

The Robustness Concern in Preference Disaggregation Approaches for Decision Aiding: An Overview

Michael Doumpos and Constantin Zopounidis

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, 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].

M. Doumpos (✉)

School of Production Engineering and Management, Technical University of Crete,
University Campus, 73100 Chania, Greece

e-mail: mdoumpos@dpem.tuc.gr

C. Zopounidis

School of Production Engineering and Management, Technical University of Crete,
University Campus, 73100 Chania, Greece

Audencia Group, Nantes School of Management, Nantes, France

e-mail: kostas@dpem.tuc.gr

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: (1) the way imperfect knowledge is treated, (2) the inappropriate preferential interpretation of certain types of data (e.g., transformations of qualitative attributes), (3) the use of modeling parameters to grasp complex aspects of reality, and (4) the introduction of technical parameters with no concrete meaning. An recent example of robustness in the context of multi-objective 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 (PDA) 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 Sect. 4 overviews the different approaches that have been proposed to obtain robust recommendations and models in PDA. Section 5 presents the statistical learning perspective and discusses its connections to the MCDA disaggregation framework. Finally, Sect. 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) $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots\}$ 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. 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 [19]. 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 $\hat{\beta}^*$ that approximate, as accurately as possible, the actual preferential system of the DM as represented in the unknown set of parameters β , i.e.:

$$\hat{\beta}^* = \arg \min_{\hat{\beta} \in \mathcal{A}} \|\hat{\beta} - \beta\| \quad (1)$$

where \mathcal{A} is a set of feasible values for the parameters $\hat{\beta}$. With the obtained parameters, the evaluations performed with the corresponding decision model $f_{\hat{\beta}^*}$ 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:

$$\hat{\beta}^* = \arg \min_{\hat{\beta} \in \mathcal{A}} L[\mathcal{D}(X'), \hat{\mathcal{D}}(X')] \quad (2)$$

where $\hat{\mathcal{D}}(X')$ denotes the recommendations of the model $f_{\hat{\beta}}$ for the alternatives in X' and $L(\cdot)$ is a function that measures the differences between $\mathcal{D}(X')$ and $\hat{\mathcal{D}}(X')$.

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 be 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 (nonnegative) 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 pre-defined 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, \dots, 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 $t_1 > t_2 \dots > t_{q-1}$ are thresholds that distinguish the groups. Alternative classification rules can also be employed such as the example-based approach of Greco et al. [21] 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 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:

$$\begin{aligned} \min \quad & \sigma_1 + \sigma_2 + \dots + \sigma_m \\ \text{s.t.} \quad & \sum_{k=1}^n w_k [v_k(x_{ik}) - v_k(x_{i+1,k})] + \sigma_i - \sigma_{i+1} \geq \delta \quad \forall \mathbf{x}_i \succ \mathbf{x}_{i+1} \\ & \sum_{k=1}^n w_k [v_k(x_{ik}) - v_k(x_{i+1,k})] + \sigma_i - \sigma_{i+1} = 0 \quad \forall \mathbf{x}_i \sim \mathbf{x}_{i+1} \\ & w_1 + w_2 + \dots + w_n = 1 \\ & v_k(x_{ik}) - v_k(x_{jk}) \geq 0 \quad \forall x_{ik} \geq x_{jk} \\ & v_k(x_{k*}) = 0, v_k(x_k^*) = 1 \quad k = 1, \dots, n \\ & w_k, v_k(x_{ik}), \sigma_i \geq 0, \quad \forall i, k \end{aligned} \tag{6}$$

where $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ and $\mathbf{x}_* = (x_{1*}, \dots, x_{n*})$ 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 $\sigma_1, \dots, \sigma_m$, 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 are 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:

