Better decisions with less cognitive load
The Parsimonious BWM

Corrente, Salvatore; Greco, Salvatore; Rezaei, Jafar

DOI
10.1016/j.omega.2024.103075

Publication date
2024

Document Version
Final published version

Published in
Omega (United Kingdom)

Omega 126 (2024) 103075

Contents lists available at ScienceDirect

Omega

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

Better decisions with less cognitive load: The Parsimonious BWM ☆

Salvatore Corrente [a,*], Salvatore Greco [a], Jafar Rezaei [b]

[a] Department of Economics and Business, University of Catania, Corso Italia, 55, 95129 Catania, Italy
[b] Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

ARTICLE INFO

Keywords:
Best–Worst Method
Alternatives' priorities
Parsimonious elicitation

ABSTRACT

Despite its recent introduction in literature, the Best–Worst Method (BWM) is among the most well-known and applied methods in Multicriteria Decision-Making. The method can be used to elicit the relative importance (weight) of the criteria as well as to get the priorities of the alternatives on the criteria at hand. In this paper, we will present an extension of the method, namely, the parsimonious Best–Worst-Method (P-BWM) permitting to apply the BWM to get the priorities of the alternatives in case they are in a large number. At first, the Decision-Maker (DM) is asked to give a rating to the alternatives under consideration; after, the classical BWM is applied to a set of reference alternatives to get their priorities used to compute, then, the priorities of all the alternatives under consideration. We propose also a procedure to select reference alternatives, possibly in cooperation with the DM, providing a well-distributed coverage of the rating range. The new proposal requires the DM a fewer number of pairwise comparisons than the original BWM. Another contribution of the paper is related to the comparison between BWM, P-BWM, the Analytic Hierarchy Process (AHP), and the parsimonious AHP in terms of the amount of preference information provided by the DM in each method to apply it. In addition to the standard approach, we propose one alternative way of inferring the priority vectors in BWM and P-BWM based on the barycenter of the space of alternatives priorities compatible with the preferences given by the DM. Finally, an experiment with university students has been conducted to test the new proposal. Results of the experiments show that P-BWM performs better than BWM in terms of capability to represent the DM's preferences and the difference between the results of the two methods is significant from the statistical point of view. The new proposal will permit to use the potentialities of the BWM to get the alternatives' priorities in real-world decision-making problems where a large number of alternatives must be taken into account.

1. Introduction

We, as individuals, groups, or organizations, make many decisions in our lifetime, which means we should choose from among several available options to achieve our different goals. Among so many examples, we could think of choosing a university to study by a student, buying a house by a family, or selecting a supplier by a retailer. Although selection might be the goal of most decision problems, it is not the only one. Sometimes, the goal is to rank the options or sort them, such as ranking universities or sorting the hospitals in a country. In general, we name a decision problem a multi-criteria decision-making (MCDM) problem, where the options (alternatives) need to be evaluated with respect to a set of (conflicting) criteria (attributes). Such problems are usually challenging to handle for decision-makers as the alternatives – to be considered – are non-dominated (one is better than the other with respect to some criteria (at least one) and worse (or equal) with respect to the others). Such a challenge has motivated many scientists to develop methods to help the Decision-Maker (DM) to make more informed decisions. Determining the relative importance of the criteria and the (overall) value of alternatives, which are usually named weight and priority, respectively, is perhaps a part that has gained more attention in the existing literature on MCDM. Among the more popular methods, we could refer to Multi-Attribute Value Theory [1], ELECTRE [2], PROMETHEE [3], Analytic Hierarchy Process (AHP) [4], and the Best–Worst Method (BWM) [5]. The main focus of the current study is the youngest of this set, the BWM.

BWM is a recently developed multi-criteria decision-making method that uses a structured pairwise comparison system to find the relative importance (weight) of the decision criteria (and alternatives). According to BWM, the decision-maker needs to choose the Best (e.g., most important), and the Worst (e.g., least important) criteria (alternatives),

☆ Area: Data-Driven Analytics. This manuscript was processed by Associate Editor E. Triantaphyllo.

* Corresponding author.
E-mail addresses: salvatore.corrente@unict.it (S. Corrente), salgreco@unict.it (S. Greco), j.rezaei@tudelft.nl (J. Rezaei).

https://doi.org/10.1016/j.omega.2024.103075
Received 30 August 2023; Accepted 7 March 2024
Available online 13 March 2024

and then conduct a pairwise comparison between the Best and all the other criteria (alternatives), and between the other criteria (alternatives) and the Worst. These two vectors of pairwise comparison judgments are then used to infer the weights. There exist several optimization models for this step, including the original non-linear model [5], a linear model [6], a multiplicative model [7], a Bayesian model [8], a nonadditive model [9] and a fuzzy one [10] to name a few. Due to its attractive features, including its data efficiency, simplicity in revising inconsistent comparisons, and its debiasing mechanism against some cognitive biases (see, 11,12), BWM has gained considerable attention among researchers and practitioners. Just to cite a few recent contributions about BWM, [13] studied the BWM determining the analytical form of criteria weights without the help of any optimization software; [14] ranked the risks associated with big data analytics implementation in Indian automotive manufacturing industry by BWM; a three-phase methodology for supplier selection, where the last is done by BWM, has been proposed by [15]; in a similar context, [16] used BWM to get the weights of criteria necessary to evaluate third-party logistics providers for sustainable last-mile delivery (for a recent survey about BWM see [17]. A full list of contributions related to BWM can be found at bestworstmethod.com).

While the BWM has been mainly used in determining the weights of the criteria, it can also be used to determine the priority of the alternatives [18]. In determining the weights, one of the core assumptions of the method (which is rooted in psychological studies related to human brain capabilities) is that the DM does not do the pairwise comparison among more than nine criteria [19,20]. While the assumption works very well for the criteria (as for most decision problems, we have a handful of relevant criteria, or in case of more than 9, we could cluster them), when it comes to the alternatives, it might work as a limitation as in many real-world decision-making problems, one might deal with a large set of alternatives (think of, for instance, ranking the countries based on their sustainability performance, or sorting the schools in a country). The main aim of the current study is to work on this limitation and empower the BWM to determine the priority of the alternatives when the set is large. We develop a parsimonious version of the method, which calls for rating all alternatives and conducting pairwise comparisons among only few well-distributed reference alternatives (instead of all). The priority to the other alternatives is assigned by linear interpolation of the reference alternative priorities based on the DM's rating. The basic idea here is that the rating provided by the DM is corrected using the priority for reference alternatives obtained through the BWM. In fact, with this procedure, on the one hand, we correct the possible errors linked to the necessity to evaluate many alternatives together, while, on the other hand, we avoid the possible unreliability of the DM's preference information linked to the large number of pairwise comparisons required by the BWM in the case of many alternatives. The introduction of the reference alternatives has specific importance also from the point of view of the decision biases [21]. Indeed, the presence of a multiplicity of reference points permits to counterbalance the anchoring biases, i.e. the tendency to base judgments on an initial piece of information [22]. As shown by [11] using BWM, the alternatives at the extreme points of the scale are over-evaluated when compared with respect to the highest (best) point and under-evaluated when compared with the lowest (worst) point. Moreover, the middle points are over-evaluated when compared with the lowest point and under-evaluated with the highest point. The compensation of the best-to-others and others-to-worst comparisons permits the BWM to mitigate the biases related to the above over-evaluations and under-evaluations. In this perspective, the direct rating conjugated with BWM further mitigates these biases because it avoids any anchor.

Observe also that the presence of intermediate reference points, in addition to the best and worst points, are beneficial from the point of view of the contextual effects of scaling and rating, according to which they depend on the whole set of stimuli. In particular, in the adaptation level theory of Helson [23,24] each judgment is defined with respect to an average of past stimuli, while in the range-frequency theory of Parducci [25,26] the judgment depends on the range of the scale and the distribution of the stimuli. In both cases, to propose a whole set of well-distributed reference points rather than simply the extreme highest and lowest points seems beneficial because it prevents anchoring the judgments on some specific points, such as one low or one high extreme point, as it is the case for SMART and SWING procedures [11]. This is also beneficial with respect to the BWM because presenting a whole range of reference points to the DM mitigates the above-mentioned over-evaluations and under-evaluation effects due to the comparisons with the best and the worst points, whose anchoring effect can be smoothed by the consideration of the other reference points. The parsimonious BWM we are proposing has a specific relevance with respect to real-world applications. Indeed, very often, the alternatives that have to be compared in a decision problem can be very numerous, in the order of tens, hundreds, or even thousands. Imagine, for example, an application in the domains of ranking of Universities [27], well-being ranking of countries [28], healthcare system assessment [29], sustainable development [30] and so on. In this case, it is not reasonable to pairwise compare all the units to be evaluated with the best and the worst units on all considered criteria. Therefore, considering the growing interest in such types of ranking, the parsimonious BWM permits extending the application of the BWM to relevant domains that, otherwise, would not be possible to consider. Another relevant type of application of the parsimonious BWM is the repetitive assessment of units that cannot be known ex-ante as it is the case for multi-criteria financial scoring or rating [31], and, more in general, multi-criteria assessment in different domains such as building performances [32], housing evaluation [33], sustainability evaluation [34], Environmental, Social and Governance (ESG), or Corporate Social Responsibility (CSR) score [35]. In all these cases, as well as in similar situations, it is not possible to apply BWM because it is not possible to compare the units to be evaluated with the best and the worst. After all, beyond being a large number, they are known time by time when the assessment is required. However, the parsimonious BWM permits always a linear interpolation with the corrected rating assigned to the reference alternatives.

