INFORMATICA, 2026, Vol. 0, No. 0, 1–36

DOI: https://doi.org/10.15388/26-INFOR637

A Consensus-Based Best–Worst Method for Linguistic Multi-Criteria Group Decision-Making

Álvaro LABELLA*, Diego GARCÍA-ZAMORA, Bapi DUTTA, Luis MARTÍNEZ
Campus Las Lagunillas s/n 23071, University of Jaén, Spain
e-mail: alabella@ujaen.es, dgzamora@ujaen.es, bdutta@ujaen.es, martin@ujaen.es

Received: May 2026; accepted: June 2026

Abstract. Determining criteria importance is a crucial task in multi-criteria decision-making problems, and the Best-Worst Method (BWM) has emerged as an effective weighting technique due to its reduced number of pairwise comparisons. Although BWM has been extended to handle group decision-making and linguistic information, existing approaches do not adequately address disagreements among evaluators, which may lead to dissatisfaction with the resulting criteria weights. To address this limitation, this paper proposes a consensus-based BWM for linguistic multi-criteria group decision-making. The proposed approach obtains consensual collective weights by minimally modifying the evaluators' initial linguistic preferences. The resulting solution can support moderators and evaluators during consensus-reaching processes by facilitating the identification of disagreements and the generation of appropriate recommendations. The proposal is developed within the 2-tuple linguistic framework and provides both numerical and linguistic representations of the resulting weights. In addition, a novel stochastic consistency index is introduced to assess the reliability of evaluators' preferences. The feasibility and effectiveness of the proposal are illustrated through a real-world case study and analysed by means of comparative and sensitivity analyses.

Key words: Multi-criteria group decision-making, consensus-reaching process, Best-Worst Method, 2-tuple linguistic model, linear optimization.

1. Introduction

Making decisions is a routine activity for human beings. Typically, a decision-making problem involves selecting the best solution among several alternatives. However, decision-making problems become more complex when multiple criteria must be considered. These types of problems are referred to as multi-criteria decision-making (MCDM) problems and arise in a variety of contexts, such as supplier selection (Kabadayi and Dehghanimohammadabadi, 2022; Sahoo *et al.*, 2024), renewable energy (Asakereh *et al.*, 2022; Jameel *et al.*, 2026), healthcare (Bouraima *et al.*, 2024; Puška *et al.*, 2022), and education (Srivastava *et al.*, 2024). Multi-criteria group decision-making (MCGDM) problems occur when the decision-making process for MCDM problems is carried out by taking into account the preferences of several evaluators.

*Corresponding author.

The participation of multiple evaluators in an MCGDM problem allows the decision problem to be analysed from different perspectives. In the specialized literature, there are multiple rules for solving MCGDM problems. Among others, Authority Rule (Pingle, 1997), in which a leader has the authority to make the final decision, Majority Rule (Straffin Jr, 1977) based on a voting process, Minority Rule (Butler and Rothstein, 2006), in which the decision is assigned to a subgroup of people, Negative Minority Rule (Butler and Rothstein, 2006), where the group votes for the least popular alternatives to be removed and several rounds are carried out, or Consensus Rules, in which all group members agreed that the solution chosen is a satisfactory solution (García-Zamora *et al.*, 2023a). This paper follows the latter type of rules and the concept of consensus-reaching process (CRP) to address MCGDM problems. In a CRP, usually supervised by a moderator, evaluators discuss among themselves and try to increase the level of agreement within the group to reach a solution that satisfies its members. However, CRPs are usually time-consuming and become increasingly complex as the number of evaluators grows. Under these circumstances, automatic CRP models have proven to facilitate the task of the moderator, supporting the consensus process in an effective and accurate way, because these models are able to detect disagreements and provide agreed solutions without evaluators' feedback (García-Zamora *et al.*, 2023b; García-Zamora *et al.*, 2025). Obviously, these solutions are valid from a theoretical point of view, but they can serve as a guide for moderators to guide a CRP through the feedback process. In this regard, our contribution addresses the handling of disagreements between evaluators in MCGDM problems from an automatic CRP perspective and subsequently helps the moderators to conduct CRPs with feedback in a faster and more efficient way if it is necessary.

Solving MCGDM problems is becoming increasingly complex due to the lack of information and uncertainty often present in decision-making contexts. This can lead to doubts and vague judgments that cannot be easily modelled by using numerical values. To address this challenge, fuzzy linguistic approaches have been widely adopted (Zadeh, 1975), allowing evaluators to express their preferences by using linguistic expressions that align better with human communication patterns. MCGDM problems in which evaluators provide their opinions by means of linguistic expressions are known as linguistic MCGDM (LiMCGDM) problems. Several linguistic models have been proposed in the literature (Du *et al.*, 2023; Martínez *et al.*, 2015; Wang and Hao, 2006), but the 2-tuple linguistic model (Martínez *et al.*, 2015) is particularly noteworthy for its ability to perform precise computing with words (CW) processes.

One of the most common tasks in MCDM problems is evaluating the importance of the criteria involved in the problem. Various decision methods have been developed to derive the importance of criteria from evaluators' opinions, such as the expert judgment method proposed by Kendall based on expert questioning (Kendall, 1970), the Analytic Hierarchy Process (AHP) introduced by Saaty (1977, 1990) that obtains criteria weights from pairwise comparisons among the criteria, or the step-wise weight assessment ratio analysis (SWARA) by Keršuliene *et al.* (2010) to estimate the opinions of experts or interest groups on the significance ratio of criteria.

The BWM has gained popularity due to its simplicity and reliability in deriving criteria weights (Rezaei, 2015). The BWM relies on pairwise comparisons between the best and worst criteria, which are chosen by the evaluator, and the remaining criteria. It addresses some behavioural errors related to the consistency of the preferences that arise in similar MCDM methods. Despite its effectiveness, the first version of this method has limitations:

- Individual decision-making: the BWM was proposed to deal with MCDM problems in which only one person is involved. Therefore, MCGDM problems were initially omitted.
- Numerical assessments: in the BWM, evaluators use the well-known Saaty's scale, a numerical scale in 1–9 (Saaty, 1977, 1990). However, these numerical assessments cannot model the inherent uncertainty in many MCDM problems. Therefore, the LiMCGDM problems were also initially ignored.
- Numerical results: the BWM provides numerical weights for the criteria, but although for some evaluators the numerical weights may be interpretable enough, others could prefer a representation closer to their common way of thinking, using words.

