Decision Support Systems 72 (2015) 33–43

Contents lists available at ScienceDirect

Decision Support Systems

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

Parametric evaluation of research units with respect to reference profiles

Miłosz Kadziński [a,*], Roman Słowiński [a,b]

[a] Institute of Computing Science, Poznań University of Technology, Poznań, Poland
[b] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form 30 January 2015
Accepted 3 February 2015
Available online 8 February 2015

Keywords:
Multiple criteria decision aiding
Reference profiles
Outranking
Inference procedure
Class cardinalities
Parametric evaluation

ABSTRACT

We introduce a method that jointly considers multiple criteria sorting and ranking. The method derives from a real-world problem of parametric evaluation of research units carried out by the Polish Ministry of Science and Higher Education. It assigns the units to three classes representing different qualities of both acquired effects and activities undertook in the evaluation period. Although units placed in the same class are guaranteed the same level of funding, they are not considered indifferent in the subsequent analysis and, thus, need to be ordered from the best to the worst in each class. A proposed outranking relation compares the units pairwise and the result is exploited so as to get a ranking of the units. The ranking is transformed to class assignments based on the attained comprehensive scores and ranks. To enhance interpretability of the results, we infer two reference profiles (artificial reference research units) separating the classes so that each class accumulates units ranked not worse than the corresponding lower profile and worse than the respective upper profile. The procedure takes into account desired cardinalities of classes, i.e., shares of units that are judged as leading, average, or weak. We discuss several procedures with different ways of inferring the reference profiles and scoring the units. We also analyze robustness of the results.

1. Introduction

Each multiple criteria decision aiding (MCDA) method is distinguished by the type of admitted preference information, the procedures used to construct a preference model, and the techniques used to work out a final recommendation [36]. Usually, these methods are designed for dealing with either ranking and choice (e.g., [8,20,37,46]) or sorting problems (e.g., [9,19,32]). In this paper, we introduce a novel MCDA method able to deal with multiple criteria sorting and ranking considered jointly. Its development has been motivated by the specific requirements of the Polish Ministry of Science and Higher Education facing a real-world problem of the parametric evaluation of research units.

Every 3 years, the ministry is carrying out an evaluation of research units applying for the statutory activity funds. This evaluation, called categorization, is performed within groups of few tens of units having similar activity profiles, called groups of joint evaluation (GJE). The categorization consists in assigning each unit of a GJE to one of three classes corresponding to different qualities of both acquired effects and activities undertook in the evaluation period. These effects and activities are represented by four independent criteria. The assignment procedure needs to respect desired cardinalities of classes, i.e., shares of alternatives that can be judged as leading, average, or weak units (see [33,40,50]). Let us emphasize that multiple criteria evaluation of education and/or research quality of different units, universities, cities, and countries is an appealing issue that has recently motivated a wide variety of studies (see, e.g., [10,23,38]).

In our application, the research units from a GJE assigned to the same class are getting the same funding level. However, they are not considered indifferent in the subsequent analysis. It is the case since the Ministry would like to differentiate over- and underperforming units within each class, to potentially distinguish a small subset of the leading research units that merit additional funds in case they prove clearly better than the remaining units, and to provide all of them with a feedback on their effectiveness against all other units. This indicates the need for ordering the units within a given GJE from the best to the worst one.

The ranking is transformed to class assignments based on the attained comprehensive scores and ranks. To enhance the interpretability of the results, some reference profiles (artificial reference research units) separating the classes need to be constructed so that each class accumulates the units not worse than the corresponding lower profile and worse than the respective upper profile. The need for inclusion of the reference profiles in the method was indicated by the representatives of the Ministry. Moreover, within the method, the existing research units need to be considered jointly with the reference profiles. This requirement implies that one cannot first rank the existing units and only then discover the profiles dividing the ranking into pre-defined proportions so that to separate the classes. Instead, the ranking and

* Corresponding author at: Institute of Computing Science, Poznań University of Technology, Piotrowo 2, Poznań 60-965, Poland. Tel.: +48 61 665 3022.
E-mail addresses: milosz.kadzinski@cs.put.poznan.pl (M. Kadziński), roman.slowinski@cs.put.poznan.pl (R. Słowiński).

http://dx.doi.org/10.1016/j.dss.2015.02.004
0167-9236/

class assignments need to be constructed simultaneously. Respecting desired class cardinalities imposes constraints on the ranks attained by the reference profiles.

Let us note that decision aiding in the context of traditional sorting problems with unsized classes is based on the absolute evaluation of each alternative to be assigned. Considering the alternatives' intrinsic values, e.g., all of them can be assigned to the same class while some other classes may remain empty [2]. When taking into account desired class cardinalities, this formulation of the sorting problem does not hold. Considering such requirements creates a partial dependence between the alternatives and implies the need for introducing a relative comparison approach. This can be achieved by integrating the constraints on the class cardinality into the assignment process. Even if the relative comparisons need to be performed, this does not contradict, however, the interpretability of the pre-defined and ordered decision classes. First, it is the particular decision aiding context that provides constraints on the size of the classes. Second, the definition of decision classes in sorting problems is first and foremost related to the way in which alternatives assigned to each class would be further processed. This treatment needs to be the same for all alternatives assigned to the same class, which holds for our problem in the phase related to granting the funds.

When comparing the alternatives in a pairwise fashion with respect to their performances on all criteria, we wish to avoid compensatory aggregation of scientific achievements measured on different scales. Thus, each criterion is characterized by the following parameters: its weight, expressing its relative importance with respect to other criteria, as well as by its indifference and preference thresholds corresponding to the differences between performances of units compared pairwise on this criterion that are negligible or significant, respectively. In other words, the indifference and preference thresholds permit to discriminate between indifference, weak preference, and strict preference in a pairwise comparison of units on this criterion. The comparison of a pair of units on all criteria is then summarized by a valued outranking relation defined in a specific way.

The rank of each unit (including existing research units and reference profiles) and, thus, the corresponding assignment is determined by its comprehensive score resulting from exploiting the outranking relation on the set of all units using the net flow score (NFS) procedure (see, e.g., [4,45]). Generally speaking, this procedure assigns to each alternative $a \in A$ a "measure of its desirability" by aggregating arguments which are in favor of its strength and weakness. We discuss different scoring procedures, which may be divided into two groups. On the one hand, a unit may get a score of one when it outranks another unit in the pairwise comparison, or no score, otherwise. Alternatively, it may be assigned a score between zero and one, corresponding to the degree of credibility of the outranking. In any case, a comprehensive score of each unit is obtained as the sum of scores corresponding to the outranking of this unit over all the others. The comprehensive score thus represents the relative power of a unit derived from its pairwise comparisons with all remaining units. Since the existing research units and reference profiles are considered jointly in the ranking procedure, let us emphasize that each existing research unit (reference profile) is compared against all reference profiles (existing units) and the remaining existing units (reference profiles).

Let us remind that reference profiles have been already used in different contexts in MCDA. For example, in the ELECTRE Tri sorting method (see, e.g., [48,14]), the class profiles are interpreted as bounds between the classes. Traditionally, these profiles had to be provided directly by the decision maker (DM), but various elicitation techniques for admitting indirect preference information have been proposed. In particular, Mousseau and Słowiński [41] suggest to infer the ELECTRE Tri preference model parameters from the assignment examples given by the DM, using non-linear optimization. Further, Ngo The and Mousseau [42] use mixed-integer linear programming (MILP) to infer these class profiles, considering other parameters as fixed. Moreover, Cailloux et al. [7] propose elicitation procedure to infer class profiles from assignment examples provided by multiple DMs. On the other hand, in ELECTRE Tri-C [3] and ELECTRE Tri-rC [35], the alternatives are not compared against the class boundaries but rather with characteristic profiles that contain the representative description of each class. Finally, Rolland [43] introduce decision rules using reference profiles (levels) for multiple criteria ranking. The results show that employing reference levels overcomes the usual weakness of the ranking methods based on pairwise comparisons, which is the sensitivity of the ranking to the change of the considered set of alternatives. The disaggregation approach for inferring these profiles in an indirect way is discussed by Zheng [49].

The organization of the paper is as follows. In the next section, we introduce notation that will be used along the paper. The decision aiding process with the proposed method is described in Section 3. The details of mathematical preference modeling underlying the introduced approach are outlined in Sections 4 and 5. They concern the definition of the employed model, procedures for deriving recommendation with the use of reference profiles selected according to some pre-defined rules, as well as algorithms for analyzing robustness of the suggested recommendation. The use of the presented method is illustrated on a problem of parametric evaluation of research units in Poland (see Section 6). Although the study consists in assigning the units to three classes, when introducing the method, we discuss a more general case with any number of classes greater than one. The last section concludes the paper.

2. Notation and basic concepts

We shall use the following notation:

- $A = \{a_1, a_2, \ldots, a_i, \ldots, a_n\}$—a finite set of n alternatives (research units);
- $G = \{g_1, g_2, \ldots, g_j, \ldots, g_m\}$—a finite set of m evaluation criteria, $g_j : A \to \mathbb{R}$ for all $j \in J = \{1, 2, \ldots, m\}$;
- $X_j = \{x_j \in \mathbb{R} : g_j(a_i) = x_j, a_i \in A\}$—the set of all different evaluations on g_j, $j \in J$; we assume, without loss of generality, that the greater $g_j(a_i)$, the better alternative a_i on criterion g_j, for all $j \in J$;
- $x_j^1, x_j^2, \ldots, x_j^{n_j(A)}$—the ordered values of X_j, $x_j^k < x_j^{k+1}$, $k = 1, \ldots, n_j(A) - 1$, where $n_j(A) = |X_j|$ and $n_j(A) \leq n$;
- C_h, $h = 1, \ldots, p$—pre-defined preference ordered classes such that C_{h+1} is preferred to C_h, $h = 1, \ldots, p - 1$; $H = \{1, 2, \ldots, p\}$;
- $R = \{r_1, \ldots, r_{p-1}\}$—reference profiles separating the classes; they are unknown a priori and need to be constructed according to some rules;
- $B = A \cup R$—a set of existing alternatives and reference profiles which are all treated equally in the ranking procedure.

