or high) sum of errors.

Another fitting indicator which overcomes these disadvantages is proposed by Grigoroudis and Siskos (2010). This indicator examines separately every level of global satisfaction, and calculates the maximum possible error value for each one of these levels. As shown in Figure 3, for the estimation of y^{*m}, $0 \le y^{*m} \le 100$ holds and thereby, the maximum overestimation σ^+ and underestimation σ^- errors are $100 - y^{*m}$ and y^{*m}, respectively. Thus, the overall maximum error for every global satisfaction level is the maximum of the previous expressions. Based on this approach, the average fitting index AFI_3 can be calculated according to the following formula:

$$AFI_3 = 1 - \frac{F^*}{M \sum_{m=1}^{\alpha} p^m \max\{y^{*m}, 100 - y^{*m}\}} \tag{7}$$

where p^m is the frequency of customers belonging to the y^m satisfaction level.

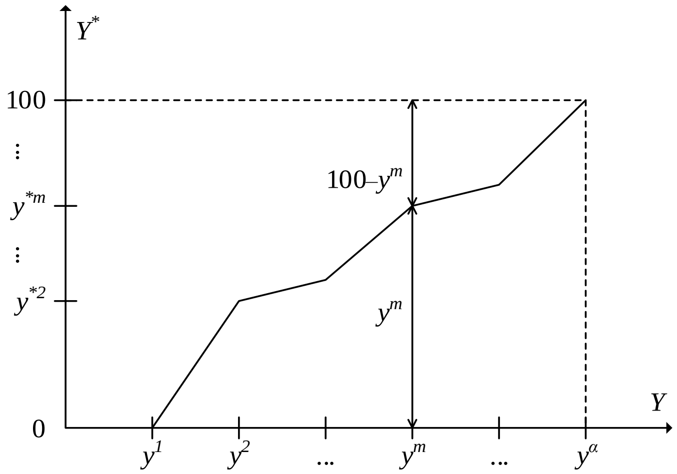


Figure 3. Maximum error values for the *m*-th global satisfaction level (Grigoroudis and Siskos, 2010)

AFI_3 is actually an alternative formulation of AFI_1, which takes into account the maximum values of the error variables for every global satisfaction level, as well as the number of customers that belongs to this level. Although AFI_3 appears more reliable, all of the aforementioned average fitting indicators are highly affected by potential inconsistencies in customer satisfaction judgments. Therefore, the examination of all these indices may give a more complete view for the fitting ability of the MUSA method.

Regarding robustness analysis, as already mentioned, the MUSA method consists of a post-optimality analysis stage. During this post-optimality stage, n LPs are formulated and solved, which maximise repeatedly the weight of each criterion. The mean value of the weights of these LPs is taken as the final solution, and the observed variance in the post-optimality matrix indicates the degree of instability of the results. Thus, the mean value of the normalised standard deviation of the estimated weights can be used as an average stability index (*ASI*) of the method:

$$ASI = 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{n \sum_{t=1}^{n} (b_i^t)^2 - \left(\sum_{t=1}^{n} b_i^t \right)^2}}{100\sqrt{n-1}} \tag{8}$$

where b_i^t is the estimated weight of the *i*-th criterion in the *t*-th post-optimality analysis LP.

ASI is normalised in the interval $[0,1]$, and when this index takes its maximum value, then $ASI = 1 \Leftrightarrow b_i^t = b_i \quad \forall i,t$ (where b_i is the final estimated weight for criterion *i*). On the other hand, if *ASI* takes its minimum value, then:

$$ASI = 0 \Leftrightarrow b_i^t = \begin{cases} 1 & \text{if } i = t \\ 0 & \text{if } i \neq t \end{cases} \quad \forall i,t$$

Grigoroudis and Siskos (2010) discuss additional stability measures in the context of the MUSA method. For example, the range of the weights during post-optimality analysis is also able to provide valuable information for the robustness of the provided results. These ranges may give a confidence interval for the estimated weights, and can identify possible competitiveness in the criteria set, i.e., the existence of certain customer groups with different importance levels for the satisfaction criteria.

2.4 Results of the MUSA method

The estimated value/satisfaction functions are the most important results of the MUSA method, considering that they show the real value, in a normalised interval $[0,100]$, that customers give for each level of the global or marginal ordinal satisfaction scale. The form of these functions indicates the customers' degree of demanding. Furthermore, the assessment of a performance norm, globally and per satisfaction criteria as well, may be very useful in customer satisfaction analysis and benchmarking. The average global and partial satisfaction indices, S and S_i, respectively, are used for this purpose, and may be assessed according to the following equations (see also Figure 4):

$$\begin{cases} S = \dfrac{1}{100} \displaystyle\sum_{m=1}^{\alpha} p^m y^{*m} \\ S_i = \dfrac{1}{100} \displaystyle\sum_{k=1}^{\alpha_i} p_i^k x_i^{*k} \quad \text{for } i = 1, 2, \ldots, n \end{cases} \tag{9}$$

where p^m and p_i^k are the frequencies of customers belonging to the y^m and x_i^k satisfaction levels, respectively. The average satisfaction indices are basically the mean values of the global or partial value functions and they are normalised in the interval $[0,100\%]$.

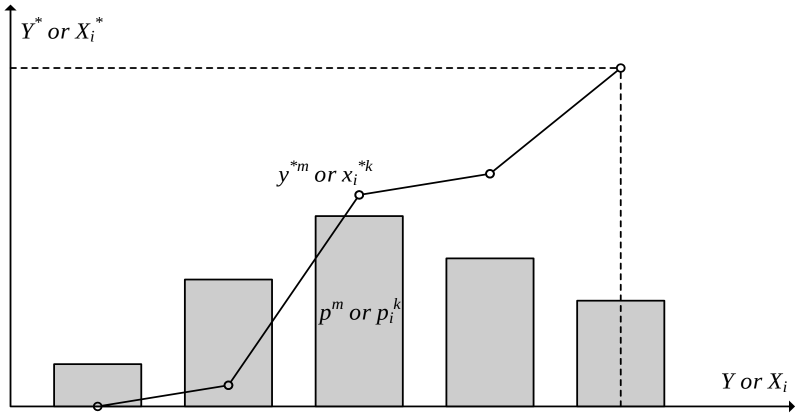


Figure 4. Assessing average satisfaction indices (Grigoroudis and Siskos, 2010)

Other important results of the MUSA method refer to the average global and partial demanding indices, which represent the average deviation of the estimated value curves from a “normal” (linear) function. These indices are normalised in the interval $[-1,+1]$ and reveal the demanding level of customers. They are assessed based on the following formulas:

$$\begin{cases} D = \dfrac{\sum_{m=1}^{\alpha-1}\left(\dfrac{100(m-1)}{\alpha-1} - y^{*m}\right)}{100\sum_{m=1}^{\alpha-1}\dfrac{m-1}{\alpha-1}} & \text{for } \alpha > 2 \\ D_i = \dfrac{\sum_{k=1}^{\alpha_i-1}\left(\dfrac{100(\mathrm{k}-1)}{\alpha_i-1} - x_i^{*k}\right)}{100\sum_{k=1}^{\alpha_i-1}\dfrac{k-1}{\alpha_i-1}} & \text{for } \alpha > 2 \text{ and } i = 1,2,\ldots,n \end{cases} \qquad (10)$$

where D and D_i are the average global and partial demanding indices, respectively.

Grigoroudis and Siskos (2002) note that these indices represent the average deviation of the estimated value curves from a “normal” (linear) function (Figure 5), and distinguish the following possible cases:

a. $D = 1$ or $D_i = 1$: customers have the maximum demanding level.

b. $D = 0$ or $D_i = 0$: this case refers to the neutral customers.

c. $D = -1$ or $D_i = -1$: customers have the minimum demanding level.

Demanding indices are used in customer behaviour analysis. They may also indicate the extent of company’s improvement efforts: the higher the value of the demanding index, the more the satisfaction level should be improved in order to fulfil customers’ expectations

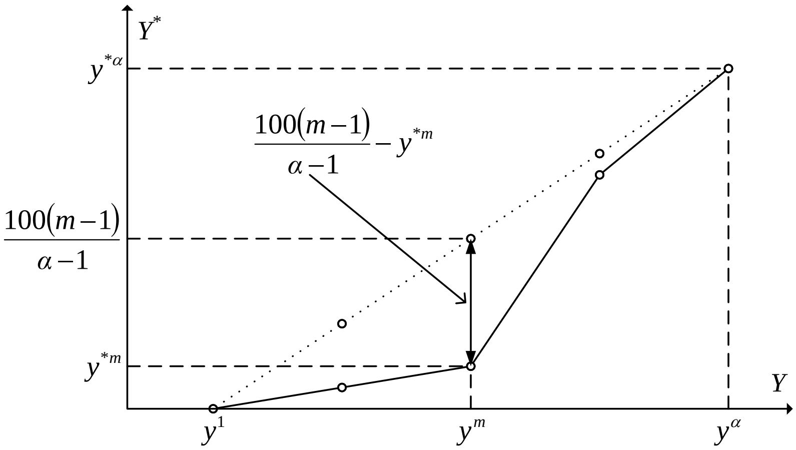


Figure 5. Assessing average demanding indices (Grigoroudis and Siskos, 2010)

3 Modelling additional information and properties

Although several extensions of the MUSA method have been proposed (see Grigoroudis and Siskos, 2010 for a comprehensive review), in this section, two cases of modelling additional information and properties in order to improve the robustness of the MUSA results are examined. The first refers to the desired properties of the collective preference system (i.e., additional properties for the assessed average indices), while the second case concerns customer preferences on satisfaction criteria importance

3.1 Desired properties of model variables

As already noted, the LP formulation of the MUSA method gives the ability to consider additional constraints regarding special properties of the assessed model variables. One of the most interesting extensions concerns additional properties for the assessed average indices.

In the case of average satisfaction indices, which are considered as the main performance indicators of a business organisation, a reasonable approach is to assume that the global average satisfaction index S is an aggregation of the partial average satisfaction indices S_i. If a weighted sum aggregation formula is used, then the following property occurs:

$$S = \sum_{i=1}^{n} b_i S_i \Leftrightarrow \sum_{m=1}^{\alpha} p^m y^{*m} = \sum_{i=1}^{n} b_i \sum_{k=1}^{\alpha_i} p_i^k x_i^{*k} \tag{11}$$

Taking into account the transformation variables z_m and w_{ik}, the previous formula (11) can be rewritten as follows

$$\sum_{m=2}^{\alpha} p^m \sum_{t=1}^{m-1} z_t = \sum_{i=1}^{n} \sum_{k=2}^{\alpha_i} p_i^k \sum_{t=1}^{k-1} w_{it} \tag{12}$$

Similarly, a weighted sum formula may be assumed for the average demanding indices:

$$D = \sum_{i=1}^{n} b_i D_i \tag{13}$$

Formula (13) can be rewritten in terms of the MUSA variables z_m and w_{ik} using equations (2) and (10):

$$\frac{\sum_{m=1}^{\alpha-1}\left(100(m-1)-(\alpha-1)\sum_{t=1}^{m-1} z_t\right)}{\alpha(\alpha-1)} = \sum_{i=1}^{n}\frac{\sum_{k=1}^{\alpha_i-1}\left((k-1)\sum_{t=1}^{\alpha_i-1} w_{it}-(\alpha_i-1)\sum_{t=1}^{k-1} w_{it}\right)}{\alpha_i(\alpha_i-1)} \tag{14}$$

Equations (12) and (14) may be easily introduced as additional constraints in the basic LP of the MUSA. However, such constraints should be used carefully, since their form does not guarantee a feasible solution of the LP, especially in case of inconsistencies between global and partial satisfaction judgments. For this reason, these constraints may be rewritten using a goal programming formulation. Using a double error variable for each constraint, the following formulas occur:

$$\begin{cases} \sum_{m=2}^{\alpha} p^m \sum_{t=1}^{m-1} z_t = \sum_{i=1}^{n}\sum_{k=2}^{\alpha_i} p_i^k \sum_{t=1}^{k-1} w_{it} - e_s^+ + e_s^- \\ \frac{1}{\alpha(\alpha-1)}\sum_{m=1}^{\alpha-1}\left(100(m-1)-(\alpha-1)\sum_{t=1}^{m-1} z_t\right) = \\ \sum_{i=1}^{n}\frac{1}{\alpha_i(\alpha_i-1)}\sum_{k=1}^{\alpha_i-1}\left((k-1)\sum_{t=1}^{\alpha_i-1} w_{it}-(\alpha_i-1)\sum_{t=1}^{k-1} w_{it}\right) - e_d^+ + e_d^- \end{cases} \tag{15}$$

where e_s^+ and e_s^- are the respective overestimation and underestimation errors of the constraints on the desired properties of S, and e_d^+ and e_d^- are the respective overestimation and underestimation errors of the constraints on the desired properties of D.

Constraints (15) can be easily modelled in a mathematical programming framework as constraints, minimising simultaneously the sum of the errors e_s^+, e_s^-, e_d^+, and e_d^-.

3.2 Preferences on criteria importance

As already mentioned, a customer satisfaction survey may include, besides the usual performance questions, preferences about the importance of the criteria. Using such questions, customers are asked either to judge the importance of a satisfaction criterion using a predefined ordinal scale, or rank the set of satisfaction criteria according to their importance (Grigoroudis and Siskos, 2010).

In any case, in order to model such importance judgements, each one of the satisfaction criteria should be assigned in a predefined ordered importance category (e.g., very important, important, less important). Assume a set of q such importance classes $C_1, C_2, \ldots, C_q$, where C_1 is the most important criterion class and C_q is the less important criterion class. Considering that C_h, with $h = 1, 2, \ldots, q$, are ordered in a 0-100% scale, there are $q-1$ thresholds, which define the rank and, therefore, label each one of these ordered categories.

Figure 6 presents these thresholds T_h $(h = 1, 2, \ldots, q-1)$ and the aforementioned importance classes.

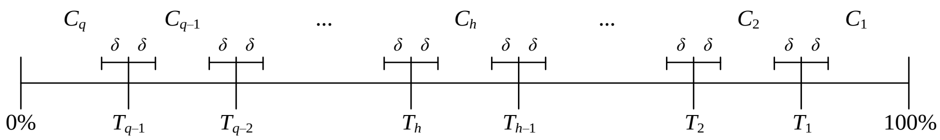


Figure 6. Preference importance classes (Grigoroudis and Spiridaki, 2003)

Based on the previous modelling, the evaluation of the preference importance classes C_h is similar to the estimation of thresholds T_h. An ordinal regression approach may also be used in order to develop the weights estimation model. Using the notation of the MUSA method and assuming that $\hat{b}_{ij}$ is the preference of customer *j* about the importance of criterion *i*, the following cases may occur (Grigoroudis and Spiridaki, 2003):

a. If $\hat{b}_{ij} \in C_1$, that is customer *j* considers criterion *i* as the most important, then:

$$\sum_{k=1}^{\alpha_i - 1} w_{ik} - 100T_1 - \delta + s_{ij}^- \geq 0$$

b. If $\hat{b}_{ij} \in C_h$, that is customer *j* considers criterion *i* in the importance class h $(h = 2, 3, \ldots, q-1)$, then:

$$\begin{cases} \sum_{k=1}^{\alpha_i - 1} w_{ik} - 100T_{h-1} + \delta - s_{ij}^+ \leq 0 \\ \sum_{k=1}^{\alpha_i - 1} w_{ik} - 100T_h - \delta + s_{ij}^- \geq 0 \end{cases}$$

c. If $\hat{b}_{ij} \in C_q$, that is customer *j* considers criterion *i* as the least important, then:

$$\sum_{k=1}^{\alpha_i - 1} w_{ik} - 100T_{q-1} + \delta - s_{ij}^+ \leq 0$$

In the previous formulas, s_{ij}^+ and s_{ij}^- are the overestimation and underestimation error, respectively, for the *j*-th customer and the *i*-th criterion. Also, δ is a small positive number, which is used in order to avoid cases where $b_{ij} = T_h \ \forall h$. In addition, using the MUSA variables, criteria weights are calculated using the following formula:

$$\sum_{k=1}^{\alpha_i-1} w_{ik} = 100 b_i \tag{16}$$

Finally, a minimum value may be assumed for the thresholds T_h in order to increase the discrimination of the importance classes. Thus, the following conditions occur:

$$\begin{cases} T_{q-1} \geq \lambda \\ T_{q-2} - T_{q-1} \geq \lambda \\ \vdots \\ T_1 - T_2 \geq \lambda \end{cases} \tag{17}$$

where λ is a positive number with $\lambda \leq 100/n$, since the maximum value that λ may take cannot exceed the criteria weights (if they are all of equal importance).

Based on the previous notations and assumptions, the estimation of criteria weights may be modelled through a LP which minimises the sum of s_{ij}^{+} and s_{ij}^{-} under the following constraints:

$$\begin{cases} \left. \begin{array}{l} \sum_{k=1}^{\alpha_i-1} w_{ik} - 100T_1 - \delta + s_{ij}^{-} \geq 0 \ \text{ if } \hat{b}_{ij} \in C_1 \\ \left. \begin{array}{l} \sum_{k=1}^{\alpha_i-1} w_{ik} - 100T_{h-1} + \delta - s_{ij}^{+} \leq 0 \\ \sum_{k=1}^{\alpha_i-1} w_{ik} - 100T_h - \delta + s_{ij}^{-} \geq 0 \end{array} \right\} \text{ if } \hat{b}_{ij} \in C_h, \ \ h = 2,3,\ldots,q-1 \\ \sum_{k=1}^{\alpha_i-1} w_{ik} - 100T_{q-1} + \delta - s_{ij}^{+} \leq 0 \ \text{ if } \hat{b}_{ij} \in C_q \end{array} \right\} \forall i, j \\ T_{q-1} \geq \lambda \\ T_{h-1} - T_h \geq \lambda \ \text{ for } \ h = 2,3,\ldots,q-1 \\ w_{ik}, s_{ij}^{+}, s_{ij}^{-} \geq 0 \ \ \forall i, j, k \end{cases} \tag{18}$$

A detailed presentation and discussion of the previous weights estimation model may be found in Grigoroudis and Spiridaki (2003) and Grigoroudis et al. (2004), including some real-world applications.

4 Extension of the MUSA method

Taking into account the previous modelling, which refers to the desired properties of the collective preference system and the customer preferences on the importance of the satisfaction criteria, an extension of the MUSA method may be modelled as a MOLP problem:

$$\begin{cases} [\min] F = \sum_{j=1}^{M} \sigma_j^+ + \sigma_j^- \\ [\min] F_w = \sum_{i=1}^{n} \sum_{j=1}^{M} s_{ij}^+ + s_{ij}^- \\ [\min] F_{sd} = e_s^+ + e_s^- + e_d^+ + e_d^- \\ \text{subject to} \\ \text{all the constraints of LP(3)} \\ \text{constraints (15)} \\ \text{constraints (18)} \\ \sigma_j^+, \sigma_j^-, s_{ij}^+, s_{ij}^-, e_s^+, e_s^-, e_d^+, e_d^- \quad \forall i, j \end{cases} \tag{19}$$

Usually the competitive nature of the multiple objective in MOLP problems does not allow to find a solution that optimises all objective functions. This competiveness may be observed in the proposed model particularly if there are inconsistencies between satisfaction performance and satisfaction importance judgments, as directly expressed by customers.

MOLP (19) may be solved using any MOLP technique (e.g. compromise programming, global criterion approach, etc.). Here, a heuristic approach based on a lexicographic concept is applied, consisting of the following steps.

Step 1

Minimise the sum of the errors σ_j^+ and σ_j^- by solving the following LP:

$$\begin{cases} [\min] F = \sum_{j=1}^{M} \sigma_j^+ + \sigma_j^- \\ \text{subject to} \\ \text{all the constraints of MOLP(19)} \end{cases} \tag{20}$$

Step 2

Minimise the sum of the errors s_{ij}^+ and s_{ij}^- by solving the following LP:

$$\begin{cases} [\min] F_w = \sum_{i=1}^{n} \sum_{j=1}^{M} s_{ij}^+ + s_{ij}^- \\ \text{subject to} \\ F \leq F^* + \varepsilon_1 \\ \text{all the constraints of MOLP(19)} \end{cases} \tag{21}$$

where F^* is the optimal value of the objective function of LP(20) and ε_1 is a small percentage of F^*.

Step 3

Minimise the sum of errors e_s^+, e_s^-, e_d^+, and e_d^- by solving the following LP:

$$\begin{cases} [\min] F_{sd} = e_s^+ + e_s^- + e_d^+ + e_d^- \\ \text{subject to} \\ F \le F^* + \varepsilon_1 \\ F_w \le F_w^* + \varepsilon_2 \\ \text{all the constraints of MOLP(19)} \end{cases} \tag{22}$$

where F_w^* is the optimal value of the objective function of LP(21) and ε_2 is a small percentage of F_w^*.

Step 4

Perform stability analysis by formulating and solving n LPs, where each one maximises the weight of a criterion:

$$\begin{cases} [\max] F' = \sum_{k=1}^{\alpha_i - 1} w_{ik} \quad \text{for } i = 1, 2, \ldots, n \\ \text{subject to} \\ F \le F^* + \varepsilon_1 \\ F_w \le F_w^* + \varepsilon_2 \\ F_{sd} \le F_{sd}^* + \varepsilon_3 \\ \text{all the constraints of MOLP(19)} \end{cases} \tag{23}$$

where F_{sd}^* is the optimal value of the objective function of LP(22) and ε_3 is a small percentage of F_{sd}^*.

Similarly to the basic MUSA method, the final solution is calculated as the average of the optimal solutions given during the post-optimality analysis stage.

The objective function of the basic MUSA model is chosen to be optimised in the first step of the proposed procedure, considering that it is most important to produce a model as consistent as possible with the customers' performance judgments. In the second step, the procedure optimises the objective function of the weights estimation model, implying that the next important optimality criterion refers to inferring a preference model as consistent as possible with the customers' importance judgments. The additional desired properties of the MUSA variables are considered as the less important optimality criterion, thus in the last step, the objective function of the model referring to the desired properties of S and D is chosen to be optimised. This assumed importance of the optimality criteria in the proposed lexicographic approach may be modified. For example, the order of steps 2 and 3 can be reversed, taking into account an alternative importance given to the two objective functions, as well as the stability and fitting level that the produced results achieve.

In any case, it should be emphasised that the main purpose of the proposed extension is to examine whether additional information about the weights of the criteria and additional constraints regarding the desired properties of S and D can improve the results of the MUSA method.

5 Numerical example

An illustrative simple example is presented in this section in order to examine whether the different extensions of the MUSA method can improve the fitting and the stability level of the results.

In this example, it is assumed that 20 customers have participated in a customer satisfaction survey expressing their satisfaction performance and importance judgments. It is also assumed that customers' global satisfaction depends on three main criteria. In addition, it is assumed that customers are asked to express their satisfaction using a 3-level ordinal scale (very satisfied, satisfied, dissatisfied) and their preferences about the importance of the criteria using a similar 3-level ordinal scale (very important, important, unimportant). It should be noted that this dataset is used to illustrate the implementation of the extensions of the MUSA method, although the number of satisfaction criteria, the length of the evaluation scales, and the number of customers are larger in real-world applications.

Table 1 shows the customer performance judgments (globally and per criterion, while importance judgements for the examined set of 20 customers are given Table 2.

Table 1. Customer performance judgments

Customer #	*Global*	*Criterion 1*	*Criterion 2*	*Criterion 3*
1	Satisfied	Satisfied	Very Satisfied	Very Satisfied
2	Satisfied	Dissatisfied	Satisfied	Very Satisfied
3	Satisfied	Dissatisfied	Very Satisfied	Satisfied
4	Very Satisfied	Very Satisfied	Very Satisfied	Very Satisfied
5	Very Satisfied	Very Satisfied	Very Satisfied	Very Satisfied
6	Satisfied	Very Satisfied	Satisfied	Satisfied
7	Satisfied	Satisfied	Very Satisfied	Very Satisfied
8	Satisfied	Satisfied	Very Satisfied	Satisfied
9	Satisfied	Satisfied	Satisfied	Satisfied
10	Dissatisfied	Very Satisfied	Satisfied	Dissatisfied
11	Very Satisfied	Very Satisfied	Very Satisfied	Satisfied
12	Satisfied	Satisfied	Very Satisfied	Dissatisfied
13	Satisfied	Very Satisfied	Satisfied	Dissatisfied
14	Very Satisfied	Very Satisfied	Satisfied	Very Satisfied
15	Dissatisfied	Dissatisfied	Dissatisfied	Dissatisfied
16	Satisfied	Satisfied	Satisfied	Satisfied
17	Very Satisfied	Satisfied	Very Satisfied	Very Satisfied
18	Satisfied	Very Satisfied	Dissatisfied	Satisfied
19	Satisfied	Very Satisfied	Dissatisfied	Very Satisfied

20	Very Satisfied	Very Satisfied	Very Satisfied	Satisfied

In this numerical example, four different variations of the MUSA method are examined:

a. The original MUSA method as given in LP(3), including the post-optimality analysis of LPs(4).
b. The extension of the MUSA (namely $MUSA_{SD}$) that considers the desired properties of S and D, including a post-optimality analysis step, similar to the other variations of the MUSA method. $MUSA_{SD}$ is given by the following MOLP:

$$\begin{cases} [\min]F = \sum_{j=1}^{M} \sigma_j^+ + \sigma_j^- \\ [\min]F_{sd} = e_s^+ + e_s^- + e_d^+ + e_d^- \\ \text{subject to} \\ \text{all the constraints of LP(3)} \\ \text{constraints (15)} \\ \sigma_j^+, \sigma_j^-, e_s^+, e_s^-, e_d^+, e_d^- \quad \forall j \end{cases}$$

This problem is solved using the lexicographic approach proposed in section 4 (omitting step 2).

c. The extension of the MUSA (namely $MUSA_W$) that considers customer preferences on criteria importance, including a post-optimality analysis step, similar to the other variations of the MUSA method. $MUSA_W$ is given by the following MOLP:

$$\begin{cases} [\min]F = \sum_{j=1}^{M} \sigma_j^+ + \sigma_j^- \\ [\min]F_w = \sum_{i=1}^{n} \sum_{j=1}^{M} s_{ij}^+ + s_{ij}^- \\ \text{subject to} \\ \text{all the constraints of LP(3)} \\ \text{constraints (18)} \\ \sigma_j^+, \sigma_j^-, s_{ij}^+, s_{ij}^- \quad \forall i, j \end{cases}$$

Similarly to the previous model, this problem is solved using the lexicographic approach proposed in section 4 (omitting step 3).

d. The extension of the MUSA (namely $MUSA_{SDW}$) that considers customer preferences on criteria importance and the desired properties of S and D as given in MOLP(19) and solved using the lexicographic approach proposed in section 4, including the post-optimality analysis of LPs(23).

Table 3 presents the fitting and stability results of the original MUSA method and its different extensions. As shown therein, there is a significant increase of *ASI* with the introduction of additional constraints regarding special properties for the assessed average indices (+38.07%) or additional information about the importance of the criteria (+57.63%). The stability level of the provided results is improved further when the aforementioned constraints are introduced together in the original MUSA method (+58.28%). These findings are justified by the fact that the incorporation of additional constraints restricts the polyhedron of the feasible solutions, producing more robust results.

On the other hand, there seems to be a decrease of the fitting indices with the introduction of additional constraints. This decrease is rather low, regarding AFI_1 and AFI_3, while it appears to be larger for AFI_2. However, as already mentioned, AFI_2 is a rather strict fitting index, examining only the existence of non-zero errors, without taking into account the values of these error variables. As a result, examining only AFI_2 underestimates in several cases the MUSA's fitting level. Therefore, the large decrease of this index should be examined along with the low decrease reduction of AFI_1 and AFI_3 in order to have a comprehensive view of the fitting ability of the MUSA's extensions.

Table 2. Customer importance judgments

Customer #	*Criterion 1*	*Criterion 2*	*Criterion 3*
1	Important	Important	Important
2	Unimportant	Important	Very important
3	Important	Important	Important
4	Important	Unimportant	Unimportant
5	Important	Important	Very important
6	Very important	Important	Unimportant
7	Very important	Important	Important
8	Important	Important	Very important
9	Very important	Important	Important
10	Very important	Very important	Important
11	Important	Very important	Important
12	Important	Important	Very important
13	Very important	Very important	Unimportant
14	Important	Important	Important
15	Important	Important	Very important
16	Important	Very important	Important
17	Very important	Unimportant	Unimportant
18	Unimportant	Important	Very important
19	Very important	Important	Unimportant
20	Unimportant	Important	Important

Table 3. Fitting and stability comparison results

Index	*Original MUSA method*	*$MUSA_{SD}$ method*	*$MUSA_{W}$ method*	*$MUSA_{SDW}$ method*
AFI_1	91.67%	90.42% (-1.36%)	91.67% (0%)	90.42% (-1.36%)
AFI_2	80.00%	65.00% (-18.75%)	80.00% (0%)	50.00% (-37.50%)
AFI_3	89.58%	88.64% (-1.05%)	89.58% (0%)	88.59% (-1.11%)

ASI	59.60%	82.29% (+38.07%)	93.95% (+57.63%)	94.93% (+59.28%)

The results of Table 3 reveal a competitive relation between the fitting and the stability indices. This is justified by the fact that the introduction of additional constraints in the MUSA method may increase the possibility of inconsistencies (i.e., increase the optimal values of the different error variables), while reducing the polyhedron of feasible solutions (i.e., increase robustness of results). Therefore, searching for more robust solutions with the introduction of additional constraints means that there must be a compromising increase of the original MUSA's overestimation and underestimation errors.

The previous findings regarding the improvement of robustness in the examined extensions of the MUSA model are also confirmed by the variance of weights during the post-optimality analysis step. For example, as shown in Table 4, the weight of the first criterion in the post-optimality analysis of the original MUSA method varies from 0.083 to 0.458 (with an average of 0.271), while this range is significantly reduced in all the examined extensions of the method. Therefore, the final (average) solutions of $MUSA_{SD}$, $MUSA_{W}$, and $MUSA_{SDW}$ are more representative.

Table 4. Variance of weights during post-optimality analysis

Criterion		*Original MUSA method*	*$MUSA_{SD}$ method*	*$MUSA_{W}$ method*	*$MUSA_{SDW}$ method*
1	min	0.083	0.260	0.358	0.337
	max	0.458	0.438	0.375	0.370
	average	0.271	0.319	0.364	0.355
2	min	0.083	0.260	0.283	0.283
	max	0.458	0.438	0.371	0.361
	average	0.271	0.319	0.334	0.324
3	min	0.271	0.271	0.271	0.302
	max	0.833	0.479	0.358	0.358
	average	0.458	0.361	0.302	0.321

Of course, it is possible that different datasets may behave different regarding the increase and decrease of the fitting and stability indices. However, it is expected that the introduction of additional constraints will produce at least as stable results as the ones of the original MUSA method. Furthermore, it is worth noticing that reversing steps 2 and 3 of the proposed heuristic method, (see section 4), may improve stability results. Performing a simulation with different datasets with distinctive characteristics (e.g., number of customers, number of criteria, length of ordinal satisfaction scales, etc.) may further confirm the findings of this study.

6 Concluding remarks

The MUSA method is a rather flexible approach and thus several extensions may be developed taking into account additional information or data. This study evaluates different extensions of the MUSA method with the introduction of additional constraints regarding customer preferences about the importance of the criteria and the desired properties for the global average satisfaction and demanding indices. These extensions are modelled as MOLP problems and a heuristic procedure is proposed for finding a satisfactory solution.

The introduction of additional constraints and information in the original MUSA method seems to significantly improve the stability of the provided results. The observed decrease of the fitting level is very low and thus, the reliability of the value system that is evaluated with the extensions of the MUSA method does not deteriorate.

Since the numerical example presented in this study refers to a small and simple set of customers' data, a simulation model with different datasets having distinctive characteristics may further evaluate the extensions of the MUSA method in terms of fitting and stability. Such simulation results may also study how the parameters of the method (number of customers, number of criteria, etc.) preference or post-optimality thresholds) may affect the stability of the results. Other future research directions include the selection of appropriate values regarding the parameters of the model (e.g., post-optimality analysis thresholds ε or ε_i) and the examination of alternative MOLP techniques. Finally, the development of additional fitting and stability measures may also facilitate the investigation of various extensions of the MUSA method.

Acknowledgment

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program ''Education and Lifelong Learning'' of the National Strategic Reference Framework (NSRF) – Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

References

Angilella, S., S. Corrente, S. Greco, and R. Slowinski (2014). MUSA-INT: Multicriteria customer satisfaction analysis with interacting criteria, *Omega*, 42(1), 189-200.

Corrente, S., S. Greco, and R. Slowinski (2012). Multiple criteria hierarchy process in robust ordinal regression, *Decision Support Systems*, 53(3), 660-674.

Corrente, S., S. Greco, M. Kadzinski, and R. Slowinski (2013). Robust ordinal regression in preference learning and ranking, *Machine Learning*, 93(2-3), 381-422.

Figueira, J.R., S. Greco, and R. Slowinski (2009). Building a set of additive value functions representing a reference preorder and intensities of preference: GRIP method, *European Journal of Operation Research*, 195(2), 460-486.

Greco, S., R. Slowinski, J.R. Figueira, and V. Mousseau (2010). Robust ordinal regression, in: M. Ehrgott, S. Greco and J. Figueira (eds.), *Trends in multiple criteria decision analysis*, Springer, Berlin, 241-283.

Greco, S., V. Mousseau, and R. Slowinski (2008). Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions, *European Journal of Operation Research*, 191(2), 416-436.

Grigoroudis, E. and O. Spiridaki (2003). Derived vs. stated importance in customer satisfaction surveys, *Operational Research: An International Journal*, 3(3), 229-247.

Grigoroudis, E. and Y. Siskos (2002). Preference disaggregation for measuring and analysing customer satisfaction: The MUSA method, *European Journal of Operational Research*, 143(1), 148-170

Grigoroudis, E. and Y. Siskos (2010). *Customer satisfaction evaluation: Methods for measuring and implementing service quality*, Springer, New York.

Grigoroudis, E., Y. Politis, O. Spiridaki, and Y. Siskos (2004). Modelling importance preferences in customer satisfaction surveys, in: C.H. Antunes, J. Figueira, and J. Climaco (eds.), *Proceedings of the 56th Meeting of the European Working Group "Multiple Criteria Decision Aiding"*, INESC Coimbra, 273-291.

Jacquet-Lagrèze, E. and Y. Siskos (1982). Assessing a set of additive utility functions for multicriteria decision-making: The UTA method, *European Journal of Operational Research*, 10(2), 151-164.

Kadzinski, M., S. Greco, and R. Slowinski (2012b). Selection of a representative value function in robust multiple criteria ranking and choice, *European Journal of Operational Research*, 217(3), 541-553.

Kadzinski, M., S. Greco, and R. Slowinski (2013). RUTA: A framework for assessing and selecting additive value functions on the basis of rank related requirements, *Omega*, 41(3), 735-751.

Siskos, J. (1984). Le traitement des solutions quasi-optimales en programmation linéaire continue: Une synthèse, *RAIRO Recherche Opérationnelle*, 18(4), 381-401.

Siskos, J. (1985). Analyses de régression et programmation linéaire, *Révue de Statistique Appliquée*, 23(2), 41-55.

Siskos, Y. and D. Yannacopoulos (1985). UTASTAR: An ordinal regression method for building additive value functions, *Investigaçao Operacional*, 5(1), 39-53.

Siskos, Y. and E. Grigoroudis (2010). New trends in aggregation-disaggregation approaches, in: C. Zopounidis and P.M. Pardalos (eds.), *Handbook of multicriteria analysis*, Springer, Heidelberg, 189-214.

Siskos, Y., E. Grigoroudis, and N.F. Matsatsinis (2005). UTA methods, in: J. Figueira, S. Greco, and M. Ehrgott (eds.), *Multiple criteria analysis: State of the art surveys*, Springer, New York, 297-344.

Siskos, Y., E. Grigoroudis, C. Zopounidis, and O. Saurais (1998). Measuring customer satisfaction using a collective preference disaggregation model, *Journal of Global Optimization*, 12(2), 175-195.

E-government Benchmarking in European Union: A Multicriteria Extreme Ranking Approach

Eleftherios Siskos[1], Michail Malafekas[2], Dimitris Askounis[1], and John Psarras[1]

[1] Decision Support Systems Laboratory, School of Electrical & Computer Engineering, National Technical University of Athens, Greece
{lsiskos,askous,john}@epu.ntua.gr

[2] Decision Support Systems Laboratory, School of Electrical & Computer Engineering, National Technical University of Athens, Greece
mmalafekas@gmail.com

Abstract. E-government benchmarking is being conducted by various organizations but its assessment is based on a limited number of indicators and does not highlight the multidimensional nature of the electronically provided services. This paper outlines a multicriteria evaluation system based on four points of view: (1) infrastructures, (2) investments, (3) e-processes, and (4) users' attitude in order to evaluate European Union countries. In this paper, twenty one European Union countries are evaluated and ranked over their e-government progress. Their ranking is obtained through an additive value model which is assessed by an ordinal regression method and the use of the decision support system MIIDAS. In order to obtain robust evaluations, given the incomplete determination of inter-criteria model parameters, the extreme ranking analysis method, based on powerful mathematical programming techniques, has been applied to estimate each country's best and worst possible ranking position.

Keywords: E-government, Multicriteria analysis, Robustness, Benchmarking, Ordinal regression, Extreme ranking analysis.

1 Introduction

E-government benchmarks are used to assess the progress made by an individual country over a period of time and compare its growth against other countries. A focused assessment of e-government and other initiatives such as e-commerce, e-education, e-health, and e-science is essential if a country is to make substantial progress (Ojo et al., 2007). Benchmarks can have a significant practical impact, both political and potentially economic and can influence the development of e-government services (Bannister, 2007). The results of benchmarking and ranking studies, particularly global projects conducted by international organizations, attract considerable interest from a variety of observers, including governments (ITU, 2009).

Indices and indicators used in benchmarks are generally quantitative in nature, and collectively form a framework for assessment and ranking. Rorissa et al. (2011)

C. Douligeris et al. (Eds.): I3E 2013, IFIP AICT 399, pp. 338–348, 2013.

classified the frameworks to those based on measurable characteristics of the entities; those that use one or more subjective measures; and the rest few that employ a combination of both. They also state that frameworks based on grounded and broadly applicable measures tend to attract fewer criticisms. On the other hand, the frameworks based on subjective measures are liable to controversies and complaints, especially from those countries or institutions who consider their characterization as inaccurate. Therefore, rankings should be supported by well understood and clarified frameworks and indices as well as transparent computational procedures to maximize their acceptability by the governments and the scientific community. Two among others well-known e-government benchmarks still being conducted are those of the United Nations and the European Commission (see United Nations, 2012 and European Commission, 2010 for more details).

Recently, Siskos et al. (2012) proposed a multicriteria evaluation system based on four points of view: (1) infrastructures, (2) investments, (3) e-processes, and (4) users' attitude from which eight criteria are modeled (see below) to evaluate twenty one European Union countries for which the related data are available. Their ranking is obtained through an additive value model which is assessed by an ordinal regression method. The whole approach consists of helping decision makers (DM), experts or potential evaluators in determining their own country evaluations, based on their own value systems and their own ways of preferring, in order to propose alternative evaluation solutions in contrast to standard published benchmarks.

The aim of this paper is to extend this decision support methodology to take into account the incomplete determination of inter-criteria model parameters and to obtain a robust evaluation of the countries. These targets were achieved with the aid of the extreme ranking analysis (Kadzinski et al., 2012) that estimates each country's best and worst possible rank. This method is based on powerful mathematical programming techniques.

The paper is organized as follows: In section 2 the consistent family of criteria is briefly outlined. Section 3 presents the assessment of the multicriteria evaluation model for a single specific decision maker-evaluator while section 4 presents the extreme ranking analysis method and the obtained results. Finally section 5 concludes the paper.

2 Multicriteria Benchmarking Modeling

In order to achieve an overall assessment of global e-government, a consistent family of criteria was built according to the classical modeling methodology of Roy (1985), in the following way:

Infrastructure Criteria

g_1: Access to the web. This criterion expresses the percentage of households and businesses that have access to the web by any means.

g_2: Broadband internet connection. It shows the percentage of each country's households and businesses with a fixed broadband internet connection.

Investments Criterion

g_3: Percentage of gross domestic product (GDP) spent on information & communications technology (ICT) and research & development (R&D). ICT and R&D expenditure data were retrieved from Eurostat and the International Monetary Fund (IMF), respectively.

E-Processes Criteria

g_4: Online sophistication. It shows each country's maturity on online service delivery. The data composing this criterion stem from the European Commission's 9th Benchmark Measurement published in 2010 (Digitizing Public Services in Europe: Putting ambition into action).

g_5: E-participation. It expresses the interaction achieved between governments and citizens in a manner of information sharing, e-consultation and e-decision making. The source of the e-participation criterion is the United Nation's survey on e-government published in 2012.

Users' Attitude Criteria

g_6: Citizens' online interaction with authorities. This criterion indicates the percentage of citizens that are already using the web to interact with the authorities.

g_7. Businesses' online interaction with authorities. It indicates the percentage of businesses that are using the web to interact with the authorities.

g_8: Users' experience. This criterion expresses citizens' experience over the 20 e-services and the national portal. The data composing this criterion stem from the European Commission's 9th Benchmark Measurement.

Table 1. Criteria evaluation scales and sources

Criterion	Index	Worst level	Best level	Source
g_1	% population	0	100	Eurostat
g_2	% population	0	100	Eurostat
g_3	% GDP	0	5	Eurostat and IMF
g_4	%	0	100	European Commission
g_5	index [0-1]	0	1	United Nations
g_6	% citizens	0	100	Eurostat
g_7	% businesses	0	100	Eurostat
g_8	% index	0	100	European Commission

Table 2. Multicriteria evaluation of twenty one European countries

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
Belgium	86.5	81.5	3.3	92	0.59	28	15	68
Czech Rep.	94	90	3.3	85	0.13	35	47	22
Denmark	83.5	85.5	3.8	95	0.64	11	31	30
Germany	71.5	80.5	3.8	99	0.61	16	26	55
Estonia	86	78	4.0	97	0.69	35	18	21
Ireland	77	97	3.5	100	0.44	40	33	49
Greece	86.5	80.5	2.4	70	0.26	57	62	33
Spain	76.5	85	3.1	98	0.83	75	74	86
France	84	92	3.7	94	0.60	65	83	83
Italy	93.5	94.5	2.7	99	0.21	84	67	85
Hungary	90	80	3.4	80	0.31	57	79	80
Netherlands	75	87.5	3.5	97	0.60	80	83	80
Austria	83	78	3.3	100	0.50	85	87	90
Poland	77.5	60.5	2.7	87	0.24	88	81	81
Portugal	79	81	3.8	100	0.27	79	59.5	75.5
Slovenia	68	72.5	3.4	99	0.51	91	87	68.5
Slovakia	87	77	3.1	81	0.07	91	93	94.5
Finland	67	70	4.3	96	0.41	51	80	93.5
Sweden	79.5	65.5	4.1	99	0.49	70	88	92.5
Norway	88.5	90	2.5	92	0.50	83.5	89	87
Un. Kingdom	83.5	87.5	4,3	97	0,77	83.5	76.5	91

All details about the criteria construction techniques are thoroughly described in Siskos et al. (2012). Tables 1 and 2 present the criteria evaluation scales and the criteria scores achieved by the twenty one European countries, respectively.

3 Assessing an Overall Evaluation Model

The main target of the proposed methodological frame is the assessment of a multicriteria additive value system, for a single decision maker, that is described by the following formulae:

$$u(\boldsymbol{g}) = \sum_{i=1}^{n} p_i u_i(g_i) \tag{1}$$

$$u_i(g_{i*}) = 0, u_i(g_i^*) = 1, \text{ for } i = 1,2, \dots, n \tag{2}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{3}$$

$$p_i \geq 0 \text{ , for } i = 1,2, \dots, n \tag{4}$$

where $\boldsymbol{g} = (g_1, g_2, \dots, g_n)$ is the performance vector of a country on the n criteria; g_{i*} and g_i^* are the least and most preferable levels of the criterion g_i, respectively; $u_i(g_i)$, $i = 1,2,\dots,n$ are non decreasing marginal value functions of the performances g_i, $i = 1,2,\dots,n$; and p_i is the relative weight of the i-th function $u_i(g_i)$. Thus, for a given country a, $\boldsymbol{g}(a)$ and $u[\boldsymbol{g}(a)]$ represent the multicriteria vector of performances and the global value of the country a respectively.

Both the marginal and the global value functions have the monotonicity property of the true criterion. For instance, in the case of the global value function, given two countries a and b the following properties hold:

$$u[\boldsymbol{g}(a)] > u[\boldsymbol{g}(b)] \Leftrightarrow a\ P\ b \quad \text{(Preference)} \tag{5}$$

$$u[\boldsymbol{g}(a)] = u[\boldsymbol{g}(b)] \Leftrightarrow a\ I\ b \quad \text{(Indifference)} \tag{6}$$

The necessary hypothesis to validate an additive value function for a given decision maker (DM) is the preferential independence of all the criteria (see Keenney and Raiffa, 1976, Keeney, 1992 for instance). A pair of criteria (g_i , g_j) is preferentially independent from the rest of the criteria when the trade-offs between the g_i and g_j criteria are not dependent on the values of the rest of the criteria. All criteria are supposed to be preferentially independent when the same condition holds for all pairs of criteria. When the u_i functions in formula (1) are already assessed, the linear model (1)-(4) exists if and only if the inter-criteria parameters (weighs) p_i are constant substitution rates (value trade-offs) between u_i.

This value system can be obtained utilizing various methods (see Keeney and Raiffa, 1976, Keeney, 1992, Figueira et al., 2005). Because of the objective difficulties to convince decision makers in externalizing tradeoffs between heterogeneous criteria and verify the preferential conditions cited above, decision analysts usually prefer to infer the DM's additive value function from global preference structures, by applying disaggregation or ordinal regression methods (see Jacquet-Lagrèze and Siskos, 1982, 2001, Greco et al., 2008, 2010).

In this study the disaggregation UTA II method is implemented by assessing the additive model (1)-(4) in two phases: In the first phase the expert is asked to assign some value points $u_i(g_i)$ of the corresponding evaluation scale for every criterion separately. Then, each marginal value function is optimally fitted (see Fig. 1) and accepted by the DM. In a second phase, the criteria weights p_i are estimated using inference procedures (see next section).

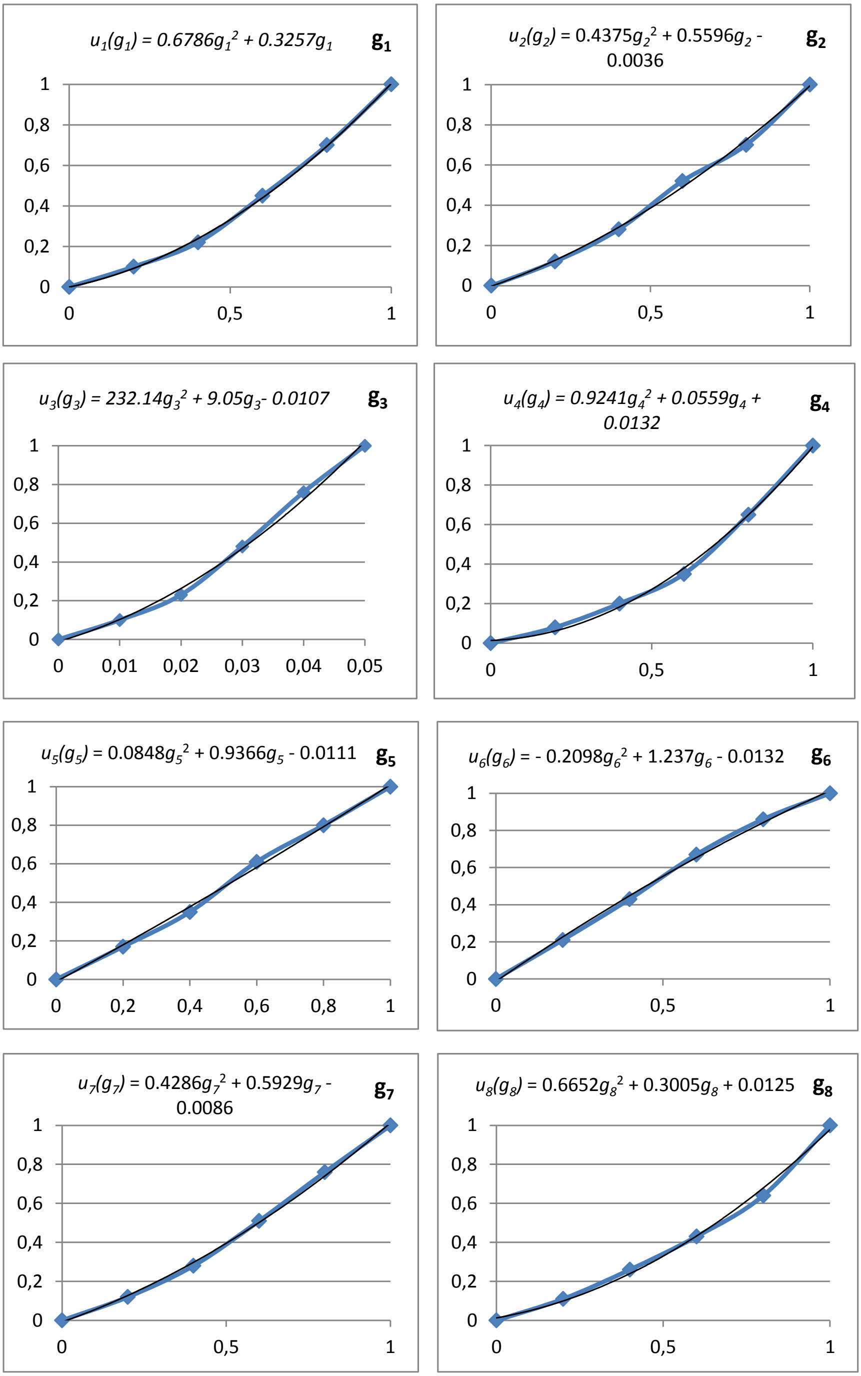


Fig. 1. Marginal value functions of the eight criteria

4 Estimation of Weights and Extreme Ranking Analysis

Through this phase the UTA II disaggregation procedure is used to infer the inter-criteria parameters p_i, $i = 1,2,\dots,n$. Specifically, the DM-evaluator is asked to give a ranking (weak order) on a set of reference countries $A_r = (a_1, a_2, \dots, a_k)$, that are fictitious country profiles differing on two or at most three criteria values. The reference countries are ranked by the DM in such a way that a_1 is the head and a_k the tail of the ranking.

Therefore, for every pair of consecutive countries (a_m, a_{m+1}) holds, either $a_m P a_{m+1}$ (preference of a_m) or $a_m I a_{m+1}$ (indifference). UTA II solves the linear program (7)-(11) below that has k constraints because of the transitivity of the (P, I) preference system.

$$[min]F,\ F = \sum_{\iota=1}^{k}\left(\sigma^+(\alpha_\iota) + \sigma^-(\alpha_\iota)\right) \tag{7}$$

Subject to:

for $m = 1, 2, \dots, k-1$

$$\sum_{i=1}^{n} p_i\, u_i[g_i(\alpha_m)] - \sigma^+(\alpha_m) + \sigma^-(\alpha_m) - [\sum_{i=1}^{n} p_i\, u_i[g_i(a_{m+1})] - \sigma^+(a_{m+1}) + \sigma^-(a_{m+1})] \geq \delta,\ \text{if } a_m \text{ P } a_{m+1} \tag{8}$$

or

$$= 0 \quad \text{if } a_m \text{ I } a_{m+1} \tag{9}$$

$$\sum_{\iota=1}^{n} p_i = 1 \tag{10}$$

$$p_i \geq 0, for\ \ i = 1,2,\dots,n;\ \sigma^+(a_j) \geq 0, \sigma^-(a_j) \geq 0, for\ j = 2,\dots,k \tag{11}$$

where δ is a small positive number, equal to 0.001 for instance; $g_i(a_m)$ is the evaluation of the a_m country on the i-th criterion and $u_i[g_i(a_m)]$ its corresponding marginal value; $\sigma^+(a_j)$, $\sigma^-(a_j)$ are the over-estimation and the under-estimation errors concerning the j-th country's position, respectively.

This technique was applied for a set of thirteen country profiles and a zero error sum was obtained ($F = 0$). The optimally most characteristic weighting factors are reported in formula (12) while the corresponding ranking of the European countries is presented in Table 3. More details about UTA II illustration is given in Siskos et al. (2012).

$$u(\boldsymbol{g}) = 0.1276u_1(g_1) + 0.1607u_2(g_2) + 0.1097u_3(g_3) + 0.2579u_4(g_4) + 0.0743u_5(g_5) + 0.1209u_6(g_6) + 0.0536u_7(g_7) + 0.0952u_8(g_8) \tag{12}$$

$\boldsymbol{g} = (g_1, g_2, \dots, g_8)$, is the performance vector of a country on the eight criteria.

Table 3. E-government ranking of the twenty one European countries

Rank position	European Country	Global Value
1	Sweden	0.825
2	Denmark	0.821
3-4	Finland	0.796
3-4	Netherlands	0.796
5	Norway	0.765
6-7	Germany	0.745
6-7	United Kingdom	0.744
8	France	0.738
9	Estonia	0.729
10-11	Austria	0.701
10-11	Slovenia	0.701
12	Spain	0.693
13	Belgium	0.686
14	Ireland	0.679
15	Portugal	0.633
16	Czech Republic	0.582
17	Slovakia	0.578
18	Hungary	0.568
19	Poland	0.548
20	Italy	0.533
21	Greece	0.467

However, the above estimation procedure bears robustness issues. In fact, there exists an infinite number of weighting vectors that are optimally consistent with the whole set of constraints (8)-(11). In order to study the impact of this indetermination on the ranking of the countries the extreme ranking analysis of Kadzinski et al. (2012) has been applied with the aid of the GAMS platform.

The extreme ranking analysis algorithm examines each country individually and estimates the best and worst possible rank it can achieve. The methodology leading to the estimation of the best possible rank of each country is outlined below.

In order to determine the best possible rank of a country A, taking into consideration all the possible combinations of the criteria weighting factors, the number N_A^* of the countries that surpass country A in the ranking under any circumstances is calculated. The countries that surpass country A in the ranking for a limited number of combinations of the criteria weights are not included in the N_A^* set. Therefore, the best possible ranking position that can be achieved by the country A is $N_A^* + 1$.

Thus, the problem is reduced to the calculation of the N_A^* set for each individual country. This set is calculated through the modeling and the solution of the mixed integer programming problem presented below:

$$[min]F = \sum_{b \in A \setminus \{a\}} u_b \tag{13}$$

Subject to:

$$\text{Constraints (8) – (11)} \tag{14}$$

$$U(\alpha) \geq U(b) - Mu_b, \forall\ b \in A \setminus \{\alpha\} \tag{15}$$

where M is an auxiliary variable equal to a big positive value, and u_b is a binary variable associated with comparison of the country *A* to another country *B*. There exist $N - 1$ such variables, each corresponding to $b \in A \setminus \{\alpha\}$. N is the total number of the countries under evaluation, i.e. 21.

The determination of the worst possible ranking of a country *A*, requires a similar procedure. In this case, however, it is estimated the number of countries N_{A*} that achieve a worse ranking position for all possible combinations of the weighting factors. Therefore, the worst possible rank a country *A* can achieve is $N - N_{A*}$.

N_{A*} set is calculated through the solution of the integer programming problem outlined below:

$$[min]F = \sum_{b \in A \setminus \{a\}} u_b \tag{16}$$

Subject to:

$$\text{Constraints (8) - (11)} \tag{17}$$

$$U(b) \geq U(\alpha) - Mu_b, \forall\ b \in A \setminus \{\alpha\} \tag{18}$$

The extreme ranking positions of the 21 European countries assessed, are graphically presented in Fig. 2.

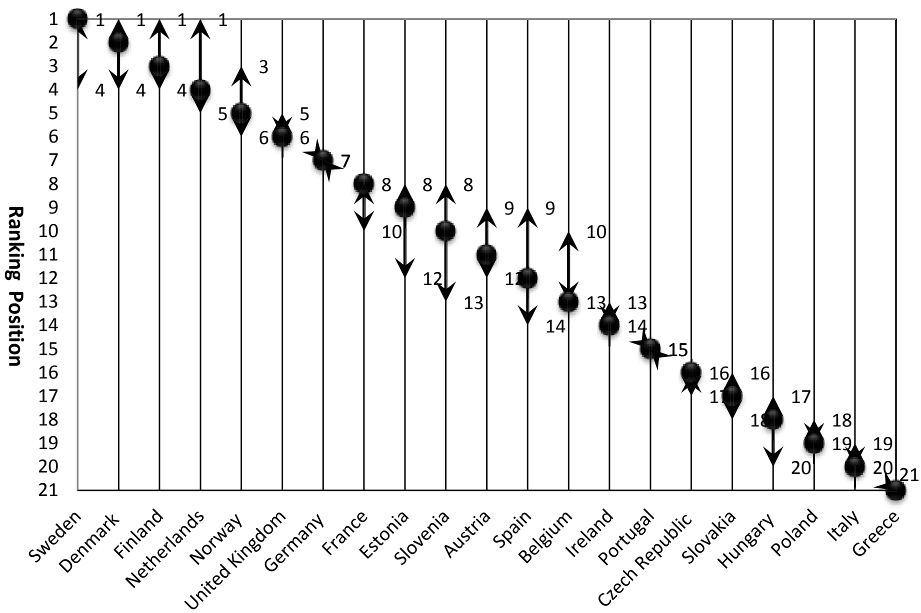


Fig. 2. Extreme ranking positions of twenty one European countries

5 Conclusion

The paper addressed the robust assessment of global e-government based on multiple criteria and special extreme ranking procedures. The proposed approach focused on the evaluation of European countries according to the standards of a benchmark.

The e-government evaluation process is an independent procedure enabling each individual to specify his (her) own preferences on criteria value functions and weights, and results in a personalized ranking of the countries. In other words, each evaluator has control over his (her) set of criteria and the assessment of his (her) own evaluation model. The proposed multicriteria techniques offer the possibility to combine different preferences and considerations of multiple decision makers and merge them easily through interactive iterative processes.

The next research steps include the development of robustness control procedures based on cardinal and visualization measures as well as the development of a decision support system aiding anyone to form his own e-government benchmarking.

Acknowledgements. This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

References

1. Bannister, F.: The curse of the benchmark: an assessment of the validity and value of e-government comparisons. International Review of Administrative Sciences 73(2), 171–188 (2007)
2. European Commission: Digitizing public services in Europe: Putting ambition into action, 9th benchmark measurement, Directorate General Information Society and Media, Brussels (2010)
3. Figueira, J., Greco, S., Ehrgott, M.: State-of-Art of Multiple Criteria Decision Analysis. Kluwer Academic Publishers, Dortrecht (2005)
4. Greco, S., Mousseau, V., Słowiński, R.: Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions. European Journal of Operational Research 191(2), 416–436 (2008)
5. Greco, S., Słowiński, R., Figueira, J., Mousseau, V.: Robust ordinal regression. In: Ehrgott, M., Figueira, J., Greco, S. (eds.) Trends in Multiple Criteria Decision Analysis, ch. 9, pp. 241–283. Springer, New York (2010)
6. ITU: Measuring the information society: The ICT development index (2009), `http://www.itu.int/ITU-D/ict/publications/idi/2009/material/IDI2009_w5.pdf` (retrieved November 10, 2012)
7. Jacquet-Lagrèze, E., Siskos, J.: Assessing a set of additive utility functions for multicriteria decision making. European Journal of Operational Research 10(2), 151–164 (1982)
8. Jacquet-Lagrèze, E., Siskos, J.: Preference disaggregation: 20 years of MCDA experience. European Journal of Operational Research 130, 233–245 (2001)
9. Kadzinski, M., Greco, S., Slowinski, R.: Extreme ranking analysis in robust ordinal regression. Omega 40, 488–501 (2012)
10. Keeney, R.L.: Value-focused thinking: A path to creative decision making. Harvard U.P., London (1992)
11. Keeney, R.L., Raiffa, H.: Decisions with multiple objectives: Preferences and value trade-offs. John Wiley and Sons, New York (1976)
12. Ojo, A., Janowski, T., Estevez, E.: Determining progress towards e-Government - What are the core indicators? (2007), `http://www.iist.unu.edu/newrh/III/1/docs/techreports/report360.pdf` (retrieved November 10, 2012)
13. Rorissa, A., Demissie, D., Pardo, T.: Benchmarking e-Government: A comparison of frameworks for computing e-Government index and ranking. Government Information Quarterly 28, 354–362 (2011)
14. Roy, B.: Méthodologie multicritère d'aide à la décision. Economica, Paris (1985)
15. Siskos, E., Askounis, D., Psarras, J.: Multicriteria decision support for global e-government evaluation, Working paper, Decision Support Systems Laboratory, School of Electrical & Computer Engineering. National Technical University of Athens, Greece (2012)
16. United Nations: E-government survey 2012, E-government for the people (2012), `http://www2.unpan.org/egovkb/global_reports/12report.htm` (retrieved November 10, 2012)

The Robustness Concern in Preference Disaggregation Approaches for Decision Aiding: An Overview

Michael Doumpos and Constantin Zopounidis

Abstract In multiple criteria decision aid, preference disaggregation techniques are used to facilitate the construction of decision models, through regression-based approaches that enable the elicitation of preferential information from a representative set of decision examples provided by a decision-maker. The robustness of such approaches and their results is an important feature for their successful implementation in practice. In this chapter we discuss the robustness concern in this context, overview the main methodologies that have been recently developed to obtain robust recommendations from disaggregation techniques, and analyze the connections with statistical learning theory, which is also involved with inferring models from data.