To test the performance of the proposed approach, we also conducted an experiment on student subjects, based on which we found promising results. We think that this is a significant contribution to the existing literature on BWM as the new model opens up a new exciting area of applications, namely the application of the BWM for problems with a large set of alternatives.

In the next section, we introduce the background of the study describing the BWM and two ways of inferring alternatives' priorities, followed by Section 3 presenting our proposed parsimonious BWM, with an illustrative example. In Section 4, we conduct a comprehensive analysis comparing the amount of data needed for the four methods of BWM, parsimonious BWM, AHP, and parsimonious AHP. In Section 5, we perform some experimental studies, an essential part of our study to show the performance of the parsimonious BWM, which is evaluated by some statistical metrics. Finally, conclusion and future research directions are presented in Section 6.

2. An overview of BWM and proposing a new approach for getting a representative vector from the results of non-linear BWM

2.1. Multiple criteria decision-making

In Multiple Criteria Decision-Making (MCDM; 1,36,37), a set of alternatives $A = \{a_1, a_2, \dots, a_n\}$ has to be evaluated on a coherent family of criteria $G = \{g_1, \dots, g_m\}$ [2] to deal with choice, ranking or sorting problems. In this paper, we are interested in ranking problems in which one has to rank all alternatives from the best to the worst. The only objective information that can be gathered from the performance matrix, where the evaluations of the alternatives on the criteria at hand are collected, is the dominance relation for which an alternative a_i dominates an alternative a_j if a_i is no worse than a_j on all criteria and better for at least one of them. However, since this relation is quite poor (in general, there are criteria for which a_i is better than a_j and vice versa criteria for which a_j is better than a_i) there is the necessity to aggregate the alternatives' evaluations. This can be done by value functions [1], outranking relations [2] or decision rules methods [38]. Value functions assign a unique numerical evaluation to each alternative being representative of its goodness with respect to the problem at hand; outranking relations compare alternatives pairwise to define if one is at least as good as another, and, finally, decision rules link the global preferences expressed by the DM on the considered alternatives to their performances on the criteria.

2.2. The best–worst method

Despite its recent introduction in literature, the Best–Worst Method (BWM, 5) is nowadays one of the most applied MCDM methods to deal with decision-making problems [17]. The method can be used to get the weights of criteria to be used as tradeoffs in value functions or, analogously, to get the alternatives' priorities on the criteria under consideration. In this paper, we are interested in its application to get alternatives' priorities. Regarding its application, at first, the DM is asked to define the *Best* and the *Worst* alternative on the criterion under consideration. Secondly, they have to pairwise compare the Best alternative with all the other alternatives and the other alternatives with the Worst one, using the traditional 1–9 scale considered in the Analytic Hierarchy Process (AHP; 4). Denoting by a_{Bj} the pairwise comparison between the Best alternative and the alternative a_j, by a_{jW} the pairwise comparison between the same alternative and the Worst one, and by $w_1, \ldots, w_n$ the alternatives' priorities, to get them one has to solve the following problem [5]

$$\begin{array}{l} \min \ \max_j \left\{ \left| \frac{w_B}{w_j} - a_{Bj} \right|, \left| \frac{w_j}{w_W} - a_{jW} \right| \right\}, \\ \text{s.t.} \\ \sum_{j=1}^{n} w_j = 1, \\ w_j \geqslant 0, \text{ for all } j = 1, \ldots, n, \end{array}$$

that can be equivalently written in the following way

$$\min \xi, \ \text{ subject to}$$

$$\left.\begin{array}{l} \left| \frac{w_B}{w_j} - a_{Bj} \right| \leqslant \xi, \text{ for all } j = 1, \ldots, n, \\ \left| \frac{w_j}{w_W} - a_{jW} \right| \leqslant \xi, \text{ for all } j = 1, \ldots, n, \\ \sum_{j=1}^{n} w_j = 1, \\ w_j \geqslant 0, \text{ for all } j = 1, \ldots, n. \end{array}\right\} E_{BWM}^{DM} \quad (1)$$

The pairwise comparisons provided by the DM are perfectly consistent iff $a_{Bj} \cdot a_{jW} = a_{BW}$ for all $j = 1, \ldots, n$. However, to check for the consistency of the given preference information before solving the mathematical problems described above, [39] proposed the following *Global Input-based Consistency Ratio* $CR^I = \max_{j=1,\ldots,n} CR_j^I$ where

$$CR_j^I = \begin{cases} \frac{|a_{Bj} \times a_{jW} - a_{BW}|}{a_{BW} \times a_{BW} - a_{BW}} & \text{if} \quad a_{BW} > 1, \\ 0 & \text{if} \quad a_{BW} = 1. \end{cases} \quad (2)$$

If CR^I is not greater than a specific threshold that depends on the size of the problem (m) and on the magnitude of a_{BW} as specified by [39], then, the pairwise comparisons provided by the DM are reliable enough. In the opposite case, there is the necessity to revise the given information to make it consistent.

2.3. Two ways of inferring the priorities in BWM

In this section, we present two different ways to infer a priority vector in BWM starting from the preference information provided by the DM (both methods can also be applied to the P-BWM that will be presented in the next section).

As explained in [6], solving the mathematical problem (1) and denoted by ξ^* the optimal value so obtained, two different cases can occur:

- $\xi^* = 0$: there is only one priority vector $(w_1, \ldots, w_n)$ satisfying all constraints in E_{BWM}^{DM} and, therefore, compatible with the information provided by the DM,
- $\xi^* > 0$: there is more than one priority vector compatible with the information given by the DM and they satisfy the constraints

$$\left.\begin{array}{l} \left| \frac{w_B}{w_j} - a_{Bj} \right| \leqslant \xi^*, \text{ for all } j = 1, \ldots, n, \\ \left| \frac{w_j}{w_W} - a_{jW} \right| \leqslant \xi^*, \text{ for all } j = 1, \ldots, n, \\ \sum_{j=1}^{n} w_j = 1, \\ w_j \geqslant 0, \text{ for all } j = 1, \ldots, n, \end{array}\right\} \quad (3)$$

that can also be written in a linear way as follows:

$$\left.\begin{array}{l} -\xi^* \cdot w_j \leqslant w_B - a_{Bj} \cdot w_j \leqslant \xi^* \cdot w_j, \text{ for all } j = 1, \ldots, n, \\ -\xi^* \cdot w_W \leqslant w_j - a_{jW} \cdot w_W \leqslant \xi^* \cdot w_W, \text{ for all } j = 1, \ldots, n, \\ \sum_{j=1}^{n} w_j = 1, \\ w_j \geqslant 0, \text{ for all } j = 1, \ldots, n. \end{array}\right\} E_{BWM}^{Opt} \quad (4)$$

We can now consider two methods, alternative to the standard approach, aiming to select one compatible priority vector among those satisfying constraints in E_{BWM}^{Opt}:

- *Central priority vector* [6]: we can find the minimum and maximum priority w_j under the constraints E_{BWM}^{Opt}. Formally, one has to compute the following LP problems for all $j = 1, \ldots, n$,

$$\left.\begin{array}{l} \min w_j = w_j^{min}, \text{ subject to} \\ E_{BWM}^{Opt} \end{array}\right\} \qquad \left.\begin{array}{l} \max w_j = w_j^{max}, \text{ subject to} \\ E_{BWM}^{Opt} \end{array}\right\}$$

  and, then, the middle point $w_j^{Int} = \frac{1}{2} w_j^{min} + \frac{1}{2} w_j^{max}$ of the intervals $\left[w_j^{min}, w_j^{max} \right]$ obtaining the priority vector $\mathbf{w}^{Int} = \left(w_1^{Int}, \ldots, w_n^{Int} \right)$;
- *Barycenter priority vector:* as underlined above, if $\xi^* > 0$, an infinite number of priority vectors satisfy all constraints in E_{Opt}^{BWM}. Since the constraints in E_{Opt}^{BWM} are linear, one can sample a certain number S of priority vectors from the space they define by using, for example, the Hit-And-Run (HAR; 40,41) method. Denoting by $\mathbf{w}^k = \left(w_1^k, \ldots, w_n^k \right)$, with $k = 1, \ldots, S$, the sampled priority vectors, their barycenter can be computed $\mathbf{w}^{Bar} = \left(w_1^{Bar}, \ldots, w_n^{Bar} \right)$ where, formally, $w_j^{Bar} = \frac{1}{S} \sum_{k=1}^{S} w_j^k$ for all $j = 1, \ldots, n$. In general, the barycenter of a space of compatible models can well represent the preferences of the DM [42].

3. The parsimonious Best–Worst method: P-BWM

Even if the BWM application involves fewer pairwise comparisons than AHP and this makes BWM more reliable than AHP in real-world applications, the amount of preference information asked to the DM by the BWM can make difficult its application when the number of alternatives is big. For this reason, in this section, we present a parsimonious version of the BWM, making it applicable to deal with big-size problems

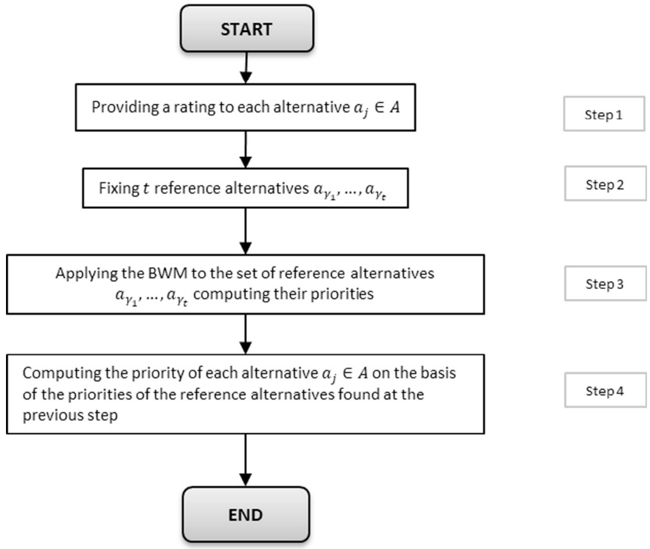

Fig. 1. Flowchart of the P-BWM.

without renouncing its basic principle. As underlined before, we are interested in the application of the BWM to get the priorities of the alternatives under consideration. Let us observe that the first proposal of the method dates back to the 2021 MCDM conference [43] and the first trial of presenting it was done by [44].