Outranking relation is a preference model intended to represent preferences of a DM on a set of alternatives by a pairwise comparison function:

$$s(g_1(a), g_1(b), \ldots, g_m(a), g_m(b)) : \mathbb{R}^{2m} \to \mathbb{R}, \quad \text{for } a, b \in A.$$

In this study, we adopt the procedure for construction of the outranking relation used in the PROMETHEE method (see, e.g., [5,6,17,18]). PROMETHEE and its further extensions have proven to be well suited for real-world multiple criteria problems in various areas such as, e.g., stock trading [1], equipment selection [47], bank rating [15], infrastructure assessment [21], energy market [24], outsourcing in information systems [11], or climate protection [39]. In this method, for each criterion g_j, $j = 1, \ldots, m$, one considers a preference function $\pi_j(a, b)$, such that for all $a, b \in B$:

$$\pi_j(a, b) = F_j\Big(d_j(a, b)\Big) \in [0, 1],$$

where $d_j(a, b) = g_j(a) - g_j(b)$.

Let us denote by w_j the weight assigned to criterion g_j, $j = 1, \ldots, m$, expressing relative importance of g_j in set G. Without loss of generality, we assume that the weights of criteria sum up to one, i.e., $\sum_{j=1}^{m} w_j = 1$. Remark that these weights are not interpreted here as weights of criteria in a weighted sum of criteria. In the latter preference model, alike in the whole family of compensatory preference models it belongs, the weights play the role of substitution rates among criteria. In the outranking model, the weights are not multiplied by performances on the corresponding criteria but, instead, they underline the relative strength of criteria in a voting-like procedure for or against outranking of one alternative over another.

The criteria are also associated with indifference q_j and preference p_j thresholds. For consistency, $p_j \geq q_j \geq 0$, $j = 1, \ldots, m$. Knowing the difference between evaluations of alternatives $a, b \in B$ on a particular criterion g_j, one is able to represent situations of weak or strict preference and indifference among a and b on g_j. The type of relation implies value assigned to a marginal preference function $\pi_j(a, b)$:

$$\pi_j(a,b) = \begin{cases} w_j, & \text{if } g_j(a) - g_j(b) \geq p_j, \\ w_j\left[\left(g_j(a) - g_j(b)\right) - q_j\right] \big/ \left(p_j - q_j\right) & \text{if } p_j \geq g_j(a) - g_j(b) \geq q_j, \\ 0, & \text{if } g_j(a) - g_j(b) \leq q_j. \end{cases} \quad (1)$$

Note that in case a is weakly preferred to b, we may employ different non-linear scoring schemes. In particular, we may define a middle level for the preference function, e.g., if $p_j \geq g_j(a) - g_j(b) \geq q_j$, then $\pi_j(a, b) = 0.5w_j$. Furthermore, if the DM provides a preference threshold p_j only, then:

$$\pi_j(a,b) = \begin{cases} w_j, & \text{if } g_j(a) - g_j(b) \geq p_j, \\ 0, & \text{if } g_j(a) - g_j(b) < p_j. \end{cases} \quad (2)$$

To express the degree in which a is preferred to b on all criteria, we will refer to an aggregated preference index:

$$\pi(a,b) = \sum_{j=1}^{m} \pi_j(a,b) \ \text{ for all } \ (a,b) \in B \times B.$$

Remark that $\pi(a, b) \in [0, 1]$, where $\pi(a, b) = 0$ if $g_j(a) - g_j(b) \leq q_j$, $j = 1, \ldots, m$ (a is at most indifferent to b on all criteria), and $\pi(a, b) = 1$ if $g_j(a) - g_j(b) \geq p_j$, $j = 1, \ldots, m$ (a is strictly preferred to b on all criteria).

3. Decision aiding with the proposed approach

Parametric evaluation of research units can be aided with the proposed approach through the five-step process illustrated in Fig. 1.

Step 1. The process begins by defining the problem: a set of alternatives (research units) A (i.e., a group of joint evaluation), a set of criteria G, the units' performances on the criteria representing

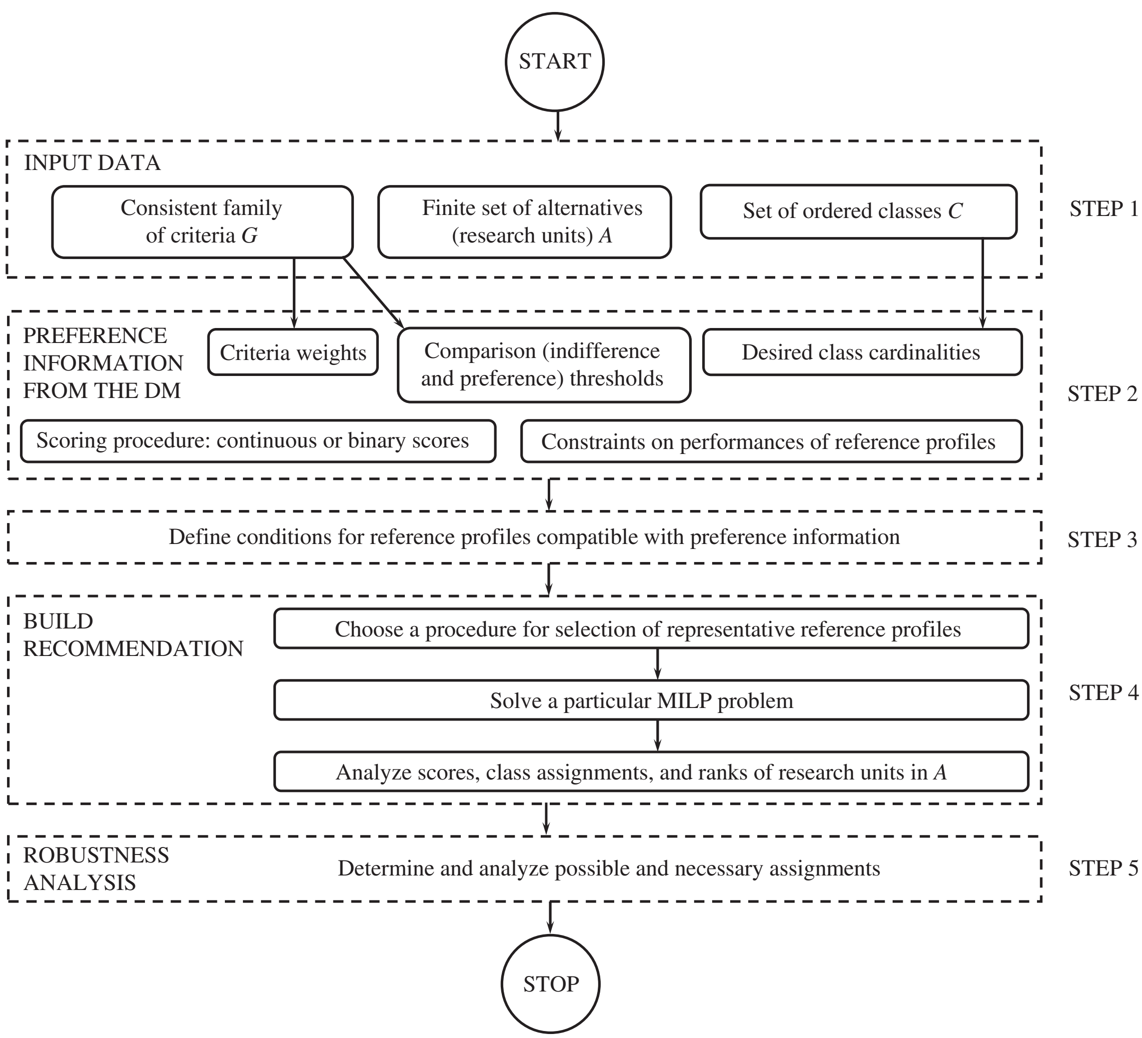

Fig. 1. Decision aiding process for the proposed approach.

the quality of their effects in the evaluation period, and a set of ordered classes C.

Step 2. Then in Step 2, the preference information is elicited. To enable comparison of units with respect to their performances, we assume that the DM provides for each criterion $g_j \in G$ its weight w_j as well as the indifference q_j and preference p_j thresholds. Let us remind that w_j should be interpreted as the voting power of g_j, which depends neither on the range of the criterion scale nor on the encoding chosen. For the problem of parametric evaluation of research units, the DM has a clear understanding of the importance of each criterion. Further, the indifference and preference thresholds are interpreted as, respectively, the greatest performance difference for which the situation of indifference holds on g_j and the smallest performance difference for which the situation of preference occurs on g_j.

Remark that instead of the two discrimination thresholds, (s)he may provide preference threshold p_j only. Such model is certainly easier to explain since a sharp transition from indifference to strict preference is more intuitive for the DMs who are not familiar with the outranking methods. Nevertheless, in some decision making situations, using pseudo-criteria with a pair of comparison thresholds is appealing. In particular, it allows accounting for a situation of weak preference where neither of the two preceding situations can be distinguished as appropriate (i.e., when there are insufficient reasons to deduce the strict preference in favor of one alternative, but they are clear and positive enough to invalidate the indifference between the alternatives). The experience of using outranking methods in real-world decision problems shows that the assumption about providing values for the indifference and preference thresholds is not unrealistic. Once their meaning is explained, the experts, being aware what is the precision of criteria, are able to indicate how much difference is negligible or significant.

