Omega 144 (2026) 103607

Contents lists available at ScienceDirect

Omega

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

Pivot-Linked Elicitation: A unified approach to hierarchical weight elicitation and its application to best–worst methods ☆

River Huang [a,*], Matteo Brunelli [b]

[a] Laboratory for Energy Systems Analysis, PSI Centers for Nuclear Engineering & Sciences and Energy & Environmental Sciences, Forschungsstrasse 111, Villigen PSI, 5232, Switzerland

[b] Department of Industrial Engineering, University of Trento, Via Sommarive 9, 38123 Trento, Italy

ARTICLE INFO

Keywords:
Weight elicitation
Multiple criteria decision analysis
PIvot-linked elicitation
Best–worst method

ABSTRACT

We study weight elicitation in Multiple Criteria Decision Analysis (MCDA) when decision criteria are organized hierarchically. We show that the conventional practice of eliciting weights separately at each level can lead to practical and conceptual limitations, including reliance on abstract upper-level judgments and the fragmentation of consistency across independent sub-models. To address these issues, we propose the PIvot-Linked Elicitation (PILE) framework, which links groups through a small set of pivot criteria and integrates intra- and inter-group preference information into a single model. We instantiate PILE for best–worst methods, deriving nonlinear and linear PILE–Best-Worst Method (BWM) formulations, and illustrate the approach on two published use cases. The results demonstrate how PILE preserves the cognitive benefits of hierarchical structuring while enabling a more holistic and consistency-aware elicitation.

1. Introduction

Most of the Multiple Criteria Decision Analysis (MCDA) methods require, as input, the weights of the criteria so that the contributions of different facets of the alternatives are properly scaled. Consequently, methodologies to find appropriate weights (often called weighting methods) are fundamental. Weighting methods can be classified into three categories: subjective, objective and mixed. *Subjective methods* [1] are based on the principle that the determination of weights cannot be detached from (and must depend on) the subjective judgments of Decision Makers (DMs) and experts. Conversely, *objective methods* suggest to do without subjective judgments and base the determination of the weights only on the characteristics of alternatives. *Mixed methods*, as their name suggests, include elements of both subjective and objective methods. Our focus is on subjective methods: whereas we acknowledge that criteria weights can have multiple interpretations [2], we refrain from believing that they should be extracted from the alternatives and we reckon that they should exclusively reflect the preferences of the DM.

A number of methods have been proposed to elicit weights from a DM. The most straightforward one is the direct assessment method. However, this method is naive and it does not cope well with complex problems and judgment biases. One more preferable approach is based on pairwise comparisons. These comparisons are intuitive for the DM and useful for breaking down complex problems into simple questions involving only two criteria [3]. Given n criteria, a properly chosen set of $(n-1)$ pairwise comparisons between them suffices to find their weights. We acknowledge the large number of methods that require only $(n-1)$ comparisons and that have been recently proposed, but an analyst should require more comparisons for two main reasons [4]: more comparisons create redundancy and this (i) increases the stability of the results and (ii) allows the estimation of the consistency of judgments, which could be considered as one of the indicators of the expertise of DMs [5]. Albeit not free from criticism, methods such as the Analytic Hierarchy Process (AHP) have been proposed in accordance to this logic and are nowadays established and used to elicit weights for criteria [6]. The trade-off method [7], widely used in Multi-Attribute Value Theory (MAVT), is another well-known method for determining criteria weights.

Thanks to its simplicity and its capacity to mitigate biases [8], the Best-Worst Method (BWM) has emerged as a promising alternative for criteria weighting. Introduced by Rezaei [9], BWM uses a structured pairwise comparison approach that requires fewer comparisons than AHP but guides the DMs towards more consistent comparisons. In BWM, the DM identifies the best (most important) and worst (least important) criteria, and provides two sets of comparisons: the preference of the best criterion over each of the other criteria, and the preference

☆ Area: Decision Analysis and Preference Modeling. This manuscript was processed by Associate Editor Evangelos Triantaphyllou.

* Corresponding author.
E-mail addresses: river.huang@psi.ch (R. Huang), matteo.brunelli@unitn.it (M. Brunelli).

https://doi.org/10.1016/j.omega.2026.103607
Received 17 December 2025; Accepted 28 May 2026
Available online 2 June 2026

of each criterion over the worst criterion. These two vectors of comparisons are then used in an optimization model to derive criteria weights. The original BWM formulation is typically solved via a min–max linear programming model that minimizes the maximum deviation between implied weight ratios and the DM's comparisons [10].

However, as the number of criteria increases, the number of pairwise comparisons in either AHP and BWM increases the DM's cognitive burden. As recalled by Smith and Dyer [11], there are decision problems with a very large number of criteria/attributes. For example, Parnell et al. [12] studied a problem with 134 attributes. In similar cases, criteria are often arranged hierarchically: high-level objectives are decomposed into more specific sub-criteria. For instance, in an energy decision-support problem, the top level may follow the three sustainability pillars: environmental, economical, and social, with each pillar further broken down into several sub-criteria [13]. The hierarchical structure helps the DM elicit weights in a more organized, cognitively manageable way [14]. Also, it reduces the number of required pairwise comparisons. Let us note that, as we shall discuss later, at present some well-known methods still seem to be ill-equipped to work on hierarchies of criteria. For example, on a recent analysis of the BWM, Rezaei [15] indicated hierarchical modeling as one future research direction. More specifically, he claimed that "A promising research direction is therefore to operationalize BWM across an entire multi-level structure [...]".

In hierarchical settings, a straightforward approach is to conduct pairwise comparisons separately at each level. In a two-level hierarchy, a DM first compares higher-level criteria to obtain their weights; then, within each one of them, he/she compares the sub-criteria to derive local weights. The global weights of criteria are the product of their local weight and the weight of their parent criteria [16]. However, this hierarchical approach raises several issues. First, high-level criteria are often abstract and cognitively difficult to compare, e.g., the three pillars of sustainability. Second, judgments at each level are elicited in isolation, so cross-group importance is inferred only indirectly, with no direct linkage between sub-criteria in different groups. For instance, even if the top economic sub-criterion receives a higher weight than the top environmental sub-criterion, it does not follow that, overall, the economic pillar should carry more weight than the environmental pillar.

In this paper, we examine hierarchical weight elicitation in MCDA with a focus on BWM methods. We first analyze the conventional hierarchical approach and identify its main conceptual and practical limitations. We then introduce PIvot-Linked Elicitation (PILE), a unified framework for hierarchically structured criteria that relocates elicitation to the lowest operational level while linking groups through a small set of pivot criteria. Next, we instantiate this framework for BWM methods by deriving nonlinear and linear PILE–BWM formulations. Finally, we illustrate the approach through two use cases and discuss its broader applicability to hierarchical weight elicitation.

The remainder of the paper is organized as follows. Section 2 reviews hierarchical criteria structures in practice and discusses the main drawbacks of conventional hierarchical weight elicitation. Section 3 revisits BWM methods and introduces the modeling background used in the paper. Section 4 presents the PILE framework and its application to BWM methods, including the nonlinear and linear formulations. Section 5.1 reports the use cases and presents the resulting weights, consistency analysis, and interval analysis. Section 6 discusses the implications, advantages, and limitations of the proposed framework. Section 7 concludes, and Appendix provides the simulation-based evidence supporting the compact linear formulation.

2. Hierarchical criteria in practice

Many MCDA methods welcome a top-down approach to the definition of criteria. First, the overarching goal is defined and then, thanks to approaches like value focused thinking [17], it is the main objective is decomposed into sub-objectives which, in turn, can be split into further sub-objectives, until these latter fulfill the necessary requirements to be criteria [18]. In a simplified setting, which is the most commonly used in practice, there is a set of m objectives (often called higher-level criteria) under which n criteria are arranged (with $m \leq n$). This corresponds to considering a set of criteria $C = \{c_1, \dots, c_n\}$, whose elements can be relabeled and partitioned into m disjoint subsets,

$$\begin{aligned} G_1 &= \{c_{1,1}, \dots, c_{1,|G_1|}\} \\ \vdots &= \quad \vdots \\ G_m &= \{c_{m,1}, \dots, c_{m,|G_m|}\} \end{aligned}$$

with $G_i \cap G_j = \emptyset$ for $i \neq j$ and $\bigcup_{g=1}^{m} G_g = C$. We call $M = \{1, \dots, m\}$ its index-set. Fig. 1 shows an example of hierarchical structure for a problem that considers the 3P paradigm (Profit, Planet, People).

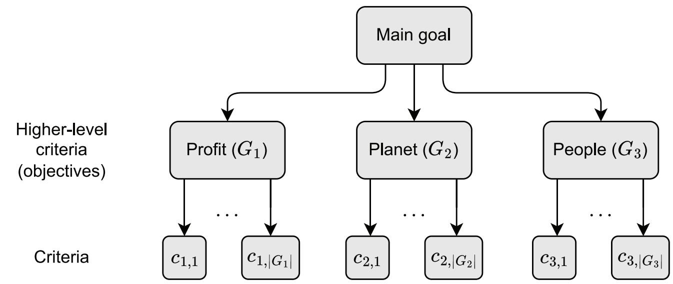


Fig. 1. A typical hierarchical structure for problems including economical, environmental and social aspects.

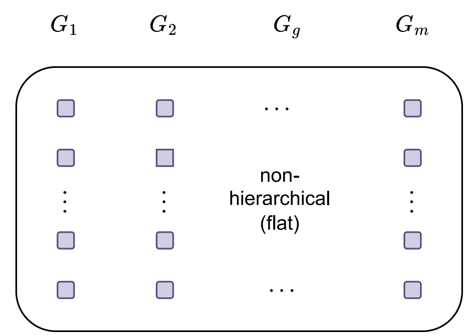


Fig. 2. If the hierarchical structure is not considered, then the hierarchy is flattened and criteria are not distinguished according to the higher-level criteria.

Determining the criteria weights is a crucial and necessary step to rescale each criterion's contribution to the main objective. Before implementing a method to carry out this step, one needs to declare whether he/she wants to consider the set of criteria as *flat* or to exploit their *hierarchical* arrangement.

Flat. One possible approach is to discard the hierarchical structure of the criteria and pairwise compare the criteria regardless of the higher-level criteria to which they belong. When a sufficient number of comparisons is collected, normalized weights can be obtained by using one of the many methods proposed in the literature. For later convenience and for comparison with other approaches, we sketch this scheme in Fig. 2.

Hierarchical. The alternative approach exploits the hierarchical arrangement of criteria. More formally, under hierarchical weight elicitation, the DM first pairwise compares criteria within each group G_g to obtain *local weights*. The local weight of criterion $c_{g,i}$ is defined as $w_{g,i}^{\text{loc}} \geq 0$ where $\sum_{i=1}^{|G_g|} w_{g,i}^{\text{loc}} = 1$ for every $g \in M$. The DM then assesses, still by means of pairwise comparisons, the importance of the

Table 1
Survey of recent BWM (or variants) with >10 criteria.

Study	Topic	#Crit.	Hier.	#Grp.	Sub-criteria per group
Güler et al. [19]	Organic wastes	28	✓	4	7/9/7/5
Huang et al. [20]	Air freight	12	✓	4	4/4/4/4
Berberoglu et al. [21]	Mining industry	11	✗	–	–
Gholamizadeh et al. [22]	Rail network safety	29	✓	8	2/3/5/5/4/4/3/3
Hossain et al. [23]	Manufacturing waste	17	✓	4	5/4/4/4
Sarwar et al. [24]	Logistics management	20	✓	5	4/3/3/4/3
Sharma and Gupta [25]	Industrialization strategies	12	✓	4	3/3/3/3
Singh and Kumar [26]	Blockchain technology	20	✓	5	4/4/3/3/3
Wu et al. [27]	Energy storage	20	✓	4	7/5/3/5

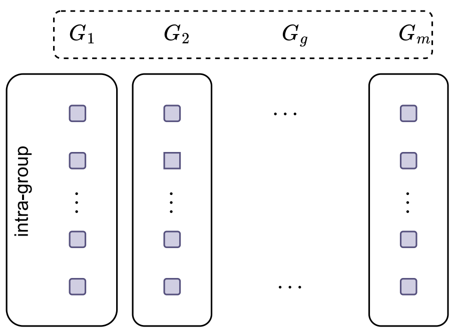


Fig. 3. Criteria weight elicitation scheme when the hierarchical structure is exploited.

groups themselves, producing higher-level criteria weights $w_g \geq 0$ with $\sum_{g\in M} w_g = 1$. Finally, the *global weight* of criterion $c_{g,i}$ is defined as:

$$w_{g,i} = w_g \cdot w^{\text{loc}}_{g,i}, \quad \forall g \in M, \ \forall i \in \{1, \ldots, |G_g|\} \tag{1}$$

which implies $\sum_{g\in M} \sum_{i=1}^{|G_g|} w_{g,i} = 1$. The set of entities that are considered in this approach is represented in Fig. 3.