$$\begin{aligned}
 \min \quad & \sum_{\ell=1}^q \frac{1}{m_\ell} \sum_{\mathbf{x}_i \in G_\ell} (\sigma_i^+ + \sigma_i^-) \\
 \text{s.t.} \quad & \sum_{k=1}^n w_k v_k(x_{ik}) + \sigma_i^+ \geq t_\ell + \delta & \forall \mathbf{x}_i \in G_\ell, \ell = 1, \dots, q-1 \\
 & \sum_{k=1}^n w_k v_k(x_{ik}) - \sigma_i^- \leq t_\ell - \delta & \forall \mathbf{x}_i \in G_\ell, \ell = 2, \dots, q \\
 & t_\ell - t_{\ell+1} \geq \varepsilon & \ell = 1, \dots, q-2 \\
 & w_1 + w_2 + \dots + w_n = 1 \\
 & v_k(x_{ik}) - v_k(x_{jk}) \geq 0 & \forall x_{ik} \geq x_{jk} \\
 & v_k(x_{k*}) = 0, v_k(x_k^*) = 1 & k = 1, \dots, n \\
 & w_k, \sigma_i^+, \sigma_i^- \geq 0 & \forall i, k
 \end{aligned} \tag{7}$$

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, \dots, 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 piecewise 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 a 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 [41]). 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.

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 be 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. These may involve the existence of noisy data, outliers, the existence of correlated criteria, etc. [12].

Table 1 An illustrative classification problem

Alternatives	Criteria			Classification
	x_1	x_2	x_3	
x_1	7	1	8	G_1
x_2	4	5	8	G_1
x_3	10	4	2	G_1
x_4	2	4	1	G_2
x_5	4	1	1	G_2
x_6	1	2	5	G_2

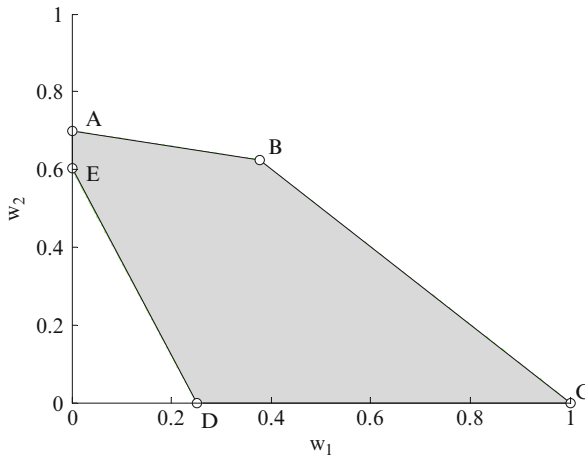


Fig. 1 The feasible set for the criteria trade-offs that are compatible with the classification of the example data of Table 1

The 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 nonlinear 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, Fig. 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 nonlinear 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

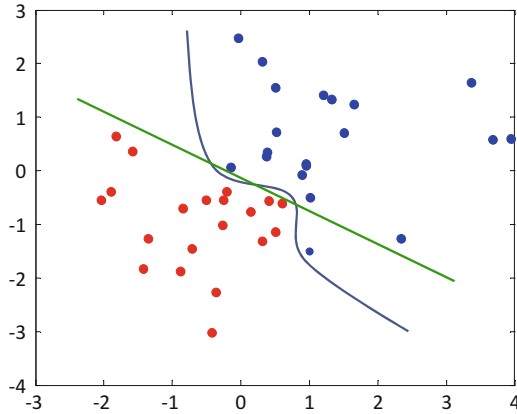


Fig. 2 A linear vs a nonlinear classification model

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].

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 the other types of decision models with some simplifications on the parameters that are inferred (see, for example, [40]).

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 Sect. 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 minimizing 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. [20] 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 [21] as well as for outranking models [10, 22] and nonadditive 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 [23].

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 Sect. 3 and its graphical illustration in Fig. 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.