Moreover, we assume that the DM specifies shares of alternatives in A that should be assigned to each class C_h, $h = 1, \ldots, p$. Let us denote the minimal and maximal bounds for such a share by $N_h^{\text{perc - min}}$ and $N_h^{\text{perc - max}}$ (in %). In case these are equal, we may denote the required share by N_h^{perc}. Remark that such constraints are not related only with the preferences of the DM but also with the particular sorting context. For our problem, they are implied by a limited budget of the Ministry granting funds for a statutory activity of research units.

Prior to discovering the reference profiles separating the classes, the DM may define constraints with respect to their performances on different criteria. These may rely either on expert knowledge about the targets that should be satisfied by a unit assigned to a given class, or on results of the statistical analysis of performances attained by the units within GJE.

Finally, the proposed approach requires to assign a comprehensive score to each unit (including existing research units and reference profiles). First, the scoring procedure constructs an outranking relation on the set of units. Then this relation is exploited to compute for each unit its comprehensive score as a sum of scores derived from unit's pairwise comparisons against all other units. We propose several exploitation procedures, out of which the DM should choose one to be used for scoring the units. On the one hand, (s)he may wish to use a procedure with binary (win/no win) scores. It assigns a single score to a unit in case it proves better when compared pairwise with some other unit, and no score otherwise. On the other hand, (s)he may employ a scheme with continuous scores corresponding to the degree of credibility of an outranking relation. In this case, a unit is rewarded for each individual aspect in which it proves its superiority over another unit. These procedures are discussed in detail in Section 4.2.

Step 3. Step 3 consists of constructing the disaggregation model for inferring reference profiles compatible with the preference information provided by the DM. This model is composed of two types of constraints:

- these concerning performances of reference profiles to be inferred and their comparison with the existing research units, and
- constraints that guarantee that the scoring procedure would work as intended and that the resulting class assignments would respect desired class cardinalities.

For clarity, we discuss these different types of constraints separately in Sections 4.1 and 4.2. Nevertheless, they are subsequently incorporated into a single model so that both the inference of reference profiles and their comparison with the existing units within the joint ranking/sorting procedure are conducted simultaneously.

The desired result of applying this procedure for the case of three decision classes is presented in Fig. 2. Let us remind that the role of reference profiles is to transform the ranking into class assignments. Precisely, each class accumulates existing research units which are scored not worse than the respective lower profile and worse than the upper profile. To respect desired class cardinalities, the construction of the profiles needs to account for the ranks they attain. This can be achieved by controlling the number of existing research units ranked at least as good and lower than each profile.

In general, there may exist more than one compatible set of reference profiles. In this perspective, in MCDA, one employs

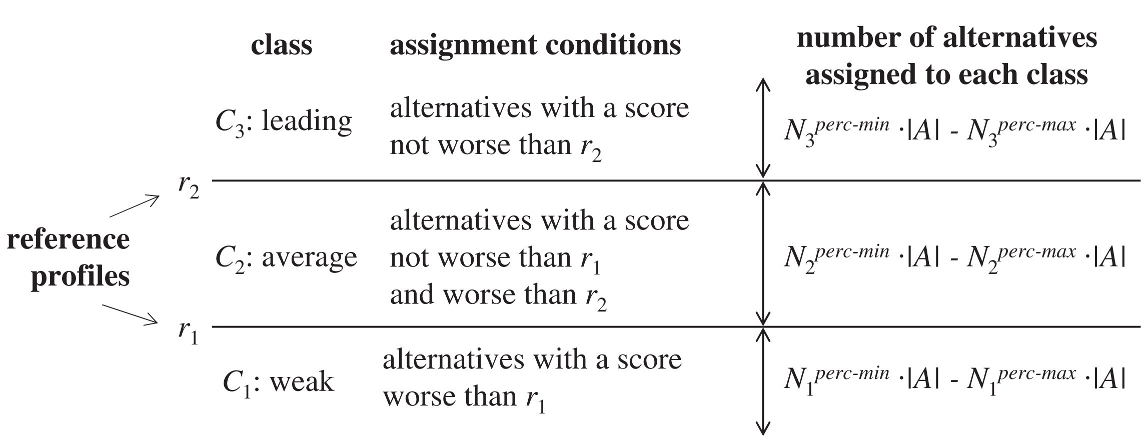

Fig. 2. Desired result of the joint ranking/sorting procedure for the case of three quality classes.

two different approaches for deriving a recommendation. One of them concerns selection of a single preference model instance that matches preferences and requirements of the DM in the "best" way (for a discussion, see, e.g., [16,28,31]). We take advantage of this approach in Step 4. The other approach, which is further employed in Step 5, takes into account all compatible preference model instances and investigates robustness of the delivered recommendation (see, e.g., [12–14,25]).

Step 4. Step 4 consists of building a recommendation with respect to a single representative set of reference profiles. We propose some pre-defined rules for its selection. In particular, the DM may wish the profiles to be either as good or as bad as possible (see Section 5.1 for details). Then the selection procedure consists in solving a single MILP problem, which leads to indicating:

- the performances of reference profiles which can be interpreted as requirements that a research unit should satisfy to be assigned to a particular class;
- the assignments of all units which determine their funding level for the next evaluation period;
- the scores of all units reflecting their desirability and effectiveness against the remaining units; these scores along with the ranks can be used to distinguish over- and underperforming units within each class.

Step 5. The analysis of a single representative set of reference profiles is surely less abstract than that of the whole set of compatible sets of reference profiles. In such a way, the DM can see a score of each unit along with the univocal recommendation. Nevertheless, the selection algorithm from Step 4 introduces some degree of arbitrariness, which may affect the results. To verify how fragile they are, in Step 5, we may conduct robustness analysis taking into account the recommendation obtained for all compatible sets of reference profiles. We suggest to focus on the possible and necessary assignments, which are confirmed by, respectively, at least one and all compatible sets (see Section 5.2 for details). On the one hand, the necessary assignment is robust, which means that the recommendation is the same whatever compatible set of reference profiles. On the other hand, the possible assignment, in case of being imprecise, reveals that the recommendation may vary if some other procedure for selection of a representative set of reference profiles was chosen.

4. Disaggregation model for inferring compatible reference profiles

In this section, we present a mathematical program for inferring reference profiles separating the classes. This model is constructed in Step 3 of the decision aiding process presented in Section 3.

4.1. Performances of reference profiles

First, we define a set of constraints concerning performances of reference profiles $g_j(r), j \in J$. We distinguish two cases depending on whether the DM provides for criterion g_j a preference threshold p_j only, or both indifference q_j and preference p_j thresholds.

$$\left.\begin{array}{l}
\pi_j(a,b) \text{ for } (a,b) \in A \times A,\ j \in J, \text{ is known and computed with (2),} \\
\text{for } r \in R,\ a \in B \text{ and } j \in J: \\
\quad \left.\begin{array}{l}
[C_1]\ v_{r,a}^j \leqslant 1/M \cdot \left(g_j(r) - g_j(a) - p_j\right) + 1, \\
[C_2]\ v_{r,a}^j \geqslant 1/M \cdot \left(g_j(r) - g_j(a) - p_j + \varepsilon\right), \\
[C_3]\ \pi_j(r,a) \leqslant v_{r,a}^j, \\
[C_4]\ \pi_j(r,a) \geqslant v_{r,a}^j + w_j - 1, \\
[C_5]\ \pi_j(r,a) \geqslant 0, \\
[C_6]\ \pi_j(r,a) \leqslant w_j, \\
[C_7]\ v_{r,a}^j \in \{0,1\},
\end{array}\right\} E(r,a) \\
\text{for } r \in R,\ a \in B \text{ and } j \in J: \\
\quad E(a,r) \text{ (corresponding to } E(r,a) \text{ with inverse positions of } a \text{ and } r), \\
[PR_1]\ g_j(r_h) \geqslant g_j(r_{h-1}) + \varepsilon,\ h = 2, \ldots, p-1, \\
[PR_2]\ x_j^{n_j(A)} \geqslant g_j^{\max}(r_h) \geqslant g_j(r_h) \geqslant g_j^{\min}(r_h) \geqslant x_j^1,\ h = 1, \ldots, p-1, \\
M \text{ is arbitrarily large positive value greater than } x_j^{n_j(A)} - x_j^1, \\
\varepsilon \text{ is an arbitrarily small positive value, e.g., } 0.001.
\end{array}\right\} E_{\text{profiles}}^{\text{pref}}$$

Let us first consider the case of a given preference threshold p_j only (see constraint set $E_{\text{profiles}}^{\text{pref}}$). Since performances $g_j(r)$, $j \in J$, are unknown, we need to relate the difference $g_j(r) - g_j(a)$, for r being a reference profile in R and a being either an alternative in A or another reference profile, with marginal preference index $\pi_j(r, a)$. This relation is represented by constraint set $E(r, a)$ in the following way. With each pair $(r, a) \in R \times B$ and criterion $g_j \in G$, we associate a binary variable $v_{r,a}^j$. It is equal to 0, if r is not preferred to a on g_j, i.e., $g_j(r) < g_j(a) + p_j$ (see constraint $[C_1]$). In case r is preferred to a on g_j, i.e., $g_j(r) \geq g_j(a) + p_j$, it is equal to 1 (see constraint $[C_2]$). Now, if $v_{r,a}^j = 1$, then marginal preference index $\pi_j(r, a)$ is equal to w_j (see constraints $[C_4]$ and $[C_6]$), while if $v_{r,a}^j = 0$, then $\pi_j(r, a)$ is equal to 0 (see constraints $[C_3]$ and $[C_5]$). This is represented graphically in Fig. 3. Additionally, we require performances of an upper profile of each class to be better than the performances of the corresponding lower profile on each criterion (see $[PR_1]$). Finally, in case the DM specified some real interval $[g_j^{\min}(r_h), g_j^{\max}(r_h)]$ allowed for the performance $g_j(r_h)$, this is respected with $[PR_2]$.