By definition, methods like the AHP privilege a hierarchical arrangement of criteria. However, in practice, even methods that are, from the theoretical point of view, agnostic on this issue, seem to be used considering the hierarchical arrangement of criteria. One example is the well-known BWM, in all its variants. Table 1 shows that, from a sample of nine studies that use BWM or its variants with more than ten criteria, eight adopted a hierarchical structure.

2.1. Possible drawbacks of hierarchical weight elicitation

The hierarchical weight elicitation offers cognitive structure but, in the current usage, it may introduce the following pitfalls:

1. Keeney and Raiffa [7, §2.3.3] noted that "we can use the hierarchy in a manner that is convenient to ourselves", but they also emphasized that when higher-level criteria and objectives are introduced, the hierarchy "is often too vague for any operational purpose". In practice, answering pairwise comparisons between higher-level criteria (as it is often done in the AHP) can be cognitively challenging. Unlike lower-level criteria, higher-level ones tend to be more abstract and multidimensional. While ordinal judgments may still be sensible, cardinal assessments often demand a level of precision that experts cannot realistically provide. This difficulty is exacerbated when pairwise comparisons involve trade-offs between high-level objectives, which frequently lack the desirable properties of measurable attributes [18]. For instance, even an experienced expert may struggle to determine the extent to which an improvement in one objective (e.g., environmental quality) can compensate for a decline in another (e.g., economic performance).
2. Each time an expert articulates a preference between higher-level criteria (for instance, between economic and environmental dimensions), he/she is, in effect, taking a political stance. As observed by Favargiotti et al. [28], this can pose challenges for the experts involved. Conversely, when an expert restricts her judgments to lower-level criteria, even a statement such as "the most important economic subcriterion should receive a higher weight than the most important environmental one" does not necessarily imply that economic considerations should, overall, outweigh environmental ones.
3. The final decision outcome may become overly sensitive to pairwise comparisons made at higher levels of the criteria hierarchy. When using Saaty's scale, even a minor adjustment—such as changing a numerical judgment from "1" (indifferent) to "2" (between indifferent and moderately more preferred)—can lead to a substantial shift in results. This change suggests to double the relative weight of all criteria associated with the first dimension compared to those of the second. Combined with the inherent difficulty of comparing abstract, high-level criteria, this sensitivity poses a significant challenge that can undermine decision processes relying on hierarchical comparisons.
4. As shown in Fig. 3, the comparisons of criteria and higher-level criteria are done in isolation. Namely, the relative importance of criteria from different groups is only indirectly determined by their separate evaluations, without any direct comparison linking those groups of criteria. For example, a criterion that belongs to the dimension "environment" is never put into a direct comparison with a criterion belonging to the "social" dimension, and their weights are determined through the mediation of the comparison between their respective higher-level criteria (with all the already mentioned drawbacks).
5. Real-world optimization problems are usually approached in the most holistic and comprehensive way. For example, recent works combined maintenance and routing optimization [29], as well as tactical and strategic decisions [30]. Although decomposing problems into subproblems is often necessary to manage computational complexity, this concern rarely applies to handling preference relations. Therefore, it is difficult to justify solving separate weight-elicitation problems at different levels of a hierarchy rather than addressing a single holistic formulation.
6. If new criteria (assume at the lower level) are introduced in the hierarchy, it is necessary to compare them with criteria at the same level. Moreover, due to a possibly changed perception of higher level criteria, it may be necessary to reconsider the comparisons between higher level criteria too, with all the above mentioned complications.

3. Revisiting best–worst methods: Problem statement and motivation

BWM is a pairwise comparison technique to derive weights for a set of criteria using two specially constructed comparison vectors [9]. We shall here discuss its fundamentals, as it will serve as a starting point to develop and illustrate our approach. Suppose a DM has identified a set of n criteria $C = \{c_1, c_2, \ldots, c_n\}$. The steps of BWM can be summarized as follows:

1. **Select Best and Worst Criteria**: The DM specifies which criterion is considered the most important (best), and which is the least important (worst).
2. **Best-to-Others (BO) Comparisons**: Using a predefined scale, e.g., 1 to 9, where 1 means equal importance and 9 means extreme importance of the best over the other, the DM provides the preference of the best criterion over each of the other criteria. This yields the Best-to-Others (BO) vector $\mathbf{A}_B = (a_{B,1}, a_{B,2}, \ldots, a_{B,n})$, where $a_{B,i}$ denotes the comparison value of the best criterion B with respect to criterion c_i. By definition, $a_{B,B} = 1$.
3. **Others-to-Worst (OW) Comparisons**: Similarly, the DM provides the preference of each criterion over the worst criterion. This yields the Others-to-Worst (OW) vector $\mathbf{A}_W = (a_{1,W}, a_{2,W}, \ldots, a_{n,W})$, where $a_{i,W}$ denotes the comparison value of criterion c_i with respect to the worst criterion W. By definition, $a_{W,W} = 1$.[1]
4. **Optimal Weight Determination**: Let w_i denote the weight of the ith criterion. The goal is to find weights $w_1, \ldots, w_n$ that are as consistent as possible with the comparisons $\mathbf{A}_B$ and $\mathbf{A}_W$. In an ideal (fully consistent) case, we have $w_B/w_i = a_{B,i}$ and $w_i/w_W = a_{i,W}$ for all i. In practice, such equalities may not hold, so Rezaei [9] suggested to solve the following non-linear optimization problem

$$\left.\begin{aligned}
\text{minimize}\quad & \xi \\
\text{s.t.}\quad & \left| \frac{w_B}{w_i} - a_{B,i} \right| \le \xi, \qquad \forall i = 1, \ldots, n, \\
& \left| \frac{w_i}{w_W} - a_{i,W} \right| \le \xi, \qquad \forall i = 1, \ldots, n, \\
& \sum_{i=1}^{n} w_i = 1, \\
& w_i \ge 0, \qquad \forall i = 1, \ldots, n\ .
\end{aligned}\right\} E(\text{Non-linear BWM}) \tag{2}$$

Alternatively, Rezaei [10] proposed to replace the previous model with the following linear one.

$$\left.\begin{aligned}
\text{minimize}\quad & \xi \\
\text{s.t.}\quad & | w_B - a_{B,i}\, w_i | \le \xi, \qquad \forall i = 1, \ldots, n, \\
& | w_i - a_{i,W}\, w_W | \le \xi, \qquad \forall i = 1, \ldots, n, \\
& \sum_{i=1}^{n} w_i = 1, \\
& w_i \ge 0, \qquad \forall i = 1, \ldots, n\ .
\end{aligned}\right\} E(\text{Linear BWM}) \tag{3}$$

In both BWM formulations, the variable ξ denotes the maximum deviation from perfect consistency and is minimized by the optimization. The constraints ensure that, for every BO and OW statement, the discrepancy between the weight ratios implied by w and the DM's stated ratios does not exceed ξ. Solving the model yields the optimal weights w_i^* and the inconsistency index ξ^*; smaller values of ξ^* indicate greater internal consistency.

The BWT method [32] is a variant of the BWM that interprets each comparison as a tradeoff between two attribute levels: the DM stipulates an indifference relation between two hypothetical alternatives differing with respect to only two attributes, thereby identifying the ratio of two weights. However, the formulation of the final model can also be represented by models (2) and (3). Overall, BWT is inspired by MAVT and seems more coherent with the interpretation of weights as scaling constants and their ratios as rates of substitution, whereas the questions that are asked in BWM seem more in line with approaches where weights represent the relative importance of criteria. On top of this, multiple variants of the BWM have been proposed in the last years and it may be more appropriate to consider it as a metamodel and talk of best-worst methods instead of best-worst method. By the same token, we propose a new framework that can be applied to the family of best-worst methods and is intended to address the challenges posed by hierarchical criteria structures [15].

4. PIvot-Linked Elicitation (PILE) and its application on best-worst methods

From a methodological perspective, the main limitation of conventional hierarchical weight elicitation is that preference information is usually elicited locally and only synthesized globally afterward. Within each group, the DMs provide judgments that produce local preference information, often normalized only with respect to the criteria in that group. A separate elicitation step is then used to assess the relative importance of the groups, and the final global weights emerge only after these distinct outputs are combined. Although this procedure is operationally convenient, it does not construct a single preference system spanning all lower-level criteria. In particular, criteria from different groups are usually not connected by direct preference statements, so their relative importance is inferred only indirectly through higher-level group assessments. As a result, cross-group trade-offs remain mediated by abstract upper-level judgments, while consistency and sensitivity are handled only locally rather than in an integrated manner.

To enable a more holistic treatment of hierarchical weight elicitation, we propose a new approach, called PIvot-Linked Elicitation (PILE). The core idea is to define *pivots*: for each higher-level criterion $G_g \in \mathcal{G}$ we select a few key lower-level criteria, typically the most and the least important ones.[2] These pivot criteria simultaneously summarize the structure within each group and link different groups through inter-group comparisons.

As illustrated in Fig. 4, PILE proceeds in four stages. Since PILE is a general framework, the exact form of the information elicited and synthesized at each stage depends on the underlying weighting method.

1. **Step 1: Intra-group elicitation.** For each group G_g, the DM provides within-group preference information using the elicitation protocol of the chosen method, such as pairwise comparisons, trade-offs, swing judgments, or card placements. The output of this step is therefore the set of intra-group preference statements for each group. In methods that admit a preliminary within-group solution, this step may also yield provisional local weights, but this is not required by PILE.
2. **Step 2: Pivot selection.** For each group G_g, one or more representative pivot criteria are selected, typically the most important criterion and, when useful for the chosen method, also the least important criterion. The output of this step is thus one or two pivot sets, for example the set of group-wise best criteria and, optionally, the set of group-wise worst criteria.
3. **Step 3: Inter-group elicitation on pivots.** The selected pivots are then compared across groups using the same elicitation protocol as in Step 1. The output of this step is a set of inter-group preference statements on the pivot sets. Importantly, this

[1] Let us note that identifying the best and worst criteria (those with the highest and lowest weights) can be supported by methods such as swing weighting [31, p. 139].

[2] For some elicitation schemes, such as the linear PILE-BWM, using only the most important criteria is already sufficient to obtain a holistic assessment; see the corresponding example in the following subsection.

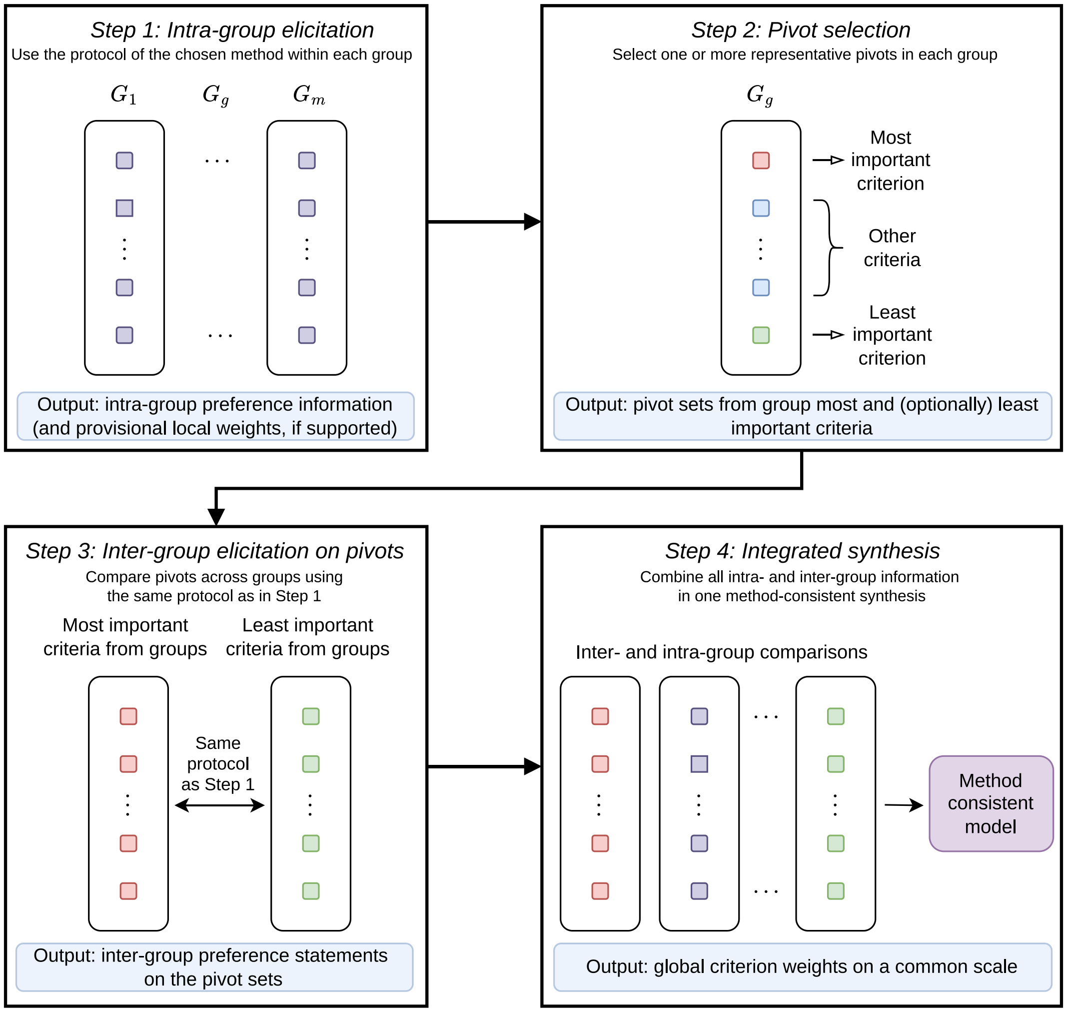


Fig. 4. Generic PILE workflow. Step 1 produces intra-group preference information within each group G_g. Step 2 identifies one or more pivots per group, typically the most important criterion and, when needed, also the least important one. Step 3 produces inter-group preference information on these pivots using the same elicitation protocol as in Step 1. Step 4 integrates all intra- and inter-group information in a single method-consistent model or synthesis procedure to obtain global criterion weights.

step does not necessarily produce a complete ranking; rather, it collects the method-specific information needed to align the groups on a common scale.