1 Introduction

Managers, analysts, policy makers, and regulators are often facing multiple technical, socio-economic, and environmental objectives, goals, criteria, and constraints, in a complex and ill-structured decision making framework, encountered in all aspects of the daily operation of firms, organizations, and public entities. Coping with such a diverse and conflicting set of decision factors poses a significant burden to the decision process when ad-hoc empirical procedures are employed.

Multiple criteria decision aid (MCDA) has evolved into a major discipline in operations research/management science, which is well-suited for problem structuring,

Michael Doumpos
Technical University of Crete, Dept. of Production Engineering and Management, University Campus, 73100 Chania, Greece, e-mail: mdoumpos@dpem.tuc.gr

Constantin Zopounidis
Technical University of Crete, Dept. of Production Engineering and Management, University Campus, 73100 Chania, Greece & Audencia Group, Nantes School of Management, France e-mail: kostas@dpem.tuc.gr

In: Rassias, Th.M., Floudas, C.A., and Butenko, S. (2014), Optimization in Science & Engineering, Springer (forthcoming).

modeling, and analysis in this context. MCDA provides a wide arsenal of methodologies and techniques that enable the systematic treatment of decision problems under multiple criteria, in a rigorous yet flexible manner, taking into consideration the expertise, preferences, and judgment policy of the decision makers (DMs) involved. The MCDA framework is applicable in a wide range of different types of decision problems, including deterministic and stochastic problems, static and dynamic problems, as well as in situations that require the consideration of fuzzy and qualitative data of either small or large scale, by a single DM or a group of DMs. A comprehensive overview of the recent advances in the theory and practice of MCDA can be found in the book of Zopounidis and Pardalos [68].

Similarly to other OR and management science modeling approaches, MCDA techniques are also based on assumptions and estimates on the characteristics of the problem, the aggregation of the decision criteria, and the preferential system of the DM. Naturally, such assumptions and estimates incorporate uncertainties, fuzziness, and errors, which affect the results and recommendations provided to the DM. As a result, changes in the decision context, the available data, or a reconsideration of the decision criteria and the goals of the analysis, may ultimately require a very different modeling approach leading to completely different outputs. Thus, even if the results may be judged satisfactory when modeling and analyzing the problem, their actual implementation in practice often leads to new challenges not taken previously into consideration.

In this context, robustness analysis has emerged as a major research issue in MCDA. Robustness analysis seeks to address the above issues through the introduction of a new modeling paradigm based on the idea that the multicriteria problem structuring and criteria aggregation process should not be considered in the context of a well-defined, strict set of conditions, assumptions, and estimates, but rather to seek to provide satisfactory outcomes even in cases where the decision context is altered.

Vincke [61] emphasized that robustness should not be considered in the restrictive framework of stochastic analysis (see also [34] for a discussion in the context of discrete optimization) and distinguished between robust solutions and robust methods. He further argued that although robustness is an appealing property, it is not a sufficient condition to judge the quality of a method or a solution. Roy [45], on the other hand, introduced the term *robustness concern* to emphasize that robustness is taken into consideration a priori rather than a posteriori (as is the case of sensitivity analysis). In the framework of Roy, the robustness concern is raised by *vague approximations* and *zones of ignorance* that cause the formal representation of a problem to diverge from the real-life context, due to: (i) the way imperfect knowledge is treated, (ii) the inappropriate preferential interpretation of certain types of data (e.g., transformations of qualitative attributes), (iii) the use of modeling parameters to grasp complex aspects of reality, and (iv) the introduction of technical parameters with no concrete meaning. An recent example of robustness in the context of multiobjective linear programming can be found in Georgiev et al. [18]. The framework for robust decision aid has some differences compared to the traditional approach to robustness often encounter in other OR areas. A discussion of these differences (and similarities) can be found in Hites et al. [28].

The robustness concern is particularly important in the context of the preference disaggregation approach of MCDA, which is involved with the inference of preferential information and decision models from data. Disaggregation techniques are widely used to facilitate the construction of multicriteria evaluation models, based on simple information that can the DM can provide [30], without requiring the specification of complex parameters whose concept is not clearly understood by the DMs. In this chapter we provide an overview of the robustness concern in the preference disaggregation context, covering the issues and factors that affect the robustness of disaggregation methods, the approaches that have been proposed to deal with robustness in this area, and the existing connections with concepts and methodologies from the area of statistical learning.

The rest of the chapter is organized as follows. Section 2 presents the context of preference disaggregation analysis with examples from ordinal regression and classification problems. Section 3 discusses the concept of robustness in disaggregation methods and some factors that affect it, whereas section 4 overviews the different approaches that have been proposed to obtain robust recommendations and models in preference disaggregation analysis. Section 5 presents the statistical learning perspective and discusses its connections to the MCDA disaggregation framework. Finally, section 6 concludes the chapter and proposes some future research directions.

2 Preference Disaggregation Analysis

2.1 General Framework

A wide class of MCDA problems requires the evaluation of a discrete set of alternatives (i.e., ways of actions, options) $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \}$ described on the basis of n evaluation criteria. The DM may be interested in choosing the best alternatives, ranking the alternatives from the best to the worst, or classifying them into predefined performance categories.

In this context, the construction of an evaluation model that aggregates the performance criteria and provides recommendations in one of the above forms, requires some preferential information by the DM (e.g., the relative importance of the criteria). This information can be specified either through interactive, structured communication sessions between the analyst and the DM or it can be inferred from a sample of representative decision examples provided by the DM. Preference disaggregation analysis (PDA) adopts the latter approach, which is very convenient in situations where, due to cognitive or time limitations, the DM is unwilling or unable to provide the analyst with specific information on a number of technical parameters (which are often difficult to understand) required to formulate the evaluation model.

PDA provides a general methodological framework for the development of multicriteria evaluation models using examples of decisions taken by a DM (or a group of DMs), so that DM's system of preferences is represented in the models as accurately as possible. The main input used in this process is a reference set of alternatives evaluated by the DM (decision examples). The reference set may consist of past decisions, a subset of the alternatives under consideration, or a set of fictitious alternatives which can be easily judged by the DM [30]. Depending on the decision problematic, the evaluation of the reference alternatives may be expressed by defining an order structure (total, weak, partial, etc.) or by classifying them into appropriate classes.

Formally, let $\mathcal{D}(X')$ denote the DM's evaluation of a set X' consisting of m reference alternatives described over n criteria (the description of alternative i on criterion k will henceforth be denoted by x_{ik}). The DM's evaluation is assumed to be based (implicitly) on a decision model f_β defined by some parameters β, which represent the actual preferential system of the DM. Different classes of models can be considered. Typical examples include:

- Value functions defined such that $V(\mathbf{x}_i) > V(\mathbf{x}_j)$ if alternative i is preferred over alternative j and $V(\mathbf{x}_i) = V(\mathbf{x}_j)$ in cases of indifference [33]. The parameters of a value function model involve the criteria trade-offs and the form of the marginal value functions.
- Outranking relations defined such that $\mathbf{x}_i S \mathbf{x}_j$ if alternative i is at least as good as alternative j. The parameters of an outranking model, may involve the weights of the criteria, as well as preference, indifference and veto thresholds, etc. (for details see [44, 60]).
- "If ... then ..." decision rules [21]. In this case the parameters of the model involve the conditions and the conclusions associated to each rule.

The objective of PDA is to infer the "optimal" parameters $\widehat{\beta}^*$ that approximate, as accurately as possible, the actual preferential system of the DM as represented in the unknown set of parameters β, i.e.:

$$\widehat{\beta}^* = \arg\min_{\widehat{\beta} \in \mathcal{A}} \|\widehat{\beta} - \beta\| \tag{1}$$

where $\mathcal{A}$ is a set of feasible values for the parameters $\widehat{\beta}$. With the obtained parameters, the evaluations performed with the corresponding decision model $f_{\widehat{\beta}^*}$ will be consistent with the evaluations actually performed by the DM for any set of alternatives.

Problem (1), however, cannot be solved explicitly because β is unknown. Instead, an empirical estimation approach is employed using the DM's evaluation of the reference alternatives to proxy β. Thus, the general form of the optimization problem is expressed as follows:

$$\widehat{\beta}^* = \arg\min_{\widehat{\beta} \in \mathcal{A}} L[\mathcal{D}(X'), \widehat{\mathcal{D}}(X')] \tag{2}$$

where $\widehat{\mathcal{D}}(X')$ denotes the recommendations of the model $f_{\widehat{\beta}}$ for the alternatives in X' and $L(\cdot)$ is a function that measures the differences between $\mathcal{D}(X')$ and $\widehat{\mathcal{D}}(X')$.

2.2 Inferring Value Function Models for Ordinal Regression and Classification Problems

The general framework of PDA is materialized in several MCDA methods that enable the development of decision models in different forms [14,50,67]. To facilitate the exposition we shall focus on functional models expressed in the form of additive value functions, which have been widely used in MCDA.

A general multiattribute value function aggregates all the criteria into an overall performance index V (global value) defined such that:

$$\begin{aligned} V(\mathbf{x}_i) > V(\mathbf{x}_j) \Leftrightarrow \mathbf{x}_i \succ \mathbf{x}_j \\ V(\mathbf{x}_i) = V(\mathbf{x}_j) \Leftrightarrow \mathbf{x}_i \sim \mathbf{x}_j \end{aligned} \tag{3}$$

where $\succ$ and $\sim$ denote the preference and indifference relations, respectively. A value function may expressed in different forms, depending on the criteria independence conditions [33]. Due to its simplicity, the most widely used form of value function is the additive one:

$$V(\mathbf{x}_i) = \sum_{k=1}^{n} w_k v_k(x_{ik}) \tag{4}$$

where w_k is the (non-negative) trade-off constant of criterion k (the trade-offs are often normalized to sum up to one) and $v_k(\cdot)$ is the marginal value functions of the criterion, usually scaled such that $v_k(x_{k*}) = 0$ and $v_k(x_k^*) = 1$, where x_{k*} and x_k^* are the least and the most preferred levels of criterion k, respectively.

Such a model can be used to rank a set of alternatives or to classify them in predefined groups. In the ranking case, the relationships (3) provide a straightforward way to compare the alternatives. In the classification case, the simplest approach is to define an ordinal set of groups $G_1, G_2, \ldots, G_q$ on the value scale with the following rule:

$$t_\ell < V(\mathbf{x}_i) < t_{\ell-1} \Leftrightarrow \mathbf{x}_i \in G_\ell \tag{5}$$

where $t_1 > t_2 \cdots > t_{q-1}$ are thresholds that distinguish the groups. Alternative classification rules can also be employed such as the example-based approach of Greco et al. [23] or the hierarchical model of Zopounidis and Doumpos [66].

The construction of a value function from a set of reference examples can be performed with mathematical programming formulations. For example, in an ordinal regression setting, the DM's defines a weak-order of the alternatives in the reference set, by ranking them from the best (alternative $\mathbf{x}_1$) to the worst one (alternative $\mathbf{x}_m$). Then, the general form of the optimization problem for inferring a decision model from the data can be expressed as in the case of the UTA method [29] as follows:

$$
\begin{aligned}
\min \quad & \sigma_1+\sigma_2+\cdots+\sigma_m \\
\text{s.t.} \quad & \sum_{k=1}^{n} w_k[v_k(x_{ik})-v_k(x_{i+1,k})]+\sigma_i-\sigma_{i+1}\geq\delta && \forall \mathbf{x}_i \succ \mathbf{x}_{i+1} \\
& \sum_{k=1}^{n} w_k[v_k(x_{ik})-v_k(x_{i+1,k})]+\sigma_i-\sigma_{i+1}=0 && \forall \mathbf{x}_i \sim \mathbf{x}_{i+1} \\
& w_1+w_2+\ldots+w_n=1 \\
& v_k(x_{ik})-v_k(x_{jk})\geq 0 && \forall x_{ik}\geq x_{jk} \\
& v_k(x_{k*})=0,\, v_k(x_k^*)=1 && k=1,\ldots,n \\
& w_k,\, v_k(x_{ik}),\, \sigma_i\geq 0, && \forall i,k
\end{aligned}
\tag{6}
$$

where $\mathbf{x}^*=(x_1^*,\ldots,x_n^*)$ and $\mathbf{x}_*=(x_{1*},\ldots,x_{n*})$ represent the ideal and anti-ideal alternatives, respectively. The solution of this optimization problem provides a value function that reproduces the DM's ranking of the reference alternatives as accurately as possible. The differences between the model's recommendations and the DM's weak-order are measured by the error variables $\sigma_1,\ldots,\sigma_m$, which are defined through the first two constraints (with δ being a small positive constant). The third constraint normalizes the trade-off constants, whereas the fourth constraint ensures that the marginal value functions and non-decreasing (assuming that the criteria are expressed in maximization form).

For classification problems, the optimization formulation for inferring a classification model from the reference examples using the threshold-based rule (5) can be expressed as follows:

$$
\begin{aligned}
\min \quad & \sum_{\ell=1}^{q}\frac{1}{m_\ell}\sum_{\mathbf{x}_i\in G_\ell}(\sigma_i^+ + \sigma_i^-) \\
\text{s.t.} \quad & \sum_{k=1}^{n} w_k v_k(x_{ik})+\sigma_i^+\geq t_\ell+\delta && \forall \mathbf{x}_i\in G_\ell,\, \ell=1,\ldots,q-1 \\
& \sum_{k=1}^{n} w_k v_k(x_{ik})-\sigma_i^-\leq t_\ell-\delta && \forall \mathbf{x}_i\in G_\ell,\, \ell=2,\ldots,q \\
& t_\ell-t_{\ell+1}\geq\varepsilon && \ell=1,\ldots,q-2 \\
& w_1+w_2+\ldots+w_n=1 \\
& v_k(x_{ik})-v_k(x_{jk})\geq 0 && \forall x_{ik}\geq x_{jk} \\
& v_k(x_{k*})=0,\, v_k(x_k^*)=1 && k=1,\ldots,n \\
& w_k,\, \sigma_i^+,\, \sigma_i^-\geq 0 && \forall i,\, k
\end{aligned}
\tag{7}
$$

The objective function minimizes the total weighted classification error, where the weights are defined on the basis of the number of reference alternatives from each class ($m_1,\ldots,m_q$). The error variables σ^+ and σ^- are defined through the first two constraints as the magnitude of the violations of the classification rules, whereas the third constraint ensures that the class thresholds are non-increasing (with ε being a small positive constant).

For the case of an additive value function, the above optimization problems can be re-expressed in linear programming form with a piece-wise linear modeling of the marginal values function (see for example [29]).

3 Robustness in Preference Disaggregation Approaches

The quality of models resulting from disaggregation techniques is usually described in terms of their accuracy, which can be defined as the level of agreement between the DM's evaluations and the outputs of the inferred model. For instance, in ordinal regression problems rank correlation coefficients (e.g., the Kendall's τ or Spearman's ρ) can be used for this purpose, whereas in classification problems the classification accuracy rate and the area under the receiver operating characteristic curve are commonly used measures. Except for accuracy-related measures, however, the robustness of the inferred model is also an crucial feature. Recent experimental studies have shown that robustness and accuracy are closely related [59]. However, accuracy measurements are done ex-post and rely on the use of additional test data, while robustness is taken into consideration ex-ante, thus making it an important issue that is taken into consideration before a decision model is actually put into practical use.

The robustness concern in the context of PDA arises because in most cases multiple alternative decision models can be inferred in accordance with the information embodied in the set of reference decision examples that a DM provides. This is particularly true for reference sets that do not contain inconsistencies, but it is also relevant when inconsistencies do exist (in the PDA context, inconsistencies are usually resolved algorithmically or interactively with the DM before the final model is built; see for instance [40]). With a consistent reference set, the error variables in formulations (6)–(7) become equal to zero and consequently these optimization models reduce to a set of feasible linear constraints. Each solution satisfying these constraints corresponds to a different decision model and even though all the corresponding feasible decision models provide the same outputs for the reference set, their recommendations can differ significantly when the models are used to perform evaluations for other alternatives.

For instance, consider the example data of Table 1 for a classification problem where a DM classified six references alternatives in two categories, under three evaluation criteria. Assuming a linear weighted average model of the form $V(\mathbf{x}_i) = w_1x_{i1} + w_2x_{i2} + w_3x_{i3}$, with $w_1 + w_2 + w_3 = 1$ and $w_1, w_2, w_3 \geq 0$, the model would be consistent with the classification of the alternatives if $V(\mathbf{x}_i) \geq V(\mathbf{x}_j) + \delta$ for all $i = 1,2,3$ and $j = 4,5,6$, where δ is a small positive constant (e.g., $\delta = 0.01$). Figure 1 illustrates graphically the set of values for the criteria trade-offs that comply with the classification of the reference alternatives (the shaded area defined by the corner points A-E). It is evident that very different trade-offs provide the same results for the reference data. For example, the trade-off w_1 of the first criterion may vary anywhere from zero to one, whereas w_2 may vary from zero up to 0.7.

Table 1: An illustrative classification problem

	Criteria			
Alternatives	x_1	x_2	x_3	Classification
$\mathbf{x}_1$	7	1	8	G_1
$\mathbf{x}_2$	4	5	8	G_1
$\mathbf{x}_3$	10	4	2	G_1
$\mathbf{x}_4$	2	4	1	G_2
$\mathbf{x}_5$	4	1	1	G_2
$\mathbf{x}_6$	1	2	5	G_2

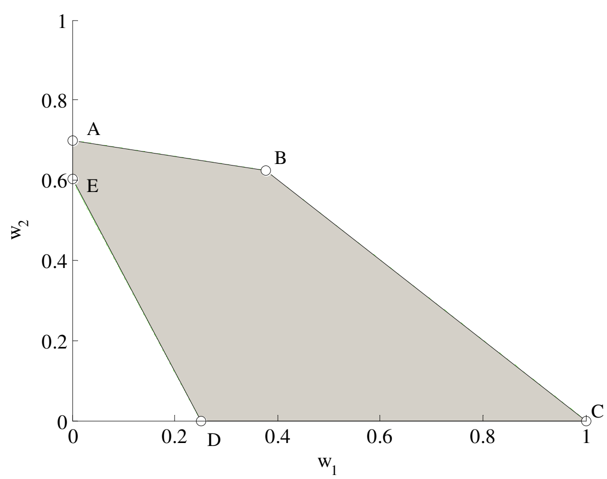


Fig. 1: The feasible set for the criteria trade-offs that are compatible with the classification of the example data of Table 1

The size of the polyhedron defined by a set of feasible constraints of formulations such as (6) and (7) depends on a number of factors, but the two most important can be identified to the adequacy of set of reference examples and the complexity of the selected decision modeling form. The former is immediately related to the quality of the information on which model inference is based. Vetschera et al. [59] performed an experimental analysis to investigate how the size of the reference set affects the robustness and accuracy of the resulting multicriteria models in classification problems. They found that small reference sets (e.g., with a limited number of alternatives with respect to the number of criteria) lead to decision models that are neither robustness nor accurate. Expect for its size other characteristics of the reference set are also relevant, may involve the existence of noisy data, outliers, the existence of correlated criteria, etc. [12].

The complexity of the inferred decision model is also an issue that is related to its robustness. Simpler models (e.g., a linear value function) are more robust compared to more complex non-linear models. The latter are defined by a larger number of parameters and as a result the inference procedure becomes less robust and more sensitive to the available data. For instance, Figure 2 illustrates a two-class classification problem with two criteria (which correspond to the axes of the figure). The linear classification model (green line) is robust; with the available data only marginal changes can be made in this model (separating line) without affecting its classification results for the data shown in the figure. On the other hand, a non-linear model (blue line) is not robust, particularly in the areas where the data are sparse (i.e., the upper left and lower right parts of the graph). Therefore, care should be given to the selection of the appropriate modeling taking into account both the DM's system of preferences as well as the available data. This issue has been studied extensively in areas such as the statistical learning theory [47, 56, 57].

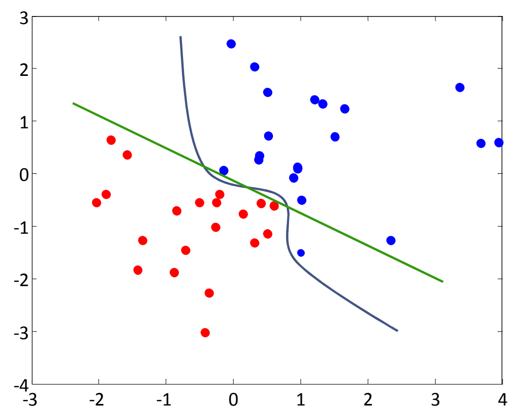


Fig. 2: A linear vs a non-linear classification model

4 Robust Disaggregation Approaches

The research in the area of building robust multicriteria decision models and obtaining robust recommendations with disaggregation techniques can be classified into three main directions. The first, involves approaches that focus on describing the set of feasible decision models with analytic or simulation techniques. The second direction focuses on procedures for formulating robust recommendations through multiple acceptable decision models, whereas a third line of research has focused on techniques for selecting the most characteristic (representative) model from the set of all models compatible with the information provided by the reference set. The following subsections discuss these approaches in more detail.

4.1 Describing the Set of Acceptable Decision Models

The DM's evaluations for the reference alternatives provide information on the set of acceptable decision models that comply with these evaluations. Searching for different solutions within this feasible set and measuring its size provides useful information on the robustness of the results. Analytic and simulation-based techniques have been used for this purpose, focusing on convex polyhedral sets for which the analysis is computationally feasible. As explained in the previous section, for decision models which are linear with respect to their parameters (such as additive value functions) the set of acceptable decision models is a convex polyhedron. The same applies to other types of decision models with some simplifications on the parameters that are inferred (see for example [41]).

Jacquet-Lagrèze and Siskos [29] were the first to emphasize that the inference of a decision model through optimization formulations such as the ones described in section 2.2, may not be robust thus suggesting that the existence of multiple optimal solutions (or even alternative near-optimal ones in the cases of inconsistent reference sets) should be carefully explored. The approach they suggested was based on a heuristic post-optimality procedure seeking to identify some characteristic alternative models corresponding to corner points of the feasible polyhedron. In the context of inferring an ordinal regression decision model, this approach is implemented in two phases. First, problem (6) is solved and its optimal objective function value F^* (total sum of errors) is recorded. In the second phase, $2n$ additional optimization problems are solved by maximizing and minimization the trade-offs of the criteria (one at a time), while ensuring that the new solutions do not yield an overall error larger than $F^*(1+\alpha)$, where α is a small percentage of F^*. While this heuristic approach does not fully describe the polyhedron that defines the parameters of the decision model, it does give an indication of how much the relative importance of the criteria deviates within the polyhedron. Based on this approach, Grigoroudis and Siskos [24] developed a measure to assess the stability and robustness of the inferred model as the normalized standard deviation of the results obtained from the post-optimality analysis.

Despite their simplicity, post-optimality techniques provide only a limited partial view of the complete set of models that are compatible with the DM's preferences. A more thorough analysis requires the implementation of computationally intensive analytic or simulation approaches. Following the former direction, Vetschera [58] developed a recursive algorithm for computing the volume of the polyhedron that is derived from preferential constraints in the case of a linear evaluation model, but the algorithm was applicable to rather small problems (e.g., up to 20 alternatives and 6 criteria). Similar, but computationally more efficient algorithms, are available in the area of computational geometry, but they have not yet been employed in the context of MCDA. For instance, Lovász and Vempala [38] presented a fast algorithm for computing the volume of a convex polyhedron, which combines simulated annealing with multi-phase Monte Carlo sampling.

The computational difficulties of analytic techniques have led to the adoption of simulation approaches, which have gained much interest in the context of robust decision aiding. Originally used for sensitivity analysis [7] and decision aiding in stochastic environments [37], simulation techniques have been recently employed to facilitate the formulation of robust recommendations under different decision modeling forms. For instance, Tervonen et al. [52] used such an approach in order to formulate robust recommendations with the ELECTRE TRI multicriteria classification method [16], whereas Kadziński and Tervonen [31,32] used a simulation-based approach to enhance the results of robust analytic techniques obtained with additive value models in the context of ranking and classification problems.

Simulation-based techniques were first based on rejection sampling schemes. Rejection sampling is a naïve approach under which a random model is constructed (usually from a uniform distribution [46]) and tested against the DM's evaluations for the reference alternatives. The model is accepted only if it is compatible with the DM's evaluations and rejected otherwise. However, the rejection rate increases rapidly with the dimensionality of the polyhedron (as defined by the number of the model's parameters). As a result the sampling of feasible solutions becomes intractable for problems of realistic complexity. Hit-and-run algorithms [35, 53] are particularly useful in reducing the computational burden, thus enabling the efficient sampling from high-dimensional convex regions.

4.2 Robust Decision Aid with a Set of Decision Models

Instead of focusing on the identification of different evaluation models that can be inferred from a set of reference decision examples through heuristic, analytic, or simulation approaches, a second line of research has been concerned with how robust recommendations can be formulated by aggregating the outputs of different models and exploiting the full information embodied in a given set of decision instances.

Siskos [49] first introduced the idea of building preference relations based on a set of decision models inferred with a preference disaggregation approach for ordinal regression problems. In particular, he presented the construction of a fuzzy preference relation based on the results of a post-optimality procedure. The fuzzy preference relation allows the evaluation of the alternatives through the aggregation of the outputs of multiple characteristic models (additive value functions) inferred from a set of decision instances.

Recently, this idea has been further extended to consider not only a subset of acceptable models but all models that can be inferred from a given reference set (without actually identifying them). Following this approach and in an ordinal regression setting, Greco et al. [22] defined necessary and possible preference relations on the basis of the DM's evaluations on a set of reference alternatives, as follows:

- Weak necessary preference relation: $\mathbf{x}_i \succsim^N \mathbf{x}_j$ if $V(\mathbf{x}_i) \geq V(\mathbf{x}_j)$ for all decision models $V(\cdot)$ compatible with the DM's evaluations on a set of reference alternatives.
- Weak possible preference relation: $\mathbf{x}_i \succsim^P \mathbf{x}_j$ if $V(\mathbf{x}_i) \geq V(\mathbf{x}_j)$ for at least one decision model $V(\cdot)$ compatible with the DM's evaluations on a set of reference alternatives.

From these basic relations preference, indifference, and incomparability relations can be built allowing the global evaluation of any alternative using the full information provided by the reference examples. The above relations can be checked through the solution of simple optimization formulations, without actually requiring the enumeration of all decision models that can be inferred from the reference examples. This approach was also used for multicriteria classification problems [23] as well as for outranking models [10, 19] and non-additive value models [1].

4.3 Selecting a Representative Decision Model

Having an analytic or simulation-based characterization of all compatible models (e.g., with approaches such as the ones described in the previous subsections) provides the DM with a comprehensive view of the range of possible recommendations that can be formed on the basis of a set of models implied from some decision examples. On the other hand, a single representative model is easier to use as it only requires the DM to "plug-in" the data for any alternative into a functional, relational, or symbolic model. Furthermore, the aggregation of all evaluation criteria in a single decision model enables the DM to get insight into the role of the criteria and their effect on the recommendations formulated through the model [20].

In the above context several approaches have been introduced to infer a single decision model that best represents the information provided by a reference set of alternatives. Traditional disaggregation techniques such as the family of the UTA methods [50] use post-optimality techniques based on linear programming in order to build a representative additive value function defined as an average solution of some characteristic models compatible with the DM's judgments, defined by maximizing and minimizing the criteria trade-offs. Such an averaging approach provides a proxy of the center of the feasible region.

However, given that only a very few number of corner points are identified with this heuristic post-optimality process (at most $2n$ corner points), it is clear that the average solution is only a very rough "approximation" of the center of the polyhedron. Furthermore, the optimizations performed during the post-optimality analysis may not lead to unique results. For instance, consider again the classification example discussed in section 3 and its graphical illustration in Figure 1 for the feasible set for the criteria trade-offs which are compatible with the DM's classification of the reference alternatives (Table 1). The maximization of the trade-off constant w_1 leads to corner point C, the maximization of w_2 leads to point A, whereas the maximization of w_3 (which corresponds to the minimization of $w_1 + w_2$) leads to point D. However, the minimization of the two trade-offs does not lead to uniquely defined solutions. For instance, the minimization of w_1 may lead to point A or point E, the minimization of w_2 leads either to C or D, and the minimization of w_3 (i.e., the maximization of $w_1 + w_2$) may lead to points B or C. Thus, depending on which corner solutions are obtained, different average decision models can be constructed. Table 2 lists the average criteria trade-offs corresponding to different centroid solutions. It is evident that the results vary significantly depending on the obtained post-optimality results.

Table 2: The post-optimality approach for constructing a centroid model within the polyhedron of acceptable models for the data of Table 1

	Post-optimality solutions							
max w_1	C	C	C	C	C	C	C	C
min w_1	E	A	E	A	E	A	E	A
max w_2	A	A	A	A	A	A	A	A
min w_2	D	D	C	C	D	D	C	C
max w_3 (min $w_1 + w_2$)	D	D	D	D	D	D	D	D
min w_3 (max $w_1 + w_2$)	B	B	B	B	C	C	C	C
	Centroid solutions							
w_1	0.31	0.31	0.44	0.44	0.42	0.42	0.54	0.54
w_2	0.32	0.34	0.32	0.34	0.22	0.23	0.22	0.23
w_3	0.37	0.35	0.24	0.23	0.37	0.35	0.24	0.23

A number of alternative approaches have been proposed to address the ambiguity in the results of the above post-optimality process. Beuthe and Scannella [4] presented different post-optimality criteria in an ordinal regression setting to improve the discriminatory power of the resulting evaluating model. Similar criteria were also proposed by Doumpos and Zopounidis [12] for classification problems.

Alternative optimization formulations have also been introduced allowing the construction of robust decision models without requiring the implementation of post-optimality analyses. Following this direction, Doumpos and Zopounidis [13] presented simple modifications of traditional optimization formulations (such as the ones discussed in section 2.2) on the grounds of the regularization principle which is widely used in data mining and statistical learning [57]. Experimental results on artificial data showed that new formulations can provide improved results in ordinal regression and classification problems. On the other hand, Bous et al. [5] proposed a non-linear optimization formulation for ordinal regression problems that enables the construction of an evaluation model through the identification of the analytic center of the polyhedron form by the DM's evaluations on some reference decision instances. Despite its non-linear character, the proposed optimization model is easy to solve with existing iterative algorithms. In a different framework, Greco et al. [20] considered the construction of a representative model through an interactive process, which is based on the grounds of preference relations inferred from the full set of models compatible with the DM's evaluations [22]. During the proposed interactive process, different targets are formulated, which can be used by the DM as criteria for specifying the most representative evaluation model.

5 Connections with Statistical Learning

5.1 Principles of Data Mining and Statistical Learning

Similarly to disaggregation analysis, statistical learning and data mining are also involved with learning from examples [25, 26]. Many advances have been made within these fields for regression, classification, and clustering problems. Recently there has been a growing interest among machine learning researchers towards preference modeling and decision-making. Some interest has also been developed by MCDA researchers on exploiting the advances in machine learning.

Hand et al. [25] define data mining as "*the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner*". Statistical learning plays an important role in the data mining process, by describing the theory that underlies the identification of such relationships and providing the necessary algorithmic techniques. According to Vapnik [56, 57] the process of learning from examples includes three main components:

1. A set X of data vectors $\mathbf{x}$ drawn independently from a probability distribution $P(\mathbf{x})$. This distribution is assumed to be unknown, thus implying that there is no control on how the data are observed [51].
2. An output y from a set Y, which is defined for every input $\mathbf{x}$ according to an unknown conditional distribution function $P(y \mid \mathbf{x})$. This implies that the relationship between the input data and the outputs is unknown.
3. A learning method (machine), which is able to assign a function $f_\beta : X \to Y$, where β are some parameters of the unknown function.

The best function f_β is the one that best approximates the actual outputs, i.e., the one that minimizes:

$$\int L[y, f_\beta(\mathbf{x})]dP(\mathbf{x}, y) \tag{8}$$

where $L[y, f_\beta(\mathbf{x})]$ is a function of the differences between the actual output y and the estimate $f_\beta(\mathbf{x})$,[1] and $P(\mathbf{x}, y) = P(\mathbf{x})P(y \mid \mathbf{x})$ is the joint probability distribution of $\mathbf{x}$ and y. However, this joint distribution is unknown and the only available information is contained in a training set of m objects $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$, which are assumed to be generated independently from this unknown distribution. Thus, the objective (8) is substituted by an empirical risk estimate:

$$\frac{1}{m}\sum_{i=1}^{m} L[y_i, f_\beta(\mathbf{x}_i)] \tag{9}$$

For a class of functions f_β of a given complexity, the minimization of (9) leads to the minimization of an upper bound for (8).

A comparison of (2) and (9) shows that PDA and statistical learning are concerned with similar problems from different perspectives and focus (for a discussion of the similarities and differences of the two fields see [14, 62]).

5.2 Regularization and Robustness in Learning Machines

In the context of data mining and statistical learning, robustness is a topic of fundamental importance and is directly linked to the theory in these fields. Robustness in this case has a slightly different interpretation compared to its used in MCDA. In particular, from a data mining/statistical learning perspective robustness involves the ability of a prediction model (or learning algorithm) to retain its structure and provide accurate results in cases where the learning process is based on data that contain imperfections (i.e., errors, outliers, noise, missing data, etc.). Given that the robustness of a prediction model is related to its complexity, statistical learning has been founded on a rigorous theoretical framework that connects robustness, complexity, and the empirical risk minimization approach.

The foundations of this theoretical framework are based on Tikhonov's regularization principle [54], which involves systems of linear equations of the form $\mathbf{Ax} = \mathbf{b}$. When the problem is ill-posed, such a system of equations may not have a solution and the inverse of matrix $\mathbf{A}$ may exhibit instabilities (i.e., $\mathbf{A}$ may be singular or ill-conditioned). In such cases, a numerically robust solution can be obtained through the approximate system $\mathbf{Ax} \approx \mathbf{b}$, such that the following function is minimized:

$$\|\mathbf{Ax} - \mathbf{b}\|^2 + \lambda\|\mathbf{x}\|^2 \tag{10}$$

[1] The specification of the loss function L depends on the problem under consideration. For instance, in a regression setting it may correspond to the mean squared error, whereas in a classification context it may represent the accuracy rate.

where $\lambda > 0$ is a regularization parameter that defines the trade-off between the error term $\|\mathbf{Ax}-\mathbf{b}\|^2$ and the "size" of the solution (thus controlling the solution for changes in $\mathbf{A}$ and $\mathbf{b}$).

With the introduction of statistical learning theory Vapnik [56] developed a general framework that uses the above idea to relate the complexity and accuracy of learning machines. In particular, Vapnik showed that under a binary loss function, the expected error $E(\beta)$ of a decision model defined by some parameters β, is bounded (with probability $1-\alpha$) by:

$$E(\beta) \leq E_{emp}(\beta) + \sqrt{\frac{h[\log(2m/h)+1]-\log(\alpha/4)}{m}} \tag{11}$$

where E_{emp} is the empirical error of the model as defined by equation (9) and h is the Vapnik-Chervonenkis dimension, which represents the complexity of the model. When the size of the training data set in relation to the complexity of the model is large (i.e., when $m/h \gg 1$), then the second term in the left-hand side of (11) decreases and the expected error is mainly defined by the empirical error. On the other hand, when $m/h \ll 1$ (i.e., the number of training observations is too low compared to the model's complexity), then the second term increases and thus becomes relevant for the expected error of the model.

This fundamental result constitutes the basis for developing decision and prediction models in classification, regression, and clustering tasks. For instance, assume a binary classification setting where a linear model $f(\mathbf{x}) = \mathbf{wx} - \gamma$ should be developed to distinguish between a set of positive and negative observations. In this context, it can be shown that if the data belong in a ball of radius R, the complexity parameter h of a model with $\|\mathbf{w}\| \leq L$ (for some $L > 0$) is bounded as follows [56, 57]:

$$h \leq \min\{L^2R^2, n\} + 1 \tag{12}$$

Thus, with a training set consisting of m positive and negative observations ($y = 1$ and $y_i = -1$, respectively), the optimal model that minimizes the expected error can be obtained from the solution of the following convex quadratic program:

$$\begin{aligned} \min \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\sigma_i \\ \text{s.t.} \quad & y_i(\mathbf{wx}_i - \gamma) + \sigma_i \geq 1 \qquad \forall\, i = 1, \ldots, m \\ & \sigma_i \geq 0 \qquad\qquad\qquad\quad\; \forall\, i = 1, \ldots, m \\ & \mathbf{w}, \gamma \in \mathbb{R} \end{aligned} \tag{13}$$

The objective function of this problem is in accordance with the Tikhonov regularization function (10). In particular, the sum of classification errors $\sigma_1, \ldots, \sigma_m$ is used as a substitute for the error term $\|\mathbf{Ax}-\mathbf{b}\|^2$ in (10), whereas the regularization parameter λ in (10) is set equal to $0.5/C$. The minimization of $\|\mathbf{w}\|^2$ in the objective

[1] Although this is not a restricted assumption, as the theory is general enough to accommodate other loss functions as well.

function of the above problem corresponds to the minimization of the complexity bound (12), which in turn leads to the minimization of the second term in the error bound (11). On the other hand, the minimization of the sum of the classification errors corresponds to the minimization of the empirical error E_{emp}.

This framework is not restricted to linear models, but it also extends to nonlinear models of arbitrary complexity and it is applicable to multi-class problems [6], regression problems [9, 39], and clustering problems [2]. Similar, principles and approaches have also been used for other types of data mining models such as neural networks [17].

The development of data mining and statistical learning models with optimization with mathematical programming techniques has received much attention [43]. In this context, robust model building has been considered from the perspective of robust optimization. Bertsimas et al. [3] expressed a robust optimization model in the following general form:

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}, \mathbf{u}_i) \leq \mathbf{0} \qquad \forall \mathbf{u}_i \in \mathcal{U}_i,\ i = 1, \dots, m \\ & \mathbf{x} \in \mathbb{R} \end{aligned} \tag{14}$$

where $\mathbf{x}$ is the vector of decision variables, $\mathbf{u}_i \in \mathbb{R}^k$ are perturbation vectors associated with the uncertainty in the parameters that define the constraints, and $\mathcal{U}_i \subseteq \mathbb{R}^k$ are uncertainty sets in which the perturbations are defined (for an overview of the theory and applications of robust optimization in design problems see [36]). For instance, a robust linear program can be expressed as follows:

$$\begin{aligned} \min \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^\top \mathbf{x} \leq b_i \qquad \forall \mathbf{a}_i \in \mathcal{U}_i,\ i = 1, \dots, m \\ & \mathbf{x} \in \mathbb{R} \end{aligned} \tag{15}$$

where the coefficients of the decision variables in the constraints take values from the uncertainty sets $\mathcal{U}_i \subseteq \mathbb{R}^n$. Thus, a constraint $\mathbf{a}_i^\top \mathbf{x} \leq b_i$ is satisfied for every $\mathbf{a}_i \in \mathcal{U}_i$ if and only if $\max_{\mathbf{a}_i \in \mathcal{U}_i} \{\mathbf{a}_i^\top \mathbf{x}\} \leq b_i$.

The framework of robust optimization has been used to develop robust decision and prediction models in the context of statistical learning. For instance, assuming that the data for observation i are subject to perturbations defined by a stochastic vector δ_i from some distribution, bounded such that $\|\delta_i\|^2 \leq \eta_i$, the constraints of problem (13) can be re-written as:

$$y_i[\mathbf{w}(\mathbf{x}_i + \delta_i) - \gamma] + \sigma_i \geq 1 \tag{16}$$

Such methodologies for developing robust learning machines have been presented in several works (see for instance [48, 55, 63, 64]). Caramanis et al. [8] as well as Xu and Mannor [65] provide comprehensive overviews of robust optimization in the context of statistical learning and data mining.

5.3 Applications in MCDA Disaggregation Approaches

The principles and methodologies available in the areas of data mining and statistical/machine learning have recently attracted interest for the development of enhanced approaches in MCDA. In this context, Herbrich et al. [27] explored how the modeling approach described in the previous section can be used to develop value function models in ordinal regression problems and analyzed the generalization ability of such models in relation to the value differences between alternatives in consecutive ranks.

Evgeniou et al. [15] also examined the use of the statistical learning paradigm in an ordinal regression setting. They showed that the development of a linear value function model of the form $V(x) = \mathbf{wx}$ that minimizes $\|\mathbf{w}\|^2$ leads to robust results, as the obtained model corresponds to the center of the largest sphere that can be inscribed by preferential constraints of the form $\mathbf{w}(\mathbf{x}_i - \mathbf{x}_j) \geq 1$ for pairs of alternatives such that $\mathbf{x}_i \succ \mathbf{x}_j$.

Doumpos and Zopounidis [13] followed a similar approach for the development of additive function functions using the L_1 norm for the vector of parameters of the model. Thus, they augmented the objective function of problems (6)–(7) considering not only the error variables, but also the complexity of the resulting value function. Through this approach, they described the relationship between the accuracy of the decision model and the quality of the information provided by the reference data. Empirical analyses on ranking and classification problems showed that the new formulations provide results that best describe the DM's preferences, are more robust to changes of the reference data, and have higher generalization performance compared to existing PDA approaches. A similar approach for constructing additive value functions was also proposed by Dembczynski et al. [11] who combined a statistical learning algorithm with a decision rule approach for classification problems.

Except for functional decision models, similar approaches have also been used for relational models, which are based on pairwise comparisons between the alternatives. For instance, Waegeman et al. [62] used a kernel approach for constructed outranking decision models and showed that such an approach is general enough to accommodate (as special cases) a large class of different types of decision models, including value functions and the Choquet integral. Pahikkala et al. [42] extended this approach to intransitive decision models.

6 Conclusions and Future Perspectives

PDA techniques greatly facilitate the development of multicriteria decision aiding models, requiring the DM to provide minimal information without asking for the specification of complex technical parameters which are often not well-understood by DMs in practice. However, using such a limited amount of data should be done with care in order to derive meaningful and really useful results.

Robustness is an important issue in this context. Addressing the robustness concern enables the formulation of recommendations and results that are valid under different conditions with respect to the modeling conditions and the available data. In this chapter we discussed the main aspects of robustness in PDA techniques and provided an up-to-date overview of the different lines of research and the related advances that have been introduced in this area. We also discussed the statistical learning perspective for developing robust and accurate decision models, which has adopted a different point of view in the analysis of robustness compared to MCDA.

Despite their different philosophies, PDA and statistical learning share common features and their connections could provide further improved approaches to robust decision aiding. Future research should also focus on the further theoretical and empirical analysis of the robustness properties of PDA formulations, the introduction of meaningful measures for assessing robustness, and the development of methodologies to improve the robustness of models and solutions in decision aid.

Acknowledgements This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning".

References

1. S. Angilella, S. Greco, and B. Matarazzo. Non-additive robust ordinal regression: A multiple criteria decision model based on the Choquet integral. *European Journal of Operational Research*, 201(1):277–288, 2010.
2. A. Ben-Hur, D. Horn, H.T. Siegelmann, and V. Vapnik. Support vector clustering. *Journal of Machine Learning Research*, 2:125–137, 2001.
3. D. Bertsimas, D.B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.
4. M. Beuthe and G. Scannella. Comparative analysis of UTA multicriteria methods. *European Journal of Operational Research*, 130(2):246–262, 2001.
5. B. Bous, Ph. Fortemps, F. Glineur, and M. Pirlot. ACUTA: A novel method for eliciting additive value functions on the basis of holistic preference statements. *European Journal of Operational Research*, 206(2):435–444, 2010.
6. E.J. Bredensteiner and K.P. Bennett. Multicategory classification by support vector machines. *Computational Optimization and Applications*, 12(1-3):53–79, 1999.
7. J. Butler, J. Jia, and J. Dyer. Simulation techniques for the sensitivity analysis of multi-criteria decision models. *European Journal of Operational Research*, 103(3):531–546, 1997.
8. C. Caramanis, S. Mannor, and H. Xu. Robust optimization in machine learning. In S. Sra, S. Nowozin, and S. Wright, editors, *Optimization for Machine Learning*, pages 369–402. MIT Press, Cambridge, 2011.
9. W. Chu and S.S. Keerthi. Support vector ordinal regression. *Neural Computation*, 19(3):792–815, 2007.
10. S. Corrente, S. Greco, and R. Słowiński. Multiple criteria hierarchy process with ELECTRE and PROMETHEE. *Omega*, 41(5):820–846, 2013.
11. K. Dembczynski, W. Kotlowski, and R. Słowiński. Additive preference model with piecewise linear components resulting from dominance-based rough set approximations. In L. Rutkowski, R. Tadeusiewicz, L. Zadeh, and J. Zurada, editors, *Artificial Intelligence and Soft Computing - ICAISC 2006*, volume 4029 of *Lecture Notes in Computer Science*, pages 499–508. Springer, Berlin - Heidelberg, 2006.

12. M. Doumpos and C. Zopounidis. *Multicriteria Decision Aid Classification Methods*. Kluwer Academic Publishers, Dordrecht, 2002.
13. M. Doumpos and C. Zopounidis. Regularized estimation for preference disaggregation in multiple criteria decision making. *Computational Optimization and Applications*, 38(1):61–80, 2007.
14. M. Doumpos and C. Zopounidis. Preference disaggregation and statistical learning for multicriteria decision support: A review. *European Journal of Operational Research*, 209(3):203–214, 2011.
15. T. Evgeniou, C. Boussios, and G. Zacharia. Generalized robust conjoint estimation. *Marketing Science*, 24(3):415–429, 2005.
16. J.R. Figueira, S. Greco, B. Roy, and R. Słowiński. ELECTRE methods: Main features and recent developments. In C. Zopounidis and P.M. Pardalos, editors, *Handbook of Multicriteria Analysis*, pages 51–89. Springer, New York, 2010.
17. S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58, 1992.
18. P.G. Georgiev, D. The Luc, and P.M. Pardalos. Robust aspects of solutions in deterministic multiple objective linear programming. *European Journal of Operational Research*, 229(1):29–36, 2013.
19. S. Greco, M. Kadziński, V. Mousseau, and R. Słowiński. ELECTRE$^{\mathrm{GKMS}}$: Robust ordinal regression for outranking methods. *European Journal of Operational Research*, 214(1):118–135, 2011.
20. S. Greco, M. Kadziński, and R. Słowiński. Selection of a representative value function in robust multiple criteria sorting. *Computers and Operations Research*, 38(11):1620–1637, 2011.
21. S. Greco, B. Matarazzo, and R. Słowiński. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1):1–47, 2001.
22. S. Greco, V. Mousseau, and R. Słowiński. Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research*, 191(2):416–436, 2008.
23. S. Greco, V. Mousseau, and R. Słowiński. Multiple criteria sorting with a set of additive value functions. *European Journal of Operational Research*, 207(3):1455–1470, 2010.
24. E. Grigoroudis and Y. Siskos. Preference disaggregation for measuring and analysing customer satisfaction: The MUSA method. *European Journal of Operational Research*, 143(1):148–170, 2002.
25. D. Hand, H. Mannila, and P. Smyth. *Principles of Data Mining*. MIT Press, Cambridge, 2001.
26. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2001.
27. R. Herbrich, T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. In A.J. Smola, P.L. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 115–132. MIT Press, Cambridge, MA, 2000.
28. R. Hites, Y. De Smet, N. Risse, M. Salazar-Neumann, and P. Vincke. About the applicability of MCDA to some robustness problems. *European Journal of Operational Research*, 174(1):322–332, 2006.
29. E. Jacquet-Lagrèze and Y. Siskos. Assessing a set of additive utility functions for multicriteria decision making: The UTA method. *European Journal of Operational Research*, 10:151–164, 1982.
30. E. Jacquet-Lagrèze and Y. Siskos. Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research*, 130:233–245, 2001.
31. M. Kadziński and T. Tervonen. Robust multi-criteria ranking with additive value models and holistic pair-wise preference statements. *European Journal of Operational Research*, 228(1):169–180, 2013.
32. M. Kadziński and T. Tervonen. Stochastic ordinal regression for multiple criteria sorting problems. *Decision Support Systems*, 55(1):55–66, 2013.
33. R.L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, Cambridge, 1993.

34. P. Kouvelis and G. Yu. *Robust Discrete Optimization and Its Applications*. Kluwer Academic Publishers, Dordrecht, 1997.
35. D.P. Kroese, T. Taimre, and Z.I. Botev. *Handbook of Monte Carlo Methods*. Wiley, New York, 2011.
36. A.J. Kurdila, P.M. Pardalos, and M. Zabarankin. *Robust Optimization-Directed Design*. Springer, New York, 2006.
37. R. Lahdelma and P. Salminen. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. *Operations Research*, 49(3):444–454, 2001.
38. L. Lovász and S. Vempala. Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *Journal of Computer and System Sciences*, 72(2):392–417, 2006.
39. O.L. Mangasarian and D.R. Musicant. Robust linear and support vector regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9):950–955, 2000.
40. V. Mousseau, J. Figueira, L. Dias, C. Gomes da Silva, and J. Clímaco. Resolving inconsistencies among constraints on the parameters of an MCDA model. *European Journal of Operational Research*, 147(1):72–93, 2003.
41. V. Mousseau, J. Figueira, and J.-Ph. Naux. Using assignment examples to infer weights for ELECTRE TRI method: Some experimental results. *European Journal of Operational Research*, 130(2):263–275, 2001.
42. T. Pahikkala, W. Waegeman, W. Tsivtsivadze, B. De Baets, and T. Salakoski. Learning intransitive reciprocal relations with kernel methods. *European Journal of Operational Research*, 206(3):676–685, 2010.
43. P.M. Pardalos and P. Hansen. *Data Mining and Mathematical Programming*. American Mathematical Society, Providence, 2008.
44. B. Roy. The outranking approach and the foundations of ELECTRE methods. *Theory and Decision*, 31:49–73, 1991.
45. B. Roy. Robustness in operational research and decision aiding: A multi-faceted issue. *European Journal of Operational Research*, 2010(3):629–638, 2010.
46. R.Y. Rubinstein. Generating random vectors uniformly distributed inside and on the surface of different regions. *European Journal of Operational Research*, 10(2):205–209, 1982.
47. B. Schölkopf and A.J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, Cambridge, Massachusetts, 2002.
48. P.K. Shivaswamy, C. Bhattacharyya, and A.J. Smola. Second order cone programming approaches for handling missing and uncertain data. *Journal of Machine Learning Research*, 6:1283–1314, 2006.
49. J. Siskos. A way to deal with fuzzy preferences in multicriteria decision problems. *European Journal of Operational Research*, 10(3):314–324, 1982.
50. Y. Siskos and E. Grigoroudis. New trends in aggregation-disaggregation approaches. In C. Zopounidis and P.M. Pardalos, editors, *Handbook of Multicriteria Analysis*, pages 189–214. Springer, Berlin Heidelberg, 2010.
51. I. Steinwart and A. Christmann. *Support Vector Machines*. Springer, New York, 2008.
52. T. Tervonen, J.R. Figueira, R. Lahdelma, J.A. Dias, and P. Salminen. A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research*, 192(1):236–242, 2009.
53. T. Tervonen, G. van Valkenhoef, N. Basturk, and D. Postmus. Hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research*, 224(3):552–559, 2013.
54. A.N. Tikhonov, A.V. Goncharsky, V.V. Stepanov, and Yagola A.G. *Numerical Methods for the Solution of Ill-Posed Problems*. Springer, Dordrecht, 1995.
55. T.B. Trafalis and R.C. Gilbert. Robust support vector machines for classification and computational issues. *Optimization Methods and Software*, 22(1):187–198, 2007.
56. V.N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, 1999.
57. V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 2nd edition, 2000.

58. R. Vetschera. A recursive algorithm for volume-based sensitivity analysis of linear decision models. *Computers and Operations Research*, 24(4):477–491, 1997.
59. R. Vetschera, Y. Chen, K.W. Hipel, and D. Marc Kilgour. Robustness and information levels in case-based multiple criteria sorting. *European Journal of Operational Research*, 202(3):841–852, 2010.
60. Ph. Vincke. *Multicriteria Decision Aid*. John Wiley and Sons, New York, 1992.
61. Ph. Vincke. Robust solutions and methods in decision-aid. *Journal of Multi-Criteria Decision Analysis*, 8(3):181–187, 1999.
62. W. Waegeman, B. De Baets, and B. Boullart. Kernel-based learning methods for preference aggregation. *4OR*, 7:169–189, 2009.
63. P. Xanthopoulos, M.R. Guarracino, and P.M. Pardalos. Robust generalized eigenvalue classifier with ellipsoidal uncertainty. *Annals of Operations Research*, 2013. forthcoming.
64. H. Xu, C. Caramanis, and S. Mannor. Robustness and regularization of support vector machines. *Journal of Machine Learning Research*, 10:1485–1510, 2009.
65. H. Xu and S. Mannor. Robustness and generalization. *Machine Learning*, 86:391–423, 2012.
66. C. Zopounidis and M. Doumpos. Building additive utilities for multi-group hierarchical discrimination: the M.H.Dis method. *Optimization Methods and Software*, 14(3):219–240, 2000.
67. C. Zopounidis and M. Doumpos. Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research*, 138(2):229–246, 2002.
68. C. Zopounidis and P.M. Pardalos. *Handbook of Multicriteria Analysis*. Springer, Berlin Heidelberg, 2010.

Robustness Analysis in Multicriteria Disaggregation – Aggregation Approaches for Group Decision Making

Denis Yannacopoulos, Athanasios Spyridakos and Nikos Tsotsolas

Technological Educational Institute of Piraeus
Department of Business Administration
P. Ralli and Thivon 250, 12244, Aigaleo, Athens
dgian,tspyr@teipir.gr, ntsotsol@teipir.gr

Abstract. Multicriteria Disaggregation - Aggregation (D-A) approaches results to the estimation of Decision Makers' preference models (usually of additive value) through interactive procedures, where the global preferences of DMs are analysed. The low robustness of preference models, presented in many cases, can be the result of the ill-structured problem formulation or can reflect the real thoughts of DMs. The case of collaborative decision making presents more complicated situations. This research work describes the use of visual techniques based on 3d graphs and a set of indexes, which can be used for picturing and comprehension of the low robustness in collaborative decision making problems. Also, the frame of feedbacks which can be utilised for the reducing and the exploitation of the low robustness is described and illustrated through a case study.

Keywords: Multicriteria Decision Aid, Collaborative Decision Making, Robustness Analysis

1 Introduction

The cases of collective decision making is more complex than to the individual ones, given that, the different perspectives increase the ill-structured nature into decision problems of the real world. Multicriteria Disaggregation Aggregation (D-A) approaches for discrete alternative actions aim to the estimation of preference models (usually of additive value) based on Decision Makers (DMs) global preferences [7][11].

The additive value model is described in the following formulae:

$$U(\mathbf{g}) = \sum_{i=1}^{n} p_i u_i(g_i)$$

$$u(g_{i*}) = 0. \quad u(g_i^{*}) = 1, \qquad \text{for } i=1, 2,..., n$$

$$\sum_{i=1}^{n} p_i = 1$$

$p_i \geq 0$, for i=1, 2,..., n

where:

- $\mathbf{g} = (g_1, g_2, ..., g_n)$ is the evaluation vector of an alternative action on the n criteria,
- g_{i*} and g_i^* are the least and most preferable levels of the criterion g_i respectively and
- $u_i(g_i)$, p_i are the value function and the relative weight of the i-th criterion.

DM's global preferences are expressed by rank-ordering (pre-ranking) of a representative and familiar to the DMs subset of the alternative actions, called reference set. Special Linear Programming (LP) techniques are utilised in order to estimate an additive value model, that produces a ranking of the reference actions as close as possible to the DM's pre-ranking. The alternative actions of the reference set are rearranged in such a way that a_1 is the head and a_k is the tail of the ranking and for every pair of consecutive actions (a_m, a_{m+1}) holds, either $a_m\mathbf{P}a_{m+1}$ (preference) either $a_m\mathbf{I}a_{m+1}$. For the estimation of the additive value model, UTA methods solve the following LP problem:

$$[\min]F,\ F = \sum_{i=1}^{k}(\sigma^+(a_i) + \sigma^-(a_i))$$

subject to

for m=1 , 2, ..., k-1

$$\sum_{i=1}^{n} p_i u_i[g_i(a_m)] - \sigma^+(a_m) + \sigma^-(a_m) - [\sum_{i=1}^{n} p_i u_i[g_i(a_{m+1})] - \sigma^+(a_{m+1}) + \sigma^-(a_{m+1})] \geq \delta \quad \text{if } a_m \text{ P } a_{m+1}$$

or

$$\sum_{i=1}^{n} p_i u_i[g_i(a_m)] - \sigma^+(a_m) + \sigma^-(a_m) - [\sum_{i=1}^{n} p_i u_i[g_i(a_{m+1})] - \sigma^+(a_{m+1}) + \sigma^-(a_{m+1})] = 0 \quad \text{if } a_m \text{ I } a_{m+1}$$

$$\sum_{i=1}^{n} p_i = 1$$

$p_i \geq 0$, for i=1, 2,..., n

$\sigma^+(a_j) \geq 0$, $\sigma^-(a_j) \geq 0$, for j = 1,2, ..., k

where δ is a small positive number; $g_i(a_m)$ the evaluation of the a_m action on the i-th criterion and $u_i[g_i(a_m)]$ the corresponding marginal value; and $\sigma+(a_j)$, $\sigma-(a_j)$ the under (over)estimation errors concerning the j-th action.

The results of the LP could be the estimation of:

1. One and only one solution indicating a robust preference model.

2. Infinite solutions which are bordered into a hyper-polyhedron (low robustness). In this case the beaten track is to move to post-optimal analysis in order to estimate a mean solution of the LP. The most familiar approach for post-optimal analysis, used in MINORA [13] and MIIDAS [12] systems, is oriented to the approximation of a barycenter solution maximizing the weights of every one of the criteria [11]. This barycenter solution is used as the working preference model for the next steps of D-A approach.
3. No solution, where the DM's preferences cannot lead to the estimation of a curved hyper-polyhedron.

The case of low robustness is the most frequently observed. Many recent research studies [5] focuses on the avoidance of low robustness, intervening into the initial preferences of the DMs so as to increase robustness a priori. Low robustness can be the result of bad structuring of the problem formulation (absence of one or more criteria, non-rational evaluation of the alternative actions on the criteria, non effective selection of the reference set etc.). In many cases, low robustness can be a mirroring of what DM has really in his/her mind. Low robustness is not necessarily a bad situation, since very useful information can be uncovered about DM’s preference structure, through an in depth analysis and investigation of the estimated preference model and the post-optimal analysis results.

The Disaggregation - Aggregation (D-A) approaches and especially UTA methods can be applied in collective decision making situations, either by the construction of a collective additive value preference model, incorporating techniques of the Social Choice Theory (a priori aggregation of individual preferences) or by the estimation and composition of the individual preference models [15],[16] (a posteriori aggregation of individual preference models). The main target of D-A approach is not to indicate the decision to be taken, but to support the analysis of the individual and collective preference models. \this would lead to knowledge improvement about the preferences' structure of the involved stakeholders, so as to support the actions of the next step such as compromises and negotiations in a collective decision making environment. Low robustness usually presented in the estimated preference models (collective and individuals) increases the complexity of the process for the approximation of collective preference models. Estimated preference models with low robustness, on the one hand complicates the decision problem while on the other hand helps the analysis of DMS preferences' structure and supports the negotiation processes and the approximation of areas of convergence in sensitive points of the decision space.

The interactive nature of D-A approach could be more strengthened by the utilisation of the robustness analysis of the individual and collective estimated preference models. Robustness analysis can enrich knowledge concerning the preference attitudes of the DMs. This enrichment of the knowledge about the preference structures of the involved stakeholders can trigger a set of interactive feedbacks, which could be either simple adaptation of the problem formulation and the DM's global preferences or more complicated ones leading to the refining of the preference models, using additional preference information.

There are a lot of cases where D-A approaches were utilised in order to support collaborative decision making, mainly in cases of small group of DMs. Spyridakos et al [14] applied UTA II and Cook and Seiford model [4] in order to assess a common accepted value function, evaluating all the executive positions in a large organization. Beuthe and Scanella [1] undertook an exhaustive analysis showing the relation between the "quality" of results and the value of the parameters involved in the different UTA methods. Kadziński et al [8] proposed the concept of a representative value function applied in robust multicriteria problems and solved with the aid of an extension of UTAGMS and GRIP methods. Kerstens et al [9] develop geometric representations of the mean-variance-skewness (MVS) portfolio frontier using the shortage function and related approaches. Spyridakos and Yannacopoulos [15] proposed the RACES software, incorporating the social choice functions for aggregating individual rankings with MINORA [13] and MIIDAS [12] systems, in order to assess value function(s), as compatible as possible with a collective ranking.

This research work is oriented on the Robustness analysis of the estimated preference models for small group collaborative decision making. The main aim of this study is to utilise the results of robustness analysis, in order to support the deeper analysis of the different preference structures of the DMs participating into the decision making and identify points or areas of differentiation and convergence. Also, a set of new tools for the measurement and visualization of the preference models robustness is proposed, in order to explain and survey the different attitudes of the participating stakeholders. Through this process a comparison among the stakeholders' preferences structures can be achieved. In addition, new interactive feedbacks, triggered by the robustness analysis results, are designed and proposed aiming to achieve a better convergence among the participating DMs.