To make clearer this description, Fig. 1 shows a flowchart of the method which steps are articulated as follows:

Step (1) The DM assigns a rating $r(a_j) \in \mathbb{R}$ to each alternative $a_j \in A$. Since, as previously underlined, we are interested in getting the priorities of the alternatives in case a large number of alternatives is taken into account, $r(a_j)$ represents the goodness of a_j with respect to all the others,

Step (2) The DM, in accordance with the analyst, fixes t reference alternatives $(a_{\gamma_1}, \ldots, a_{\gamma_t})$ in A ordered from the least to the most preferred, that is, $a_{\gamma_1} \precsim a_{\gamma_2} \precsim \cdots \precsim a_{\gamma_t}$, where $a_{\gamma_s} \precsim a_{\gamma_{s+1}}$ means that $a_{\gamma_{s+1}}$ is at least as good as a_{γ_s},

Step (3) The DM, with the support of the analyst, applies the BWM to the set of reference alternatives $\{a_{\gamma_1}, \ldots, a_{\gamma_t}\}$. At first, they choose the best a_B and the worst a_W among $a_{\gamma_1}, \ldots, a_{\gamma_t}$ and, then, they compare a_B and a_W with the other reference alternatives. If CR^I is not greater than the considered threshold, then, the priorities of $\{a_{\gamma_1}, \ldots, a_{\gamma_t}\}$, that is, $u(r(a_{\gamma_1})), \ldots, u(r(a_{\gamma_t}))$, are computed. In the opposite case, one has to revise the preference information given by the DM to find the reason for the obtained inconsistency and, after that, apply the BWM to get the priorities of the reference evaluations. Let us observe that in addition to the original way of obtaining the priorities presented by [5], the two alternative methods presented in the previous version could be applied to obtain the reference alternatives priorities,

Step (4) For each $a_j \in A$ such that $r(a_j) \in \left[r(a_{\gamma_s}), r(a_{\gamma_{s+1}})\right]$, its priority $w_j = u(r(a_j))$ is computed by interpolating the priorities $u(r(a_{\gamma_s}))$ and $u(r(a_{\gamma_{s+1}}))$ obtained at the previous step. From a mathematical point of view,

$$w_j = u(r(a_j)) = u(r(a_{\gamma_s})) + \frac{u(r(a_{\gamma_{s+1}})) - u(r(a_{\gamma_s}))}{r\left(a_{\gamma_{s+1}}\right) - r\left(a_{\gamma_s}\right)}\left(r(a_j) - r\left(a_{\gamma_s}\right)\right). \quad (5)$$

Let us remark that in the proposed procedure a very important role is played by the reference points. In fact, due to the context dependence effect of rating [45,46], reference points not well distributed could bias the rating, for example, because a concentration of low-value reference points could induce a systematic over-evaluation as well as a concentration of high-value reference points could induce a systematic under-evaluation (see, e.g., [47]). To reduce and control the context effect we consider the recommendation of [48]: *"Several precautions have been standard with rating scales in functional measurement. First, is the use of preliminary practice, which has several functions. The general range of stimuli is not known to the subject initially, and the rating scale is arbitrary. Accordingly, the subject needs to develop a frame of reference for the stimuli and correlate it with the given response scale"*. In this perspective, we propose a specific procedure to select a well-distributed set of reference alternatives from A. Ideally, we want to maximize the distance in terms of rating between reference alternatives. With this aim, we propose to select the set $R = \{a_{\gamma_1}, \ldots, a_{\gamma_t}\} \subset A$ that maximizes the minimal absolute difference $\left|r\left(a_{\gamma_r}\right) - r\left(a_{\gamma_s}\right)\right|$. Some specific constraints can be taken into account in this optimization problem, and for example, reordering the alternatives from R such that $r\left(a_{\gamma_1}\right) < \cdots < r\left(a_{\gamma_t}\right)$,

- the alternatives with the minimal rating and the maximal rating are included among the reference points, that is,

  $r(a_{\gamma_1}) = min\{r(a_j), a_j \in A\}$ and $r(a_{\gamma_t}) = max\{r(a_j), a_j \in A\}$,

- to avoid concentration in some parts of the rating sequence, the set of reference alternatives cannot include three consecutive rating alternatives, that is, if the alternatives a_j from A are reordered in the sequence

  $a_{(1)}, \ldots, a_{(n)}$

  such that $r\left(a_{(j)}\right) \leqslant r\left(a_{(j+1)}\right)$, no triple $\left\{a_{(j)}, a_{(j+1)}, a_{(j+2)}\right\}$ is contained in the set of reference alternatives R.

On the basis of previous considerations, the set of reference alternatives $R = \{a_{\gamma_1}, \ldots, a_{\gamma_t}\} \subset A$ can be obtained by solving the following MILP problem where M is a large number and the unknown variables are ε and $\rho_j, j = 1, \ldots, n$, with ρ_j a binary variable such that $\rho_j = 1$ if a_j is selected as a reference alternative and $\rho_j = 0$ otherwise:

$$\max \varepsilon = \varepsilon^*, \text{ subject to}$$

$$\left.\begin{array}{l} \rho_j \cdot r(a_j) - \rho_{j'} \cdot r(a_{j'}) \geqslant \varepsilon - (2 - \rho_j - \rho_{j'}) \cdot M \\ \quad \text{for all } j, j' = 1, \ldots, n, \text{ such that } r(a_j) \geqslant r(a_{j'}) \\ \sum_{j=1}^{n} \rho_j = t, \\ \rho_{(1)} = 1, \rho_{(n)} = 1, \\ \rho_{(j)} + \rho_{(j+1)} + \rho_{(j+2)} \leqslant 2, \text{ for all } j = 1, \ldots, n-2, \\ \rho_j \in \{0, 1\} \text{ for all } j = 1, \ldots, n. \end{array}\right\} \quad (6)$$

Let us comment on the MILP problem (6). Taking $M > \max_{a_j \in A} |r(a_j)|$, the constraint

$$\rho_j \cdot r(a_j) - \rho_{j'} \cdot r(a_{j'}) \geqslant \varepsilon - (2 - \rho_j - \rho_{j'}) \cdot M$$

is always satisfied if at least one between a_j and $a_{j'}$ is not selected as a reference point, that is, if $\rho_j = 0$ or $\rho_{j'} = 0$. On the contrary, if $\rho_j = \rho_{j'} = 1$, the constraint becomes

$$r(a_j) - r(a_{j'}) \geqslant \varepsilon$$

which ensures that the minimal absolute difference between ratings of reference alternatives is ε^*. The constraint $\sum_{j=1}^{n} \rho_j = t$ fixes in t the number of reference alternatives, while the constraints $\rho_{(1)} = 1$ and $\rho_{(n)} = 1$ are required to include the alternatives with the smallest and the highest rating in the set R of reference alternatives. Finally, the

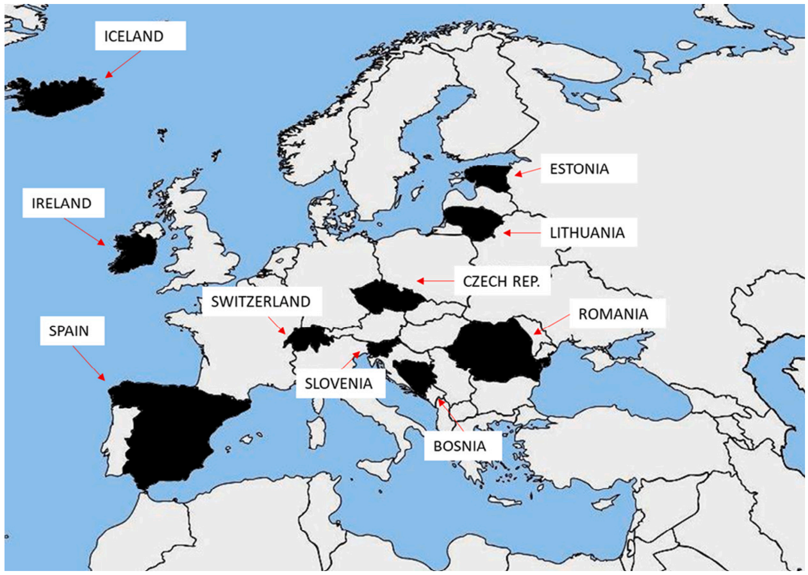

Fig. 2. Ten Countries of which the DM would like to find the area.