4. **Step 4: One-model synthesis.** All intra- and inter-group information is then integrated in a single model or synthesis procedure consistent with the chosen method. The output of this step is a set of global criterion weights on a common scale. When useful, group weights and within-group local weights can subsequently be recovered from these global weights.

In this way, PILE preserves the hierarchical organization of the criteria, but links groups through pivots and performs weight elicitation in one integrated step rather than through multiple isolated hierarchical stages.

PILE is particularly natural for BWM-type methods, which already identify best and worst criteria. In what follows, we instantiate PILE with the BWM methods to illustrate how the framework applies to specific elicitation schemes.

4.1. PILE -BWM methods

In the case of BWM, this general issue takes a specific form. The standard hierarchical implementation preserves the best–worst logic only within each separate block: the DM identifies a best and a worst criterion inside each group, solves the corresponding local model, and then repeats the procedure at the higher level for the groups themselves. Consequently, the best and worst criteria identified in one group are methodologically unrelated to those identified in another group, because they are calibrated in different models. The synthesis of global weights is therefore obtained by multiplying local and group-level results rather than by estimating all lower-level criteria jointly from one coherent system of BO and OW comparisons. This means that the relative position of a criterion in one group with respect to a criterion in another group is not directly informed by best–worst judgments at the criterion level, but only reconstructed afterward through the group weights. This is precisely the gap that PILE–BWM is designed to fill.

The core idea of PILE–BWM is therefore to introduce additional comparisons that directly link the groups, so that all lower-level criteria can be calibrated within one integrated optimization framework. Instead of relying only on separate local BWM models and a top-level BWM on higher-level criteria, PILE–BWM asks the DM to compare representative criteria across groups. More specifically, the best and worst criteria identified within each group are used as pivots that connect the groups at the criterion level. Intuitively, comparisons among the best criteria of different groups help align the upper end of the importance scale across the hierarchy, while comparisons among the worst criteria help align the lower end. In this way, PILE–BWM preserves the cognitive advantages of hierarchical structuring, but avoids

$$
E\,(\text{Non-linear PILE-BWM}) \left\{
\begin{array}{lll}
\text{minimize } \xi & & \\
\text{s.t.} & & \\
\left.\begin{array}{l}
\left| a_{B_g, c_{g,i}} - \dfrac{w_{B_g}}{w_{g,i}} \right| \le \xi, \\
\left| a_{C_{g,i}, W_g} - \dfrac{w_{g,i}}{w_{W_g}} \right| \le \xi,
\end{array}\right\} \forall g \in M \ \forall i \in \{1, \dots, |G_g|\}, & \text{(Intra-group)} \\
\left.\begin{array}{l}
\left| a_{\bar{B}, B_g} - \dfrac{w_{\bar{B}}}{w_{B_g}} \right| \le \xi, \\
\left| a_{B_g, \underline{B}} - \dfrac{w_{B_g}}{w_{\underline{B}}} \right| \le \xi,
\end{array}\right\} \forall g \in M, & \text{(Inter-group } \mathcal{B}) \\
\left.\begin{array}{l}
\left| a_{\bar{W}, W_g} - \dfrac{w_{\bar{W}}}{w_{W_g}} \right| \le \xi, \\
\left| a_{W_g, \underline{W}} - \dfrac{w_{W_g}}{w_{\underline{W}}} \right| \le \xi,
\end{array}\right\} \forall g \in M, & \text{(Inter-group } \mathcal{W}) \\
\sum_{g \in M} \sum_{i=1}^{|G_g|} w_{g,i} = 1, & & \\
w_{g,i} \ge 0, \ \forall g \in M \ \forall i \in \{1, \dots, |G_g|\}. & &
\end{array}
\right. \tag{4}
$$

Box I.

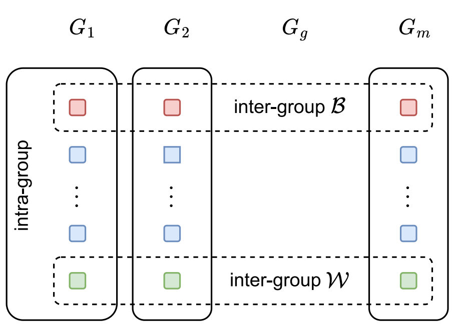


Fig. 5. Visual interpretation of the sets of pairwise comparisons in PILE-BWM.

inferring inter-group relations solely from a separate set of higher-level judgments.

As shown in Fig. 5, in the PILE-BWM approach, DM will provide the following inputs:

- *Intra-group comparison* within the gth group, the DM identifies the best sub-criterion B_g and worst sub-criterion W_g, along with comparisons $a_{B_g, c_{g,i}}$ (best B_g vs. each sub-criterion $c_{g,i}$ in that group) and $a_{c_{g,i}, W_g}$ (each sub-criterion vs. the worst W_g in that group). These are the same type of intra-group comparisons as in the conventional approach.
- *Inter-group comparison* across groups: let $\mathcal{B} = \{B_1, \dots, B_m\}$ and $\mathcal{W} = \{W_1, \dots, W_m\}$. Within $\mathcal{B}$, the DM selects the *best-of-bests* $\bar{B} \in \mathcal{B}$ and the *worst-of-bests* $\underline{B} \in \mathcal{B}$. Similarly, considering $\mathcal{W}$, the DM selects the *best-among-worsts* $\bar{W} \in \mathcal{W}$ and the *worst-among-worsts* $\underline{W} \in \mathcal{W}$. Then elicit the inter-group comparisons on $\mathcal{B}$: $\{a_{\bar{B}, B_g}\}_{g \in M}$ and $\{a_{B_g, \underline{B}}\}_{g \in M}$, and on $\mathcal{W}$: $\{a_{\bar{W}, W_G}\}_{g \in M}$ and $\{a_{W_G, \underline{W}}\}_{g \in M}$.

When the comparisons are obtained from either BWM scales or BWT indifference tradeoffs, we can construct the PILE-BWM model:

Similar to the original formulation of the BWM, (4) is given in Box I is also nonlinear. As done by Rezaei [10] for BWM, we present the linear version of the PILE-BWM:

In the optimization problems (4) and (5), the optimal ξ^* equals the maximum absolute residual across all constraints. Let us focus on the linear problem (5) is given in Box II and consider the effect of the comparison $a_{ij} = 2$ when the weights are $w_i = 0.03$ and $w_j = 0.01$ and when they are, instead, $w_i' = 0.3$ and $w_j' = 0.1$. Either way, the ratio between weights is 3 and in the nonlinear PILE-BWM they would induce the same error, i.e., $\xi \ge |2 - 3|$. In contrast, for the linear PILE-BWM the error induced by the case with greater weights ($\xi \ge |0.3 - 2 \cdot 0.1| = 0.1$) is larger than the one induced by the smaller ones ($\xi \ge |0.03 - 2 \cdot 0.01| = 0.01$). This means that, ceteris paribus, in the linear model, the constraints based on smaller weights (typically those on $\mathcal{W}$) are less likely to be binding in the optimal solution. Consequently, to lessen elicitation burden, we argue (see also Appendix) that the inter-$\mathcal{W}$ block can be treated as optional in the linear PILE–BWM formulation, unless lower-tail calibration is specifically required. Hence, we suggest that, for practical purposes, the model (5) can be replaced by model (6) in Box III.

5. Application of PILE to best–worst methods

This section illustrates the application of PILE to two members of the best–worst family of methods. The purpose is not to provide an empirical validation against a ground-truth set of weights, but to show how the same pivot-linked logic can be instantiated under different elicitation protocols and optimization models. In the first use case, we apply the nonlinear PILE–BWM formulation to a published study on the prioritization of strategies for inclusive and sustainable industrialization. This example illustrates how PILE modifies the conventional hierarchical use of BWM by replacing direct comparisons between broad dimensions with pivot-based criterion-level links. In the second use case, we apply a linear PILE–BWT formulation to a port-selection problem. This example shows how PILE can also be combined with a tradeoff-based elicitation scheme grounded in MAVT, where the comparison inputs explicitly account for attribute ranges. Together, the two cases demonstrate that PILE is not tied to a single method but can be adapted to different weight elicitation approaches while preserving the same four-stage workflow.

$$
E(\text{Linear PILE-BWM}) \quad
\left\{
\begin{aligned}
&\text{minimize } \xi \\
&\text{s.t.} \\
&\left.\begin{aligned}
&\left| w_{B_g} - a_{B_g, c_{g,i}}\, w_{g,i} \right| \le \xi, \\
&\left| w_{g,i} - a_{c_{g,i}, W_g}\, w_{W_g} \right| \le \xi
\end{aligned}\right\} \forall g \in M \;\; \forall i \in \{1, \dots, |G_g|\}, && \text{(Intra-group)} \\
&\left.\begin{aligned}
&\left| w_{\bar{B}} - a_{\bar{B}, B_g}\, w_{B_g} \right| \le \xi, \\
&\left| w_{B_g} - a_{B_g, \underline{B}}\, w_{\underline{B}} \right| \le \xi,
\end{aligned}\right\} \forall g \in M, && \text{(Inter-group } \mathcal{B}\text{)} \\
&\left.\begin{aligned}
&\left| w_{\bar{W}} - a_{\bar{W}, W_g}\, w_{W_g} \right| \le \xi, \\
&\left| w_{W_g} - a_{W_g, \underline{W}}\, w_{\underline{W}} \right| \le \xi,
\end{aligned}\right\} \forall g \in M, && \text{(Inter-group } \mathcal{W}\text{)} \\
&\sum_{g \in M} \sum_{i=1}^{|G_g|} w_{g,i} = 1, \\
&w_{g,i} \ge 0, \;\; \forall g \in M \;\; \forall i \in \{1, \dots, |G_g|\}.
\end{aligned}
\right.
\tag{5}
$$

Box II.

$$
E(\text{Compact linear PILE-BWM}) \quad
\left\{
\begin{aligned}
&\text{minimize } \xi \\
&\text{s.t.} \\
&\left.\begin{aligned}
&\left| w_{B_g} - a_{B_g, c_{g,i}}\, w_{g,i} \right| \le \xi, \\
&\left| w_{g,i} - a_{c_{g,i}, W_g}\, w_{W_g} \right| \le \xi
\end{aligned}\right\} \forall g \in M \;\; \forall i \in \{1, \dots, |G_g|\}, && \text{(Intra-group)} \\
&\left.\begin{aligned}
&\left| w_{\bar{B}} - a_{\bar{B}, B_g}\, w_{B_g} \right| \le \xi, \\
&\left| w_{B_g} - a_{B_g, \underline{B}}\, w_{\underline{B}} \right| \le \xi,
\end{aligned}\right\} \forall g \in M, && \text{(Inter-group } \mathcal{B}\text{)} \\
&\sum_{g \in M} \sum_{i=1}^{|G_g|} w_{g,i} = 1, \\
&w_{g,i} \ge 0, \;\; \forall g \in M \;\; \forall i \in \{1, \dots, |G_g|\}.
\end{aligned}
\right.
\tag{6}
$$

Box III.