Extant works have already presented several attempts to address the previous limitations, either by adapting the BWM to MCGDM (Moreno-Albarracín *et al.*, 2020; Safarzadeh *et al.*, 2018) or even by using linguistic information (Licerán-Gutiérrez *et al.*, 2022). Safarzadeh *et al.* (2018) proposed a Group BWM (G-BWM) to manage MCGDM problems. However, it still has some drawbacks, such as the imposition that all evaluators select the same best and worst criterion, the omission of the LiMCGDM problems, or its inability to deal with the disagreements that may arise among evaluators. On the other hand, the Minimum Cost Consensus BWM (MCC-BWM) versions introduced in Licerán-Gutiérrez *et al.* (2022); Moreno-Albarracín *et al.* (2020) follow a similar approach. First, they applied an optimization BWM model for each evaluator to obtain individual weights. Afterward, such individual weights are used as inputs of an MCC optimization model (Ben-Arieh *et al.*, 2008) to obtain collectively agreed weights. The MCC model allowed for deriving collective weights by changing as little as possible the individual weights and obtaining an agreed solution. However, this way of dealing with disagreements can lead to inconsistent results.

Therefore, this paper aims to support CRPs with feedback and deal properly with the conflicts that may appear among the participants in a LiMCGDM problem to fix the importance of the criteria. A novel MCC-BWM method is then proposed to adapt the traditional BWM to a 2-tuple linguistic setting, and generates optimal agreed collective weights based on pairwise linguistic comparisons of their views on the best to others, and others to the worst criteria. These collective weights are directly obtained as numeric values from a single mathematical programming model that integrates consensus constraints within classic BWM to obtain the theoretical optimum for the CRP. To retrieve a linguistic output, an unbalanced 2-tuple linguistic scale (Licerán-Gutiérrez *et al.*, 2022; Martínez *et al.*, 2015) is then used to guarantee the interpretability of the results. In comparison to a balanced linguistic scale, unbalanced scales allow describing situations in which the linguistic labels are not symmetrically distributed in the universe of discourse in which they are defined (Herrera *et al.*, 2008). These linguistic outputs can be used by the moderators involved in a CRP to (i) identify disagreements among the evaluators, (ii) identify the changes in the preferences to be suggested, and, ultimately, (iii) to reach the desired consensus, keeping the evaluators' initial opinions as much as possible. In addition, the proposal includes a novel consistency measure to quantify the consistency of the preferences provided by the evaluators.

To sum up, the main novelties of the proposal are listed below:

- Group decision-making: the proposal can deal with BWM MCGDM problems in which several evaluators' opinions are considered.
- Linguistic information: the proposal replaces the numerical Saaty's scale used in BWM with a linguistic scale to model linguistic evaluators' preferences. In addition, the results of the optimization model are translated into 2-tuple linguistic values by using the fuzzy unbalanced linguistic approach, which facilitates their interpretability and keeps the accuracy of the numerical results (Licerán-Gutiérrez *et al.*, 2022).
- Consensual solution: the proposal detects disagreements among evaluators, smooths them out, and obtains agreed optimal collective weights for the criteria that will be analysed by the moderator in a CRP with feedback to provide proper suggestions to the evaluators if necessary, and reach the desired agreed solution faster.
- Stochastic consistency index: the proposal provides a novel consistency index to measure the reliability of evaluators' opinions with statistical insight.

This paper is structured as follows: Section 2 provides a review of fundamental concepts related to the proposed method. Section 3 presents an extension of the BWM for handling disagreements in LiMCGDM problems. In Section 4, the feasibility and applicability of the proposed method are illustrated through a case study related to sustainable packaging. Section 5 further validates the proposal by means of a comprehensive numerical analysis, including computational scalability experiments, comparative analyses with related BWM approaches, sensitivity analyses, and a discussion of the main findings. Finally, Section 6 summarizes the main conclusions and identifies avenues for future research.

2. Background

In this section, some preliminary concepts related to the proposal are revised. Firstly, the notion of LiMCGDM is reviewed. Afterward, the 2-tuple linguistic model is described. Then, CRPs are introduced as a tool to soften the disagreements in MCGDM processes. Finally, the classical BWM and some of its extensions are discussed.

2.1. *Linguistic MCGDM*

An MCGDM problem consists of several elements (Ishizaka and Nemery, 2013):

- A finite set of m, ($m \in \mathbb{N}, m \geqslant 2$), evaluators $E = \{E_1, E_2, \dots, E_m\}$.
- A finite set of r, ($r \in \mathbb{N}, r \geqslant 2$), alternatives $A = \{A_1, A_2, \dots, A_r\}$.
- A finite set of n, ($n \in \mathbb{N}, n \geqslant 2$), criteria $C = \{C_1, C_2, \dots, C_n\}$.

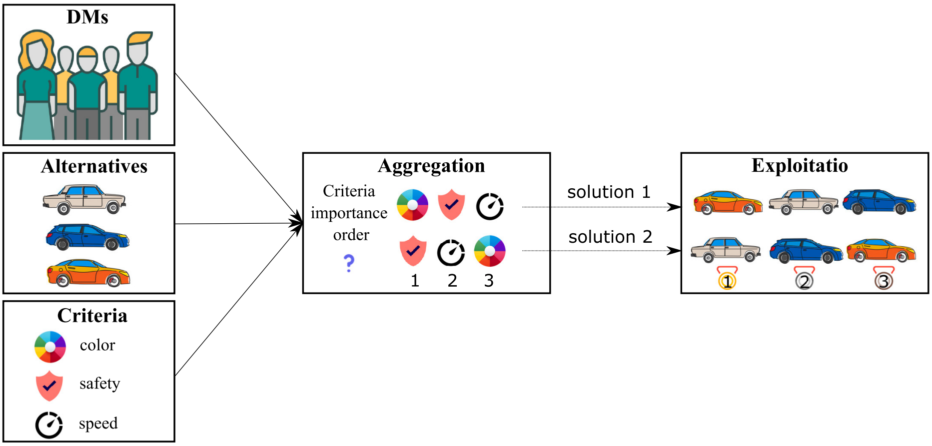

Fig. 1. MCGDM resolution scheme.