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

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	C
min w_1	E	A	E	A	E	A	E	A	A
max w_2	A	A	A	A	A	A	A	A	A
min w_2	D	D	C	C	D	D	C	C	C
max w_3 (min $w_1 + w_2$)	D	D	D	D	D	D	D	D	D
min w_3 (max $w_1 + w_2$)	B	B	B	B	C	C	C	C	C
Centroid solutions									
w_1	0.31	0.31	0.44	0.44	0.42	0.42	0.54	0.54	0.54
w_2	0.32	0.34	0.32	0.34	0.22	0.23	0.22	0.23	0.23
w_3	0.37	0.35	0.24	0.23	0.37	0.35	0.24	0.23	0.23

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 Sect. 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 nonlinear 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 nonlinear character, the proposed optimization model is easy to solve with existing iterative algorithms. In a different framework, Greco et al. [23] 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 [20]. 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 $P(y | \mathbf{x})$. 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 $f_\beta : X \rightarrow Y$, 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:

$$\int L[y, f_\beta(\mathbf{x})] dP(\mathbf{x}, y) \quad (8)$$

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 $P(\mathbf{x}, y) = P(\mathbf{x})P(y | \mathbf{x})$ 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 $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}$, which are assumed to be generated independently from this unknown distribution. Thus, the objective (8) is substituted by an empirical risk estimate:

$$\frac{1}{m} \sum_{i=1}^m L[y_i, f_\beta(\mathbf{x}_i)] \quad (9)$$

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.

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 use 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:

$$\|\mathbf{Ax} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}\|^2 \quad (10)$$

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 function², the expected error $E(\beta)$ of a decision model defined by some parameters β , is bounded (with probability $1 - \alpha$) by:

$$E(\beta) \leq E_{\text{emp}}(\beta) + \sqrt{\frac{h[\log(2m/h) + 1] - \log(\alpha/4)}{m}} \quad (11)$$

where E_{emp} is the empirical error of the model as defined by Eq. (9) and h is the Vapnik–Chervonenkis dimension, which represents the complexity of the model.

² Although this is not a restricted assumption, as the theory is general enough to accommodate other loss functions as well.

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 $f(\mathbf{x}) = \mathbf{w}\mathbf{x} - \gamma$ 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]:

$$h \leq \min\{L^2 R^2, n\} + 1 \quad (12)$$

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:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \sigma_i \\ \text{s.t.} \quad & y_i(\mathbf{w}\mathbf{x}_i - \gamma) + \sigma_i \geq 1 \quad \forall i = 1, \dots, m \\ & \sigma_i \geq 0 \quad \forall i = 1, \dots, m \\ & \mathbf{w}, \gamma \in \mathbb{R} \end{aligned} \quad (13)$$

The objective function of this problem is in accordance with the Tikhonov regularization function (10). In particular, the sum of classification errors $\sigma_1, \dots, \sigma_m$ is used as a substitute for the error term $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ 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 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:

$$\begin{aligned}
& \min && f(\mathbf{x}) \\
& \text{s.t.} && g_i(\mathbf{x}, \mathbf{u}_i) \leq \mathbf{0} \quad \forall \mathbf{u}_i \in \mathcal{U}_i, i = 1, \dots, m \\
& && \mathbf{x} \in \mathbb{R}
\end{aligned} \tag{14}$$

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:

$$\begin{aligned}
& \min && \mathbf{c}^\top \mathbf{x} \\
& \text{s.t.} && \mathbf{a}_i^\top \mathbf{x} \leq b_i \quad \forall \mathbf{a}_i \in \mathcal{U}_i, i = 1, \dots, m \\
& && \mathbf{x} \in \mathbb{R}
\end{aligned} \tag{15}$$