5.1. Use case 1: Prioritizing strategies for inclusive and sustainable industrialization

To illustrate how the PILE framework compares with the conventional hierarchical approach, we revisit an existing application that already relies on a hierarchical criteria structure. As a use case, we adopt the study by Sharma and Gupta [25], which evaluates key criteria (strategies) through which Cognitive Digital Twins (CDT) can support inclusive and sustainable industrialization in the context of Industry 5.0. The authors structure twelve CDT criteria into four main dimensions: technological, organizational, environmental, and human. We retain this hierarchy, summarized in Table 2.

For illustration, we use the BWM inputs of Expert 1 (see Table 3). At the main level, G_2 (OD) is judged best and G_3 (ED) worst among the four dimensions. Within each dimension, the expert selects a local best and worst criterion ($c_{1,1}/c_{1,3}$ in G_1, $c_{2,2}/c_{2,1}$ in G_2, $c_{3,1}/c_{3,2}$ in G_3, and $c_{4,1}/c_{4,2}$ in G_4) and provides the corresponding comparison vectors.

The resulting weights, calculated from the input of Expert 1's using BWM, are reported in Table 4. More specifically, the weights of the criteria are normalized to sum to one and are obtained by multiplying each group weight by the corresponding local weight, as in Eq. (1). Note that all results are obtained with the non-linear BWM, which preserves the direct interpretation of the declared ratios in terms of elicited weight ratios and therefore offers a more intuitive basis for comparison afterward.

5.1.1. Applying PILE-BWM

We now re-analyze Expert 1's judgments using the nonlinear PILE–BWM formulation. As in the hierarchical BWM case, we retain the same twelve CDT criteria and the same four groups, $G_1, \dots, G_4$.

In terms of the generic PILE workflow as shown in Fig. 4, Step 1 consists of the intra-group elicitation. Here, the intra-group inputs are exactly those reported in Table 3. Within each group $G_g = \{c_{g,1}, c_{g,2}, c_{g,3}\}$, the expert has already identified a local best criterion B_g and a local worst criterion W_g, and has provided the corresponding BO and OW vectors on the 1–9 BWM scale.

Step 2 consists of pivot selection. From the local best and worst criteria identified in Step 1, we construct the two pivot sets

$$\mathcal{B} = \{B_1, B_2, B_3, B_4\} = \{c_{1,1}, c_{2,2}, c_{3,1}, c_{4,1}\},$$
$$\mathcal{W} = \{W_1, W_2, W_3, W_4\} = \{c_{1,3}, c_{2,1}, c_{3,2}, c_{4,2}\}.$$

Thus, $\mathcal{B}$ collects the group-wise best criteria, whereas $\mathcal{W}$ collects the group-wise worst criteria.

Step 3 consists of inter-group elicitation on the pivot sets. To complete the nonlinear PILE–BWM model, Expert 1 is asked to provide additional BWM judgments on $\mathcal{B}$ and $\mathcal{W}$. On $\mathcal{B}$, the expert identifies the best-of-bests and the worst-of-bests, and then provides the corresponding BO and OW judgments among the four elements B_g. Similarly, on $\mathcal{W}$, the expert identifies the best-among-worsts and the worst-among-worsts, and then provides the corresponding BO and OW judgments among the four elements W_g. In our use case, these additional inter-group inputs are reported in Table 5. They are expressed on the same

Table 2
Dimensions and CDT criteria used in the use case.
Source: Adapted from Sharma and Gupta [25].

Higher-level criterion	Criterion
G_1: Technological Dimension (TD)	$c_{1,1}$: Real-time Optimization (TD1) $c_{1,2}$: Data Analytics (TD2) $c_{1,3}$: IoT and Connectivity (TD3)
G_2: Organizational Dimension (OD)	$c_{2,1}$: Change Management (OD1) $c_{2,2}$: Strategic Alignment (OD2) $c_{2,3}$: Governance Structures (OD3)
G_3: Environmental Dimension (ED)	$c_{3,1}$: Emission Reduction (ED1) $c_{3,2}$: Resource Conservation (ED2) $c_{3,3}$: Environmental Compliance (ED3)
G_4: Human Dimension (HD)	$c_{4,1}$: Job Security and Skill Augmentation (HD1) $c_{4,2}$: Workplace Safety Monitoring (HD2) $c_{4,3}$: Employee Autonomy (HD3)

Table 3
BWM inputs of Expert 1 for main dimensions and dimensional criteria.
Source: Adapted from Sharma and Gupta [25].

Level	Criteria set	Best	Worst	BO vector $a_{B,\cdot}$	OW vector $a_{\cdot,W}$
Main dimensions	$\{G_1, G_2, G_3, G_4\}$	G_2	G_3	$a_{G_2,\cdot} = (3,1,9,5)$	$a_{\cdot,G_3} = (7,9,1,3)$
Technological (TD)	$\{c_{1,1}, c_{1,2}, c_{1,3}\}$	$c_{1,1}$	$c_{1,3}$	$a_{c_{1,1},\cdot} = (1,4,9)$	$a_{\cdot,c_{1,3}} = (9,4,1)$
Organizational (OD)	$\{c_{2,1}, c_{2,2}, c_{2,3}\}$	$c_{2,2}$	$c_{2,1}$	$a_{c_{2,2},\cdot} = (9,1,6)$	$a_{\cdot,c_{2,1}} = (1,9,3)$
Environmental (ED)	$\{c_{3,1}, c_{3,2}, c_{3,3}\}$	$c_{3,1}$	$c_{3,2}$	$a_{c_{3,1},\cdot} = (1,9,7)$	$a_{\cdot,c_{3,2}} = (9,1,2)$
Human (HD)	$\{c_{4,1}, c_{4,2}, c_{4,3}\}$	$c_{4,1}$	$c_{4,2}$	$a_{c_{4,1},\cdot} = (1,9,7)$	$a_{\cdot,c_{4,2}} = (9,1,2)$

Table 4
Hierarchical BWM weights for Expert 1.

Higher-level criterion	Weight	Criterion	Local weight	Global weight
G_1	0.316	$c_{1,1}$	0.692	0.169
		$c_{1,2}$	0.231	0.047
		$c_{1,3}$	0.077	0.017
G_2	0.508	$c_{2,1}$	0.084	0.044
		$c_{2,2}$	0.759	0.449
		$c_{2,3}$	0.156	0.083
G_3	0.056	$c_{3,1}$	0.789	0.040
		$c_{3,2}$	0.088	0.004
		$c_{3,3}$	0.123	0.006
G_4	0.119	$c_{4,1}$	0.789	0.111
		$c_{4,2}$	0.088	0.012
		$c_{4,3}$	0.123	0.017

1–9 BWM scale and are consistent with the ordering implied by the hierarchical global weights.

Step 4 is the one-model synthesis. In this use case, it is performed by solving Eq. (4) using all intra-group inputs from Table 3 together with all inter-group inputs from Table 5. The solution of this single optimization model yields the global weights $w_{g,i}$ of all criteria directly.

The global criterion weights $w_{g,i}$ obtained from PILE–BWM can also be interpreted hierarchically. In particular, the group weights w_g and the within-group local weights $\tilde{w}_{g,i}$ are recovered as

$$w_g = \sum_{i=1}^{|G_g|} w_{g,i}, \quad \forall g \in M, \qquad \tilde{w}_{g,i} = \frac{w_{g,i}}{w_g}, \quad \forall g \in M, \ \forall i \in \{1, \dots, |G_g|\}. \tag{7}$$

Table 6 reports the resulting group, local, and global weights, rounded to three decimal places. The global criterion weights obtained directly from PILE–BWM range approximately from 0.005 to 0.416. In a conventional non-hierarchical BWM, it would be cognitively demanding for the DM to express such large relative differences explicitly within a single elicitation exercise. By contrast, PILE–BWM infers these differences endogenously from intra-group and pivot comparisons, without requiring the DM to specify extreme ratios between broad dimensions.

5.1.2. Consistency analysis

In pairwise comparison methods, *consistency* reflects the highest level of rationality and can be defined in several equivalent ways [33]. It may be seen as a stricter extension of transitivity to valued preferences or as the existence of a priority vector that perfectly matches the expert's estimations of weight ratios. In the case of the BWM, this corresponds to achieving $\xi^* = 0$. In this study, we interpret ξ^* as an inconsistency index, with higher values denoting greater inconsistency, but we do not establish specific thresholds or cutoff rules. This choice is motivated by two factors: first, the complexity and variable structure of the problem, which depends on the criteria hierarchy, make threshold determination difficult; second, the intrinsic challenge of applying precise and indisputable thresholds to an inherently imprecise and continuous phenomenon. Instead, as suggested by other authors [34,35], we adopt a graphical approach.

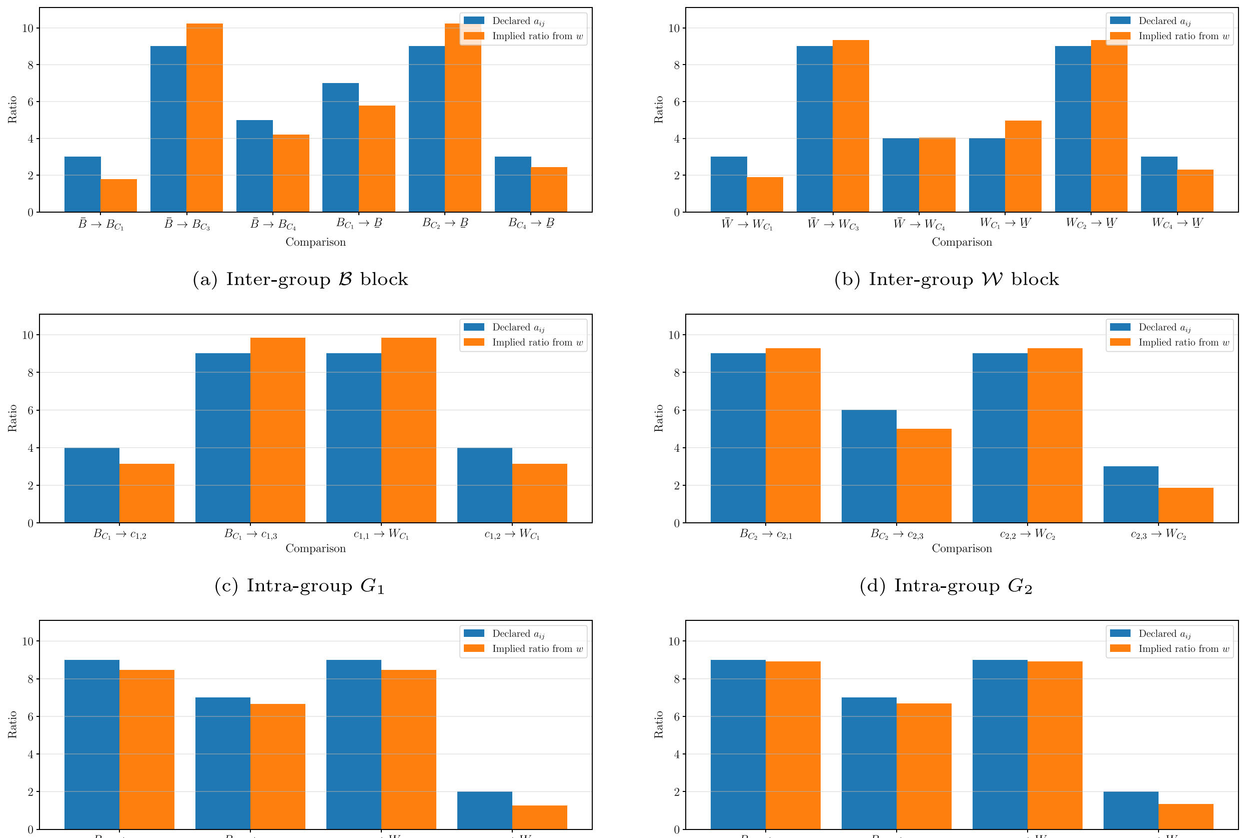
Fig. 6 shows the declared BWM comparisons next to the ratios from the elicited weights. The largest difference between these two values offers a geometric interpretation of ξ^*. To supplement this figure with more information, Fig. 7 reports these differences for every BO/OW comparison in the model: consistency corresponds to the horizontal reference line at 0. Comparisons whose lollipops lie far from this line are those that contribute most to the overall inconsistency and thus deserve closer scrutiny. These figures point out the most inconsistent pairs and provide insight for improving the consistency. Based on the figures, within-group judgments for G_3 and G_4 still display somewhat larger deviations than those for G_1 and G_2, but the single most inconsistent comparison arises from the inter-group $\mathcal{B}$ comparisons across groups, rather than from an intra-group assessment.