The group of evaluators evaluates the alternatives based on some criteria to select the best alternative/s as the solution to the problem (Tüysüz and Kahraman, 2023). In this sense, the importance of the criteria plays a key role in the resolution of an MCGDM problem (Rezaei, 2015; Saaty, 1977). The classical MCGDM resolution scheme carries out an aggregation process of the evaluators' assessments before selecting the best alternative, which should properly rate the relevance of the criteria to reach reliable results (see Fig. 1). For instance, it is logical to assume that when buying a car, safety has a greater weight than colour in the decision-making process.

Inevitably, society evolves and demands even more complex decision problems to be solved. Consequently, evaluators have to deal with problems in which there is a lack of information, or that are defined in vague and imprecise contexts. Under these uncertain conditions, evaluators may not feel comfortable providing numerically precise assessments, and they may prefer to use expressions closer to their natural way of thinking. The fuzzy linguistic approach addresses the uncertainty modelling in human beings' opinions through the use of linguistic variables defined by Zadeh as "*a variable whose values are not numbers, but words or sentences in a natural or artificial language*" (Zadeh, 1975). When evaluators face MCGDM problems by using linguistic information, we talk about LiMCGDM. Classically, in the resolution of a LiMCGDM problem, evaluators use a linguistic term set, i.e. an ordered set of linguistic descriptors, $S = \{s_0, s_1, \dots, s_G\}$, to provide their assessments (see Fig. 2).

The use of linguistic information for uncertainty modelling involves performing computations based on it. The aim of the CW approach (Zadeh, 1996) is to emulate the reasoning process of humans by generating linguistic results from linguistic inputs, which are easily comprehensible and accurately depicted. Various CW proposals have been put forth in the literature (González and Lantigua, 2024; Li *et al.*, 2023; Martínez *et al.*, 2015; Wang and Hao, 2006). In particular, this paper adopts the 2-tuple linguistic representation model (Martínez *et al.*, 2015) for modelling linguistic information and performing CW

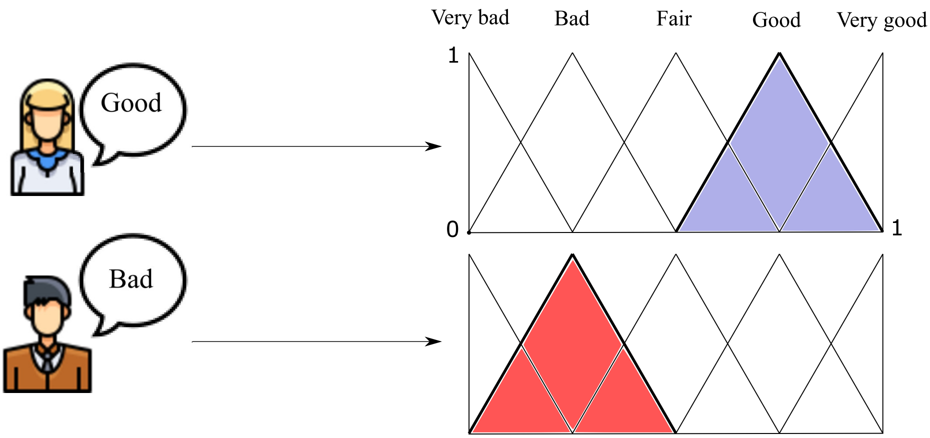

Fig. 2. Linguistic variable.

processes. This approach allows the modelling of uncertainty using the fuzzy linguistic approach while retaining the integrity, accuracy, and interpretability of the results.

2.2. *2-Tuple Linguistic Model*

The 2-tuple linguistic model (Martínez *et al.*, 2015) uses the fuzzy linguistic approach to enable linguistic computations in a precise and understandable manner. This model represents linguistic information using a pair of values $(s_l, \alpha) \in \overline{S}$, where s_l corresponds to a linguistic term from a pre-defined set of terms $S = \{s_0, s_1, \dots, s_G\}$, and $\alpha \in [-0.5, 0.5)$ represents the displacement of the membership function of the linguistic term s_l to the right or left. Specifically, the symbolic translation is defined as follows:

$$\alpha \in \begin{cases} [-0.5, 0.5), & \text{if } s_l \in \{s_1, s_2, \dots, s_{G-1}\}, \\ [0, 0.5), & \text{if } s_l = s_0, \\ [-0.5, 0], & \text{if } s_l = s_G. \end{cases} \tag{1}$$

One significant feature of 2-tuple linguistic expressions is their ability to be converted into a numerical value $x \in [0, G]$, which streamlines the computational process.

Proposition 1 (Martínez *et al.*, 2015). *Let $S = \{s_0, \dots s_G\}$ be a linguistic term set. Then, the function $\Delta_S^{-1} : \overline{S} \to [0, G]$ defined by*

$$\Delta_S^{-1}(s_l, \alpha) = l + \alpha, \quad \forall (s_l, \alpha) \in \overline{S} \tag{2}$$

is a bijection whose inverse $\Delta_S : [0, G] \to \overline{S}$ is given by

$$\Delta_S(x) = \big(s_{round(x)}, x\text{-}round(x)\big) \quad \forall x \in [0, G], \tag{3}$$

where $round(\cdot)$ is the function that assigns the closest integer number $l \in \{0, \dots, G\}$.