The paper comprises an introduction and four sections. The additive value models robustness analysis is presented in the second section as well as the way knowledge can be extracted concerning the DMs' preferences status, who participate in collaborative decision making. The third section includes a set of indexes and 3-D visual techniques that support the robustness analysis of the estimated preference models by UTA methods for collaborative decision making. The process is illustrated in the next section through a real world case study, which are demostrates the new interactive capabilities of the proposed approach for small group decision aid. Finally conclusion and further suggestions are outlined in the last section.

2 Robustness Analysis of Collective and Individual Preference Models

Spyridakos and Yannacopoulos [15] presented a methodological frame and a software (RACES) for small group collaborative decision making, utilising the UTA STAR Method [11]. The process includes two alternative paths. The first one concerns am a priori aggregation of the individual global preferences to a collective pre-ranking of the alternative actions exploiting one of the social choice functions [2],[3], [6]. Following that, the estimation of collective preference additive value models is triggered, taking advantages from the collective pre-ranking. The second (a posteriori aggregation of individual preferences) functions in the opposite direction, the individual additive value models being estimated for every one of the individual pre-rankings and then synthesised to a collective weighted one. The main aim of this approach is to provide machanisms for the support of the analysis of the individual and collective preference models and to enrich the knowledge of the decision problem and the structure of DM's preferences.

The estimated additive value models (Collective and Individual) present (in most of the cases) low robustness. The post optimal techniques of MINORA and MIIDAS systems are activated in cases of low robustness, aiming to approximate the convex hyper-polyhedron of the LP solutions by applying one of the following algorithms:

- Maximisation of the criteria weights which leads to the estimation of a hyper-polyhedron with at most n vertices (n the number of criteria).
- Maximasation and minimisation of criteria weights which leads to the estimation of a hyper-polyhedron with at most 2n vertices.
- Manas-Nedoma [10] Algorithm which estimates all the vertices of the hyperpolyhedron through a set of relating steps.

The mean solution (barycenter,Figure 1) is selected as a working solution for the further exploitation.

Let be V_i^t the vertices of the hyper-polyhedra of the individual preference models, $V_i^t = \{p_{i1}^t, p_{i2}^t ... p_{in}^t\}$, i=1..k (k the number of vertices) and n the number of criteria and t=1...m, m the number of Decision makers and $V_i = \{p_{i1}, p_{i2} ... p_{cn}\}$, i=1...k the vertices of the Collective Preference Model.

The barycenter solution is estimated by

$V_b^t = (p_{b1}^t, p_{b2}^t ... p_{bn}^t)$ where

$p_{bj}^t = (\sum_{i=1}^{k} p_{ij}^t)/k$, j=1... n, t=1...m for the individual preference models

and

$V_b = (p_{b1}, p_{b2} ... p_{bn})$ where

$p_{bj} = (\sum_{i=1}^{k} p_{ij})/k$, j=1... n for the collective preference model.

The robustness analysis and feedbacks provided in this work have two directions:

- to increase robustness in the individuals and collective preference models in order to lead to a more stable decision making
- to determine significant factors of the preferences among the DMs which differentiate them or common points thaat may be a starting point, triggering convergences in a compromising process.

For the analysis, a set of indexes are used measuring the degree of robustness of individual preference models:

A. The range between minimum and maximum values of the criteria weights estimated by the post optimality analysis) for the t individual preference model is estimated by

$\mu_i^t = (Max(p_{ji}^t) - Min(p_{ji}^t))$, i=1 ... k, j=1... n, t=1...m.

The index μ_i^t is used in order to identify variations in the ranges of criteria weights and is a major measure for the robustness.

B. The normalized Euclidean distances of the vertexes of the high dimensional hyper-polyhedron.

$$D_{ir}^t = \sqrt{\sum_{q=1}^{n} (p_{iq}{}^t - p_{rq}{}^t)^2 / n} \quad \iota, r=1\ldots k \text{ and } i \neq r, \ t=1\ldots m$$

This index is used in order to identify vertices which are far away from each other.

C. The normalized Euclidean distance between the vertices of the hyper-polyedron and the Barycenter.

$$D_{bi}^t = \sqrt{\sum_{q=1}^{n} (p_{bq}{}^t - p_{iq}{}^t)^2 / n}, \quad i=1\ldots n, \ t=1\ldots m$$

It is used for the comparison of the individual and collective preference models.

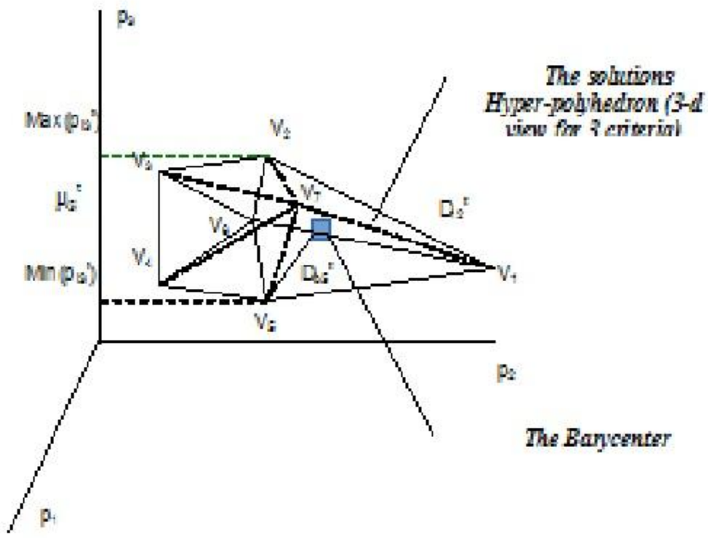


Fig. 1. Geometric representation of Solutions hyper-polyedron for 3 criteria

.

D. Euclidean distance among the Barycenters (Mi) of the Individual Preference Models

$$DM_{ij} = \sqrt{\sum_{q=1}^{n} (p_{bq}{}^i - p_{iq}{}^j)^2 / n} \quad i,j=1,2, \ r \text{ (r: Decision Makers)}$$

E. Euclidean Distance among the Barycenters of the Collective preference Model (M0) and the Individual ones (Mi).

$$\mathrm{DM_i} = \sqrt{\sum_{q=1}^{n} (p_{bq} - p_{bq}{}^{i})^2 / n}\ ,\ \ \text{i=1,2, r (r: Decision Makers)}$$

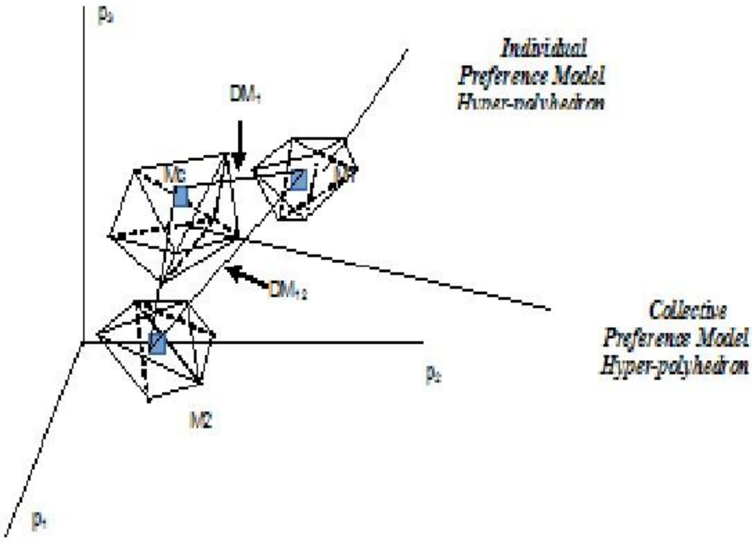


Fig. 2. Individuals and Collective Hyper-polyhedra for three criteria

3 Robust Analysis through Visual and Interactive (RAVI) approaches

The RAVI approach is tre product of the research effort to further supporting of the interactive feature of D-A approach and constitutes one more way for the analysis of the assessed additive value preference models. This robustness analysis operates in a synergistic manner with the existing functions of MINORA and MIIDAS systems and enriches the feedbacks and the interactivity of the preference models' construction. The aim of these feedbacks is to estimate more robust preference models with better convergence among them. This can be achieved through processes that reduce the volumes of the hyper-polyhedra (increase the robustness) and simultaneously converge their centers, exploiting new preference information which can be derived by the DMs for specific ranges of the Decision Space.

A crucial point of this process is to examine how the following questions can be answered:

- How robust are the assessed individual preference models?

- Which criteria make DMs more or less sensitive to changes?
- How far away lie the individual preference models into the decision space and which are their major differences?
- Which criteria present higher or lower differentiations among the DMs.
- Which are the criteria with accepted regions of weights for all or the majority of DMs.

The search for answers to the above mentioned questions can result in a set of interactive processes aimed to assessing preference models that meet best the needs of the examined case. The estimation of the previous indices as well as the visual 3-d graphics of the RAVI system (new software developed for the needs of this research work) can provide a frame to work, exploiting the results of robustness analysis. The interactive processes can be one or more of the existing ones in the systems MINORA and MIIDAS such as:

- The adaptation of problem formulation such as the criteria modelling, the alternatives' evaluation on the criteria and the selection of the reference set.
- The reformulation of DMs preferences as they were expressed in the pre-ranking or in the evaluation of the alternatives on the criteria.

In addition interactions can be implemented aiming to reduce the low robustness of individuals and collective preference models in order to achieve a better convergence. This includes two interactive processes. In the first one, through the analysis of the indexes and the 3-d visualisation of the hyper-polyhedra, common ranges of the criteria weights among the DMs can be identified. Following, lower ranges of the criteria weights could be obtained, exploiting focused dialogues with the DMs expressing their attitudes,. New conditions in the form of $p_{ji} > q_{ji}$ and $p_{ji} < q_{ji}'$, i=1...n, where [qi, q'i] are the new ranges of the criteria weights so as $[q_{ji}, q_{ji}'] = (Max(p_{ji}^t) - Min(p_{ji}^t))$, i=1 ... k, j=1... n are inserted into UTA LP programme, which will lead to the estimation of new preference for both individuals and collective models. Therefore higher robustness and a better convergence among the DMs will be achieved.

The second interactive process concerns the estimation of inter-priorities of the criteria weights either for a couple of criteria, a subset or all of them. This is easily achieved by utilizing a small set of virtual or real alternative actions very carefully selected in order to efficiently determine preferences concerning the importance of the criteria by the individual DMs. The additional preference information acquired by the DMs can support the identification of priorities between two or more criteria and enrich the UTA linear Programmes with one or more conditions in the form:

$p_i > p_j$, i,j=1...n and $i \neq j$, in preference models where the intersection of the criteria weights ranges is not null.

The estimation of the new preference models is expected to be more robust, leading to a shrinking of the hyper-polyhedra.

The above described interactive feedbacks can be implemented for these preference models that exhibit low robustness and high range of the criteria weights. The main target of this process is to shrink the Hyper-polyhedra through a refining process where the global preferences of the DMs cannot be changed.

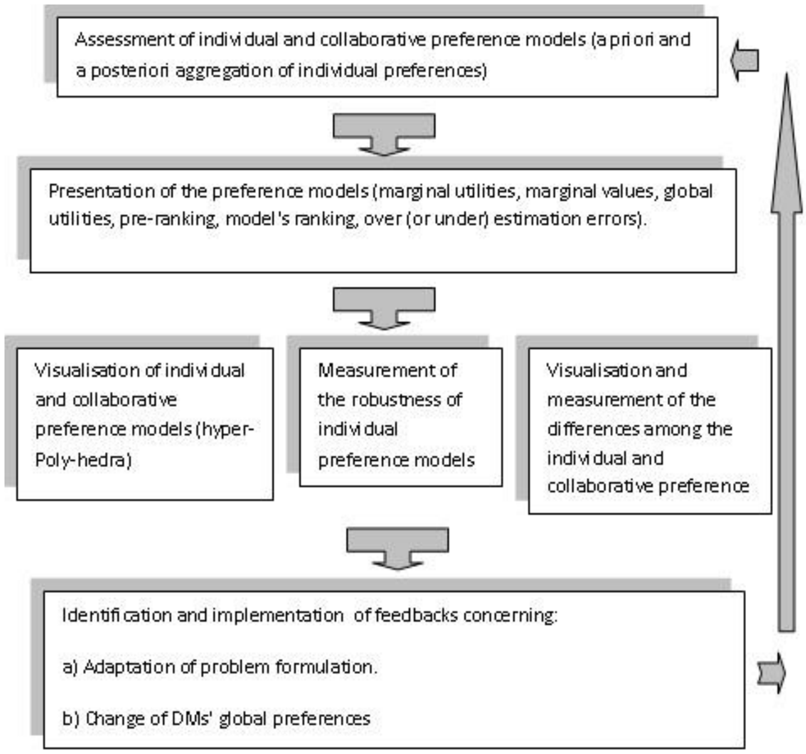


Fig. 3. *Steps of the Preference Models Robustness Analysis and Feedbacks*

The system RAVII includes a set of graphical presentations which facilitates the picturing of the robustness of the estimated preference models. Specifically it enables:

- Presentations in three-dimensional form of the hyper-polyhedra for three criteria, which are selected by the user in an interactive manner.
- Rotation of the hyper-polyhedra both horizontally and vertically, so as the DM can have full view of their shape, using polar coordinates and a set of scroll bars which allow the change of the view position.
- Calculation and presentation of Robustness indices in tables beside the graphic presentations,

Figure 2 presents the four steps of the proposed process. In the first we estimate the preference models (individuals and collective) and in the second stage the we present and analyse the results to the DMs similarly to the way we adopt in MINORA and MIIDAS systems. Provided that , the results will be clear and understandable to DMs we can move forward in the robustness analysis. The next two steps involve the robustness analysis of the estimated preference models and the identification and implementation of feedbacks with the procedures described above for the assessment of new individual and collective preference models. Through the new preference models we intend to achieve higher robustness and possible better convergence among DMs preference models.

4 Illustration Example

The proposed methodological approach will be presented through an example concerning the job evaluation (positions of authority). The evaluation was commissioned by four experts (Decision Makers) and their duty was to evaluate 25 jobs (p-1, p-2, , .., p-25). For the evaluation of the jobs the committee used 6 criteria (qualifications required by the position holder, the staff, decisions taken, multiplicity of the tasks, responsibility of the position and budget handling). From a set of 25 positions, 13 were selected as a reference set and the DMs rank-order them separately, expressing their global preferences. Also, a collective ranking was calculated using the Borda function, a simple and credible method for the purposes of this illustration (Table 1).

Table 1. DMs' pre-rankings (global preferences) and collective raking (Borda Social Choice function)

Jobs	DM1	DM2	DM3	DM4	Collective
p3	1	2	4	6	3
P13	2	1	2	1	1
P7	3	3	1	2	2
P17	4	4	3	3	4
P4	5	6	5	7	5
P22	6	5	7	5	5
P16	7	7	8	9	7
P10	8	8	6	4	6
P6	9	9	10	8	8
P18	10	10	11	10	9
P23	11	11	12	11	10
P9	12	12	9	13	11
p19	13	13	13	12	12

The four individual additive value models were estimated as well as the collective utilising the DMs' pre-rankings and the collective one. Table 2 presents the weights of the assessed individuals and collective preference models (minimum, maximum and average weights) as well as an index μ indicating the robustness measurement. Figure 3 presents in 3-d graph, views of the hyper-polyhedra where M is the barycenter of the collective preference model and M1, M2, M3, M4 the barycenters of the individuals ones.

Table 2. Criteria weights (after Post Optimal Analysis) for individuals and collective preference models

Crite-	DM1 (weights)	DM2(weights)	DM3 (weights)

ria												
	min	Mean	max	μ_i^1	min	mean	max	μ_i^2	min	mean	max	μ_i^3
Cr 1	0	0.036	0.211	0.121	0	0.023	0.121	0.054	0.398	0.426	0.452	0.054
Cr 2	0	0.12	0.282	0.279	0	0.135	0.279	0.016	0.144	0.152	0.16	0.016
Cr 3	0.211	0.257	0.333	0.115	0.112	0.189	0.227	0.052	0.162	0.198	0.214	0.052
Cr 4	0.198	0.266	0.336	0.093	0.217	0.244	0.31	0.032	0.058	0.074	0.09	0.032
Cr 5	0.158	0.189	0.264	0.096	0.167	0.223	0.263	0.044	0.105	0.127	0.149	0.044
Cr 6	0	0.131	0.199	0.171	0.14	0.187	0.311	0.002	0.032	0.033	0.034	0.002

Criteria	DM4 (weights)			Collective (Borda)			
	min	mean	max	min	mean	max	μ_i^{Ψ}
Cr 1	0.058	0.058	0.058	0.051	0.061	0.165	0.114
Cr 2	0.275	0.275	0.275	0.03	0.185	0.279	0.249
Cr 3	0.081	0.081	0.081	0.115	0.184	0.231	0.116
Cr 4	0.102	0.102	0.102	0.165	0.191	0.22	0.055
Cr 5	0.227	0.227	0.227	0.178	0.204	0.235	0.057
Cr 6	0.257	0.257	0.257	0.114	0.173	0.23	0.116

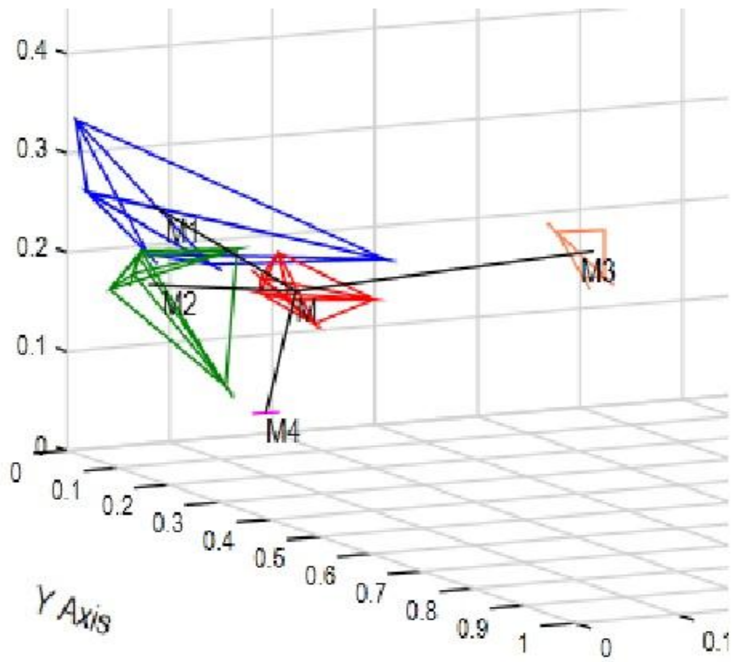


Fig. 4. ***3-D views of individuals and collective hyper-polyedra before the post analysis feedbacks (Color Lines (Blue - DM 1, Green - DM 2, Orange - DM3, Coral DM 4, Red - Collective).***

The analysis of the robustness brings to the fore some interesting results. DM4 preference model is totally robust, while all the others present a high level of low robustness. The higher low robustness presented in DM 1 followed by DM 2. The preference structures of DM1 and DM2 are quite close to each other which is pictured in the 3-d graph and the distance among their barycenters is small. Opposite, DM3 and DM4 seem to be differentiated from the first two DMs. For the needs of the presentation of the proposed approach, the focus will centre on DM3 given that:

The assessed preference model of DM3 has some inconsistencies with the pre-ranking since alternatives p-6 and p-16 are ranked in other positions in relation to what DM ranked them (Figure 5).

Table 3. : Euclidean Dinstances of the BaryCenters

Barycenters	Euclidean Distance (before Feedback	Euclidean Distance (After FeedBacks)
M1 - M	0.0546	0,0436
M2 - M	0.03353	0,03517
M3-M	0.17004	0,13662
M4-M	0.07561	0,076
M1-M2	0.00987	0,00456
M1-M3	0.20613	0,10703
M1- M4	0.0999	0,08348
M2-M3	0.21308	0,13863
M2-M4	0.05674	0,06275
M3-M4	0.22312	0,20535

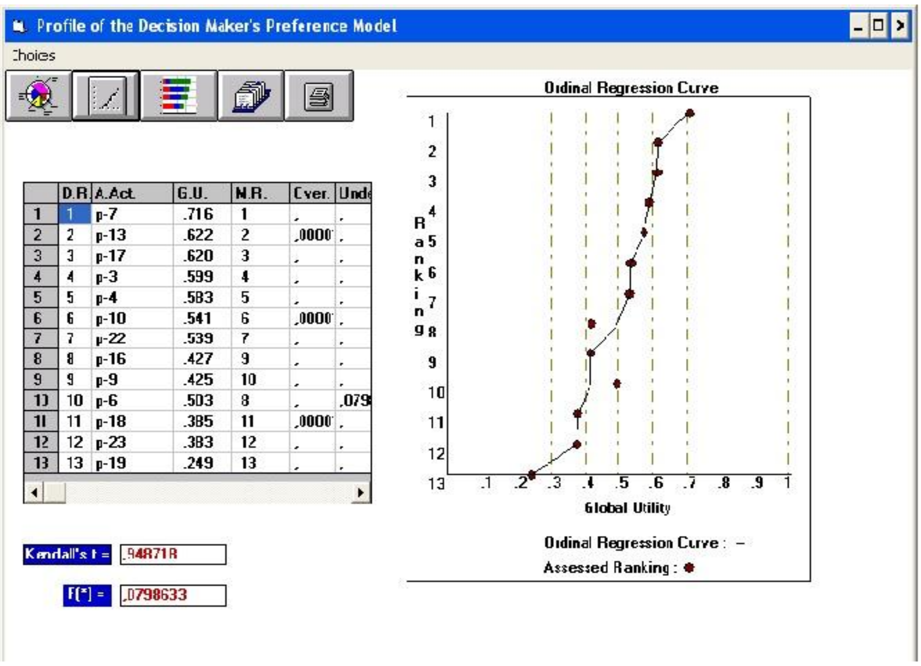


Fig. 5. Additive value model and ordinal regression curve for DM3 (Initial)

The Hyper-polyhedron of DM3 is far away from the hyper-polyhedra of the other DMs (Figure 4) and the collective one. Moreover, the distance of the Barycenter of DM3 is higher from the Collective Preference Model (0.17004) as well as from all the other DMs preference models (0.20613, 0.22803 and 0.22313 correspondingly) (Table 3). On the other hand DM3 seems to have a level of convergence with DM1 and DM2 in criterion2, since the intersection of the weights ranges is [0.144, 0.16]. The same is presented into other criteria but in a pair-wise manner. The above lead us to re-examine the case of DM 3.

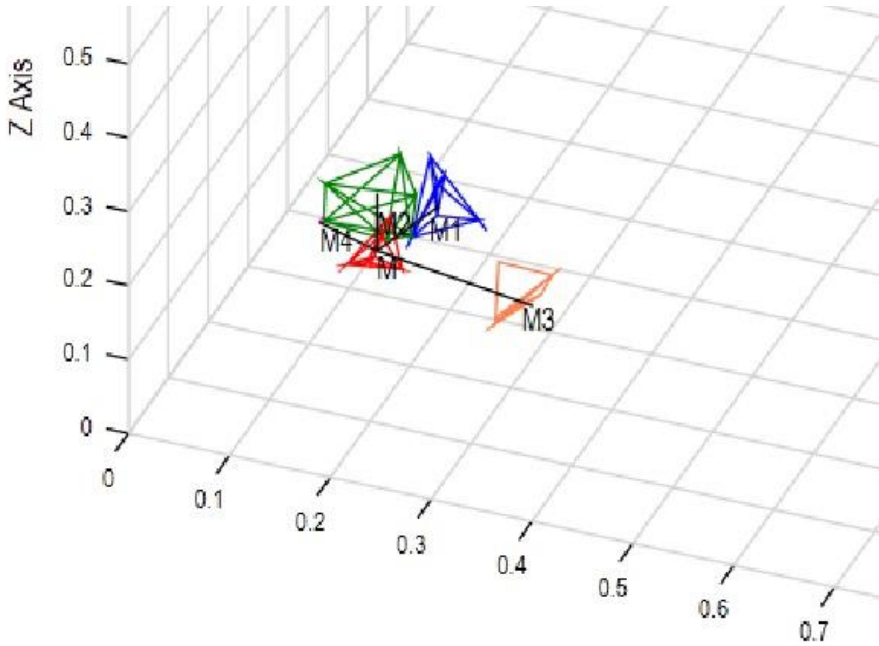


Fig. 6. ***3-D views of individuals and collective hyper-polyedra, after the post analysis feedbacks (Color Lines (Blue - DM 1, Green - DM 2, Orange - DM3, Coral DM 4, Red - Collective).***

An extensive dialogue was initiated with the DM3 trying to explain the inconsistencies of the estimated individual preference model, the low robustness and the high differentiation from the other DMs. The dialogue led to an alternation of DM3 global preferences, which resulted in the estimation of a new collective ranking (Table 4) and a new preference model of DM3, impacting also the new collective model (Figure 5). Also, another dialogue took place with DM1, DM2 in order to check the acceptance of the extreme (min, max) weights of Criterion 2. The dialogues point out to more strict ranges of the criteria weights for these two DMs. In our illustration example the new ranges of the weight for criterion 2 are [0.12, 0.2] and [0.1, 019] for DM1 and DM2 correspondingly. These conditions were inserted into the LP of UTA method and new preference models were estimated. These interventions bring out the estimation of new ranges of criteria weights, without affecting the final ranking of alternative actions. This constitutes a tool to bring closer the DMs preferences structures without changing their essential global preferences and the calculated ranking by the preference models. The new results presented in Figure 6 and Table 4 provide a better robustness of the preference models as well as convergence of the individual preference models.

Table 4. Criteria weights for individuals DM1, DM2 and DM3 after feedbacks

Crite-ria	DM1 (weights)				DM2(weights)				DM3 (weights)			
	min	mean	max	μ_i^1	min	mean	max	μ_i^2	min	mean	max	μ_i^3
Cr 1	0	0.034	0,2	0.2	0	0.0271	0.117	0.117	0.262	0.2767	0.321	0.059
Cr 2	0.12	0.15	0.2	0.08	0,1	0.1452	0,19	0.09	0.123	0.156	0.195	0.072
Cr 3	0,22	0.239	0,3	0.08	0.12	0.189	0.214	0.094	0.3106	0.328	0.343	0.0324
Cr 4	0.211	0.256	0.321	0.11	0.23	0.254	0.301	0.071	0.143	0.158	0.169	0.026
Cr 5	0.158	0.195	0.264	0.106	0.175	0.226	0.251	0.076	0	0.0221	0.053	0.053
Cr 6	0	0.126	0.167	0.167	0.14	0.158	0.289	0.149	0.032	0.056	0.091	0.059

5 Conclusions

The proposed approaches for handling preference models with low robustness provide new capabilities for the DMs' profiling in collaborative decision making. The interactivity in the processes for the estimation of the individual and collective preference models can be more efficiently supported, provided the knowledge of the preference structures is improved. In the case of collective decision-making, finding a better level of convergence among participants is often requested in order to avoid unpleasant situations, delays and cost increases. The visualization of the robustness and the comparison of the individual preference models with the collective one can provide an easy way to picture and contrast profiles of preference. The new feedbacks are included in the systems MINORA and MIIDAS for increasing the robustness of individual preferences models and convergence among them, as well as for impoving the interactive nature of D-A approach to collective decision-making environment. There is a lot of work that remains to be done yet, since robustness analysis of the preference models opens new promising directions for further development of the interactive nature of D-A approaches. Also, a lot of opportunities emerge from the exploitation of robustness in collective decision making and particularly in situations where conflicts among the stakeholders must be resolved. We address these unresolved in our future research undertakings.

Acknowledgement

This research has been co ‑ financed by the European Union (European Social Funds (ESF) and Greek national funds through the Operational Program "Education and Life long Learning" of the National Strategic Reference Framework (NSRF) ‑ Research

Funding Program: THALES. Investing in knowledge society through the European So cial Funds.

References

1. Beuthe, M. and G. Scannella (2001), Comparative analysis of UTA Multicriteria methods, European Journal of Operational Research, 130, 246-262
2. Black, D. (1958), The theory of Committees and Elections, Cambridge University Press, Cambridge.
3. Borda, J.C. (1781). Memoire sur les elections au scrutin. Histoire de L'Academie Royale des Sciences.
4. Cook W.D. and L.M Seiford (1978), Priorities ranking and consensus formation, Management Science, vol. 20, No. 16, pp. 59-73.
5. Greco, S., M. Kedzinski, V. Mousseau and R. Slowinski (2012), Robust Ordinal regression for multiple criteria group decision, UTA^GMS-GROUP and UTADIS^GMS-GROUP, Decision Support Systems, 52(3) 549-561
6. Hwang, C-l. and M-J. Lin (1988), Group Decision Making under Multiple Criteria in Lectures Notes in Economics and Mathematical Systems, Springer-Verlag,, Berlin, Heidelberg.
7. Jacquet-Lagrèze, E. and Y. Siskos (1982), Assessing a set of additive utility functions for multicriteria decision making, European Journal of Operational Research, Vol. 10, No.2, pp. 151-164.
8. Kadzinski, M. S. Greco and R. Slovinski (2012), Selection of a representative function in robust multiple criteria ranking and choice, European Journal of Operational Research, 217, 541-553.
9. Kersten K., A. Mounir and I. Van de Woestyne (2011), Geometric representation of the mean-variance-skewness portfolio frontier based upon the shortage function, European Jurnal of Operational Research 210, 81-94.
10. Manas, M. and J. Nedoma (1968) Finding all vertexes of a convex polyhedron, Numerische Mathematik, 12, 226-229.
11. Y. Siskos, and D. Yannacopoulos, (1985), UTASTAR, an ordinal regression method for building additive value functions, Investigacao Operacional 5, no 1, pp 39-53.
12. Siskos, Y., A. Spyridakos and D. Yannacopoulos, (1998), Using artificial intelligent techniques into preference disaggregation analysis: The MIIDAS system, European Journal of Operational Research, Vol. 113, 281-299.
13. Siskos, Y., A. Spyridakos and D. Yannacopoulos, (1993) MINORA: A multicriteria decision aiding system for discrete alternatives, Journal of Information Science and Technology, in: Y. Siskos and C. Zopounidis (eds), Special Issue on Multicriteria Decision Support Systems, vol. 2, no 2, pp. 136-149.
14. Spyridakos, A., Y. Siskos, D. Yannacopoulos and A. Skouris, (2001), Multicriteria Job evaluation for large organizations, European Journal of Operational Research, Vol. 130, pp. 375 - 387.
15. Spyridakos, A (2012), aggregation of individual preference models in collaborative decision making through disaggregation-aggregation approach with

the RACES system in A. Respicio and F. Burstein (Eds) Fusing decision support systems into the fabric of the contex, IOS Press, pp.241-252.

16. Spyridakos, A and D. Yannacopoulos (2013), Incorporating collective functions to multicriteria disaggregation – aggregation approaches for small group decision making, Annals of Operation Research, submitted.

Robust Discovery of Coordinated Patterns in a multi-Actor Business Process

Pavlos Delias[1], Michael Doumpos[2], and Nikolaos Matsatsinis[2]

[1] Department of Accounting and Finance, Eastern Macedonia and Thrace Institute of Technology, Kavala, Greece,
pdelias@teikav.edu.gr
[2] School of Production Engineering and Management, Technical University of Crete, Chania, Greece,
mdoumpos, nikos@dpem.tuc.gr

Abstract. In this work we propose a methodology based on the process mining approach to discover coordinated patterns of behavior in a customer service request handling process. We analyze a real dataset containing events from an incident and a problem management information system, and deliver results that eventually can raise the capacity of the company to manage the process. The core of the work comprises the looking for coordinated patterns among involved actors, a discrepancy analysis and a robust classification technique.

1 Introduction

Customer service request handling is a reactive business process that is triggered when a customer submits a service request to the help desk of a company. It has been identified as a core function of modern organizations, due to its tight relationship with their marketing function [1]. Establishing a service response capability includes a number of actions [2], like creating a service response policy, setting guidelines for communicating with outside parties regarding customer requests, selecting a team structure and staffing model, establishing relationships between the help desk team and other groups, both internal (e.g., technical support teams) and external, determining what services the incident response team should provide and staffing and training the incident response team.

There are multiple factors that affect the complexity of the process, such as the number of support teams involved, the organizational hierarchy, the number of products / product categories being served, special business rules etc. Due to the complexity of this process, special IT systems are often employed. A common practice reference model that introduces standard best practices for IT service management is the Information Technology Infrastructure Library [3]. Nevertheless, the processes described in ITIL are deliberately non-prescriptive. In practice, the actual behavior can significantly vary, not just according to the organizational implementation but because of a plethora of other implementation parameters as well (e.g. the resource performing the activities). Process mining

[4] is a promising approach to expose the real behavior of the process from IT systems' logs.

The process mining approach has recently attracted researchers for the service request management process analysis [5]. Since the respective process takes place in a highly flexible environment, multiple techniques are typically combined to deliver a solution. In [6], authors propose a combination of trace clustering and text mining to enhance process discovery techniques with the purpose of retrieving more useful insights from process data, while in [7] process mining is used to assess whether a business process is implemented according to ITIL guidelines. In this work we propose a methodology based on the process mining approach to discover coordinated patterns of behavior in a customer service request handling process. The process perspective is a necessary dimension of the proposed methodology, since ordinary data mining techniques would fail to capture the sequencing of the related events. Eventually, the results of this methodology can be used to raise the capability of the company to handle service requests by *i)* establishing more robust response policies and procedures and *ii)* aid the teams' structure decision, including outsourcing considerations. The basic steps of the proposed methodology is to arrange data with a process perspective (yet over multiple views), to draw the pertinent social networks, to perform a discrepancy analysis for the observed behavioral variation, and to apply a robust classification technique to explain the factors affecting the behavior and to deliver a predictive model for undesired behaviors as well.

2 Case Study

2.1 Description of the Case and the Dataset

Volvo IT Belgium provided a dataset[3] from its information system that supports the incidents management for the 2013 edition of the BPI challenge. The dataset contains events from an incident and a problem management information system. The primary goal of the incident management process is restoring a customer's normal service operation as quickly as possible when incidents arise ensuring that the best possible levels of service quality and availability are maintained. The dataset contains 65533 timestamped events related to the incident management process. Each record contains a number of variables such as the unique ticket number of the service request, the impact of the case (a measure of the business criticality of the incident), the case status (queued, accepted, completed or closed) and sub-status (assigned, awaiting assignment, cancelled, closed, in progress, wait or unmatched), the business area of the user reporting the incident, the technology-wise division of the organization, the support team that will try to respond to the service request and the location that takes the ownership of the support team.

The process is roughly the following: A customer submits a service request. The process reactively triggers a "first line" response, in other words, the Service

[3] doi:10.4121/500573e6-accc-4b0c-9576-aa5468b10cee

Desk or the Expert Help Desk tries to resolve the issue. When this is not possible, the case should be escalated to Second Line and/or Third Line teams. The quick resolution of the issue is defined within Service Level Agreement of the company.

2.2 Description of Patterns

There is an announced policy of the company that most of the incidents need to be resolved by the first line support teams (mainly service desks). This is called "Push to Front" tactic and it is mostly a matter of efficiency. Pushing to Front, allows the 2nd and 3rd line support teams to focus on their special, more demanding tasks (usually not related to customer service support). Unless this tactic is consistently applied a lot of 'easy', big volume cases will end up in those lines. The definition of push to front in this paper refers to the case when the 1st line support teams can resolve the service request without interference of a 2nd or 3rd line support team. As such, pushing to front is an important coordinated pattern that may arise during the process execution.

Besides pushing work towards the front, any team upon receiving a task can either try to resolve the issue by itself or hand over the task to another team (of the same or of another line). Handover of work is an ordinary action, however if this is excessively used it may have an inadmissible effect on process efficiency. Namely, extensive handover may reveal dodging or deferring behavior. The opposite (extensive takeover) may also reveal some undesired elements like lack of collaboration mentality of lack of knowledge transferring. Therefore, the inter-team handovers may also include coordinated patterns of (social) behavior.

A special case of handover of work is when support teams send the same case to each other again and again. We shall call this undesirable situation "Ping Pong". The definition of "Ping Pong" that we use in this work is that a Ping Pong occurs when a support team is revisited during the case, after it has passed the work to another team. However, we count a single Ping Pong per support team, even if this is revisited multiple times. This definition allows for a numeric representation of the Ping Pong behavior (a case may have multiple Ping Pongs, yet attributed to different teams). Ping Pong is also an undesirable coordinated behavior that may affect significantly the process performance.

3 Looking for Patterns

The dataset in its original format contains a list of timestamped events. It is quite hard to elicit patterns of behavior from within this format, since the sequencing of events and their aggregation per case are not exploited. Therefore, the leading step is to reach a process perspective for the dataset. In particular, the methodology unfolds in the following stages:

1. Commit data to process format
   (a) Control flow-wise (trajectories of status / substatus changes)
   (b) Social-wise (transactions among support teams or lines)

2. Discover the process map and check the flows.
3. Get the social networks for the social-wise process view and analyze social behavioral patterns
4. Perform a discrepancy analysis to analyze how the state sequences are related to one or more covariates
5. Apply a robust classification technique for both explanatory and predictive purposes.

3.1 Control Flow-wise Patterns

Control flow refers to how the status / substatus of a case changes during its lifecycle. There are 13 distinct alternatives for the status / substatus of a case (presented in Table 1). Although the set of activities (status changes) is small, we noticed that there are 2278 different variants of the same process (for a dataset of 7554 cases). Out these 2278 variants, just 88 have a frequency higher than 100, while the dominant variant represents just a 23% of total cases, a fact that confirms that the process environment is highly flexible.

Since there is no strict sequencing rule, discovering an exact behavior would not reflect the real situation, and would probably be of little importance. In general terms, cases go from some *Accepted* substatus to either a *Completed* substatus or to *Queued*. In the latter option, the case returns to an *Accepted* substatus. A process map is depicted in Fig. 1, where some labels for performance measures are printed. In particular, the heavier the weight of an edge, the worst its performance. The illustration has been created using Disco® [8] and it is a direct way to visualize the process' bottlenecks. The largest delays happen between Completed-Resolve and Completed-Closed (7.2 days), Accepted-Wait User and Completed-Resolve (5.3 days) and Accepted-Wait Implementation and Completed-Resolved (4.7 days). It is also interesting to regard that there is a meantime of 4.3 days between the Completed-Closed status and the Accepted-In Progress status, a fact that indicates that some cases are closed only to be re-inititiated after 4-5 days.

Status	Substatus
Accepted	Assigned, In Progress, Wait, Wait-User, Wait-Customer, Wait-Implementation, Wait Vendor
Queued	Awaiting Assignment
Completed	In Call, Resolved, Closed, Cancelled
Unmatched	Unmatched

Table 1: Status and Substatus alternatives

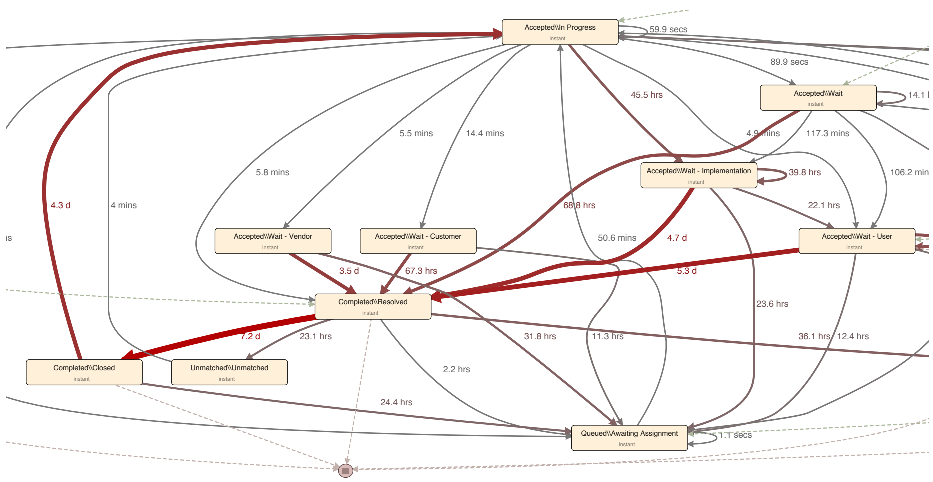


Fig. 1: Process Performance Map

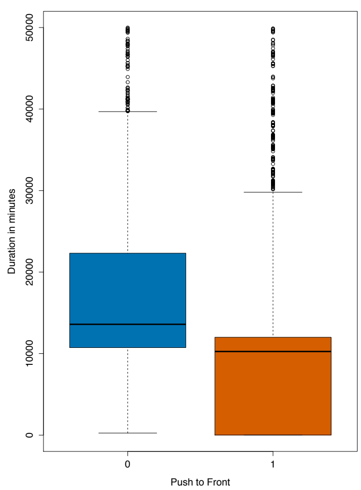
3.2 Social-wise Patterns

First of all, we need to evaluate the "Ping Pong" and the "Push to Front" patterns for each case, based on the descriptions of section 2.2. To this end, the following R [9] script was developed.

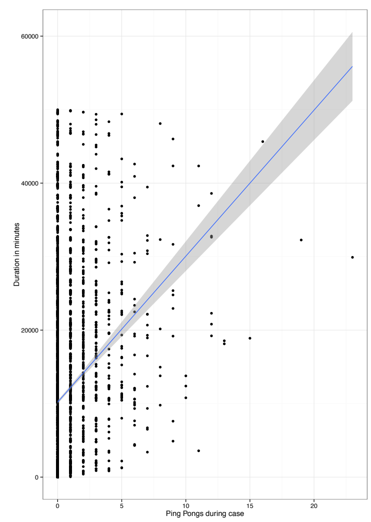
```
#---Evaluate Ping Pong behavior---
PingPong<-c();
#...Loop over traces....(traces contain Support Teams as activities)
Rle<-rle(traceRow)
#Does the case Ping Pong?
PingPong<-c(PingPong,sum(duplicated(Rle$values)))

#---Evaluate Push to Front behavior---
PushtoFront<-c();
#...Loop over traces....(traces contain Lines as activities)
Rle<-rle(traceRow)
#Does the case Push to Front?
if(Rle$values[1]=="1st" & length(tempRle$values)>1){
   PushtoFront<-c(PushtoFront,0)
}else {
        PushtoFront<-c(PushtoFront,1)
}
```

As expected, both behaviors have a negative effect on the case duration. Figure 2 illustrates these effects for the mainstream cases (outliers, i.e. cases that last more than 50000 minutes are removed). While for *Push to Front* a binary variable is sufficient, for *Ping Pong* a numerical scale is preferred. An illustrative argument for this choice is presented in Fig. 3. In this point we shall remind that a Ping Pong is assigned per team, i.e., even if a pair of teams handover their work multiple times during a case, that will still count for two (one for each team that is revisited).



(a) Duration of cases that Push to Front (1) or not (0)



(b) Duration of cases over the number of Ping Pongs they contain

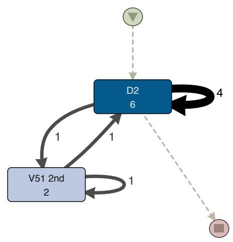
Fig. 2: The effect on case duration

4 Analyzing the Relevance of Factors

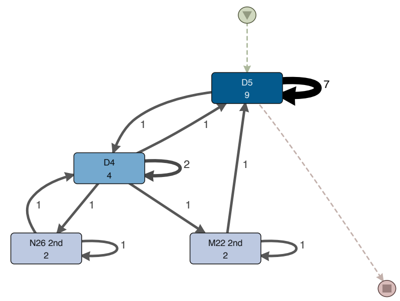
4.1 Discrepancy Analysis

In a case evolving framework, discrepancy measures the between-case variability of the case lifecycle trajectories. Therefore, higher discrepancy, for example, would reflect a greater level of uncertainty about the path followed by the cases. The discrepancy of sequences will be defined from their pairwise dissimilarities. Perhaps the most popular dissimilarity measure used for sequence analysis is the generalized Levenshtein distance. It is defined as the lowest cost of transforming one sequence into the other by means of state insertions–deletions and state substitutions.

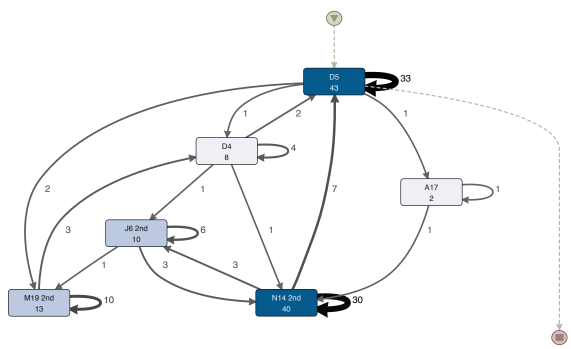
In this section, we integrate the sequence discrepancy analysis with the regression tree method introduced in [10]. The intuition of this regression tree method is the following: Start with all cases grouped in an initial node. Then,



(a) Single Ping Pong



(b) Double Ping Pong



(c) Multiple Ping Pong

Fig. 3: A numerical scale for the Ping Pong behavior is preferable

recursively partition each node using values of another variable. At each node, the variable and the split are chosen in such a way that the resulting child nodes differ as much as possible from one another or have, more or less equivalently, lowest within-group discrepancy. The process is repeated on each new node until a certain stopping criterion is reached. For the implementation of this method, we used the TraMineR [11] package of R.

As illustrated in Fig. 4, both social patterns (Push to Front and Ping Pong) result in clustered behaviors. In particular, the first split is among cases that Ping Pong or not (0 and greater than 0). Cases of the later category (no Ping Pong) last significantly less and visit a lot less frequently the "*Queued*" status. At the second level, leftmost the split is among cases that Push to Front (>0) and not (0). We regard that cases that Push to Front reach a "*Completed*" status earlier, and that their average duration is smaller. The rightmost split is again based on the Ping Pong behavior, but this time the critical value is two. Cases that Ping Pong more than twice spend an important percentage of their lifetime in a "*Queued*" status, and are naturally prolonged.

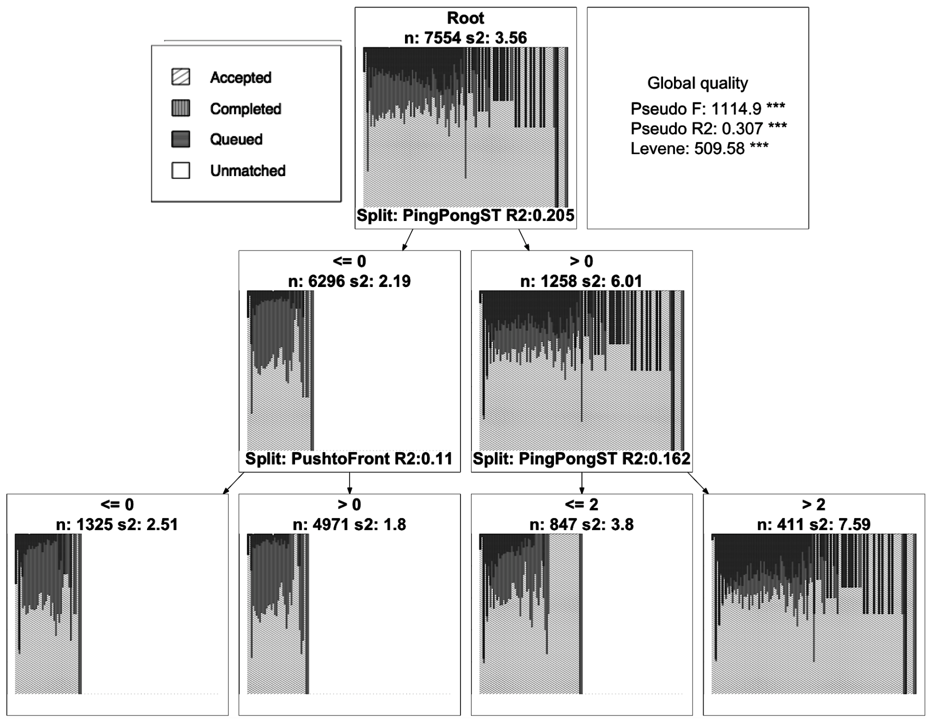


Fig. 4: Discrepancy Analysis for cases lifecycle trajectories

4.2 Binary Classification

Support Vector Machines (SVM) is one of the most well-known supervised classification algorithms. It was originally proposed by Vapnik [12]. The intuition of SVM is that the goal is to get an hyperplane that optimally distinguishes two classes of data. The major advantage of SVM is its minimal generalization error (at least in the case of binary classification - two classes of data) reached computationally efficiently. The SVM is one of the most applied algorithm of robust optimization in data mining. For a thorough exploration of theoretical and practical issues, we cite the classic work [13] and the works of Trafalis et al. [14] and Xu et al. [15]. We used 10-fold cross validation on a training data set of case-label pairs $(x_i, y_i), i = 1, \ldots, 7$, where $x_i \in \Re^n$and $y \in \{-1, 1\}^7$. Number 7 indicate that seven factors (Country, Impact, Line, Function, Organization, number of Events and Push to Front) were examined to predict the Ping Pong behavior. We used a linear kernel, as implemented by the LIBSVM library [16]. The overall accuracy of the model (for all folds, both classes) was 89.48%, but what is more important is to try to explain the factors that appear to be the most critical. According to [17], in linear SVMs, the use of w_i^2 can be justified as a feature ranking criterion. Therefore, the following interesting points emerged:

- We identified that there are 3 countries (China, Sweden and U.S.A.) whose support teams are more prone to Ping Pong.
- The impact of cases does not appear to have an effect
- Ping Pong appears the most when cases are initiated in the front line.
- There are some particular Function Divisions and Organizations that are more prone to Ping Pong behavior
- Pushing to Front seems to have a negative impact
- As expected, the number of events per case is the most critical predictor of Ping Pong behavior

Overall, this paper applied a process mining approach to explore a real case study with the goal to provide insights to this implicit business process and to raise the capability of the company to handle service requests. The results presented in the previous sections allow the company to reach evidence-based response policies. In addition, since the identified issues are localized (certain support teams, certain divisions etc.), the evidence provided could aid company's decision about the teams' structure.

Acknowledgement

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning"

References

1. Alan Wilson, Valarie A Zeithaml, Mary Jo Bitner, and Dwayne D Gremler. *Services marketing: Integrating customer focus across the firm.* Number 2nd Eu. McGraw Hill, 2012.
2. Tim Grance, Karen Kent, and Brian Kim. Computer security incident handling guide. *NIST Special Publication*, pages 800–61, 2004.
3. A. Hochstein, R. Zarnekow, and W. Brenner. Itil as common practice reference model for it service management: formal assessment and implications for practice. In *e-Technology, e-Commerce and e-Service, 2005. EEE '05. Proceedings. The 2005 IEEE International Conference on*, pages 704–710, March 2005.
4. Wil Aalst, Arya Adriansyah, Ana Karla Alves Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose, Peter Brand, Ronald Brandtjen, Joos Buijs, Andrea Burattin, Josep Carmona, Malu Castellanos, Jan Claes, Jonathan Cook, Nicola Costantini, Francisco Curbera, Ernesto Damiani, Massimiliano Leoni, Pavlos Delias, Boudewijn F. Dongen, Marlon Dumas, Schahram

Dustdar, Dirk Fahland, Diogo R. Ferreira, Walid Gaaloul, Frank Geffen, Sukriti Goel, Christian Günther, Antonella Guzzo, Paul Harmon, Arthur Hofstede, John Hoogland, Jon Espen Ingvaldsen, Koki Kato, Rudolf Kuhn, Akhil Kumar, Marcello Rosa, Fabrizio Maggi, Donato Malerba, Ronny S. Mans, Alberto Manuel, Martin McCreesh, Paola Mello, Jan Mendling, Marco Montali, Hamid R. Motahari-Nezhad, Michael Muehlen, Jorge Munoz-Gama, Luigi Pontieri, Joel Ribeiro, Anne Rozinat, Hugo Seguel Pérez, Ricardo Seguel Pérez, Marcos Sepúlveda, Jim Sinur, Pnina Soffer, Minseok Song, Alessandro Sperduti, Giovanni Stilo, Casper Stoel, Keith Swenson, Maurizio Talamo, Wei Tan, Chris Turner, Jan Vanthienen, George Varvaressos, Eric Verbeek, Marc Verdonk, Roberto Vigo, Jianmin Wang, Barbara Weber, Matthias Weidlich, Ton Weijters, Lijie Wen, Michael Westergaard, and Moe Wynn. Process mining manifesto. In Wil Aalst, John Mylopoulos, Michael Rosemann, Michael J. Shaw, Clemens Szyperski, Florian Daniel, Kamel Barkaoui, and Schahram Dustdar, editors, *Business Process Management Workshops*, volume 99, pages 169–194. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

5. Business process intelligence challenge 2013. In Boudewijn van Dongen, Barbara Weber, Diogo Ferreira, and Jochen De Weerdt, editors, *Proceedings of the 3rd Business Process Intelligence Challenge*, volume 1052. CEUR.
6. J. De Weerdt, S.K.L.M. vanden Broucke, J. Vanthienen, and B. Baesens. Leveraging process discovery with trace clustering and text mining for intelligent analysis of incident management processes. In *Evolutionary Computation (CEC), 2012 IEEE Congress on*, pages 1–8, June 2012.
7. Diogo R. Ferreira and Miguel Mira da Silva. Using process mining for ITIL assessment: a case study with incident management. Bournemouth University, April 2008.
8. Fluxicon. *Disco*. Fluxicon, 2012.
9. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
10. Matthias Studer, Gilbert Ritschard, Alexis Gabadinho, and Nicolas S. Müller. Discrepancy analysis of state sequences. *Sociological Methods and Research*, 40(3):471–510, 2011.
11. Alexis Gabadinho, Gilbert Ritschard, Nicolas S Müller, and Matthias Studer. Analyzing and visualizing state sequences in r with traminer. *Journal of Statistical Software*, 40(4):1–37, 4 2011.
12. Vladimir Naoumovitch Vapnik. *The nature of statistical learning theory*. Springer.
13. ChristopherJ.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.
14. Theodore B. Trafalis and Robin C. Gilbert. Robust classification and regression using support vector machines. *European Journal of Operational Research*, 173(3):893 – 909, 2006.
15. Huan Xu, Constantine Caramanis, and Shie Mannor. Robustness and regularization of support vector machines. *J. Mach. Learn. Res.*, 10:1485–1510, December 2009.
16. Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.
17. Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3):389–422, 2002.

ΕΠΙΛΟΓΗ ΕΥΡΩΣΤΟΥ ΧΑΡΤΟΦΥΛΑΚΙΟΥ ΕΠΕΝΔΥΤΙΚΩΝ ΣΧΕΔΙΩΝ ΜΕ ΜΑΘΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟ

Γ. Μαυρωτάς, Ο. Πετσάκ, Β. Παππά
Εργαστήριο Βιομηχανικής & Ενεργειακής Οικονομίας, Σχολή Χημικών Μηχανικών, Ε.Μ.Π., Ηρώων Πολυτεχνείου 9, Ζωγράφος, 157 80 Αθήνα

ΠΕΡΙΛΗΨΗ
Στη συγκεκριμένη εργασία μελετάται το πρόβλημα της επιλογής χαρτοφυλακίου επενδυτικών σχεδίων για ένα εύρος προϋπολογισμού. Τα επενδυτικά σχέδια αξιολογούνται μέσω πολυκριτηριακής ανάλυσης και μέσω της προκύπτουσας πολυκριτηριακής επίδοσης συμμετέχουν στη μία αντικειμενική συνάρτηση του προβλήματος. Το πρόβλημα τίθεται ως πρόβλημα Ακέραιου Προγραμματισμού με αντικειμενικές συναρτήσεις τη μεγιστοποίηση της πολυκριτηριακής επίδοσης του χαρτοφυλακίου και την ελαχιστοποίηση του κόστους του χαρτοφυλακίου. Με μια επαναληπτική διαδικασία στην οποία συμμετέχει ο αποφασίζων εκφράζοντας τις προτιμήσεις του, μειώνεται σταδιακά το εύρος του προϋπολογισμού έως ότου καταλήξουμε από ένα μεγάλο αριθμό αρχικών υποψήφιων χαρτοφυλακίων στο τελικό χαρτοφυλάκιο. Η επαναληπτική αυτή διαδικασία ονομάζεται ITA (Iterative Trichotomic Approach) και μας δίνει πληροφορίες για τον βαθμός εμπιστοσύνης με τον οποίο συμμετέχει το κάθε επενδυτικό σχέδιο στο τελικό χαρτοφυλάκιο. Επίσης μας δίνει πληροφορίες για την ευρωστία του τελικού χαρτοφυλακίου μέσω του δείκτη ευρωστίας.

Λέξεις κλειδιά: *Χαρτοφυλάκιο, Επενδυτικά σχέδια, Βελτιστοποίηση, Μαθηματικός Προγραμματισμός, Αβεβαιότητα, Ευρωστία*

ΕΙΣΑΓΩΓΗ
Το πρόβλημα της επιλογή χαρτοφυλακίου επενδυτικών σχεδίων (project portfolio selection) είναι μια διαδικασία που συναντάται πολύ συχνά σε διάφορους οργανισμούς. Το πρόβλημα τίθεται ως εξής: Να επιλεγεί το χαρτοφυλάκιο εκείνο των επενδυτικών σχεδίων που μεγιστοποιεί μια αντικειμενική συνάρτηση επίδοσης και συγχρόνως υπακούει σε συγκεκριμένες απαιτήσεις που μεταφράζονται σε κατάλληλους περιορισμούς. Ο κυριότερος περιορισμός έχει να κάνει με το κόστος γιατί, ο προϋπολογισμός δεν επαρκεί για το σύνολο των υποψήφιων επενδυτικών σχεδίων και πρέπει να γίνει επιλογή του καλύτερου υποσυνόλου αυτών (χαρτοφυλακίου). Επίσης υπάρχουν περιορισμοί πολιτικής (γεωγραφικοί, τεχνολογικοί κλπ), αλληλεξαρτήσεις μεταξύ των σχεδίων (αμοιβαία αλληλοαποκλειόμενα ή προαπαιτούμενα). Οι αλληλεξαρτήσεις μεταξύ των επενδυτικών σχεδίων είναι αυτές που δίνουν στο πρόβλημα έναν συνδυαστικό χαρακτήρα έτσι ώστε τα σχέδια να μην είναι τελικά ανεξάρτητα μεταξύ τους και η απλή ιεράρχηση (όπως π.χ. λαμβάνεται από την πολυκριτηριακή ανάλυση) να μην είναι αρκετή. Στην προκειμένη περίπτωση το πρόβλημα της συνδυαστικής βελτιστοποίησης μοντελοποιείται κι επιλύεται με ένα κατάλληλο μοντέλο Μαθηματικού προγραμματισμού και συγκεκριμένα ακέραιου προγραμματισμού.
Οι πρώτες εφαρμογές μαθηματικού προγραμματισμού σε τέτοια προβλήματα έγιναν τη δεκαετία του ’50 από τους Lorie and Savage [1]. Έκτοτε έχει χρησιμοποιηθεί πολλές φορές σε διάφορα προβλήματα [2]. Έχει μάλιστα συνδυαστεί με πολυκριτηριακή ανάλυση για την εξαγωγή των συντελεστών της αντικειμενικής συνάρτησης [3-7]. Η ενσωμάτωση της αβεβαιότητας ως προς κάποιες παραμέτρους έχει επίσης εξετασθεί από διάφορους συγγραφείς [8-10].

Στη συγκεκριμένη περίπτωση θεωρούμε ότι υπάρχει αβεβαιότητα ως προς τον διαθέσιμο προϋπολογισμό για τον οποίο ξέρουμε όχι την ακριβή τιμή του αλλά ένα διαθέσιμο εύρος. Μέσα στο εύρος υπάρχουν πολλά βέλτιστα χαρτοφυλάκια που για κάθε ύψος προϋπολογισμού δίνουν την καλύτερη λύση. Σκοπός μας είναι να δημιουργήσουμε μια αλληλεπιδραστική, επαναληπτική διαδικασία που σε συνεργασία με τον αποφασίζοντα να καταλήγει καταρχήν στο προτιμότερο χαρτοφυλάκιο εντός του εύρους και κατά δεύτερο λόγο θα μας δίνει πληροφορίες για το βαθμό εμπιστοσύνης στη συμμετοχή του κάθε σχεδίου στο χαρτοφυλάκιο. Τέλος, θα δίνει και πληροφοίες για την συνολική ευρωστία του επιλεχθέντος χαρτοφυλακίου. Όπως θα περιγραφεί στη συνέχεια θα χρησιμοποιήσουμε καταρχήν πολυκριτηριακό ακέραιο προγραμματισμό για να παράξουμε τα υποψήφια χαρτοφυλάκια εντός του εύρους του προϋπολογισμού και στη συνέχεια με την μέθοδο ΙΤΑ θα συγκλίνουμε στο προτιμότερο.
Η διάρθρωση της εργασίας έχει ως εξής: Στο δεύτερο κεφάλαιο θα περιγραφεί η μεθοδολογία της ΙΤΑ, στο τρίτο κεφάλαιο θα περιγραφή η μελέτη περίπτωσης με τα 133 σχέδια από τον ενεργειακό τομέα. Το 4^{o} κεφάλαιο είναι αφιερωμένο στην ανάπτυξη του μοντέλου και το 5^{o} κεφάλαιο στη συζήτηση των αποτελεσμάτων. Τέλος στο 6^{o} κεφάλαιο θα παρουσιαστούν τα βασικά συμπεράσματα.