5.1.3. Interval analysis of the feasible weight space

Similar to that happens with the nonlinear BWM, the optimal solution of PILE-BWM may not be unique. Proceeding in parallel to what is commonly done for BWM [10] we quantify the stability of results by performing an interval analysis on the criteria weights $w_{g,i}$. As before, we let ξ^* denote the optimal worst-case deviation of the nonlinear PILE-BWM model. For each weight $w_{g,i}$, we solve two auxiliary

Table 5
Inter-group PILE–BWM inputs of Expert 1 on best and worst pivot sets. BO denotes the best-to-others vector, and OW denotes the others-to-worst vector.

Pivot set	Elements	Best	Worst	BO vector $a_{B,\cdot}$	OW vector $a_{\cdot,W}$
$\mathcal{B}$	$(c_{1,1}, c_{2,2}, c_{3,1}, c_{4,1})$	$\bar{B} = c_{2,2}$	$\underline{B} = c_{3,1}$	$a_{\bar{B},\cdot} = (3, 1, 9, 5)$	$a_{\cdot,\underline{B}} = (7, 9, 1, 3)$
$\mathcal{W}$	$(c_{1,3}, c_{2,1}, c_{3,2}, c_{4,2})$	$\bar{W} = c_{2,1}$	$\underline{W} = c_{3,2}$	$a_{\bar{W},\cdot} = (3, 1, 9, 4)$	$a_{\cdot,\underline{W}} = (4, 9, 1, 3)$



(a) Inter-group $\mathcal{B}$ block (b) Inter-group $\mathcal{W}$ block (c) Intra-group G_1 (d) Intra-group G_2 (e) Intra-group G_3 (f) Intra-group G_4

Fig. 6. Declared BWM ratios versus implied ratios from the elicited weights for inter-group blocks and intra-group blocks.

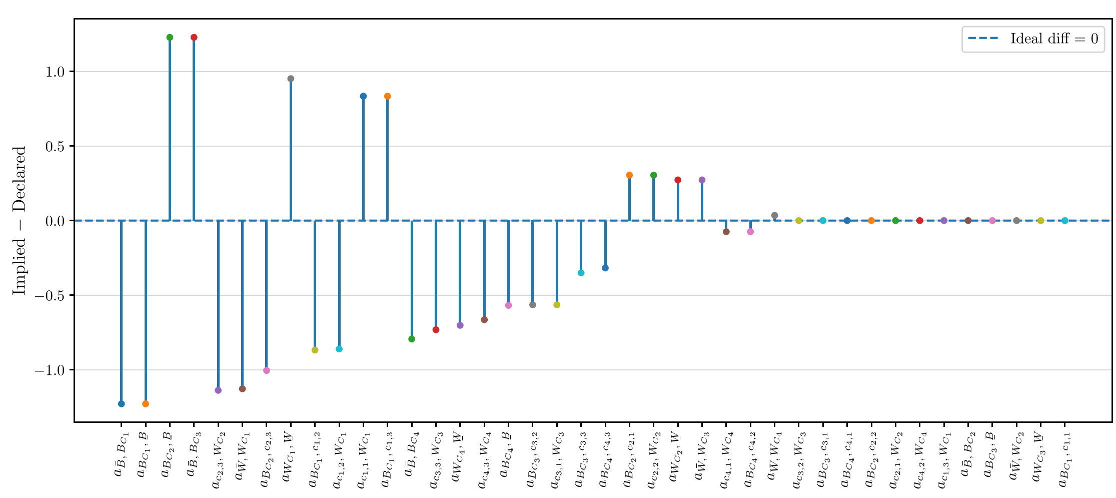


Fig. 7. Lollipop chart of (implied) – (declared) for all BO/OW comparisons, sorted by absolute deviation (ideal value = 0).

Table 6
Group, local, and global weights for the CDT criteria (PILE–BWM). The sub-criteria global weights are elicited directly, while local and criterion weights are obtained indirectly.

Higher-level criterion	Group weight	Criterion	Local weight	Global weight
G_1	0.317	$c_{1,1}$	0.704	0.223
		$c_{1,2}$	0.225	0.071
		$c_{1,3}$	0.072	0.023
G_2	0.544	$c_{2,1}$	0.082	0.045
		$c_{2,2}$	0.765	0.416
		$c_{2,3}$	0.153	0.083
G_3	0.049	$c_{3,1}$	0.788	0.039
		$c_{3,2}$	0.093	0.005
		$c_{3,3}$	0.119	0.006
G_4	0.118	$c_{4,1}$	0.793	0.094
		$c_{4,2}$	0.089	0.011
		$c_{4,3}$	0.119	0.014

Table 7
Criteria weight intervals obtained from nonlinear PILE-BWM.

Higher-level criterion	Criterion	$w_{\min}$	w_{center}	$w_{\max}$	Uncertainty
G_1	$c_{1,1}$	0.214	0.226	0.237	0.023
	$c_{1,2}$	0.059	0.071	0.084	0.024
	$c_{1,3}$	0.021	0.023	0.027	0.006
G_2	$c_{2,1}$	0.038	0.042	0.048	0.011
	$c_{2,2}$	0.380	0.400	0.421	0.041
	$c_{2,3}$	0.068	0.078	0.087	0.019
G_3	$c_{3,1}$	0.037	0.039	0.041	0.004
	$c_{3,2}$	0.004	0.005	0.005	0.001
	$c_{3,3}$	0.005	0.006	0.007	0.003
G_4	$c_{4,1}$	0.069	0.087	0.107	0.038
	$c_{4,2}$	0.008	0.011	0.014	0.006
	$c_{4,3}$	0.008	0.013	0.018	0.010

problems: one minimizing $w_{g,i}$ and one maximizing $w_{g,i}$, subject to the original constraints and the additional condition that the deviation does not exceed ξ^*. Repeating this procedure for all criteria produces $2n$ feasible weight vectors (in our example $n = 12$). For each criterion $c_{g,i}$, we then define

$$w^{\min}_{g,i} := \min_{\ell=1,\dots,2n} w^{(\ell)}_{g,i}, \qquad w^{\max}_{g,i} := \max_{\ell=1,\dots,2n} w^{(\ell)}_{g,i},$$

and we take the *center* weight as the empirical average over all $2n$ values:

$$w^{\text{center}}_{g,i} := \frac{1}{2n}\sum_{\ell=1}^{2n} w^{(\ell)}_{g,i}$$

Finally, we quantify the uncertainty of a weight by means of the length of the interval $[w^{\min}_{g,i}, w^{\max}_{g,i}]$ (see Table 7). Fig. 8 visualizes the resulting intervals for all 12 criteria as a dumbbell plot (min-center-max for each $w_{g,i}$).

The intervals represented in Fig. 8 show that the ranking is quite robust at the top: $c_{2,2}$ is consistently the most important, with a center around 0.40, and $c_{1,1}$ remains clearly in the second tier. Several lower-weight criteria (e.g. $c_{3,2}$, $c_{3,3}$, $c_{4,2}$, $c_{4,3}$) exhibit relatively larger relative ranges, indicating that their exact positions within the bottom tier are less stable. However, their intervals remain well separated from those of the leading criteria, so the main conclusions regarding which CDT criteria are most influential are not sensitive to the particular optimal solution selected within the feasible weight space.

5.2. Use case 2: Port evaluation problem

To illustrate that PILE can also be applied to a tradeoff-based member of the best–worst family, we revisit the port-selection case study

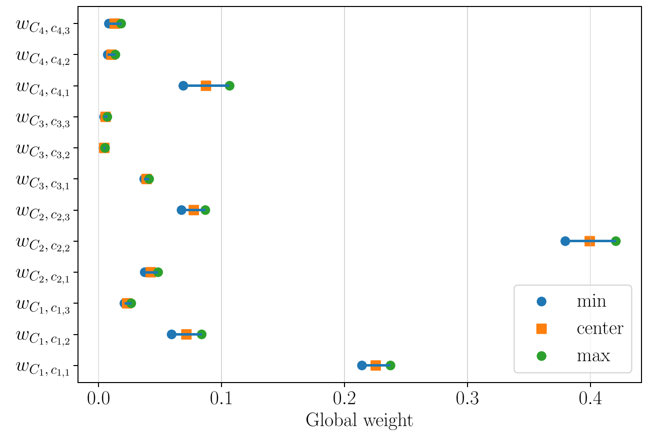


Fig. 8. Weight intervals for all $w_{g,i}$, with markers at $w_{\min}, w_{\max}, w_{\text{center}}$.

used by Liang et al. [32] to demonstrate the Best-Worst Tradeoff (BWT) method. The original case evaluates seven ports using six criteria: (C_1) Terminal handling charges, (C_2) International Ship and Port Facility Security Code charges, (C_3) Customs service, (C_4) Port reputation, (C_5) Satisfaction with terminal operations, and (C_6) Number of container terminals. Terminal handling charges and International Ship and Port Facility Security Code charges are cost criteria, while the remaining four are benefit criteria.

In the original study, the criteria were treated as a flat set, and the expert identified C_1 and C_3 as the best and worst criteria, respectively. After the consistency-improvement process, the reported BWT tradeoff values are

$$r^{BO} = (1, 0.20, 0.11, 0.37, 0.28, 0.32), \tag{8}$$

and

$$r^{OW} = (0.11, 0.41, 1, 0.13, 0.18, 0.16). \tag{9}$$

Here, the entries of r^{BO} are proportional to w_j/w_B, while the entries of r^{OW} are proportional to w_W/w_j. Hence, the comparison coefficients used in the linear BWT model are obtained by taking their reciprocals. The resulting flat BWT weights reported in the original study are

$$w = (0.39, 0.10, 0.03, 0.18, 0.14, 0.16). \tag{10}$$

For the present use case, we reorganize the six criteria into two groups, as reported in Table 8. The first group contains the two monetary cost criteria and is therefore labeled *Economic burden*. The second group contains service, operational, reputational, and capacity-related criteria and is therefore labeled *Operational and technical performance*.

5.2.1. Applying PILE –BWT

We now apply the PILE workflow to the hierarchical structure in Table 8. Since the original study did not elicit preferences under a hierarchical PILE structure, the following inputs should be interpreted as an illustrative hierarchical adaptation. More specifically, we assume that another expert, following the same preference logic as the expert in the original case, provides tradeoff judgments whose implied comparison coefficients are consistent with the ratios encoded in the revised BWT tradeoff values reported above.

Let r_i^{BO} denote the best-to-others tradeoff value associated with criterion C_i, and let r_i^{OW} denote the others-to-worst tradeoff value associated with criterion C_i. When a comparison is needed within a subgroup whose local best criterion is B_g and local worst criterion is W_g, the corresponding linear BWT coefficients are obtained as

$$a^{BO}_{B_g,i} = \frac{r^{BO}_{B_g}}{r^{BO}_i}, \qquad a^{OW}_{i,W_g} = \frac{r^{OW}_{W_g}}{r^{OW}_i}. \tag{11}$$

Table 8
Hierarchical grouping of the port-selection criteria.

Higher-level criterion	Lower-level criteria
G_1: Economic burden	$c_{1,1} = C_1$: Terminal handling charges $c_{1,2} = C_2$: International Ship and Port Facility Security Code (ISPS)
G_2: Operational and technical performance	$c_{2,1} = C_3$: Customs service $c_{2,2} = C_4$: Port reputation $c_{2,3} = C_5$: Satisfaction with terminal operations $c_{2,4} = C_6$: Number of container terminals

Table 9
PILE–BWT inputs for the port-selection case. BO denotes the best-to-others comparison vector, and OW denotes the others-to-worst comparison vector.

Block	Elements	Best	Worst	BO vector	OW vector
G_1	(C_1, C_2)	C_1	C_2	$(1, 5.000)$	$(3.727, 1)$
G_2	(C_3, C_4, C_5, C_6)	C_4	C_3	$(3.364, 1, 1.321, 1.156)$	$(1, 7.692, 5.556, 6.250)$
$\mathcal{B}$	(C_1, C_4)	C_1	C_4	$(1, 2.703)$	$(1.182, 1)$

Table 10
Group, local, and global weights for the port-selection criteria obtained with compact linear PILE–BWT. The global criteria weights are obtained directly, while group and local weights are recovered afterward.