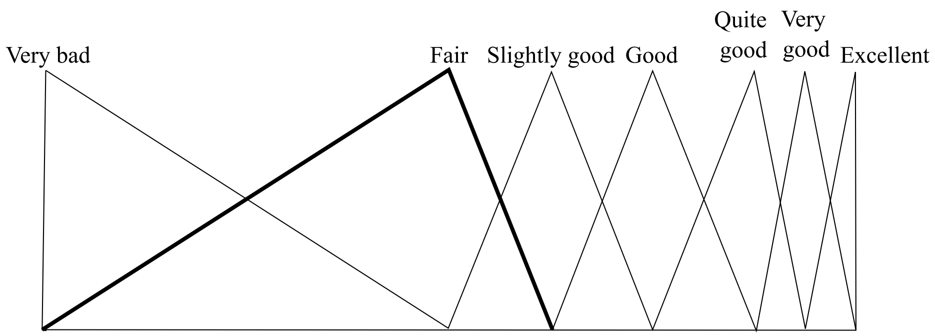

Fig. 3. Unbalanced linguistic scale (Herrera *et al.*, 2008).

Remark 1. It is important to note that a linguistic term $s_l \in S$ can be converted into a 2-tuple linguistic value in $\overline{S}$ by including a zero as a symbolic translation to the linguistic term.

$$s_l \in S \to (s_l, 0) \in \overline{S}. \tag{4}$$

Consequently, to make computations with linguistic terms using the 2-tuple linguistic model, it suffices to use Proposition 1 to transform the linguistic inputs into numerical values, carry out the corresponding computational processes, and use Proposition 1 to re-translate the numeric results into linguistic expressions. Note that the resulting 2-tuple linguistic value will be of the form (s_l, α), and usually, it will not correspond to one of the original linguistic terms in S, i.e. $\alpha \neq 0$. In that case, the numeric value $\Delta_S^{-1}(s_l, \alpha)$ will be located between two consecutive linguistic terms, namely s_{l-1} and s_l, or s_l and s_{l+1}, and therefore it will have a positive membership degree for the triangular fuzzy numbers associated to these two labels, which depend on the value of α.

Many problems that involve modelling information by using linguistic assessments typically rely on linguistic variables with uniformly and symmetrically distributed values (see Fig. 2). However, some problems require the use of unbalanced linguistic term sets, which are asymmetrically distributed (Herrera *et al.*, 2008; Licerán-Gutiérrez *et al.*, 2022). Unbalanced linguistic scales are necessary when the linguistic variables involved in the problem require different degrees of specificity on each side of a central label. Figure 3 illustrates an example of an unbalanced linguistic scale asymmetrically distributed, with the central label being represented by the linguistic term "Fair".

2.3. *Consensus-Reaching Process*

When multiple evaluators with diverse viewpoints and expertise are involved in an MCGDM process, conflicts can arise among them. However, the resolution scheme depicted in Fig. 1 overlooks such potential conflicts and performs an aggregation process that yields a solution based on the majority opinion, which may not be agreeable to all evaluators (Butler and Rothstein, 2006). To address this issue, CRPs were introduced as an additional step in resolving group decision-making problems (García-Zamora *et al.*, 2023a). In a CRP, evaluators (or decision-makers) express and justify their opinions and strive to converge on a common ground before reaching a decision that is acceptable to everyone.

In the specialized literature, CRPs have been classified into two main categories: CRPs with and without a feedback process (García-Zamora *et al.*, 2023a). The CRPs with the feedback process are usually guided by a moderator who identifies the changes to make over the evaluators' preferences and asks evaluators if they want to accept or reject such changes. The CRPs without a feedback process replace the role of the moderator with an automatic process that detects the changes to the evaluators' preferences and applies such changes automatically without checking with them. Among the latter proposals, MCC models (Ben-Arieh *et al.*, 2008; García-Zamora *et al.*, 2023b) are especially convenient when used as a consensus support system to guide (decision) analysts/moderators in interactive CRPs to quickly detect disagreements and suggest changes in a more precise way (García-Zamora *et al.*, 2025).

CRPs can help to enhance the level of consensus among groups of evaluators. However, the notion of consensus is not straightforward. Traditionally, it has been defined as the unanimous agreement of all evaluators involved in a problem (Butler and Rothstein, 2006), which is practically unattainable in real-world decision-making scenarios. Therefore, an alternative concept of consensus has been proposed in the literature. This paper focuses on the soft consensus introduced by Kacprzyk and Fedrizzi (1988), which posits that consensus exists when "*most of the important individuals agree as to (their testimonies concerning) almost all the relevant options*".

Therefore, instead of a unanimous consensus, we consider here a more flexible consensus definition based on consensus measures (Palomares *et al.*, 2014). Let us denote by $\mathcal{P}$ the set consisting of all the possible ratings that a group of evaluators may provide. Then, a consensus measure is a mathematical function $\kappa : \mathcal{P} \to \mathbb{R}_0^+$ that, intuitively, allows computing the distance/dissimilarity between these opinions (García-Zamora *et al.*, 2023a). Therefore, given the opinions of a certain group of evaluators $P \in \mathcal{P}$, the value $\varepsilon = \kappa(P)$ is known as the consensus degree of such preferences, and the closer to one, the higher the similarity (consensus) between the opinions P. In the CRPs that follow the consensus measure-based approach, a certain consensus threshold ε_0 is fixed a priori, and the goal of the CRP is to modify the initial evaluators' preferences P_0 into new values P that satisfy $\kappa(P) \leqslant \varepsilon_0$. This approach enables the assessment of how far the opinions of evaluators are from unanimity. Note that ε_0 is a unitless magnitude that represents the desired level of consensus to achieve in the group. The range of values is the positive real numbers, in which 0 means total agreement. Consequently, this value should not be predefined by the evaluators involved in the negotiation but by a selected person who could be an analyst, the moderator of the process, or even the decision-makers. The value assigned depends on the context of the problem. Some aspects could be considered to set a proper value of ε_0, such as the number of evaluators, which plays an important role here; a decision involving hundreds of evaluators cannot be expected to reach a very high consensus within a few modifications. Another aspect to consider may be the initial consensus degree $\kappa(P_0)$ within the group before the CRP.