In case the DM provided both indifference q_j and preference threshold p_j, with each pair $(r, a) \in R \times B$ and criterion $g_j \in G$, we associate three binary variables: $v_{r,a}^{p,j}$, $v_{r,a}^{i,j}$, and $v_{r,a}^{w,j}$ (see constraint set $E'(r, a)$). They correspond to zones of strict preference, indifference, and weak preference between r and a, respectively. Only one of these variables may be instantiated with one, while the remaining ones are set to zero. This is guaranteed by constraint $[CV_1]$. Which binary variable is set to one is determined by a comparison of $g_j(r) - g_j(a)$ with indifference q_j and preference p_j thresholds. For example, if $g_j(r) - g_j(a) \geqslant p_j$, then $v_{r,a}^{p,j} = 1$, $v_{r,a}^{i,j} = 0$, and $v_{r,a}^{w,j} = 0$ (see $[CP_1]$ and $[CV_1]$). This, in turn, implies that $\pi_j(r, a) = w_j$ (see $[CP_2]$ and $[CP_3]$). Since $v_{r,a}^{i,j} = 0$ and $v_{r,a}^{w,j} = 0$, constraints $[CI_1 - CI_3]$ and $[CW_1 - CW_4]$ are always satisfied, being eliminated. On the other hand, if $v_{r,a}^{i,j} = 1$, then $\pi_j(r, a) = 0$, and

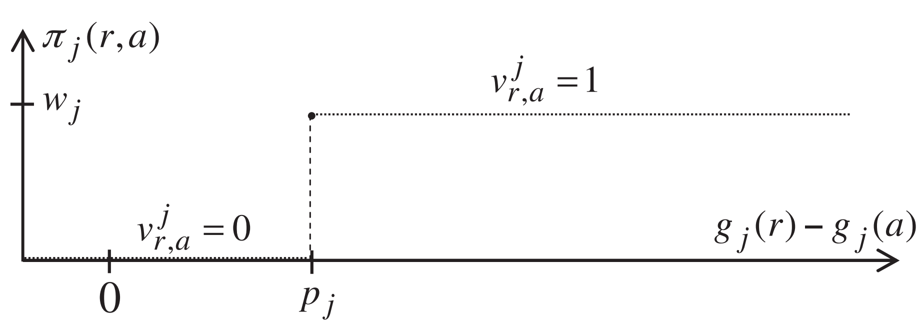

Fig. 3. Marginal preference function in case the DM provides a preference threshold only.

if $v_{r,a}^{w,j} = 1$, then $\pi_j(r, a) = w_j \cdot [(g_j(r) - g_j(a)) - q_j]/(p_j - q_j)$. This is represented graphically in Fig. 4.

$$
\left.\begin{array}{l}
\pi_j(a,b) \text{ for } (a,b) \in A \times A,\ j \in J, \text{ is known and computed with (1)}, \\
\text{for } r \in R,\ a \in B \text{ and } j \in J: \\
\left.\begin{array}{l}
[CP_1]\ g_j(r) - g_j(a) \geqslant p_j - M\left(1 - v_{r,a}^{p,j}\right), \\
[CP_2]\ \pi_j(r,a) \geqslant w_j - M\left(1 - v_{r,a}^{p,j}\right), \\
[CP_3]\ \pi_j(r,a) \leqslant w_j + M\left(1 - v_{r,a}^{p,j}\right), \\
[CI_1]\ g_j(r) - g_j(a) \leqslant q_j + M\left(1 - v_{r,a}^{i,j}\right), \\
[CI_2]\ \pi_j(r,a) \geqslant 0 - M\left(1 - v_{r,a}^{i,j}\right), \\
[CI_3]\ \pi_j(r,a) \leqslant 0 + M\left(1 - v_{r,a}^{i,j}\right), \\
[CW_1]\ g_j(r) - g_j(a) \leqslant p_j + M\left(1 - v_{r,a}^{w,j}\right), \\
[CW_2]\ g_j(r) - g_j(a) \geqslant q_j - M\left(1 - v_{r,a}^{w,j}\right), \\
[CW_3]\ \pi_j(r,a) \geqslant w_j \cdot \left[\left(g_j(r) - g_j(a)\right) - q_j\right] / \left(p_j - q_j\right) - M\left(1 - v_{r,a}^{w,j}\right), \\
[CW_4]\ \pi_j(r,a) \leqslant w_j \cdot \left[\left(g_j(r) - g_j(a)\right) - q_j\right] / \left(p_j - q_j\right) + M\left(1 - v_{r,a}^{w,j}\right), \\
[CV_1]\ v_{r,a}^{p,j} + v_{r,a}^{i,j} + v_{r,a}^{w,j} = 1, \\
[CV_2]\ v_{r,a}^{p,j}, v_{r,a}^{i,j}, v_{r,a}^{w,j} \in \{0,1\},
\end{array}\right\} E'(r,a) \\
\text{for } r \in R,\ a \in B \text{ and } j \in J: \\
\quad E'(a,r) \text{ (corresponding to } E(r,a) \text{ with inverse positions of } a \text{ and } r), \\
[PR_1], [PR_2].
\end{array}\right\} E_{\text{profiles}}^{\text{ind-pref}}
$$

4.2. Scoring procedures

Let us now discuss exemplary scoring procedures and the way of accounting for desired class cardinalities specified by the DM.

First, we refer to a scoring procedure, which compares alternatives pairwise and grants each alternative a score in the range [0, 1]. This score is equal to a comprehensive preference index $\pi(a, b)$. Then a comprehensive score of $a \in B$ is computed as $Sc(a) = \sum_{b \in B} \pi(a, b)$; see constraint set $E_{\text{scores}}^{\text{continuous}}$. This scoring procedure is represented graphically in Fig. 5a). To respect desired cardinalities for each class C_h, we need to ensure that only a pre-defined share of the whole set of research units $a \in A$ is at least as good as the reference profile r_{h-1} and worse than r_h. This is achieved with constraints $[P_3 - P_8]$. Precisely, if $Sc(a) \leqslant Sc(r_{h-1})$ and $Sc(a) < Sc(r_h)$, then the binary variable v_{a,C_h}^{COMP} is set to one. Thus, the sum of binary variables v_{a,C_h}^{COMP} for all $a \in A$ corresponds to the number of existing research units, which are ranked at least as good as reference profile r_{h-1} and worse than r_h. This sum needs to be not less than $\lceil N_h^{\text{perc - min}} \cdot n \rceil$ and not greater than $\lceil N_h^{\text{perc - max}} \cdot n \rceil$ (or equal to $\lceil N_h^{\text{perc}} \cdot n \rceil$ if the DM provided precise desired class cardinality).

$$
\left.\begin{array}{l}
\text{for } a, b \in B: \\
\quad [P_1]\ \pi(a,b) = \sum_{j=1}^{m} \pi_j(a,b), \\
\text{for } a \in B: \\
\quad [P_2]\ Sc(a) = \sum_{b \in B} \pi(a,b), \\
\left.\begin{array}{l}
\text{for } a \in A: \\
\quad [P_3]\ \sum_{h=1}^{p} v_{a,C_h}^{\text{COMP}} = 1, \\
\text{for } a \in A,\ h = 1, \ldots, p: \\
\quad \text{if } h > 1: \\
\qquad [P_4]\ Sc(a) \geqslant Sc(r_{h-1}) - M\left(1 - v_{a,C_h}^{\text{COMP}}\right), \\
\quad \text{if } h < p: \\
\qquad [P_5]\ Sc(a) + \varepsilon \leqslant Sc(r_h) + M\left(1 - v_{a,C_h}^{\text{COMP}}\right), \\
\quad [P_6]\ v_{a,C_h}^{\text{COMP}} \in \{0,1\}, \\
\text{for } h = 1, \ldots, p: \\
\quad [P_7]\ \sum_{a \in A} v_{a,C_h}^{\text{COMP}} \geq \lceil N_h^{\text{perc-min}} \cdot n \rceil, \\
\quad [P_8]\ \sum_{a \in A} v_{a,C_h}^{\text{COMP}} \leq \lfloor N_h^{\text{perc-max}} \cdot n \rfloor
\end{array}\right\} E_{\text{card}}^{\text{class}}
\end{array}\right\} E_{\text{scores}}^{\text{continuous}}
$$

A different scoring procedure assumes that alternatives are compared pairwise and the alternative which proves better is granted a score of one (see constraint set $E_{\text{scores}}^{\text{binary}}$). For example, if the strength of arguments in favor of a when compared to b, which is materialized with $\pi(a, b)$, is greater than the strength of arguments in favor of b when compared to a ($\pi(b, a)$), then binary variable $v_{a,b}$ in constraint $[W_4]$ is set to one. Otherwise, it is instantiated with zero. This scoring procedure is represented graphically in Fig. 5b). Alternatively, a may be granted a score only if arguments supporting its strength are sufficiently great (e.g., if $\pi(a, b)$ is not less than some cutting level $\lambda > 0.5$ pre-defined by the DM; see $[W_1']$). In any case, a comprehensive score $Sc(a)$ of each alternative or reference profile $a \in B$ is computed by summing up scores resulting from pairwise comparisons with all remaining $b \in B$, i.e., $Sc(a) = \sum_{b \in B} v_{a,b}$. Then desired class cardinalities are accounted analogously as in $E_{\text{scores}}^{\text{continuous}}$. The procedure with binary scores is interesting because when computing $Sc(a)$, for $a \in B$, it eliminates the undesired compensation between a single large value $\pi(a, b)$ and several small values $\pi(c, a)$, for $b, c \in B$.