Higher-level criterion	Group weight	Criterion	Local weight	Global weight
G_1	0.487	C_1	0.818	0.398
		C_2	0.182	0.089
G_2	0.513	C_3	0.056	0.029
		C_4	0.309	0.159
		C_5	0.299	0.154
		C_6	0.335	0.172

Thus, for example, because the original best-to-others tradeoff value for C_2 is 0.20 while the value for the best criterion C_1 is 1, the corresponding comparison coefficient is

$$a^{BO}_{C_1,C_2} = \frac{1}{0.20} = 5.$$

Similarly, if the local best criterion in a group is C_4, then the comparison coefficient between C_4 and C_5 is

$$a^{BO}_{C_4,C_5} = \frac{0.37}{0.28} = 1.321.$$

In terms of the generic PILE workflow, Step 1 consists of intra-group elicitation. In the economic group G_1, the local best criterion is C_1 and the local worst criterion is C_2. In the operational and technical performance group G_2, the local best criterion is C_4 and the local worst criterion is C_3. Step 2 then yields the pivot sets

$$\mathcal{B} = \{C_1, C_4\}, \qquad \mathcal{W} = \{C_2, C_3\}. \tag{12}$$

Since we use the compact linear PILE–BWT formulation, Step 3 only requires inter-group comparisons among the best pivots in $\mathcal{B}$. Hence, no inter-group comparison among the worst pivots in $\mathcal{W}$ is elicited.

Within $\mathcal{B}$, C_1 is the best-of-bests and C_4 is the worst-of-bests. The resulting intra-group and inter-group inputs are reported in Table 9. The values in the table are the comparison coefficients used in the linear BWT model, not the raw tradeoff values; therefore, values greater than one are expected.

Step 4 is the one-model synthesis. Here, it is performed by solving the compact linear PILE–BWT model, which has the same algebraic structure as model (6) is given in Box III, using the two intra-group blocks and the inter-$\mathcal{B}$ block in Table 9. In line with the linear BWT interpretation, the intra-group constraints are written as

$$\left| w_{B_g} - a^{BO}_{B_g,i} w_{g,i} \right| \le \xi, \qquad \left| w_{g,i} - a^{OW}_{i,W_g} w_{W_g} \right| \le \xi, \tag{13}$$

for each criterion $c_{g,i}$ in group G_g. The inter-$\mathcal{B}$ constraints are written analogously for the best pivots:

$$\left| w_{\bar{B}} - a^{BO,\mathcal{B}}_{\bar{B},B_g} w_{B_g} \right| \le \xi, \qquad \left| w_{B_g} - a^{OW,\mathcal{B}}_{B_g,\underline{B}} w_{\underline{B}} \right| \le \xi. \tag{14}$$

Together with the normalization condition

$$\sum_{g \in M} \sum_{i=1}^{|G_g|} w_{g,i} = 1, \tag{15}$$

and non-negativity constraints, the model yields the global criterion weights directly.

The resulting weights are reported in Table 10. As in the previous use case, the global criterion weights are obtained directly from the compact linear PILE–BWT model, while the group weights and local weights are recovered from the global weights. The results show that the two groups receive comparable weights, with a slightly larger weight assigned to G_2.

To inspect the consistency of the resulting PILE–BWT solution, Fig. 9 compares the declared comparison values with the ratios implied by the elicited weights for each comparison block. The first panel reports the inter-group comparison among the best pivots, while the second and third panels report the intra-group comparisons within G_1 and G_2, respectively. Fig. 10 complements this block-wise view by reporting the difference between the implied and declared values for all comparisons. Values closer to zero indicate better agreement between the elicited judgments and the weight vector obtained from the optimization model.

Compared with the original flat BWT application, the present PILE–BWT use case changes how the hierarchy is handled. In the original study, all six criteria were treated in a single flat elicitation problem. By contrast, PILE–BWT allows the same type of tradeoff-based elicitation to be used with a hierarchical criteria structure. The criteria are first organized into two meaningful groups, namely *Economic burden* and *Operational and technical performance*, while the final global weights are still obtained from one integrated model.

This use case highlights a specific contribution of PILE for tradeoff-based methods. In a conventional hierarchical weight-elicitation approach, the analyst would normally need to elicit weights for the higher-level criteria and then combine them with local weights. For BWT, however, this is not directly possible because the method relies on explicit value functions and tradeoff questions defined over criterion ranges. The higher-level criteria, such as *Economic burden* and *Operational and technical performance*, do not have explicit value functions or directly measurable attribute ranges in the original problem. Therefore, a conventional hierarchical BWT procedure cannot easily ask the expert to trade off these two higher-level criteria in the same way as it asks tradeoff questions for concrete lower-level criteria.

PILE makes this hierarchical extension possible by avoiding direct tradeoffs between higher-level criteria. Instead, the connection between groups is established through pivot criteria selected from the lower level. In the present case, the inter-group comparison is performed between the best pivots of the two groups, C_1 and C_4. These are concrete criteria with explicit value functions and observable ranges, so the tradeoff logic of BWT remains applicable. The pivot comparison then supplies the missing inter-group link and allows the model to place all lower-level criteria on a common scale.

Thus, the main role of PILE in this use case is not merely to reorganize the original flat BWT problem. Rather, it provides a way

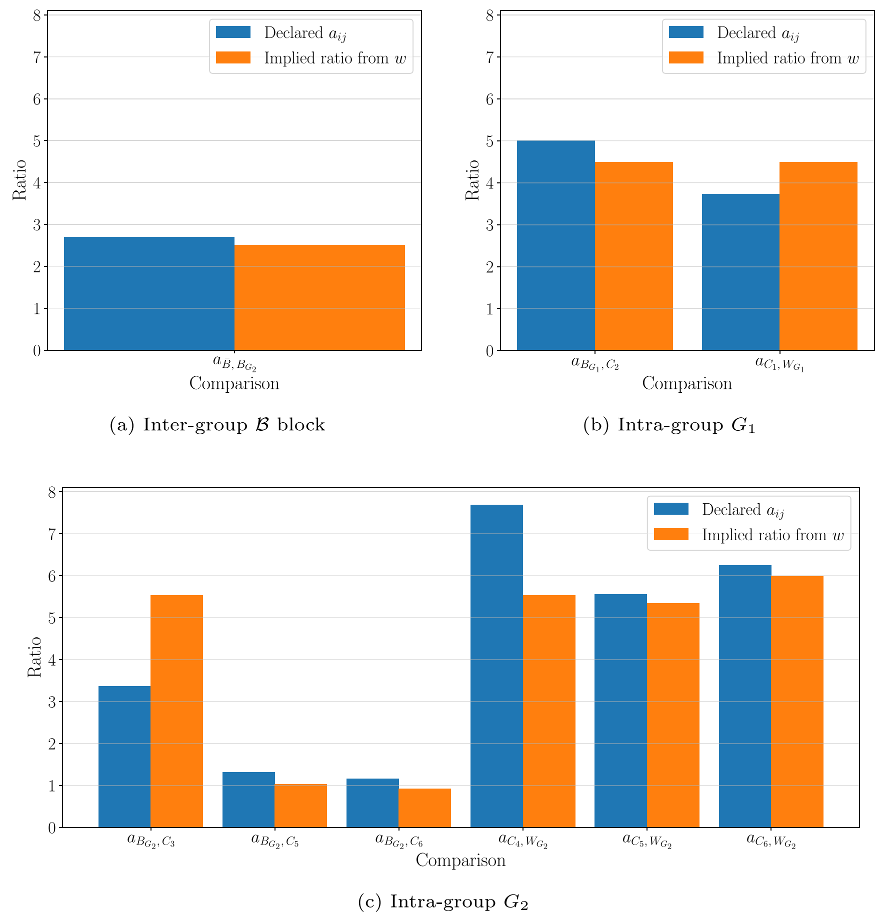


Fig. 9. Declared BWT comparison values versus implied ratios from the elicited weights for the inter-group and intra-group blocks.

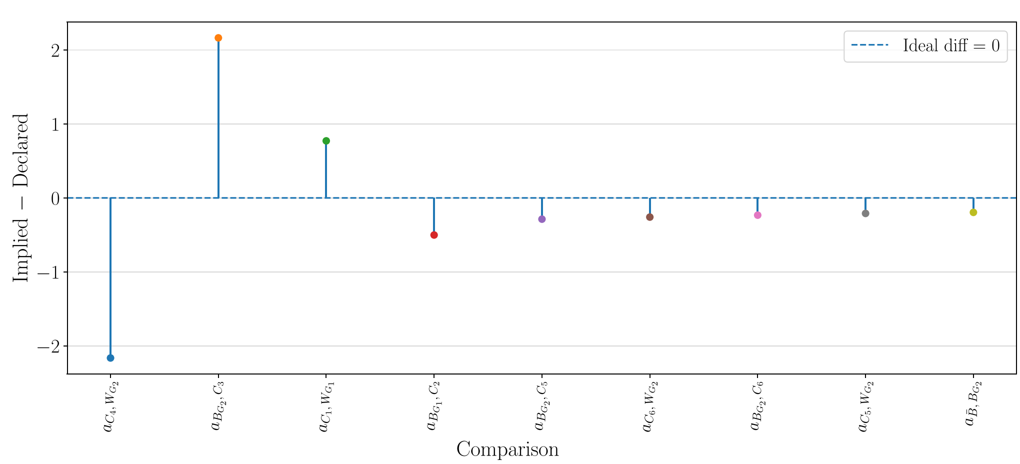


Fig. 10. Lollipop chart of (implied) – (declared) for all PILE–BWT comparisons, sorted by absolute deviation. The ideal value is 0.

to apply a tradeoff-based method to a hierarchical criteria structure when conventional hierarchical elicitation would be infeasible because no explicit value functions are available for the higher-level criteria. In this sense, PILE turns an otherwise unavailable hierarchical BWT formulation into an operational one while preserving the range-sensitive interpretation of the original method.

6. Discussion

With the demonstrations of PILE with BWM and BWT method, we illustrated how PILE mitigates the challenges posed by the conventional hierarchical weight elicitation approach. In Section 2.1, we discussed some of them and here we revisit the same numbered structure to discuss the positive effects of PILE.

1. PILE keeps the elicitation at the level of criteria, which are typically more concrete, operational, and easier to assess than broad objectives. Instead of asking the expert to compare high-level dimensions directly, PILE relies on judgments about lower-level criteria and lets the higher-level structure emerge indirectly through the pivots and their cross-group comparisons. This preserves the hierarchical interpretation while grounding the elicitation in objects that the expert can assess more meaningfully.
2. By avoiding direct trade-offs between politically sensitive objectives (e.g., "economic" vs. "environmental") and focusing on criteria, in some situations PILE may help reduce the risk that the elicitation process forces the DM to make explicit political statements. Such benefit can be boosted when PILE is equipped with BWT: in this case the underlying theory is MAVT, for which ratios between weights are interpreted as rates of substitution and weights cannot be interpreted as levels of importance [31, p. 151].
3. PILE alleviates the oversensitivity of results to small changes in higher-level comparisons. Since PILE does not rely on a separate set of dimension-level comparisons, there is no single upper-level block whose small perturbations can rescale entire groups. When BWT is used within PILE, this effect is further mitigated by grounding preferences in explicit tradeoffs over attribute ranges, which constrains the numerical meaning of the inputs.
4. PILE explicitly links groups within the same elicitation model. In the conventional hierarchical approach, criteria belonging to different groups are related only indirectly through the weights assigned to their parent dimensions. By construction, PILE introduces inter-group constraints through the pivot criteria, so that criteria from different groups are calibrated on a common importance scale without requiring exhaustive all-to-all comparisons.
5. PILE adopts a holistic view. PILE pools all available intra-and inter-group comparisons into a single model, instead of having separate models. This unified formulation also enables analyses that are difficult to conduct under the conventional hierarchical workflow. For instance, in the nonlinear BWM, interval analysis is cumbersome when weights are obtained from separate, level-specific models and then synthesized. In contrast, PILE-BWM produces global weights directly within one nonlinear model, making interval analysis of the feasible weight space straightforward, as illustrated in the case study.
Not surprisingly, the holistic approach may also present more consistent results. To test our approach, we also perform a block-wise inconsistency analysis based on the nonlinear BWM applied to the use case. Specifically, we took the full set of BO/OW inputs used in PILE-BWM (intra-group comparisons for G_1–G_4 and the inter-group best and worst blocks, inter-$\mathcal{B}$ and inter-$\mathcal{W}$) and evaluated two situations:
   (a) the integrated PILE-BWM solution, in which all blocks share a single deviation variable ξ, and
   (b) a collection of *individual* non-linear BWMs, where each block (each G_g, inter-$\mathcal{B}$, inter-$\mathcal{W}$) is solved separately with the same inputs. For the intra-group blocks G_1—G_4 and the inter-$\mathcal{B}$ block, these individual models coincide with the inputs of the conventional hierarchical BWM. Thus, the comparison also provides a direct assessment of the consistency achieved by PILE-BWM versus the standard hierarchical implementation.