2.4. *Best-Worst Method*

The BWM (Brunelli and Rezaei, 2019; Rezaei, 2015; Ortega-Rodríguez and Labella, 2026) was proposed as a prioritization methodology for deriving the weights of the criteria by reducing inconsistencies in the elicitation process and obtaining more reliable solutions. To do so, the evaluator chooses the best and the worst criteria, which are denoted by C_B and C_W, respectively. Afterward, the evaluator must point out the comparisons of C_B regarding the remaining criteria and all the criteria regarding C_W. To model such comparisons, BWM makes use of Saaty's multiplicative scale (Saaty, 1977; Adali and Tuş, 2026), which is decomposed into two parts: numerical scale and linguistic scale. Whereas the linguistic scale consists of 9 gradations, i.e. 1: Equally important, 2: Weakly more important, 3: Moderately more important, 4: Moderately plus more important, 5: Strongly more important, 6: Strongly plus more important, 7: Demonstrated more important, 8: Very, very strongly more important, and 9: Extremely more important, the numerical scale is composed of 17 numerical values described as:

$$\left\{\frac{1}{f_t}, f_1, f_t\right\}, \quad t = 2, 3, \dots, 9,$$

where $f_1 = 1$ and $f_{t+1} > f_t > 1$ and the value of f_t correspond to the t grade of the linguistic scale.

The pairwise comparisons among C_B and the remaining criteria are collected in the Best-to-Others (BO) vector,

$$BO = (b_{B1}, b_{B2}, \dots, b_{Bn}),$$

where $b_{Bi} \in [1, 9] \cap \mathbb{N}$ denotes the degree of preference of C_B over the criterion C_i.

In the same way, the evaluator's pairwise comparisons regarding all the criteria with C_W are collected in the Others-to-Worst (OW) vector,

$$OW = (b_{1W}, b_{2W}, \dots, b_{nW}),$$

where $b_{iW} \in [1, 9] \cap \mathbb{N}$ denotes the preference degree of the criterion C_i over C_W. These values are then used as the input of an optimization model to obtain the weights for the criteria $\{w_1^*, w_2^*, \dots, w_n^*\}$:

$$\begin{aligned} &\min_{w} \max_{i=1,2,\dots,n} \left\{\left|b_{Bi} - \frac{w_B}{w_i}\right|, \left|b_{iW} - \frac{w_i}{w_W}\right|\right\}, \\ &\text{s.t.} \quad \begin{cases} \sum_{i=1}^{n} w_i = 1, \\ w_i \geqslant 0, \quad \forall i = 1, 2, \dots, n. \end{cases} \end{aligned} \tag{5}$$

The BWM produces weights that allow constructing a fully consistent multiplicative pairwise comparison matrix ($\hat{b}_{ij} := \frac{w_i}{w_j}$) close to the original preferences BO and OW specified by the evaluator, thanks to the full consistency property of the multiplicative

Table 1
Unbalanced linguistic scale, S_n^U, for weights in proposal.

Linguistic terms	Membership function
Unimportant (U)	$(0, 0, \frac{1}{n})$
Average Important (AI)	$(0, \frac{1}{n}, \frac{n+2}{3n})$
Important (I)	$(\frac{1}{n}, \frac{n+2}{3n}, \frac{2n+1}{3n})$
Very Important (VI)	$(\frac{n+2}{3n}, \frac{2n+1}{3n}, 1)$
Extremely Important (EI)	$(\frac{2n+1}{3n}, 1, 1)$

pairwise comparison matrix in 1-9 Saaty's scale (Saaty, 1977; Majumder *et al.*, 2026), where $b_{ij} = b_{ik}b_{kj}$ for $i, j, k \in 1, 2, \dots, n$. Additionally, the BWM optimization model proposes a consistency index to measure the reliability of evaluators' preferences (see Rezaei, 2015 for further details).

Lately, the classical BWM has been extended to versions with additional features. Particularly, this proposal takes advantage of one of them introduced in Licerán-Gutiérrez *et al.* (2022), which is the use of an unbalanced linguistic scale (Herrera *et al.*, 2008) to represent the BWM weights linguistically. This proposal argued that to represent the importance of the BWM weights in this format, it is necessary to define a suitable unbalanced linguistic scale (Licerán-Gutiérrez *et al.*, 2022), here noted as S_n^U, whose semantics depends on the n number of the elements to be weighted (see Table 1).

Remark 2. It is important to note that the average value of a family of weights is indeed determined by the number of elements to compare:

$$\frac{1}{n}\sum_{i=1}^{n} w_i = \frac{1}{n}. \tag{6}$$

For example, the average importance of 3 criteria is 0.33, while that of 7 elements is approximately 0.143. Therefore, to adequately express the importance of elements by linguistic terms, an unbalanced linguistic scale is utilized, which is constructed based on the number of elements n being evaluated.

3. Consensus BWM Under 2-Tuple Environment

This section first introduces the selected linguistic additive scale that replaces Saaty's multiplicative scale used in the classical BWM to provide the evaluators' preferences. Afterward, a novel consensus BWM optimization model is introduced, which allows facing LiMCGDM problems regarding evaluating criteria importance and obtaining consensual solutions represented both numerically and linguistically. Finally, a novel consistency index is provided to evaluate the reliability of evaluators' preferences.

3.1. *Linguistic Scale*

The 2-tuple linguistic approach is usually based on a symmetrically distributed additive scale, which suggests that the pairwise comparisons necessary to apply the linguistic BWM should also be modelled as an additive scale (Orlovski, 1978), rather than the classic multiplicative Saaty's scale (Saaty, 1977). The use of the 2-tuple linguistic model within the BWM method allows not only the evaluators to provide their preferences using a linguistic scale but also offers an output in terms of linguistic values. This is extremely important to develop a feedback mechanism, since the analyst/moderator may provide the recommendations using expressions such as "the importance of the criterion C1 is average" or "criterion C2 is unimportant".