$$
\left.\begin{array}{l}
\text{for } (a,b) \in A \times A: \\
\quad [W_1]\ v_{a,b} = 1, \quad \text{if } \pi(a,b) > \pi(b,a) \\
\qquad [W_1']\ (\text{or } \pi(a,b) \geq \lambda), \\
\quad [W_2]\ v_{a,b} = 0, \quad \text{if } \pi(a,b) \leq \pi(b,a) \\
\qquad [W_2']\ (\text{or } \pi(a,b) < \lambda), \\
\text{for } (a,b) \in R \times A \text{ or } A \times R: \\
\quad [W_3]\ \pi(a,b) = \sum_{j=1}^{m} \pi_j(a,b), \\
\quad [W_4]\ \pi(a,b) \geqslant \pi(b,a) + \varepsilon - M\left(1 - v_{a,b}\right), \\
\quad [W_4]\ \pi(a,b) \leqslant \pi(b,a) + M v_{a,b}, \\
\qquad [W_4']\ \left(\text{or } \pi(a,b) \geqslant \lambda - M\left(1 - v_{a,b}\right)\right), \\
\qquad [W_4']\ \left(\text{or } \pi(a,b) + \varepsilon \leqslant \lambda + M v_{a,b}\right), \\
\text{for } a \in B: \\
\quad [W_5]\ Sc(a) = \sum_{b \in B} v_{a,b}, \\
E_{\text{card}}^{\text{class}}.
\end{array}\right\} E_{\text{scores}}^{\text{binary}}
$$

Let us denote by $\mathcal{R}^{DM}$ the set of preference model instances (in particular, reference profiles) compatible with the DM's preference information. It is defined by a set of constraint $E_{\text{DM}} = E_{\text{profiles}} \cup E_{\text{scores}}$, where E_{profiles} is equivalent to $E_{\text{profiles}}^{\text{ind - pref}}$ or $E_{\text{profiles}}^{\text{pref}}$ depending on whether the DM provided indifference thresholds or not and E_{scores} is equivalent to $E_{\text{scores}}^{\text{binary}}$ or $E_{\text{scores}}^{\text{continuous}}$ depending on the selected scoring procedure.

The following variables are involved in formulation of E_{DM}:

- $g_j(r_h)$ for $h = 1, \ldots, p-1$, involved in E_{profiles} (these variables represent the performances of the reference profiles),
- $\pi_j(r, a)$ and $\pi_j(a, r)$ for $r \in R$, $a \in B$, and $j \in J$, involved in E_{profiles} (these variables represent marginal preference indices for pairs consisting of a profile and a unit),
- $v_{r,a}^j$ and $v_{a,r}^j$ used in case $E_{\text{profiles}} = E_{\text{profiles}}^{\text{pref}}$, or $v_{r,a}^{p,j}, v_{r,a}^{i,j}, v_{r,a}^{w,j}, v_{a,r}^{p,j}, v_{a,r}^{i,j}$, and $v_{a,r}^{w,j}$ used in case $E_{\text{profiles}} = E_{\text{profiles}}^{\text{ind - pref}}$ for $r \in R$, $a \in B$, and $j \in J$ (these binary variables represent either preference or its lack, or preference, weak preference, and indifference on a given criterion for pairs consisting of a profile and a unit);
- $\pi(a, r)$ and $\pi(r, a)$ for $r \in R$ and $a \in B$, involved in E_{scores} (these variables represent comprehensive preference indices for pairs consisting of a profile and a unit; note that $\pi(a, b)$ for $a, b \in A$, are constants);
- $Sc(a)$ for $a \in B$, involved in E_{scores} (these variables represent comprehensive scores attained by the units);
- v_{a,C_h}^{COMP} for $a \in A$ and $h \in H$, involved in E_{scores} (these binary variables represent assignment of a unit $a \in A$ to class C_h);
- $v_{a,r}$ and $v_{r,a}$ for $a \in A$ and $r \in R$, involved in $E_{\text{scores}}^{\text{binary}}$ (these binary variables represent superiority of a unit over a profile, or vice versa; note that $v_{a,b}$ for $a, b \in A$, are constants).

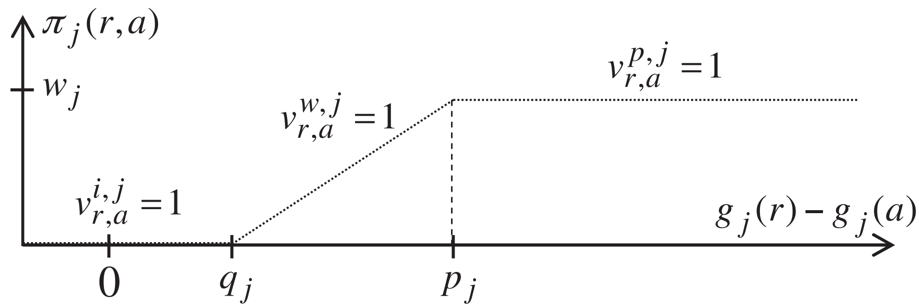

Fig. 4. Marginal preference function in case the DM provides indifference and preference thresholds.

5. Multiple criteria ranking and sorting with inferred reference profiles

In this section, we discuss procedures for both selection of a single set of representative reference profiles as well as robustness analysis. They are employed in Steps 4 and 5 of the decision aiding process presented in Section 3.

5.1. Selection of a representative set of reference profiles

When selecting a single set of representative reference profiles, one may consider different requirements. In particular, one may select the profiles which are either as good or as bad as possible. Such profiles can be interpreted as performance vectors that a research unit should attain to be assigned to a particular class. For interpretability of the results, it is also reasonable to require the performances of the profiles to be balanced with respect to the extreme performances of alternatives on different criteria. In this way, we avoid selection of the profiles, which are relatively good on some criteria, while being relatively bad on the others. This is achieved by solving problem (3).

$$\text{Maximize (or Minimize)}: \sum_{h=1}^{p-1}\delta_h+\gamma\sum_{h=1}^{p-1}\sum_{j=1}^{m}\delta_{hj} \tag{3}$$

$$\left.\begin{array}{l} E^{DM}, \\ \text{for } h=1,\ldots,p-1, j\in J: \\ \quad \delta_{hj}\leq\left(g_j(r_h)-x_j^1\right)\Big/\left(x_j^{n_j(A)}-x_j^1\right), \\ \quad \left(\text{or } \delta_{hj}\geq\left(g_j(r_h)-x_j^1\right)\Big/\left(x_j^{n_j(A)}-x_j^1\right)\right), \\ \quad \delta_h\leq\delta_{hj}, \\ \quad \left(\text{or } \delta_h\geq\delta_{hj}\right), \end{array}\right\} E^{\text{inference}}$$

where γ is an arbitrarily small positive value, e.g., 0.001.

Now, let us explain the objective function in case the selected profiles are required to be as good as possible (an explanation for the case of profiles required to be as bad as possible can be formulated analogously). A coefficient $\left(g_j(r_h)-x_j^1\right)\Big/\left(x_j^{n_j(A)}-x_j^1\right)$ for $j\in J$ and $h\in\{1,\ldots,p-1\}$, which is used in the above constraint set, represents the location of $g_j(r_h)$ on a scale delimited by the extreme performances of existing research units: $x_j^{n_j(A)}$ and x_j^1. For example, if it is equal to 0.75, the distance of $g_j(r_h)$ from x_j^1 is three times greater than its distance from $x_j^{n_j(A)}$. Apart from the variables included in E^{DM} (these are listed in Section 4.2), the constraint set $E^{\text{inference}}$ involves the following variables:

- δ_{hj} for $h=1,\ldots,p-1$ and $j\in J$, which bound the aforementioned coefficients from below; consequently, the greater δ_{hj}, the better r_h on g_j;
- δ_h for $h\in\{1,\ldots,p-1\}$, which bound the values of respective variables δ_{hj} for $j\in J$, from below; thus, the greater δ_h, the better the worst performance of r_h on any criterion g_j.

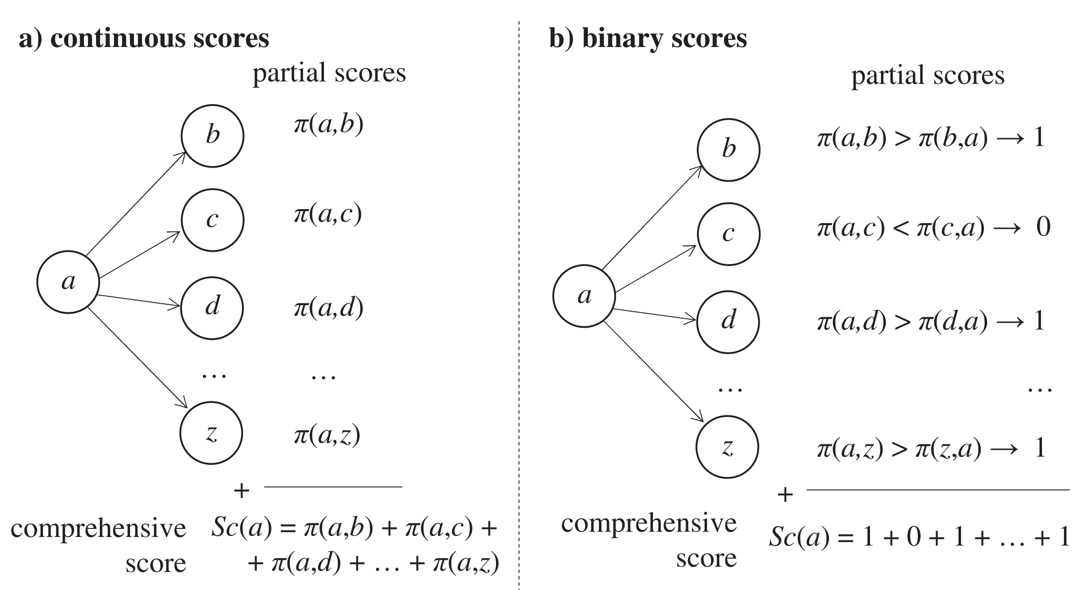

Fig. 5. Scoring procedure with (a) continuous scores and (b) binary scores.