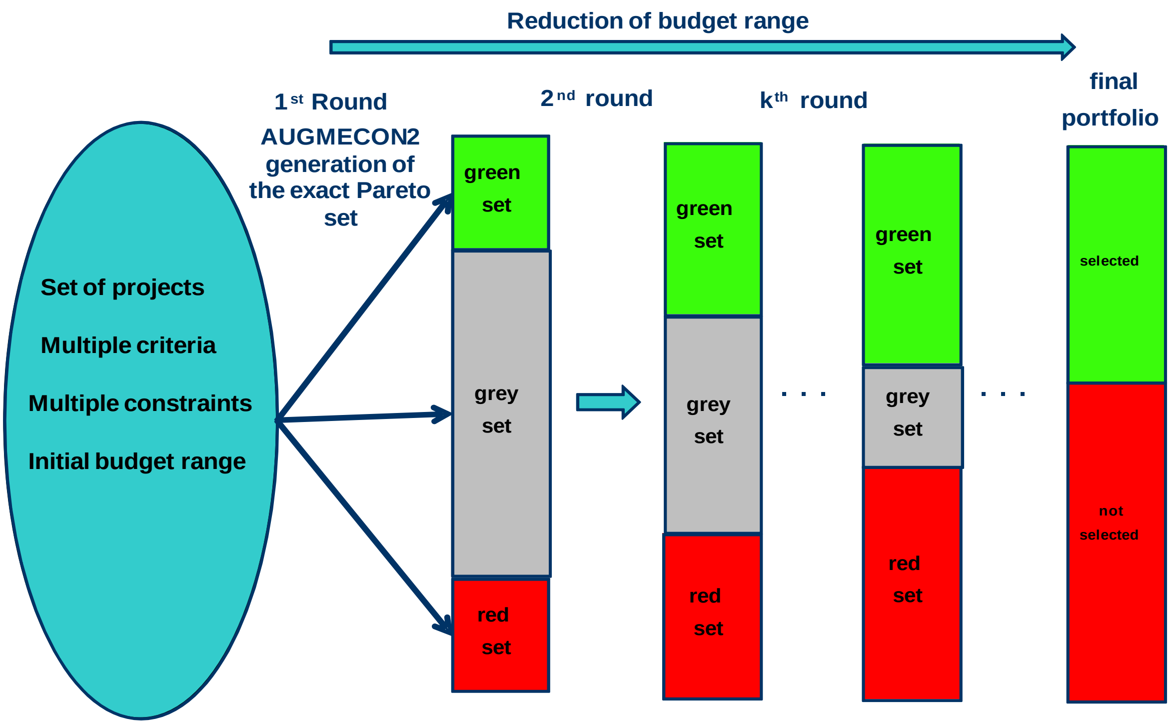
ΜΕΘΟΔΟΛΟΓΙΚΟ ΜΕΡΟΣ

Η μέθοδος ΙΤΑ

Η μέθοδος ΙΤΑ (Iterative Trichotomic Approach) είναι μια μέθοδος που αναπτύχθηκε για την επιλογή χαρτοφυλακίου επενδυτικών σχεδίων σε συνθήκες αβεβαιότητας [11]. Η αβεβαιότητα αφορά τους συντελεστές της αντικειμενικής συνάρτησης του μοντέλου ακέραιου προγραμματισμού που δημιουργείται. Συγκεκριμένα, η αβεβαιότητα έχει να κάνει με τους συντελεστές βαρύτητας των κριτηρίων που χρησιμοποιούνται για την πολυκριτηριακή επίδοση. Οι αβέβαιες παράμετροι χαρακτηρίζονται από συγκεκριμένες κατανομές πιθανότητας (στοχαστική προσέγγιση). Σχεδιάζεται στη συνέχεια μια διαδικασία προσομοίωσης Monte Carlo - Βελτιστοποίησης όπου γίνεται σε κάθε επανάληψη δειγματοληψία από αυτές τις κατανομές κι επίλυση του αντίστοιχου προβλήματος ακέραιου προγραμματισμού. Με τη διαδικασία αυτή και μετά από π.χ. 1000 επαναλήψεις παράγεται ένας μεγάλος αριθμός βέλτιστων χαρτοφυλακίων. Στη συνέχεια τριχοτομούμε (εξού και το όνομα της μεθόδου) το σύνολο των επενδυτικών σχεδίων σε τρία σύνολα. Το πράσινο σύνολο που περιλαμβάνει τα σχέδια εκείνα που εμφανίζονται παρόντα σε όλα τα παραγόμενα χαρτοφυλάκια, το κόκκινο σύνολο που περιλαμβάνει τα σχέδια εκείνα που δεν εμφανίζονται σε κανένα από τα παραγόμενα χαρτοφυλάκια και το γκρι σύνολο που περιλαμβάνει τα σχέδια εκείνα που εμφανίζονται σε μερικά από τα χαρτοφυλάκια. Τα επενδυτικά σχέδια του πράσινου συνόλου θεωρούνται ότι σε κάθε περίπτωση επιλέγονται, τα σχέδια του κόκκινου συνόλου θεωρούνται ότι σε κάθε περίπτωση απορρίπτονται και τα σχέδια του γκρί συνόλου χρειάζονται περαιτέρω διερεύνηση. Η διερεύνηση αυτή γίνεται με έναν επαναληπτικό τρόπο. Στον επόμενο γύρο μειώνεται η αβεβαιότητα των παραμέτρων στενεύοντας της αντίστοιχες κατανομές κι επαναλαμβάνεται η διαδικασία προσομοίωσης Monte Carlo – Βελτιστοποίησης. Σε κάθε επόμενο γύρο το πλήθος του πράσινου και του κόκκινου συνόλου αυξάνεται ενώ μειώνεται το πλήθος του γκρί συνόλου. Τελικά η διαδικασία συγκλίνει στο βέλτιστο χαρτοφυλάκιο, προμηθεύοντας όμως τον αποφασίζοντα και με πληροφορίες για την βεβαιότητα με την οποία εγκρίνεται ή απορρίπτεται κάποιο σχέδιο ανάλογα με τον γύρο στον οποίο εντάχθηκε στο πράσινο ή κόκκινο σύνολο.

Η μέθοδος ΙΤΑγια εύρος προϋπολογισμού

Στην προκειμένη περίπτωση θα θεωρηθεί ότι η αβεβαιότητα χαρακτηρίζει τον διαθέσιμο προϋπολογισμό. Ο προϋπολογισμός είναι γνωστός ως εύρος τιμών κι όχι ως συγκεκριμένη τιμή, με άλλα λόγια είναι χαλαρός περιορισμός του μοντέλου. Επιδιώκεται λοιπόν να βρεθεί το βέλτιστο χαρτοφυλάκιο εντός του εύρους τιμών έτσι ώστε ο λόγος αποτελεσματικότητας/κόστους να είναι αυτός που επιθυμεί ο αποφασίζων. Η αποτελεσματικότητα του χαρτοφυλακίου εκφράζεται ως το άθροισμα των επιδόσεων του κάθε σχεδίου που τελικά περιλαμβάνεται σε αυτό. Η επίδοση του κάθε σχεδίου είναι αποτέλεσμα πολυκριτηριακής ανάλυσης. Χρησιμοποιώντας πολυκριτηριακό μαθηματικό προγραμματισμό και συγκεκριμένα την μέθοδο AUGMECON2 [12] παράγουμε το σύνολο των χαρτοφυλακίων που θεωρούνται βέλτιστα για κάθε τιμή του συνολικού προϋπολογισμού μέσα στο συγκεκριμένο εύρος (η μέθοδος AUGMECON2 έχει τη δυνατότητα να παράγει το σύνολο των κατά Pareto βέλτιστων λύσεων σε προβλήματα πολυκριτηριακού ακέραιου προγραμματισμού). Στη συνέχεια παρουσιάζονται οι λύσεις αυτές στον αποφασίζοντα ο οποίος επαναληπτικά επιλέγει όλο και στενότερες περιοχές εύρους του προϋπολογισμού μέχρι να καταλήξει στο τελικό χαρτοφυλάκιο. Η περιοχή που επιλέγει έχει κάθε φορά το μισό εύρος από την προηγούμενη έτσι ώστε η διαδικασία να συγκλίνει σχετικά γρήγορα. Η διαδικασία φαίνεται στο Σχήμα 1.



Σχήμα 1. Σχηματική παράσταση της επαναληπτικής διαδικασίας

Στον πρώτο γύρο με τη μέθοδο AUGMECON2 παράγονται όλα τα Pareto βέλτιστα χαρτοφυλάκια με αντικειμενικές συναρτήσεις την μεγιστοποίηση της συνολικής επίδοσης και την ελαχιστοποίηση του συνολικού κόστους. Για την δεύτερη αντικειμενική συνάρτηση βάζουμε άνω και κάτω όρια αυτά που υπαγορεύονται από το εύρος του προϋπολογισμού. Στους επόμενους γύρους δεν χρειάζεται να ξαναπαράξουμε τα Pareto βέλτιστα χαρτοφυλάκια, απλώς διαλέγει ο αποφασίζων τα χαρτοφυλάκια που βρίσκονται σε συγκεκριμένη περιοχή του εύρους του προϋπολογισμού. Για την περιοχή αυτή υπολογίζουμε το πράσινο, κόκκινο και γκρι σύνολο και συνεχίζουμε τη διαδικασία έως ότου συγκλίνουμε σε ένα τελικό χαρτοφυλάκιο. Σε κάθε γύρο ο αποφασίζων βλέπει πληροφορίες για το πώς εξελίσσεται η συνολική επίδοση ως προς το κόστος, τον αριθμό των σχεδίων ανά χαρτοφυλάακιο κλπ. Όπως είναι φυσικό όσο στενεύει το εξεταζόμενο εύρος τόσο αυξάνονται τα έργα στο πράσινο και στο κόκκινο σύνολο και μειώνονται αυτά στο γκρι σύνολο.

ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ

Η μελέτη περίπτωσης αφορά 133 επενδυτικά σχέδια για ανανεώσιμες πηγές ενέργειας (ΑΠΕ) από τρείς τεχνολογίες (αιολικά, μικρά υδροηλεκτρικά και φωτοβολταϊκά) στις 13 περιφέρειες της Ελλάδας [13]. Η κατανομή ανά περιοχή και ανά τεχνολογία δίνεται στον Πίνακα 1.

Πίνακας 1. Γεωγραφική και τεχνολογική κατανομή των επενδυτικών σχεδίων

	W	SH	PV	**ΣΥΝΟΛΟ**
ΑΝΑΤ. ΜΑΚΕΔΟΝΙΑΣ-ΘΡΑΚΗΣ (EMD)	3		2	**5**
ΑΤΤΙΚΗΣ (ATT)		1		**1**
ΒΟΡΕΙΟΥ ΑΙΓΑΙΟΥ (NAG)			6	**6**
ΔΥΤΙΚΗΣ ΕΛΛΑΔΟΣ (WGR)			1	**1**
ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ (WMD)	3		6	**9**
ΗΠΕΙΡΟΥ (EPR)		3	8	**11**
ΘΕΣΣΑΛΙΑΣ (THE)	1	7	9	**17**
ΙΟΝΙΩΝ ΝΗΣΙΩΝ (ION)	1			**1**
ΚΕΝΤΡΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ (CMD)	3	5	6	**14**
ΚΡΗΤΗΣ (CRE)			4	**4**
ΝΟΤΙΟΥ ΑΙΓΑΙΟΥ (SAG)	1			**1**
ΠΕΛΟΠΟΝΝΗΣΟΥ (PEL)	8	1	3	**12**
ΣΤΕΡΕΑΣ ΕΛΛΑΔΟΣ (STE)	33	13	5	**51**
ΣΥΝΟΛΟ	**53**	**30**	**50**	**133**

W: αιολικά, SH: μικρά υδροηλεκτρικά, PV Φωτοβολταϊκά

Το συνολικό κόστος των 133 σχεδίων είναι 659 εκ. €. Ο διαθέσιμος προϋπολογισμός κυμαίνεται από 75 ως 125 εκ. €. Επίσης τίθενται οι εξής επιπλέον περιορισμοί πολιτικής:

- Το σύνολο των χρημάτων που θα διατεθούν σε έργα για την Στερεά Ελλάδα πρέπει να είναι μικρότερο από το 30% του συνολικού κόστους
- Το σύνολο των χρημάτων που θα διατεθούν σε έργα για την Πελλοπόνησο πρέπει να είναι μικρότερο από το 15% του συνολικού κόστους
- Το σύνολο των χρημάτων που θα διατεθούν σε έργα για την Αν. Μακεδονία-Θράκη, Βόρειο Αιγαίο, Δυτική Μακεδονία, Ήπειρο, Νότιο Αιγαίο πρέπει να είναι μεγαλύτερο από το 10% του συνολικού κόστους
- Ο αριθμός των έργων από κάθε τεχνολογία πρέπει να είναι ανάμεσα στο 20% και στο 60% του συνολικού αριθμού των επιλεχθέντων έργων
- Η συνολική ισχύς του τελικού χαρτοφυλακίο να είναι μεγαλύτερη από 170 MW.

Η αξιολόγηση του κάθε επενδυτικού σχεδίου γίνεται με 5 κριτήρια (1) την περιφερειακή ανάπτυξη (2) τις θέσεις εργασίας (3) την οικονομική αξιολόγηση (IRR) (4) τις εκπομπές CO2 που αποφεύγονται με βάση το ενεργειακό μίγμα της περιφέρειας (5) τη δέσμευση γης.

ΚΑΤΑΣΚΕΥΗ ΜΟΝΤΕΛΟΥ

Αξιολόγηση των σχεδίων με πολυκριτηριακή ανάλυση

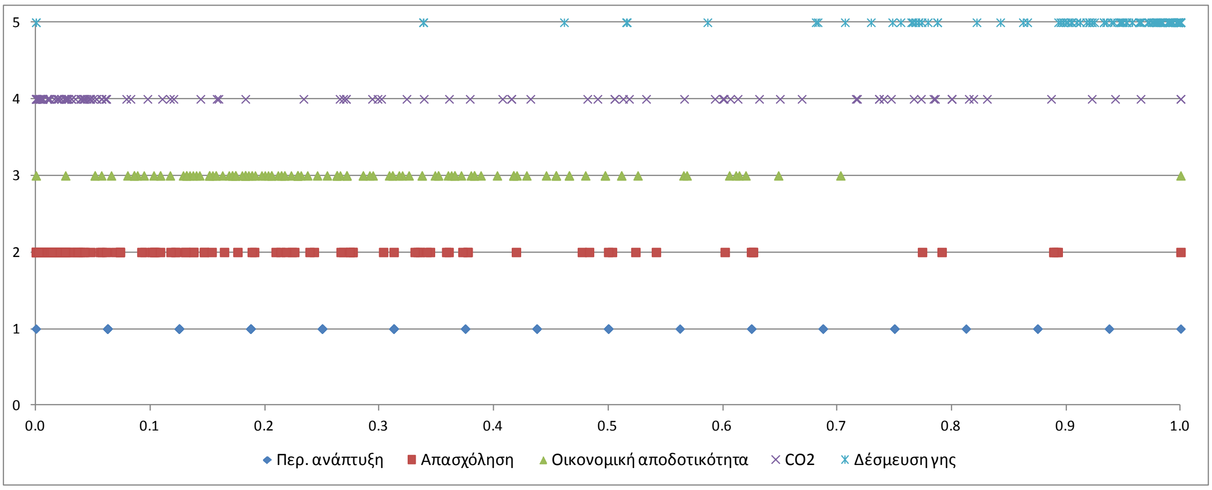
Για την πολυκριτηριακή ανάλυση χρησιμοποιείται η μέθοδος των συναρτήσεων χρησιμότητας. Για κάθε κριτήριο επιλέγεται η μορφή της συνάρτησης χρησιμότητας ανάλογα με τη συσσώρευση των τιμών έτσι ώστε να μεγιστοποιηθεί η διαχωριστική ικανότητα του κάθε κριτηρίου. Ο τρόπος που γίνεται αυτό είναι να κανονικοποιηθούν οι επιδόσεις των κριτηρίων στο διάστημα [0,1] με μια γραμμική σχέση ως εξής:

$$ls_{ij} = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}}$$

Όπου ls_{ij} η γραμμική επίδοση του i-σχεδίου στο j-κριτήριο, x_{ij} είναι η τιμή του i-σχεδίου στο j- κριτήριο και με *jmin* και *jmax* συμβολίζουμε το μέγιστο και το ελάχιστο κάθε κριτηρίου. Για κριτήρια προς ελαχιστοποίηση η κανονικοποίηση γίνεται με το (1-ls_{ij}) έτσι ώστε η καλύτερη επίδοση να είναι πάντα η μονάδα και η χειρότερη το μηδέν.
Στη συνέχεια εφαρμοζουμε κοίλες ή κυρτές συναρτήσεις χρησιμότητες με βάση τη σχέση:

$$s_{ij} = \frac{1 - e^{c_j \times ls_{ij}}}{1 - e^{c_j}}$$

Όπου s_{ij} είναι η επίδοση του i-σχεδίου στο j-κριτήριο και c_j η παράμετρος της συνάρτησης χρησιμότητας για το j-κριτήριο που παίρνει τιμές στο [-5, 5] και υποδεικνύει την κλίση της. Η επιλογή των c_j γίνεται έτσι ώστε να μεγιστοποιηθεί η διαχωριστική ικανότητα κάθε κριτηρίου. Στο Σχήμα 2 φαίνεται η διασπορά των τιμών σε κάθε κριτήριο και στον Πίνακα 2 το είδος, οι παράμετροι κλίσης της συνάρτησης χρησιμότητας και οι συντελεστές βαρύτητας κάθε κριτηρίου w_j (όπου θεωρούνται ισοβαρή τα κοινωνικά κριτήρια (1,2), το οικονομικό (3) και τα περιβαλλοντικά (4,5)



Σχήμα 2. Διασπορά τιμών ανά κριτήριο

Πίνακας 2. Χαρακτηριστικά κριτηρίων πολυκριτηριακής ανάλυσης επενδυτικών σχεδίων

		Συνάρτηση χρησιμότητας	Συντελεστής κλίσης c_j	Συντελεστής βαρύτητας w_j
1	Περιφερειακή ανάπτυξη	Γραμμική	0.001	0.167
2	Απασχόληση	Κοίλη	-5	0.166
3	Οικονομική αποδοτικότητα	Κοίλη	-3	0.334
4	Αποφυγή εκπομπών CO_2	Κοίλη	-1	0.167
5	Δέσμευση γης	Κυρτή	5	0.166

Η συνολική πολυκριτηριακή επίδοση για κάθε επενδυτικό σχέδιο i δίνεται από τη σχέση:

$$ms_i = \sum_{j=1}^{5} w_j s_{ij}$$

Κατασκευή μοντέλου Ακέραιου προγραμματισμού

Το μοντέλου του πολυκριτηριακού ακέραιου προγραμματισμού έχει ως εξής με βάση τους περιορισμούς της προηγούμενης παραγράφου:

$$\max Z_1 = \sum_{i=1}^{133} ms_i X_i$$

$$\min Z_2 = \sum_{i=1}^{133} \cos t_i X_i \qquad with\ Z_2 \in [lb, ub]$$

$$st$$

$$\sum\nolimits_{i \in STE} \cos t_i X_i \le 0.3 \times Z_2$$

$$\sum\nolimits_{i \in PEL} \cos t_i X_i \le 0.15 \times Z_2$$

$$\sum\nolimits_{i \in EMD, NAG, WMD, EPR, SAG} \cos t_i X_i \ge 0.1 \times Z_2$$

$$0.2 \times \sum_{i=1}^{133} X_i \le \sum_{i \in W} X_i \le 0.6 \times \sum_{i=1}^{133} X_i$$

$$0.2 \times \sum_{i=1}^{133} X_i \le \sum_{i \in SH} X_i \le 0.6 \times \sum_{i=1}^{133} X_i$$

$$0.2 \times \sum_{i=1}^{133} X_i \le \sum_{i \in PV} X_i \le 0.6 \times \sum_{i=1}^{133} X_i$$

$$\sum_{i=1}^{133} mw_i X_i \ge 170$$

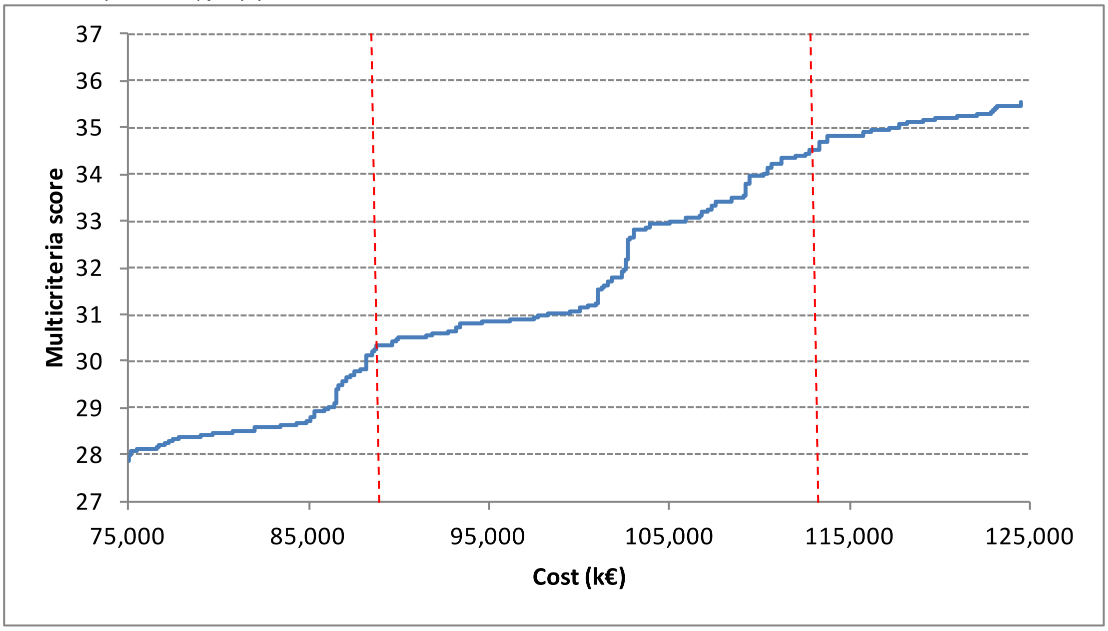
Όπου X_i είναι η δυαδική μεταβλητή απόφασης που παίρνει την τιμή X_i=1 αν το i-σχέδιο ενταχθεί στο χαρτοφυλάκιο αλλιώς X_i=0, ms_i είναι η πολυκριτηριακή επίδοση για το i-σχέδιο, $cost_i$ είναι το κόστος για το i-σχέδιο και mw_i είναι η ισχύς του i-σχεδίου σε MW.

ΑΠΟΤΕΛΕΣΜΑΤΑ

Η εφαρμογή της μεθόδου AUGMECON2 μας παράγει το σύνολο των κατά Pareto βέλτιστων χαρτοφυλακίων στο εύρος προϋπολογισμού 75–125 εκ. ευρώ. Η μεθοδος AUGMECON2 υλοποιήθηκε σε περιβάλλον GAMS και χρησιμοποιείται ο επιλύτης CPLEX 12.2 για την επίλυση. Ο χρόνος επίλυσης είναι 175 sec σε ένα Intel Core i5 στα 2.5 GHz. Το μέτωπο Pareto φαίνεται στο Σχήμα 3.

Το μέτωπο Pareto αποτελείται από 411 Pareto βέλτιστες λύσεις-χαρτοφυλάκια που κάθε ένα αποτελείται από διαφορετικό συνδυασμό έργων. Κάθε χαρτοφυλάκιο περιλαμβάνει από 60 έως 77 έργα. Παρατηρούμε από το Σχήμα 3 ότι η γραμμή δεν είναι ομαλή αλλά υπάρχουν αυξομειώσεις της κλίσης λόγω του ασυνεχούς (μη κυρτού) πεδίου ορισμού για τα X_i. Είναι προφανές ότι ο ορθολογικός αποφασίζων θα εστιάσει την προσοχή του στις περιοχές με μεγάλη κλίση όπου με μια μικρή αύξηση του κόστους επιτυγχάνεται μεγάλη αύξηση του δείκτη πολυκριτηριακής επίδοσης του χαρτοφυλακίου. Στην περιοχή αυτή προέκυψαν 35 πράσινα, 46 κόκκινα και 52 γκρί σχέδια.

Στη συνέχεια ο αποφασίζων επιλέγει μια περιοχή του συνολικού κόστους που έχει το μισό εύρος από το αρχικό. Η περιοχή που επιλέχθηκε φαίνεται στο Σχήμα 3 μεταξύ των διακεκομένων γραμμών..



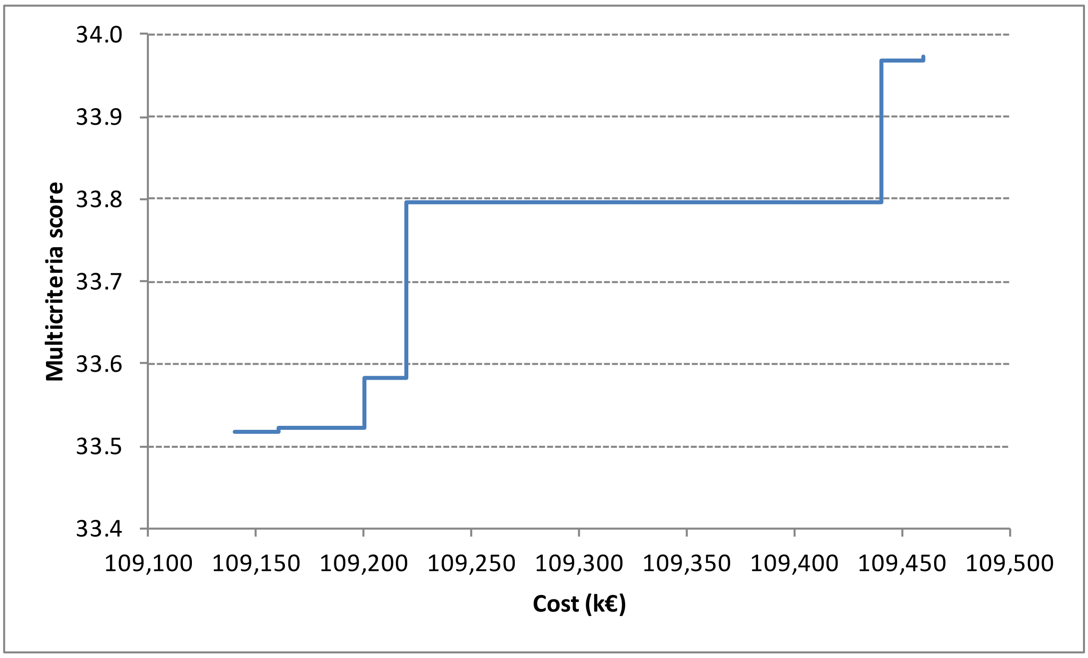
Σχήμα 3. Μέτωπο Pareto των χαρτοφυλακίων στο διάστημα 75-125 εκ. € και μείωση εύρους στην πρώτη επανάληψη

Για την περιοχή αυτή που περιλαμβάνει 185 Pareto βέλτιστα χαρτοφυλάκια καταγράφονται 41 πράσινα, 50 κόκκινα και 42 γκρί σχέδια. Η επαναληπτική διαδικασία συνεχίζεται για 8 γύρους μέχρι το τελικό χαρτοφυλάκιο και τα χαρακτηριστικά της φαίνονται στον Πίνακα 3.

Πίνακας 3. Χαρακτηριστικά της επαναληπτικής διαδικασίας επιλογής επενδυτικών σχεδίων

Γύρος	Εύρος προϋπολο-γισμού (M€)	Αριθμός Pareto χαρτοφυλακίων	Πράσινο σύνολο	Κόκκινο σύνολο	Γκρι σύνολο
1	75 - 125	411	35	46	52
2	88 - 113	185	41	50	42
3	100-113	78	45	50	38
4	106-112	28	58	52	23
5	107-110.5	15	64	56	13
6	108.5-110.5	8	64	56	13
7	109-109.8	6	67	57	9
8	109.2-109.4	1	73	60	0

Στον 7ο γύρο η εικόνα που βλέπει ο αποφασίζων σχετικά με το μέτωπο Pareto μετά από την σταδιακή εστίαση είναι αυτή που φαίνεται στο Σχήμα 4.
Το τελικά επιλεχθέν χαρτοφυλάκιο είναι αυτό με πολυκριτηριακή επίδοση 33.797 και κόστος 109.22 M€ το οποίο περιλαμβάνει 73 επενδυτικά σχέδια. Είναι αξιοσημείωτο ότι από το Σχήμα 4 βλέπουμε ότι το συγκεκριμένο χαρτοφυλάκιο προκύπτει από μια μεγάλη κάθετη μετατόπιση και η παρατεταμένη οριζόντια γραμμή δείχνει ότι είναι αρκετά εύρωστο αφού είναι βέλτιστο για μεγάλο σχετικά εύρος κόστους.



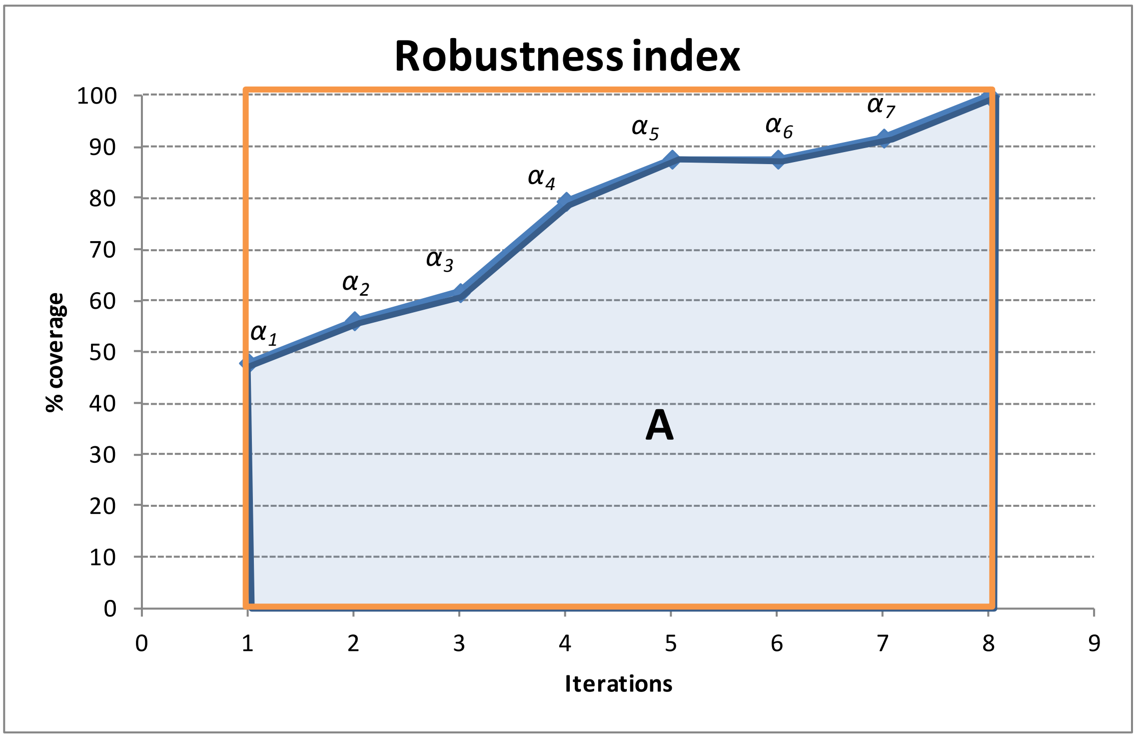
Σχήμα 4. Το μέτωπο Pareto στον 7° γύρο

Η συνολική εικόνα του χαρτοφυλακίου φαίνεται με χρωματικούς κωδικούς στο Σχήμα 5. Όσο πιο σκούρο είναι το χρώμα τόσο πιο σίγουροι είμαστε για την κατάσταση του συγκεκριμένου έργου. Τα πράσινα είναι τα έργα που τελικά εντάσσονται και τα κόκκινα αυτά που απορρίπτονται. Το κέρδος σε σχέση με τις συμβατικές προσεγγίσεις είναι ότι με την ΙΤΑ ο αποφασίζων δεν καταλήγει απλώς στο τελικό χαρτοφυλάκιο έργων αλλά βλέπει και την «ένταση» της έγκρισης ή «απόρριψης» του κάθε έργου.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133							

Σχήμα 5. Χρωματικός χάρτης των υπό εξέταση έργων

Με βάση το συγκεκριμένο εύρος προϋπολογισμού μπορούμε να υπολογίσουμε και τον βαθμό ευρωστίας του συγκεκριμένου χαρτοφυλακίου με βάση το πόσο νωρίς στην επαναληπτική διαδικασία εντάσσονται τα έργα. Αν για παράδειγμα από τον πρώτο γύρο είχαμε βρεί το 100% των έργων του τελικού χαρτοφυλακίου τότε το τελικό χαρτοφυλάκιο θα είχε την απόλυτη ευρωστία (θα ήμασταν απόλυτα σίγουροι για τη σύνθεσή του). Αντίθετα αν όλα τα έργα του χαρτοφυλακίου προέκυπταν στον τελευταίο γύρο τότε θα είχε την ελάχιστη ευρωστία. Ο δείκτης ευρωστίας που υπολογίζεται στη συνέχεια ποσοτικοποιεί την ευρωστία ανάλογα με το τι ποσοστό των έργων εντάσσονται σε κάποιον γύρο της διαδικασίας. Ο δείκτης ευρωστίας προκύπτει ως το εμβαδό κάτω από την αθροιστική καμπύλη του χαρτοφυλακίου προς το εμβαδό της περιοχής απόλυτης ευρωστίας (βλ. Σχήμα 6).



Σχήμα 6. Σχηματική παράσταση του δείκτη ευρωστίας

Η αθροιστική καμπύλη του χαρτοφυλακίου ορίζεται από τα σημεία $α_r$. Ως $α_r$ ορίζεται το ποσοστό των έργων του τελικού χαρτοφυλακίου που έχουν προκύψει ως τον γύρο r. Αν λοιπόν $α_1$, $α_2$, $...α_n$ είναι οι τεταγμένες των σημείων καμπής στο Σχήμα 7 και n ο αριθμός των γύρων τότε ο δείκτης ευρωστίας RI υπολογίζεται ως:

$$RI = [\frac{a_1 + a_2}{2} + \frac{a_2 + a_3}{2} + ... + \frac{a_{n-1} + a_n}{2}] / (n-1)$$

$$RI = [\frac{a_1}{2} + \sum_{i=2}^{n-1} a_i + \frac{a_n}{2}] / (n-1) = [\frac{a_1}{2} + \sum_{i=2}^{n-1} a_i + \frac{1}{2}] / (n-1)$$

Με τον τρόπο αυτό μπορούμε να συγκρίνουμε διάφορα χαρτοφυλάκια ως προς τον δείκτη ευρωστίας τους που ποσοτικοποιεί τον βαθμό ευρωστίας ως προς την αβεβαιότητα σε σχέση με το εύρος του διαθέσιμου προϋπολογισμού.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η επιλογή χαρτοφυλακίου επενδυτικών έργων μπορεί να αναχθεί σε ένα πρόβλημα με δύο φάσεις. Στην πρώτη φάση αξιολογούνται τα σχέδια με κάποια μέθοδο πολυκριτηριακής ανάλυσης και στη δεύτερη φάση εφαρμόζεται ένα μοντέλο μαθηματικού προγραμματισμού που ενσωματώνει τους όποιους περιορισμούς και περιλαμβάνει στην αντικειμενική συνάρτηση τα αποτελέσματα της πρώτης φάσης. Όταν υπάρχει ελαστικότητα στον συνολικό προϋπολογισμό και δίνεται με την μορφή εύρους τιμών μπορούμε να ακολουθήσουμε μια επαναληπτική διαδικασία υποστήριξης απόφασης που καταλήγει στην καλύτερη δυνατή εκμετάλλευση του προϋπολογισμού με βάση τις προτιμήσεις του αποφασίζοντα. Η Iterative Trichotomic Approach (ITA) αποτελεί ένα χρήσιμο εργαλείο υποστήριξης αποφάσεων σε παρόμοια προβλήματα βοηθώντας αλληλεπιδραστικά τον αποφασίζοντα και δίνοντας χρήσιμες πληροφορίες για την βεβαιότητα με την οποία εντάσσεται ή απορρίπτεται κάποιο σχέδιο.

ΕΥΧΑΡΙΣΤΙΕΣ

Μέρος του έργου υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος "Εκπαίδευση και Δια Βίου Μάθηση" και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους (Εθνικό Στρατηγικό Πλαίσιο Αναφοράς 2007-2013)" – Πρόγραμμα Θαλής "Robust MCDA" Η κ. Πετσάκ ευχαριστεί το ΙΚΥ για τη συνδρομή του στις διδακτορικές της σπουδές.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Lorie, J. H., Savage, L. J. Journal of Business, Vol. 28, No. 4, pp. 229-239, (1955)
[2] Zanakis, S.H., Mandakovic, T., Gupta, S.K., Sahay, S., and Hong, S., Socio Economic Planning Sciences, Vol. 29, No. 1, pp. 59-79, (1995)
[3] Abu-Taleb, M., Mareschal, B., European Journal of Operational Research, Vol. 81, No. 3, pp. 500-511, (1995)
[4] Belton, V., Stewart, T. 'Multiple Criteria Decision Analysis. An Integrated Approach', Kluwer Academic Publishers, UK, (2002).
[5] Mavrotas, G., Diakoulaki, D., Capros, P., Annals of Operations Research, Vol. 120, No. 1-4, pp. 159-170, (2003).
[6] Mavrotas, G., Diakoulaki, D., Caloghirou, Y., European Journal of Operational Research, Vol. 171, No. 1, pp. 296-308, (2003).
[7] Mavrotas, G., Diakoulaki, D., Kourentzis, A., European Journal of Operational Research, Vol. 187, No. 1, pp. 177-192, (2008).
[8] Liesio, J., Mild, P., Salo, A., European Journal of Operational Research, Vol. 181, No. 3, pp. 1488–1505, (2007)
[9] Liesio, J., Mild, P., Salo, A. European Journal of Operational Research, Vol. 190, No. 3, pp. 679–695, (2008).
[10] Shakhsi-Niaei, M., Torabi, S.A., Iranmanesh, S.H., Computers and Industrial Engineering, Vol. 61, No. 1, pp. 226-237, (2011).
[11] Mavrotas G. and Pechak O., International Journal of Multiple Criteria Decision Making 3(1), 79-96 (2013)
[12] Mavrotas G. and Florios K., "An improved version of the augmented ε-constraint method (AUGMECON2) for finding the exact Pareto set in Multi-Objective Integer Programming problems", Applied Mathematics & Computation 219, pp. 9652-9669, (2013).
[13] Μακρυβέλιος Ε. «Πολυκριτηριακή αξιολόγηση επενδύσεων σε έργα ανανεώσιμων πηγών ενέργειας» Διπλωματική εργασία στο ΔΠΜΣ Οργάνωση και Διοίκηση Βιομηχανικών Συστημάτων, Αθήνα 2011

Project Portfolio Selection in a Group Decision Making Environment: Aiming at Convergence with the Iterative Trichotomic Approach

G. Mavrotas*, O. Pechak, D. Siatras
School of Chemical Engineering,
National Technical University of Athens
Iroon Polytechniou, 9, 15780 Athens
**mavrotas@chemeng.ntua.gr*

E. Siskos, J. Psarras
School of Electrical and Computer Engineering,
National Technical University of Athens
Iroon Polytechniou, 9, 15780 Athens

Abstract

Project portfolio selection is the problem of selecting a subset of projects from a wider set, optimizing one or more criteria and satisfying specific constraints. The basic tools are usually Multiple Criteria Decision Analysis and mathematical programming. In the presence of multiple decision makers the preferences are not unique and there must be a negotiation approach taking into account all the points of view. In the present work we use the Iterative Trichotomic Approach (ITA) in order to seek convergence. With ITA we can draw conclusions for the acceptance of each individual project as well as for the robustness of the final portfolio. The weights of evaluation criteria differ among the decision makers so that each one of them finally selects a different "optimal" portfolio. ITA can classify the projects into three sets: the green projects (selected in the "optimal" portfolio by all the decision makers), the red projects (not selected in the "optimal" portfolio by any of the decision makers) and the grey projects which are selected by some (but not all) the decision makers. A converging Delphi like process is designed for the weights so that in the next round new weights are calculated for every decision maker. The mathematical model is updated according to the new weights and solved. As the iterative process moves from round to round the green and the red set are enriched and the grey projects are reduced. The iterative process terminates when the calculated weights for all the decision makers provide the same "optimal" portfolio. The above method is illustrated with an example involving 133 energy projects. The final outcome is the final portfolio as compromise among the decision makers as well as the degree of accordance on each one of the projects that are finally selected. Finally a consensus index for the final portfolio can be extracted according to the progress of the converging process.

KEYWORDS

Project portfolio selection, MCDA, Integer Programming, Group Decision Making

1. INTRODUCTION

Project portfolio selection is defined as the problem of selecting one or a subset from a set of projects (a subset of projects is considered as a "portfolio of projects"). In the latter case, the usual approach is to rank projects using one or more criteria and select the top ranked ones that cumulatively satisfy a budget limitation. However, in real world decision making there are two concepts that complicate the process like e.g. the existence of constraints imposed by the decision maker. The existence of constraints to be satisfied by the final selection destroys the independence of projects, which is one of the main assumptions in Multiple Criteria Decision Analysis (MCDA) ranking (see e.g. Belton and Stewart, 2002). In other words, the top ranked projects may only by chance satisfy the imposed constraints. For such cases Integer Programming (IP) is an appropriate tool that performs optimization under specific constraints. In case of project selection, the combinatorial character of the problem implies the use of IP with 0-1 (binary) variables expressing the incorporation (Xi=1) or not (Xi=0) of the i-th project in the portfolio. The earliest contributions were published under the title of capital budgeting (see e.g. Lorie and Savage, 1955), using strictly financial measures to measure the value of projects and portfolios, giving emphasis to the budget constraint. From early sixties, the so called capital budgeting problem was recognized as equivalent to the popular in Operational Research (OR) knapsack paradigm. The incorporation of multiple criteria can also been found in literature using Goal Programming (see e.g for a review Zanakis et al., 1995), combinations of

MCDA with IP (see e.g. Golabi et al., 1981; Abu Taleb and Mareschal, 1995; Mavrotas et al., 2003; Mavrotas et al., 2006; Mavrotas et al., 2008). In our case the problem is even more complicated as we have more than one decision makers (Group Decision Making). The preference of each one of the decision makers is expressed by assigning their own weights of importance to the criteria. The result is that each decision maker has his/her own optimal portfolio of projects. In order to achieve a consensus a convergence process is designed based on the Iterative Trichotomic Approach (ITA) as described in (Mavrotas and Pechak, 2013a and 2013b).

In the second section we describe the methodology and in the third section we present an application that illustrates the method. Finally, in section 4 the main conclusions are presented.

2. METHODOLOGY

2.1. The ITA Method

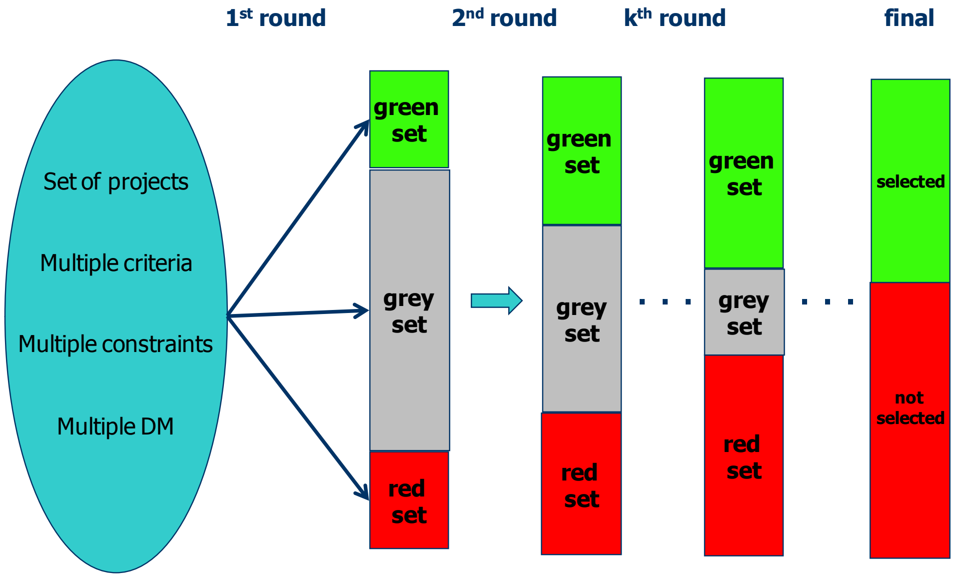
The basic idea of the ITA method is the separation of a set into three parts (trichotomy). In our case the set of projects is divided into three subsets (classes): green projects that are present in the final portfolio under all circumstances, red projects that are absent from the final portfolio under all circumstances, and grey projects that are present in part of the final portfolios. The classification in three subsets is not new in the literature. Liesio et al. (2007) used a similar approach in the framework of robust programming. However, the way the projects are assigned to each set is different. In addition, Mavrotas and Rozakis (2009) used similar concepts in a student selection problem for a post graduate program. The term "iterative" indicates that the proposed process is developed in a series of decision rounds (or cycles). A predetermined number of decision rounds may be defined from the beginning and every round feeds its subsequent until a convergence to the final portfolio is attained. From round to round the grey set is reduced as a result of convergence or uncertainty reduction. The first applications of ITA had to do with stochastic uncertainty in the parameters of the model and a series of Monte Carlo simulation-optimization steps (Mavrotas and Pechak, 2013a, 2013b). In the present case we adjust it to the group decision making context.

2.2. Adaptation of ITA to the Group Decision Making Context

In the present case the set of projects is divided into three subsets (classes): green projects that are present in the final portfolio according to all the decision makers, red projects that are absent from the final portfolio according to all the decision makers, and grey projects that are present in the final portfolios according to some decision makers.

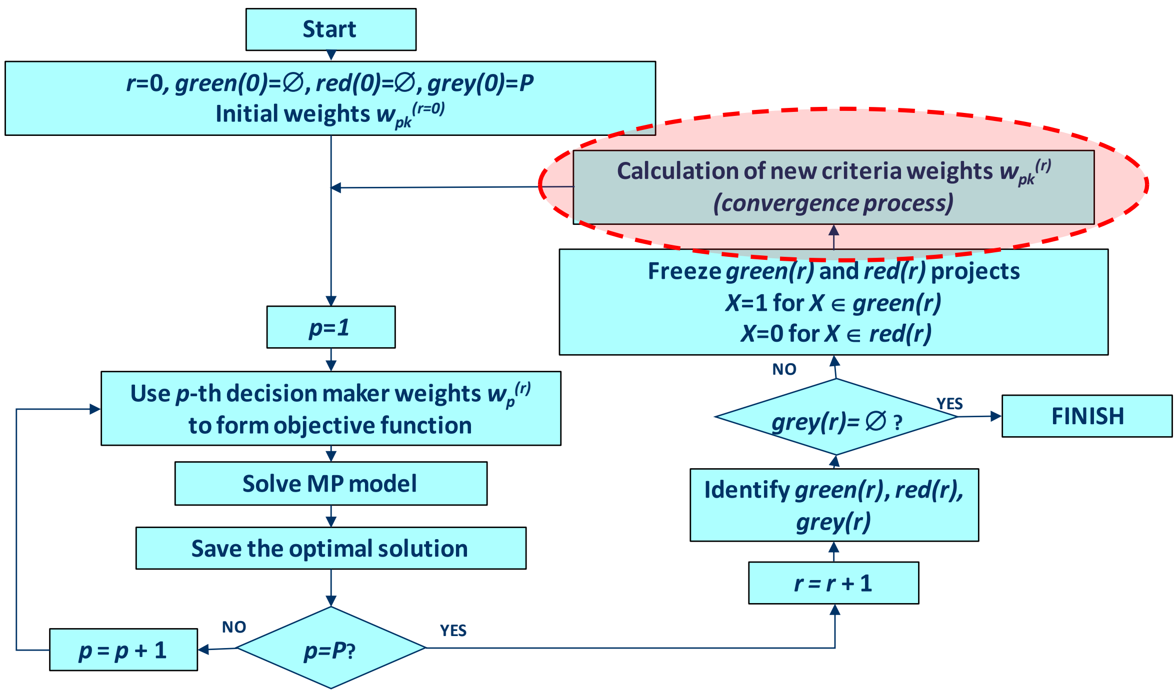
ISBN: 978-618-80361-1-6

Figure 1. Illustration of the method



The flowchart of the method is shown in Figure 2.

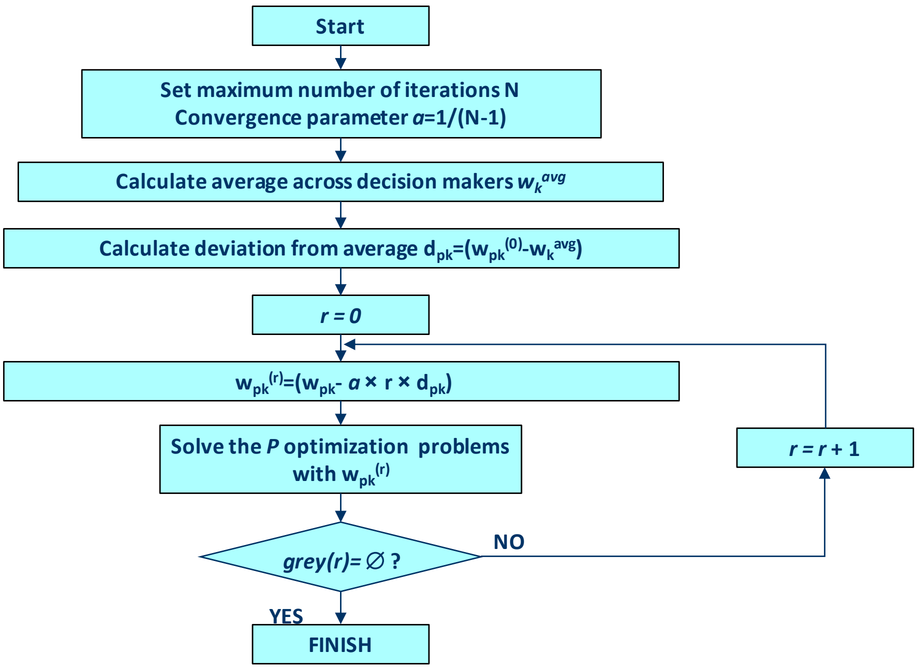
Figure 2. The flowchart of the method Group ITA



2nd International Symposium and 24th National Conference on Operational Research
ISBN: 978-618-80361-1-6

The convergence process is designed to converge the weights of the criteria and is depicted in Figure 3. We adjust the weights of importance of the decision makers from round to round in order to converge as we move in the iterative process. First, we select the maximum number of rounds that we are going to perform (N) and we accordingly assign the convergence parameter α as α=1/(N-1). Then we calculate the deviation of each one of the weights from their average across the decision makers.

Figure 3. The convergence process of the criteria weights



Where p =index for Decision Makers (DM) p=1..P, k=index for criteria k=1..K, r=index for rounds in the iterative process, $w_{pk}^{(r)}$ =weight of p-th DM for k-th criterion in r-th round. It must be noted that from round to round the weights keep their property to sum up to 1 as it is shown from the following equations.
Given that for the original weights w_{pk} we have:

$$\sum_{k=1}^{K} w_{pk} = 1 \quad p = 1..P, \qquad and \qquad w_k^{avg} = \frac{\sum_{p=1}^{P} w_{pk}}{K}$$

Then, we have for the new weights w'_{pk}:

$$\sum_{k=1}^{K} w'_{pk} = \sum_{k=1}^{K} (w_{pk} - a(w_{pk} - w_k^{avg})) = \sum_{k=1}^{K} (1-a) w_{pk} + a \cdot \sum_{k=1}^{K} w_k^{avg} =$$

$$(1-a)\sum_{k=1}^{K} w_{pk} + a \cdot \sum_{k=1}^{K} w_k^{avg} = (1-a) + a = 1$$

3. APPLICATION

We applied the method in a group decision making problem dealing with energy projects. There are 133 energy projects from three RES technologies (wind, small hydro, photovoltaic) distributed across the 13

ISBN: 978-618-80361-1-6

regions of Greece. There are 5 criteria according to which the projects are evaluated, namely, (1) Regional development (2) CO2 emission reduction (3) Economic efficiency (4) Employment (5) Land use. There are 12 decision makers that give the weights of importance shown in Table 1.

Table 1. Weights of importance for the 12 decision makers

			criteria		
DM	1	2	3	4	5
1	0.14	0.13	0.46	0.13	0.14
2	0.25	0.37	0.15	0.08	0.15
3	0.41	0.21	0.03	0.14	0.21
4	0.07	0.41	0.35	0.16	0.01
5	0.02	0.02	0.50	0.33	0.13
6	0.20	0.20	0.20	0.20	0.20
7	0.15	0.25	0.40	0.02	0.18
8	0.08	0.28	0.35	0.17	0.12
9	0.22	0.25	0.28	0.17	0.08
10	0.15	0.35	0.25	0.20	0.05
11	0.21	0.30	0.15	0.15	0.19
12	0.20	0.20	0.30	0.25	0.05
average	0.1750	0.2475	0.2850	0.1667	0.1258

There are also the following constraints that must be fulfilled. The total cost of the 133 projects is 659 M€ and the available budget is 150 M€. The cost of projects in Central Greece should be less than 30% of the total cost, the cost of projects in Pelopponese should be less than 15% of the total cost, the cost of projects in East & West Macedonia, Northern & Southern Aegean, Epirus should be greater than 10% of the total cost. In addition, the number of projects from each technology should be between 20% and 60% of the selected projects and the total capacity of the selected projects should be greater than 300 MW. We apply the group ITA method with convergence parameter α=0.1 an the results are depicted in the following Figure 4. The great advantage of the group –ITA method is that we see information about the consensus of each one of the projects that participate or not in the final portfolio as well as the consensus on the final portfolio.

4. CONCLUSIONS

The project portfolio selection problem can be effectively addressed with a combination of MCDA and IP. In the presence of multiple decicion makers ITA can provide a sound framework for an iterative convergence process that can be used in Delphi like approaches. The main advantage is that we can measure the consensus per project as well as for the final portfolio. For future research we will examine the incorporation of additional decision parameters that express the decision maker preferences in the group-ITA e.g. the shape of utility function or the policy constraints. In this case, the convergence process must be adjusted also to these parameters that may vary from decision maker to decision maker.

ISBN: 978-618-80361-1-6

Figure 4: The results from group – ITA method

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133							

iter 1:	73
iter 5:	1
iter 7:	3
iter 9:	1
iter 11:	5

iter 1:	40
iter 5:	3
iter 7:	1
iter 9:	1
iter 11:	5

We see information about the intensity (consensus) for every project that is selected or rejected from the final portfolio

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning"

REFERENCES

Abu-Taleb M, Mareschal B (1995) *Water resources planning in the Middle East: application of the PROMETHEE V multicriterion method*. Eur J Oper Res 81:500-511

Belton V, Stewart T (2002) *Multiple Criteria Decision Analysis. An Integrated Approach*. Kluwer Academic Publishers, UK

Golabi K, Kirkwood CW, Sicherman A (1981) *Selecting a portfolio of Solar Energy Projects Using Multiattribute Preference Theory*. Manage Sci 27:174-189

Mavrotas G, Diakoulaki D, Capros P (2003) *Combined MCDA – IP Approach for Project Selection in the Electricity Market.* Ann Oper Res 120:159-170

Mavrotas G, Diakoulaki D, Caloghirou Y (2006) *Project prioritization under policy restrictions. A combination of MCDA with 0–1 programming.* Eur J Oper Res 171:296-308

Mavrotas G, Diakoulaki D, Kourentzis A (2008) *Selection among ranked projects under segmentation, policy and logical constraints.* Eur J Oper Res 187:177-192

Mavrotas G, Rozakis S (2009) *Extensions of the PROMETHEE method to deal with segmenta-tions constraints*. J Decis Syst 18:203-229

Mavrotas, G., Pechak, O. (2013a) *The trichotomic approach for dealing with uncertainty in project portfolio selection: Combining MCDA, mathematical programming and Monte Carlo simulation*. International Journal of Multiple Criteria Decision Making 3(1) 79-97

ISBN: 978-618-80361-1-6

Mavrotas, G., Pechak, O. (2013b) "*Combining Mathematical Programming and Monte Carlo simulation to deal with uncertainty in energy project portfolio selection*" Chapter 16 in F. Cavallaro (ed) Assessment and Simulation Tools for Sustainable Energy Systems Springer-Verlag, London

Liesio J, Mild P, Salo A (2008) *Robust portfolio modeling with incomplete cost information and project interdependencies*. Eur J Oper Res 190(3):679–695

Lorie J H, Savage LJ (1955) *Three problems in rationing capital*. J Business 28(4): 229-239

Zanakis SH, Mandakovic T, Gupta SK, Sahay S, Hong S (1995) *A Review of Program Evaluation and Fund Allocation Methods Within the Service and Government Sectors.* Socio-Econ Plan Sci 29:59-79

ISBN: 978-618-80361-1-6

A Robust Extension of the MUSA Method Based on Desired Properties of the Collective Preference System

E. Grigoroudis
School of Production Engineering and Management
Technical University of Crete
University Campus, 73100 Chania, Greece

Y. Politis
School of Science and Technology
Hellenic Open University
Parodos Aristotelous 18,
GR26 335, Patra, Greece

Abstract

The MUSA method is a collective preference disaggregation approach following the main principles of ordinal regression analysis under constraints using linear programming techniques. The method has been developed in order to measure and analyze customer satisfaction and it is used for the assessment of a set of marginal satisfaction functions in such a way that the global satisfaction criterion becomes as consistent as possible with customer's judgments. Thus, the main objective of the method is to assess collective global and marginal value functions by aggregating individual judgments. This study presents an extension of the MUSA method based on desired properties of the inferred preference system. In particular, the linear programming formulation of the method gives the ability to consider additional constraints regarding special properties of the assessed model variables. One of the most interesting extensions concerns additional properties for the assessed average indices. These indices refer to the average satisfaction indices, which are the mean value of the global and marginal value functions and can be considered as the basic performance norms and the average demanding indices, which indicate customers' demanding level and represent the average deviation of the estimated value functions from a "normal" (linear) function. The main aim of the study is to show how incorporating these additional constraints in the linear program of the original MUSA method, the robustness of the estimated results may be improved. In addition, the study presents potential problems in the aforementioned approach, especially in case of inconsistencies between global and partial judgments, and proposes alternative modeling techniques based on goal programming that may be used in the post-optimality analysis step of the method.

KEYWORDS

MUSA method, robustness analysis, ordinal regression, satisfaction analysis

1. INTRODUCTION

The MUSA (MUlticriteria Satisfaction Analysis) method is a preference disagregation approach following the main principles of ordinal regression analysis. It measures and analyzes customer satisfaction assuming that customer's global satisfaction is based on a set of criteria representing service characteristic dimensions. The main object of the MUSA method is the aggregation of individual judgments into a collective value function.

Different extensions of the method for the improvement of the provided results include additional DMs' preferences or desired properties of the inferred preference system. These extensions concern properties for the estimated value functions, hierarchy or interaction of criteria, alternative objective functions (during the post-optimality analysis), different types of input data (ordinal/cardinal), etc.

This study presents an extension of the MUSA method based on desired properties of the inferred preference system and the examination of whether the introduction of additional constraints in the linear program of the original MUSA method may improve the robustness of the estimated results.

2. THE MUSA METHOD

2.1. Mathematical Development

The MUSA method assesses global and partial satisfaction functions Y^* and X_i^* respectively, given customers' judgments Y and X_i. The method follows the principles of ordinal regression analysis under constraints using linear programming techniques (Jacquet-Lagrèze and Siskos, 1982; Siskos and Yannacopoulos, 1985; Siskos, 1985). The ordinal regression analysis equation has the following form:

$$\begin{cases} Y^* = \sum_{i=1}^{n} b_i X_i^* \\ \sum_{i=1}^{n} b_i = 1 \end{cases} \quad (1)$$

where b_i is the weight of the i-th criterion and the value functions Y^* and X_i^* are normalised in the interval [0, 100]. The main objective of the method is to achieve the maximum consistency between the value function Y^* and the customers' judgments Y. Based on the above modelling and by introducing two error variables σ^+ and σ^-, the ordinal regression equation becomes as follows:

$$\tilde{Y}^* = \sum_{i=1}^{n} b_i X_i^* - \sigma^+ + \sigma^- \quad (2)$$

where $\tilde{Y}^*$ is the estimation of global value function Y^* and σ^+ and σ^- are the overestimation and underestimation error, respectively.

The final form of the linear programming problem is as follows:

$$[\min]F = \sum_{j=1}^{M} \sigma_j^+ + \sigma_j^-$$

under the constraints

$$\begin{aligned} &\sum_{i=1}^{n} \sum_{k=1}^{t_{ji}-1} w_{ik} - \sum_{m=1}^{t_j-1} z_m - \sigma_j^+ + \sigma_j^- = 0, \text{ for } j = 1,2,\ldots,M \\ &\sum_{m=1}^{a-1} z_m = 100 \\ &\sum_{i=1}^{n} \sum_{k=1}^{a_i-1} w_{ik} = 100 \\ &z_m \geq 0,\ w_{ik} \geq 0,\ \forall m,i,k \\ &\sigma_j^+ \geq 0,\ \sigma_j^- \geq 0, \text{ for } j = 1,2,\ldots,M \end{aligned} \quad (3)$$

where t_j and t_{ji} are the judgments of the j-th customer globally and partially for each criterion i=1,2,…,n, M is the number of customers and z_m, w_{ik} are a set of transformation variables such us:

$$\begin{cases} z_m = y^{*m+1} - y^{*m} & \text{for } m = 1,2,\ldots,\alpha - 1 \\ w_{ik} = b_i x_i^{*k+1} - b_i x_i^{*k} & \text{for } k = 1,2,\ldots,\alpha_i - 1 \text{ and } i = 1,2,\ldots,n \end{cases} \quad (4)$$

where α and α_i is the evaluation ordinal scale for the global assessment and for the assessment of the i-th criterion, respectively.