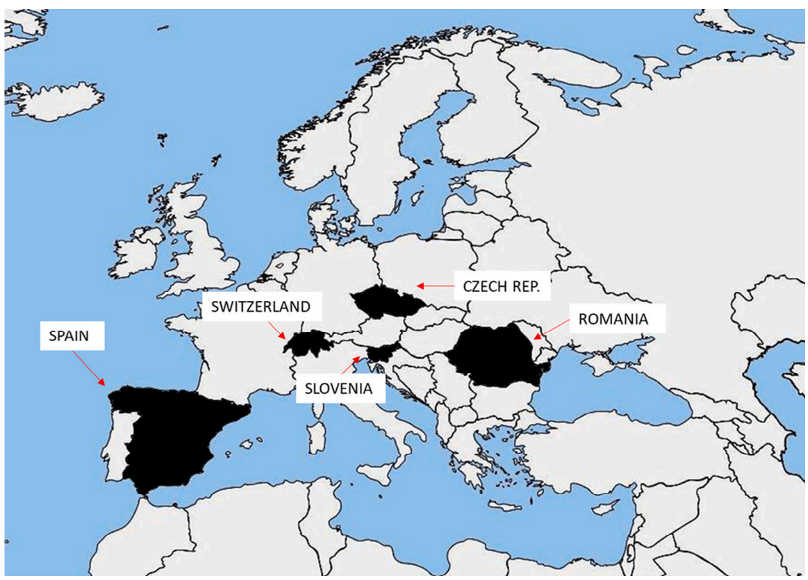

Fig. 3. Five Countries selected as reference alternatives.

constraint $\rho_{(j)} + \rho_{(j+1)} + \rho_{(j+2)} \leqslant 2, j = 1, \ldots, n-2$, prevents that three consecutive alternatives could be taken as reference alternatives.

In a more simplified version, one could also consider the possibility of choosing the t reference members with the alternatives characterized by the ratings closest to the following values:

$$r(a_{(1)}),\ r(a_{(1)}) + \frac{1}{t-1}\left[r(a_{(n)}) - r(a_{(1)})\right],\ r(a_{(1)}) + \frac{2}{t-1}\left[r(a_{(n)}) - r(a_{(1)})\right],\ \ldots\ , r(a_{(n)}).$$

One could also consider the possibility to select the t reference members in cooperation with the DM among the alternatives with the rating closest to the above-mentioned values. In this way, the DM could select the alternatives for which they will be more confident in providing the required judgments which should increase the reliability of the elicited preference information.

Observe also that, instead of the DM's rating, when present, one could consider some underlying value on which the scaling is based. For example, in a decision problem related to university students' assessment, the rating is based on the grades in the considered subjects, so that, in the above procedure to select the reference alternatives, the grades, rather than the DM's ratings, can be considered. In this case, one could also consider some transformations of the underlying value, taking into account the relationship between the stimulus magnitude I and the sensation S, such as:

- the Weber–Fechner Law [49], for which there is a logarithmic relation, that is, $S = a \cdot log\frac{I}{I_0}$, with a and I_0 positive constant and I_0 referred as detection level, or
- the Stevens law [50], for which there is a power relation, that is, $S = b \cdot I^n$, with b and n positive, and, in general, $n < 1$ because basically the increase in sensation intensity decreases with increasing stimulus magnitude.

In fact, to propose a set of reference alternatives well-distributed with respect to the sensation, the Weber–Fechner law suggests to apply the above procedure on a logarithmic transformation of the underlying value, while the Stevens law suggests to consider a power transformation.

Table 1
Rating of all Countries provided by the DM.

	Bosnia	Estonia	Ireland	Iceland	Lithuania	Czech Rep.	Romania	Slovenia	Spain	Switzerland
$r(\cdot)$	3	2.7	3.5	5	3.7	4	6	**1**	7	3
$u(r(\cdot))$	0.076	0.071	0.110	0.205	0.123	0.144	0.266	0.044	0.470	0.076

Table 2
Pairwise comparisons provided by the DM between Spain (the best Country) and the reference Countries (a_{BO}) as well as between the reference Countries and Slovenia (the Worst one) (a_{OW}). Rating given by the DM to the reference Countries and priorities obtained by the BWM application to the set composed of the reference Countries only.

	Czech Rep. (a_{γ_3})	Romania (a_{γ_4})	Slovenia (a_{γ_1})	Spain (a_{γ_5})	Switzerland (a_{γ_2})
a_{BO}	5	3	9	1	7
a_{OW}	5	7	1	9	3
$r(\cdot)$	$r(a_{\gamma_3}) = 4$	$r(a_{\gamma_4}) = 6$	$r(a_{\gamma_1}) = 1$	$r(a_{\gamma_5}) = 7$	$r(a_{\gamma_2}) = 3$
$u(r(\cdot))$	0.144	0.266	0.044	0.470	0.076

3.1. The P-BWM application: An example

In this section, we present how to apply the P-BWM described above, introducing the problem on which the experiments in Section 5 are based.

Let us assume we would like to find the area of the ten European Countries shown in Fig. 2. To this aim, let us use the P-BWM presented above showing, in detail, the steps on which the method application is based.

Step (1) The DM has to provide a rating to the considered Countries. In order to facilitate the DM's task, let us assume that the area of Slovenia is 1 and that the rating of the other Countries should be given on the basis of this assumption. The rating given by the DM is shown in Table 1.

Step (2) Let us consider the five Countries shown in Fig. 3 that will act as reference alternatives. They have been selected using the procedure described in Section 3, solving the MILP problem (6), considering as rating the real area of the ten European Countries shown in Table 6 of Section 4. In other words, in the MILP problem (6), the constraint

$$\rho_j \cdot r(a_j) - \rho_{j'} \cdot r(a_{j'}) \geqslant \varepsilon - (2 - \rho_j - \rho_{j'}) \cdot M$$
$$\text{for all } \; j, j' = 1, \dots, n, \; \text{such that } r(a_j) \geqslant r(a_{j'})$$

is replaced by the constraint

$$\rho_j \cdot Area(a_j) - \rho_{j'} \cdot Area(a_{j'}) \geqslant \varepsilon - (2 - \rho_j - \rho_{j'}) \cdot M$$
$$\text{for all } j, j' = 1, \dots, n, \; \text{such that } Area(a_j) \geqslant Area(a_{j'})$$

and, with reference to constraints

$$\rho_{(1)} = 1, \quad \rho_{(n)} = 1,$$

and

$$\rho_{(j)} + \rho_{(j+1)} + \rho_{(j+2)} \leqslant 2, \; j = 1, \dots, n-2$$

the alternatives a_j from A are reordered in the sequence

$$a_{(1)}, \dots, a_{(n)}$$

such that $Area(a_{(j)}) \leqslant Area(a_{(j+1)}), \; j = 1, \dots, n-1$. Instead of utilizing the rating, we opted for the real area because our goal was to evaluate P-BWM by presenting the same set of reference alternatives to all participants. Conversely, employing the rating method would have necessitated providing each participant with a unique reference set tailored to their rating.

Step (3) In this step, the DM is asked to select the *Best* and the *Worst* Countries among the reference ones where here, Best and Worst refer to their size, that is, the biggest and the smallest ones. The DM selects Spain as the Best Country and Slovenia as the Worst one. Then, they are asked to apply the BWM to the five reference Countries comparing the Best and the Worst with the other reference Countries using the classical 1–9 Saaty scale. In comparing Countries A and B, assuming that the DM retains A not smaller than B, the points in the scale have the following interpretation:

1- A and B have the same size,

3- A is moderately bigger than B,

5- A is strongly bigger than B,

7- A is very strongly bigger than B,

9- A is extremely bigger than B.

Values 2, 4, 6 and 8 denote a hesitation between 1–3, 3–5, 5–7 and 7–9, respectively. Let us assume that the vectors a_{BO} and a_{WO} containing the pairwise comparisons between the Best country (i.e., the biggest) and the reference ones as well as the pairwise comparisons between the reference Countries and the Worst (i.e., the smallest) one are those shown in Table 2.

Computing CR^I as shown in Eq. (2) and observing that its value (0.2222) is lower than the threshold considered in this case (0.3062), one can apply the BWM finding the priorities of the reference Countries shown in the last row of Table 2. Let us observe that [39] provide the CR^I threshold for problems composed of at most 9 criteria. In our case, even if the number of alternatives (acting as criteria) is 10, we considered the thresholds defined in the paper for the case $n = 9$. This is an even more restrictive assumption observing that for a fixed value of a_{BW} the CR^I threshold is not decreasing with n. Therefore, one would expect that passing from $n = 9$ to $n = 10$ the threshold used to check if the pairwise comparisons provided by the DM are consistent enough should increase.

Step (4) The size of all the Countries under consideration is obtained by Eq. (5) using the priorities of the reference Countries found in the previous step. For example, considering the rating assigned to Iceland by the DM, that is, 5 (see Table 1) and observing that this rating belongs to the interval rating $[r(a_{\gamma_3}), r(a_{\gamma_4})] = [4, 6]$ assigned to reference Countries (see the third row of Table 2) for which the priorities $u(r(a_{\gamma_3})) = u(4) = 0.144$ and $u(r(a_{\gamma_4})) = u(6) = 0.266$ have been obtained, one gets

$$u(5) = u(4) + \frac{u(6) - u(4)}{6 - 4}(5 - 4) = 0.144 + \frac{0.266 - 0.144}{2} = 0.205.$$

The priorities obtained for the ten Countries under consideration are therefore shown in the last row of Table 1. In Appendix D, we provide details on the programming problem to be solved to get the priorities of the reference alternatives as well as all the computations done to obtain the priorities of all the alternatives as shown in Table 1. All the mathematical problems related to the BWM and P-BWM application have been solved using the commercial software MATLAB 2021.

Table 3
Pieces of preference information involved in the four considered methods: n is the number of alternatives at hand, while r is the number of reference alternatives considered in P-AHP as well as in P-BWM methods.

AHP	BWM	P-AHP	P-BWM
$\frac{n(n-1)}{2}$	$2n-3$	$n+\frac{r(r-1)}{2}$	$n+(2r-3)$

Table 4
Comparison between methods with respect to the number of pieces of preference information asked to the DM to apply them. Each value in the table represents under which condition the application of the method in the row asks for a lower number of pieces of preference information than the application of the method in the column.