To select the profiles that are as good as possible with relatively balanced performances on different criteria, we maximize $\sum_{h=1}^{p-1}\delta_h$, thus requiring the worst performance of each profile to be as good as possible. As a secondary target, we optimize $\sum_{h=1}^{p-1}\sum_{j=1}^{m}\delta_{hj}$, thus maximizing the profiles' performances on the individual criteria.

5.2. Robustness analysis

Robustness analysis is understood as a theoretical basis and a diversity of particular multiple criteria decision support methods that take into account internal and external uncertainties observed in the actual decision situations. In this paper, we are interested in investigating the robustness of the provided recommendation, i.e., whether it is valid for all or for the most plausible sets of model parameters. We focus on the assignment-related results. In this case, it is reasonable to consider the possible and necessary assignments. Precisely, given a set $\mathcal{R}^{\mathrm{DM}}$ of compatible preference model instances, for each alternative $a \in A$, the possible assignment $C_P(a)$ is defined as the set of indices of classes C_h for which there exists at least one compatible preference model instance assigning a to C_h, and the necessary assignment $C_N(a)$ as the set of indices of classes C_h for which all compatible preference models assign a to C_h (see [26, 32]).

Computing the possible assignment for $a \in A$ requires considering problem (4) for each $h \in H$. It verifies whether score $Sc(a)$ of $a \in A$ can be simultaneously at least as good as score of the lower profile r_{h-1} and less than score of the upper profile r_h of class C_h, in the set of compatible preference model instances (defined with the constraint set E_{DM}). We assume here that ε, which is used in both E_{DM} and newly added constraints, is a variable rather than a constant (as stated previously in Section 4.1); it is used to transform strict inequalities into non-strict ones.

$$\text{Maximize}: \quad \varepsilon \tag{4}$$

$$\left.\begin{array}{l} E_{\mathrm{DM}}, \\ Sc(a) \geqslant Sc(r_{h-1}), \ \text{ if } \ h > 1, \\ Sc(a) + \varepsilon \leqslant Sc(r_h), \ \text{ if } \ h < p. \end{array}\right\} E^P(a, h)$$

If $\varepsilon^* = \max \varepsilon$, s.t. $E^P(a, h)$, is greater than 0, and $E^P(a, h)$ is feasible, then a is possibly assigned to C_h. Note that if the possible assignment $C_P(a)$ is univocal, then it is equivalent to the necessary assignment $C_N(a)$. On the contrary, if the possible assignment is non-univocal, the necessary assignment is empty ($C_N(a) = \varnothing$). This holds for all sorting procedures where assignments are derived from the comparison of comprehensive scores with the thresholds between consecutive classes (see [34]). Note that when investigating robustness of the delivered recommendation, we may additionally account for the truth of the necessary and possible assignment-based preference relations as well as the observed extreme class cardinalities as proposed in [29].

6. Case study

Let us consider exemplary data concerning twenty Polish research units to be assigned to one of the three classes C_1–C_3 (with C_1 being the worst, and C_3 the best one). The units are evaluated on the following four criteria:

- scientific activity (g_1), including scientific publications in journals included in the *Journal Citation Reports* (JCR), Polish ministerial lists, or European Reference Index for the Humanities as well as monographs, chapters, and number of patents; the evaluation reflects an average number of points gained for publications by a single researcher of the unit;
- scientific potential (g_2), including the ability to grant PhD and habilitation degrees, number of PhDs, habilitations, and professor titles granted in the evaluation period as well as prestigious memberships (e.g., being an editor of a JCR journal, member of an editorial board of such a journal, or coordinator of an international working group or institution); all achievements are scored and these scores are summed up to get an evaluation;
- material effects of unit's activities (g_3)—representing money acquired from grants or cooperation with industry (not included in the statutory grant from the Ministry);
- remaining (non-material) effects of unit's activities (g_4)—subjective evaluation of ten most important achievements of unit's members conducted by experts of the evaluation team.

The performances of 20 considered research units are given in Table 1. As recently noted in [22], a problem representation has a significant impact on decision processes. However, in case of outranking-based methods, we suggest using a table representation (such as Table 1) or parallel coordinate plots instead of heat maps. It is the case since specification of comparison thresholds is easier when analyzing the original performances of alternatives instead of their mapping into colors representing different performance intervals or levels.

We assume the DM to have provided weight w_j and preference threshold p_j for each criterion g_j (see Table 2). According to the DM's preferences, the scientific activity of a research unit (g_1) is the most important criterion. The desired distribution of class cardinalities is as follows: 25 % and 40 % of research units should be assigned, resepectively, to C_3 or C_2, and the remaining 35 % should go to C_1.

First, let us focus on the procedure with continuous scores. Solving problem (3), we inferred representative reference profiles which are as good as possible (see Table 3). To respect the desired class cardinalities, the profiles are such that there are 5 research units at least as good as PR2, which separates classes C_2 and C_3, and 13 units not worse than profile PR1 separating C_1 and C_2. The scores, class assignments, and ranks of the units and profiles are provided in Table 1 (continuous scores). Units with ranks 1–5 (RU5, RU19. RU9,

Table 1
Performance matrix for 20 research units. Scores, class assignment, and ranks of research units and discovered profiles according to two different scoring procedures.

					Continuous scores		Binary scores	
	g_1	g_2	g_3	g_4	$Sc(a)$	Class (rank)	$Sc(a)$	Class (rank)
RU1	90	86	46	30	15.05	C_3 (5)	18	C_3 (4)
RU2	40	90	14	48	9.65	C_2 (12)	6	C_1 (16)
RU3	88	40	50	12	9.70	C_2 (11)	15	C_2 (7)
RU4	82	94	26	48	15.40	C_3 (4)	16	C_3 (5)
RU5	94	100	40	36	17.60	C_3 (1)	19	C_3 (3)
RU6	78	76	30	50	13.05	C_2 (7)	14	C_2 (8)
RU7	74	70	50	20	10.40	C_2 (10)	11	C_2 (11)
RU8	80	64	32	38	10.85	C_2 (9)	12	C_2 (10)
RU9	100	74	48	40	16.20	C_3 (3)	20	C_3 (2)
RU10	60	60	30	30	6.25	C_1 (17)	7	C_2 (13)
RU11	64	72	12	46	8.80	C_2 (14)	7	C_2 (13)
RU12	78	76	36	12	9.45	C_2 (13)	13	C_2 (9)
RU13	50	80	20	18	6.65	C_1 (16)	6	C_3 (16)
RU14	62	88	22	48	10.95	C_2 (8)	10	C_2 (12)
RU15	30	44	30	18	2.50	C_1 (21)	3	C_1 (19)
RU16	40	54	40	32	5.75	C_1 (18)	4	C_1 (18)
RU17	70	30	12	12	4.30	C_1 (19)	1	C_1 (20)
RU18	32	18	28	22	2.45	C_1 (22)	1	C_1 (20)
RU19	100	80	40	42	16.90	C_3 (2)	21	C_3 (1)
RU20	24	58	30	18	2.55	C_1 (20)	1	C_1 (20)
PR1					8.45	(15)	7	(15)
PR2					14.95	(6)	16	(6)

Table 2
Preference thresholds and weights provided by the decision maker.

	g_1	g_2	g_3	g_4
p_j	4	4	2	2
w_j	0.45	0.25	0.1	0.2

RU4, and RU1) are assigned to C_3; their scores are not less than $Sc(\text{PR2}) = 14.95$. Further, units with ranks 7–14 are assigned to C_2; their scores are at least as good as $Sc(\text{PR1}) = 8.45$ but less than $Sc(\text{PR2})$. Note that units assigned to the same class remain comparable. For example, RU5 proves to be the best with a score of 17.60, while the last unit in C_3, RU1, has the score of 15.05. In the same spirit, the range of scores in C_2 is between 13.05 (RU6) and 8.80 (RU11).

Profiles PR2 and PR1 can be interpreted as balanced performance vectors representing minimal requirements that a research unit should satisfy to be assigned to class C_3 or C_2, respectively. To support this claim, we present the inferred profiles in Fig. 6, along with the performances of all research units.

To verify robustness of the delivered recommendation, we need to consider problem (4) for each pair of an existing research unit $a \in A$ and a class C_h, $h = 1, 2, 3$. The analysis performed on this data set reveals that in case all different reference profiles respecting desired class cardinalities are used, the assignments presented in Table 1 can be considered as robust. Thus, for each unit, the assignments obtained with the representative preference profiles are equal to both the necessary and possible ones. This means that although the ranks and scores attained by the research units may vary, their class assignments remain the same, not being influenced by the procedure for selection of the reference profiles.