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, 65]). Caramanis et al. [8] as well as Xu and Mannor [64] 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 $V(x) = \mathbf{w}\mathbf{x}$ 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 $w(\mathbf{x}_i - \mathbf{x}_j) \geq 1$ 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 functions using the L_1 norm for the vector of parameters. 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. Angilella, S., Greco, S., Matarazzo, B.: Non-additive robust ordinal regression: a multiple criteria decision model based on the Choquet integral. *Eur. J. Oper. Res.* **201**(1), 277–288 (2010)
2. Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V.: Support vector clustering. *J. Mach. Learn. Res.* **2**, 125–137 (2001)
3. Bertsimas, D., Brown, D.B., Caramanis, C.: Theory and applications of robust optimization. *SIAM Rev.* **53**(3), 464–501 (2011)
4. Beuthe, M., Scannella, G.: Comparative analysis of UTA multicriteria methods. *Eur. J. Oper. Res.* **130**(2), 246–262 (2001)
5. Bous, B., Fortemps, Ph., Glineur, F., Pirlot, M.: ACUTA: a novel method for eliciting additive value functions on the basis of holistic preference statements. *Eur. J. Oper. Res.* **206**(2), 435–444 (2010)
6. Bredensteiner, E.J., Bennett, K.P.: Multicategory classification by support vector machines. *Comput. Optim. Appl.* **12**(1–3), 53–79 (1999)
7. Butler, J., Jia, J., Dyer, J.: Simulation techniques for the sensitivity analysis of multi-criteria decision models. *Eur. J. Oper. Res.* **103**(3), 531–546 (1997)
8. Caramanis, C., Mannor, S., Xu, H.: Robust optimization in machine learning. In: Sra, S., Nowozin, S., Wright, S. (eds.) *Optimization for Machine Learning*, pp. 369–402. MIT Press, Cambridge (2011)
9. Chu, W., Keerthi, S.S.: Support vector ordinal regression. *Neural Comput.* **19**(3), 792–815 (2007)
10. Corrente, S., Greco, S., Słowiński, R.: Multiple criteria hierarchy process with ELECTRE and PROMETHEE. *Omega* **41**(5), 820–846 (2013)
11. Dembczynski, K., Kotłowski, W., Słowiński, R.: Additive preference model with piecewise linear components resulting from dominance-based rough set approximations. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L., Zurada, J. (eds.) *Artificial Intelligence and Soft Computing - ICAISC 2006. Lecture Notes in Computer Science*, vol. 4029, pp. 499–508. Springer, Berlin/Heidelberg (2006)
12. Doumpos, M., Zopounidis, C.: *Multicriteria Decision Aid Classification Methods*. Kluwer Academic, Dordrecht (2002)
13. Doumpos, M., Zopounidis, C.: Regularized estimation for preference disaggregation in multiple criteria decision making. *Comput. Optim. Appl.* **38**(1), 61–80 (2007)
14. Doumpos, M., Zopounidis, C.: Preference disaggregation and statistical learning for multicriteria decision support: a review. *Eur. J. Oper. Res.* **209**(3), 203–214 (2011)
15. Evgeniou, T., Boussios, C., Zacharia, G.: Generalized robust conjoint estimation. *Mark. Sci.* **24**(3), 415–429 (2005)
16. Figueira, J.R., Greco, S., Roy, B., Słowiński, R.: ELECTRE methods: main features and recent developments. In: Zopounidis, C., Pardalos, P.M. (eds.) *Handbook of Multicriteria Analysis*, pp. 51–89. Springer, New York (2010)
17. Geman, S., Bienenstock, E., Doursat, R.: Neural networks and the bias/variance dilemma. *Neural Comput.* **4**(1), 1–58 (1992)
18. Georgiev, P.G., The Luc, D., Pardalos, P.M.: Robust aspects of solutions in deterministic multiple objective linear programming. *Eur. J. Oper. Res.* **229**(1), 29–36 (2013)
19. Greco, S., Matarazzo, B., Słowiński, R.: Rough sets theory for multicriteria decision analysis. *Eur. J. Oper. Res.* **129**(1), 1–47 (2001)