	AHP	BWM	P-AHP	P-BWM
AHP	■	✗	$r > \frac{1+\sqrt{4n^2-12n+1}}{2}$	$r > \frac{n^2-3n+6}{4}$
BWM	$n > 3$	■	$r > \frac{1+\sqrt{8n-23}}{2}$	$r > \frac{n}{2}$
P-AHP	$r < \frac{1+\sqrt{4n^2-12n+1}}{2}$	$r < \frac{1+\sqrt{8n-23}}{2}$	■	✗
P-BWM	$r < \frac{n^2-3n+6}{4}$	$r < \frac{n}{2}$	$r > 3$	■

4. Comparing the amount of preference information involved in AHP, BWM, P-AHP and P-BWM methods

In this section, we perform a comparison between four MCDM methods, namely, AHP, BWM, Parsimonious AHP [51] (denoted by P-AHP) and Parsimonious BWM (denoted by P-BWM) with respect to the number of pieces of preference information asked to the DM to apply it. Let us remember that P-AHP differs from P-BWM in the way the priorities of the reference alternatives are computed. Indeed, in P-AHP, they are obtained using the AHP method, that is, computing the eigenvector of the pairwise comparison matrix filled by the DM considering the reference alternatives only.

We think the comparison is fair as the type of needed information is the same (pairwise comparisons using the same scale) for all the methods in question. Let us also assume that the DM retains equivalent the cognitive effort involved in a pairwise comparison of alternatives or in a rating of one alternative on the considered criteria. Therefore, denoting by n the number of alternatives in the MCDM problem and by r the number of reference alternatives in the P-AHP and P-BWM methods (there is not any particular reason for which the number of reference alternatives in P-AHP should be different from the number of reference alternatives in P-BWM), the pieces of preference information involved in each of the considered methods is shown in Table 3.

To compare these four methods with respect to the number of pieces of preference information asked to the DM to apply them, in Table 4 we report in which case the application of the method in the row involves a lower number of pieces of preference information than the application of the method in the column. To perform such a comparison, we assume that $n \geqslant 3$. For example, the ✗ in correspondence of the (AHP, BWM) pair means that there is not any value of n for which AHP involves a lower number of pieces of preference information than BWM. Viceversa, $n > 3$ in correspondence of the (BWM, AHP) pair means that BWM application involves a lower number of pieces of preference information than AHP in all test problems having more than three alternatives. Let us observe that if $n = 3$, then, both AHP and BWM applications involve three pairwise comparisons only.

A detailed description of the way the values in the table are obtained is provided in Appendix A.

To better compare the considered methods, in Table 5 we show the number of pieces of preference information asked to the DM to apply them for some specific configurations (n, r) of MCDM problems. Analogously, to take into account a greater number of configurations, in Fig. 4 we show the number of pieces of preference information involved in the four considered methods for different (n, r) configurations, where, $n = 5, \ldots, 40$, and $r = 3, \ldots, \left\lceil \frac{n}{2} \right\rceil$. In particular, on the x-axis, we report the (n, r) configuration, while, on the y-axis, we show the number of pieces of preference information involved in the application of the four considered methods.

Table 5
Number of pieces of preference information asked the DM to apply each method for some specific (n, r) configuration.

	(5, 3)	(7, 3)	(7, 4)	(9, 4)	(9, 5)	(10, 4)	(10, 5)	(20, 5)	(30, 7)	(30, 10)
AHP	10	21	21	36	36	45	45	190	435	435
BWM	7	11	11	15	15	17	17	37	57	57
P-AHP	8	10	13	15	19	16	20	30	51	75
P-BWM	8	10	12	14	16	15	17	27	41	47

As one can see, AHP is the most expensive among the four considered methods from the cognitive point of view for the DM. Moreover, the following points can be observed:

- as shown in Table 4, the P-BWM application is always preferable to the P-AHP one,
- BWM and P-BWM applications involve a similar cognitive effort for problems with a number of alternatives up to ten, while P-BWM is more parsimonious than BWM for problems with a great number of alternatives ((20,5), (30,7) and (30,10)). Of course, as shown in Table 4, BWM could involve a lower cognitive effort than P-BWM but, only if the number of reference levels was greater than $n/2$ being counter-intuitive,
- for small problems, BWM application involves a cognitive effort comparable to that one of the P-AHP application and in few cases lower; for big size problems (considering $n \geqslant 20$), the comparison between the two methods is strictly dependent on the number of reference evaluations used in P-AHP. For example, considering $n = 30$, if $r = 7$, then, P-AHP is better than BWM, while, if $r = 10$, then, BWM is much better than P-AHP.

The comparison between the considered methods is performed on the assumption that a pairwise comparison of alternatives or a rating of one alternative on a particular criterion involve the same cognitive effort from the part of the DM. For this reason, looking at Table 5 and, in particular, at the (10,5) configuration, we stated that the cognitive effort asked the DM in the BWM and P-BWM application is the same. Indeed, in both cases, 17 pieces of preference information are asked to the DM. The main difference is, however, that the BWM application asks for 17 pairwise comparisons, while, the P-BWM application asks for the rating of the 10 alternatives and 7 pairwise comparisons to the DM. As we show in the next section, our claim is that providing a pairwise comparison or a rating is not the same for the DM and mixing them (as done in the P-BWM method) could be beneficial for the application of the method.

5. Experiments and detailed comparison

To check the reliability of the proposed method, we performed a comparison between the BWM and the P-BWM methods on the same problem presented in Section 3.1, that is, estimating the size of the ten European Countries shown in Fig. 2. To this aim, we submitted two questionnaires to some students of the Department of Economics and Business of the University of Catania and, in particular, students attending the *Marketing* course of the Business Economics bachelor degree (Group 1) and students attending the *Financial Mathematics* course of the Economics bachelor degree (Group 2). Students of the two groups did not have any knowledge about MCDM. We decided to submit the questionnaires to two different sets of students since they have different backgrounds and academic experiences: students of Group 1 are 3rd year students (their mean age is approximately 22 years) and, therefore, close to complete their academic studies, while, students of Group 2, are 2nd year students (their mean age is approximately 21 years) and, therefore, in the middle of their academic career. Each of the two groups was split in two parts to fill out the two questionnaires included in Appendix B:

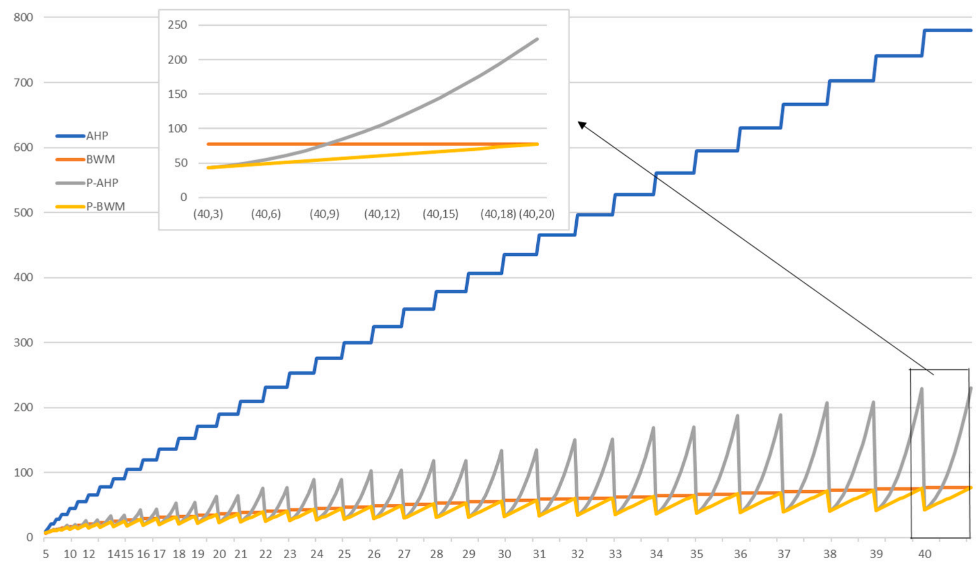

Fig. 4. Comparison between methods with respect to the amount of involved preference information in their application. In the x-axis, (n, r) configurations with $n = 5, \ldots, 40$ and $r = 3, \ldots, \left\lceil \frac{n}{2} \right\rceil$. In the y-axis, the number of pieces of preference information involved in each of the four methods for considered (n, r) configuration is shown.

Table 6
Real and normalized areas of the ten considered European Countries.

	Bosnia	Estonia	Ireland	Iceland	Lithuania	Czech Rep.	Romania	Slovenia	Spain	Switzerland
Real area km^2	51,209	45,227	70,273	103,000	65,300	78,871	238,397	20,273	505,992	41,284
$Area^{real}_{j,norm}$	2.526	2.231	3.466	5.081	3.221	3.890	11.759	1	24.959	2.036

- 54 students of Group 1 were asked to apply the BWM to the Countries shown in Fig. 2 to get their priorities, while 40 students of Group 1 were asked to apply the P-BWM described in Section 3,
- 41 students of Group 2 were asked to apply the BWM, while 47 students belonging to Group 2 were asked to apply the P-BWM.

The real size of the ten European Countries subject of the two questionnaires is shown in the first row of Table 6.