Let us also consider a scoring procedure with binary scores. Each alternative a (research unit or reference profile) is granted a score of one if it proves better in the pairwise comparison against another alternative b, i.e., $\pi(a, b) > \pi(b, a)$. The reference profiles inferred in this case are provided in Table 3 (binary scores). Compared to the previous ones, performances of PR2 are slightly better, and performances of PR1 slightly worse. Again, scores, class assignments, and ranks attained by the alternatives are given in Table 1 (binary scores). Obviously, all scores are integer values. Five research units with the score not less than $Sc(\text{PR2}) = 16$ are assigned to C_3; another eight units with the score at least as good as $Sc(\text{PR1}) = 7$, but worse than 16 are placed in C_2.

7. Conclusions

In this paper, we introduced a novel approach to joint sorting and ranking of research units. These units are evaluated on multiple criteria representing the level of effects they acquired and activities they undertook in the evaluation period. For interpretability of the results, we constructed reference profiles separating the classes. To leave units assigned to the same class comparable, they are first ranked and then assigned to the respective class based on the attained score. Ranking procedure is based on non-compensatory pairwise comparisons including both existing research units and reference profiles to be discovered. To account for desired class cardinalities and taking into account the role played by the reference profiles, we imposed constraints on the ranks they need to attain. We discussed different inference procedures conditioned by the preference information of the DM, as well as a set of procedures with both continuous and binary scores gained by the alternatives. We also proposed some rules for selecting precise and representative performances of the reference profiles, and we have shown how to conduct robustness analysis of the delivered recommendation. We demonstrated practical use of the approach by considering a case study of sorting/ranking 20 Polish research units with respect to 4 criteria.

Let us mention that the current procedure used by the Polish Ministry for parametric evaluation of research units is also based on non-compensatory pairwise comparisons of research units including reference profiles. It is also using the procedure with continuous scores. However, the reference profiles are defined prior to calculation, as multiples of median values of performance distributions of considered units on particular criteria, and the desirable class cardinalities are not specified.

Obviously, in this paper, we have not exhausted all possible ways of constructing and exploiting the outranking relation. While maintaining the main idea of joint multiple criteria ranking and sorting with reference profiles, it may be appealing, e.g., to account for the veto phenomenon in the construction phase or to use a different net flow score procedure in the exploitation phase.

Finally, let us remark that the ranking and sorting result can be interpreted a posteriori in terms of decision rules involving elementary conditions on a subset of criteria in the premise and specifying a binary relation or unit assignment in the conclusion (see [27,44]). Such explanations are important for justifying that the final recommendation is logical, valid, and correct because they prove to be useful in making explicit the experts logic and assumptions [30].

The proposed inference and scoring procedures are based on the use of MILP. In particular, constraint set $E^{\text{pref}}_{\text{profiles}}$ involves $2(p-1)m(n+2)$ binary variables ($2(p-1)m(n+2) = 2 \times (p-1)$ profiles $\times\, m$ criteria $\times\, (n+2)$ alternatives), while $E^{\text{ind}-\text{pref}}_{\text{profiles}}$ three times as many of them. Further, $E^{\text{continuous}}_{\text{scores}}$ involves $p(n+2)$ binary variables, whereas $E^{\text{binary}}_{\text{scores}}$ employs additional $2(p-1)n$ binary variables. Thus, for example, to infer reference profiles in our illustrative study, we solved MILPs with 418 and 498 binary variables for scoring procedures with continuous and binary scores, respectively. The execution time on Intel Atom CPU D325 1:80 GHz with 4GB RAM and GLPK solver was a few hours. Considering the typical size of groups of common evaluation of research units, the respective MILP problems are still manageable with the existing solvers. Nevertheless, as proved by the extensive experiments conducted by Cailloux et al. [7], nowadays problems with a few thousands of binary variables cannot be solved in a reasonable time of several hours. Taking this into account, the practical usefulness of the proposed approach is limited to sets consisting of several tens of alternatives.

Table 3
Inferred representative reference profiles.

	g_1	g_2	g_3	g_4
Continuous scores				
PR1	68.00	65.47	34.00	34.00
PR2	85.16	84.00	42.58	42.58
Binary scores				
PR1	64.00	61.20	32.00	32.00
PR2	86.00	84.90	43.00	43.00

Acknowledgments

The first author wishes to acknowledge financial support from the Polish National Science Centre (grant SONATA, no. DEC-2013/11/D/ST6/03056). The authors thank two anonymous referees for their remarks which helped us to significantly improve the paper.

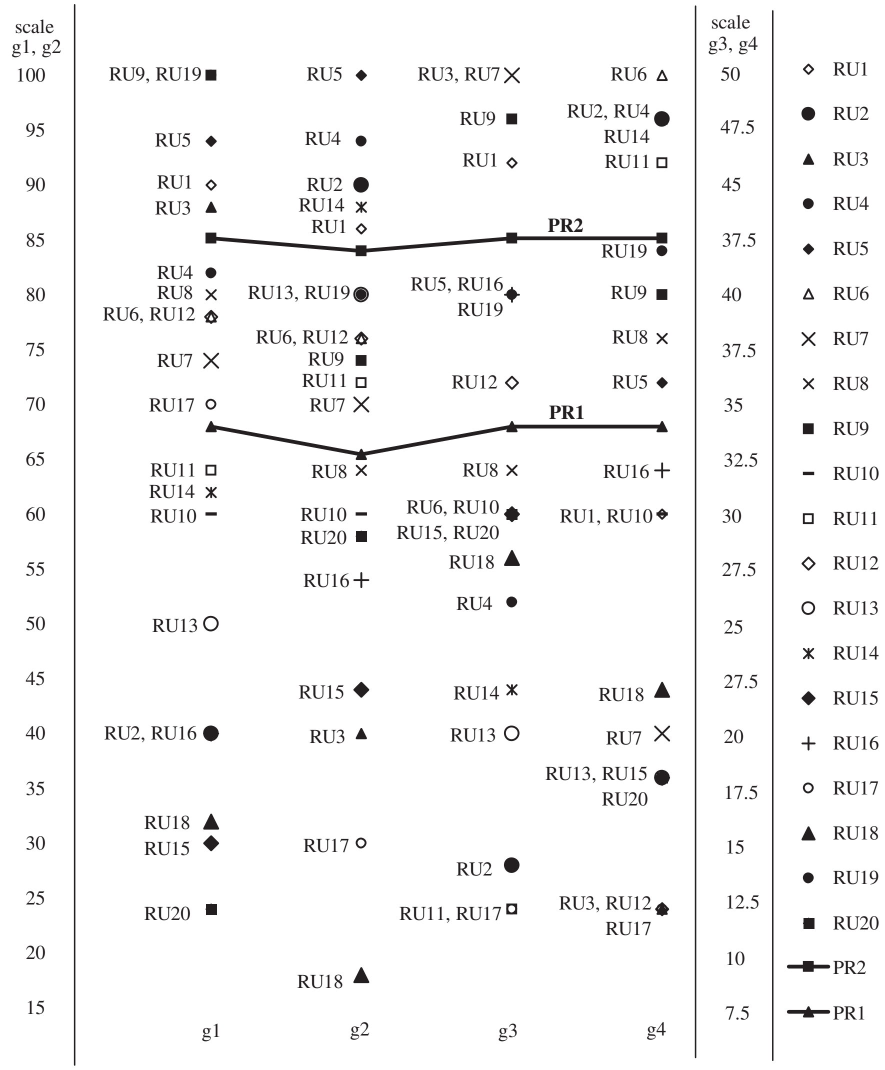

Fig. 6. Research units' performances and inferred profiles for the procedure with continuous scores.

References

[1] A. Albadvi, S.K. Chaharsooghi, A. Esfahanipour, Decision making in stock trading: An application of PROMETHEE, European Journal of Operational Research 177 (2) (2007) 673–683.

[2] J. Almeida Dias, Multiple criteria decision aiding for sorting problems: Concepts, methodologies, and applications(PhD Dissertation) Technical University of Lisbon, 2011.

[3] J. Almeida Dias, J. Figueira, B. Roy, Electre Tri-C: A multiple criteria sorting method based on characteristic reference actions, European Journal of Operational Research 204 (3) (2010) 565–580.

[4] D. Bouyssou, Ranking methods based on valued preference relations: A characterization of the net flow method, European Journal of Operational Research 60 (1) (1992) 61–67.

[5] J. Brans, B. Mareschal, The PROMCALC & GAIA decision support system for multicriteria decision aid, Decision Support Systems 12 (1994) 297–310.

[6] J. Brans, P. Vincke, B. Mareschal, How to select and how to rank projects: The PROMETHEE method, European Journal of Operational Research 24 (2) (1986) 228–238.

[7] O. Cailloux, P. Meyer, V. Mousseau, Eliciting ELECTRE TRI category limits for a group of decision makers, European Journal of Operational Research 223 (1) (2012) 133–140.

[8] A. Certa, M. Enea, T. Lupo, ELECTRE III to dynamically support the decision maker about the periodic replacements configurations for a multi-component system, Decision Support Systems 55 (1) (2013) 126–134.

[9] S. Chakhar, I. Saad, Dominance-based rough set approach for groups in multicriteria classification problems, Decision Support Systems 54 (2012) 372–380.

[10] J.-K. Chen, I.-S. Chen, Inno-Qual efficiency of higher education: Empirical testing using data envelopment analysis, Expert Systems with Applications 38 (3) (2011) 1823–1834.

[11] Y.-S. Chen, T.-C. Wang, C.-Y. Wu, Strategic decisions using the fuzzy PROMETHEE for IS outsourcing, Expert Systems with Applications 38 (10) (2011) 13216–13222.