For each block, we computed the maximal absolute deviation ξ^* between the declared ratio and the ratio implied by the elicited weights. Because the intra-group and inter-$\mathcal{B}$ inputs are identical in both settings, this provides a direct comparison between the integrated and decomposed uses of the same information. The resulting values are reported in Table 11.
Table 11 shows that, when all blocks are estimated jointly, the worst deviation across the entire system decreases from about 1.39 (for the individual inter-$\mathcal{B}$ model) to about 1.23 (for the integrated PILE-BWM solution).

Table 11
Block-wise maximal deviation ξ^* for nonlinear PILE–BWM vs. individual nonlinear BWMs in the use case.

Block	PILE–BWM ξ^*	Individual BWMs ξ^*
C_1 intra-group	0.8678	1.0000
C_2 intra-group	1.1374	1.1459
C_3 intra-group	0.7314	0.5949
C_4 intra-group	0.6644	0.5949
Inter-$\mathcal{B}$ (bests)	1.2280	1.3944
Inter-$\mathcal{W}$ (worsts)	1.1272	0.4586
Overall maximum	1.2280	1.3944

6. PILE handles the evolution of the criteria set more gracefully. Under PILE, adding a new criterion primarily affects the local group: it must be compared to the existing best and worst in its group, and it may or may not become a pivot in the inter-group sets $\mathcal{B}$ or $\mathcal{W}$. However, the higher-level structure does not need to be rebuilt from scratch, and previously collected comparisons remain valid.

In summary, PILE preserves the cognitive advantages of a hierarchical organization of criteria, but it relocates the elicitation effort to the level of tangible criteria and links groups via pivots within a single optimization model. This design addresses the main conceptual and practical shortcomings of conventional hierarchical weight elicitation, while remaining compatible with established methods such as BWM and BWT.

7. Conclusion

In this paper we revisited hierarchical weight elicitation in MCDA and highlighted several conceptual and practical drawbacks of the conventional scheme, in which local criteria weights are first elicited within each group and then aggregated via a separate set of comparisons on higher-level criteria. We showed that this practice can force DM to formulate judgments on vague, multidimensional objectives, invite politically sensitive comparison at the level of broad dimensions, treats groups in isolation, and fragments preference information across multiple independent models. These features may amplify sensitivity to upper-level judgments and make the overall elicitation process less coherent than intended.

To address these issues, we proposed the PIvot-Linked Elicitation (PILE) framework. PILE preserves the interpretability and convenience of hierarchical criteria organization, but relocates elicitation to the most operational level, namely the lowest criteria level, while linking groups through a small set of pivot criteria (typically the most and least important representatives of each group). All intra-group and inter-group information is then combined in a single integrated model. Instantiating PILE for best–worst methods yielded PILE-BWM, for which we provided both nonlinear and linear formulations. In the use cases, PILE-BWM methods produced interpretable global weights in one step and enabled consistency and interval analyses within the same modeling framework.

Bearing in mind that PILE is an approach to weight elicitation rather than an elicitation method in the narrow sense, its comparative advantages can be illustrated by coupling it with two methods, BWM and BWT, as shown in the two use cases above. Table 12 compares

Table 12
Comparison of selected weight-elicitation methods with respect to weight interpretation, debiasing strategy, treatment of the hierarchy, and consistency analysis.

Method	Ref.	Weight interpretation	Debiasing strategy	Hierarchical structure	Consistency analysis
Tradeoff	[7]	Ratios between weights represent substitution rates between criterion levels	No debiasing strategy is implemented in the basic version [36]	The hierarchical structure of criteria must be flattened and cannot be directly exploited	Consistency checks are suggested [7, p. 123], but inconsistency quantification is not implemented
AHP	[3,37]	Criteria weights express relative importance, typically through Saaty's scale	No explicit debiasing strategy; biases may arise, especially with incomplete pairwise comparison matrices	Criteria can be arranged hierarchically, but higher-level criteria are directly compared with one another	Several inconsistency indices are available in the literature [38]
BWM	[9,15]	As in AHP, weights express the relative importance of criteria; Saaty's scale is often adopted	The consider-the-opposite strategy is used to mitigate anchoring bias [8,39]	As in the AHP, the hierarchy can be used, but objectives or higher-level criteria still need to be compared	Input-based and output-based consistency indices, as well as ordinal consistency checks, are available
BWT	[32]	As in the tradeoff method, ratios between weights represent rates of substitution		The hierarchy cannot be directly exploited unless higher-level criteria are associated with explicit value functions and measurable ranges	
PILE-BWM	This paper	It inherits the semantics of BWM: weights express the relative importance of criteria	As in BWM and BWT, the elicitation design can help mitigate anchoring bias by relying on best–worst comparisons	The hierarchical structure can be exploited without directly comparing objectives or higher-level criteria	The optimal value ξ^* of the associated minimization problem is a global inconsistency index; its structure can be decomposed and analyzed graphically, e.g., Figs. 6 and 7
PILE-BWT	This paper	It inherits the semantics of BWT: ratios between weights represent rates of substitution		The hierarchy can be exploited without asking tradeoffs on objectives	

PILE-BWM and PILE-BWT with some well-known elicitation methods with respect to selected key features.

A key advantage of PILE is that it is independent of the method used. The framework specifies where comparisons should be placed (criteria level and pivot links) and how groups should be connected (via pivots), but it does not prescribe a particular elicitation protocol. This allows for several natural extensions. For value-based elicitation schemes, such as Swing Weighting, PILE can be implemented by performing Swing judgments within each group. Then, a small number of additional Swing Comparisons are conducted across the group pivots to align all groups on a common scale. The same principle applies to card-based procedures such as Simos–Roy–Figueira (SRF). Huang et al. [40] proposed an SRF scheme in which all preference information is integrated within a single model. PILE provides another way of comparison across different criteria groups but integrate information in a single model.

More generally, the contribution of PILE depends on the method to which it is applied. For methods that already allow for some form of hierarchical implementation, PILE provides a new elicitation design that shifts the assessment to the criterion level and introduces pivot-based links across groups within a unified formulation. For methods without a principled way to handle hierarchical criteria structures, PILE plays an enabling role by extending them to such settings while preserving their original elicitation logic. Thus, PILE is not tied to a single weighting procedure but rather contributes a general scheme for adapting diverse elicitation methods to hierarchical criteria structures. We view PILE as a complementary design that yields a unified formulation while avoiding the need for the DM to compare the most and least important criteria from all groups in a single step. In all cases, PILE aims to reduce reliance on abstract, politically loaded, or overly sensitive upper-level assessments while exploiting the structure that motivates hierarchies.

Several directions for future work follow from these results. We conjecture that the application of PILE on BWM (PILE-BWM) inherits the good behavioral properties of BWM as, for instance, its ability to neutralize the anchoring bias. Nevertheless, we could not provide conclusive evidence, and this could be the object of future works. Also, the choice and number of pivots merit further investigation: while best/worst pivots are natural, other schemes may benefit from multiple representatives per group or from adaptive pivot selection driven by inconsistency diagnostics.

CRediT authorship contribution statement

River Huang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Matteo Brunelli:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Conceptualization.

Declaration of competing interest

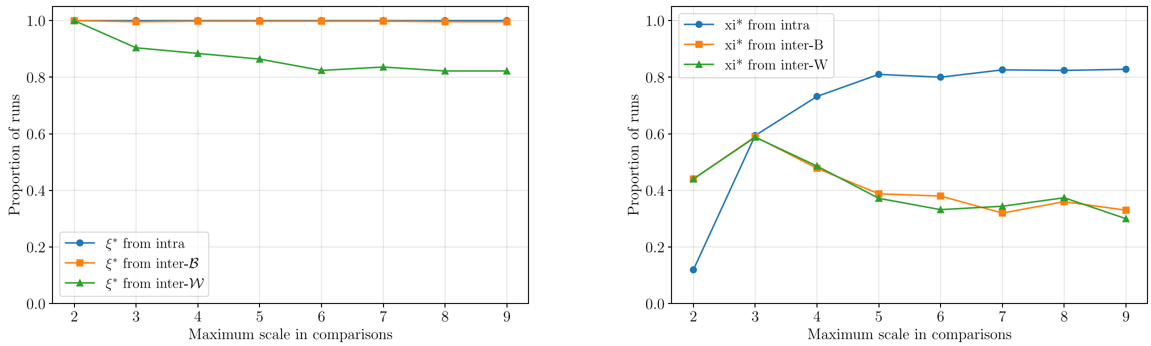
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work of River Huang published in this article, was carried out with the support of the Swiss Federal Office of Energy SFOE as part of the SWEET project SURE. The authors bear sole responsibility for the conclusions and results presented in this publication.

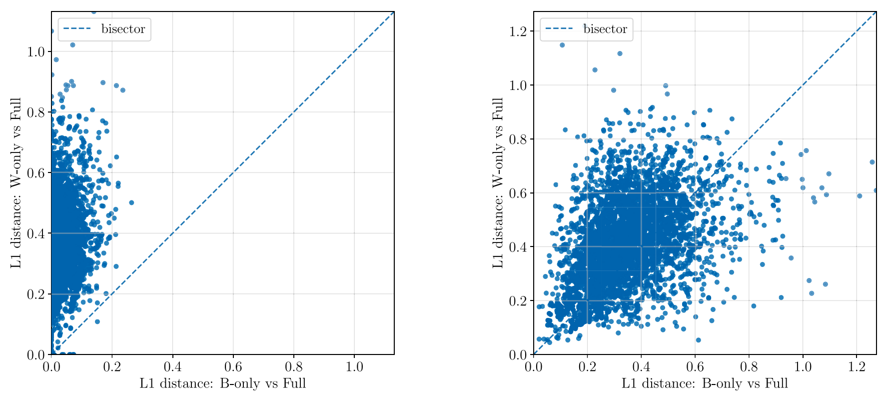
Appendix. Evidence for redundancy in the linear PILE-BWM model: Simulation protocol and results

We examine whether inter-group comparisons on worst representatives (the inter-$\mathcal{W}$ block) are redundant in the linear PILE–BWM formulation, and whether any such redundancy persists in the nonlinear model. Our working hypothesis is that, because of the linear (additive) form in (5), inter-$\mathcal{W}$ constraints have intrinsically low leverage. The intra-group and inter-$\mathcal{B}$ blocks involve the most influential (larger) weights and therefore dominate the absolute residuals that define ξ^*, whereas inter-$\mathcal{W}$ constraints primarily couple smaller weights and seldom become binding in the metric. Empirically, removing the inter-$\mathcal{W}$ block from the linear model yields global weights that are numerically similar to the full specification, while removing inter-$\mathcal{B}$ does not. By contrast, in the nonlinear ratio formulation (4), best and worst enter symmetrically through reciprocal ratios, so inter-$\mathcal{W}$ generally cannot be discarded without altering the solution.



(a) Linear *PILE-BWM*. (b) Non-linear *PILE-BWM*.

Fig. A.11. Block-wise frequency with which intra-group, inter-$\mathcal{B}$ and inter-$\mathcal{W}$ constraints attain the maximal deviation ξ^* in the full model (possibly overlapping), as a function of the maximum scale L.



(a) Linear *PILE-BWM*. (b) Non-linear *PILE-BWM*.

Fig. A.12. Scatter of ℓ_1 distances between the full model and the B-only model (horizontal axis) and between the full model and the W-only model (vertical axis), for the linear and non-linear PILE-BWM formulations. The diagonal line corresponds to equal distances.

A.1. Simulation protocol

To test these hypotheses, a large-scale Monte Carlo experiment was carried out for both the linear and non-linear PILE–BWM/BWT models. For each replication, the following steps were performed.

1. *Random problem size.* The number of groups m was drawn uniformly from $\{2, \dots, 6\}$, and the number of sub-criteria per group was drawn uniformly from $\{3, \dots, 7\}$. The total number of lower level criteria thus varied between 6 and 42.
2. *Random comparison scales.* A maximum scale L was chosen from $\{2, \dots, 9\}$. Within each group, best-to-others $a_{B_g,c_{g,i}}$ and others-to-worst $a_{c_{g,i},W_g}$ intensities were generated as independent integers in $\{1, \dots, L\}$, with $a_{B_g,B_g} = a_{W_g,W_g} = 1$, and at least one entry per instance was forced to take the value L to ensure that the entire range of the scale is used. Similarly, inter-group comparisons $a_{\bar{B},B_g}$, $a_{B_g,\underline{B}}$, $a_{\bar{W},W_g}$ and $a_{W_g,\underline{W}}$ were generated in $\{1, \dots, L\}$.
3. *Model variants.* For each randomly generated instance and each scale L, three models were solved:
   - **Full model:** the complete linear or non-linear PILE formulation (Models (4) and (5)), with intra-group constraints and both inter-$\mathcal{B}$ and inter-$\mathcal{W}$ blocks.
   - **B-only model:** the same formulation but omitting the inter-$\mathcal{W}$ constraints, i.e., retaining intra-group and inter-$\mathcal{B}$ constraints.
   - **W-only model:** the same formulation but omitting the inter-$\mathcal{B}$ constraints, i.e., retaining intra-group and inter-$\mathcal{W}$ constraints.