This section aims to clarify how to remap linguistic preferences into the interval [0, 1] to compute weights using BWM.

Let us consider a linguistic term set $S = \{s_0, \dots, s_G\}$ of granularity $G + 1 \in \mathbb{N}$. To remap the linguistic labels into a numeric additive scale, we can define the mapping $\hat{\Delta}_S : [0, 1] \to \overline{S}$ as

$$\hat{\Delta}_S(\beta) = \big(s_{round(G\beta)}, G\beta - round(G\beta)\big), \quad \forall\beta \in [0, 1], \tag{7}$$

whose inverse $\hat{\Delta}_S^{-1} : \overline{S} \to [0, 1]$ is

$$\hat{\Delta}_S^{-1}(s_l, \alpha) = (l + \alpha)/G, \quad \forall(s_l, \alpha) \in \overline{S}. \tag{8}$$

Note that the mapping $\hat{\Delta}_S$ is slightly different from the mapping Δ_S. Whereas the input of the first one is the centroid of the triangular fuzzy number associated with the linguistic label, the input of the second one is the index of the linguistic label.

The linguistic scale S represents the element importance in both directions, so that the preference degree between the criterion C_i and C_j is s_l, then the inverse preference is s_{G-l}. Nevertheless, the linguistic scale S consists of a set of linguistic descriptors with an uncertain and imprecise nature that should be represented in some way. In this sense, the fuzzy linguistic approach uses the fuzzy sets theory to handle vague and imprecise information (Parsons, 1996) (see Section 2.1). To do so, a fuzzy membership function is associated with each linguistic descriptor, now a fuzzy linguistic term, resulting in a fuzzy linguistic term set. Therefore, we set a fuzzy membership function for each linguistic descriptor of S to define a fuzzy linguistic term set S. For instance, using these mappings for $G = 16$ may lead to the linguistic scale given in Table 2, whose semantics are represented in Fig. 4.

Consequently, the evaluators may use the linguistic scale S to provide their BWM preferences. However, because of the nature of the BWM performance in which the evaluators compare how important the best criterion is with the remainder and how important the rest of the criteria are concerning the worst criterion, the evaluators will only use the subset $\{EI, WMI, MMI, MPMI, SMI, SPMI, DMI, VSMI, EMI\}$ of the scale.

In addition, after applying BWM to derive numerical weights in [0, 1], it will be necessary to retrieve the linguistic values and represent such weights in a format closer to

Table 2
Fuzzy linguistic scale.

Membership function	Linguistic term
(0, 0, 0.062)	Extremely less important (ELI)
(0, 0.062, 0.125)	Very, very, strongly less important (VLI)
(0.062, 0.125, 0.188)	Demonstratedly less important (DLI)
(0.125, 0.188, 0.25)	Strongly plus less important ($SPLI$)
(0.188, 0.25, 0.312)	Strongly less important (SLI)
(0.25, 0.312, 0.375)	Moderately plus less important ($MPLI$)
(0.312, 0.375, 0.438)	Moderately less important (MLI)
(0.375, 0.438, 0.5)	Weakly less important (WLI)
(0.438, 0.5, 0.562)	Equally important (EI)
(0.5, 0.562, 0.625)	Weakly more important (WMI)
(0.562, 0.625, 0.688)	Moderately more important (MMI)
(0.625, 0.688, 0.75)	Moderately plus more important ($MPMI$)
(0.688, 0.75, 0.812)	Strongly more important (SMI)
(0.75, 0.812, 0.875)	Strongly plus more important ($SPMI$)
(0.812, 0.875, 0.938)	Demonstratedly more important (DMI)
(0.875, 0.938, 1)	Very, very strongly more important ($VSMI$)
(0.938, 1, 1)	Extremely more important (EMI)

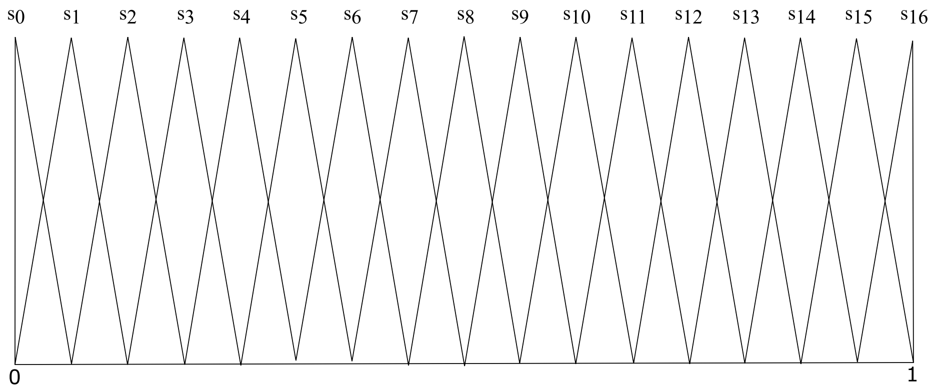

Fig. 4. Fuzzy linguistic scale.

how human beings express their opinions. Therefore, such numerical weights must be remapped into 2-tuple linguistic values in the unbalanced linguistic scale S_n^U (see Table 1) according to the following process (Licerán-Gutiérrez *et al.*, 2022):

Let $\overline{x}_h$ represent the centroid (Cheng, 1998) of the linguistic term s_h, which is assumed to be an element of the unbalanced scale S_n^U with granularity G. Consider the mapping $\hat{\Delta}_U : [0, 1] \rightarrow S_n^U$ defined as $\hat{\Delta}_U(w) = (s_l, \alpha), \forall\, w \in [0, 1]$ where:

$$l = \operatorname*{arg\,min}_{h\in\{0,\dots,G\}} |w - \overline{x}_h|, \tag{9}$$