ISBN: 978-618-80361-1-6

Furthermore, taking into account the hypothesis of strict preferential order of the scales of some or all the dimensions/criteria, the following conditions are met:

$$\begin{cases} y^{*m} < y^{*m+1} \Leftrightarrow y^m \prec y^{m+1} & \text{for } m = 1,2,\ldots,\alpha \\ x_i^{*k} < x_i^{*k+1} \Leftrightarrow x_i^k \prec x_i^{k+1} & \text{for } k = 1,2,\ldots,\alpha_i - 1 \text{ and } i = 1,2,\ldots,n \end{cases} \tag{5}$$

where $\prec$ means “strictly less preferred”. Based on (5) the following conditions occur:

$$\begin{cases} y^{*m+1} - y^{*m} \geq \gamma \Leftrightarrow z_m \geq \gamma \Leftrightarrow z_m' \geq 0 & \text{for } m = 1,2,\ldots,\alpha \\ x_i^{*k+1} - x_i^{*k} \geq \gamma_i \Leftrightarrow \mathrm{w}_{ik} \geq \gamma_i \Leftrightarrow \mathrm{w}_{ik}' \geq 0 & \text{for } k = 1,2,\ldots,\alpha_i - 1 \text{ and } i = 1,2,\ldots,n \end{cases} \tag{6}$$

where γ and γ_i are the preference thresholds (minimum step of increase) for the value functions Y^* and X_i^*, respectively, with γ, $\gamma_i > 0$, and it is set that:

$$\begin{cases} z_m = z_m' + \gamma & \text{for } m = 1,2,\ldots,\alpha \\ w_{ik} = w_{ik}' + \gamma_i & \text{for } k = 1,2,\ldots,\alpha_i - 1 \text{ and } i = 1,2,\ldots,n \end{cases} \tag{7}$$

The proposed extension affects the initial LP and constitutes the *generalized* form of the MUSA method (Grigoroudis and Siskos, 2002; Grigoroudis and Siskos, 2010).

2.2. Post Optimality Analysis

The previous (initial) LP is considered as producing an optimal solution with many degrees of freedom. Therefore, a stability analysis, as a post-optimality analysis, follows. During the post-optimality analysis stage of the MUSA method, *n* linear programs (equal to the number of criteria) are formulated and solved. Each linear program maximizes the weight of a criterion and has the following form:

$$\begin{cases} [\max] F' = \sum_{k=1}^{\alpha_i - 1} w_{ik} \quad \text{for } i = 1,2,\ldots,n \\ \text{subject to} \\ F \leq F^* + \varepsilon \\ \text{all the constraints of LP (3)} \end{cases} \tag{8}$$

where F^* is the optimal value of the objective function of LP (3) and ε is a small percentage of F^*. The average of the optimal solutions given by the n LPs (8) may be considered as the final solution of the problem. In case of instability, a large variation of the provided solutions appears and the final average solution is less representative.

The stability of the results provided by the post-optimality analysis is calculated with the Average Stability Index *(ASI)*. *ASI* is the mean value of the normalized standard deviation of the estimated weights b_i and is calculated as follows:

$$ASI = 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{n \sum_{j=1}^{n} \left(b_i^j\right)^2 - \left(\sum_{j=1}^{n} b_i^j\right)^2}}{100\sqrt{n-1}} \tag{9}$$

ISBN: 978-618-80361-1-6

where b_i^j is the estimated weight of the criterion *i*, in the *j*-th post-optimality analysis LP (Grigoroudis and Siskos, 2002; Grigoroudis and Siskos, 2010).

Furthermore, the fitting level of the MUSA method refers to the assessment of a preference collective value system (value functions, weights, etc.) for the set of customers with the minimum possible errors. For this reason, the optimal values of the error variables indicate the reliability of the value system that is evaluated.

The Average Fitting Index (*AFI*) depends on the optimum error level and the number of customers:

$$AFI = 1 - \frac{F^*}{100 \cdot M} \tag{10}$$

where F^* is the minimum sum of errors of the initial LP, and M is the number of customers.

The *AFI* is normalised in the interval [0, 1], and it is equal to 1 if $F^* = 0$, i.e. when the method is able of evaluating a preference value system with zero errors. Similarly, the *AFI* takes the value 0 only when the pairs of the error variables σ_j^+ and σ_j^- take the maximum possible values.

An alternative fitting indicator is based on the percentage of customers with zero error variables, i.e. the percentage of customers for whom the estimated preference value systems fits perfectly with their expressed satisfaction judgments. This average fitting index AFI_2 is assessed as follows:

$$AFI_2 = \frac{M_0}{M} \tag{11}$$

where M_0 is the number of customers for whom $\sigma^+ = \sigma^- = 0$.

A final average fitting index AFI_3 takes into account the maximum values of the error variables for every global satisfaction level, as well as the number of customers that belongs to this level:

$$AFI_3 = 1 - \frac{F^*}{M \sum_{m=1}^{\alpha} p^m \max\left\{ y^{*m}, 100 - y^{*m} \right\}} \tag{12}$$

where p^m is the frequency of customers belonging to the y^m satisfaction level.

2.3. Results

The most important results of the method are the estimated value functions. Another interesting result concerns the average global and partial satisfaction indices, S and S_i, which can be assessed according to the following equations:

$$\begin{cases} S = \frac{1}{100} \sum_{m=1}^{\alpha} p^m y^{*m} \\ S_i = \frac{1}{100} \sum_{k=1}^{\alpha_i} p_i^k x_i^{*k} \quad \text{for } i = 1, 2, \ldots, n \end{cases} \tag{13}$$

ISBN: 978-618-80361-1-6

where p^m and p_i^k are the frequencies of customers belonging to the y^m and x_i^k satisfaction levels respectively. It should be noted that the average satisfaction indices are basically the mean value of the global or partial value functions and they are normalised in the interval [0, 100%].

The average global and partial demanding indices, D and D_i, represent the average deviation of the estimated value curves from a "normal" (linear) function and reveal the demanding level of customers. They are normalised in the interval [–1, 1] and, are assessed as follows:

$$\begin{cases} D = \dfrac{\sum_{m=1}^{\alpha-1}\left(\dfrac{100(m-1)}{\alpha-1} - y^{*m}\right)}{100\sum_{m=1}^{\alpha-1}\dfrac{m-1}{\alpha-1}} & \text{for } \alpha > 2 \\ D_i = \dfrac{\sum_{k=1}^{\alpha-1}\left(\dfrac{100(k-1)}{\alpha_i-1} - x_i^{*k}\right)}{100\sum_{k=1}^{\alpha_i-1}\dfrac{k-1}{\alpha_i-1}} & \text{for } \alpha_i > 2 \text{ and } i = 1,2,\ldots,n \end{cases} \tag{14}$$

3. MODELING ADDITIONAL PROPERTIES

3.1. Model Development

The LP formulation of the MUSA method gives the ability to consider additional constraints regarding special properties of the assessed model variables. One of the most interesting extensions concerns additional properties for the assessed average indices.

For example, a linkage between global and partial average satisfaction indices may be assumed, since these indices are considered as the main performance indicators of the business organization. In particular, the global average satisfaction index *S* is assessed as a weighted sum of the partial satisfaction indices S_i:

$$S = \sum_{i=1}^{n} b_i S_i \Leftrightarrow \sum_{m=1}^{\alpha} p^m y^{*m} = \sum_{i=1}^{n} b_i \sum_{k=1}^{\alpha_i} p_i^k x_i^{*k} \tag{15}$$

Taking into account the transformation variables z_m, w_{ik} and using formula (13), the above equation becomes as follows:

$$\sum_{m=2}^{\alpha} p^m \sum_{t=1}^{m-1} z_t = \sum_{i=1}^{n}\sum_{k=2}^{\alpha_i} p_i^k \sum_{t=1}^{k-1} w_{it} \tag{16}$$

In the case of the generalized MUSA method, the preference thresholds γ and γ_i should be introduced, and equation (16) is written:

$$\sum_{i=1}^{n}\sum_{k=2}^{\alpha_i} p_i^k \sum_{t=1}^{k-1} w_{it} - \sum_{m=2}^{\alpha} p^m \sum_{t=1}^{m-1} z_t = \sum_{m=2}^{\alpha} p^m \gamma(m-1) - \sum_{i=1}^{n}\sum_{k=2}^{\alpha_i} p_i^k \gamma_i (k-1) \tag{17}$$

Similarly, a weighted sum formula may be assumed for the average demanding indices:

$$D = \sum_{i=1}^{n} b_i D_i \tag{18}$$

ISBN: 978-618-80361-1-6

Taking into account the transformation variables z_m, w_{ik} and using formula (14), the previous equation can be written in terms of the MUSA variables:

$$\frac{\sum_{m=1}^{\alpha-1} 100(m-1) - (\alpha-1)\sum_{t=1}^{m-1} z_t}{\alpha(\alpha-1)} = \sum_{i=1}^{n} \frac{\sum_{k=1}^{\alpha_i-1}(k-1)\sum_{t=1}^{\alpha_i-1} w_{it} - (\alpha_i-1)\sum_{t=1}^{k-1} w_{it}}{\alpha_i(\alpha_i-1)} \qquad (19)$$

In the case of the generalized MUSA method, equation (19) should be modified by introducing the variables z'_m and w'_{ik} (see formula (7)).

The equations (17) and (19) may be introduced as additional constraints in the LP (3). However, these additional properties of average indices should be used carefully, since their form does not guarantee a feasible solution of the LP, especially in case of inconsistencies between global and partial satisfaction judgments. For this reason, the aforementioned equations may be written using a goal programming formulation and used alternatively as post-optimality criteria.

3.2. Numerical Example

A hypothetical numerical example has been used in order to examine whether the introduction of additional constraints in the linear program of the original MUSA method may improve the robustness of the estimated results. In this example 20 customers express their satisfaction globally and for 3 different criteria using a three-level qualitative scale (see Table 1).

Table 1 Performance Judgments

Table 1: Performance Judgments

Overall	Criterion 1	Criterion 2	Criterion 3
Satisfied	Satisfied	Very Satisfied	Very satisfied
Satisfied	Dissatisfied	Satisfied	Very satisfied
Satisfied	Dissatisfied	Very satisfied	Satisfied
Very satisfied	Very satisfied	Very satisfied	Very satisfied
Very satisfied	Very satisfied	Very satisfied	Very satisfied
Satisfied	Very satisfied	Satisfied	Satisfied
Satisfied	Satisfied	Very satisfied	Very satisfied
Satisfied	Satisfied	Very satisfied	Satisfied
Satisfied	Satisfied	Satisfied	Satisfied
Dissatisfied	Very satisfied	Satisfied	Dissatisfied
Very satisfied	Very satisfied	Very satisfied	Satisfied
Satisfied	Satisfied	Very satisfied	Dissatisfied
Satisfied	Very satisfied	Satisfied	Dissatisfied
Very satisfied	Very satisfied	Satisfied	Very satisfied
Dissatisfied	Dissatisfied	Dissatisfied	Dissatisfied
Satisfied	Satisfied	Satisfied	Satisfied
Very satisfied	Satisfied	Very satisfied	Very satisfied
Satisfied	Very satisfied	Dissatisfied	Satisfied
Satisfied	Very satisfied	Dissatisfied	Very satisfied
Very satisfied	Very satisfied	Very satisfied	Satisfied

Table 2 presents the results of the original MUSA method and of the extension of the method with the successive introduction of additional constraints for the average satisfaction and demanding indices. According to Table 2, there is a remarkable increase of the *ASI and* AFI_2 indices when the additional constraints are introduced in the original MUSA method. The slight decrease (<1%) that appears for the other two fitting indices (AFI_1.and AFI_3) is not sufficient to reduce the considerably improved results.

ISBN: 978-618-80361-1-6

Table 2 Results - comparison

Table 2: Results - comparison			
	Original MUSA method	**MUSA method + Constraint for average satisfaction indices**	**MUSA method + Constraint for average satisfaction / demanding indices**
AFI_1	**92.00%**	**91.67% (-0.36%)**	**91.67% (-0.36%)**
AFI_2	**60.00%**	**75.00% (+25.00%)**	**75.00% (+25.00%)**
AFI_3	**89.47%**	**89.04% (-0.49%)**	**89.04% (-0.49%)**
ASI	**71.95%**	**80.07% (+11.29%)**	**80.56% (+11.98%)**

4. CONCLUDING REMARKS

The introduction of additional constraints in the linear program of the original MUSA method seems to increase its stability. The MUSA method is a rather flexible approach and thus several extensions may be developed taking into account additional information or data. Additional information that can be introduced in the original MUSA method and improve its robustness may include preferences for the importance of the criteria or other model properties. A simulation study can also be performed in order to study the impact of model parameters and whether appropriate combinations of these parameters can improve the stability of the provided results. Additional measures of robustness may facilitate the investigation of various extensions of the MUSA model.

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

REFERENCES

Grigoroudis, E. and Y. Siskos (2002). *Preference disaggregation for measuring and analysing customer satisfaction: The MUSA method,* European Journal of Operational Research, 143 (1), 148-170.

Grigoroudis, E. and Y. Siskos (2010). *Customer Satisfaction Evaluation: Methods for Measuring and Implementing Service Quality*, Springer, New York.

Jacquet-Lagrèze, E. and J. Siskos (1982). *Assessing a set of additive utility functions for multicriteria decision-making: The UTA method,* European Journal of Operational Research, 10 (2), 151-164.

Siskos, J. (1985). *Analyses de régression et programmation linéaire*, Révue de Statistique Appliquée, 23 (2), 41-55.

Siskos, Y. and D. Yannacopoulos (1985). *UTASTAR: An ordinal regression method for building additive value functions*, Investigaçao Operacional, 5 (1), 39-53.

Combining Performance and Importance Judgment in the MUSA Method

Y. Politis
School of Science and Technology
Hellenic Open University
Parodos Aristotelous 18,
GR26 335, Patra, Greece

E. Grigoroudis
School of Production Engineering and Management
Technical University of Crete
University Campus, 73100 Chania, Greece

Abstract

The multicriteria method MUSA (MUlticriteria Satisfaction Analysis) is a preference disaggregation model following the principles of ordinal regression analysis (inference procedure). The method is used for measuring and analyzing customer satisfaction and aims at evaluating the satisfaction level of a set of individuals (customers, employees, etc.) based on their values and expressed preferences. This study presents an extension of the MUSA method based on additional customer preferences. In particular, a customer satisfaction survey may include, besides the usual performance questions, preferences about the importance of the criteria. Using such questions, customers are asked either to judge the importance of a satisfaction criterion using a predefined ordinal scale, or rank the set of satisfaction criteria according to their importance. All these performance and importance preferences are modeled using linear programming techniques in order to assess a set of marginal satisfaction functions in such a way that the global satisfaction criterion and the importance preferences become as consistent as possible with customer's judgments. Based on these optimality criteria, the extension of the MUSA method is modeled as a Multiobjective Linear Programming (MOLP) problem. The main aim of the study is to show how combining customers' performance and importance preferences, the robustness of the estimated results may be improved compared to the original MUSA method. An illustrative example is presented in order to show the applicability of this approach, while several MOLP techniques (e.g., heuristic method, compromise programming, global criterion approach) and post-optimality approaches are applied.

KEYWORDS

MUSA extension, robustness improvement

1. INTRODUCTION

The MUSA (MUlticriteria Satisfaction Analysis) method is a preference dissagregation model for measuring and analyzing customer satisfaction. It follows the principles of ordinal regression analysis and aims at evaluating the satisfaction level of a set of individuals (customers, employees, etc) based on their values and expressed preferences. Considering that the MUSA method is based on a linear programming modeling the problem of multiple or near optimal solutions appears in several cases. This has an impact on the stability level of the provided results. Additional customer preferences such as preferences about the importance of the criteria may improve the stability of the basic MUSA model.

2. THE MUSA METHOD

The main objective of the MUSA method is the aggregation of individual judgments into a collective value function assuming that client's global satisfaction depends on a set of n criteria or variables representing service characteristic dimensions. According to the model, each customer is asked to express his/her preferences, namely his/her global satisfaction and his/her satisfaction with regard to the set of discrete criteria. MUSA assesses global and partial satisfaction functions Y^* and X_i^* respectively, given customers' judgments Y and X_i. The method follows the principles of ordinal regression analysis under constraints using linear programming techniques (Jacquet-Lagrèze and Siskos, 1982; Siskos and Yannacopoulos, 1985; Siskos, 1985). The ordinal regression analysis equation with the introduction of a double-error variable representing the overestimation and underestimation error has the following form:

$$\begin{cases} Y^* = \sum_{i=1}^{n} b_i X_i^* - \sigma^+ + \sigma^- \\ \sum_{i=1}^{n} b_i = 1 \end{cases} \tag{1}$$

where the value functions Y^* and X_i^* are normalised in the interval [0, 100], and b_i is the weight of the i-th criterion.

Removing the monotonicity constraints, the size of the previous LP can be reduced in order to decrease the computational effort required for optimal solution search. This is effectuated via the introduction of a set of transformation variables, which represent the successive steps of the value functions Y^* and X_i^* (Siskos and Yannacopoulos, 1985; Siskos, 1985). The transformation equation can be written as follows (see also Figure 3):

$$\begin{cases} z_m = y^{*m+1} - y^{*m} & \text{for } m=1,2,\dots,\alpha-1 \\ w_{ik} = b_i x_i^{*k+1} - b_i x_i^{*k} & \text{for } k=1,2,\dots,\alpha_i-1 \text{ and } i=1,2,\dots,n \end{cases} \tag{2}$$

According to the aforementioned definitions and assumptions, the basic estimation model can be written in a linear program formulation as it follows:

[min]F = $\sum_{j=1}^{M} \sigma_j^+ + \sigma_j^-$

under the constraints

$$\begin{aligned} & \sum_{i=1}^{n}\sum_{k=1}^{t_{ji}-1} w_{ik} - \sum_{m=1}^{t_j-1} z_m - \sigma_j^+ + \sigma_j^- = 0, \text{ for } j=1,2,\dots,M \\ & \sum_{m=1}^{a-1} z_m = 100 \\ & \sum_{i=1}^{n}\sum_{k=1}^{a_i-1} w_{ik} = 100 \\ & z_m \geq 0,\ w_{ik} \geq 0,\ \forall m,i,k \\ & \sigma_j^+ \geq 0,\ \sigma_j^- \geq 0, \text{ for } j=1,2,\dots,M \end{aligned} \tag{3}$$

where *M* is the number of customers.

The preference disaggregation methodology consists also of a post optimality analysis stage in order to face the problem of multiple or near optimal solutions. The MUSA method applies a heuristic method for near optimal solutions search (Siskos, 1984). The final solution is obtained by exploring the polyhedron of near optimal solutions, which is generated by the constraints of the above linear program. During the post optimality analysis stage of the MUSA method, *n* linear programs (equal to the number of criteria) are formulated and solved. Each linear program maximizes the weight of a criterion and has the following form:

ISBN: 978-618-80361-1-6

$$\begin{cases} [\max] F' = \sum_{k=1}^{\alpha_i - 1} w_{ik} \quad \text{for } i = 1, 2, \ldots, n \\ \text{subject to} \\ F \le F^* + \varepsilon \\ \text{all the constraints of LP (3)} \end{cases} \tag{4}$$

where F^* is the optimal value of the objective function of LP (3) and ε is a small percentage of F^*.

The average of the optimal solutions given by the n LPs (4) may be considered as the final solution of the problem. In case of instability, a large variation of the provided solutions appears and the final average solution is less representative.

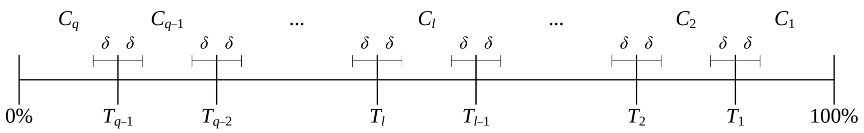
3. MODELING PREFERENCES ON CRITERIA IMPORTANCE

3.1. Model Development

A customer satisfaction survey may include, besides the usual performance questions, preferences about the importance of the criteria. Using such questions, customers are asked either to judge the importance of a satisfaction criterion using a predefined ordinal scale, or rank the set of satisfaction criteria according to their importance.

Based on such importance questions, each one of the satisfaction criteria can be placed in one of the following categories C_1, C_2, ..., C_q, where C_1 is the most important criterion class and C_q is the less important criterion class. Considering that C_l, with l the class index, are ordered in a 0-100% scale, there are T_{q-1} thresholds, which define the rank and, therefore, label each one of the classes (see Figure 1). Thus, the evaluation of preference importance classes C_l is similar to the estimation of thresholds T_l.

Figure1 Preference importance classes



An ordinal regression approach may also be used in order to develop the weights estimation model. The WORT (Weights evaluation using Ordinal Regression Techniques) model is presented in LP (5) in which the goal is the minimization of the sum of errors under a set of constraints according to the importance class that each customer j considers that a criterion i belongs (Grigoroudis and Spiridaki, 2003).

ISBN: 978-618-80361-1-6

$$
\begin{aligned}
&[\min] \quad F_2 = \sum_j \sum_i S_{ij}^+ + S_{ij}^- \\
&\left.\begin{aligned}
&\sum_{t=1}^{a_i-1} w_{it} - 100 \ T_1 - \delta + S_{ij}^- > 0, \quad \hat{b}_{ij} \in C_1 \\
&\left.\begin{aligned}
&\sum_{t=1}^{a_i-1} w_{it} - 100 \ T_{l-1} + \delta - S_{ij}^+ < 0 \\
&\sum_{t=1}^{a_i-1} w_{it} - 100 \ T_l - \delta + S_{ij}^- \geq 0
\end{aligned}\right\}, \hat{b}_{ij} \in C_l, \ l = 2,...,q-1 \\
&\sum_{t=1}^{a_i-1} w_{it} - 100 \ T_{q-1} + \delta - S_{ij}^+ < 0, \quad \hat{b}_{ij} \in C_q
\end{aligned}\right\} \quad \forall \quad i = 1,2,...,n \quad \text{και} \quad j = 1,2,...,M \\
&\sum_{i=1}^{n} \sum_{k=1}^{a_i-1} w_{ik} = 100 \\
&T_{q-1} \geq \lambda \\
&T_{q-2} - T_{q-1} \geq \lambda \\
&\vdots \\
&T_1 - T_2 \geq \lambda \\
&w_{ik}, S_{ij}^+, S_{ij}^- \geq 0, \forall \ i, \ j, k
\end{aligned}
\tag{5}
$$

Here, $\hat{b}_{ij}$ is the preference of customer *j* about the importance of criterion *i*., δ is a positive number, which is used in order to avoid cases where $b_{ij} = T_l \ \forall \ l$ and λ a minimum value introduced to increase the discrimination of the importance classes.

3.2. Combining Performance and Importance Judgments

Using together customers' performance and importance judgments, an extension of the MUSA method may be modeled as a Multiobjective Linear Programming (MOLP) problem (see MOLP(6)). The main purpose of this analysis is to examine whether additional information about the weights of the criteria can improve the results of the MUSA method.

$$
\begin{cases}
[\min] F = \sum_{j=1}^{M} \sigma_j^+ + \sigma_j^- \\
[\min] \Phi = \sum_{i=1}^{n} \sum_{j=1}^{M} S_{ij}^+ + S_{ij}^- \\
\text{subject to} \\
\text{all the constraints of LPs (3) and (5)}
\end{cases}
\tag{6}
$$

The examination of possible improvement is done through the Average Stability Index *(ASI)*. *ASI* is the mean value of the normalized standard deviation of the estimated weights b_i and is calculated as follows:

$$
ASI = 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{n \sum_{j=1}^{n} \left(b_i^j\right)^2 - \left(\sum_{j=1}^{n} b_i^j\right)^2}}{100\sqrt{n-1}}
\tag{7}
$$

2nd International Symposium and 24th National Conference on Operational Research
ISBN: 978-618-80361-1-6

where b_i^j is the estimated weight of the criterion *i*, in the j-th post-optimality analysis LP (Grigoroudis and Siskos, 2002).

Since competitiveness of the objective functions is the main characteristic of MOLP problems, searching for a solution that optimizes both *F* and *Φ* is rather pointless. The above problem may be solved using any MOLP technique (e.g. compromise programming, global criterion approach, etc.). Here, an alternative heuristic method, consisting of three steps, is presented (Grigoroudis et al., 2004):

Step 1:
Solve the following

$$\text{LP:}\begin{cases}[\min]F=\sum_{j=1}^{M}\sigma_j^{+}+\sigma_j^{-}\\ \text{subject to}\\ \text{all the constraints of LPs (3) and (5)}\end{cases} \tag{8}$$

Step 2:
Minimize the errors S_{ij}^{+} and S_{ij}^{-} using the following LP:

$$\begin{cases}[\min]\Phi=\sum_{i=1}^{n}\sum_{j=1}^{M}S_{ij}^{+}+S_{ij}^{-}\\ \text{subject to}\\ F\leq F^{*}+\varepsilon_1\\ \text{all the constraints of LPs (3) and (5)}\end{cases} \tag{9}$$

where F^* is the optimal value of the objective function of LP (8), and ε_1 is a small percentage of F^*.

Step 3:
Perform stability analysis (formulate and solve *n* LPs where each one maximizes the weight of a criterion):

$$\begin{cases}[\min]F'=\sum_{k=1}^{\alpha_i-1}w_{ik}\quad\text{for } i=1,2,\ldots,n\\ \text{subject to}\\ F\leq F^{*}+\varepsilon_1\\ \Phi\leq\Phi^{*}+\varepsilon_2\\ \text{all the constraints of LPs (3) and (5)}\end{cases} \tag{10}$$

where F^*, Φ^* are the optimal values of the objective functions of LPs (8)-(9), and ε_1, ε_2 are small percentages of F^* and Φ^*, respectively; similarly to the basic MUSA method, the final solution is calculated as the average of the optimal solutions of the previous LPs.

A detailed discussion about modelling preferences on criteria importance in the framework of the MUSA method, as well as real-world applications of the aforementioned approaches are given by Grigoroudis and Spiridaki (2003) and Grigoroudis et al. (2004).

4. NUMERICAL EXAMPLE

A hypothetical numerical example has been used in order to examine whether additional information about the weights of the criteria can improve the stability of the provided results. In this example 20 customers express their satisfaction globally and for 3 different criteria using a three-level qualitative scale (see Table 1). Similarly, according to Table 2, customers are called to express their preferences about the importance of the criteria using three importance classes.

2nd International Symposium and 24th National Conference on Operational Research
ISBN: 978-618-80361-1-6

Table 1 Performance Judgments

Overall	Criterion 1	Criterion 2	Criterion 3
Satisfied	Satisfied	Dissatisfied	Very satisfied
Satisfied	Dissatisfied	Satisfied	Very satisfied
Satisfied	Dissatisfied	Very satisfied	Satisfied
Very satisfied	Very satisfied	Very satisfied	Very satisfied
Very satisfied	Very satisfied	Very satisfied	Very satisfied
Satisfied	Very satisfied	Satisfied	Satisfied
Satisfied	Satisfied	Very satisfied	Dissatisfied
Satisfied	Satisfied	Very satisfied	Satisfied
Satisfied	Satisfied	Satisfied	Satisfied
Dissatisfied	Dissatisfied	Dissatisfied	Dissatisfied
Very satisfied	Very satisfied	Very satisfied	Satisfied
Satisfied	Satisfied	Very satisfied	Dissatisfied
Satisfied	Very satisfied	Satisfied	Dissatisfied
Very satisfied	Very satisfied	Satisfied	Very satisfied
Dissatisfied	Dissatisfied	Dissatisfied	Dissatisfied
Satisfied	Satisfied	Satisfied	Satisfied
Very satisfied	Satisfied	Very satisfied	Very satisfied
Satisfied	Very satisfied	Dissatisfied	Satisfied
Satisfied	Very satisfied	Dissatisfied	Very satisfied
Very satisfied	Very satisfied	Very satisfied	Satisfied

Table 2 Importance Judgments

Criterion 1	Criterion 2	Criterion 3
Important	Important	Important
Important	Very important	Unimportant
Important	Important	Important
Important	Very important	Unimportant
Important	Very important	Important
Important	Very important	Unimportant
Very important	Important	Unimportant
Very important	Important	Unimportant
Important	Very important	Important
Important	Very important	Unimportant
Important	Very important	Important
Important	Important	Important
Important	Very important	Unimportant
Very important	Important	Important
Important	Very important	Important
Important	Very important	Important
Important	Very important	Important
Unimportant	Important	Very important
Important	Important	Important
Important	Very important	Unimportant

Table 3 presents the results for a single set of values for the parameters of the MUSA extension model ($\gamma=\gamma_i=2$, $\lambda=0.1$, $\delta=0.015$, $\varepsilon=\varepsilon_1=\varepsilon_2=0.1$). Different scenarios with various values for the parameters have also been examined with similar results. According to Table 3, criterion 2 is the most important one and criterion 3 is the less important in all cases. But considering the *ASI* index there is a remarkable increase when additional information about the weights of the criteria is introduced regardless the MOLP technique chosen for the solution of the problem and it reaches almost 100% when using compromise programming.

Table 3 Results - comparison

	Basic MUSA model	Compromise programming	Global criterion	Heuristic method
Criterion 1 weight	30.51%	36.63%	36.04%	36.30%
Criterion 2 weight	40.12%	36.69%	37.27%	37.02%
Criterion 3 weight	29.37%	26.68%	26.69%	26.68%
ASI	**83.74%**	**99.98%**	**98.81%**	**99.31%**

5. CONCLUDING REMARKS AND FUTURE RESEARCH

The introduction of additional information about the importance of the criteria seems to increase the stability of the MUSA model. Further investigation with the simulation of data sets with various combinations of the parameters values or data with different statistical characteristics is also needed in order to validate the previous assumption. Using different criteria for the post-optimality analysis of the basic or the WORT MUSA models should also be investigated as it can give different results. The introduction of other additional information or constraints such as the linkage between global and partial satisfaction and demanding indices may also decrease instability of the MUSA method and should also be examined. Finally, developing additional measures of robustness may facilitate the investigation of various extensions of the MUSA model.

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

REFERENCES

Grigoroudis, E. and O. Spiridaki (2003). *Derived vs. stated importance in customer satisfaction surveys*, Operational Research: An International Journal, 3 (3), 229-247.

Grigoroudis, E. and Y. Siskos (2002). *Preference disaggregation for measuring and analysing customer satisfaction: The MUSA method,* European Journal of Operational Research, 143 (1), 148-170.

Grigoroudis, E., Y. Politis, O. Spiridaki, and Y. Siskos (2004). *Modelling importance preferences in customer satisfaction surveys,* in: C.H. Antunes, J. Figueira, and J. Climaco (eds.), Proceedings of the 56th Meeting of the European Working Group “Multiple Criteria Decision Aiding”, INESC Coimbra, 273-291.

Jacquet-Lagrèze, E. and J. Siskos (1982). *Assessing a set of additive utility functions for multicriteria decision-making: The UTA method,* European Journal of Operational Research, 10 (2), 151-164.

Siskos, J. (1984). *Le traitement des solutions quasi-optimales en programmation linéaire continue: Une synthèse*, RAIRO Recherche Opérationnelle, 18, 382-401.

Siskos, J. (1985). *Analyses de régression et programmation linéaire*, Révue de Statistique Appliquée, 23 (2), 41-55.

Siskos, Y. and D. Yannacopoulos (1985). *UTASTAR: An ordinal regression method for building additive value functions*, Investigaçao Operacional, 5 (1), 39-53.

ISBN: 978-618-80361-1-6

Constructing Robust Efficient Frontiers for Portfolio Selection Under Various Future Returns Scenarios

G. Mavrotas
School of Chemical Engineering,
National Technical University of Athens
Iroon Polytechniou, 9, 15780 Athens
**mavrotas@chemeng.ntua.gr*

P. Xidonas, H. Doukas, J. Psarras
School of Electrical and Computer Engineering,
National Technical University of Athens
Iroon Polytechniou, 9, 15780 Athens

Abstract

An efficient frontier in the typical portfolio selection problem consists an illustrative way to express the tradeoffs between return and risk. Following the basic ideas of modern portfolio theory as introduced by markowitz the security returns are usually extracted from past data. This work is an attempt to incorporate future returns scenarios in the investment decision. For representative points of the efficient frontier, the minimax regret portfolio is calculated on the basis of the aforementioned scenarios. These points correspond to specific weight combinations. in this way, the areas of the efficient frontier that are more robust than others are identified. An example with the 50 securities from eurostoxx 50 is also presented to illustrate the method.

KEYWORDS

Multiobjective programming; Robust programming; Financial modeling

1. INTRODUCTION

In financial theory, models allowing the selection of an optimal portfolio are all inspired from the classical theory of Markowitz (1952, 1959), which is exclusively based on the criteria of expected value and variance of the return distribution. In this regard, an investor considers expected return as desirable and variance of return as undesirable. The Markowitz's theory describes how we calculate a portfolio which exhibits the highest expected return for a given level of risk, or the lowest risk for a given level of expected return (efficient portfolio). Then, according to the theory, the problem of portfolio selection is a single-objective quadratic programming problem, which consists in minimizing risk, while keeping in mind an expected return which should be guaranteed. Thus, the solution of the original bi-objective model is reduced to the parametric solution of a single objective problem, providing the efficient (or Pareto optimal) portfolios (Xidonas et al, 2010, 2011).

In Modern Portfolio Theory (MPT) the return and risk for each stock in the investment universe are usually extracted from past data. In this work we try to incorporate future scenarios for the return and risk that is mainly based on the perspectives of the investor/decision maker. It is an attempt to show how this information may be exploited and produce robust portfolios against the future scenarios. We deal with future scenarios using the concept of the minimax regret criterion as it was proposed for Mathematical Programming problems by Kouvelis and Yu (1997).

The methodological contribution of the present work is that it expands the concept of the robust solution as it was proposed by Kouvelis and Yu to the multi-objective case. Namely, we use the minimax regret criterion in order to identify robust Pareto optimal solutions in the Pareto front of multi-objective problems. In this way we can identify areas of the Pareto front that are more robust than other. The specific areas of the Pareto front are characterized by the weight combination used in the objective functions.

The remainder of the paper is as follows: In the second section we describe the methodology and in the third section we present the application that illustrates the method. Finally, in section 4 the main conclusions are presented.

2. METHODOLOGY

2.1. The Minimax Regret Criterion in Mathematical Programming

It is well known that the minimax regret criterion is among the most popular criteria in Decision Science along with the maximax, maximin, Hurcwitz criterion etc) where different scenarios are present. It was introduced by Savage (Savage, 1954). It actually aims ate selecting the solution or alternative which is under the worst case closer to its scenario optimum. Assume that we have the following payoff table with 3 alternatives and 5 scenarios (assume that the optimum is the maximum which are the yellow cells).

Table 1 Example of Payoff matrix (a) and regret matrix (b)

	(a) Payoff matrix					(b) Distance from optimum (regret matrix)					
	Scen1	Scen2	Scen3	Scen4	Scen5	Scen1	Scen2	Scen3	Scen4	Scen5	Worst case
Solution 1	3	12	14	6	6	8	0	0	3	4	8
Solution 2	11	8	7	6	10	0	4	7	3	0	7
Solution 3	8	8	11	9	5	3	4	3	0	5	5

In the rightmost column we calculate the maximum of each row to find the worst case scenario for each solution (worst case in the sense that is more distant from the optimal of the scenario). The minimax regret solution is the one that has the minimum among the worst case values which is Solution 3 in our case. Kouvelis & Yu (1997) accomplish this task not for a finite number of solutions but for an infinite number of solutions according to the feasible region of the problem. Other attempts include the works of Mausser and Laguna (1999) as well as Loulou and Kanudia (1999). According to Kouvelis and Yu (1997) assume the following mathematical programming problem:

$$\max z = cx$$
$$st$$
$$x \in F$$

Assume now that we have S scenarios for the objective function coefficients which means that the corresponding objective function coefficient vectors are denoted as c^s. The minimax regret solution is calculated from the following problems:

(a) Relative regret

$$z_{MMR} = \min y$$
$$st$$
$$c^s x \geq (1-y) z^s \quad s \in S$$
$$x \in F$$

(b) Deviation regret

$$z_{MMR} = \min y$$
$$st$$
$$c^s x \geq z^s - y \quad s \in S$$
$$x \in F$$

ISBN: 978-618-80361-1-6

Where z^s is the optimal value for the s-scenario and y is the variable that expresses the minimax regret.

2.2. Extension of the Minimax Regret Criterion to the Multi-objective Context

In the present work we extend the formulation of Kouvelis and Yu to the multi-objective case. Specifically we use the weighting method in order to calculate the Pareto optimal solutions of the Pareto front. In the weighting method we solve one single-objective problem that has as objective function the weighting sum of the objective functions at hand (assume all objectives are for maximization).

$$\max z = \sum_{p=1}^{P} w_p \times \frac{c_p x - f_{p,\min}}{f_{p,\max} - f_{p,\min}} \qquad st \qquad x \in F$$

Where $f_{p,min}$ and $f_{p,max}$ are the minimum and the maximum values of the objective functions as obtained from the payoff table. The solution of this problem corresponds to a Pareto optimal solution of the multi-objective problem. Varying the weights we obtain a representative set of the Pareto optimal solutions of the multi-objective problem. The concept of our method is to apply the Kouvelis and Yu formulation to every combination of the weights. In this way we obtain the minimax regret solutions that correspond to different areas of the Pareto front. Assuming that we have S scenarios for the objective function coefficients, we discretize the weight space to g weight combinations (vectors) and we solve the following problem:

$$MMR(g) = \min y_g$$

$$st$$

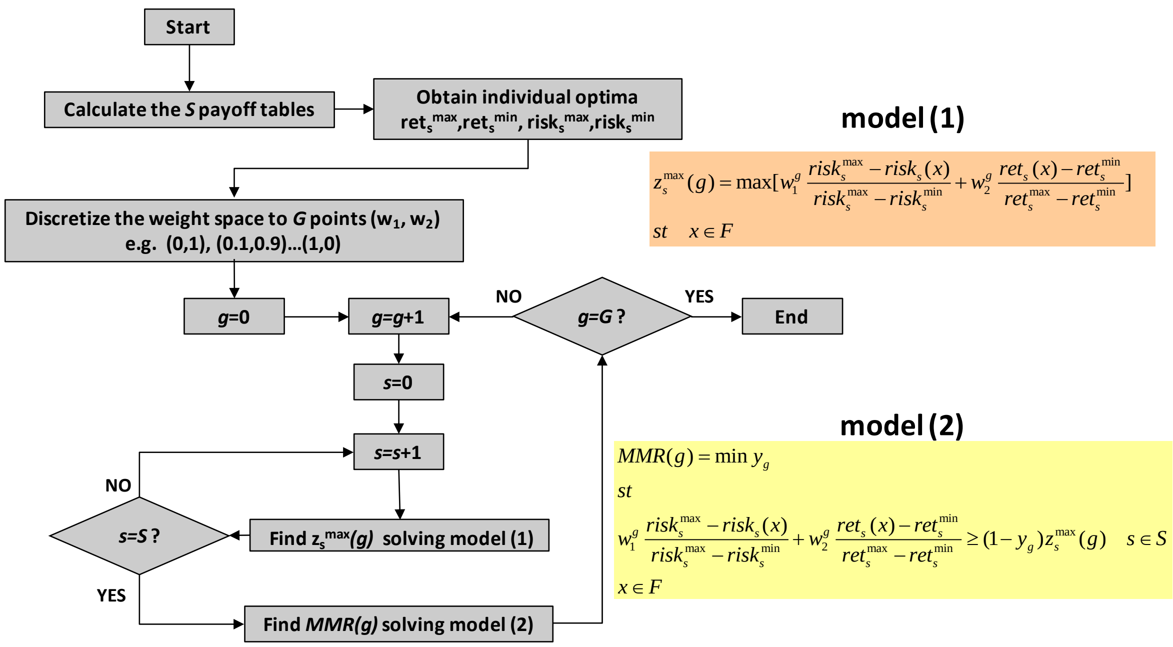
$$\sum_{p=1}^{P} w_p^g \times \frac{c_p^s x - f_{p,\min}^s}{f_{p,\max}^s - f_{p,\min}^s} \geq (1 - y_g) z_g^s \qquad s \in S$$

$$x \in F$$

and we solve the above problem for every g obtaining the MMR solution at representative points of the Pareto front. According to the value of the minimax regret solution (y_g) we can draw conclusions about the areas of higher or lower robustness of the Pareto front. The flowchart of the algorithm adjusted to the specific case study where the objective functions are maximize return (*ret*) and minimize risk is depicted in Figure 1.

2nd International Symposium and 24th National Conference on Operational Research
ISBN: 978-618-80361-1-6

Figure 1. The flowchart of the method for minimax regret criterion

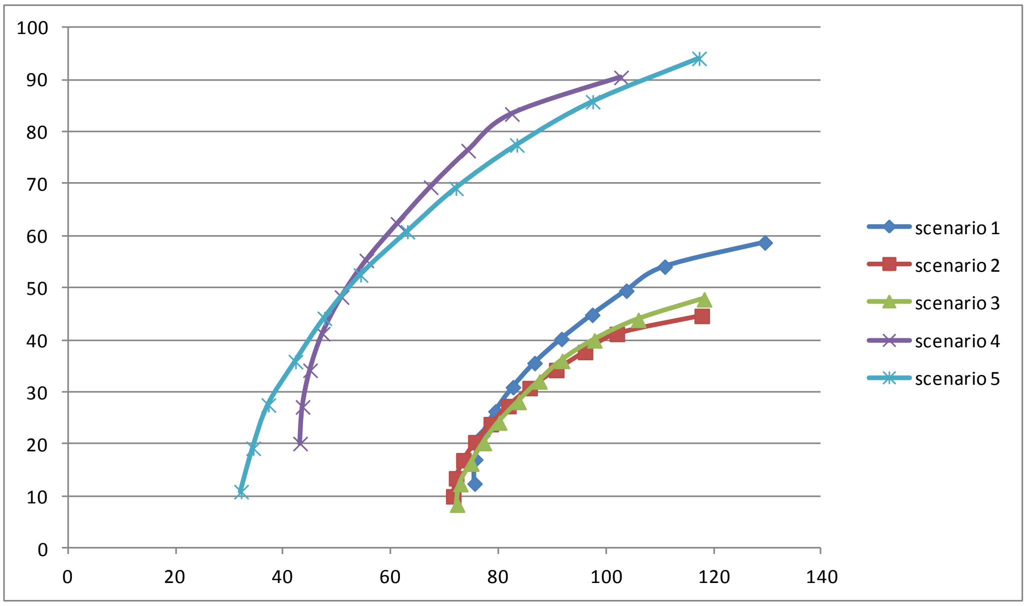


3. APPLICATION

We use the 50 stocks of the Eurostoxx 50 the Europe's leading blue-chip index for the Eurozone, provides a high capitalization representation of supersector leaders in the Eurozone. The index covers 50 stocks from 12 Eurozone countries. The model is presented in Xidonas and Mavrotas (2013). We use five scenarios of return and risk evolution. In the absence of actual decision makers we create 5 scenarios for the return and the risk as follows: We used historical data of 80, 60, 40, 20, and 10 weeks extracting the average return and Mean Absolute Deviation (MAD) as a risk measure from the corresponding data. Therefore scenario 1 that corresponds to 80 weeks past horizon denotes a more long-term point of view than scenario 2, 3, 4 and 5 that denotes a short-term behavior. The five efficient frontiers are depicted in Figure 2.

ISBN: 978-618-80361-1-6

Figure 2: The five efficient frontiers



The obtained results using the minimax regret model are shown in Table 2. We used 11 weight combinations, namely (0,1), (0.1,0.9), (0.2,0.8)…(0.9,0.1), (1,0). The optimum of each scenario for the weight combinations (0,1), (0.1,0.9) and (1,0) along with the minimax regret solution are shown in Table2 (the first objective function is the minimization of risk and the second one is the maximization of return)

Table 2: The output format of the obtained solutions for three weight coefficients

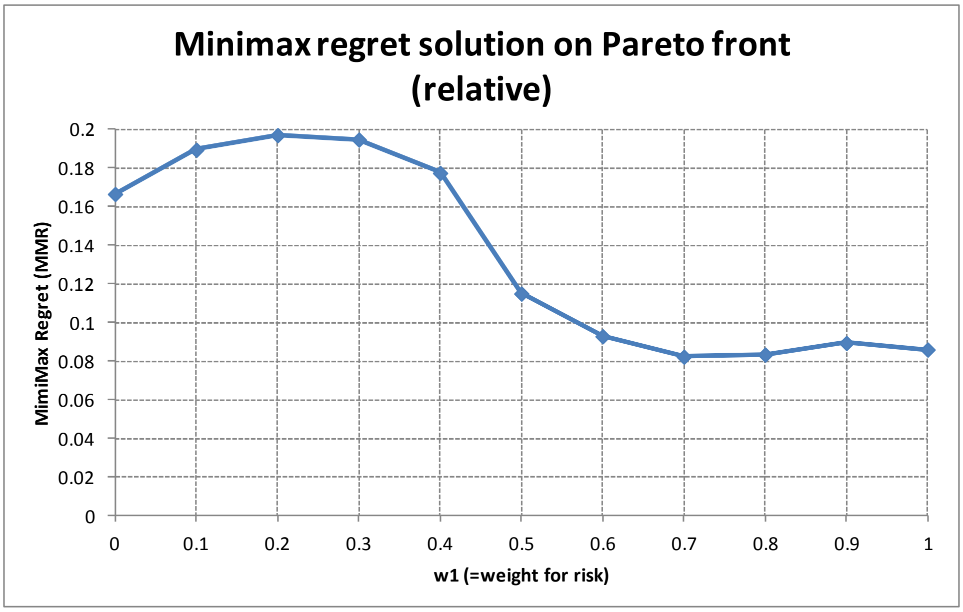
w1=	0.0																
SCEN#	WSUM	MAD	Return	Stck/Portf	1	2	3	4	5	6	7	8	9	10	11	......	50
1	0.999	2.4902	58.731	10	0	0.1	0	0	0	0	0	0.1	0	0	0.1		0
2	0.999	2.265	44.632	10	0	0.1	0	0	0	0	0	0.1	0	0	0.1		0
3	0.999	2.2734	47.821	10	0	0.1	0	0	0	0	0	0.1	0	0	0.1		0
4	0.999	1.9764	90.459	10	0	0.1	0	0	0	0	0	0	0	0	0.1		0
5	0.999	2.2553	94.112	10	0	0.1	0	0	0	0.1	0	0	0	0	0.1		0
MMR=	0.1666*		*	11	0	0.1	0	0	0	0	0	0	0	0	0.1		0
w1=	0.1																
SCEN#	WSUM	MAD	Return	Stck/Portf	1	2	3	4	5	6	7	8	9	10	11		50
1	0.9061	2.2974	58.089	10	0	0.1	0	0	0	0	0	0.1	0	0	0.1		0.1
2	0.9073	2.0772	44.1	10	0	0.1	0	0	0	0	0	0.1	0	0	0.1		0
3	0.9	2.2734	47.821	10	0	0.1	0	0	0	0	0	0.1	0	0	0.1		0
4	0.9097	1.8486	90.348	10	0	0.1	0	0	0	0	0	0	0	0	0.1		0
5	0.9	2.2553	94.112	10	0	0.1	0	0	0	0.1	0	0	0	0	0.1		0
MMR=	0.1895*		*	12	0	0.1	0	0	0	0.06	0	0.088	0	0	0.1		0
w1=	1.0																
SCEN#	WSUM	MAD	Return	Stck/Portf	1	2	3	4	5	6	7	8	9	10	11		50
1	0.999	1.4528	12.402	12	0	0	0.1	0	0	0.1	0	0.042	0	0	0		0
2	0.999	1.3768	9.962	12	0	0	0.047	0	0	0.099	0	0.1	0	0	0		0
3	0.999	1.3905	8.3581	13	0	0	0.06	0	0	0.07	0	0.1	0	0	0		0.067
4	0.999	0.8291	20.1063	13	0	0	0	0	0.01	0.057	0	0	0	0	0		0
5	0.999	0.6182	10.8648	11	0	0.1	0	0	0	0	0	0.1	0	0	0		0
MMR=	0.086*		*	13	0	0.02	0	0	0	0.068	0	0.1	0	0	0		0

ISBN: 978-618-80361-1-6

We can observe that the minimax regret solution in all the weight combinations contains more stocks in the final portfolio than the individual scenarios optima. This means that we have a more dispersed allocation of the total investment.

The minimax regret (MMR) solution across the Pareto front is obtained from the MMR values for the specific weight combinations. Hence we can detect areas of the Pareto front that present relatively increased robustness in relation to other areas. We calculate the MMR solutions for the 11 weight coefficients for the relative and the deviation regret criterion (see section 2). The results are shown in Figure 3 for the relative MMR.

Figure 3 The MMR values across the Pareto front (relative MMR)



In Figure 3 the lower the MMR the more robust is the specific area of the Pareto front. Hence It is obvious that there are areas in the Pareto front with higher robustness based on the 5 scenarios. For example the Pareto optimal solutions that correspond to weights varying from 0.1 to 0.4 are less robust than the Pareto optimal solutions that correspond to weights varying from 0.6 to 1 (robust area of the Pareto front).

4. CONCLUSIONS

We extend the Kouvelis Yu formulation for the minimax regret criterion in multi-objective programming problems. Using the weighting method for generating the Pareto optimal solutions we can detect the robust areas of the efficent frontier. Future research can be driven towards examining the effectiveness of the method for more objective functions and also in other applications.

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning"

REFERENCES

Markowitz, H., 1952. *Portfolio selection*. The Journal of Finance, 7 (1), 77-91.

Markowitz, H., 1959. *Portfolio Selection: Efficient Diversification of Investments*. John Wiley and Sons, New York.

Kouvelis, P., Yu, G. 1997. *Robust discrete optimization and its applications*. Kluwer, Amsterdam

Loulou R and Kanudia A., 1999 *Minimax regret strategies for greenhouse gas abatement: methodology and application*. Operations Research Letters 25; 219-230

Mausser HE, Laguna M., 1999 *Minimising the maximum relative regret for linear programmes with interval objective function coefficients*. Journal of Operational Research Society 50; 1063-1070.

Savage LJ. *The Foundations of Statistics*. Wiley: New York; 1954.

Xidonas, P., Mavrotas, G., 2013. *Multiobjective portfolio optimization with non-convex policy constraints: Evidence from the Eurostoxx 50*. European Journal of Finance, To appear.

Xidonas, P., Mavrotas, G., Zopounidis, C., Psarras, J., 2011. *IPSSIS: An integrated multicriteria decision support system for equity portfolio construction and selection*. European Journal of Operational Research, 210 (2), 398-409.

Xidonas, P., Mavrotas, G., Psarras, J., 2010. *Equity portfolio construction and selection using multiobjective mathematical programming*. Journal of Global Optimization, 47 (2), 185-209.

ISBN: 978-618-80361-1-6

Value Focused Pharmaceutical Strategy Determination with Multicriteria Decision Analysis Techniques

K. Mastorakis
University of Piraeus
80, Karaoli and Dimitriou str., 18534, Piraeus, Greece
kostmast28@yahoo.gr

E. Siskos
National Technical University of Athens
9, Iroon Polytechniou str., 15780, Zografou, Athens, Greece
lsiskos@epu.ntua.gr

Y. Siskos
University of Piraeus
80, Karaoli and Dimitriou str., 18534, Piraeus, Greece
yannsiskos@gmail.com

Abstract

The size of the pharmaceutical market and its contribution, both in national and global level, to the regional, national and international economic development is widely recognized. This fact signifies that the supported and efficient decision making in the sector is a matter of paramount importance. This paper refers to a multicriteria assessment system for portfolio optimization in the Hellenic pharmaceutical market. The evaluation criteria are extracted from three points of view, namely: i) current market situation, ii) development of the sector over the recent years, and iii) comparison with other European countries. The overall objective of this research work is the assessment and ranking of 192 therapeutic categories for investment purposes in the Hellenic pharmaceutical market. The ranking of these categories is obtained through the utilization of an additive value model which is assessed by the ordinal regression method utastar, implemented in three phases. in the first phase, the decision maker (dm) is asked to rank a sample of these alternatives, to infer an additive value system which should be as close as possible to the dm's ranking and as robust as possible. In a second phase, all the alternatives are evaluated and a complete ranking is obtained. Finally, in order to analyze the robustness of the model, given the incomplete determination of inter-criteria model parameters, a random weighting sampling technique is utilized, to obtain the probability that an alternative maintains its initial position in the ranking.

KEYWORDS

Multiple Criteria Decision Analysis, Pharmaceutical market strategy, Ordinal regression, Robustness analysis

1. INTRODUCTION

Greece throughout the last three years is experiencing an economic crisis and recession in a scale unprecedented within Eurozone. Data analyzing and decoding the messages that this murky economic environment is signaling constitutes the black box that everyone is search of, in order to optimize the outcome of a forecast. So here lays the question of what a company, which is willing to invest, can really do?

Even though the attempt to invest within an economic environment, with the aforementioned characteristics, is a risk, investments and investors will always emerge when opportunities are likely to appear. In such an economic turmoil opportunities will arise but the key element that will minimize the risk factor, in terms of capital and labor forces loss, is diversification of the investment portfolio along with in depth analysis of the data of the market in which the investments will take place.

One sector that has succeeded in withstanding the economic crisis is the pharmaceutical. Having said that the sector has suffered lately a considerable downsize in terms of gross revenue and labor force reduction (Athanasiadis et al., 2012). Despite the latter, the pharmaceutical sector, which is consisted of the companies that manufacture or import drugs or non-pharmaceutical products, is one of the main pillars of the Greek industry and one of the largest employers in the Greek economy. According to the Greek Pharmaceutical Companies Association the expenditure on pharmaceutical products represents 2.5% of the

GDP from 2007 and onwards, while pharmaceutical products constituted the 5% of the gross exports of the country and more than 6% of the gross imports in 2010 (Athanasiadis et al., 2012).

The contribution of the pharmaceutical sector in the domestic economy is not restricted at the commercial aspect, it affects considerable the labor factors as well. According to data provided by the Division of Labor force of the Hellenic Statistics Agency, in 2011 approximately 13.600 persons were employed at the sector of pharmaceutical manufacture, a number which renders the sector a main player in the overall labor within the Greek economy (Bank of Piraeus, 2011).

The ongoing downturn of the pharmaceutical market since 2009 with a rate of 12% per annum, with the situation being exacerbated the last two, with a downturn rate of 18% per annum (see Figure 1), cannot be interpreted that the companies should halt any investment program. Within this environment the companies should have a balanced scorecard and carefully assess the investment targets.

Figure 1: Pharmaceutical Sector - value sales evolution (Data source: IMS Hellas).

The purpose of this paper is to provide the appropriate tools and methodologies to assist the managers of a pharmaceutical company in investing on new pharmaceutical products. The paper proposes an integrated multicriteria decision aid (MCDA) methodology to extract and imprint the preferences of the decision makers-managers of a pharmaceutical company.

The study described below elaborates with the involvement of two real experts in the pharmaceutical market in the role of the decision maker. In the end, a personalized ranking of the possible therapeutic categories for investment emerges, from which the company will select those top ranked to form its investment portfolio. At the end of the paper, the robustness of the model and its results is analyzed, to check for differentiations over the final ranking, supporting therefore further the investment decision of the company.

2. PROBLEM STATEMENT AND THE MULTICRITERIA EVALUATION SYSTEM

In the case of a pharmaceutical company, there is a plethora of decisions that managers are obligated to take. Through this paper, the problematic concerning the launch of a prototype drug in the Greek market is

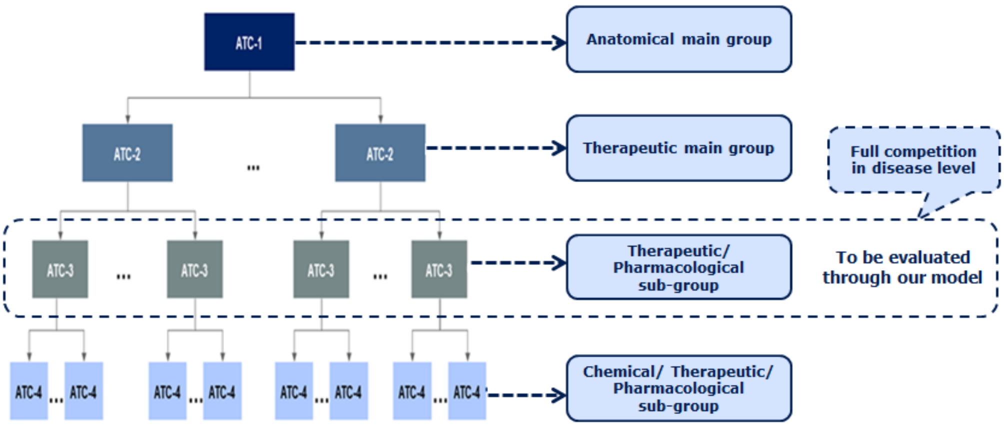
ISBN: 978-618-80361-1-6

analyzed, which the company already has launched abroad or may purchase it from another company and then import it in the Greek market.

The success or failure of this new launch is affected by many different parameters and criteria that differentiate according to which strategy the company is willing to follow. For instance, some would opt the strategy to target large therapeutic categories with low pricing relying on their organized production and brand recognition, in order to eliminate head on the competition, while others would prefer to invest in smaller therapeutic categories where the competition is less dynamic and precise (Jones, 2002).

Besides the scope under which the decision makers set their preferences, the main purpose of this simulation is the assessment of all 192 existing therapeutic categories (EphMRA, 2012) of level three (ATC-3s, see Figure 2), through a number of evaluation criteria to be modeled in the following section. Consequently, their preferences can be prioritized on a scale of importance to select the therapeutic category they will be interested to invest upon.

Figure 2: Anatomic Therapeutic Classification



After reviewing the literature for similar surveys, focusing on the evaluation of therapeutic categories, a problem of confidentiality occurred since these projects/surveys constitute a main tool of corporate strategy in real corporate life.

For instance, major consulting groups are involved in the assessment of pharmaceutical categories in order to provide consistent and credible data to assist their clients to invest in the most profitable way. In Greece, companies like Boston Consulting Group, McKinsey and especially IMS Consulting Group are the main players in providing services concerning similar problems with the ones this paper is attempting to address.

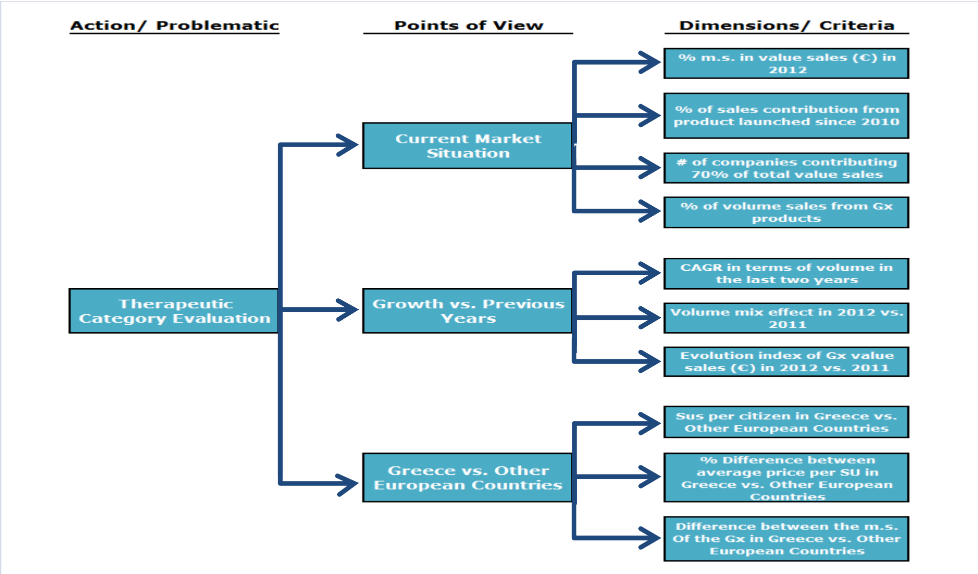
2.1. Evaluation Criteria Modeling

Pharmaceuticals are a market with a continuous flow of sales data, numerous products, in depth historical data, constituting the market an open field for analysis, deployment of statistical model and in general of any tool of study. To achieve an overall assessment of the pharmaceutical categories, a consistent family of criteria is built according to the classical modeling methodology of Roy (1985), taking into consideration the severe Greek economic crisis. More specifically, the problematic of the paper, namely the evaluation of therapeutic categories is divided into three points of view: (i) the current market situation in Greece, (ii) the growth of the categories over the previous years, and (iii) the financial differentiations between Greece and

ISBN: 978-618-80361-1-6

the rest EU countries. These points of view are farther analyzed to form the ten evaluation criteria of the study which are presented in Figure 3 and briefly described below.

Figure 3: The multicriteria evaluation system of Therapeutic Category Investment.