To compare the priorities of the Countries obtained applying the BWM or the P-BWM methods with the real area of the ten considered European Countries, we perform a normalization of these areas by dividing the area of each Country for the area of Slovenia so that its normalized value becomes 1 (the normalized area of the ten Countries is shown in the second row of Table 6). To perform such a comparison, after applying the BWM or the P-BWM, the following steps have to be done:

(1) For each questionnaire, the priorities of the ten Countries obtained by the BWM (w_j^{BWM}) or by the P-BWM (w_j^{P-BWM}) are normalized by dividing them by the Slovenia priority, that is:
$$w_{j,norm}^{BWM} = \frac{w_j^{BWM}}{w_{Slovenia}^{BWM}}, \qquad w_{j,norm}^{P-BWM} = \frac{w_j^{P-BWM}}{w_{Slovenia}^{P-BWM}}.$$
Considering the example shown in Section 3.1, the priorities obtained by the P-BWM application and reported in the first row of Table 7 are normalized obtaining the values in the second row of the same Table.

(2) The Mean Squared Error (MSE) between the vector of the normalized areas in the last column of Table 6 and the vector of the normalized estimated areas $w_{j,norm}$ ($w_{j,norm}^{BWM}$ or $w_{j,norm}^{P-BWM}$) obtained as described above is computed as follows:
$$MSE = \frac{1}{10}\sum_{j=1}^{10}\left(Area^{real}_{j,norm} - w_{j,norm}\right)^2. \tag{7}$$

For example, the MSE computed between the vector of normalized areas shown in the last row of Table 6 and the vector of normalized priorities obtained by the P-BWM application and shown in the second row of Table 7 is equal to 23.78. Moreover, the Maximum Absolute Error (MAE) between the normalized areas and the same vectors is computed as shown in Eq. (8)

$$MAE = \max_{j=1,\ldots,10}\left|Area^{real}_{j,norm} - w_{j,norm}\right|. \tag{8}$$

5.1. Results

After removing the questionnaires filled by the students presenting a CR^I greater than the considered threshold defined by [39], we applied the BWM or the P-BWM to the information included in the questionnaires for the two different groups. In Table 8 we report the average and standard deviation of the MSE and of the MAE computed between the vector of the normalized real areas and the vector of the normalized priorities obtained by the different versions of the BWM and P-BWM. Let us observe that BWM differs from $BWM_{Interval}$ and $BWM_{Barycenter}$ (analogously $P-BWM$ differs from $P-BWM_{Interval}$ and $P-BWM_{Barycenter}$) only on the way the Best-Other and Other-Worst vectors provided by the DM are used to infer the priority vector and not on the way the method is applied to infer the DM's preferences.

As one can see from the values reported in the Table, even if the BWM and P-BWM involve the same number of pieces of preference information (17, as shown in Table 5), the results obtained by P-BWM are clearly better than those obtained by BWM and this does not depend on the specific group to which the two questionnaires have been submitted. Indeed, at the global level (considering Groups 1 and 2 together) the average MSE between the vectors of normalized real areas and the normalized priorities obtained by P-BWM is 25.350, while, the average MSE obtained by BWM is more than double (51.809).

Table 7
Priorities of the ten Countries and normalized values.

Bosnia	Estonia	Ireland	Iceland	Lithuania	Czech Rep.	Romania	Slovenia	Spain	Switzerland	
$u(r(\cdot))$	0.076	0.071	0.110	0.205	0.123	0.144	0.266	0.044	0.470	0.076
$w_{j,norm}^{P-BWM}$	1.740	1.629	2.508	4.665	2.815	3.275	6.055	1.000	10.725	1.740

Table 8
Mean and standard deviation of the MSE and MAE computed between the vector of the normalized real areas and the vector of the normalized priorities obtained by the BWM or the P-BWM.

(a) Global (Group 1 + Group 2)				
	Average MSE	StD MSE	Average MAE	StD MAE
BWM	51.809	7.621	19.835	1.344
$BWM_{Interval}$	51.022	5.442	21.112	0.946
$BWM_{Barycenter}$	47.681	5.958	19.920	1.204
P-BWM	25.350	3.757	14.473	0.850
$P-BWM_{Interval}$	24.906	3.937	14.473	0.850
$P-BWM_{Barycenter}$	24.956	4.122	14.473	0.850
(b) Group 1				
	Average MSE	StD MSE	Average MAE	StD MAE
BWM	52.252	6.671	19.829	1.178
$BWM_{Interval}$	51.010	4.215	21.076	0.758
$BWM_{Barycenter}$	47.378	4.823	19.719	1.055
P-BWM	26.137	4.827	14.564	1.099
$P-BWM_{Interval}$	25.739	5.007	14.564	1.099
$P-BWM_{Barycenter}$	26.081	4.944	14.564	1.099
(c) Group 2				
	Average MSE	StD MSE	Average MAE	StD MAE
BWM	51.351	8.589	19.842	1.517
$BWM_{Interval}$	51.034	6.553	21.149	1.122
$BWM_{Barycenter}$	47.996	7.017	20.128	1.328
P-BWM	24.594	2.166	14.385	0.520
$P-BWM_{Interval}$	24.105	2.369	14.385	0.520
$P-BWM_{Barycenter}$	23.875	2.842	14.385	0.520

Moreover, the P-BWM presents a lower standard deviation than the BWM (3.757 vs 7.621). The same can be stated considering the MAE since applying the BWM, the average MAE is 19.835, while, it is 14.473 applying the P-BWM. Also in this case, the standard deviation of the MAE for the P-BWM is lower than the one observed for the BWM (0.850 vs 1.344) showing a greater stability in the obtained results.

As previously underlined, the same behavior can be observed in both groups where the values of the average MSE and MAE as well as the standard deviation of the MSE and MAE obtained by applying the P-BWM are lower than the corresponding values obtained by BWM. In Group 2 the average MSE obtained by the P-BWM is 24.594, while, the one obtained by the BWM application is 51.351. As to the standard deviation of the MSE, the one observed for the P-BWM is 2.166, while the one observed for the BWM is 8.580. Analogously, the average MAE obtained by the P-BWM is 14.385 with a standard deviation of 0.520, while the ones obtained applying the BWM are 19.842 and 1.517, respectively.

Regarding Group 1, the average MSE obtained by the P-BWM application is almost half of the same index obtained by the BWM (26.137 vs 52.252) and the standard deviation of the MSE obtained by the P-BWM application is 4.827 against the 6.671 obtained in correspondence of the BWM. Looking at the MAE, on the one hand, the average values obtained by the two methods are very close to the ones observed for Group 2, while, on the other hand, the standard deviation of the MAE values for the P-BWM is 1.099 versus the 1.178 obtained in correspondence of the BWM.

As described in the previous section, we checked also the results obtained inferring differently the priority vectors. On the one hand, we denoted by $BWM_{Interval}$ and $P-BWM_{Interval}$ the methods based on BWM and P-BWM in which the inferred priority vector is $\mathbf{w}^{Int}$; on the other hand, we denoted by $BWM_{Barycenter}$ and $P-BWM_{Barycenter}$ the methods based on BWM and P-BWM in which the inferred priority vector is $\mathbf{w}^{Bar}$.

The first thing that can be observed looking at Table 8 is that in all cases (Global, Group 1 and Group 2) the P-BWM versions ($P-BWM$, $P-BWM_{Interval}$ and $P-BWM_{Barycenter}$) perform better than the corresponding BWM versions (BWM, $BWM_{Interval}$ and $BWM_{Barycenter}$). Going more in depth comparing the three parsimonious versions of BWM, we cannot observe a very big difference. At global level, $P-BWM_{Interval}$ performs slightly better than $P-BWM$ and $P-BWM_{Barycenter}$ both, in terms of average MSE and MAE, while $P-BWM$ presents the lowest standard deviation for both indicators. Considering the two groups separately, again, the results obtained by the three methods are quite similar.

As to the comparison between the BWM versions, instead, one can observe a significant advantage of $BWM_{Barycenter}$ with respect to both $BWM_{Interval}$ and BWM at global level as well as considering the two groups differently. Indeed, while the average MSE observed for $BWM_{Interval}$ is lower than the one observed for BWM but very similar, the difference between the MSE values obtained by $BWM_{Interval}$ and $BWM_{Barycenter}$ is greater than 3 in all cases. This shows that $BWM_{Barycenter}$ performs well than the other two considered versions of BWM.

5.2. Statistical tests on the obtained results

To check if the difference between the results obtained by the different versions of BWM and P-BWM are significant from the statistical point of view, we performed two versions of the 2-sample Kolmogorov–Smirnov test at the 5% significance level [52]: (i) in the first version, we test the null hypothesis that the cumulative distribution functions of the MSE values obtained by the two methods are equal versus the alternative hypothesis that the cumulative distribution functions are different (we call this first version, "Equal" Test); (ii) in the second version, we test the null hypothesis that the cumulative distribution function of the MSE values obtained by the first method is smaller or equal than the cumulative distribution function of MSE values obtained by the second method versus the alternative hypothesis that the first cumulative distribution function is greater than the second (we call this first version, "Larger" Test). Let us observe that, in this case, the fact that the cumulative distribution function of the MSE values obtained by a method is greater than the cumulative distribution function of the MSE values obtained by a second method means that the first is better than the second. The two tests are also performed on the MAE distributions obtained by the different versions of BWM and P-BWM.

Considering the results obtained by the equal test (see Table 9), we can observe the following:

- At the global level the difference between BWM and $BWM_{Interval}$ is not significant from the statistical point of view, while, the difference between all other pairs of methods is statistically significant;
- Considering Group 1, again, the difference between BWM and $BWM_{Interval}$ is not statistically significant and the same happens in comparing the three versions of $P-BWM$. The difference between all the other pairs of methods is, instead, significant from the statistical point of view;

Table 9
First version of the Kolmogorov–Smirnov Test for the MSE values: "Equal" Test. $h = 0$ means that the null hypothesis is not rejected (the cumulative distributions of the MSE values obtained by the two methods are equal), while, $h = 1$ means that the null hypothesis is rejected in favor of the alternative hypothesis (the cumulative distributions of MSE values obtained by the two methods are different).