20. Greco, S., Mousseau, V., Słowiński, R.: Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions. *Eur. J. Oper. Res.* **191**(2), 416–436 (2008)
21. Greco, S., Mousseau, V., Słowiński, R.: Multiple criteria sorting with a set of additive value functions. *Eur. J. Oper. Res.* **207**(3), 1455–1470 (2010)
22. Greco, S., Kadziński, M., Mousseau, V., Słowiński, R.: ELECTRE^{GKMS}: robust ordinal regression for outranking methods. *Eur. J. Oper. Res.* **214**(1), 118–135 (2011)
23. Greco, S., Kadziński, M., Słowiński, R.: Selection of a representative value function in robust multiple criteria sorting. *Comput. Oper. Res.* **38**(11), 1620–1637 (2011)
24. Grigoroudis, E., Siskos, Y.: Preference disaggregation for measuring and analysing customer satisfaction: the MUSA method. *Eur. J. Oper. Res.* **143**(1), 148–170 (2002)
25. Hand, D., Mannila, H., Smyth, P.: *Principles of Data Mining*. MIT Press, Cambridge (2001)
26. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York (2001)
27. Herbrich, R., Graepel, T., Obermayer, K.: Large margin rank boundaries for ordinal regression. In: Smola, A.J., Bartlett, P.L., Schölkopf, B., Schuurmans, D. (eds.) *Advances in Large Margin Classifiers*, pp. 115–132. MIT Press, Cambridge (2000)
28. Hites, R., De Smet, Y., Risse, N., Salazar-Neumann, M., Vincke, P.: About the applicability of MCDA to some robustness problems. *Eur. J. Oper. Res.* **174**(1), 322–332 (2006)
29. Jacquet-Lagrèze, E., Siskos, Y.: Assessing a set of additive utility functions for multicriteria decision making: the UTA method. *Eur. J. Oper. Res.* **10**, 151–164 (1982)
30. Jacquet-Lagrèze, E., Siskos, Y.: Preference disaggregation: 20 years of MCDA experience. *Eur. J. Oper. Res.* **130**, 233–245 (2001)
31. Kadziński, M., Tervonen, T.: Robust multi-criteria ranking with additive value models and holistic pair-wise preference statements. *Eur. J. Oper. Res.* **228**(1), 69–180 (2013)
32. Kadziński, M., Tervonen, T.: Stochastic ordinal regression for multiple criteria sorting problems. *Decis. Support Syst.* **55**(1), 55–66 (2013)
33. Keeney, R.L., Raiffa, H.: *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. Cambridge University Press, Cambridge (1993)
34. Kouvelis, P., Yu, G.: *Robust Discrete Optimization and Its Applications*. Kluwer Academic, Dordrecht (1997)
35. Kroese, D.P., Taimre, T., Botev, Z.I.: *Handbook of Monte Carlo Methods*. Wiley, New York (2011)
36. Kurdila, A.J., Pardalos, P.M., Zabaranin, M.: *Robust Optimization-Directed Design*. Springer, New York (2006)
37. Lahdelma, R., Salminen, P.: SMAA-2: stochastic multicriteria acceptability analysis for group decision making. *Oper. Res.* **49**(3), 444–454 (2001)
38. Lovász, L., Vempala, S.: Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *J. Comput. Syst. Sci.* **72**(2), 392–417 (2006)
39. Mangasarian, O.L., Musicant, D.R.: Robust linear and support vector regression. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(9), 950–955 (2000)
40. Mousseau, V., Figueira, J., Naux, J.-Ph.: Using assignment examples to infer weights for ELECTRE TRI method: some experimental results. *Eur. J. Oper. Res.* **130**(2), 263–275 (2001)
41. Mousseau, V., Figueira, J., Dias, L., Gomes da Silva, C., Clímaco, J.: Resolving inconsistencies among constraints on the parameters of an MCDA model. *Eur. J. Oper. Res.* **147**(1), 72–93 (2003)
42. Pahikkala, T., Waegeman, W., Tsivtsivadze, W., De Baets, B., Salakoski, T.: Learning intransitive reciprocal relations with kernel methods. *Eur. J. Oper. Res.* **206**(3), 676–685 (2010)
43. Pardalos, P.M., Hansen, P.: *Data Mining and Mathematical Programming*. American Mathematical Society, Providence (2008)
44. Roy, B.: The outranking approach and the foundations of ELECTRE methods. *Theory Decis.* **31**, 49–73 (1991)
45. Roy, B.: Robustness in operational research and decision aiding: a multi-faceted issue. *Eur. J. Oper. Res.* **200**(3), 629–638 (2010)