The indices measuring the criteria along with their evaluation scales are presented in Table 1. The data of all ten criteria stem from IMS Hellas (see data sources).

g_1: % market share in value sales (€) in 2012: It is the value market share that each ATC-3 captures within the total pharmaceutical market for 2012 calendar year.

g_2: % of sales contribution from products launched since 2010: The value sales market share of new launched products (products launched within the last three years) for each ATC-3 is calculated. It is a way to calculate how easily a new product launch can capture market share within the category.

g_3: # of companies contributing 70% of total value sales: The number of companies that all together contribute the 70% of total value sales for this category. In this way the level of concentration in sales within category is quantified.

g_4: % of volume sales from Generics products: This criterion is a measure to evaluate the penetration of Gx products within the category. The volume market share is calculated in order to eliminate the price effect from the original products which are more expensive.

g_5: % CAGR in terms of volume sales over the last two years: Volume evolution of the ATC-3 category within the last two years.

g_6: Volume mix effect in 2012 vs. 2011: It is a metric calculating the switch from more expensive to less expensive products revealing how easily the patients switch their brand of choice.
g_7: % Evolution index of Generics in value sales (€) in 2012 vs. 2011: Calculation of the growth of the category vs. the growth of the total market.

ISBN: 978-618-80361-1-6

g_8: Number of Standard Units per capita in Greece vs. Other European countries: This criterion measures the deviation in consumption in terms of dosages per capita in Greece vs. Other European countries.

g_9: % of difference between average price per Standard Unit in Greece vs. Other European countries: This criterion measures the differentiation between the average price of the category in Greece vs. Other European countries.

g_{10}: Difference between the Generics market share (in Standard Units) in Greece vs. Other European countries: The delta between the penetration of Gx in Greece and Other European countries is calculated.

Table 1: Evaluation scales of the criteria

Criterion	Index	Worst Level	Best Level	Optimize	Point of View
g_1	% of sales	0.02%	7.57%	maximize	Current Market Situation
g_2	% of sales	0%	100%	maximize	
g_3	Integer	1	13	maximize	
g_4	% of sales	0%	85.4%	maximize	
g_5	% of sales	-71.8%	100%	maximize	Growth vs. Previous Years
g_6	% of sales	0%	61.4%	maximize	
g_7	% ppts	100%	0%	minimize	
g_8	% ppts	-100%	100%	maximize	Greece vs. Other EU Countries
g_9	% ppts	-100%	100%	maximize	
g_{10}	% ppts	92.5%	0%	minimize	

Following the discussion with the decision makers, the alternatives selected for evaluation in this study are the level three therapeutic categories (ATC-3) and more specifically the ones that exhibited gross sales above 500,000 euros in 2012. Their preferences revealed that they are mostly interested to invest on mature categories in terms of Generics penetration in order to avoid potential instability due to the latest changeable legislation (Maniadakis et al., 2010).

2.2. The Additive Value Model

The evaluation methodological frame to aggregate the above criteria proposes the assessment of a mulicriteria additive value system, which is described by the following mathematical formulae:

$$u(\boldsymbol{g}) = \sum_{i=1}^{n} p_i \, u_i(g_i)$$

$$u_i(g_{i*}) = 0, \quad u_i(g_i^*) = 1, \qquad 0 \leq u_i(g_i) \leq 1 \qquad \forall\, i = 1,2,\dots,n$$

$$\sum_{i=1}^{n} p_i = 1, \; p_i \geq 0, \; \forall\, i = 1,2,\dots,n$$

where $\boldsymbol{g} = (g_1, g_2, \dots, g_n)$ is the performance vector of an action on the *n* criteria; g_{i*} and g_i^* are the least and most preferable levels of the criterion g_i, respectively; $u_i(g_i)$, i =1,2,…,n are non-decreasing marginal

ISBN: 978-618-80361-1-6

value functions of the performances g_i, $i = 1,2, \dots, n$; and p_i is the relative weight of the i-th function $u_i(g_i)$.

An additive value function is valid in the case of an individual decision maker (DM) if the criteria are preferentially independent from each other (see Keenney and Raiffa, 1976, Keeney, 1992, for instance). A number of different methods may be utilized to obtain the aforementioned additive value system (see Keeney and Raiffa, 1976, Figueira et al., 2005). In this study the ordinal regression UTASTAR method (Siskos and Yannacopoulos, 1985) is implemented to assess and construct the additive value system.

2.3. The UTASTAR Method

The ordinal regression or disaggregation UTASTAR method is an enhanced version of the original UTA model (Jacquet-Lagrèze and Siskos, 1982). It infers one or more piecewise linear value functions from pairwise comparisons or a ranking of *m* reference actions given by the decision maker. The estimation of each marginal value function is made on a number of discrete points g_i^j, $j = 2, 3, \dots, \alpha_i - 1$. Assuming that a_1 is the head of the ranking and $a_k P \; or \; I a_{k+1}$ (a_k *is preferred or indifferent to* a_{k+1}), the UTASTAR algorithm is summarized in the following steps:

Step 1: Express the global value of reference actions $u[g(a_k)]$, *k =1,2, ..., m*, first in terms of marginal values $u_i(g_i^j)$, and then in terms of variables $w_{ij} = u_i(g_i^{j+1}) - u_i(g_i^j)$ $\forall i$ and j, by means of the following expressions:

$$\begin{cases} u_i(g_i^1) = 0 & \forall i = 1, 2, \dots, n \\ u_i(g_i^j) = \sum_{t=1}^{j-1} w_{it} & \forall i = 1, 2, \dots, n \; and \; j = 2, 3, \dots, a_i - 1 \end{cases}$$

Step 2: Introduce two error functions σ^+ and σ^- by writing for each pair of consecutive actions in the ranking the analytic expressions:

$$\Delta(a_k, a_{k+1}) = u[g(a_k)] - \sigma^+(a_k) + \sigma^-(a_k) - u[g(a_{k+1})] + \sigma^+(a_{k+1}) - \sigma^-(a_{k+1})$$

Step 3: Solve the linear program:

$$\begin{cases} [\min] z = \sum_{k=1}^{m} [\sigma^+(a_k) + \sigma^-(a_k)] \\ subject\ to \\ \left.\begin{array}{ll} \Delta(a_k, a_{k+1}) \geq \delta & if\ a_k P\, a_{k+1} \\ \Delta(a_k, a_{k+1}) = 0 & if\ a_k I\, a_{k+1} \end{array}\right\} \quad \forall k \\ \sum_{i=1}^{n} \sum_{j=1}^{\alpha_i - 1} w_{ij} = 1 \\ w_{ij} \geq 0, \sigma^+(a_k) \geq 0, \sigma^-(a_k) \geq 0 \; \forall i, j \; and \; k \end{cases}$$

with δ being a small positive number.

Step 4: Robustness analysis (Centroid value function based on multiple characteristic value functions, robust ordinal regression analysis, see section 4 below).

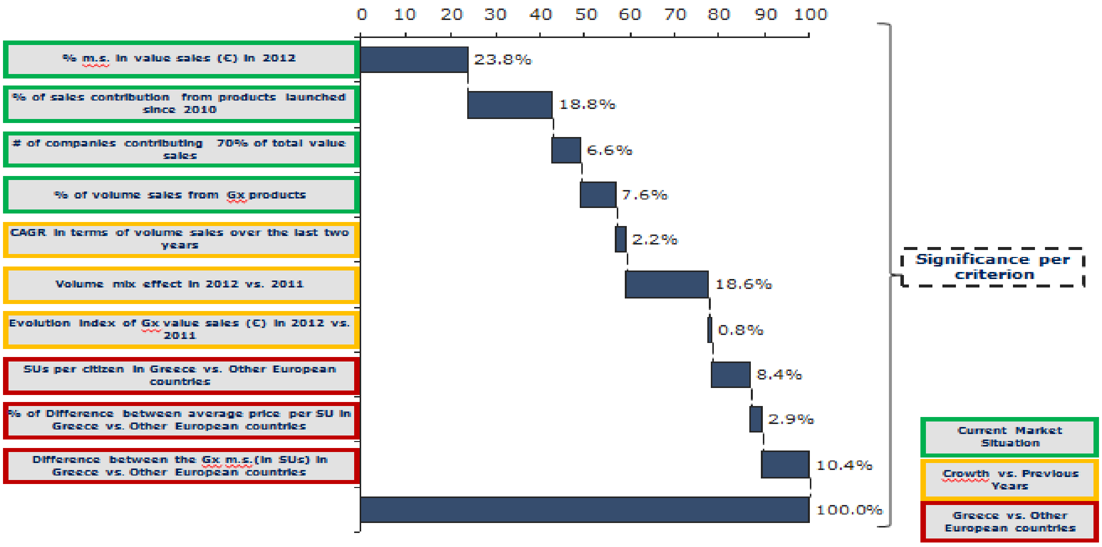
ISBN: 978-618-80361-1-6

3. IMPLEMENTATION OF THE MODEL– RESULTS

Finalizing the structure of the evaluation model, the parameters of the model are calculated before its implementation. A sub-set of thirty alternatives from 192 was selected to form the reference set. Much effort was spent to construct this set in order to be representative of the whole entity of pharmaceutical categories and to avoid being very complicated for the DMs to rank. The constructed reference set, its evaluation by the DMs and the confirmation of this evaluation by the UTASTAR method are presented in Table 2.

Figure 4 presents the estimated weights $p_i, i = 1,2,\dots,8$ of the model that emerged after the solution of the linear program of the UTASTAR method, for the calculated centroid that resulted by the two rankings.

Figure 4: Weight per criterion based on the calculated centroid



The final stage of the evaluation process is the implementation of the already constructed model on all 192 alternatives to achieve their evaluation and ranking. The results of this stage, also known as extrapolation, that emerge after the implementation of the additive value model are presented clearly in Table 3. As normally expected the 30 alternatives maintain their initial ranks on the final complete ranking, as shown in Table 2.

Table 2: Reference actions the joint DMs' ranking and its confirmative values given by UTASTAR method

DMs Ranking	ATC-3 Short	Global value	Final Ranking	Sample Ranking
1	C10A	0,501924	1	1
2	M04A	0,429213	3	2
3	J06C	0,401858	5	3
4	M05B	0,391837	7	4
5	A02B	0,381771	9	5
6	A10S	0,371679	10	6
7	J07A	0,361775	11	7
8	J02A	0,351731	12	8

Table 3: Sample of the final ranking of the 192 alternatives

Final Ranking	ATC-3 Short	Global Value
1	C10A	0,501924
2	N05A	0,429786
3	M04A	0,429213
4	C09D	0,409905
5	J06C	0,401858
6	B01C	0,394716
7	M05B	0,391837
8	D05B	0,382879

ISBN: 978-618-80361-1-6

9	N07X	0,341827	**14**	9
10	M01C	0,330631	**16**	10
11	A10N	0,319845	**18**	11
12	L01D	0,308773	**23**	12
13	A11C	0,302083	**24**	13
14	C04A	0,292183	**27**	14
15	G04C	0,282283	**32**	15
16	R03A	0,272234	**35**	16
17	L01X	0,262334	**38**	17
18	A10C	0,252306	**40**	18
19	D07A	0,239034	**47**	19
20	T02D	0,230811	**52**	20
21	L01C	0,222591	**63**	21
22	L04X	0,212686	**69**	22
23	B01B	0,202786	**80**	23
24	G04D	0,192885	**97**	24
25	T01A	0,190632	**102**	25
26	J07B	0,185825	**111**	26
27	B03C	0,162716	**136**	27
28	L03B	0,151817	**147**	28
29	C06A	0,141331	**153**	29
30	L04B	0,138498	**154**	30

9	A02B	0,381771
10	A10S	0,371679
11	J07A	0,361775
12	J02A	0,351731
13	H01C	0,343724
14	N07X	0,341827
15	R03H	0,341042
…	**…**	**…**
…	**…**	**…**
181	R03X	0,102531
182	H03B	0,102266
183	C01A	0,101084
184	G01B	0,100916
178	A10M	0,106144
185	A10H	0,093215
186	B01X	0,086983
187	N01B	0,085897
188	A10J	0,085659
189	J01A	0,084111
190	D04A	0,081468
191	A03C	0,075055
192	M03A	0,07441

4. ROBUSTNESS ANALYSIS

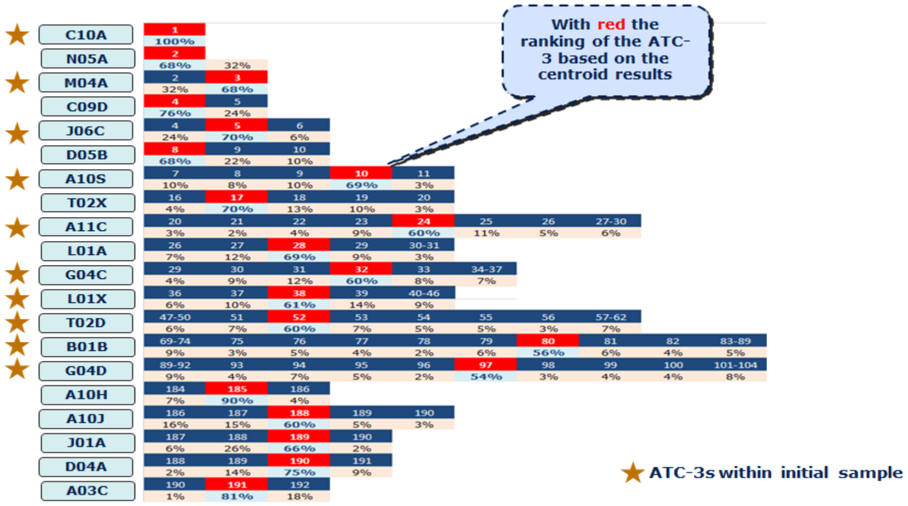
Having concluded to the ranking of the pharmaceutical categories, based on the preferences of the two DMs some questions arise concerning the robustness of our method. From the methodological point of view, the UTASTAR linear programming problem, described in paragraph 2.3 does not guarantee a single specific solution (Kadzinski et al., 2012). On the other hand, there exists an infinite number of evaluation parameters that are optimally consistent with the set of UTASTAR constraints that arises by the DMs' rankings. All these different parameters, although compatible with the DMs' rankings over the reference set of alternatives, may cause differentiations in the final ranking. Consequently, it must be estimated to what extent the overall hierarchy of the alternatives could be affected by random solutions of the UTASTAR linear problem.

Under this context, a random sampling algorithm was implemented to extract a big number of different random solutions of the UTASTAR linear programming problem. It is a weighting vector generating algorithm proposed by Tervonen et al. (2012) who adapted the "Hit and Run" sampling algorithm of Lovasz (1999) to multiple criteria decision analysis problems. This algorithm is forced to stop when the desired number of solutions to be generated is reached.

The weighting sets produced by the algorithm, led to the construction of 1,000 individual random rankings through the implementation of the additive value model. The statistical procession of these rankings is presented in Figure 5.

ISBN: 978-618-80361-1-6

Figure 5: Statistical analysis of the 1,000 rankings and positions obtained



The output of the analysis reveals that the position, each alternative holds based on the centroid, is also the position in the majority among the 1,000 different random solutions. In addition to this, it has been revealed that differentiations between different ranking positions are more frequent for alternatives belonging in the top or bottom of the ranking. On the other hand, greater alternations occur for the alternatives being ranked in the middle positions.

5. CONCLUSION

Through this paper a solution on the identification of the ideal therapeutic category for investment was given implementing the UTASTAR method based on the feedback given from two decision makers. The final output was a ranking of all therapeutic categories based on the revealed preferences of the two DMs providing a stable and robust solution for this problem and a powerful and stable tool in the decision making process. The ranking obtained based as defined by the DM's preferences can be easily adjusted to the preferences of different DMs and in that way different preferences can be incorporated in the model.

The next research steps include the expansion of the proposed approach to the Over the Counter and Personal Care market, the inclusion of more criteria as well as the involvement of more DMs on the procedure.

ACKNOWLEDGEMENT

The authors wish to acknowledge financial support from the Ministry of Education, Religious Affairs, Culture and Sports of Greece and the European Social Fund, under Grant THALIS, MIS 377350: Methodological approaches for studying robustness in multiple criteria decision making projects.

REFERENCES

[1] Athanasiadis, T., Maniatis, G., Demousis, F., 2012. *Pharmaceutical Market in Greece*, Foundation for Economic & Industrial Research, Athens, Greece.

[2] Bank of Piraeus, 2011. *Production and trading of pharmaceutical products*, Sectorial report, No. 5.

[3] EphMRA, 2012. *Anatomical Classification guidelines 2012*, Retrieved October 31, 2013, from http://www.ephmra.org/pdf/ATCGuidelines2012Final%20V2%20revised.pdf.

[4] Figueira, J., Greco, S., Ehrgott, M., Eds. 2005. *State-of-Art of Multiple Criteria Decision Analysis*, Kluwer Academic Publishers, Dortrecht.

[5] Jacquet-Lagrèze, E., Siskos, J., 1982. *Assessing a set of additive utility functions for multi-criteria decision making: The UTA method*, European Journal of Operational Research, Vol. 10, pp. 151-164.

[6] Jones, D., 2002. *Pharmaceutical Statistics*, Pharmaceutical Press, London, Great Britain.

[7] Kadzinski, M, Greco, S., Slowinski, R., 2012. *Selection of a representative value function in robust multiple criteria ranking and choice,* European Journal of Operational Research, Vol. 217, No. 3, pp. 541–553.

[8] Keeney, R.L., 1992. *Value-focused thinking: A path to creative decision making*, Harvard U.P., London.

[9] Keeney, R.L., Raiffa, H., 1976. *Decisions with multiple objectives: Preferences and value trade-offs*, John Wiley and Sons, New York.

[10] Lovαsz, L., 1999. *Hit-and-run mixes fast*, Mathematical Programming, 86, pp. 443–461.

[11] Maniadakis, N., 2010. *Study on the Generic products market and the legislation across Europe*, Travel Times Publishing, Athens, Greece.

[12] Roy, B., 1985. *Méthodologie multicritère d'aide à la décision*, Economica, Paris.

[13] Siskos, Y., Yanacopoulos D., 1985. *UTA STAR—an ordinal regression method for building additive value function*, Investigacao Operational, Vol. 5, pp. 39–53.

[14] Tervonen, T., Valkenhoef, B.N., Postmus, D., 2013. *Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis*, European Journal of Operational Research, Vol. 224, No. 3, pp. 552–559.

DATA SOURCES

IMS Health Hellas, Retrieved February 01, 2013, from http://www.imshealth.com/portal/site/imshealth. .

ISBN: 978-618-80361-1-6

Dealing with Robustness in Government Decision-Making using Facilitated Modelling

Nikos Tsotsolas *
*University of Piraeus **
80, Karaoli & Dimitriou str.,
*GR-18534 Piraeus, Greece **

Spyros Alexopoulos *
*University of Piraeus **
80, Karaoli & Dimitriou str.,
*GR-18534 Piraeus, Greece **

`

ABSTRACT

Policy decisions (e.g. economic, fiscal, development) are often complex and multifaceted and involve many different stakeholders with different objectives and priorities. Very often decision-makers, when confronted with such problems, attempt to use intuitive or heuristic approaches to simplify the complexity until the problem seems more manageable. In this process, important information may be lost, opposing points of view may be discarded, and elements of uncertainty may be ignored. A crucial issue, when dealing with political decisions, is the radical uncertainty about the present (e.g. lack or poor quality of information) and also about the future. The latter one addresses the seeming paradox - how can we be rational in taking decisions today if the most important fact that we know about future conditions is that they are unknowable? Robustness Analysis is a way of supporting government decision making when dealing with uncertainties and ignorance. In the present research we discuss the different definitions and approaches of Robustness Analysis in government decision-making concerning the present and the future as a way to support the identification of potential robust strategies in policy circles. We also initiate the discussion on how facilitated forms of MCDA could tackle different aspects associated with government decision making and provide effective support in dealing with robustness of strategic decisions in designing complex policies with long-term consequences.

KEYWORDS

Robustness Analysis, Government Decision-Making, Facilitated Modelling

1. INTRODUCTION

The government decision-making processes have always been a subject of philosophical enquiry and an open field for debating in OR community for the last 40 years. Nowadays, this specific topic is at the forefront of the discussion because of the crisis in Europe which emerges the urgent need for defining "efficient policy" concepts. The produced policies should be the rational and robust outcome of a collective decision-making process, based on a well-defined and sound framework of rules and methods.

Policy decisions (e.g. economic, fiscal, development) are often complex and multifaceted and involve many different stakeholders with different priorities or objectives. Furthermore, policy-making consists of several sequential actions focusing on the achievement of a specific goal with societal, economic and political implications, with several feedbacks and loops, so we may describe the whole process as a policy-circle (Lasswell, 1956). Very often decision-makers, when confronted with such problems, attempt to use heuristic and intuitive approaches to simplify the complexity until the problem seems more manageable. In this process, important information may be lost, opposing points of view may be discarded, and elements of uncertainty may be ignored. In short, there are many reasons to expect that during the evolution of a policy circle the involved stakeholders will often experience difficulty making informed, thoughtful choices in a complex decision-making environment involving value trade-offs and uncertainty (McDaniels et al, 1999). The problem is even more complex if we consider that, according to principles of sociotechnical design, specific - most of the times conflicting - objectives shall be best met by the joint optimization of technical and social aspects. Of course, the absence of certainty, the interference of political-power and the presence of complexity shall not be an excuse of inaction in this field.

The well-known area of Multi-criteria Decision Aiding (MCDA) (Roy, 2005; Montibeller and Franco, 2010) offers techniques designed to deal with situations, as the aforementioned, in which there are multiple conflicting goals for reaching strategic decisions. Furthermore, the process of creating, evaluating and implementing strategic political decisions is typically characterised by the consideration of potential synergies between different options, long term consequences, and the need of key stakeholders to engage in significant psychological and social negotiation about the strategic decision under consideration. MCDA can efficiently tackle all these issues. However, the political decision and negotiation process among members of the governmental committees does not take place in a political vacuum and political conflict is a reality. Thus, certain adaptations to the methods, tools and processes of MCDA are required if it is to be effectively applied in such a context (Tsoukias et al., 2013).

In the present research we initiate the discussion concerning the application of robustness analysis as a way to support the identification of potential robust strategies. We also discuss how facilitated forms of MCDA, where the model is created and analysed directly with a group of decision-makers in a decision conference, could tackle different aspects associated with government decision making and provide effective support in dealing with robustness of strategic decisions in designing complex policies with long-term consequences.

2. GOVERNMENT DECISION-MAKING

Political behaviour is argued to result from interacting political and information-processing mechanisms. The measures taken in order to meet a certain political goal can create conflict when simultaneously trying to achieve other goals. Thus, government policy makers seldom seek to maximize a single welfare objective; typically they are concerned about a bundle of policy objectives, expressed by contributing variables or indicators, conditional on and constrained by, applicable legislation (André García and Cardenete Flores, 2008). Another important characteristic of the mechanism of political actions is that politics is a game among forward-looking stakeholders (Lempert and Collins, 2007). As a result, the government's current payoffs are equal to the "net present value" of its anticipated future actions (and resulting victories/losses), not just its present and past policy. In this context actions of government policy makers can be interpreted as efforts to:

(a) design "efficient" policies (those for which every objective is reached with the minimum loss for the other relevant objectives) to improve government performance, as measured by well-defined indicators, while at the same time
(b) maintain a political behaviour true to their "political identity" (ideology, values, interests, influences).

Political mechanisms are intended to explain how the interests of the participants get balanced, compromised, suppressed, or are met to varying degrees. Equally important, these mechanisms are intended to account for whether and how the values and interests of the participants get transformed into the goals of the political unit as a whole. These political mechanisms, however, are only one part of a full account of the political decision-making process. Political actors are also problem solvers, who make decisions by exploring different choices, who plan and execute various actions, who make use of their knowledge and experience in the pursuit of their interests and goals (Sylvan *et al*, 1990). This is the area of political decision making we are studying in this work. Nevertheless, we keep in mind that even the most systematic operations of a government decision making are not implemented in a political vacuum, and this reality shall not be absent in the decision models which are proposed by the analysts.

There is an on-going discussion on whether the performance-based policies could reach optimum results. Performance-based government efforts aim to:

- clarify the mission and prioritize objectives with an emphasis on the expected results,

ISBN: 978-618-80361-1-6

- develop mechanisms for monitoring and reporting the achievement of those objectives, and
- use this information to make decisions about government activities, including making government more accountable.

A government is consider informed if it simultaneously presents informed options to decision-makers internally, and candid assessments of plans and performance externally, exploiting high quality information on context, activities and results which it collects and analyses (National Audit Office[10]).

The quantitative, decision-analytic framework of the Multi-criteria Decision Aid (MCDA) discipline, to be presented in more detail in Section 3, offers a wide pallet of techniques designed to deal with problems which face multiple, conflicting goals, such as the government policy-making objectives. The target of joint optimization of technical and social aspects indicates the need of a collective, participating sociotechnical approach informed by both MCDA and "facilitated modeling" (presented in more detail in Section 5), focusing not only on addressing challenges involved, but also on exploiting the adaptability and innovativeness of stakeholders in achieving goals instead of over-determining technically the matter in which these goals should be attained. It should also be noted that MCDA techniques are particularly appropriate for servicing the need of accountability of government, through the measurement of performance.

3. USING MCDA APPROACH

Modelling and analysis play a key role in the interventions between the discipline of Operational Research (OR) and strategic decision-making. Political Decisions are indeed Strategic Decisions because they share the same characteristics as described in Montibeller & Franco (2010):

- ❖ A strategic decision has been defined as one that is "important, in terms of the actions taken, the resources committed, or the precedents is sets"
- ❖ Strategic decisions are "infrequent decisions made by the top leaders of an organisation that critically affect organizational health and survival"
- ❖ The process of creating, evaluating and implementing strategic decisions is typically characterised by the consideration of high levels of uncertainty, potential synergies between different options, long term consequences, and the need of key stakeholders to engage in significant psychological and social negotiation about the strategic decision under consideration

In order to define the "main strategic choices" of an organization Richard (1983) takes into account the organization's mode of action and suggests a set of "strategic criteria" that permit an assessment of the organization's possibilities for survival and success, and verify the limitations of the economic system. These criteria can be apportioned into three groups or points of view, according to the organization's provisional horizon and the subsystem under study, namely:

- Competitiveness (analysis of the current external environment – a known field).
- Effectiveness (internal analysis of the company – a known field).
- Flexibility (analysis of the future external environment – unknown field that cannot be modelled).

The performance of an organization's (or equally of a government) should aim at improving each of the above three groups of criteria. In other words, an organization is engaged in a strategic path whenever it chooses to alter the balance of its available resources with the environment. Furthermore, according to the principles of sociotechnical design, organizational objectives are best met by the joint optimization of the technical and the social aspects (Cherns, 1976). For the complex, multidimensional process of strategic

[10] Option Appraisal: Making informed decisions in government, 2011, National Audit Office, www.nao.org.uk

ISBN: 978-618-80361-1-6

decision-making, David (2009) proposed a three-stage decision-making framework in which important multicriteria strategy-formulation techniques can be integrated.

Complex decision problems need a multicriteria decision analysis (MCDA) approach to be adopted, in order to take into account all the criteria/options involved in the analytical process of defining the scope of the decision, to construct a preference model, and to support the decision (see Roy, 1985; Roy and Bouyssou, 1993; Belton and Stewart, 2002; Figueira et al., 2005; Siskos, 2008). A collection of papers dealing with new trends in multicriteria analysis theory and practice was presented by Zopounidis and Pardalos (2010).

4. ROBUSTNESS ANALYSIS

An essential issue that shall be taken into consideration when implementing a political decision making process is the need for robustness analysis of the results of this process, given broad issues and multiple values being considered (Tsoukias et al, 2013). The stability of a model or/and of a solution should be assessed and evaluated each time so that the analyst shall be able to have a clear picture regarding the reliability and robustness of the produced results. Stability and reliability shall be expressed using measures which are understandable by the analyst and the decision maker and based on these measures the decision maker may accept or reject or adapt the proposed decision model. Given the fact that that uncertainty is present and has an influence on every decision-making context and that it appears in several different ways, it shall be neither omitted, nor relegated it. Its importance shall be realized and it shall be considered in an appropriate manner. As robustness allows us to experiment with uncertainty, it is necessary to define its concept, its significance and to emphasize its importance in the MCDA field.

Robustness analysis has achieved a remarkable importance in recent years. However, there is some confusion about the different meanings that the term robustness has received. For that reason it is necessary to consider the different notions behind the word "*robustness*" according to Vincke's approach (2003):

- Robust conclusion – valid in all or most pairs (version, procedure) – dealing with system values and gap from reality (Roy, 2010)
- Robust solution – good in all or most cases– dealing with uncertainty of external environment and external factors (Kouvelis and Yu, 1997)
- Robust decision in dynamic context – keep open as many good plans as possible for the future – dealing with the unknown future (Rosenhead *et al*, 1972; Rosenhead, 2003; Haasnoot *et al*, 2013)

The question is how somebody can tackle the aforementioned robustness issues. It is believed that specification and studying of robustness issues are often best achieved in some kind of an interactive mode with those who are faced with the need to decide. That is, the analysis is carried out by and under the control of the relevant policy group, with the assistance of one or more consultants (facilitators). Furthermore, robustness shall be expressed using measures which are understandable by the analyst and the decision maker. In this view it is suggested that visual tools may serve this need in a better way.

Robust strategic approaches in political decision-making are commonly expressed by trading some optimal performance for less sensitivity to assumptions, satisficing over a wide range of futures, and keeping options open. Relevant research suggests that this often adopted strategy is also usually identified as the most robust choice. Robust political strategies may be preferable to optimum strategies when the uncertainty is sufficiently deep and the set of alternative policy options is sufficiently rich (Lempert and Collins, 2007).

ISBN: 978-618-80361-1-6

5. FACILITATED MODELLING

Up to now the most usual way to conduct OR consultancy for strategic decision making support in organizations has been to adopt what is called the "expert mode", where the operational researcher uses operational research methods and models that permit an "objective" analysis of the client's problem situation, together with the recommendation of optimal solutions. The "expert mode" faces decision problems as real entities, thus the main task of the operational researcher is to represent the real problem that the client organization is dealing with, avoiding "biases" from different perspectives (Franco and Montibeller, 2010).

Yet, more often than not, problems are socially constructed, thus the operational researcher has to help a policy-makers team in negotiating a problem definition that can accommodate their different perspectives. This process is a participative one, in the sense that participants are able to:

- jointly define the situation, structure it, and agree in a focus,
- negotiate a shared problem definition by developing a model of organizational objectives,
- create, refine and evaluate a portfolio of options/priorities, and
- develop action plans for subsequent implementation.

In the cases of a policy analysis circle a participative operational research consultancy process for strategic decision aid support, other than "expert mode", needs to be applied. This process should incorporate the exploration of the notions of strategic decisions and the decision aid process, the examination of interconnectedness and long-term consequences as key characteristics of strategic decisions, and the consideration of the discursive nature of the processes within which strategic decisions are made. Such a process was proposed by Franco and Montibeller (2010) incorporating "facilitated decision modelling" (Eden, 1990; Phillips 2007). In facilitated modelling, a management team or a group of policy-makers is typically placed as responsible for scoping, analysing and solving the problem situation of interest. The operational researcher acts not only as an analyst, but also as a facilitator to this team. Participants' interaction with the model reshapes the analysis, and the model analysis reshapes the group discussion.

Facilitated modelling is used as an intervention tool, which requires the operational researcher to carry out the whole intervention jointly with the client, and enables the accommodation of multiple and differing positions, possible objectives and strategies among participants (Checkland, 1981; Eden and Ackermann, 2004; Rosenhead and Mingers, 2001; Williams, 2008). As a result, strategic problems frequently require the facilitated mode, due to their complex social nature and qualitative dimensions, their uniqueness, and the need to engage a management team in the decision making process (Ackermann and Eden, 2001; Friend and Hickling, 2005).

6. TOWARDS A HOLISTIC FRAMEWORK FOR GOVERNMENTAL DECISION MAKING

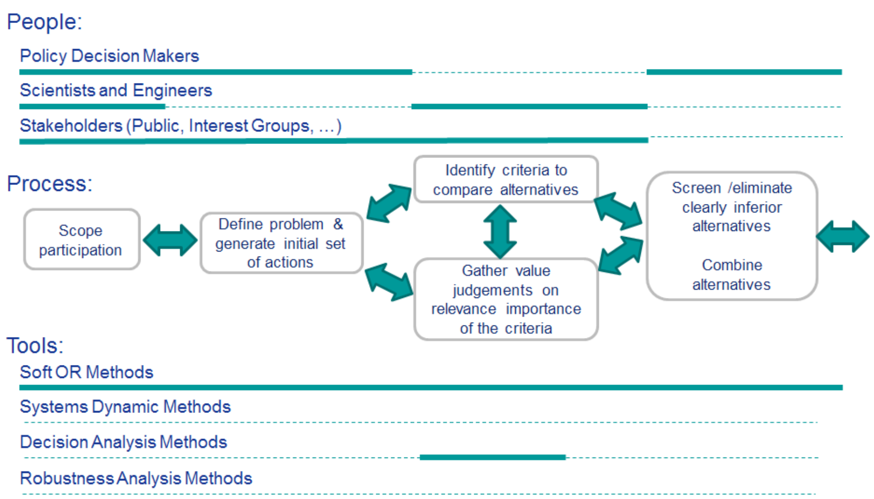
Governmental decision-making is a specific type of decision making which needs a holistic framework (tools and methodologies) that can successfully tackle its complex and multifaceted nature and at the same time to be transparent and easy to use by several stakeholders with diverse backgrounds. This framework shall be able to deal with technical complexity, which includes uncertainty (''epistemic'', policy values, political power) and decision complexity (inter-related choices, stakeholders' variety), as well as with social complexity, which includes social representations and communication channels.

Given the nature of the governmental decision-making we think that an adapted approach of a facilitated MCDA model, based on the ideas of Montibeller and Franco (2010), could efficiently support the following tasks: defining the problem, scoping participation, tackling uncertainty with future scenarios, considering multiple objectives, designing and appraising complex strategic options, and finally considering long term consequences. Special attention shall be paid on how to address issues of robustness at the different stages and mostly in scenario planning (Tsoukias et al, 2013). Furthermore, the framework shall describe in details the procedure of identifying, prioritizing and using multiple objectives (Keeney, 2013), as well as the procedure of choosing a multicriteria decision aiding method well adapted to each political decision context (Roy and Slowinski, 2013).

The types of facilitated modelling that we are including in our approach for political decision making (Figure 1a and Figure 1b) are the following:

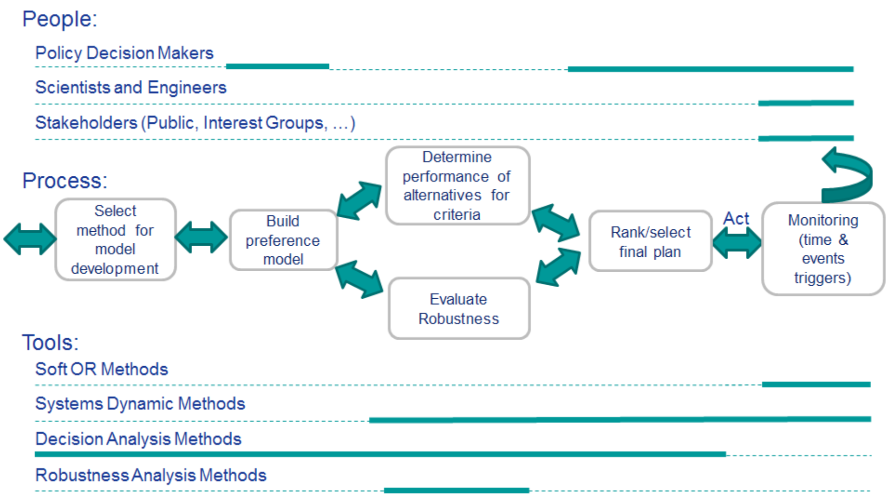
- Facilitated Problem Structuring: A set of modelling methods collectively known as 'Soft OR' methods (tools: Cognitive maps, DFDs, Strategic Options Development and Analysis, Soft Systems Methodology, Strategic Choice Approach, Organizational Knowledge Management Systems)
- Facilitated System Dynamics: Originated in the system dynamics field, it supports the modelling of systems where dynamics and feedback loops are important in understanding the impact of decision policies/options over time (tools: Workflow Diagrams, Spiral Method, Adaptive Policy making, Adaptation Pathways, Dynamic Adaptive Policy Pathways)
- Facilitated Decision Analysis: A set of methods that help modelling decisions that involve multiple objectives and/or uncertainty of outcomes.
- Facilitated Robustness Analysis: A set of tools supporting the comprehension of robustness and its handling

Figure 16a Facilitated Approach in Political Decision Making (1st part)



ISBN: 978-618-80361-1-6

Figure 17b Facilitated Approach in Political Decision Making (2nd part)



7. CONCLUSIONS

Is it a utopia to have informed government using scientific tools of OR trying to reach and to produce viable and robust strategic plans? For sure it is not a straightforward task (Andersson *et al*, 2013) but efforts are undertaken towards that direction. There are organisations such as GORS[11] (Government Operational Research Service in UK - supports policy-making, strategy and operations in many different departments and agencies and employs around 400 analysts) and RAND[12] Corporation (a non-profit institution that since 1948 helps improve policy and decision making through research and analysis, with approximately 1,700 people from more than 50 countries) which insist to use scientific tools in the political decision field. If a government intends to take decisions based on fairness, objectivity and thoroughness, then the proposed holistic approach has more strengths in relation to no scientific at all approaches such us TINA (stands for There is No Alternative, see Guardian, 4/7/2013[13]).

Concluding, we believe that given the importance of political strategic decision making for the survival of any political system, further developments in this field could, therefore, not only bring opportunities for research on the several challenges we highlighted here, but also have a real impact on MCDA practice. More studies on robustness of strategic options under multiple scenarios are required; for example, about suitable operators and graphical displays for interacting with policy makers. As far as the design of complex policies is concerned, structuring policies composed by options that are interconnected is an area almost unexplored, from a decision analysis perspective, and ideas from the field of problem structuring methods may be relevant for this intent. Furthermore, given the special nature of the political decision making, it would be interesting to assess the impacts of the framework we are suggesting and the effectiveness of strategy workshops, as well as the overall usefulness of the framework to increase our understanding of decision analytical support at the political strategic level.

[11] GORS: http://www.operational-research.gov.uk/recruitment
[12] RAND: http://www.rand.org/
[13] Guardian: http://www.theguardian.com/science/life-and-physics/2013/may/04/no-alternative-bayes-penalties-philosophy-thatcher-merkel

ISBN: 978-618-80361-1-6

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning".

REFERENCES

Ackermann F. and Eden C., 2001. Contrasting single user and networked group decision support systems for strategy making. *Group Decision and Negotiation*, Vol. 10, No. 1, pp. 47–66

Andersson A., Grönlund A. and Åström J., 2012. "You can't make this a science!"—Analyzing decision support systems in political contexts. *Government Information Quarterly*, Vol. 29, pp. 543–552

André García F.J. and Cardenete Flores M.A., 2008. Economic and environmental efficient policies in an applied general equilibrium framework. *Ekonomiaz: revista vasca de economía, Vitoria-Gasteiz: Servicio Central de Publ. del Gobierno Vasco*, Vol. 67, 1, pp. 72-91. (in Spanish)

Belton V. and Stewart T., 2002. *Multiple criteria decision analysis: An integrated approach*. Kluwer Academic Publishers, Dordrecht

Checkland P., 1981. *Systems Thinking, Systems Practice*. Wiley, Chichester

Cherns A., 1976. The Principles of sociotechnical design. *Human Relations*. Vol. 29, No. 8, pp. 783-792

David F.R., 2009. *Strategic Management: Concepts and Cases*, 12th edition. Prentice Hall, Upper Saddle River

Eden C., 1990. The unfolding nature of group decision support: Two dimensions of Skill. In: Eden C. and Radford J. (eds.), *Tackling Strategic Problems: The Role of Group Decision Support*. Sage, London, pp. pp.48–52

Eden C. and Ackermann F., 2004. Use of 'Soft OR' models by clients: What do they want from them? In: Pidd M., (Ed.) *Systems Modeling: Theory and Practice*. Wiley, Chichester, pp. 146–163

Figueira J., Greco S. and Ehrgott, M., 2005. *State-of-Art of Multiple Criteria Decision Analysis*. Kluwer Academic Publishers, Dortrecht

Franco L.A. and Montibeller G., 2010. Facilitated modeling in operational research. *European Journal of Operational Research*, Vol. 205, pp. 489-500

Friend J. and Hickling A., 2005. *Planning Under Pressure: The Strategic Choice Approach*. Third ed. Elsevier

Haasnoot M., Kwakkel J. H., Walker Warren E. and Maat J., 2013. Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, Vol. 23, No. 2, pp. 485-498

Keeney R.L., 2013. Identifying, prioritizing, and using multiple objectives. *EURO Journal on Decision Processes*, Vol. 1, No. 1-2, pp. 45-67

Kouvelis P. and Yu G., 1997. *Robust Discrete Optimisation and Its Applications*. Kluwer Academic Publishers, Netherlands

Lasswell H.D., 1956. *The decision process: seven categories of functional analysis*. University of Maryland Press, College Park

ISBN: 978-618-80361-1-6

Lempert R.J. and Collins M.T., 2007. Managing the Risk of Uncertain Threshold Responses: Comparison of Robust, Optimum, and Precautionary Approaches. *Risk Analysis-An official publication of the Society for Risk Analysis*, vol. 27, No. 4, pp. 1009-1026

McDaniels T.L., Gregory R.S. and Fields D., 1999. Democratizing risk management: Successful public involvement in local water management decisions. *Risk Analysis*, vol. 19, pp.497–510

Montibeller G. and Franco A., 2010. Multi-Criteria Decision Analysis for Strategic Decision Making. In: Zopounidis, C. and Pardalos, P.M. (eds) *Handbook of Multicriteria Analysis*. Springer New York, pp. 25-48

Phillips L., 2007. Decision conferencing. In: Edwards W., Miles R. Jr,. and von Winterfeldt D., (eds.), *Advances in Decision Analysis: From Foundations to Applications*. Cambridge University Press, New York, pp. 375–399

Richard J.L., 1983. Aide à la Décision Stratégique en PME. In: Jacquet-Lagrèze, E. et Siskos, J. (eds.), *Méthode de Décision Multicritère*. Hommes et Techniques, Paris, pp. 119–142

Rosenhead J., 2002. Robustness Analysis. *Newsletter of the European Working Group "Multiple Criteria Decision Aiding"*, Series 3, No. 6, pp. 6-10

Rosenhead J., Elton M. and Gupta, S.K., 1972. Robustness and optimality as criteria for strategic decisions. *Operational Research Quarterly*, Vol. 23, No. 4, pp. 413-430

Rosenhead J. and Mingers J., 2001. A new paradigm of analysis. In: Rosenhead J. and Mingers J. (eds.), *Rational Analysis for a Problematic World Revisited: Problem Structuring Methods for Complexity, Uncertainty, and Conflict*. Wiley, Chichester, pp. 1–19

Roy B., 1985. *Méthodologie Multicritère d'Aide à la Décision*. Paris : Economica

Roy B., 2005. Paradigms and challenges. In: Figueira J.R., Greco J.S. and Ehrgott S. (eds) *Multiple Criteria Decision Analysis - State of the Art Surveys*. Springer, pp. 3-24

Roy B., 2010. Robustness in operational research and decision aiding: A multi-faceted issue. *European Journal of Operational Research*, Vol. 200, pp. 629-638

Roy B. and Bouyssou, D., 1993. *Aide multicritère à la décision: Méthodes et cas*. Paris: Economica

Siskos Y., 2008. *Decision Models*. New Technologies Publications, Athens (in Greek)

Roy B. and Słowiński, R., 2013. Questions guiding the choice of a multicriteria decision aiding method. *EURO Journal on Decision Processes*, Vol. 1, No. 1-2, pp. 69-97

Sylvan D.A., Goel A. and Chandrasekaran B., 1990. Analyzing political decision making from an information-processing perspective: JESSE. *American Journal of Political Science*. Vol. 34, No. 1, pp. 74-123

Tsoukias A., Montibeller G., Lucertini G. and Belton V., 2013. Policy analytics: an agenda for research and practice. *EURO Journal on Decision Processes*, Vol. 1, No. 1-2, pp. 115-134

Vincke P., 2003. About Robustness Analysis. *Newsletter of the European Working Group "Multiple Criteria Decision Aiding"*, Series 3, No. 8, pp. 7-9

Williams T., 2008. *Management Science in Practice*. Wiley, Chichester

Zopounidis C. and Pardalos P.M., 2010. *Handbook of Multicriteria Analysis*. Springer, New York

ISBN: 978-618-80361-1-6

Customer satisfaction performance and importance judgments: An application of the MUSA+ model

Ifigeneia Pologiorgi
Technical University of Crete
University Campus,,
GR73100 Chania, Greece

Evangelos Grigoroudis
Technical University of Crete
University Campus, GR73100
Chania, Greece

Stelios Tsafarakis
Technical University of Crete
University Campus, GR73100
Chania, Greece

George Baltas
Athens University of
Economics and Business
76 Patission Av., GR10434
Athens, Greece

Abstract

Customer satisfaction surveys often include questions as to how consumers weight various satisfaction criteria. The weight of a satisfaction criterion which is evaluated directly through such importance questions is termed stated importance. On the other hand, derived importance is inferred by some regression-type quantitative technique. The comparison between derived and stated importance can categorize satisfaction criteria using the attractive quality approach (Kano's model). In particular, such an approach may help business organizations to identify truly important and truly unimportant, unspoken motivators or even expected or cost of entry attributes. The aim of this study is to analyze customer satisfaction in the supermarket sector combing satisfaction performance and importance data. The analysis is based on the MUSA+ method which is a multicriteria preference disaggregation approach for customer satisfaction benchmarking analysis. The applied methodology consists of the following parts: (1) customer satisfaction analysis, which concerns the identification of customer preferences and refers to the estimation of the relative importance and the demanding level of the different satisfaction dimensions, (2) the importance analysis using a weight ordering regression technique (WORT method), (3) the satisfaction benchmarking analysis, which is mainly focused on the comparative performance evaluation of all examined companies in the sector, and (4) the categorization of the customer satisfaction criteria in the dimensions of the Kano model, using dual importance diagrams. The results demonstrate how businesses can use the model to locate their position against competition, pinpoint their weak points, and determine which characteristics will improve their global performance.

KEYWORDS

Multiple Criteria analysis; MUSA Method, WORT Method, Ordinal Regression, Customer Satisfaction; Benchmarking Analysis

1. INTRODUCTION

Within the retail industry, the nature of competition among supermarkets is a major area of research. The grocery retail market in many western countries is dominated by a few supermarket groups primarily differing in pricing policy, assortment and service levels, making the sector highly competitive and extremely concentrated (Solgaard and Hansen 2003). The competition between the various actors in the supermarket market in most western countries is very strong as consumers have a wide range of choice alternatives to cover their needs. In the UK alone, consumer choices vary a lot and can range from large mainstream supermarkets (e.g. Tesco, Sainsbury's, Morrisons, ASDA), to grocers focusing on the quality niche market (e.g. Waitrose, Marks & Spencer) and discounters such as Aldi, Netto, and Lidl (Corporate Watch 2001). Therefore, in order to survive in the long-run and build strong and lasting relationships with their customers, supermarkets need to know the key criteria for store choice amongst food shoppers.

The objective of the present exploratory study is to analyze customer satisfaction in the supermarket sector combing satisfaction performance and importance data using the MUSA+ method (Grigoroudis and Siskos 2010). The proposed model is appropriate for the purpose of this research as it allows the researcher to experiment with a large number of attributes and at the same time limit the workload of each research subject to a reasonable level.

This study develops a MUSA+ model for the estimation of the relative impact of 9 supermarket attributes on the actual supermarket choice. Increasingly, MUSA analysts are being asked to design and analyze studies involving large numbers of attributes and/or attribute levels. The application of the MUSA+ model to a Greek setting tries to shed some light on two key questions: what are the basic determinants that drive supermarket patronage, and what is the relative impact of each of them on the actual choice?

2. STORE CHOICE LITERATURE

In recent years, considerable attention has been paid to store attributes that consumers may evaluate while making a store choice decision. The main goal of store choice literature is to understand the consumer's choice process between alternative stores. All this literature studies the key store attributes that a customer takes into account while evaluating alternative stores. Much research has been conducted in order to identify the most important store attributes in retail patronage. Store choice literature suggests that generally there are seven key elements in the retail brand that can influence the choice between alternative stores and can be potentially significant drivers for the consumer's store choice (Corstjens and Corstjens 1995; Aaker 1996; Levy and Weitz 1998), namely, the physical store (e.g. store design: furniture, store layout, etc.), the services offered (e.g. the personnel, their friendliness, the atmosphere, the cleanliness, etc.), the fresh food products (e.g. quality, variety, etc.), the own-label products (e.g. variety, quality, price, etc.), the communication strategy (e.g. promotional activities, database marketing, etc.), the price positioning of the store (EDLP vs. HILO stores), and the access attributes (e.g. location, distance to the store, parking, traffic, etc.).

Other studies have examined the role of store attributes such as environment and atmosphere (e.g. Baker et al. 1994; Donovan et al. 1994), product assortment (Grewal et al. 1999), price format (Bell and Lattin 1998; Colla 2003), store brands (Burt 2000; Ailawadi 2001), customer services (Sparks 1995), salespersons (Darian et al. 2005), store promotions (Volle 2001) and other situational factors (Ryans 1977; Heeler et al. 1979; Mattson 1982).

Taking into account the aforementioned quality features, a set of satisfaction criteria can be assessed reflecting all the aspects of user perceptions about the supermarket quality. The set of satisfaction criteria used in this study include: 1. Prices, 2. Product variety, 3. Product quality, 4. Layout, 5. Easy access, 6. Offers, 7. Service, 8. Atmosphere, 9. Reputation.

3. BACKGROUND

3.1 The MUSA+ method

Customer satisfaction surveys often include collection of global and partial satisfaction judgements (for a set of service criteria/attributes/characteristics). Benchmarking analysis requires this qualitative data set to be expanded in order to include satisfaction judgement for a set of competitive business organizations. The MUSA+ method, applied in this research, is an extension of the original MUSA (MUlticriteria Satisfaction Analysis) method proposed by Grigoroudis and Siskos (2002) for the case of customer benchmarking analysis.

Grigoroudis et al. (2008) present an application of a first version of MUSA+ method to the transportation-telecommunications sector. Also an analytical presentation of the original MUSA method is given by Grigoroudis and Siskos (2002, 2010), while several applications of the ordinal regression approach to the customer satisfaction evaluation problem can be found in the literature.

Similarly to the original MUSA method, the results provided by MUSA+ include value function (global and partial), criteria weights, average satisfaction and demanding indices, as well as action and improvement diagrams. In addition, MUSA+ provides a comparative performance diagrams (average satisfaction indices of a particular company in relation to the performance of the other competitive companies).

3.2 Kano's model of customer satisfaction

Kano's model classifies the quality attributes into different quality dimensions (Kano et al., 1984):

1. *Must-be quality*: These quality attributes are taken for granted when fulfilled but result in dissatisfaction when not fulfilled. The customer expects these attributes, and thus views them as basics. Customers are unlikely to tell the company about them when asked about quality attributes; rather they assume that companies understand these fundamentals of product design.
2. *One-dimensional quality*: These attributes result in satisfaction when fulfilled and result in dissatisfaction when not fulfilled. They are also referred as the-more-the-better quality attributes. The one-dimensional attributes are usually spoken and they are those with which companies compete.

3. *Attractive quality*: These quality attributes provide satisfaction when fully achieved but do not cause dissatisfaction when not fulfilled. They are not normally expected by customer, and thus they may be described as surprise and delight attributes. For this reason, these quality attributes are often left unspoken by customers.

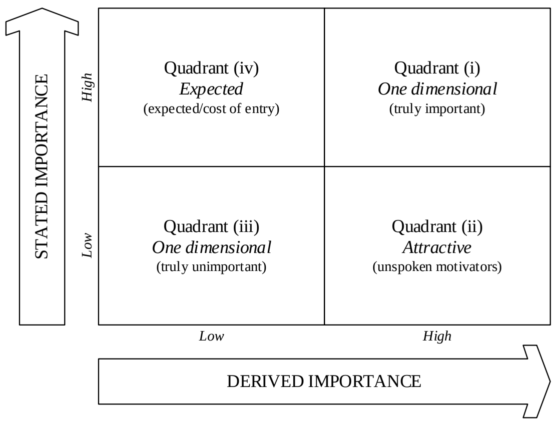
The most common approach in order to classify quality attributes into the five aforementioned dimensions is to use a specific questionnaire that contains pairs of customer requirement questions for every product attribute (i.e., ask customers how do they feel if a given feature is present or not present in the product).

3.3 Dual importance grid

Although they are several alternative approach for classifying quality attribute in the major Kano's dimensions, in the presented study we use the comparison between derived and stated importance. The straightforward customer preference for the weight of a satisfaction criterion which is evaluated through importance questions is defined as stated importance. Derived importance is estimated by a regression-type quantitative technique using customer judgments for the performance of this set of criteria.

The comparison between derived and stated importance can give valuable information. It enables a company to identify what attributes the customers rate as important and see how these agree with truly important and truly unimportant attributes. Moreover, it helps the company identify unspoken motivators or even expected or cost of entry attributes. The Dual Importance Grid is divided in four quadrants (see Figure 1).

Figure 1 Dual importance diagram



Quadrants (i) and (ii) include the dimensions that are truly important to the customers. These are the main characteristics that management and production should focus on. Quadrants (i) and (iv) include the important dimensions according to the customers' free statement. These are the dimensions that marketing should focus on. When a characteristic appears in quadrant (i) or (iii) there is an agreement between derived and stated importance. On the other hand, in quadrants (ii) or (iv) there is a disagreement between the stated and derived importance. This disagreement is an indication that these dimensions require further analysis. The dual importance diagram may be linked with the Kano's model and its three basic categories of product/service requirements.

4. RESULTS

4.1 Satisfaction and benchmarking analysis

The presented survey was conducted in a sample of 2101 customers of different Greek supermarkets. Customers were asked to evaluate their satisfaction level on each one of these criteria, as well as to express their overall judgement using a 5-point qualitative scale of the form: very satisfied, satisfied, moderately satisfied, dissatisfied, very dissatisfied. The set of companies examined in this study consists of the 7 major supermarket providers in Greece, having a total market share greater than 95%. It should be emphasized that

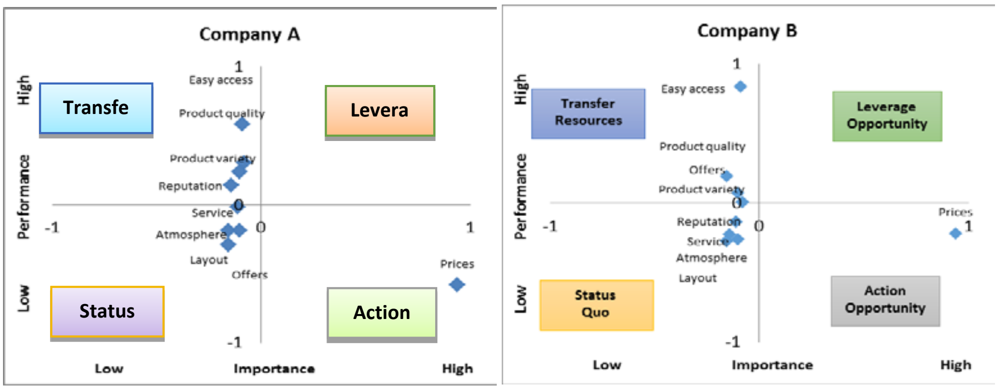
the supermarket sector in Greece is highly competitive, and there are no significant differences for these three companies concerning products and services offered, prices, etc.
According to Table 1, all the criteria seem to be of almost equal importance for the customers of supermarket providers in Greece. The "Prices", the "Product Quality" and the "Easy Access" of the supermarkets are, slightly, the three most important criteria.

Table 1 Criteria weights and average satisfaction and demanding indices

Criteria	Weights (%)	Satisfaction Indices for Company A (%)	Satisfaction Indices for Company B (%)	Demanding Indices (%)
Prices	12.32	88	87	-0.578
Product variety	10.98	92	89	-0.598
Product quality	11.01	92	88	-0.599
Layout	10.91	89	87	-0.594
Easy access	11.00	94	93	-0.600
Offers	10.91	90	89	-0.594
Service	10.97	91	88	-0.593
Atmosphere	10.98	90	87	-0.597
Image	10.93	91	87	-0.592
Overall satisfaction		92	90	-0.735

Performance analysis is mainly focused on the assessment of supermarket quality level for each one of the competitive companies. Furthermore, it includes, for each supermarket, the identification of their competitive advantages, their weak points, as well as the improvement priorities. The relative action diagrams for the major supermarket providers in Greece are shown in Figure 2. As it can be observed, there are no significant differences in these relative maps, indicating homogeneity of user perceptions about these supermarkets.

Figure 2 Action diagrams for the major supermarket providers



The most important findings are summarized as follows:

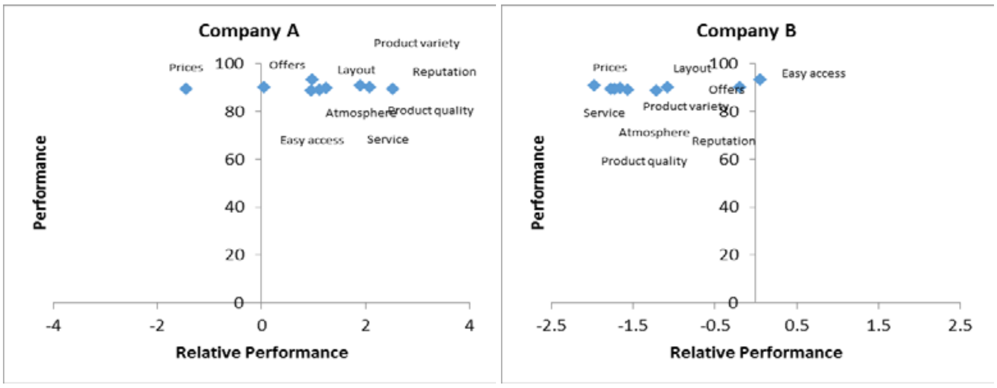
- Company A: The only quality attribute located at the action opportunity quadrant, i.e. the quadrant with low performance and high importance, is the criterion "Prices". This essentially indicates that this is an important – critical criterion that should be certainly improved so that the satisfaction level increases. There is no quality attribute located at the leverage opportunity quadrant, i.e. the quadrant with high performance and high importance. This means that no quality attribute can be used as competitive advantage against competition, since quality attributes that are located at this quadrant have especially high performance while they also considered as important from the customers. Such attributes are the basic reason and difference for choosing a particular supermarket. The criteria "Service", "Atmosphere", "Layout" and "Offers" are located at the status quo quadrant, i.e. the quadrant with low performance and low importance. Usually there is no need for further actions from the company's side, given that these particular criteria are considered as not important from customers. The rest of the criteria ("Easy access", "Product quality", "Product variety" and "Reputation") are located at the transfer resources quadrant, i.e. the quadrant with low importance and high performance. The resources and generally the company's attempt that are related with these attributes can be used in a different way.
- Company B: Similarly, the only quality attribute located at the action opportunity quadrant is the criterion "Prices". Also, there is no quality attribute located at the leverage opportunity quadrant. The criteria

"Service", "Atmosphere", "Layout" and "Reputation" are located at the status quo quadrant, while the rest of the criteria ("Easy access", "Product quality", "Product variety" and "Offers") are located at the transfer resources quadrant.

The satisfaction benchmarking analysis is mainly focused on the performance evaluation of the competitive organizations against the satisfaction criteria, as well as the identification of the competitive advantages of each company. The analysis is based on the comparative performance diagrams produced by the MUSA+ method; these diagrams can help each company to locate its position against the competition, to pinpoint its weak points and to determine which criteria will improve its global performance.

As shown in Figure 3, a large number of satisfaction criteria appear as competitive advantages for company A, while company B shows relatively low performance against the competition. None of the companies has quality attributes located in the "waiting" quadrant (this quadrant refers to the companies' weaknesses, which have performance higher than the competition). Improvement efforts should be focused on "Prices" of company A, as it is a characteristic with average satisfaction index lower than the competitors and the customers regard it as an important criterion. All the other satisfaction criteria may be considered as competitive advantages.

Figure 3 Comparative performance diagrams for the major supermarket providers



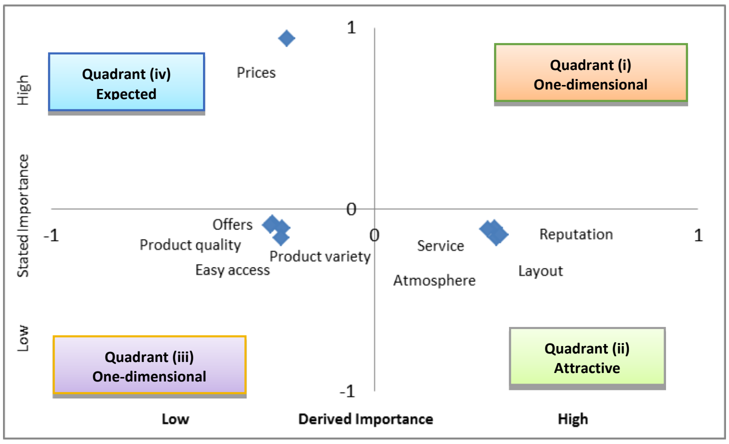
Regarding company B, direct improvement efforts should be focused on all the criteria except "Easy access". The rest of the criteria, are located at the "struggle' quadrant meaning that the performance of the characteristics is rather high but worse than competition. Hence, it seems that Company B has significant difficulties to follow other competitors.