(a) Global (Group 1 + Group 2)

h/p-value	BWM	$BWM_{Interval}$	$BWM_{Barycenter}$	$P-BWM$	$P-BWM_{Interval}$	$P-BWM_{Barycenter}$
BWM	■	0/0.3345	1/0.0115	1/0.0000	1/0.0000	1/0.0000
$BWM_{Interval}$	■	■	1/0.0062	1/0.0000	1/0.0000	1/0.0000
$BWM_{Barycenter}$	■	■	■	1/0.0000	1/0.0000	1/0.0000
$P-BWM$	■	■	■	■	1/0.0001	1/0.0008
$P-BWM_{Interval}$	■	■	■	■	■	1/0.0290
$P-BWM_{Barycenter}$	■	■	■	■	■	■

(b) Group 1

h/p-value	BWM	$BWM_{Interval}$	$BWM_{Barycenter}$	$P-BWM$	$P-BWM_{Interval}$	$P-BWM_{Barycenter}$
BWM	■	0/0.2003	1/0.0259	1/0.0000	1/0.0000	1/0.0000
$BWM_{Interval}$	■	■	1/0.0046	1/0.0000	1/0.0000	1/0.0000
$BWM_{Barycenter}$	■	■	■	1/0.0000	1/0.0000	1/0.0000
$P-BWM$	■	■	■	■	0/0.2160	0/0.2160
$P-BWM_{Interval}$	■	■	■	■	■	0/0.8608
$P-BWM_{Barycenter}$	■	■	■	■	■	■

(c) Group 2

h/p-value	BWM	$BWM_{Interval}$	$BWM_{Barycenter}$	$P-BWM$	$P-BWM_{Interval}$	$P-BWM_{Barycenter}$
BWM	■	0/0.9270	0/0.3213	1/0.0000	1/0.0000	1/0.0000
$BWM_{Interval}$	■	■	0/0.1844	1/0.0000	1/0.0000	1/0.0000
$BWM_{Barycenter}$	■	■	■	1/0.0000	1/0.0000	1/0.0000
$P-BWM$	■	■	■	■	1/0.0004	1/0.0013
$P-BWM_{Interval}$	■	■	■	■	■	1/0.0104
$P-BWM_{Barycenter}$	■	■	■	■	■	■

Table 10
Second version of the Kolmogorov–Smirnov Test for the MSE values: "Greater" Test. $h = 0$ means that the null hypothesis is not rejected (the cumulative distribution of the MSE values obtained by the method on the row is therefore smaller or equal to the cumulative distribution of the MSE values obtained by the method in the column), while, $j = 1$ means that the null hypothesis is rejected in favor of the alternative hypothesis (the cumulative distribution of the MSE values obtained by the method in the row is larger than the cumulative distribution of the MSE values obtained by the method in the column).

(a) Global (Group 1 + Group 2)

h/p-value	BWM	$BWM_{Interval}$	$BWM_{Barycenter}$	$P-BWM$	$P-BWM_{Interval}$	$P-BWM_{Barycenter}$
BWM	■	■	0/0.7517	0/1.0000	0/1.0000	0/1.0000
$BWM_{Interval}$	■	■	0/1.0000	0/1.0000	0/1.0000	0/1.0000
$BWM_{Barycenter}$	1/0.0058	1/0.0031	■	0/1.0000	0/1.0000	0/1.0000
$P-BWM$	1/0.0000	1/0.0000	1/0.0000	■	0/0.9786	0/0.3470
$P-BWM_{Interval}$	1/0.0000	1/0.0000	1/0.0000	1/0.0001	■	0/0.3470
$P-BWM_{Barycenter}$	1/0.0000	1/0.0000	1/0.0000	1/0.0004	1/0.0145	■

(b) Group 1

h/p-value	BWM	$BWM_{Interval}$	$BWM_{Barycenter}$	$P-BWM$	$P-BWM_{Interval}$	$P-BWM_{Barycenter}$
BWM	■	■	0/0.9647	0/1.0000	0/1.0000	0/1.0000
$BWM_{Interval}$	■	■	0/0.9647	0/1.0000	0/1.0000	0/1.0000
$BWM_{Barycenter}$	1/0.0129	1/0.0023	■	0/1.0000	0/1.0000	0/1.0000
$P-BWM$	1/0.0000	1/0.0000	1/0.0000	■	■	■
$P-BWM_{Interval}$	1/0.0000	1/0.0000	1/0.0000	■	■	■
$P-BWM_{Barycenter}$	1/0.0000	1/0.0000	1/0.0000	■	■	■

(c) Group 2

h/p-value	BWM	$BWM_{Interval}$	$BWM_{Barycenter}$	$P-BWM$	$P-BWM_{Interval}$	$P-BWM_{Barycenter}$
BWM	■	■	■	0/1.0000	0/1.0000	0/1.0000
$BWM_{Interval}$	■	■	■	0/1.0000	0/1.0000	0/1.0000
$BWM_{Barycenter}$	■	■	■	0/1.0000	0/1.0000	0/1.0000
$P-BWM$	1/0.0000	1/0.0000	1/0.0000	■	0/0.9574	0/0.4986
$P-BWM_{Interval}$	1/0.0000	1/0.0000	1/0.0000	1/0.0002	■	0/0.4986
$P-BWM_{Barycenter}$	1/0.0000	1/0.0000	1/0.0000	1/0.0006	1/0.0052	■

- With respect to Group 2, the difference between the three BWM versions is not significant from the statistical point of view, while the difference between all other pairs of methods is statistically significant.

Performing the greater test to the pairs of methods for which the difference between the cumulative distributions of the MSE values is significant from the statistical point of view, we can state the following (see Table 10):

- At the global level, the cumulative distribution of the MSE value obtained by $P-BWM_{Barycenter}$ is greater (therefore better) than the one obtained by all the other methods; as second best, we can consider $P-BWM_{Interval}$, followed by $P-BWM$ and, therefore, by $BWM_{Barycenter}$;
- Considering Group 1, the three P-BWM versions are better than all BWM versions, and $BWM_{Barycenter}$ is better than both BWM and $BWM_{Interval}$;

- With respect to Group 2, we have again that $P-BWM_{Barycenter}$ is better than all other methods, followed by $P-BWM_{Interval}$ and $P-BWM$.

Considering the distributions of the MAE values, we can state the following (to save space, we included the tables with the values obtained by the two tests in Appendix C):

- The three $P-BWM$ versions are equivalent considering the two groups together as well as separately; analogously, the difference between the distributions of MAE values obtained by BWM and $BWM_{Barycenter}$ is not statistically significant at global and partial way;
- Considering the larger test, both at the global and partial level, we can state that P-BWM versions are better than BWM versions and BWM is better than $BWM_{Interval}$.

5.3. Some comments

The results obtained by the BWM and P-BWM application to two questionnaires submitted to two different groups of university students evince the goodness of our proposal with respect to the original BWM. Even if the two considered questionnaires involve the same number of pieces of preference information (17) their nature is different. In the BWM application, the students were asked to provide 17 pairwise comparisons, while, in the P-BWM application, students were asked to provide 10 ratings and only 7 pairwise comparisons. Looking at the values presented in the previous section, this difference in the type of preference asked to the students made the application of the P-BWM simpler and more reliable than the BWM application.

As to the comparison between the different BWM and P-BWM versions obtained considering the interval and the barycenter as priority vectors, we observed that $P-BWM_{Barycenter}$ obtains the best results among the six methods. In particular, it is better than the other two P-BWM versions at the global level and considering Group 2, while the difference between the obtained MSE values is not statistically significant with respect to Group 1.

6. Conclusions

In this study, we introduced a new version of the Best–Worst Method (BWM), parsimonious BWM, to handle decision-making problems involving a large set of alternatives. Following this approach, a manageable subset of the alternatives is chosen for conducting the pairwise comparison following the BWM steps by the Decision-Maker (DM). The priorities found by this subset of alternatives, along with the rating of the whole set (also determined by the DM), are used to rank the whole set of alternatives. The idea behind the procedure we propose is that the errors associated with the evaluation of a large number of alternatives can be corrected by taking into account the priorities obtained from the pairwise comparison of a limited number of well-distributed reference alternatives. We conducted an experiment to test the performance of the new approach, and as the results show, it performs very well against the original BWM. We also showed that the new approach does not require more information pieces from the DM, which is in line with the main philosophy of the original BWM. We conducted a detailed analysis to reach this conclusion.

Summarizing, we detail the contributions of the paper as follows:

- We introduced a parsimonious version of the BWM called P-BWM, which allows for determining the priorities of alternatives in cases where their large quantity makes it impractical to use the original BWM method,
- We examined AHP, P-AHP, BWM, and P-BWM in terms of the amount of preference information required from the DM to utilize them. This allows them to determine the most efficient method based on the number of alternatives/criteria available,
- An experiment conducted by students was performed to compare the performance of the P-BWM with the BWM.

We have some interesting ideas for future research direction. First, although we found outstanding performance for the parsimonious BWM in an experiment (where the subset is defined by the researchers and it is fixed for all the subjects), we think the performance of the new approach might even improve if each DM is free to choose a subset herself. In real-world settings, a DM might feel more comfortable/knowledgeable about some particular alternatives. This would be a reasonable criterion to compose the subset from those alternatives. Such a choice could, in principle, lead to more reliable priorities for that subset, hence a more reliable rating of the alternatives in the whole set. This idea could be investigated in a new experimental study. Second, the composition of the subset might relate to some biases that can be investigated in future studies. For instance, having the alternatives with the best and worst performance in the subset might lead to different results than when such alternatives are not in the set. Finally, while the problem in our study is simple (for the sake of experiments), more sophisticated cases need to be studied (alternatives with different dimensions) to test the performance of the new approach. The initial experiments conducted in the paper were carried out with students, but it is recommended that future studies involve actual DMs who are experts in the relevant field of application. This will allow for feedback on any potential bias or limitations of the method when applied in a practical setting.