[12] S. Corrente, S. Greco, M. Kadziński, R. Słowiński, Robust ordinal regression in preference learning and ranking, Machine Learning 93 (2–3) (2013) 381–422.

[13] S. Corrente, S. Greco, R. Słowiński, Multiple criteria hierarchy process in robust ordinal regression, Decision Support Systems 53 (3) (2012) 660–674.

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

[15] M. Doumpos, C. Zopounidis, A multicriteria decision support system for bank rating, Decision Support Systems 50 (1) (2010) 55–63.

[16] M. Doumpos, C. Zopounidis, E. Galariotis, Inferring robust decision models in multicriteria classification problems: An experimental analysis, European Journal of Operational Research 236 (2) (2014) 601–611.

[17] S. Eppe, Y. De Smet, Approximating Promethee II's net flow scores by piecewise linear value functions, European Journal of Operational Research 223 (3) (2013) 651–665.

[18] S. Eppe, Y. De Smet, An adaptive questioning procedure for eliciting PROMETHEE II's weight parameters, International Journal of Multicriteria Decision Making 4 (1) (2014) 1–30.

[19] E. Fernandez, J. Navarro, A. Duarte, G. Ibarra, Core: A decision support system for regional competitiveness analysis based on multi-criteria sorting, Decision Support Systems 54 (3) (2013) 1417–1426.

[20] E. Fernandez, R. Olmedo, An agent model based on ideas of concordance and discordance for group ranking problems, Decision Support Systems 39 (3) (2005) 429–443.

[21] H. Gervasio, L.S. da Silva, A probabilistic decision-making approach for the sustainable assessment of infrastructures, Expert Systems with Applications 39 (8) (2011) 7121–7131.

[22] J. Gettinger, E. Kiesling, C. Stummer, R. Vetschera, A comparison of representations for discrete multi-criteria decision problems, Decision Support Systems 54 (2) (2013) 976–985.

[23] C. Giannoulis, A. Ishizaka, A web-based decision support system with ELECTRE III for a personalised ranking of British universities, Decision Support Systems 48 (3) (2010) 488–497.

[24] M. Goumas, V. Lygerou, An extension of the PROMETHEE method for decision making in fuzzy environment: Ranking of alternative energy exploitation projects, European Journal of Operational Research 123 (3) (2000) 606–613.
[25] S. Greco, M. Kadziński, V. Mousseau, R. Słowiński, Robust ordinal regression for multiple criteria group decision problems: UTA^{GMS}-GROUP and UTADIS^{GMS}-GROUP, Decision Support Systems 52 (3) (2012) 549–561.
[26] S. Greco, V. Mousseau, R. Słowiński, Multiple criteria sorting with a set of additive value functions, European Journal of Operational Research 207 (4) (2010) 1455–1470.
[27] S. Greco, R. Słowiński, P. Zielniewicz, Putting dominance-based rough set approach and robust ordinal regression together, Decision Support Systems 54 (2) (2013) 891–903.
[28] E. Jacquet-Lagrèze, Y. Siskos, Preference disaggregation: 20 years of MCDA experience, European Journal of Operational Research 130 (2) (2001) 233–245.
[29] M. Kadziński, K. Ciomek, R. Słowiński, Modeling assignment-based pairwise comparisons within integrated framework for value-driven multiple criteria sorting, European Journal of Operational Research 241 (3) (2015) 830–841.
[30] M. Kadziński, S. Corrente, S. Greco, R. Słowiński, Preferential reducts and constructs in robust multiple criteria ranking and sorting, OR Spectrum 36 (4) (2014) 1021–1053.
[31] M. Kadziński, S. Greco, R. Słowiński, Selection of a representative set of parameters for robust ordinal regression outranking methods, Computers & Operations Research 39 (11) (2012) 2500–2519.
[32] M. Kadziński, S. Greco, R. Słowiński, Robust ordinal regression for dominance-based rough set approach to multiple criteria sorting, Information Sciences 283 (2014) 211–228.
[33] M. Kadziński, R. Słowiński, DIS-CARD: A new method of multiple criteria sorting to classes with desired cardinality, Journal of Global Optimization 56 (3) (2013) 1143–1166.
[34] M. Kadziński, T. Tervonen, Stochastic ordinal regression for multiple criteria sorting problems, Decision Support Systems 55 (11) (2013) 55–66.
[35] M. Kadziński, T. Tervonen, J. Figueira, Robust multi-criteria sorting with the outranking preference model and characteristic profiles, Omega (2015), http://dx.doi.org/10.1016/j.omega.2014.06.004 (in press).
[36] G. Kou, Y. Shi, S. Wang, Multiple criteria decision making and decision support systems—guest editor's introduction, Decision Support Systems 51 (2) (2011) 247–249.
[37] M.-T. Lu, S.-W. Lin, G.-H. Tzeng, Improving RFID adoption in Taiwan's healthcare industry based on a DEMATEL technique with a hybrid MCDM model, Decision Support Systems 56 (2013) 259–269.
[38] T. Lupo, A fuzzy ServQua based method for reliable measurements of education quality in italian higher education area, Expert Systems with Applications 40 (17) (2013) 7096–7110.
[39] L. Markl-Hummel, J. Geldermann, A local-level, multiple criteria decision aid for climate protection, EURO Journal on Decision Processes 2 (1–2) (2014) 121–152.
[40] V. Mousseau, L. Dias, J. Figueira, On the notion of category size in multiple criteria sorting models, Cahier du LAMSADE 205, Université Paris-Dauphine, Paris, France, 2003.
[41] V. Mousseau, R. Słowiński, Inferring an ELECTRE TRI model from assignment examples, Journal of Global Optimization 12 (2) (1998) 157–174.
[42] A. Ngo The, V. Mousseau, Using assignment examples to infer category limits for the ELECTRE TRI method, Journal of Multi-Criteria Decision Analysis 11 (1) (2002) 29–43.
[43] A. Rolland, Reference-based preferences aggregation procedures in multi-criteria decision making, European Journal of Operational Research 225 (3) (2013) 479–486.
[44] R. Słowiński, S. Greco, B. Matarazzo, Rough set and rule-based multicriteria decision aiding, Pesquisa Operacional 32 (2) (2012) 213–269.
[45] M. Szelag, S. Greco, R. Słowiński, Rule-based approach to multicriteria ranking, in: M. Doumpos, E. Grigoroudis (Eds.), Multicriteria decision aid and artificial intelligence: Links, theory and applications, Wiley-Blackwell, London, 2013, pp. 127–160 (Chapter 6).
[46] G. van Valkenhoef, T. Tervonen, T. Zwinkels, B. de Brock, H. Hillege, ADDIS: A decision support system for evidence-based medicine, Decision Support Systems 55 (2) (2013) 459–475.
[47] B. Yilmaz, M. Dagdeviren, A combined approach for equipment selection: F-PROMETHEE method and zero–one goal programming, Expert Systems with Applications 38 (9) (2011) 11641–11650.
[48] W. Yu, ELECTRE TRI: Aspects méthodologiques et manuel d'utilisation, Document du LAMSADE no 74, Université Paris-Dauphine, 1992.
[49] J. Zheng, Preference elicitation for reference based aggregation models: Algorithms and procedures(PhD Thesis) Ecole Centrale Paris, 2012.
[50] J. Zheng, O. Cailloux, V. Mousseau, Constrained multicriteria sorting method applied to portfolio selection, in: R.I. Brafman, F.S. Roberts, A. Tsoukiàs (Eds.), Algorithmic decision theory—Second International Conference, ADT 2011, Piscataway, NJ, USA, October 26-28, 2011. Proceedings, Lecture Notes in Computer Science, vol. 6992, Springer, 2011, pp. 331–343.

Miłosz Kadziński is an Assistant Professor at the Poznan University of Technology, member of the Laboratory of Intelligent Decision Support Systems (IDSS) within the Institute of Computing Science. He defended his Ph.D. thesis concerning computer decision support in 2012. His main research interests are in Multiple Criteria Decision Making (MCDM) and exploratory data analysis. He holds the MCDM Doctoral Dissertation Award 2013 by the International Society on MCDM and the EURO Doctoral Dissertation Award 2013 – top 3 finalist – by EURO — The Association of European Operational Research Societies. He has been distinguished by the Polish Academy of Sciences (Best Paper in Technical Sciences 2013), Polish Ministry of Science and Higher Education (for scientific achievements), and Foundation for Polish Science (START Scholarship in 2013 and 2014). He has been also acknowledged with the Best Reviewer Award by EJOR in 2012, 2013 and 2014. His works have been published in journals such as Omega, EJOR, DSS, Computers & OR, GDN, OR Spectrum, JOGO, and Machine Learning.

Roman Słowiński is Professor and Founding Head of the Laboratory of Intelligent Decision Support Systems within the Institute of Computing Science, Poznan University of Technology. As member of the Polish Academy of Sciences (PAS), he is currently the president of the Poznan Branch of the PAS. His area of expertise covers multiple criteria decision aiding, preference modeling, rough set theory, granular computing and knowledge discovery. Author or co-author of 14 books and more than 200 papers in major scientific journals. Laureate of the EURO Gold Medal (1991), and Doctor Honoris Causa of PolytechMons (2000), University of Paris Dauphine (2001) and Technical University of Crete (2008). He holds, moreover, the Edgeworth-Pareto Award, by International Society on Multiple Criteria Decision Making (1997) and the 2005 Prize of the Foundation for Polish Science, regarded as the most prestigious scientific award in Poland. Since 1999, he is editor-in-chief of the European Journal of Operational Research. Senior Member of the IEEE.