46. Rubinstein, R.Y.: Generating random vectors uniformly distributed inside and on the surface of different regions. *Eur. J. Oper. Res.* **10**(2), 205–209 (1982)
47. Schölkopf, B., Smola, A.J.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, Cambridge (2002)
48. Shivaswamy, P.K., Bhattacharyya, C., Smola, A.J.: Second order cone programming approaches for handling missing and uncertain data. *J. Mach. Learn. Res.* **6**, 1283–1314 (2006)
49. Siskos, J.: A way to deal with fuzzy preferences in multicriteria decision problems. *Eur. J. Oper. Res.* **10**(3), 314–324 (1982)
50. Siskos, Y., Grigoroudis, E.: New trends in aggregation-disaggregation approaches. In: Zopounidis, C., Pardalos, P.M. (eds.) *Handbook of Multicriteria Analysis*, pp. 189–214. Springer, Berlin/Heidelberg (2010)
51. Steinwart, I., Christmann, A.: *Support Vector Machines*. Springer, New York (2008)
52. Tervonen, T., Figueira, J.R., Lahdelma, R., Dias, J.A., Salminen, P.: A stochastic method for robustness analysis in sorting problems. *Eur. J. Oper. Res.* **192**(1), 236–242 (2009)
53. Tervonen, T., van Valkenhoef, G., Basturk, N., Postmus, D.: Hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. *Eur. J. Oper. Res.* **224**(3), 552–559 (2013)
54. Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., Yagola, A.G.: *Numerical Methods for the Solution of Ill-Posed Problems*. Springer, Dordrecht (1995)
55. Trafalis, T.B., Gilbert, R.C.: Robust support vector machines for classification and computational issues. *Optim. Methods Softw.* **22**(1), 187–198 (2007)
56. Vapnik, V.N.: An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **10**(5), 988–999 (1999)
57. Vapnik, V.N.: *The Nature of Statistical Learning Theory*, 2nd edn. Springer, New York (2000)
58. Vetschera, R.: A recursive algorithm for volume-based sensitivity analysis of linear decision models. *Comput. Oper. Res.* **24**(4), 477–491 (1997)
59. Vetschera, R., Chen, Y., Hipel, K.W., Marc Kilgour, D.: Robustness and information levels in case-based multiple criteria sorting. *Eur. J. Oper. Res.* **202**(3), 841–852 (2010)
60. Vincke, Ph.: *Multicriteria Decision Aid*. Wiley, New York (1992)
61. Vincke, Ph.: Robust solutions and methods in decision-aid. *J. Multi-Criteria Decis. Anal.* **8**(3), 181–187 (1999)
62. Waegeman, W., De Baets, B., Boullart, B.: Kernel-based learning methods for preference aggregation. *4OR* **7**, 169–189 (2009)
63. Xanthopoulos, P., Guarracino, M.R., Pardalos, P.M.: Robust generalized eigenvalue classifier with ellipsoidal uncertainty. *Ann. Oper. Res.* **216**(1), 327–342 (2014)
64. Xu, H., Mannor, S.: Robustness and generalization. *Mach. Learn.* **86**, 391–423 (2012)
65. Xu, H., Caramanis, C., Mannor, S.: Robustness and regularization of support vector machines. *J. Mach. Learn. Res.* **10**, 1485–1510 (2009)
66. Zopounidis, C., Doumpos, M.: Building additive utilities for multi-group hierarchical discrimination: the M.H.Dis method. *Optim. Methods Softw.* **14**(3), 219–240 (2000)
67. Zopounidis, C., Doumpos, M.: Multicriteria classification and sorting methods: a literature review. *Eur. J. Oper. Res.* **138**(2), 229–246 (2002)
68. Zopounidis, C., Pardalos, P.M.: *Handbook of Multicriteria Analysis*. Springer, Berlin/Heidelberg (2010)