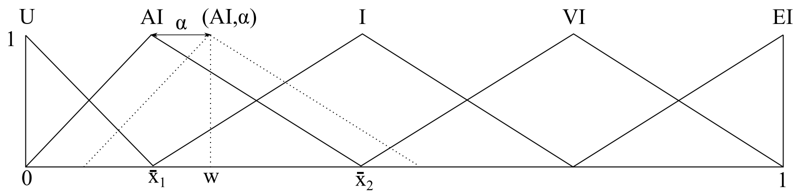

Fig. 5. 2-tuple linguistic weight computation.

$$\alpha = \begin{cases} 0, & \text{if } w = \overline{x}_l, \\ \frac{w-\overline{x}_l}{\overline{x}_{l+1}-\overline{x}_l}, & \text{if } w > \overline{x}_l, \\ \frac{w-\overline{x}_l}{\overline{x}_l-\overline{x}_{l-1}}, & \text{if } w < \overline{x}_l. \end{cases} \tag{10}$$

As an illustrative example, Fig. 5 shows the graphic representation of an unbalanced linguistic scale with 5 fuzzy linguistic terms, in which the centroid of the central label (AI) is the closest one to a numerical weight w, i.e. $\hat{\Delta}_U(w) = (AI, \alpha)$.

3.2. *Consensus Process*

MCC models have been proven to be extremely useful in supporting interactive CRPs by facilitating the achievement of consensus by the group (García-Zamora *et al.*, 2023a). Although they are automatic consensus models without feedback (and thus they cannot ensure the evaluators' acceptance of the generated collective opinion), they can help the analyst/moderator in a CRP to identify the stronger disagreement points and provide precise recommendations to smooth them. In this view, here we introduce a BWM consensus model inspired by MCC. This model could act as a consensus support tool for CRPs.

Let us suppose an MCGDM in which a group of $m \in \mathbb{N}$ evaluators $E = \{E_1, E_2, \dots, E_m\}$ wants to reach a collectively agreed solution about the importance of $n \in \mathbb{N}$ criteria. First, the evaluators provide their opinions by using BW preferences, which implies the evaluator, E_k, chooses the best (B^k) and the worst (W^k) criteria according to his/her opinion and also provides two pairwise comparison vectors:
for the best criteria C_{B^k}

$$BO^k = (a_{B^k1}, a_{B^k2}, \dots, a_{B^kn})$$

and for the worst criteria C_{W^k}

$$OW^k = (a_{1W^k}, a_{2W^k}, \dots, a_{nW^k}),$$

where $a_{B^ki}, a_{iW^k} \in S, i = 1, 2 \dots, n$

Classically, these preferences were given by using a 1–9 Saaty scale (Saaty, 1977). Nevertheless, in this proposal, the evaluators are using fuzzy linguistic terms that belong to a fuzzy linguistic term set S. To facilitate the linearization of the BWM optimization model, these linguistic preferences are remapped into a linear scale in [0, 1] by using the function $\hat{\Delta}_S^{-1}$. To simplify the notation, the information is stored as follows:

- The vectors $BO^k = (a_{B^k 1}, a_{B^k 2}, \dots, a_{B^k n})$, $k = 1, \dots, m$, containing the linguistic comparisons among the best criterion and the remainder for each evaluator,
- The vector $OW^k = (a_{1W^k}, a_{2W^k}, \dots, a_{nW^k})$, $k = 1, \dots, m$, containing the linguistic comparisons among the criteria regarding the worst for each evaluator,
- A vector $B = (B^1, B^2, \dots, B^m) \in \mathbb{N}^m$, $B^k \in \{1, 2, \dots, n\}$, containing the best criterion for each evaluator,
- A vector $W = (W^1, W^2, \dots, W^m) \in \mathbb{N}^m$, $W^k \in \{1, 2, \dots, n\}$, containing the worst criterion for each evaluator.

Consequently, the preference structure can be defined as

$$\mathbb{P}_S := \left\{(B, W, BO, OW) : B, W \in \{1, 2, \dots, n\}, BO, OW \in \{s_8, s_9, \dots, s_{16}\}^n\right\},$$

which is bijective to

$$\mathbb{P} := \left\{(B, W, BO, OW) : B, W \in \{1, 2, \dots, n\}, BO, OW \in [0.5, 1]^n\right\}.$$

In other words, the preferences of each evaluator can be stored as a list (B, W, BO, OW) containing which are the best and the worst criteria B and W and the corresponding pairwise comparison vectors BO and OW. Since we aim to propose a BWM method for groups of m evaluators, let us consider the set $\mathcal{P} = \mathbb{P}^m$ consisting of all the possible values of the preferences of such m evaluators.

Whereas the classical BWM obtains the weights for only one evaluator, our proposal derives the individual weight for each evaluator involved in the decision process and stores them in a matrix

$$U = (u_{ki}) \in \mathcal{M}_{m \times n},$$

where u_{ki} represents the individual weight corresponding to E_k for the criterion C_i. For the sake of simplicity, let us consider the preference structure consisting of the weighting vectors of dimension n $\mathbb{P}_U = \{u \in [0, 1]^n : \sum_{i=1}^n u_i = 1\}$. In such a case, our goal is to associate a weighting vector $u_k = (u_{k1}, \dots, u_{kn}) \in \mathbb{P}_U$ to each evaluator. In addition, our proposal also computes the collective weights from the aggregation of each evaluator's weights:

$$g = (g_1, g_2, \dots, g_n) \in \mathbb{P}_U,$$

where $g_i = F_i(u_1, \dots, u_m)$ for a certain fusion mapping $F : \mathbb{P}_U^m \to \mathbb{P}_U$. In such a case, after running the model, each evaluator will be assigned a weighting vector $u_k \in \mathbb{P}_U$, and we will also obtain the collective weighting vector $g \in \mathbb{P}_U$. For the sake of clarity, in this paper, we consider the arithmetic mean to compute the collective weight vector, i.e. $F_i(u_1, \dots, u_m) = \frac{1}{m} \sum_{k=1}^m u_{ki} \ \ \forall\, i \in \{1, 2, \dots, n\}$, which also allows linearizing the resulting optimization model.