4.2 Importance analysis

Customer satisfaction criteria has been categorized in the dimensions of the Kano model using the dual importance diagram. In this study, derived importance has been estimated using the MUSA+ method, while the WORT method (Grigoroudis and Spiridaki 2003) has been applied in order to asses stated importance.
The dual importance grid is shown in Figure 4, where the following can be observed:

a. The criteria "Product variety", "Product quality", "Easy access" and "Offers" are located in the Quadrant III. They are one-dimensional attributes and reflect the major competitive features of the provided services.
b. The criteria "Layout", "Service", "Atmosphere" and "Reputation" are located in the Quadrant II, where the attractive attributes correspond. This means that it is possible a high performance level of these attributes can lead to high satisfaction, while a low performance level will not imply low satisfaction index. They are called unspoken motivators and companies may take advance of these features in order to differentiate from competition.
c. The criterion "Prices" is located in Quadrant IV, so this attribute is characterized as expected according to the Kano method.

Figure 4 Dual importance grid



5. CONCLUSIONS

The main objective of this paper is the implementation of ordinal regression models for satisfaction benchmarking analysis. The main advantage of the applied approach is that it fully considers the qualitative form of customer judgements and preferences, as expressed in a customer satisfaction survey. Furthermore, MUSA+ is able to assess an integrated benchmarking system, given the wide range of results provided (e.g. value functions, criteria weights, average satisfaction indices, average demanding indices, action, improvement, and comparative performance diagrams, etc.). Thus, discussion is focused not only on the descriptive analysis of customer satisfaction data, but it is able to give emphasis on the analysis of customer preferences and expectations. The analysis indicates the improvement efforts that supermarket providers may consider for the development of their stores.

On the other hand, customer satisfaction benchmarking analysis is a useful tool for modern business organizations in order to locate their position against competition. This gives the ability to identify the most critical improvement actions and adopt the best practices of the industry.

Furthermore, the theory of attractive quality may give a valuable explanation about the relationship between the degree of sufficiency of a quality attribute and the customer satisfaction with that attribute. Based on this approach, it can be recognized that customer satisfaction is more than a one-level issue as traditionally examined. Moreover, it may not be enough to merely satisfy customers by meeting their basic and spoken requirements, particularly in a highly competitive environment.

ACKNOWLEDGEMENT

This work was partially supported by the Ministry of Education, Religious Affairs, Culture and Sports of Greece and the European Social Fund, under the Grant THALIS, MIS 377350 “Methodological approaches for studying robustness in multiple criteria decision making projects” for the second in order author.

REFERENCES

Aaker D.A., 1996. *Building store brands,* The Free Press, New York.

Ailawadi K.L., 2001. The retail power-performance conundrum: what we have learned, *Journal of Retailing*, Vol. 77, No. 3, pp. 299-318.

Baker J., Grewal D., Parasuraman A., 1994. The influence of store environment on quality inferences and store image, *Journal of the Academy of Marketing Science*, Vol. 22, No. 4, pp. 328-339.

Bell D., Lattin J., 1998. Shopping behavior and consumer preference for store price format: Why ‘large basket’ shoppers prefer EDLP, *Marketing Science*, Vol. 17, No. 1, pp.66-88.

Burt S., 2000. The strategic role of retail brands in British grocery retailing, *European Journal of Marketing*, Vol. 34, No. 8, pp. 875-890.

Colla E., 2003. International expansion and strategies of discount grocery retailers: The winning models, *International Journal of Retail and Distribution Management*, Vo. 31, No. 1, pp. 55-66.

Corstjens J., Corstjens M., 1995. *Store wars: The battle for mindspace and shelfspace*. John Wiley & Sons, New York.

Darian J., Wiman A., Tucci L., 2005. Retail patronage intentions: The relative importance of perceived prices and salesperson service attributes, *Journal of Retailing and Consumer Services*, Vol. 12, No. 1, pp. 15-23.

Donovan R., Rossister J., Marcoolyn G., Nesdale A., 1994. Store atmosphere and purchasing behavior, *Journal of Retailing*, Vol. 70, No. 3, pp. 283-294.

Grewal D., Levy M., Mehrotra A., Sharma A., 1999. Planning merchandising decisions to account for regional and product assortment differences, *Journal of Retailing*, Vol. 75, No. 3, pp. 405-424.

Grigoroudis E., Litos C., Moustakis V., Politis Y., Tsironis L., 2008. The assessment of user-perceived web quality: Application of a satisfaction benchmarking approach, *European Journal of Operational Research*, Vol. 187, No. 3, pp. 1346-1357.

Grigoroudis E., Siskos Y., 2002. Preference disaggregation for measuring and analyzing customer satisfaction: The MUSA method, *European Journal of Operational Research*, Vol. 143, No. 1, pp. 148-170.

Grigoroudis E., Siskos Y., 2010. *Customer satisfaction evaluation: Methods for measuring and implementing service quality*, Springer, New York.

Grigoroudis E., Spiridaki O., 2003. Derived vs. stated importance in customer satisfaction surveys, *Operational Research: An International Journal*, Vol. 3, No. 3, pp. 229-247.

Heeler R., Francis J., Okechuku C., Reid S., 1979. Gift versus personal use brand selection, *Advances in Consumer Research*, Vol. 6, No. 1, pp. 325-328.

Kano N., Seraku N., Takahashi F., Tsjui S., 1984. Attractive quality and must-be quality, *Hinshitsu*, Vol. 14, No. 2, pp. 147-56.

Levy M., Weitz B., 1998. *Retailing management*, 3rd edition, Irwin/McGraw-Hill, New York.

Mattson B., 1982. Situational influences on store choice, *Journal of Retailing*, Vol. 58, No. 3, pp. 46-58.

Ryans A., 1977. Consumer gift buying: an exploratory analysis, in: B.A. Greenberg, D.N. Bellenger (eds.), *Contemporary marketing thought*, American Marketing Association, Chicago, pp. 99-115.

Solgaard H., Hansen T., 2003. A hierarchical Bayes model of choice between supermarket formats, *Journal of Retailing and Consumer Services*, Vol. 10, No. 3, pp. 169-180.

Sparks L., 1995. Customer service in retailing, in: G. Akehurst, A. Nicholas (eds.), *Retail marketing*, Frank Call, London, pp. 76-95

Volle P., 2001. The short-term effect of store-level promotions on store choice, and the moderating role of individual variables, Journal of Business Research, Vol. 53, No. 2, pp. 63-73.

Ranking risks of maritime activities with multicriteria decision aid: Application to a ship-to ship transfer operation

Dimitrios I. Stavrou
National Technical University of Athens,
9 Iroon Polytechniou str.
Zografou 15773, Greece

Nikolaos P. Ventikos
National Technical University of Athens,
9 Iroon Polytechniou str.
Zografou 15773, Greece

Yannis Siskos
University of Piraeus
80 Karaoli and Dimitriou str,
Piraeus, 185 34, Greece

Abstract

A feasible way to tackle the risks of maritime activities is through the use of multicriteria decision aid (MCDA) methodologies. The aim of this paper is to develop a multicriteria model to assess and evaluate different risks of a maritime operation based on the outranking relation concept. The proposed methodology uses confidence indices to compare alternative actions with respect to a predefined consistent family of criteria. In addition, fuzzy domination relation is applied to complete the ranking of the risks. The effectiveness of the suggested model is applied on a ship-to-ship (STS) transfer operation, which refers to the transfer of cargo between seagoing ships. Different risk scenarios are evaluated from a team of experts with relative experience in STS operations, in terms of three risk factors: the likelihood of a failure occurring, the severity of the failure, and the ability to detect the failure on time. In addition, a robustness analysis is applied based on the lack of precision of criteria importance weights.

KEYWORDS

Maritime activities, Risk assessment; Multicriteria analysis; Ship-to-ship (STS) transfer.

1. INTRODUCTION

Maritime activities suffer inherently from risks that can compromise the success of the operation. The origins of these risks can be located in several factors, to name a few: the need to deal with the complex and many times hostile sea environment where the operation is conducted; the human element due to the unavoidable possibility of error during the procedures; the potential failure of the means to conduct the operation, including the vessels or the additional equipment during the operation. The risks can compromise the success of the marine operation and may lead to an accident with adverse effects on human lives or the environment in combination with the loss of property. There are many definitions of risk in maritime operations. An analytical discussion is presented by Goerlandt and Montewka (2015), as well as, from Aven (2012). With regard to the latter risk definitions, two basic features can be addressed: (1) their relation to the factors of occurrence and the consequence of a failure and, (2) the dependency of risk on uncertainty. An interesting conclusion in our previous study (Stavrou and Ventikos; 2015) highlights the importance of another factor, named detectability. Detectability is a new risk factor that refers to the ability to detect a risk on time to prevent it from happening. On the other hand, there are two fundamental types of uncertainty: aleatory and epistemic. Aleatory uncertainty is related to the inherent variability of the system under consideration. Marine systems have a stochastic nature that does not permit deterministic predictions of its behavior. This type of uncertainty is difficult to be reduced. Epistemic uncertainty, on the other hand, refers to the lack of knowledge about system parameters and can be found in errors or limitations during the process of collecting data. This lack of knowledge comes from several sources. Inadequate understanding of the underlying processes, incomplete knowledge of the phenomena, or imprecise evaluation of the related characteristics are common sources of epistemic uncertainty. Epistemic uncertainty refers to the limited knowledge we may have about the system. This type of uncertainty is tolerable. More observations or experiences mean more information, which means less level of uncertainty (Compton et al. 2009). From the definition of epistemic uncertainty, it can be inferred that the risk factors inherently have an element of uncertainty themselves. A feasible way to tackle uncertainty is through the use of multicriteria decision aid (MCDA) methodologies under uncertainty. There are several stochastic MCDA methods to deal with uncertainty; to name a few, Keeney and Raiffa (1976) introduced the multiattribute utility theory (MAUT) under uncertainty; Martel and D'Avignon (1982) proposed the use of confidence indices; D'Avignon and Vincke (1988) suggested a method to compare alternatives with preference indices, and Zaras and Martel (1994) addressed the stochastic dominance rules, which consist of pairwise comparisons of the actions and ranking of the actions according to the determined stochastic dominance relation; Fan et al. (2010) also proposed a method based on pairwise comparisons of alternatives with random evaluations using Probability Theory. This paper focuses on a stochastic approach, based on the preference relation among the actions with regard to the selected criteria. The outranking relation is defined as a binary relation of the alternative actions as a result of the pairwise comparison of the actions. Hence, the proposed methodology uses confidence indices to compare alternative actions with respect to a predefined consistent family of criteria. Moreover, the concept of fuzzy domination relation, borrowed from Siskos and Hubert (1983), is applied to complete the ranking of the risks. Lastly, the relative importance (weights) of the selected criteria is determined by means of applying a mathematical programming process.

The study of the risk factors can reduce uncertainty and lead to safe conclusions regarding the evaluation of the risks of maritime activities. It should be conducted in a methodological and analytical way. To do so, risk scenarios are composed after taking into account several guidelines, recommendations and other studies relevant to the activity (Skjong, 2007; IMO, 2010; OCIMF, 2013*)*. Risk scenarios are narrated descriptions of causes and effects from different circumstances that may be combined and lead to an unfortunate event called an 'accident'. Risk scenarios have been successfully applied to solve MCDA problems in previous studies (Islei et al. 1999; Schoemaker, 1995; Van der Heijden, 1996; Durbach I.N, 2014) because of their ability to give to the decision-makers a good understanding of the problem at hand and to unusual insights into possible courses of action (Durbach & Stewart, 2012). To control the stability of the model an illustrative example is also proposed. In particular, a ship-to-ship transfer operation, which refers to the transfer of oil cargo between seagoing ships positioned alongside each other, is thoroughly analyzed. Several risk scenarios regarding the operation are composed and evaluated in terms of three different risk factors: the likelihood of a failure occurring, the severity of the failure, and the ability to detect the failure on time. The risk scenarios are assessed from a team of experts with relevant experience in such operations. In addition, a robustness analysis is performed to control the stability of the proposed model.

The aim of this paper is to develop a model to assess and evaluate different risks of a maritime operation based on the outranking relation among the risk factors that affect the activity. To do so, the paper is structured as follows: Section 2 presents the general steps of the implemented methodology; Section 3 refers to the implementation of the proposed methodology by applying an illustrative example of an STS transfer operation; and in Section 4 the results are presented followed by a brief discussion of the results. Section 5 concludes the paper.

2. MODEL THEORETICAL BACKGROUND

In MCDA modelling under uncertainty, a set of actions $A=\{a_1,a_2,\dots,a_m\}$ is evaluated based on a consistent family of criteria $F=\{g_1,g_2,\dots g_j\dots,g_n\}$ under the assumption that each criterion is represented from a probability distribution δ_i^j. The discrete distribution is expressed from the following equation:

$$\sum_j \delta_i^a \cdot (g_i^j)=1 \tag{1}$$

Following the method of Martel and d' Avignon (1982), this paper proposes a model based on the preference ranking of a set of actions using two indices called Confidence Index and Doubt Index. More specifically, the model consist of four different steps:

Step 1: For each criterion g_i and every pair of actions $(a,b)\in A\times A$ the Confidence and Doubt Indices are calculated. The **confidence index** shows the degree of credibility of the outranking of *b* by *a* on each criterion. It refers to the relative frequency of evaluations for the criterion g_i^j that is at least as high for the action *a* as it is for the action *b*. The range of values of the degree of credibility is between 0 and 1; value 1 is when the lowest evaluation of action *a* on the criterion g_i^j is higher or equal to the highest evaluation of action *b* to the same criterion g_i^j.

$$d_i(a,b)=\sum_j \delta_i^b(g_i^j)\sum_{k\geq j}\delta_i^a(g_i^k) \tag{2}$$

The **doubt index** expresses the degree on average scale in which the experts evaluate action a higher than action b with regard to the criterion g_i^j. The doubt index can take values analogous to the range of the scale of each criterion.

$$D_i(a,b) = \frac{1}{g_i^* - g_{i*}} \sum_j \delta_i^a(g_i^j) \sum_{k>j} (g_i^k - g_i^j) \delta_i^b(g_i^k) \quad (3)$$

Step 2: The relative importance (weight) of the criteria k_1, k_2, …, k_n is determined by the use of mathematical programming. The aggregation of the weights should be equal to one.

Following, the total confidence index is calculated by the use of the weight of each criterion:

$$C(a,b) = \sum_{i=1}^{n} k_i d_i(a,b) \quad (4)$$

The total confidence index $C(a, b)$ expresses the normalized concordance of the criteria with regard to the outranking of action a against action b.

Step 3: The degree of global credibility of the outranking is calculated for the pairs of actions using the combination of the confidence and doubt index:

$$d(a,b) = \begin{cases} C(a,b) & \text{when } C(a,b) \geq D_i(a,b) \quad \forall i = 1, 2, \ldots, n \\ \dfrac{C(a,b)}{1 - C(a,b)} \prod_{i^*} [1 - D_{i^*}(a,b)] & i^* \in \{ i / D_i(a,b) > C(a,b) \} \end{cases} \quad (5)$$

Step 4: According to the matrix of the degrees of global credibility of the outranking, the fuzzy domination relation (Siskos and Hubert,1983; Siskos et al. 1984) is applied to compare the different pairs of actions to determine the outranking intensity between each pair of actions:

$$d^D(a,b) = \begin{cases} d(a,b) - d(b,a) & \text{when } d(a,b) \geq d(b,a) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

From equation 6 it can be inferred that for an action $b \in A$ the function $d^D(a, b)$ is the fuzzy set of actions $a \in A$ that are dominated by action b. Following, the fuzzy non-domination relation can be determined:

$$d^{ND}(a,b) = 1 - d^D(a,b) \quad (7)$$

Similarly, for an action $b \in A$, the $d^{ND}(b, a)$ is the fuzzy set of actions $a \in A$ that they are not dominated by action b. Thus, the cross of the fuzzy sets for the actions $b \in A$ will result in a set of actions that they are not dominated by any other action of A. The latter set of actions is named the non-dominated set of actions μ^{ND}: $A \rightarrow [0, 1]$ and is determined as follows:

$$\mu^{ND}(a) = \min_{b \in A} d^{ND}(b,a) = \min_{b \in A} [1 - d^D(b,a)] = 1 - \max_{b \in A} d^D(b,a) = 1 - \max_{b \in A} \left[d(b,a) - d(a,d) \right] \quad (8)$$

where $\mu^{ND}(a)$ refers to the non-domination degree of action a by the other actions of set A. The final ranking of the actions according to problematic-γ, in which the actions are placed in order from the best to the worst, can be accomplished by the use of the fuzzy non-domination relation. In particular, the best action is the one with the highest value of the non-fuzzy domination relation.

3. ILLUSTRATIVE EXAMPLE: STS TRANSFER OF CARGO OPERATION

The STS transfer of cargo refers to the transfer of cargo from one ship to another. The transferred cargo mainly has to do with oil and oily products, liquefied gases (LNG or LPG), or even solid bulk cargoes such as ore (OCIMF, 2013). The operation refers to the transfer of cargo between seagoing ships positioned alongside each other. A common STS transfer operation comprises four discrete phases: the preparation, the mooring procedure, the transfer of cargo and, finally, the unmooring phase. For more details, see Ventikos and Stavrou (2013). Thirteen risk scenarios regarding the STS operation are evaluated from a team of experts with relevant experience in such operations, in terms of three different risk factors: the likelihood of a failure occurring, the severity of the failure, and the ability to detect the failure on time. Finally, a robustness analysis is performed to control the stability of the proposed model. The distributions of the experts' evaluations with respect to the risk factors are depicted in Table 1. Next, the equations (1) and (2) from the first step process are applied and the results of the confidence and doubt matrixes are shown in Table 2.

Table 1: Risk scenarios and distribution of experts' evaluation of risk factors

Description of risk scenario	ID	0-2.66	2.67-5.33	5.33-8	0-3.33	3.33-6.66	6.67-10	0-1.66	1.66-3.33	3.33-5
		+	++	+++	+	++	+++	+	++	+++
Inadequate fendering resulting in low energy collision and physical damage to ships. Detection method has medium effectiveness.	M7	0	0.8	0.2	0.2	0.6	0.2	0.4	0.6	0
Pilotage error resulting in high Energy Collision and personnel injury or fatality. Detection method has medium effectiveness.	M9	0	0.8	0.2	0.4	0.4	0.2	0	0.6	0.4
Inadequate communication resulting in Low energy collision and physical damage to ships. Detection is moderately high effectiveness.	M12	0	0.8	0.2	0.4	0.4	0.2	0.6	0	0.4
Inadequate procedures resulting in high Energy Collision and personnel injury or fatality. Detection is moderately high effectiveness.	M14	0.2	0.6	0.2	0.6	0.4	0	0	0.4	0.6
Inadequate planning for emergency breakaway resulting in low energy collision and physical damage to one or both ships. Detection method has medium effectiveness.	M21	0	0.8	0.2	0	0.8	0.2	0.4	0.4	0.2
Tug/ support vessel failure resulting in low energy collision leading to significant equipment damage, fatalities, and loss of containment. Detection method has medium effectiveness.	M22	0	0.8	0.2	0	0.4	0.6	0.2	0.4	0.4
Mooring equipment failure resulting in low Energy Collision and physical damage to ships. Detection method has medium effectiveness.	M24	0	0.6	0.4	0.2	0.4	0.4	0.2	0.6	0.2
Fatigue during mooring/unmooring procedures resulting in Low Energy Collision and physical damage to one or both ships. Detection method has medium effectiveness.	M26	0.2	0.6	0.2	0.2	0.4	0.4	0.2	0.6	0.2
Inadequate procedures resulting in the loss of containment and vapour release leading to ignition back to source resulting in fatalities/injury/hardware/steelwork damage. Detection method has moderately high effectiveness.	T28	0	1	0	0.6	0.4	0	0	0.4	0.6
Inadequate experience resulting in the loss of containment and/or vapour release leading to ignition back to source resulting in fatalities/injury/hardware/steelwork damage. Detection method has medium effectiveness.	T29	0	0.8	0.2	0.4	0.2	0.4	0	0.2	0.8
Fatigue during transfer resulting in the loss of containment and/or vapour release leading to ignition back to source resulting in fatalities/ injury/ hardware/ steelwork damage. Detection method has medium effectiveness.	T30	0	0.8	0.2	0.2	0.4	0.4	0	0.6	0.4
Different roll periods during transfer phase resulting in the loss of containment and/ or vapour release leading to ignition back to source resulting in fatalities/ injury/ hardware/ steelwork damage. Detection method is unproven or unreliable; or effectiveness of detection method is unknown to detect in time.	T35	0	0.8	0.2	0	0.4	0.6	0	0.8	0.2
Damaged mooring due to chafing and cyclic loading during transfer resulting in the loss of containment leading to vapour collection in a confined space, leading to ignition with explosion causing fatalities, injury and hull damage. Detection has medium effectiveness.	T39	0.6	0.2	0.2	0.2	0.4	0.4	0	0.6	0.4

Table 2: Matrices of Confidence and Doubt Indices

d1(a,b)

	M7	M9	M12	M14	M21	M22	M24	M26	T28	T29	T30	T35	T39
M7	-	0.84	0.84	0.84	0.84	0.84	0.68	0.84	1	0.84	0.84	0.84	0.84
M9	0.84	-	0.84	0.84	0.84	0.84	0.68	0.84	1	0.84	0.84	0.84	0.84
M12	0.84	0.84	-	0.84	0.84	0.84	0.68	0.84	1	0.84	0.84	0.84	0.84
M14	0.68	0.68	0.68	-	0.68	0.68	0.56	0.72	0.8	0.68	0.68	0.68	0.8
M21	0.84	0.84	0.84	0.84	-	0.84	0.68	0.84	1	0.84	0.84	0.84	0.84
M22	0.84	0.84	0.84	0.84	0.84	-	0.68	0.84	1	0.84	0.84	0.84	0.84
M24	0.88	0.88	0.88	0.88	0.88	0.88	-	0.88	1	0.88	0.88	0.88	0.88
M26	0.68	0.68	0.68	0.72	0.68	0.68	0.56	-	0.8	0.68	0.68	0.68	0.8
T28	0.8	0.8	0.8	0.8	0.8	0.8	0.6	0.8	-	0.8	0.8	0.8	0.8
T29	0.84	0.84	0.84	0.84	0.84	0.84	0.68	0.84	1	-	0.84	0.84	0.84
T30	0.84	0.84	0.84	0.84	0.84	0.84	0.68	0.84	1	0.84	-	0.84	0.84
T35	0.84	0.84	0.84	0.84	0.84	0.84	0.68	0.84	1	0.84	0.84	-	0.84
T39	0.36	0.36	0.36	0.48	0.36	0.36	0.32	0.48	0.4	0.36	0.36	0.36	-

d2(a,b)

	M7	M9	M12	M14	M21	M22	M24	M26	T28	T29	T30	T35	T39
M7	-	0.76	0.76	0.92	0.68	0.44	0.6	0.6	0.92	0.64	0.6	0.44	0.6
M9	0.6	-	0.68	0.84	0.52	0.36	0.52	0.52	0.84	0.6	0.52	0.36	0.52
M12	0.6	0.68	-	0.84	0.52	0.36	0.52	0.52	0.84	0.6	0.52	0.36	0.52
M14	0.44	0.56	0.56	-	0.32	0.16	0.36	0.36	0.76	0.48	0.6	0.56	0.36
M21	0.84	0.84	0.84	1	-	0.52	0.68	0.68	1	0.68	0.68	0.52	0.68
M22	0.92	0.92	0.92	1	0.92	-	0.84	0.84	1	0.84	0.84	0.76	0.84
M24	0.76	0.8	0.8	0.92	0.72	0.56	-	0.68	0.92	0.72	0.68	0.56	0.68
M26	0.76	0.8	0.8	0.92	0.72	0.56	0.68	-	0.92	0.72	0.68	0.56	0.68
T28	0.44	0.56	0.56	0.76	0.32	0.16	0.36	0.36	-	0.48	0.36	0.16	0.36
T29	0.64	0.72	0.72	0.84	0.56	0.48	0.6	0.6	0.84	-	0.6	0.48	0.6
T30	0.76	0.8	0.8	0.92	0.72	0.56	0.68	0.68	0.92	0.72	-	0.56	0.68
T35	0.92	0.92	0.92	1	0.92	0.76	0.84	0.84	1	0.84	0.84	-	0.84
T39	0.76	0.8	0.8	0.92	0.72	0.56	0.68	0.68	0.92	0.72	0.68	0.56	-

d3(a,b)

	M7	M9	M12	M14	M21	M22	M24	M26	T28	T29	T30	T35	T39
M7	-	0.36	0.6	0.24	0.64	0.44	0.56	0.56	0.24	0.12	0.36	0.48	0.36
M9	1	-	0.76	0.64	0.88	0.76	0.88	0.88	0.64	0.52	0.76	0.88	0.76
M12	0.64	0.4	-	0.4	0.64	0.52	0.52	0.52	0.4	0.4	0.4	0.4	0.4
M14	1	0.84	0.84	-	0.92	0.84	0.92	0.92	0.76	0.68	0.84	0.92	0.84
M21	0.76	0.44	0.68	0.36	-	0.52	0.6	0.6	0.36	0.28	0.44	0.52	0.44
M22	0.19	0.64	0.76	0.56	0.8	-	0.76	0.76	0.56	0.48	0.64	0.72	0.64
M24	0.88	0.56	0.68	0.44	0.76	0.6	-	0.72	0.44	0.32	0.56	0.68	0.56
M26	0.88	0.56	0.68	0.12	0.76	0.6	0.72	-	0.44	0.32	0.56	0.68	0.56
T28	1	0.84	0.84	0.76	0.92	0.84	0.92	0.92	-	0.68	0.84	0.92	0.84
T29	1	0.92	0.92	0.88	0.96	0.92	0.96	0.96	0.88	-	0.92	0.96	0.92
T30	1	0.76	0.76	0.64	0.88	0.76	0.88	0.88	0.64	0.52	-	0.88	0.76
T35	1	0.68	0.68	0.52	0.84	0.68	0.84	0.84	0.52	0.36	0.68	-	0.68
T39	1	0.76	0.76	0.64	0.88	0.76	0.88	0.88	0.64	0.52	0.76	0.88	-

D1(a,b)

	M7	M9	M12	M14	M21	M22	M24	M26	T28	T29	T30	T35	T39
M7	-	0.08	0.08	0.1	0.08	0.08	0.16	0.08	0	0.1	0.08	0.08	0.08
M9	0.08	-	0.08	0.1	0.08	0.08	0.16	0.08	0	0.1	0.08	0.08	0.08
M12	0.08	0.08	-	0.1	0.08	0.08	0.16	0.08	0	0.1	0.08	0.08	0.08
M14	0.12	0.18	0.18	-	0.18	0.18	0.26	0.16	0.1	0.2	0.18	0.18	0.12
M21	0.08	0.08	0.08	0.1	-	0.08	0.16	0.08	0	0.1	0.08	0.08	0.08
M22	0.08	0.08	0.08	0.1	0.08	-	0.16	0.08	0	0.1	0.08	0.08	0.08
M24	0.06	0.06	0.06	0.1	0.06	0.06	-	0.06	0	0.1	0.06	0.06	0.06
M26	0.18	0.18	0.18	0.2	0.18	0.18	0.26	-	0.1	0.2	0.18	0.18	0.12
T28	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	-	0.1	0.1	0.1	0.1
T29	0.08	0.08	0.08	0.1	0.08	0.08	0.16	0.08	0	-	0.08	0.08	0.08
T30	0.08	0.08	0.08	0.1	0.08	0.08	0.16	0.08	0	0.1	-	0.08	0.08
T35	0.08	0.08	0.08	0.1	0.08	0.08	0.16	0.08	0	0.1	0.08	-	0.08
T39	0.38	0.38	0.38	0.3	0.38	0.08	0.46	0.32	0.3	0.4	0.08	0.08	-

D2(a,b)

	M7	M9	M12	M14	M21	M22	M24	M26	T28	T29	T30	T35	T39
M7	-	0.14	0.14	0	0.18	0.34	0.24	0.24	0.04	0.2	0.24	0.34	0.24
M9	0.24	-	0.2	0.1	0.28	0.44	0.32	0.32	0.08	0.3	0.32	0.44	0.32
M12	0.24	0.2	-	0.1	0.28	0.44	0.32	0.32	0.08	0.3	0.32	0.44	0.32
M14	0.34	0.28	0.28	-	0.4	0.6	0.44	0.44	0.12	0.4	0.32	0.6	0.44
M21	0.08	0.08	0.08	0	-	0.24	0.16	0.16	0	0.2	0.16	0.24	0.16
M22	0.04	0.04	0.04	0	0.04	-	0.08	0.08	0	0.1	0.08	0.12	0.08
M24	0.14	0.12	0.12	0	0.16	0.28	-	0.2	0.04	0.2	0.2	0.28	0.2
M26	0.14	0.32	0.12	0	0.16	0.28	0.2	-	0.04	0.2	0.2	0.28	0
T28	0.34	0.08	0.28	0.1	0.4	0.6	0.44	0.44	-	0.4	0.44	0.6	0.44
T29	0.22	0.28	0.18	0.1	0.26	0.38	0.28	0.28	0.08	-	0.28	0.38	0.28
T30	0.14	0.32	0.12	0	0.16	0.28	0.2	0.2	0.04	0.2	-	0.28	0.2
T35	0.04	0.2	0.04	0	0.04	0.12	0.08	0.08	0	0.1	0.08	-	0.08
T39	0.14	0.32	0.12	0	0.16	0.28	0.2	0.2	0.04	0.2	0.2	0.28	-

D3(a,b)

	M7	M9	M12	M14	M21	M22	M24	M26	T28	T29	T30	T35	T39
M7	-	0.4	0.28	0.5	0.22	0.36	0.26	0.26	0.5	0.6	0.4	0.3	0.4
M9	0	-	0.12	0.2	0.06	0.12	0.06	0.06	0.18	0.2	0.12	0.06	0.12
M12	0.18	0.42	-	0.5	0.24	0.36	0.3	0.3	0.48	0.5	0.42	0.36	0.42
M14	0	0.08	0.08	-	0.04	0.08	0.04	0.04	0.12	0.2	0.08	0.04	0.08
M21	0.12	0.36	0.24	0.4	-	0.32	0.24	0.24	0.44	0.5	0.36	0.28	0.36
M22	0.06	0.22	0.16	0.3	0.12	-	0.14	0.14	0.28	0.3	0.22	0.16	0.22
M24	0.06	0.26	0.2	0.3	0.14	0.24	-	0.16	0.34	0.4	0.26	0.18	0.26
M26	0.06	0.26	0.2	0.3	0.14	0.24	0.16	-	0.34	0.4	0.26	0.18	0.26
T28	0	0.08	0.08	0.1	0.04	0.08	0.04	0.04	-	0.2	0.08	0.04	0.08
T29	0	0.04	0.04	0.1	0.02	0.04	0.02	0.02	0.06	-	0.04	0.02	0.04
T30	0	0.12	0.12	0.2	0.06	0.12	0.06	0.06	0.18	0.2	-	0.06	0.12
T35	0	0.16	0.16	0.2	0.08	0.16	0.08	0.08	0.24	0.3	0.16	-	0.08
T39	0	0.12	0.12	0.2	0.06	0.12	0.06	0.06	0.18	0.2	0.16	0.08	-

During the second step the relative importance (weight) of the criteria is calculated. To do so, the experts were asked to evaluate on an aggregated level, the relative importance of the three risk factors (O, S, D). The corresponding results are shown in Table 3.

Table 3: (Left) Evaluation of risk factors from experts. (Right) Min-max range of the weight of risk factors

	Exp1	Exp2	Exp3	Exp4	Exp5
S	High	High	High	High	High
O	Medium	Medium	High	Medium	Medium
D	High	Very High	Very High	Medium	Very High

	Exp1	Exp2	Exp3	Exp4	Exp5	min	aver	max
S	0.4	0.29	0.25	0.5	0.3	0.25	**0.34**	0.5
O	0.2	0.14	0.25	0.25	0.1	0.143	**0.20**	0.25
D	0.4	0.57	0.5	0.25	0.6	0.25	**0.46**	0.571

To transform the linguistic terms used for the three risk factors (O, S, and D) to crisp values the following scale is assumed (with the various levels related in a linear way): the medium importance equals a/2; high importance equals a, and very high importance equals 2a. Hence, each expert provides, based on his/her experience, his/her overall view of the relative importance among the three risk factors (the aggregation of the three evaluations equals one), regardless of the specifics of scenarios and situations; as, for example, this metric for the expert "Exp3" is given below:

$$a + a + 2a = 1 \Rightarrow a = \frac{1}{4} \tag{9}$$

After the calculation of the weight of the risk factors, the total confidence index C(a,b) matrix is constructed as shown in Table 4 (Left). Moreover, the degree of global credibility from step 3, is calculated for the pairs of actions and the relative matrix is depicted in Table 4 (Centre). Finally, in step 4 the fuzzy domination relation is applied to compare the different pairs of actions to determine the outranking intensity between each pair of actions. The equation (6) is applied and the results are shown in Table 4 (Right).

Table 4: (Left) The matrix of the total confidence index; (Centre) Matrix of the degree of global credibility; (Right) Matrix of outranking intensity of the pairs of actions.

C(a,b)

	M7	M9	M12	M14	M21	M22	M24	M26	T28	T29	T30	T35	T39
M7		0.64	0.7	0.67	0.7	0.52	0.6	0.63	0.7	0.5	0.56	0.53	0.56
M9	0.785		0.7	0.77	0.71	0.59	0.68	0.71	0.8	0.62	0.67	0.63	0.67
M12	0.661	0.62		0.69	0.62	0.51	0.55	0.58	0.7	0.58	0.54	0.47	0.54
M14	0.68	0.68	0.7		0.6	0.5	0.59	0.62	0.8	0.59	0.7	0.71	0.61
M21	0.812	0.7	0.8	0.75		0.58	0.65	0.68	0.8	0.57	0.63	0.58	0.63
M22	0.654	0.81	0.8	0.82	0.86		0.78	0.81	0.8	0.72	0.77	0.76	0.77
M24	0.825	0.73	0.8	0.75	0.77	0.64		0.73	0.8	0.61	0.68	0.66	0.68
M26	0.786	0.69	0.7	0.61	0.73	0.6	0.67		0.7	0.57	0.64	0.62	0.66
T28	0.704	0.7	0.7	0.77	0.62	0.52	0.6	0.64		0.61	0.61	0.55	0.61
T29	0.803	0.81	0.8	0.85	0.75	0.7	0.74	0.77	0.9		0.76	0.72	0.76
T30	0.858	0.79	0.8	0.81	0.8	0.68	0.75	0.78	0.8	0.67		0.73	0.74
T35	0.932	0.82	0.8	0.8	0.88	0.75	0.81	0.84	0.8	0.67	0.78		0.78
T39	0.764	0.7	0.7	0.74	0.7	0.59	0.68	0.71	0.7	0.58	0.64	0.63	

d(a,b)

	M7	M9	M12	M14	M21	M22	M24	M26	T28	T29	T30	T35	T39
M7		0.64	0.72	0.67	0.70	0.52	0.60	0.63	0.70	0.40	0.56	0.53	0.56
M9	0.79		0.74	0.77	0.71	0.59	0.68	0.71	0.80	0.62	0.67	0.63	0.67
M12	0.66	0.62		0.69	0.62	0.51	0.55	0.58	0.72	0.58	0.54	0.47	0.54
M14	0.68	0.68	0.68		0.60	0.50	0.59	0.62	0.77	0.59	0.70	0.71	0.61
M21	0.81	0.70	0.78	0.75		0.58	0.65	0.68	0.78	0.57	0.63	0.58	0.63
M22	0.65	0.81	0.85	0.82	0.86		0.78	0.81	0.85	0.72	0.77	0.76	0.77
M24	0.82	0.73	0.77	0.75	0.77	0.64		0.73	0.77	0.61	0.68	0.66	0.68
M26	0.79	0.69	0.74	0.61	0.73	0.60	0.67		0.73	0.57	0.64	0.62	0.66
T28	0.70	0.70	0.70	0.77	0.62	0.43	0.60	0.64		0.61	0.61	0.55	0.61
T29	0.80	0.81	0.81	0.85	0.75	0.70	0.74	0.77	0.89		0.76	0.72	0.76
T30	0.86	0.79	0.79	0.81	0.80	0.68	0.75	0.78	0.84	0.67		0.73	0.74
T35	0.93	0.82	0.82	0.80	0.88	0.75	0.81	0.84	0.83	0.67	0.78		0.78
T39	0.76	0.70	0.70	0.74	0.70	0.59	0.68	0.71	0.72	0.58	0.64	0.63	

dD(a,b)

	M7	M9	M12	M14	M21	M22	M24	M26	T28	T29	T30	T35	T39
M7		0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
M9	0.15		0.12	0.09	0.005	0.00	0.00	0.01	0.10	0.00	0.00	0.00	0.00
M12	0.00	0.00		0.01	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
M14	0.01	0.00	0.00		0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
M21	0.11	0.00	0.16	0.15		0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00
M22	0.13	0.22	0.34	0.32	0.28		0.14	0.22	0.41	0.01	0.09	0.01	0.18
M24	0.22	0.06	0.22	0.15	0.11	0.00		0.06	0.17	0.00	0.00	0.00	0.00
M26	0.15	0.00	0.15	0.00	0.04	0.00	0.00		0.09	0.00	0.00	0.00	0.00
T28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00
T29	0.40	0.19	0.23	0.27	0.18	0.00	0.13	0.20	0.27		0.08	0.04	0.18
T30	0.29	0.13	0.25	0.11	0.17	0.00	0.07	0.14	0.23	0.00		0.00	0.09
T35	0.40	0.19	0.35	0.10	0.29	0.00	0.14	0.22	0.29	0.00	0.06		0.15
T39	0.20	0.03	0.16	0.12	0.08	0.00	0.00	0.05	0.11	0.00	0.00	0.00	

4. RESULTS AND ROBUSTNESS ANALYSIS

The results from the implemented methodology are shown in Table 5. The first column shows the ranking of the risk scenarios with regard to the average weights of the risk factors. The remaining columns show the results from different combinations of the weight factors according to the experts' evaluation in Table 3.

Table 5: Matrix of the ranking of actions with regard to the min-max weight limitations of the risk factors

	Mean	1	2	3	4	5	6	7	8	9	10	11
S	**0.34**	**0.25**	**0.25**	**0.50**	**0.50**	**0.40**	**0.34**	**0.29**	**0.34**	**0.50**	**0.29**	**0.25**
O	**0.20**	**0.25**	**0.20**	**0.14**	**0.20**	**0.14**	**0.14**	**0.25**	**0.25**	**0.25**	**0.14**	**0.20**
D	**0.46**	**0.50**	**0.55**	**0.36**	**0.30**	**0.46**	**0.52**	**0.46**	**0.41**	**0.25**	**0.57**	**0.57**
1	M22	M22	M22	T29	T29	T29	T29	M22	M22	T29	M22	M22
2	T29	T35	T35	M22	M22	M22	M22	T35	T29	M22	T35	T35
3	T35	T29	T29	T30	T30	T35	T35	T29	T35	T30	T29	T29
4	T30	T30	T30	T35	T35	T30	T30	T30	T30	T35	T30	T30
5	M24	M24	M24	T39	M9	T39	T39	M24	M24	M9	M24	M24
6	T39	M26	T39	M9	M14	M24	M24	T39	M9	M14	T39	T39
7	M26	T39	M26	M24	M24	M9	M9	M26	T39	M24	M26	M26
8	M9	M9	M9	M14	T28	M26	M26	M9	M26	T28	M9	M9
9	M21	M21	M21	T28	T39	M21	M21	M21	M21	M12	M21	M21
10	M14	M12	M12	M26	M12	M14	M12	M12	M12	T39	M12	M12
11	M12	M7	M7	M21	M21	M12	M7	M7	T28	M21	M7	M7
12	M7	T28	M14	M12	M26	T28	T28	T28	M14	M26	M14	M14
13	T28	M14	T28	M7	M7	M7	M14	M14	M7	M7	T28	T28

Durbach and Stewart (2012) address the importance of the weight calculation of the selected factors in their study. In order to control the stability and verify the strength of the proposed model the algorithm of section 2 is applied from the entire range of the weights of the risk factors. Figure 1 shows the fluctuation of the ranking of the risk scenarios with regard to the total range of the weight of the risk factors. On the left side the possible combinations of the weights according to the experts' evaluation is depicted. The results demonstrate the stability of the model and the accurate evaluation of the risks of the STS operation.

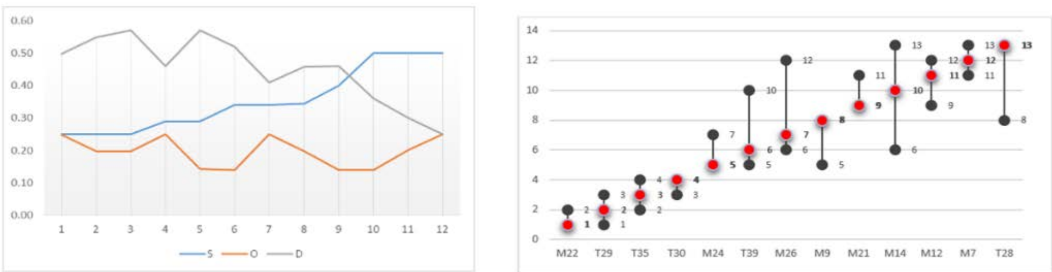


Figure 1. Left: Different combinations of risk factors according to min-max limitations. Right: Fluctuations of risk factors in robustness analysis for the min-max ranges of the weights.

An important observation from the results comes from scenario M26, which refers to fatigue during mooring/unmooring procedures. For high values of importance of detectability (D> 0.41) fatigue is a prominent factor (in seventh place), whilst for low values of detectability the risk of fatigue is underestimated. Taking into consideration that the detectability has values under 0.40 in 9 of the 12 cases, it is obvious that the important issue of the risk from fatigue is properly addressed in the model.

5. CONCLUSION

Risk evaluation of maritime activities is at the top of the agenda in the shipping business. The MCDA techniques are an efficient tool in the hands of experts to help operators to compose the true picture regarding the safety of a marine activity and the potential threats that could compromise the success of the operation. This paper proposed a combination of MCDA methods, which includes initially the outranking approach and next the fuzzy domination relation. The objective goal is the modeling of the experts' judgments under uncertainty. In particular, three risk factors were selected as the criteria to evaluate risk scenarios; the occurrence and the consequences of a failure as well as the ability to detect a failure on time. The inherent uncertainty of these factors is expressed by the probability distribution of the experts' evaluation for each risk criterion. The model of outranking approach under uncertainty, borrowed from the MCDA discipline, can give reliable and positive results when evaluating risks of maritime activities. Moreover, limitations in the number of actions comes from the model implementation due to computational complexity. Thus, in the case of multiple sets of actions, alternative methodologies should be taken into account. Nevertheless, for small number of actions the proposed methodology can work effectively on risk evaluation of maritime activities.

ACKNOWLEDGEMENT

This work was partially supported by the "IKY fellowships of excellence for postgraduate studies in the Greece-Siemens program" for the first in order author. It was also partially supported by the Ministry of Education, Religious Affairs, Culture and Sports of Greece and the European Social Fund, under the Grant THALIS, MIS 377350: Methodological approaches for studying robustness in multiple criteria decision making projects for the third in order author.

REFERENCES

Aven T. 2012. The risk concept—historical and recent development trends. Reliability Eng. System Safety, 99:33–44.

Compton K.L.,Faber R.,Ermolieva T.Y.,Linnerooth-Bayer J.,Nachtnebel H-P.,2009.Uncertainty and Disaster Risk Management Modeling the Flash Flood Risk to Vienna and Its Subway System, Int. Inst. for Applied Systems Analysis.

Durbach I.N, 2014. Outranking under uncertainty using scenarios, European J. of Operational Research, 232(1), 98–108.

Durbach, I., Stewart, T. 2012. Modelling uncertainty in multi-criteria decision analysis. EJOR, 223(1), 1–14.

Fan Z., Liu Y., Feng B., 2010. A method for stochastic multiple criteria decision making based on pairwise comparisons of alternatives with random evaluations. European Journal of Operational Research, 207(2):906{915.

Goerlandt F., Montewka J. 2015. Maritime transportation risk analysis: Review and analysis in light of some foundational issues, Reliability Engineering and System Safety 138 pp. 115-134.

IMO MERC, 2010. Manual on Oil Pollution, Section I, Prevention of pollution, IMO Publication.

Islei G.,Lockett G.,Naude P.1999. Judgemental modelling as an aid to scenario planning and analysis.Omega, 27(1):61-73.

Keeney, R., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. John Wiley & Sons, NY.

Martel J., D'Avignon G, Couillard G., 1986. A fuzzy outranking relation in multicriteria decision-making. EJOR, 25:258-271.

Martel J., D'Avignon G., 1982. Project ordering with multicriteria analysis. Multicriteria Analysis. Part II: Applications. European Journal of Operational Research, 10 (1982) 56-69.

OCIMF, 2013. Ship to Ship transfer guide for petroleum, chemicals and liquefied gases.

Ventikos N.P., Stavrou I.D., 2013., Ship to Ship (STS) Transfer of Cargo: Latest Developments and Operational Risk Assessment. SPOUDAI. JEB, Vol.63, Issue 3.

Schoemaker P., 1995. Scenario Planning: A Tool for Strategic Thinking. Sloan Management Review, 36:25, 25.

Skjong R., 2007, Hazid of tanker operations, SAFEDOR project, No IP-516278.

Siskos, J., Hubert Ph.,1983. Multicriteria analysis of the impacts of energy alternatives, EJOR 13, pp. 278-299.

Siskos, J., Lochard J., Lombard J., 1984. A multicriteria decision-making methodology under fuzziness: Application to the evaluation of radiological protection in nuclear plants, TIMS/Studies in the Management Sciences, 20, pp.261-283.

Stavrou D., Ventikos N., 2015. Risk Evaluation of Ship-to-Ship Transfer of Cargo Operations by Applying PFMEA and FIS, Proceedings of 61st Annual Reliability and Maintainability Symposium, Palm Florida, USA, pp. 619-625.

Van der Heijden K., 1996. Scenarios: The Art of Strategic Conversation. John Wiley & Sons, New York.

Zaras, K., Martel, J.M., 1994. Multiattribute analysis based on stochastic dominance. In: Munier, B., Machina, M.J. (Eds.), Models and Experiments in Risk and Rationality. Kluwer Academic Publishers, Dordrecht, pp. 225–248.

Durbach I.N, 2014. Outranking under uncertainty using scenarios

Robustness analysis approaches in political decision making

Nikos Tsotsolas, Spiros Alexopoulos
University of Piraeus
80, Karaoli & Dimitriou Str., GR-18534 Piraeus, Greece

Abstract

A crucial issue, when dealing with political decisions, is the radical uncertainty about the present (e.g. lack or poor quality of information) and also about the future. In the literature it is mentioned that robustness analysis is a way of supporting government decision making when dealing with uncertainties and ignorance. In the present research a framework to deal with robustness in policy making is proposed, in a concrete and comprehensive manner. Different definitions and approaches of Robustness Analysis in government decision-making, concerning the present and the future, as a way to support the identification of potential robust strategies in policy circles, are discussed. Robust strategic approaches in political decision-making could be perceived as an effort to trade some optimal performance for less sensitivity to assumptions, performing well over a wide range of versions and possible futures, and keeping options open. Relevant researches suggest that this often adopted strategy is also usually identified as the most robust choice. Robust political strategies may be preferable to optimum strategies when the uncertainty, usually taking the form of epistemic uncertainty (referred to a lack of complete knowledge of the political and societal environment), is sufficiently deep and the set of alternative policy options is sufficiently rich. The proposed approach is incorporating, by adopting a facilitated approach, a holistic robustness approach for policy making, using appropriate measures covering three notions of robustness namely: robust solution, robust conclusion, and robust decision in dynamic context. The framework under discussion is actually based on a robustness centre approach focusing on the determination of scenarios for the present as well as for multiple futures and on the evaluation of the options within and across the scenarios. Actually, we are suggesting the incorporation of the (procedure, version) approach with the approach of the initial commitments leading to a set of representative future states and the inter-scenario robustness approach. The corresponding robustness of all alternative actions in a policy circle across the plausible variable settings shall be evaluated using different types of measures.

KEYWORDS

Robustness Analysis, Political Decision-Making, MCDA.

1. INTRODUCTION

Political decisions are often considered as complex, multifaceted and involve many different stakeholders with different priorities and objectives. Furthermore, policy-making consists of several sequential actions focusing on the achievement of a specific goal with societal, economic and political implications, with several feedbacks and loops. So, we may describe the whole process as a policy circle (Lasswell 1956). Very often DMs, when confronted with such problems, attempt to use heuristic and intuitive approaches to simplify the complexity until the problem seems more manageable. The problem is even more complex if we consider that, according to principles of sociotechnical design, specific - most of the times conflicting - objectives must be best met by the joint optimization of technical and social aspects. Of course, the absence of certainty, the interference of political-power and the presence of complexity, as well as the unquestionable existence of wicked problems, also called social-messes (Ritchey 2011), makes it necessary the application of niche methods. Nobody could assert that a specific framework can tackle the whole extent of a policy circle or guarantee optimal or robust solutions, but it could certainly increase rationality and accountability at a certain level.

Multi-criteria Decision Aiding (MCDA) offers techniques designed to deal with situations, as the aforementioned, in which there are multiple conflicting goals for reaching strategic decisions (Roy 2005; Montibeller and Franco 2010). Furthermore, the process of creating, evaluating and implementing strategic political decisions is typically characterised by the consideration of potential synergies between different options, long term consequences, and the need of key stakeholders to engage in significant psychological and social negotiation about the strategic decision under consideration. However, certain adaptations to the methods, tools and processes of MCDA are required, if it is to be effectively applied in such a context (Tsoukias et al. 2013). These adaptations have to tackle issues such as the probabilistic nature of the data and the uncertainty of future events and system states. Thus the robustness analysis is a major issue in governmental decisions and the appropriate approaches and techniques shall be contained in different stages of any proposed methodology.

Lempert and Collins (2007) state that an important characteristic of the mechanism of political actions is that politics is a game among forward-looking stakeholders. Thus, the government's current payoffs are equal to the "net present value" of its anticipated future actions (and resulting victories/losses), not just its present and past policy choices. In this context actions of government policy makers can be interpreted as efforts to: (a) design "efficient" policies (those for which every objective is reached with the minimum loss for the other relevant objectives) to improve government performance, as measured by well-defined indicators, while at the same time, and (b) maintain a political behaviour true to their "political identity" (ideology, values, interests, influences).

An efficient political decision making process requires a robustness analysis of the results of this process, which could give broad issues and multiple values being considered (Tsoukias et al. 2013). However, robustness can be defined in many ways by putting the focus on the different elements of the decision problem, namely the model, the data, the futures, the method, the algorithm, the technical parameters. If we want to deal with robustness in a holistic way we have to tackle each one of these elements. A common point is that robustness is called to provide resistance or self-protection, as Roy mentioned (2010), against the existence of uncertainty, contingency and ambiguity of the past (historical data), the present (decision model) and the future (possible states) resulting to vague approximations and zone of ignorance. As far as the past and the present time are concerned the source of this situation might be the information shortage for the decision problem in question and the different interpretation of reality depending on the DMs' points of view. On the other hand, the uncertainty of the future states is even deeper and increased proportionally to the timeline, while on the same time is affected by the choice of each alternative decision.

2. ROBUSTNESS ANALYSIS APPROACHES

Efforts towards seeking for mere optimality might be very often a misleading approach to political decision problems by providing solutions that are not well-performed in different scenarios (or versions) of the reality and in different futures. Thus, the robustness of a model or/and of a solution should be assessed and evaluated each time so that the analyst shall be able to have a clear picture regarding the reliability and stability of the produced results. Robustness shall be expressed using measures, also referred as robustness criteria, which are understandable by the analyst and the decision maker. Based on these measures the DM may accept, or reject, or adapt the proposed decision model. Given the fact that uncertainty is present, influencing every decision-making context and that it appears in several different ways, it should be neither omitted, nor relegated. Its importance shall be realized and it shall be considered in an appropriate manner. As robustness allows us to experiment with uncertainty, it is necessary to define its concept, its significance and to emphasize its importance in the MCDA field.

For the evaluation of a solution several robustness measures have been proposed, most them based on the three standard measures (absolute robustness, absolute deviation, relative deviation) proposed by Kouvelis and Yu (1997). Since these standard measures are considered to be conservative, because they tend to give the higher importance to the worst case, several of the proposed approaches try to take into consideration other cases approaching to a median or average case. These other measures allow the DMs to express a specific degree of optimism about future outcomes by selecting some really good solutions which show remarkable bad performance only in a minimum portion of cases. Nevertheless, in cases where a good solution might have extremely bad consequences, under specific scenarios, that might be also irreversible or, in cases where uncertainty about the future is severe, the security provided by the pessimistic measures of Kouvelis and Yu is preferred.

Following that perspective, Montibeller et al (2006) discussed the Goodwin and Wright (2001) approach for evaluating the performance of alternatives in different scenarios, where each decision alternative is a combination of strategic option in a given future scenario (a_i-s_j). They also extended this approach by introducing notions such as: different priorities across scenarios, elicitation of strategies' performance, analysing inter-scenario risk and inter-scenario robustness of options. According to their proposition, each one of the *n* strategic options a_i is evaluated on the *m* criteria under each *s* scenario, using a different model for each scenario: $V_s(a_i) = \sum w_{s,k} v_{s,k}(a_i)$, where $w_{s,k}$ is the weight of the *k*-th criterion under the *s*-th scenario ($\sum w_{s,k} = 1$ for a given scenario) and $v_{s,k}(a_i)$ is the value of the *i*-th alternative on the *k*-th criterion (scaled from

0 to 100) under the *s*-th scenario. Notice that the model allows different weights for distinct scenarios, in order to reflect different future priorities.

A critical issue in evaluating the solutions over a set of cases is the generation or selection of these cases. Usually, in the literature cases are referred to as scenarios, where each scenario represents a probable instance of reality provided through a plausible set of parameters' values of the model that are considered as uncertain. Where the number of probable instances is huge then a representative set of scenarios is used.

Another approach of robustness connects it with the dynamic context in which decisions are made. This approach is of particular interest when dealing with political decision making because of the multistage nature of many strategic decisions and the radical uncertainty of a long or even middle term future of the political environment. Political decisions, in a degree higher than other strategic decisions, must be or can be staged. That is, the commitments made at the first stage of a decision do not necessarily define completely the future state of the system. There will be one or more future opportunities to modify or to define it further. These futures can be identified but their details are not known in advance and furthermore the initial commitments may affect the characteristics of the futures. The paradox, as Rosenhead (2002) points out, is how can we be rational in taking decisions today if the most important fact that we know about future conditions is that they are unknowable? The answer to the aforementioned paradox is the concept of flexibility. An initial decision (maybe the 1st stage of a multistage decision) is considered to be flexible if it keeps open attractive (i.e. good or at least acceptable) future options at specific points on a time line or when some kind of event creates a specific trigger. As Hites et al. (2006) state, the fewer obstacles a decision poses to future good decisions, the more flexible it is. Under this dynamic approach, a robust decision is one that does not undermine any possible future choice. The evaluation of the robustness of the decisions shall be done at each stage for each pair (alternative,future) by taking into consideration how these decisions will affect the context of future decisions. This specific robustness concern focuses on the continuous evolution of decisions which shall be adapted to various middle or long term futures.

According to this approach, let a_i be the initial alternative decision chosen from a set of decisions. Let S be the set of all possible plans realised in the future. Let S_i be a subset of S of attainable plan after decision a_i has been chosen. Let S^* and S_i^* be respectively the subset of S and S_i of "good" or "acceptable" plans. Then the robustness of a_i is measured in function of the subset of good plans, that is: $r_i = n(S_i^*)/n(S^*)$, where $n(S)$ is the number of elements in the set S. Obviously, the greater the value of r_i the more the decision is robust.

A sequence of actions a_i in a form of pathways could be considered as well, where an initial commitment to a short-term action may lead to another action when an adaptation tipping point is reached (Haasnoot et al. 2013). Each possible pathway could be considered from the beginning as a candidate action and evaluated as such.

3. A HOLISTIC APPROACH

Given the necessity of dealing with robustness issues in political decision making, in order to confront uncertainty about the present and the future real-world states, our approach is based on a robustness centre view, focusing on the determination of scenarios for the present, as well as for multiple futures, and on the evaluation of the options within and across the scenarios. Actually, we are suggesting the incorporation of the (procedure, version) approach of Roy (2010) with the ideas of Rosenhead et al. (1972) concerning the initial commitments leading to a set of representative future states and the inter-scenario robustness approach of Montibeller and Franco (2011). In this latter work even though the authors had discussed that one challenge of using the concept of robustness is that there are different ways of conceptualising it, they assessed robustness using only the notion of robust solution, based the work of Kouvelis and Yu (1997). Furthermore, the basis of their evaluation approach is the practical application of the multi-attribute value function of Goodwin and Wright (2001) as discussed in Montibeller et al. (2006). Our proposal is to incorporate, by adopting a facilitated approach, a holistic robustness approach in a MCDA framework, using appropriate measures covering different notions of robustness, with the evaluation of each alternative under each plausible scenario. Moreover, we propose an extension of the scope of scenarios in relation to the ones proposed in Goodwin and Wright (2001) by including the notion of pair (procedure, version) used by Roy (2010) combined with the effect of subsequent actions on future states of the problem. Our approach is consisted of the following steps:

- Define a set of k alternative actions, A= {a_1, a_2, …, a_i, … a_k}
- Set future states fs where fs=0, 1, 2, …, q, while fs=0 is referring to the present time
- Define a set of plausible variable settings s=(procedure, version), using Roy's approach as an extension to the notion of scenario, in a future state fs, $S^{fs}=\{s^{fs}_1, s^{fs}_2, \dots, s^{fs}_j, \dots, s^{fs}_{m(fs)}\}$. The cardinality of S^{fs} is probably different for each future fs, and equal to $m(fs)$. It is very likely that for the distant futures the information about plausible variable settings maybe poor, so the cardinality of the S^{fs} will be decreased.

The set A of the alternative actions may evolve during the policy circle and new, combined and updated actions might enter in A, so its cardinality |A|=k will be increased. We suggest that all previous versions of combined or updated actions shall remain in set A for comparison reasons. Furthermore, each pair (procedure, version) includes a number of parameters, called frailty points by Roy (2010), depending on the processing procedure in a certain method family and on the different version of reality's representation which is strongly connected to the decision model. The first set of parameters includes purely technical parameters (eg. thresholds for replacing strict equalities, concordance level) as well as parameters which are in one way or another connected to the real problem (eg. weights, preference and veto thresholds). This second subset of parameters could be also viewed as part of a version of the problem along with parameters directly connected with the real-life context and which could take the form of objective function and constraint matrix coefficients, constraints right-hand-side values, etc. (see Aissi and Roy 2010 for extended discussion on the notion of (p, v) pairs).

It could be even considered that each variable setting is also affected by the choice of an initial action a_{ic} at present time ($fs=0$), referred as initial commitment by Rosenhead (2002), so the set of plausible variable settings could be represented as: $S^{fs}(a_{ic})=\{s^{fs}_1(a_{ic}), s^{fs}_2(a_{ic}), \dots, s^{fs}_j(a_{ic}), \dots, s^{fs}_{m(fs)}(a_{ic})\}$ for fs=1, 2, …, q given that the selection of choice a_{ic} could affect each s in the future. For the special case of fs=0, present time, the set of plausible variable settings is not affected by the action i: $S^0=\{s^0_1, s^0_2, \dots, s^0_j, \dots, s^0_{m(0)}\}$.

For each alternative a_i an overall evaluation of its performance under each plausible variable setting $s^{fs}_j(a_{ic})$, denoted as $V(s^{fs}_j(a_{ic}), a_i)$, shall be calculated. This overall evaluation can be consisted of one or several performance measures depending on the family of procedures that are used.

Furthermore, the corresponding robustness of all alternative actions across the plausible variable settings shall be evaluated using three types of measures:

i. $R_{st}(s^{fs}_j(a_{ic}), a_i)$: The *Standard Type*, based on the definitions of Kouvelis and Yu (1997) as well as on the corresponding proposed variations by Roy (2010). This type of measures, which is the most commonly used, highlights solutions that are good enough in most scenarios and not very bad at any scenario.
ii. $R_{cr}(s^{fs}_j(a_{ic}), a_i)$: The *Credibility Type*, based on the proposals of Siskos and Grigoroudis (2010) where the robustness of a solution is evaluated during post-optimality analysis by calculating stability and credibility measures.
iii. $R_{cr}(s^{fs}_j(a_{ic}), a_i)$: The *Flexibility Type*, based on the ideas of Rosenhead et al. (1972). According to their approach an action a_i is considered to be robust, or equally flexible, if a significant number of 'good' or at least 'acceptable' plans are kept open in future states fs.

The robustness measures that belonging to the aforementioned types shall be further elaborated and specified for different methods. They shall also be presented to the DMs through visual representations using software applications for a better comprehension and more efficient feedback as argued by Montibeller and Franco (2010) and by Siskos and Grigoroudis (2010).