We plan to continue our research on parsimonious BWM in different directions:

- comparing the parsimonious BWM with other multicriteria scaling methods such as SMART [53] and SWING procedures [54],
- developing a customized version of the method for specific real-world applications in relevant domains such as multicriteria evaluation of sustainable development [55],
- improving the procedure to select the reference points, paying attention to the aspects related to the cooperation with the DM in this specific task,
- developing a procedure to apply the parsimonious BWM to the elicitation of weights in case of decision problems with many criteria, possibly hierarchically organized.

In summary, we would like to work on the idea of decision support procedures permitting to obtain better decisions by reducing the amount of the required preference information but increasing its salience. This seems a very interesting research perspective for the whole domain of multiple criteria decision aiding.

CRediT authorship contribution statement

Salvatore Corrente: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Salvatore Greco:** Methodology, Writing – original draft, Writing – review & editing. **Jafar Rezaei:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

Salvatore Corrente and Salvatore Greco acknowledge the support of the Ministero dell'Istruzione, dell'Universitá e della Ricerca (MIUR) - PRIN 2017, project "Multiple Criteria Decision Analysis and Multiple Criteria Decision Theory", grant 2017CY2NCA. The authors are grateful to students attending the Marketing and Financial Mathematics courses of the Department of Economics and Business of the University of Catania for their helpful contribution in filling out the questionnaires presented in the paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.omega.2024.103075.

References

[1] Keeney RL, Raiffa H. Decisions with multiple objectives: Preferences and value tradeoffs. New York: J. Wiley; 1976.
[2] Roy B. Multicriteria methodology for decision aiding. Nonconvex optimization and its applications, Dordrecht: Kluwer Academic Publishers; 1996.
[3] Brans J-P, Vincke P, Mareschal B. How to select and how to rank projects: The PROMETHEE method. European J Oper Res 1986;24(2):228–38.
[4] Saaty T. A scaling method for priorities in hierarchical structures. J Math Psych 1977;15(3):234–81.
[5] Rezaei J. Best-worst multi-criteria decision-making method. Omega 2015;53:49–57.
[6] Rezaei J. Best-worst multi-criteria decision-making method: Some properties and a linear model. Omega 2016;64:126–30.
[7] Brunelli M, Rezaei J. A multiplicative best–worst method for multi-criteria decision making. Oper Res Lett 2019;47(1):12–5.
[8] Mohammadi M, Rezaei J. Bayesian best-worst method: A probabilistic group decision making model. Omega 2020;96:102075.
[9] Liang Y, Ju Y, Tu Y, Rezaei J. Nonadditive best-worst method: Incorporating criteria interaction using the choquet integral. J Oper Res Soc 2023;74(6):1495–506.
[10] Guo S, Zhao H. Fuzzy best-worst multi-criteria decision-making method and its applications. Knowl-Based Syst 2017;121:23–31.
[11] Rezaei J, Arab A, Mehregan M. Analyzing anchoring bias in attribute weight elicitation of SMART, Swing, and best-worst method. Int Trans Oper Res 2022.
[12] Rezaei J, Arab A, Mehregan M. Equalizing bias in eliciting attribute weights in multiattribute decision-making: experimental research. J Behav Decis Mak 2022;35(2):e2262.
[13] Wu Q, Liu X, Zhou L, Qin J, Rezaei J. An analytical framework for the best–worst method. Omega 2024;123:102974.
[14] Kusi-Sarpong S, Orji IJ, Gupta H, Kunc M. Risks associated with the implementation of big data analytics in sustainable supply chains. Omega 2021;105:102502.
[15] Rezaei J, Nispeling T, Sarkis J, Tavasszy L. A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method. J Clean Prod 2016;135:577–88.
[16] Švadlenka L, Bošković S, Jovčić S, Simic V, Kumar S, Zanne M. Third-party logistics provider selection for sustainable last-mile delivery: A case study of E-shop in Belgrade. J Urb Dev Manag 2023;2(1):1–13.
[17] Mi X, Tang M, Liao H, Shen W, Lev B. The state-of-the-art survey on integrations and applications of the best worst method in decision making: Why, what, what for and what's next? Omega 2019;87:205–25.
[18] Mohammadi M, Rezaei J. Hierarchical evaluation of criteria and alternatives within BWM: A Monte Carlo approach. In: Advances in best-worst method: proceedings of the second international workshop on best-worst method. Springer; 2022, p. 16–28.
[19] Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 1956;63(2):81–97.
[20] Saaty TL, Ozdemir MS. Why the magic number seven plus or minus two. Math Comput Modelling 2003;38(3):233–44.
[21] Montibeller G, Von Winterfeldt D. Cognitive and motivational biases in decision and risk analysis. Risk Anal 2015;35(7):1230–51.
[22] Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. Science 1974;185(4157):1124–31.
[23] Helson H. Adaptation-level as frame of reference for prediction of psychophysical data. Am J Psychol 1947;60(1):1–29.
[24] Helson H. Adaptation-level theory: an experimental and systematic approach to behavior. New York: Harper and Row; 1964.
[25] Parducci A. Category judgment: a range-frequency model. Psychol Rev 1965;72(6):407.
[26] Parducci A. Happiness, pleasure, and judgment: The contextual theory and its applications. Lawrence Erlbaum Associates, Inc; 1995.
[27] Liu NC, Cheng Y. The academic ranking of world universities. Higher Educ Europe 2005;30(2):127–36.
[28] Rowan AN. World happiness report 2023. WellBeing News 2023;5(3):1.
[29] Haakenstad A, et al. Assessing performance of the healthcare access and quality index, overall and by select age groups, for 204 countries and territories, 1990–2019: a systematic analysis from the global burden of disease study 2019. Lancet Global Health 2022;10(12):e1715–43.
[30] Sachs JD, Kroll C, Lafortune G, Fuller G, Woelm F. Sustainable development report 2022, Cambridge University Press; 2022.
[31] Doumpos M, Zopounidis C. Multicriteria analysis in finance, Springer; 2014.
[32] Soebarto VI, Williamson TJ. Multi-criteria assessment of building performance: theory and implementation. Build Environ 2001;36(6):681–90.
[33] Natividade-Jesus E, Coutinho-Rodrigues J, Antunes CH. A multicriteria decision support system for housing evaluation. Decis Support Syst 2007;43(3):779–90.
[34] Cinelli M, Coles SR, Kirwan K. Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. Ecol Indic 2014;46:138–48.
[35] Dorfleitner G, Halbritter G, Nguyen M. Measuring the level and risk of corporate responsibility–An empirical comparison of different ESG rating approaches. J Asset Manag 2015;16:450–66.
[36] Belton V, Stewart TJ. Multiple criteria decision analysis: an integrated approach. Springer; 2002.
[37] Greco S, Ehrgott M, Figueira JR. Multiple criteria decision analysis: State of the art surveys, Berlin: Springer; 2016.
[38] Greco S, Matarazzo B, Słowiński R. Rough sets theory for multicriteria decision analysis. European J Oper Res 2001;129(1):1–47.
[39] Liang F, Brunelli M, Rezaei J. Consistency issues in the best worst method: Measurements and thresholds. Omega 2020;96.
[40] Smith RL. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. Oper Res 1984;32:1296–308.
[41] Tervonen T, Figueira JR, Lahdelma R, Almeida Dias J, Salminen P. A stochastic method for robustness analysis in sorting problems. European J Oper Res 2009;192(1):236–42.
[42] Arcidiacono SG, Corrente S, Greco S. Scoring from pairwise winning indices. Comput Oper Res 2023;157:106268.
[43] Corrente S, Greco S, Rezaei J. A parsimonious extension of the best-worst method. In: 26th International Conference on Multiple Criteria Decision Making. 2022.
[44] Moslem S. A novel parsimonious best worst method for evaluating travel mode choice. IEEE Access 2023;11:16768–73.
[45] Anderson NH. On the role of context effects in psychophysical judgment. Psychol Rev 1975;82(6):462.
[46] Mellers BA, Birnbaum MH. Loci of contextual effects in judgment. J Exp Psychol [Hum Percept] 1982;8(4):582.
[47] Riskey DR, Parducci A, Beauchamp GK. Effects of context in judgments of sweetness and pleasantness. Percept Psychophys 1979;26:171–6.
[48] Anderson NH. Algebraic models in perception. In: Handbook of perception, vol. 2, 1974, p. 215–98.
[49] Fechner GT. Elemente der psychophysik, vol. 2, Breitkopf u. Härtel; 1860.
[50] Stevens SS. On the psychophysical law. Psychol Rev 1957;64(3):153.
[51] Abastante F, Corrente S, Greco S, Ishizaka A, Lami IM. A new parsimonious AHP methodology: assigning priorities to many objects by comparing pairwise few reference objects. Expert Syst Appl 2019;127:109–20.
[52] Conover WJ. Practical nonparametric statistics, vol. 350, John Wiley & Sons; 1999.
[53] Edwards W. How to use multiattribute utility measurement for social decisionmaking. IEEE Trans Syst Man Cybern 1977;7(5):326–40.
[54] von Winterfeldt D, Edwards W. Decision analysis and behavioral research, Cambridge University Press; 1986.
[55] Boggia A, Cortina C. Measuring sustainable development using a multi-criteria model: A case study. J Environ Manag 2010;91(11):2301–6.