   For each variant, the solver returned the optimal value ξ^* and the criteria weights $\{w^*_{g,i}\}$.
4. *Distance to the full model.* For each instance, the discrepancy between the reduced models and the full model was quantified by the ℓ_1, ℓ_2 and ℓ_∞ distances

$$d_1(w^{\text{full}}, w^{\text{B}}) = \sum_{g,i} \left| w^{\text{full}}_{g,i} - w^{\text{B}}_{g,i} \right|, \qquad d_1(w^{\text{full}}, w^{\text{W}}) = \sum_{g,i} \left| w^{\text{full}}_{g,i} - w^{\text{W}}_{g,i} \right|.$$

5. *Block-wise decomposition of ξ^*.* For the full model, the residual of each constraint was evaluated at the optimal solution, and the set of constraints achieving the maximal deviation ξ^* was identified. Each of these maximizers was labeled as belonging to one of three blocks: *Intra-group*, *Inter-$\mathcal{B}$*, *Inter-$\mathcal{W}$*.
   For each instance, boolean indicators were recorded for whether ξ^* was attained by at least one constraint in each block, and whether the maximizers were purely intra, purely inter-$\mathcal{B}$, purely inter-$\mathcal{W}$, or involved a mixture of blocks.

This experiment was replicated 1000 times for each scale $L \in \{2, \dots, 9\}$, both for the linear and the non-linear PILE–BWM formulations.

A.2. Summary statistics and figures

For each model family (linear vs. non-linear), the simulation output was summarized along:

- distance between the weights of the reduced models (B-only, W-only) and the full model;

- the frequency with which the maximal deviation ξ^* in the full model is attained by intra-group, inter-$\mathcal{B}$ and inter-$\mathcal{W}$ constraints.

The block-wise decomposition of the maximal deviation ξ^* for the linear and non-linear formulations is summarized in Fig. A.11. In the linear case (Fig. A.11(a)), all three blocks, intra-group, inter-$\mathcal{B}$, and inter-$\mathcal{W}$, exhibit relatively high probabilities of containing at least one constraint that attains ξ^*. This indicates that, for a large fraction of instances, the worst deviation is shared across multiple blocks rather than being dominated by a single type of comparison; in other words, the set of active constraints at ξ^* often mixes intra-group and inter-group relations. Nevertheless, as the maximum scale L increases, the probability that inter-$\mathcal{W}$ constraints are involved in ξ^* decreases systematically, while intra-group and inter-$\mathcal{B}$ constraints remain frequently active. This suggests that, under more extreme scales, the tightest constraints in the linear model tend to be associated with best-related comparisons (intra and inter-$\mathcal{B}$), and the contribution of the worst-related inter-$\mathcal{W}$ block to the worst-case deviation becomes less pronounced. For the non-linear formulation (Fig. A.11(b)), the picture is more balanced. The three blocks still show different probabilities of being involved in the maximal deviation, but these differences are less systematic and, across scales, all blocks retain a substantial share of runs in which they participate in ξ^*. In particular, inter-$\mathcal{B}$ and inter-$\mathcal{W}$ constraints have comparable chances of being active at the optimum, and this symmetry does not decrease as the scale grows. This behavior is consistent with the structure of the non-linear constraints, where best-to-others and others-to-worst ratios enter in a more symmetric way.

The scatter plots in Fig. A.12 compare the ℓ_1 distances between the full model and the reduced B-only and W-only models. In the linear case (Fig. A.12(a)), the vast majority of points lie above the diagonal, and many points accumulate close to the vertical axis. This means that $d_1(w^{\text{full}}, w^{\text{B}})$ is typically much smaller than $d_1(w^{\text{full}}, w^{\text{W}})$, and in a non-negligible fraction of instances the B-only model reproduces the full model exactly (zero ℓ_1 distance). Since these distances are sums of absolute differences, such small values imply that the entire weight vector from the B-only model is extremely close to that of the full model. By contrast, the W-only model can deviate substantially from the full solution, with distances that cover a much wider range along the vertical axis.

The non-linear scatter plot (Fig. A.12(b)) shows a different pattern. Points are distributed more symmetrically around the diagonal, and the cloud extends similarly along both axes. This indicates that the B-only and W-only models have comparable ℓ_1 distances to the full non-linear solution: discarding inter-$\mathcal{B}$ or inter-$\mathcal{W}$ constraints leads to similar levels of approximation error in terms of the global weights. Combined with the block-share analysis, this reinforces the view that, in the non-linear formulation, the two inter-group blocks play symmetric roles and neither can be neglected without changing the solution in a systematic way.

These results provide strong evidence that, for the linear formulation of PILE-BWM, the inter-$\mathcal{W}$ block is largely redundant from a practical point of view. The full model and the compact B-only model yield almost identical global weights, while the W-only model can differ substantially. For the *non-linear* formulation, however, inter-$\mathcal{B}$ and inter-$\mathcal{W}$ constraints both contribute significantly to ξ^* and removing either block leads to comparable distortions of the weight vector. From an efficiency standpoint, this suggests that a compact only-B model (6) is a justified simplification for the linear PILE-BWM, whereas in the non-linear case both inter-$\mathcal{B}$ and inter-$\mathcal{W}$ comparisons should be retained.

Data availability

Data will be made available on request.

References

[1] Kizielewicz B, Tomczyk T, Gandor M, Sałabun W. Subjective weight determination methods in multi-criteria decision-making: A systematic review. Procedia Comput Sci 2024;246:5396–407.
[2] Choo EU, Schoner B, Wedley WC. Interpretation of criteria weights in multicriteria decision making. Comput Ind Eng 1999;37(3):527–41.
[3] Saaty TL. A scaling method for priorities in hierarchical structures. J Math Psych 1977;15(3):234–81.
[4] Brunelli M. Why should not a decision analyst be content with only (n-1) pairwise comparisons? Echoes from the literature. In: The international workshop on best-worst method. Springer; 2022, p. 33–40.
[5] Shanteau J, Weiss DJ, Thomas RP, Pounds JC. Performance-based assessment of expertise: How to decide if someone is an expert or not. European J Oper Res 2002;136(2):253–63.
[6] Saaty TL. How to make a decision: The analytic hierarchy process. European J Oper Res 1990;48(1):9–26.
[7] Keeney RL, Raiffa H. Decisions with multiple objectives: Preferences and value trade-offs. Cambridge University Press; 1993.
[8] Rezaei J, Arab A, Mehregan M. Analyzing anchoring bias in attribute weight elicitation of smart, swing, and best-worst method. Int Trans Oper Res 2024;31(2):918–48.
[9] Rezaei J. Best-worst multi-criteria decision-making method. Omega 2015;53:49–57.
[10] Rezaei J. Best-worst multi-criteria decision-making method: Some properties and a linear model. Omega 2016;64:126–30.
[11] Smith JE, Dyer JS. On (measurable) multiattribute value functions: An expository argument. Decis Anal 2021;18(4):247–56.
[12] Parnell GS, Conley HW, Jackson JA, Lehmkuhl LJ, Andrew JM. Foundations 2025: A value model for evaluating future air and space forces. Manag Sci 1998;44(10):1336–50.
[13] Wulf C, Mesa Estrada LS, Haase M, Tippe M, Wigger H, Brand-Daniels U. MCDA for the sustainability assessment of energy technologies and systems: Identifying challenges and opportunities. Energy, Sustain Soc 2025;15(1):45.
[14] Maslow AH. A theory of human motivation. Psychol Rev 1943;50(4):370.
[15] Rezaei J. Best-worst method: A decade of evolution and future prospects. Omega 2026;103546.
[16] Saaty TL, Wind Y. Marketing applications of the analytic hierarchy process. Manag Sci 1980;26(7):641–58.
[17] Keeney RL. Value-focused thinking: Identifying decision opportunities and creating alternatives. European J Oper Res 1996;92(3):537–49.
[18] Keeney RL, Gregory RS. Selecting attributes to measure the achievement of objectives. Oper Res 2005;53(1):1–11.
[19] Güler M, Samastı M, Yücesan M, Gül M, Çelik E, Ciner MN, Algorabi Ö. Selection of the central composting facility for organic wastes through the bayesian best worst method and geographic information system incorporation: The case of istanbul. Eng Sci Technol Int J 2025;70:102175.
[20] Huang S-H, Chen J-W, Nguyen HT, Hsu W-K. An assessment of service quality for cold-chain logistics in air freight: A perception-expectation gap model based on fuzzy best-worst method. Int J Refrig 2025;170:164–71.
[21] Berberoglu Y, Mangla SK, Kazancoglu Y. Towards sustainable mining in an emerging economy: Assessment of sustainability challenges. Resour Policy 2024;97:105288.
[22] Gholamizadeh K, Pamucar D, Moslem S, Basiri P, Esztergár-Kiss D, Mohammadfam I. Decoding rail derailments: Unraveling the weighted factors influencing safety and sustainability using the best-worst method. Results Eng 2024;23:102539.
[23] Hossain M, Al Aziz R, Karmaker CL, Debnath B, Bari AM, Islam ARMT. Exploring the barriers to implement industrial symbiosis in the apparel manufacturing industry: Implications for sustainable development. Heliyon 2024;10(13).
[24] Sarwar M, Akram M, Gulzar W, Deveci M. Group decision making method for third-party logistics management: An interval rough cloud optimization model. J Ind Inf Integr 2024;41:100658.
[25] Sharma R, Gupta H. Leveraging cognitive digital twins in industry 5.0 for achieving sustainable development goal 9: An exploration of inclusive and sustainable industrialization strategies. J Clean Prod 2024;448:141364.
[26] Singh AK, Kumar VP. Analyzing the barriers for blockchain-enabled bim adoption in facility management using best-worst method approach. Built Environ Proj Asset Manag 2024;14(2):164–83.
[27] Wu L, Hou Z, Luo Z, Fang Y, Mao J, Qin N, Guo Y, Zhang T, Cai N. Site selection for underground bio-methanation of hydrogen and carbon dioxide using an integrated multi-criteria decision-making (MCDM) approach. Energy 2024;306:132437.
[28] Favargiotti S, Zantedeschi G, Pianegonda A, Brunelli M, Urbani M, et al. Designing food hubs for territories of proximity: Assessing the spatial, ecological, and cultural potentials of places through multi-criteria decision support systems. Land 2024;13(8):1–24.
[29] Irawan CA, Ouelhadj D, Jones D, Stålhane M, Sperstad IB. Optimisation of maintenance routing and scheduling for offshore wind farms. European J Oper Res 2017;256(1):76–89.

[30] Bouchard M, D'Amours S, Rönnqvist M, Azouzi R, Gunn E. Integrated optimization of strategic and tactical planning decisions in forestry. European J Oper Res 2017;259(3):1132–43.

[31] Eisenführ F, Weber M, Langer T. Rational decision making. Springer; 2010.

[32] Liang F, Brunelli M, Rezaei J. Best-worst tradeoff method. Inform Sci 2022;610:957–76.

[33] Brunelli M, Cavallo B. Incoherence measures and relations between coherence conditions for pairwise comparisons. Decis Econ Finance 2020;43(2):613–35.

[34] Genest C, Zhang S-S. A graphical analysis of ratio-scaled paired comparison data. Manag Sci 1996;42(3):335–49.

[35] Mizuno T. A block chart visualizing positive pairwise comparison matrices. In: International KES conference on intelligent decision technologies. Springer; 2023, p. 282–9.

[36] Montibeller G, Von Winterfeldt D. Cognitive and motivational biases in decision and risk analysis. Risk Anal 2015;35(7):1230–51.

[37] Saaty TL. The modern science of multicriteria decision making and its practical applications: The AHP/ANP approach. Oper Res 2013;61(5):1101–18.

[38] Brunelli M. A survey of inconsistency indices for pairwise comparisons. Int J Gen Syst 2018;47(8):751–71.

[39] Sun G, Kroesen M, Rezaei J. Anchoring bias in the tradeoff procedure within multi-attribute value theory. J Behav Decis Mak 2026;39(2):e70069.

[40] Huang R, Corrente S, Siskos E, Burgherr P. The multi-level cards method: Weights elicitation, consistency checking and restoration for hierarchical criteria. Omega 2025;103438.