4. CONCLUSIONS

The proposed methodology meant to be an integrated part of an iterative, continuous process with feedback and forward loops supporting prospective and retrospective analyses based on multiple views and multiple states of the future, along with its focus on the aforementioned robustness concerns. Such a framework shall be consisted of several discrete stages which all of them shall be implemented in every case; however their duration and sequence shall vary depending on the case, given the fact that as a rule, policy analysis projects require a customised design. Towards the designing of such a framework we believe that an adapted approach of a facilitated MCDA model, based on the ideas of Montibeller and Franco (2010), could efficiently support mainly the following tasks: defining the problem, scoping participation, tackling uncertainty with future scenarios, considering multiple objectives, designing and appraising complex strategic options, and finally considering long term consequences. The decision analysts, who will act as facilitators in the policy circle, shall be highly sophisticated in order to be able to understand the political environment and the dynamics of the participatory process. The facilitator could also been regarded in this process as mediator, who designs the rules and procedures for negotiating in the policy circle and manages the interaction and progress of the process.

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning".

REFERENCES

Goodwin P., Wright G., 2001. Enhancing strategy evaluation in scenario planning: a role for decision analysis. *Journal of Management Studies*, Vol. 38, No. 1, pp.1-16

Haasnoot M., Kwakkel J.H., Walker Warren E., Maat J., 2013. Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, Vol. 23, No.2, pp.485-498

Hites R., De Smet Y., Risse N., Salazar-Neumann M., Vinkce P., 2006. About the applicability of MCDA to some robustness problems. *European Journal of Operational Research*, Vol. 174, pp.322-332

Kouvelis P., Yu G., 1997. *Robust Discrete Optimisation and Its Applications*. Kluwer Academic Publishers, Netherlands

Lasswell H.D., 1956. *The decision process: seven categories of functional analysis*. University of Maryland Press, College Park

Lempert R.J., Collins M.T. 2007. Managing the Risk of Uncertain Threshold Responses: Comparison of Robust, Optimum, and Precautionary Approaches. *Risk Analysis-An official publication of the Society for Risk Analysis*, Vol. 27, No.4. pp.1009-1026

Montibeller G., Gummer H., Tumidei D., 2006. Combining scenario planning and multi-criteria decision analysis in practice. *J Multicriteria Decis Anal*, Vol.14, No.1, pp.5–20

Montibeller G., Franco A., 2010. Multi-Criteria Decision Analysis for Strategic Decision Making. In: Zopounidis C, Pardalos PM (eds) Handbook of Multicriteria Analysis. Springer, New York, pp.25-48

Montibeller G., Franco A., 2011. Raising the bar: strategic multi-criteria decision analysis. *Journal of the Operational Research Society*, Vol. 62, pp.855-867

Ritchey T., 2011. Wicked Problems – Social Messes: Decision Support Modelling with Morphological Analysis. Springer-Verlag, Berlin, Heidelberg

Rosenhead J., 2002. Robustness Analysis. Newsletter of the European Working Group “Multiple Criteria Decision Aiding”, Series 3, No. 6, pp.6-10

Rosenhead J., Elton M., Gupta SK., 1972. Robustness and optimality as criteria for strategic decisions. *Operational Research Quarterly*, Vol.23, No.4, pp.413-430

Roy B., 2005. Paradigms and challenges. In: Figueira JR, Greco JS, Ehrgott S (eds) Multiple Criteria Decision Analysis - State of the Art Surveys. Springer, New York, pp 3-24

Roy B., 2010. Robustness in operational research and decision aiding: A multi-faceted issue. *European Journal of Operational Research*, Vol. 200, pp.629-638

Siskos Y., Grigoroudis E., 2010. New Trends in Aggregation-Disaggregation Approaches. In: Zopounidis C, Pardalos PM (eds) Handbook of Multicriteria Analysis. Springer New York, pp 189-214

Analyzing robustness of the MUSA method through a simulation model

Yannis Politis
School of Science and Technology
Hellenic Open University
Parodos Aristotelous 18, GR26 335, Patra, Greece

Evangelos Grigoroudis
School of Production Engineering & Management
Technical University of Crete
University Campus, 73100 Chania, Greece

Abstract

The MUSA method is a collective multicriteria approach which is used for measuring and analyzing customer satisfaction. It is a preference disaggregation model following the principles of ordinal regression analysis. The method has a set of parameters that may influence the stability of the provided results. These parameters depend on the characteristics of the problem (e.g., number of customers, number of criteria, length of ordinal satisfaction scales, consistency or variability of data) or they are set by the analyst (e.g., preference or post-optimality thresholds). The main aim of this study is to examine how different values of these parameters may affect the robustness of the MUSA results. For this reason an extended simulation model has been developed and different customer satisfaction data sets have been generated and solved. These data sets combine properties about the deviation level, the number of customers, the number of criteria and the number of satisfaction levels and have been tested for different values of preference and post-optimality thresholds. Different fitting and stability indicators have been used for the evaluation of the results. The analysis of the results facilitates the selection of appropriate model parameters considering the characteristic properties of the customer satisfaction data.

KEYWORDS

MUSA method, Robustness analysis, Simulation, Satisfaction data generation.

1. INTRODUCTION

The MUSA (MUlticriteria Satisfaction Analysis) method is a preference disaggregation approach following the main principles of ordinal regression analysis. It measures and analyzes customer satisfaction assuming that customer's global satisfaction is based on a set of criteria representing service characteristic dimensions. The main object of the MUSA method is the aggregation of individual judgments into a collective value function.

The method has a set of parameters that may influence the stability of the provided results. The main aim of this study is to examine how different values of these parameters may affect the robustness of the MUSA results. For this reason, a number of different customer satisfaction data sets have been generated combining different values of the method's parameters in order to examine how these can affect the fitting and stability level of the provided results.

2. THE MUSA METHOD

2.1 Mathematical Development

The MUSA method assesses global and partial satisfaction functions Y^* and X_i^* respectively, given customers' judgments Y and X_i. The method follows the principles of ordinal regression analysis under constraints using linear programming techniques (Jacquet-Lagrèze and Siskos, 1982; Siskos and Yannacopoulos, 1985; Siskos, 1985). The ordinal regression analysis equation has the following form:

$$\begin{cases} Y^* = \sum_{i=1}^{n} b_i X_i^* \\ \sum_{i=1}^{n} b_i = 1 \end{cases} \tag{1}$$

where b_i is the weight of the i-th criterion and the value functions Y^* and X_i^* are normalised in the interval [0, 100]. The main objective of the method is to achieve the maximum consistency between the value function Y^* and the customers' judgments Y. Based on the above modelling and by introducing two error variables σ^+ and σ^-, the ordinal regression equation becomes as follows:

$$\tilde{Y}^* = \sum_{i=1}^{n} b_i X_i^* - \sigma^+ + \sigma^- \tag{2}$$

where $\tilde{Y}^*$ is the estimation of global value function Y^* and σ^+ and σ^- are the overestimation and underestimation error, respectively.

The final form of the linear programming problem is as follows:

$$[\min]\mathrm{F} = \sum_{j=1}^{M} \sigma_j^+ + \sigma_j^-$$

under the constraints

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{k=1}^{t_{ji}-1} w_{ik} - \sum_{m=1}^{t_j-1} z_m - \sigma_j^+ + \sigma_j^- = 0 \text{, for } j = 1,2,...,M \\
&\sum_{m=1}^{\alpha-1} z_m = 100 \\
&\sum_{i=1}^{n}\sum_{k=1}^{\alpha_i-1} w_{ik} = 100 \\
&z_m \geq 0 \text{ , } w_{ik} \geq 0 \text{ , } \forall m,i,k \\
&\sigma_j^+ \geq 0 \text{ , } \sigma_j^- \geq 0 \text{, for } j = 1,2,...,M
\end{aligned} \tag{3}$$

where t_j and t_{ji} are the judgments of the j-th customer globally and partially for each criterion i=1,2,…,n, M is the number of customers and z_m, w_{ik} are a set of transformation variables such us:

$$\begin{cases} z_m = y^{*m+1} - y^{*m} & \text{for } m = 1,2,...,\alpha-1 \\ w_{ik} = b_i x_i^{*k+1} - b_i x_i^{*k} & \text{for } k = 1,2,...,\alpha_i - 1 \text{ and } i = 1,2,...,n \end{cases} \tag{4}$$

where α and α_i is the evaluation ordinal scale for the global assessment and for the assessment of the i-th criterion, respectively.

Furthermore, taking into account the hypothesis of strict preferential order of the scales of some or all the dimensions/criteria, the following conditions are met:

$$\begin{cases} y^{*m} < y^{*m+1} \Leftrightarrow y^m \prec y^{m+1} & \text{for } m = 1,2,\ldots,\alpha \\ x_i^{*k} < x_i^{*k+1} \Leftrightarrow x_i^k \prec x_i^{k+1} & \text{for } k = 1,2,\ldots,\alpha_i - 1 \text{ and } i = 1,2,\ldots,n \end{cases} \tag{5}$$

where $\prec$ means "strictly less preferred". Based on (5) the following conditions occur:

$$\begin{cases} y^{*m+1} - y^{*m} \geq \gamma \Leftrightarrow z_m \geq \gamma \Leftrightarrow z_m^{'} \geq 0 & \text{for } m = 1,2,\ldots,\alpha \\ x_i^{*k+1} - x_i^{*k} \geq \gamma_i \Leftrightarrow \mathrm{w}_{ik} \geq \gamma_i \Leftrightarrow \mathrm{w}_{ik}^{'} \geq 0 & \text{for } k = 1,2,\ldots,\alpha_i - 1 \text{ and } i = 1,2,\ldots,n \end{cases} \tag{6}$$

where γ and γ_i are the preference thresholds (minimum step of increase) for the value functions Y^* and X_i^*, respectively, with γ, γ_i> 0, and it is set that:

$$\begin{cases} z_m = z_m' + \gamma & \text{for } m = 1,2,\ldots,\alpha \\ w_{ik} = w_{ik}' + \gamma_i & \text{for } k = 1,2,\ldots,\alpha_i - 1 \text{ and } i = 1,2,\ldots,n \end{cases} \tag{7}$$

The proposed extension affects the initial LP and constitutes the *generalized* form of the MUSA method (Grigoroudis and Siskos, 2002; Grigoroudis and Siskos, 2010).

2.2 Stability Analysis

The preference disaggregation methodology consists also of a post optimality analysis stage in order to face the problem of multiple or near optimal solutions. During the post-optimality analysis stage of the MUSA method, *n* linear programs (equal to the number of criteria) are formulated and solved. Each linear program maximizes the weight of a criterion and has the following form:

$$\begin{cases} [\max] F' = \sum_{k=1}^{\alpha_i - 1} w_{ik} \quad \text{for } i = 1,2,\ldots,n \\ \text{subject to} \\ F \leq F^* + \varepsilon \\ \text{all the constraints of LP (3)} \end{cases} \tag{8}$$

where F^* is the optimal value of the objective function of LP (3) and ε is a small percentage of F^*. The average of the optimal solutions given by the n LPs (8) may be considered as the final solution of the problem. In case of instability, a large variation of the provided solutions appears and the final average solution is less representative.

The stability of the results provided by the post-optimality analysis is calculated with the Average Stability Index *(ASI)*. *ASI* is the mean value of the normalized standard deviation of the estimated weights b_i and is calculated as follows:

$$ASI = 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{n \sum_{j=1}^{n} \left(b_i^j\right)^2 - \left(\sum_{j=1}^{n} b_i^j\right)^2}}{100\sqrt{n-1}} \tag{9}$$

where b_i^j is the estimated weight of the criterion *i*, in the *j*-th post-optimality analysis LP (Grigoroudis and Siskos, 2002; Grigoroudis and Siskos, 2010).

Furthermore, the fitting level of the MUSA method refers to the assessment of a preference collective value system (value functions, weights, etc.) for the set of customers with the minimum possible errors. For this reason, the optimal values of the error variables indicate the reliability of the value system that is evaluated.

The Average Fitting Index (*AFI*) depends on the optimum error level and the number of customers:

$$AFI = 1 - \frac{F^*}{100 \cdot M} \tag{10}$$

where F^* is the minimum sum of errors of the initial LP, and *M* is the number of customers.

The *AFI* is normalised in the interval [0, 1], and it is equal to 1 if $F^* = 0$, i.e. when the method is able of evaluating a preference value system with zero errors. Similarly, the *AFI* takes the value 0 only when the pairs of the error variables σ_j^+ and σ_j^- take the maximum possible values.

3. SELECTION OF PARAMETERS AND THRESHOLDS

3.1 Preference Thresholds

The problem of selecting appropriate model parameters is focused on the preference values γ, γ_i, and the tradeoff threshold ε during the post-optimality analysis.

In this section, it is examined how different values of these parameters may affect the fitting and stability level of the MUSA results. For this reason, a large number of indicative customer satisfaction data sets have been used. These data sets present different characteristic properties (e.g. number of criteria, number of satisfaction levels, consistency of judgments and stability level, etc.). One of the most important results of this analysis is that the selection of preference thresholds γ and γ_i depends mainly on the stability of the results.

In particular, in case of stable results, the average fitting index *AFI*, as well the average stability index *ASI*, have high values (~100%) for $\gamma = \gamma_i = 0$. The increase of γ and γ_i will cause a relatively small reduction of the fitting and stability level of the results. This finding may be justified by the fact that the preference thresholds provide a lower bound for the model variables z_m and w_{ik}. For example, by increasing γ_i, the MUSA method is forced to assign a minimum weight of γ_i $(a_i–1)$ to each criterion. Thereby, the initially achieved fitting and stability level of the results is decreased. Consequently, in case of stable results, it is preferred to set $\gamma = \gamma_i = 0$ (or at least very small values for the preference thresholds).

In case of unstable results, *ASI* may take rather small values (e.g. <50%) for $\gamma = \gamma_i = 0$, while *AFI* may retain a relatively high level (e.g. >80%). The increase of preference thresholds γ and γ_i may improve the stability of the results, but it will decrease the fitting level of the model. As previously noted, this is justified by the fact that the preference thresholds determine the minimum value of the criteria weights. Thus, in case of instability, the increase of γ and γ_i will decrease the variability observed in the post-optimality table, and therefore, it will increase the average stability index.

Generally, the following rules should be considered when selecting appropriate values for the preference thresholds γ and γ_i:

- An arbitrarily large increase of the preference thresholds may falsify the customer satisfaction data set; large values of γ and γ_i require stronger assumptions for the preference conditions
- Based on the assessed values of γ_i the minimum weight of criterion *i* is γ_i $(a_i–1)$. This assumption should be verified by the decision-maker.

The post-optimality threshold ε does not affect the fitting ability of the model, since the Average Fitting Index does not depend on the post-optimality results. Moreover, it should be noted that usually, in real world applications, $F^* > 0$, and thus ε may be assessed as a small percentage of the optimal value of the objective function *F*.

Similarly to the previous analyses, a large number of customer satisfaction data sets have been used, in order to examine the effect of post-optimality threshold on the stability level of the MUSA results. These experiments show that the increase of ε causes a decrease of the average stability index *ASI*, regardless of the stability level of results. This is rather expected, since an increase of ε implies an increase of the near optimal solutions space.

The decrease of *ASI* is larger in case of unstable results because F^* is larger and, thus, the overall tradeoff value $(1+\varepsilon)F^*$ is larger in the post-optimality analysis.

Consequently, ε is a near optimal solutions threshold that should be always selected as a small percentage of F^*. The modification of ε should take into account the following:

- A very large value of ε will falsify the information provided by the post-optimality analysis, and decrease the stability ability of the model.
- A very low value of ε will not give the ability to explore the near optimal solutions space during post-optimality analysis.

3.2 Experimental Comparison Analysis

In order to examine how different values of the MUSA method's parameters can affect the fitting and stability level of the provided results, 243 different customer satisfaction data sets have been generated combining properties about the deviation level, the number of customers, the number of criteria, the type of customers and the number of satisfaction levels. These data sets have been tested with different values of preference thresholds (γ, γ_i) and post-optimality thresholds (ε). Table 1 summarizes the basic properties of the generated data sets.

Table 1: Properties of the generated data sets

Deviation level (D_e)	Number of customers (*M*)	Number of criteria (*n*)	Number of satisfaction levels (α=α_i)	Type of customers
0.05-0.25-0.40	100-500-100	3-5-10	3-5-7	Non demanding-Normal-Demanding

Table 2 presents a summary of the simulation results for the MUSA method. The fitting level of the MUSA method is rather high, since *AFI* ranges between 87.4% and 100%, with an average of 94.8% for the generated data sets. This justifies the ability of the MUSA method to effectively evaluate a value system for the set of customers.

ASI ranges between 32.0% and 99.5%, with an average of 85.3% for the generated data sets which shows that for particular data sets the MUSA method can achieve higher or lower levels of stability.

Table 2: Simulation results for the MUSA method

Index	Statistics	Value
AFI	Range	0.874-1.000
	Average	0.948
ASI	Range	0.320-0.995
	Average	0.853

Another important objective of the experimental analysis is to examine the influence of the parameters of the MUSA method to the fitting and the stability level of the estimated results. For this reason, a series of one-way ANOVA analyses have been performed in order to analyze the influence of each parameter of the experiment to the calculated *AFI* and *ASI* indices. Tables 3 and 4 present the summary results for this analysis of variance, from where the following points raise:

- The chosen deviation level of the experiment does not affect *ASI*, but influences *AFI*. This is more or less expected, since D_e determines the consistency of the satisfaction judgments and therefore it is strongly related with the fitting ability of the MUSA method.
- The size of the data set (number of customers) does not seem to affect the fitting and stability level of the MUSA method.
- It is confirmed that the increase of ε causes a decrease of the average stability index *ASI*, regardless of the stability level of results. As it is already mentioned, this is rather expected, since an increase of ε implies an increase of the near optimal solutions space.
- *ASI* is influenced by the chosen number of criteria and the number of satisfaction levels. Specifically, greater number of criteria and of satisfaction levels improves *ASI*. It seems that giving customers the capability to express their preferences with more choices regarding the number of criteria and the number of satisfaction levels can be considered as giving indirectly additional more specified information about the problem, which improves the stability of the provided results.

Table 3: Summary results for one-way ANOVA (*ASI*)

Factors	*SS*	*df*	*MS*	*F*	*p*-value
Deviation level	1350.65	2	675.32	4.218	0.016
Number of customers	203.13	2	101.56	0.614	0.542
Number of criteria	2391.50	2	1195.75	6.107	0.002
Number of satisfaction levels	2885.79	2	1442.89	7.674	0.000
Demanding level	757.41	2	378.70	2.325	0.100
Tradeoff threshold	7.20	3	2.40	1058.384	0.000

Table 4: Summary results for one-way ANOVA (*AFI*)

Factors	*SS*	*df*	*MS*	*F*	*p*-value
Deviation level	1006.39	2	503.19	118.778	0.000
Number of customers	0.71	2	0.35	0.040	0.961
Number of criteria	74.54	2	37.27	3.627	0.029
Number of satisfaction levels	23.08	2	11.54	1.320	0.270
Demanding level	8.85	2	4.42	0.496	0.610

There is evidence that different data sets can give better results with specific combinations of the MUSA parameters. However, the presented experiment may be considered as a pilot analysis, since a larger number of data sets is required, in order to increase the reliability of the findings.

4. CONCLUSIONS

There is evidence that different data sets can give better results with specific combinations of the MUSA parameters. However, the presented experiment may be considered as a pilot analysis, since a larger number of data sets is required, in order to increase the reliability of the findings. For this reason, future research may focus on other alternative customer satisfaction evaluation models or examination of how several combinations of these parameters may affect the reliability of the results. Robustness improvement of the MUSA method can also be achieved with the introduction of additional information or constraints. Robustness improvement with the combination of different approaches (e.g. selection of appropriate parameters, introduction of additional information and constraints, etc) should also be examined.

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES. Investing in knowledge society through the European Social Fund.

REFERENCES

Grigoroudis, E. and Y. Siskos (2002). Preference disaggregation for measuring and analysing customer satisfaction: The MUSA method, *European Journal of Operational Research*, 143 (1), 148-170.

Grigoroudis, E. and Y. Siskos (2010). *Customer Satisfaction Evaluation: Methods for Measuring and Implementing Service Quality*, Springer, New York.

Jacquet-Lagrèze, E. and J. Siskos (1982). Assessing a set of additive utility functions for multicriteria decision-making: The UTA method, *European Journal of Operational Research*, 10 (2), 151-164.

Siskos, J. (1985). Analyses de régression et programmation linéaire, *Révue de Statistique Appliquée*, 23 (2), 41-55.

Siskos, Y. and D. Yannacopoulos (1985). UTASTAR: An ordinal regression method for building additive value functions, *Investigaçao Operacional*, 5 (1), 39-53.

Estimating criteria weights exploiting priorities of the criteria and techniques of robustness analysis

Athanasios Spyridakos
Piraeus University of Applied Sciences, P. Ralli and Thivon 250, 12244 Aigaleo, Athens, Greece

Nikos Tsotsolas
University of Piraeus 80, Karaoli & Dimitriou str., GR-18534 Piraeus, Greece

Eleftherios Siskos
National Technical University of Athens 9, Iroon Polytechniou Str., GR-15780 Zografou, Athens, Greece

Dennis Yannacopoulos
Piraeus University of Applied Sciences, P. Ralli and Thivon 250, 12244 Aigaleo, Athens, Greece

Abstract

Utilising linear programming techniques in order to estimate the criteria weights based on preferences of decision makers expressed in the form which is supported by Simos method usually results in infinite number of optimal solutions and low robustness. This paper presents techniques and methods which on the one hand exploit the results of the robustness analysis in order to obtain additional and focused preference information by the decision maker (DM) while on the other estimate vectors of weights which are might be considered to be closer to DM preferences with higher level of robustness. For the presentation of the proposed approach an illustration example is used utilizing the capabilities of RAVI software.

KEYWORDS

Robustness analysis, Criteria weights, Simos method.

1. INTRODUCTION

The Simos method and its revision (Simos, 1990a,1990b, Figueira and Roy, 2002) constitutes a simple methodological tool through which the DMs express their preferences of the criteria importance and these preferences are translated to a unique vector of weights. DMs are asked to declare the relative importance of the criteria in a pair-wise manner with the use of cards by ranking the criteria from the least important to the most important. In cases of indifference DMs create subsets of ex aequo criteria. DMs can also increase the distance of the weights between criteria by introducing one or more white cards between two successive criteria (or two successive subsets of ex aequo criteria). The estimation of the criteria weights is achieved by using simple numerical techniques and normalizations. This method has been extensively used in multicriteria outranking relations methods such ELECTRE, PORMETHEE, TOPSIS, AHP (Siskos and Tsotsolas, 2015) which require the estimation of criteria weights.

Simos method provides a very simple and comprehensible tool that can result to the estimation of the criteria weights reflecting the preferences of the DM regarding the criteria priorities and it can be used to quantify the criteria priorities without requiring to attach great emphasis on accuracy. Nevertheless, the process recommended by Simos bears some robustness issues (Scharlig, 1996, Figueira and Roy, 2002), and more specifically, it calculates a unique weighting vector even though an infinite number of solutions might exist, also satisfying the requirements and constraints of the method. Furthermore, the accuracy of the calculated criteria weights is thrown into doubt. The existence of other solutions and the accuracy of the criteria weights may have significant effect on the evaluation of alternative actions, when applying these weights in decision making problems.

The revised Simos method proposed by Roy and Figuera (2002), tried to tackle some robustness issues by introducing the index $z = w_k/w_r$ where w_k corresponds to the weight of the most important criterion and w_r to the weight of the least important criterion. The value z is asked by the DM so as to provide a better accuracy of the calculated criteria weights. That is to say that the DM is asked to quantify, with great precision, how much more he/she influenced by the most important criterion in relation to the least important one. Even though this additional information increases the robustness of the estimated solution still several compatible weights vectors exist and the DMs are not informed for this issue.

The multicriteria methods are based on large extent to their interactive nature which enriches the process for the estimation of preference models, responding and reflecting the preferences of DMs. The Simos method includes limited opportunities for interactions with the DMs focused only to adaptations of the initial preference information of DMs and the level of index z (Shanian et al, 2008), without allowing interventions to the estimated weights vector.

Using the preference information of the Simos methods the weights of the criteria can be elicited by applying Linear Programming (LP) techniques (Siskos & Tsotsolas, 2015). The conditions derived by the priorities on criteria and any additional preference information provided by the DMs are leading to infinite set of weights vectors expressing low robustness, which are actually bordered into a n-dimensional hyper-polyhedron. So, our interest is focused on the estimation of the ranges of the criteria weights which can be estimated by the calculation of their minimum and maximum values. The LPs used for the estimation of the criteria weights ranges have the following form.

Let n the number of criteria g_i and m the number of white cards noted wc_h. Let us denote by $p_1, p_2, \ldots, p_n$ the unknown weights of criteria, and by $w_1, w_2, \ldots, w_k$ the unknown weights of the white cards. Then we have the following linear program (LP1):

$Min\ p_j\ \&\ Max\ p_j$, for $j = 1, 2, \ldots, n$

s.t.

$p_j = p_{j+1}$, if g_j is followed by g_{j+1}, and g_{j+1} belongs to the same importance class as g_j,

$p_j < p_{j+1} \Leftrightarrow p_{j+1} - p_j \geq \delta$, if g_j is followed by g_{j+1}, and g_{j+1} belongs to a most importance class

$p_j < w_h$ and $w_h < p_{j+1} \Leftrightarrow w_h - p_j \geq \delta$ and $p_{j+1} - w_h \geq \delta$, if between g_j and g_{j+1} a white card wc_h is placed

$p_1 + p_2 + \ldots + p_n = 1$

$p_1 \geq 0, p_2 \geq 0, \ldots, p_n \geq 0; w_1 \geq 0, w_2 \geq 0, \ldots, w_k \geq 0$

where δ is set equal to the minimal quantity, say 0.01 (1%) for instance, to differentiate two consecutive classes of the ranking. In case of the revised Simos method, the following equation should be added in the constraints' set $p_n = zp_1$, where z is the ratio introduced in revised Simos method.

The barycenter solution of the estimated hyper-polyhedron could be used as the working weights' vector for further analysis of the decision aiding process (Siskos and Yannacopoulos, 1985).

2. ROBUSTNESS ANALYSIS AND BARYCENTRIC SOLUTIONS

A key issue when applying linear programming techniques in order to estimate the criteria weights is the level of the accuracy of the barycentric solution regarding the actual preferences of DM, as well as the existence of other compatible weights' vectors which could be also used by the DMs. An important issue in such situations is the level of robustness. High level of robustness means low range of compatible weights vectors, so the barycenter is considered to be a confident representation of DMs' preference. Unlike, low robustness may generate objections to how it is appropriate to use the barycenter for further analysis of the decision problem.

The methodological framework proposed in this research work includes processes for:

a) the assessment of the level of robustness of the estimated set of criteria weights and imprinting of its structure and
b) the application of interactive feedbacks through which DMs are asked to provide additional preference information which results to the shrinkage of the estimated hyper-polyhedron and consequently to more robust preference models (Kadzinski et al., 2012).

2.1 Robustness Measures and Tomographical Approach

The measurement of the level of robustness includes a visual approach and two indices. The first step in the proposed framework is the calculation of robustness measures, which are able to provide an initial evaluation of the robustness level of the set of compatible value functions. A set of indices are proposed for this purpose.

The first type of indices used is the range between the maximum and minimum values of the criteria weights for every criterion, as these values are estimated at each vertex of the hyper-polyhedron during post optimality analysis. This index is very simple, but it gives a picture of the extent of robustness in each criterion and at the same time is fairly comprehensible to DM. For the i-th criterion the index is estimated as:

$\mu_i = \left(\max\left(p_{ij}\right) - \min\left(p_{ij}\right)\right)$, p_{ij} the weight of the *i* criterion of the *j*,

$i = 1,2,...,n$, $j = 1,2,...,\text{m}$, *n*: the number of criteria and *m*: the number of vertices of hyper-polyhedron

The second index used in this analysis represents the normalized standard deviation of the different solutions corresponding to the hyper-polyhedron vertices, where the value 1 corresponds to total robustness and 0 to complete un-robustness of the preference models (Hurson and Siskos 2014). This normalized index is called Average Stability Index (ASI):

$$ASI = 1 - \frac{\sum_{i=1}^{n}\sqrt{\left(m\left(\sum_{j=1}^{m}\left(p_i^j\right)^2\right) - \left(\sum_{j=1}^{m}p_i^j\right)^2\right)}}{m\sqrt{(n-1)}}$$

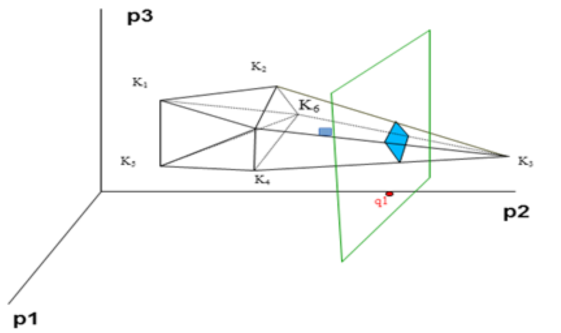
n :number of criteria ,*m*: number of vertices of hyper-polyhedron .

Apart from the indices, the tomographical technique provides the mean for scanning the level of robustness into the estimated hyper-polyhedron in a similar way to the technique used in medical practice of diagnosis. The idea of the tomographical approach is to discretize the n-dimensional estimated hyper-polyhedron of the criteria weights by using n-1 dimensional cutting hyper-polyhedra. For the presentation of the cutting tomographies the parallel coordination system is used (Figure 1). By this way, with a simple and comprehensive way we can visualise the levels of criteria weights' robustness. For the estimation of these n-1 dimensional cutting hyper-polyhedra the Linear Programme of post optimality analysis is enriched with the following conditions:

$p_i = q$, where $q = \min\left(p_i\right) + rt$, where *t* is a predeterine step,

$r = 0.1....,l$, where *l* is the total number of steps and $q \le \max\left(p_i\right)$

Figure 8 Visual example of Tomography of Hyper-polyhedron for three criteria



For the illustration of the aforementioned approach a case study is used with 6 criteria (crit.1, crit.2, crit.3, crit.4, crit.5, crit.6) and 2 white cards (wc_1 and wc_2). DM groups together the cards associated to the criteria having the same importance into six different subsets of *ex aequo* and that she/he also uses a white card, as follows: Subset of *ex aequo* criteria: {crit.3}, {wc_1}, {crit.2, crit.1}, {crit.4}, {wc_2}, {crit.5, crit.6.}.

For the use of revised Simos, DM also states that crit.3 is 6 times more important than crit.6. so *z*=6. The solutions of the LPs during the post optimality following the structure of LP1 are shown in Table 1.

Table 1 Post optimal solutions of LPs

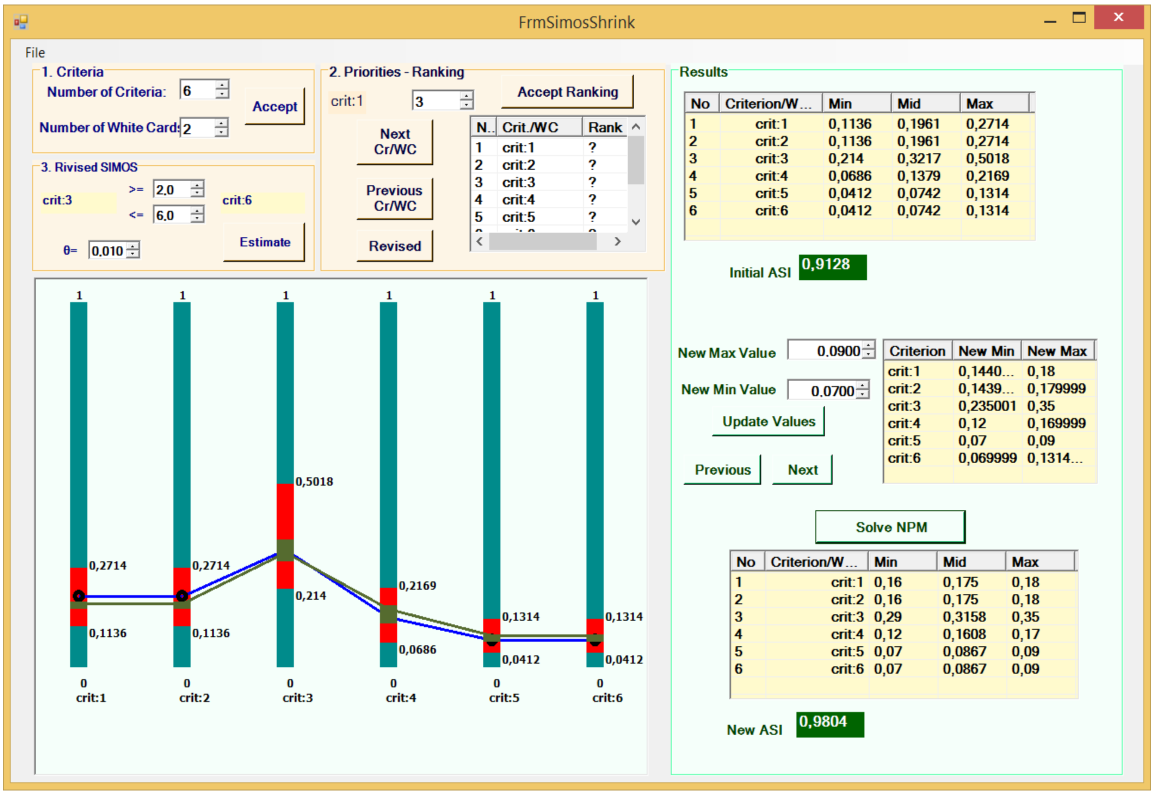
No	Criterion	Min	Max	μ_i	Barycenter	ASI
1	Crit.1	0.1136	0.2714	0.1578	0.1961	0.9128
2	Crit.2	0.1136	0.2714	0.1578	0.1961	
3	Crit.3	0.2140	0.5018	0.2878	0.3217	
4	Crit.4	0.0686	0.2169	0.1483	0.1389	
5	Crit.5	0.0412	0.1314	0.0902	0.0742	
6	Crit.6	0.0412	0.1314	0.0902	0.0742	

Furthermore, the utilisation of the tomographical approach (Figure 1) gave significant results as far as the structure of the low robustness is concerned. Higher robustness is presented in the tomographies of the lower and higher values of all the criteria weights. Also, there were tomographies for every criterion presenting total robustness.

2.2 Feedback Process

The analysis of the values of Table 1 may trigger new dialogues with the DM at a foreclosure process of values of criteria weights which are considered either extreme or not satisfying the real attitudes of DM. For example in the examined case study the maximum value of 0.5018 for the weight of criterion 3 can be considered extremely high by the DM. Also for criteria 5 and 6 the weight 0.0412 can be considered very low. DM sates that the maximum weight of criterion 3 must be reduced and the minimum weights of criteria 5 and 6 must be increased.

Figure 2 Screenshot of RAVI software with parallel co-ordination diagram setting limits for criteria



Within the feedback process additional preference information is requested by the DM concerning extreme points of criteria weights (very low or very high) and the cases of unexpected wide ranges of the criteria weights. The identification of new lower and upper limits of the criteria weights inserts new conditions and constraints in the LP programmes at the post optimality analysis process. The estimation of the new preference model may lead to a new hyper-polyhedron, which, in reality, constitutes a shrinking of the initial one. The new conditions embedded into the linear problems have the following form:

$p_i \geq r_{i\min}$, $p_i \leq r_{i\max}$ where $r_{i\min}$, $r_{i\max}$ the lower and upper calculated values of criteria weights

In our case study the dialogue with the DM lead to the identification of new accepted lower and higher values of the criteria weights. Table 2 includes the accepted by the DM minimum and maximum values of the criteria weights. The enrichment of the LPs with the new conditions resulted higher robustness of the criteria weights while the ASI climbed at 0.9804. The new barycenter (0.175, 0.175, 0.3158, 0.1608, 0.0867, 0.0867) can be considered satisfactory in order to move forward to the decision support process.

Table 2 Updated post optimal solutions of LPs after feedback

No	Criterion	Min	Max	μ_i	Barycenter	ASI
1	Crit.1	0.144	0.180	0.036	0.1750	
2	Crit.2	0.144	0.180	0.036	0.1750	
3	Crit.3	0.230	0.350	0.120	0.3158	0.9804
4	Crit.4	0.120	0.170	0.050	0.1608	
5	Crit.5	0.070	0.090	0.020	0.0867	
6	Crit.6	0.070	0.131	0.061	0.0867	

3. CONCLUSIONS

The utilisation of linear programming, for the calculation of the criteria weights based on preference information taken through Simos method, makes apparent the low level of robustness of the solutions corresponding to the set of compatible preference models. The analysis of robustness and the interactive processes of tomographical approach, as well as the subsequent feedback process concerning the shrinking ranges of weights, proposed in this research work, provides the tools to tackle the low level of robustness and simultaneously to develop a dialogue with the DM in order to improve the decision aiding process. Furthermore, the proposed framework support the DM to clarify in a better way the structure of his/her preferences and to estimate weights of criteria reflecting in a more concrete way his/her true preference attitudes.

ACKNOWLEDGEMENT

This research has been co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning".

REFERENCES

Figueira J., Roy R., 2002. Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure, *European Journal of Operational Research*, Vol. 139, pp.317-326.

Hurson C., Siskos Y., 2014. A synergy of multicriteria techniques to assess additive value models, *European Journal of Operational Research*, http://dx.doi.org/10.1016/j.ejor.2014.03.047

Kadzinski M., Greco S., Slovinski R., 2012. Selection of a representative function in robust multiple criteria ranking and choice, *European Journal of Operational Research*, Vol. 217, pp.541-553.

Scharlig A., 1996. Pratiquer Electre et PROMETHEE – Un complément à décider sur plusieurs critères, Collection Diriger L'Entreprise, *Presses Polytechniques et Universitaires Romandes*, Lausanne

Shanian A., Milani A.S., Carson C., Abeyarante R.C., 2008. A new application of ELECTRE III and revised Simos' procedure for group material selection under weighting uncertainty, *Knowledge-Based Systems*, Vol. 21, No 7, pp.709-720.

Simos J., 1990a. L'évaluation environnementale: Un processus cognitif négocié. *Thèse de doctorat*, DGF EPFL, Lausanne.

Simos J., 1990b. Evaluer l'impact sur l'environnement: Une approche originale par l'analyse multicritère et la négociation. *Presses Polytechniques et Universitaires Romandes*, Lausanne.

Siskos E., Tsotsolas N., 2015. Elicitation of criteria importance weights through the Simos method: A robustness concern, *European Journal of Operational Research*, http://dx.doi.org/10.1016/j.ejor.2015.04.037

Siskos Y., Yannacopoulos D., 1985. UTASTAR, an ordinal regression method for building additive value functions. *Investigacao Operacional*, Vol. 5, No 1, pp.39-53.

Assessing the financial performance of European banks under stress testing scenarios: A multicriteria approach

Michael Doumpos, Constantin Zopounidis, Pantelis Fragiadakis
Technical University of Crete
School of Production Engineering and Management
Financial Engineering Laboratory
University Campus, 73100 Chania, Greece

Abstract

The European banking system has been under considerable pressure since the beginning of the financial crisis in 2007-2008. Except for the global credit crunch, the European sovereign debt crisis has created additional difficulties. In response to the need for increasing the transparency and stability in the European financial/banking system and identifying weaknesses in banks' capital structures, EU-wide stress tests have been performed by the European Banking Authority (EBA), on a regular basis since 2010. In this context, the aim of this study is to examine the financial performance of the European banks that participated in the stress tests of EBA. The analysis takes into account the actual financial data of the banks, on the basis of the CAMEL framework, as well their results derived from the stress tests. The evaluation of the banks' financial strength, under the status quo, the baseline, and adverse scenarios, is performed through a multicriteria decision aid classification methodology that is used to distinguish between banks that failed to meet the minimum capital requirement conditions imposed by EBA from well-capitalized banks.

KEYWORDS

Banking, Financial risk management, Multicriteria decision aid.

1. INTRODUCTION

Banks have a prominent role in the financial and business environment. Despite the advances made in the regulatory and supervisory framework since the late 1990s with the introduction of the first Basel capital accord, the global credit crunch of 2007-2008 has put considerable the banking industry under enormous pressure, leading to an increasing number of bank defaults with a direct impact on global financial stability and economic growth. The situation in Europe has been even more difficult due to the European sovereign debt crisis.

In response, to such developments and the growing concerns about financial stability, supervisors and regulators have further intensified their actions towards strengthening the risk management and monitoring procedures used in the banking sector. In this context, the implementation of formal capital adequacy stress tests has played an important role. Stress tests exercises simulate the financial performance and capital adequacy of financial institutions over a range of scenarios about relevant macro-economic and financial variables. Such tests have been regularly conducted after the 2007-2008 crisis, in USA, Europe, and elsewhere.

Despite some criticism on the framework on which the conducted tests have been based and the value of their outcomes, empirical results provides evidence indicating the stress tests are indeed relevant. For instance, Morgan et al. (2014) found that the 2009 US stress test produced the information demanded by the markets, whereas Petrella and Resti (2013) examined the European test of 2011 and concluded that it produced valuable information for market participants (beyond simple bank-specific accounting information) and they can play a role in mitigating bank opaqueness.

In this context, analyzing and predicting the outcomes of stress tests could provide useful information to supervisors, the management of banking institutions, and investors. Following this line of research, the this study explores the possibility of construct decision models that can act as early warning systems for capital shortfalls that banks may face, based on the outcomes of stress test exercises. For that purpose, we focus on the tests conducted by the European Banking Authority (EBA) in 2010, 2011 and 2014. On the basis of the tests' results a multicriteria methodology is employed to construct classification models that discriminate banks that are likely to face capital shortfalls from banks that are well-capitalized. The model construction process is based on a robust preference disaggregation technique.

The rest of the paper is organized as follows. Section 2 presents the multicriteria methodology used in the analysis. Section 3 describes the data and the empirical results. Finally, section 4 concludes the paper and discusses some possible future research directions.

2. A ROBUST MULTICRITERIA CLASSIFICATION APPROACH

Multicriteria decision aid (MCDA) provides a variety of different modeling forms, which can be used for financial decision making purposes (Doumpos and Zopounidis, 2014). In this study we employ additive value models in the framework of the UTADIS method (Doumpos and Zopounidis, 2002). Additive models are intuitive and simple to understand and implement. They allow the modeling of non-linear preferences, while retaining the interpretability of simple linear models.

Formally, an additive value function is expressed in the following form:

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

In this model, the global value $V(\mathbf{x}_i)$ is used as composite indicator of the financial performance of bank i, which acts as a proxy of the likelihood that the bank will face capital adequacy shortfalls in the future. The global value is a weighted average of partial scores $v_1(x_{i1}),\ldots,v_n(x_{in})$ defined over a set of n performance attributes. The partial scores (marginal values) are defined through marginal value functions $v_1(\cdot),\ldots,v_n(\cdot)$, which are scaled such that $v_j(x_{j*})=0$ and $v_j(x_j^*)=w_j$, where w_j is the trade-off constant of attribute j, whereas x_{j*} and x_j^* are the most and least risky level of attribute j, respectively.

The parameters of the above model can be inferred (estimated) using a sample of m banks classified into q risk classes, based on preference disaggregation techniques (Jacquet-Lagrèze and Siskos, 2001). In this study we employ the robust formulation proposed by Doumpos and Zopounidis (2007). In the two-class setting of this study (banks that are likely to face capital shortfalls – class C_1 – versus well-capitalized banks – class C_2), the decision model can be constructed through the solution of the following linear program (for simplicity we assume a linear value function with all attributes scaled in [0, 1]):

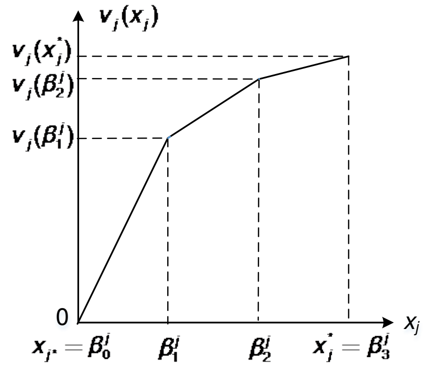
$$\begin{aligned}
\min\quad & \lambda\left(\frac{m}{2m_1}\sum\nolimits_{i\in C_1}\varepsilon_i^{+}+\frac{m}{2m_2}\sum\nolimits_{i\in C_2}\varepsilon_i^{-}\right)+\sum_{j=1}^{n}w_j \\
\text{s.t.}\quad & w_1x_{i1}+w_2x_{i2}+\ldots w_nx_{in}+\varepsilon_i^{+}\geq t+\delta \qquad \forall i\in C_1 \\
& w_1x_{i1}+w_2x_{i2}+\ldots w_nx_{in}-\varepsilon_i^{-}\leq t-\delta \qquad \forall i\in C_2 \\
& w_j,\varepsilon_i^{\pm},t\geq 0
\end{aligned}$$

where m_1 and m_2 denote the number of banks from classes C_1 and C_2, respectively, ε_i^{+} and ε_i^{-} are error variables for bank i, t is the threshold that distinguishes the two classes (a bank is classified in the low risk class C_1 if $V(\mathbf{x}_i)>t$), whereas δ and λ are user-defined positive constants. The parameter λ defines the trade-off between the fit of the model and its complexity (in this analysis we set λ=10, chosen on the basis of experimental testsing). In the solution of the above model, the attributes' trade-offs may not sum up to one, but as shown in Doumpos and Zopounidis (2007), this can be easily achieved by scaling the obtained solution. As explained in Doumpos and Zopounidis (2007), the above modeling formulation is closely related to Tikhonov's regularization principle (Tikhonov et al., 1995) and it has some additional interesting properties (increased stability, connections with the quality of the data, etc.).

A common approach to allow for non-linear marginal value functions is to assume that these can be expressed in a piecewise linear form. This can be done by splitting the range of each attribute j into k_j+1 subintervals, defined by break-points $\beta_0^j<\beta_1^j(r)<\cdots<\beta_{k_j}^j<\beta_{k_j+1}^j$ as illustrated in Figure 1, and modeling the marginal value functions an interpolation scheme (Doumpos and Zopounidis, 2002).

However, this piecewise linear modeling approach is heavily affected by the approach used to define the breakpoints in the scale of the attributes. To ameliorate this problem, we employ an approach similar to the one introduced by Doumpos et al. (2007) in the context of support vector machine learning models. This approach results in smooth marginal value functions, by averaging multiple piecewise linear functions. As the results are smooth functions, they are more robust to data changes. An outline of the algorithmic process is given below.

Figure 1 Piecewise linear modeling of a marginal value function



1. Define R sets of breakpoints $\{\beta_1^j(r), \beta_2^j(r), \ldots, \beta_{k_j}^j(r)\}_{r=1}^R$ for each attribute j
2. For each set of breakpoints:
   a. Construct an additive model with marginal value functions $v_1^r(x_1), v_2^r(x_2), \ldots, v_n^r(x_n)$
   b. Use the marginal value functions to calculate the values at all breakpoints $v_j^r(\beta_\ell^j(p))$, for all $j=1,\ldots,n$, $\ell=1,\ldots,k_j$, $p=1,\ldots,R$
3. Construct the final smoothed marginal value functions by averaging:

$$v_j(\beta_\ell^j(p)) = \frac{1}{R}\sum_{r=1}^{R} v_j^r(\beta_\ell^j(p))$$

Doumpos et al. (2007) used an adaptive algorithm to specify the breakpoints in step 1 of the above algorithm. Instead, in this study we use a simple approach and define the breakpoints as random variables uniformly distributed in the attributes' ranges (we randomly sample 50 sets of breakpoints).

3. EMPIRICAL ANALYSIS

3.1 Data

In the analysis we use all banks that participated in the stress tests of EBA in 2010, 2011 and 2014, for which financial data were available. In particular, the data consist of 72 banks from the test of 2010, 76 from the test of 2011 and 108 from the test of 2014 (i.e., a total of 256 cases). The banks are classified into two classes, namely those that failed to achieve a minimum threshold in terms of the common equity Tier 1 (CET1) capital ratio under the adverse scenario considered in the tests. In accordance with the scheme used in the 2014 test, we set the CET1 threshold at 5.5%. Banks with CET1 ratio below that threshold are classified as failed (they are likely to face capital shortfalls in the subsequent 2-3 years), whereas all other banks are classified as non-failed (i.e., well-capitalized banks). Under this rule, 53 cases are classified in the failed group.

Financial data for the banks in the sample were collected from the Orbis database of Bureau van Dijk, whereas information about the CET1 ratio of the banks was obtained from the reports of EBA. The definition of the financial performance attributes is based on the CAMEL (**C**apital, **A**sset quality, **M**anagement, **E**arnings, **L**iquidity) framework, which is commonly used for assessing bank soundness. In particular, the following ratios are used in the analysis: (i) CET1 ratio (for capital adequacy), (ii) non-performing loans / gross loans (NPL, for asset quality), (iii) cost to income ratio (CIR, for management), (iv) return on assets (ROA, for earning power), and (v) liquid assets / deposits & short-term funding (LIQ, for liquidity). Table 1 presents the means of the ratios for the two classes and their area under the receiver operating characteristic curve (AUROC), which represents the probability that a non-failed bank outperforms a failed one (according to each financial ratio). All ratios are statistically significant in discriminating the two classes of banks at the 1% level according the non-parametric Mann-Whitney test.

Table 1 Means of financial ratios and their discriminating power (AUROC)

	Class means		
	Non-failed	Failed	AUROC
CET1	11.61	7.55	0.879
NPL	6.5	12.89	0.736
CIR	59.33	64.60	0.627
ROA	0.42	-0.53	0.723
LIQ	34.63	16.33	0.736

3.2 Results

The robust MCDA methodology described in section 2 was applied to the abovementioned data set. For the purposes of the analysis, we applied the smoothing algorithm with the robust linear program formulation using different settings for the number of breakpoints (1, 3, 5) used for modeling the marginal value functions. An illustration of the obtained smooth marginal value functions, Figure 2 presents the results for the liquidity ratio. Except for the aggregate (smooth) functions, the 95% confidence bands (estimated from the 50 sets of breakpoints used in the averaging/smoothing process) are also depicted in with the dashed lines. In addition, we also present the functions resulting from the piecewise linear modeling scheme (dotted line). It is evident that as the number of breakpoints increase, the piecewise linear model becomes more complex, difficult to interpret, and (possible) sensitive to the data. The smoothed functions, on the other hand, are quite stable and robust to the specification of the number of breakpoints.

Figure 2 Examples of smooth and piecewise linear marginal value functions (LIQ ratio)

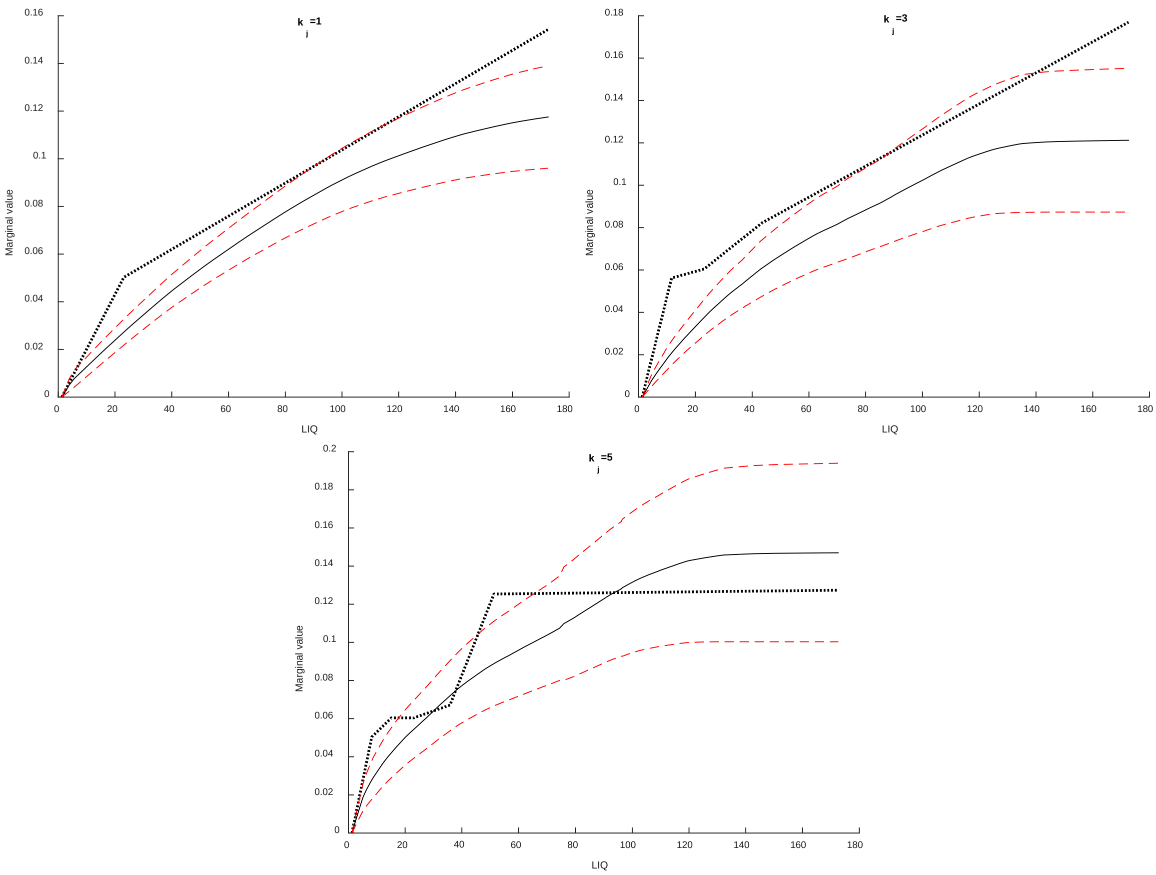


Table 2 presents the trade-offs of the financial ratios. It is evident that in all cases the CET1 ratio is the strongest variable for analyzing the financial soundness of the banks, followed by ROA, and liquidity. On the other hand, the cost to income ratio is the weakest attribute. These results are in accordance with the existing literature and other systems for assessing the soundness of banks, which emphasize the importance of capital adequacy, profitability, and liquidity, while given management efficiency a lower weight (see, for instance, Moody's, 2007).

Table 2 The trade-offs of the attributes

	Breakpoints		
	1	3	5
CET1	0.581	0.533	0.454
NPL	0.105	0.114	0.141
CIR	0.068	0.088	0.101
ROA	0.129	0.145	0.157
LIQ	0.118	0.121	0.147

To examine the discriminating power of the models, we employed follow a bootstrapping approach based 1,000 bootstrap tests. Table 3 summarizes the out-of-the-bootstrap results that focus on the generalizing (predictive) ability of the models. The results are summarized in terms of the average performance over five criteria: (i) the accuracy rate for the class of well-capitalized banks (non-falied, ACC_{NF}), (ii) the accuracy rate for the failed banks (ACC_F), the average classification accuracy (ACA, the average of ACC_{NF} and ACC_F), (iii) the overall accuracy rate (OCA), (iv) AUROC, and (v) the Kruskal-Wallis distance (the maximum difference between the cumulative probability functions of the global values/scores of the banks in the two classes). We report results for the standard UTADIS method with piecewise linear marginal value functions (UTADIS), its robust version (RUTADIS), as well as the one with the proposed smoothing procedure (RUTADIS-S). The results of logistic regression (LR) are also reported for comparison purposes (LR is widely used in finance and banking research for building classification and prediction models).

Table 3 Classification accuracy results (out-of-the-bootstrap averages)

Breakpoints		ACC_{NF}	ACC_F	ACA	OCA	AUROC	KS
0	RUTADIS	0.871	0.859	0.865	0.868	0.914	0.781
	UTADIS	0.887	0.843	0.865	0.878	0.911	0.782
1	RUTADIS-S	0.885	0.857	0.871	0.879	0.920	0.794
	RUTADIS	0.907	0.838	0.873	0.893	0.921	0.793
	UTADIS	0.892	0.820	0.856	0.877	0.902	0.764
3	RUTADIS-S	0.901	0.850	0.875	0.890	0.923	0.801
	RUTADIS	0.905	0.827	0.866	0.889	0.917	0.781
	UTADIS	0.900	0.807	0.853	0.881	0.894	0.759
5	RUTADIS-S	0.909	0.850	0.880	0.897	0.925	0.806
	RUTADIS	0.905	0.816	0.861	0.887	0.915	0.772
	UTADIS	0.901	0.807	0.854	0.881	0.896	0.760
	LR	0.861	0.854	0.857	0.859	0.902	0.769

With a linear modeling scheme (no breakpoints) both UTADIS and RUTADIS provide very similar results and they outperform LR. With more complex additive models, the performance of UTADIS appears slightly worse than RUTADIS. The results of RUTADIS, however, also follow a decreasing trend when the number of breakpoints increases (i.e., for more complex models). On the other hand, the results with the smoothing procedure (RUTADIS-S) appear quite robust to different specifications for the number of breakpoints. In fact, with more breakpoints the results slightly improve.

Table 4 presents further results about the robustness of the results, in terms of the trade-off constant of the financial ratios (averages over the 1,000 bootstrap runs). Regarding the estimates with the proposed robust and smooth approach, the bootstrap estimates are quite similar to the results from the full sample (Table 2). The stability of the estimates much higher (lower coefficients of variation) compared to the other approaches. In fact, even for the more complex models defined by multiple breakpoints, the stability of the estimates is not much different (in fact it is even better) than the simplest models with a single breakpoint. On the other hand, the variation of the estimates with the other approaches increases considerably with more complex instances, particularly for the standard version of UTADIS. These results clearly support the introduction of the robust version of UTADIS and its smooth variant introduced in this analysis.

Table 4 Bootstrap estimates for the trade-off constants of the ratios (averages and coefficients of variation in parentheses)

	One breakpoint			Three breakpoints			Five breakpoints		
	RUTADIS-S	RUTADIS	UTADIS	RUTADIS-S	RUTADIS	UTADIS	RUTADIS-S	RUTADIS	UTADIS
CET1	0.542 (0.142)	0.486 (0.217)	0.435 (0.405)	0.460 (0.140)	0.505 (0.247)	0.395 (0.566)	0.414 (0.138)	0.474 (0.291)	0.428 (0.609)
NPL	0.101 (0.426)	0.118 (0.526)	0.080 (0.648)	0.130 (0.377)	0.135 (0.695)	0.113 (1.248)	0.147 (0.389)	0.137 (0.643)	0.136 (1.474)
CIR	0.073 (0.565)	0.085 (0.503)	0.053 (0.831)	0.100 (0.463)	0.091 (0.539)	0.051 (1.158)	0.111 (0.430)	0.110 (0.559)	0.048 (1.356)
ROA	0.153 (0.450)	0.153 (0.837)	0.229 (0.766)	0.167 (0.433)	0.119 (0.900)	0.209 (1.200)	0.176 (0.404)	0.105 (0.706)	0.188 (1.386)
LIQ	0.131 (0.331)	0.157 (0.439)	0.203 (0.701)	0.144 (0.271)	0.150 (0.497)	0.232 (1.067)	0.153 (0.262)	0.174 (0.453)	0.201 (1.202)

4. CONCLUSIONS

The vulnerabilities of the global financial system to systemic crisis calls for an enhancement of the current supervisory practices. In the banking sector, the recently established formal procedures for stress testing the capital adequacy of banks is a positive step. Monitoring and analyzing the results of such tests can provide useful information to multiple stakeholders.

In this study, a robust multicriteria approach was introduced and applied to construct decision models that can be provide early warning signals for possible future capital shortfalls that banks may face. The models were constructed using as input the outcomes of the EBA stress tests. The results show that the proposed MCDA approach provides models of high discriminating power. Furthermore, the robust approach improves the quality (accuracy and stability) of the results.

Further research can be directly to a number of directions. For instance, the set performance attributes could be enriched with non-financial data (e.g., corporate governance indicators), macro-economic (country-specific) variables, as well as data from the financial markets (e.g., stock prices, CDS prices, etc.). On the methodological side other MCDA methods could be considered too, such as rule-based models and outranking techniques. Finally, implementing such techniques into integrated decision support systems could be of major help to end-users (supervisors, managers of banks, etc.).

ACKNOWLEDGEMENT

This work was partially supported by the Ministry of Education, Religious Affairs, Culture and Sports of Greece and the European Social Fund, under the Grant THALIS, MIS 377350 “Methodological approaches for studying robustness in multiple criteria decision making projects” for the first and second in order authors.

REFERENCES

Doumpos, M., Zopounidis, C. (2002). Multicriteria Decision Aid Classification Methods. Springer, New York.

Doumpos, M., Zopounidis, C. (2007). Regularized estimation for preference disaggregation in multiple criteria decision making. Computational Optimization and Applications 38(1), 61-80.

Doumpos, M., Zopounidis, C. (2014). Multicriteria Analysis in Finance. Springer, Berlin Heidelberg.

Doumpos, M., Zopounidis, C., Golfinopoulou, V. (2007). Additive support vector machines for pattern classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B 37(3), 540-550.

Jacquet-Lagrèze, E., Siskos, Y. (2001). Preference disaggregation: Twenty years of MCDA experience. European Journal of Operational Research 130(2), 233-245.

Moody’s (2007). Bank Financial Strength Ratings: Global Methodology.

Morgan, D.P., Peristiani, S., Savino, V. (2014). The information value of the stress test. Journal of Money, Credit and Banking 46(7), 1479-1500.

Petrella, G., Resti, A. (2013). Supervisors as information producers: Do stress tests reduce bank opaqueness? Journal of Banking and Finance 37(12), 5406-5420.

Tikhonov, A., Goncharsky, A., Stepanov, V., A.G., Y. (1995). Numerical Methods for the Solution of Ill-Posed Problems. Springer, Dordrecht.