To guarantee agreed weights, let us consider the consensus measure $\kappa : \mathbb{P}_U^m \to \mathbb{R}_0^+$ defined by

$$\kappa(u_1, \ldots, u_m) = \frac{1}{mm} \sum_{i=1}^{n} \sum_{k=1}^{m} |u_{ki} - F_i(u_1, \ldots, u_m)| \; \forall \, (u_1, \ldots, u_m) \in \mathbb{P}_U^m.$$

Therefore, in our model, individual and collective weights must satisfy the following consensus constraint:

$$\kappa(u_1, \ldots, u_m) \leqslant \varepsilon_0. \tag{11}$$

This constraint computes the level of agreement in the group by using a consensus measure based on the distance between the individual weights and the collective ones and ensures the achievement of a desired level of consensus ε_0 (García-Zamora *et al.*, 2023a). Note that this consensus measure could be replaced by any other with a similar interpretation. For instance, to avoid compensatory behaviours, we might define $\hat{\kappa} : \mathbb{P}_U^m \to \mathbb{R}_0^+$ by

$$\hat{\kappa}(u_1, \ldots, u_m) = \max_{\substack{i=1,\ldots,m \\ k=1,\ldots,n}} \big\{ |u_{ki} - F_i(u_1, \ldots, u_m)| \big\} \quad \forall \, (u_1, \ldots, u_m) \in \mathbb{P}_U^m.$$

On the other hand, due to the transformation to the $0 - 1$ linear scale, the distance between the weights that the method aims to obtain, and the original preferences elicited from the evaluators can be defined by $\xi : \mathcal{M}_{m,n} \times \mathcal{M}_{m,n} \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{P}_U^m) \to \mathbb{R}_0^+$

$$\begin{aligned} \xi(BO, OW, B, W, u) = {} & \frac{1}{2mn} \sum_{k=1}^{m} \sum_{i=1}^{n} \big(\big| \hat{\Delta}_S^{-1}(a_{B^k i}) - 0.5 - u_{kB^k} + u_{ki} \big| \\ & + \big| \hat{\Delta}_S^{-1}(a_{iW^k}) - 0.5 - u_{ki} + u_{kW^k} \big| \big). \end{aligned} \tag{12}$$

Therefore, for fixed values BO, OW, B, W, the proposed consensual BWM (LC-BWM) is defined as

$$\begin{aligned} & \min_{u,g} \xi(BO, OW, B, W, u), \\ & \text{s.t.} \begin{cases} g_i = F_i(u_1, \ldots, u_m), \\ \frac{1}{mn} \sum_{k=1}^{m} \sum_{i=1}^{n} |u_{ki} - g_i| \leqslant \varepsilon_0, \\ \sum_{i=1}^{n} u_{ki} = 1, \quad \forall k = 1, 2, \ldots, m. \end{cases} \end{aligned} \tag{LC-BWM}$$

REMARK 3. Note that to obtain an individual version of the previous optimization model, the so-called L-BWM, it is sufficient to define

$$\begin{aligned} & \min_{u} \xi(BO, OW, B, W, u), \\ & \text{s.t.} \sum_{i=1}^{n} u_{ki} = 1, \quad \forall k = 1, 2, \ldots, m, \end{aligned} \tag{L-BWM}$$

where BO, OW are $1 \times n$ matrices and B, W are integer numbers.

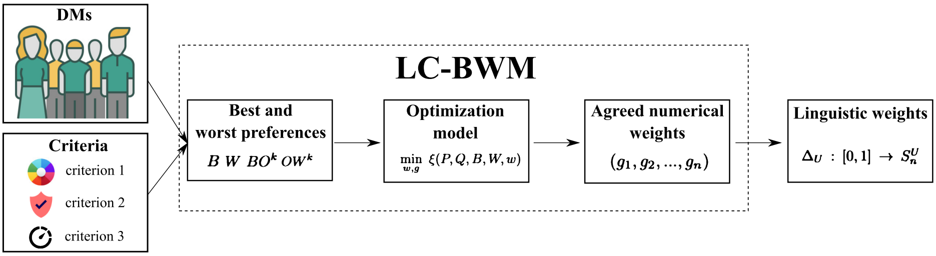

Fig. 6. LC-BWM scheme.

Here, the interpretation of the objective function ξ in both LC-BWM and L-BWM should be highlighted. This function measures the distance between the original evaluators' opinions and the output (weights) obtained from the models. In this sense, the function ξ can be seen as a consistency measure that allows measuring the magnitude of the changes applied to the original preferences. Since the goal of these models is minimizing these distance/consistency measures, the weights obtained after their resolution are guaranteed to be the ones that minimize the value of the consistency measure, or, in other words, the weights that are closest to the initial preferences and simultaneously satisfy the consensus condition.

The LC-BWM provides individual and collective values u, g, expressed in an additive scale, that represent as much as possible the BW preferences given by the evaluators, while also reaching the predefined consensus level and ε_0. It should be highlighted that LC-BWM assumes additive consistency to compute the values u, g. Therefore, to obtain weights expressed in a multiplicative scale, in the same way as traditional BWM, it is enough to transform the values:

$$w_{ki} = \frac{9^{2u_{ki}-1}}{\sum_{j=1}^{n} 9^{2u_{kj}-1}}, \quad k = 1, \ldots, m, \ i = 1, \ldots, n, \tag{13}$$

$$w_i = \frac{9^{2g_i-1}}{\sum_{j=1}^{n} 9^{2g_j-1}}, \quad i = 1, \ldots, n. \tag{14}$$

Finally, if w_i is the agreed numerical weight associated with the criterion C_i, its corresponding 2-tuple linguistic representation in S_n^U is given by $\hat{\Delta}_U(w_i)$. The scheme of the proposal is represented in Fig. 6.

3.3. *Consistency Index*

To quantify the consistency of the preferences provided by the evaluators, here we combine some statistical insight with the value of the function $\xi(BO, OW, B, W, u)$, which measures the distance between the original preferences given by the evaluators (BO, OW, B, W) and the output of the L-BWM u.

We will conduct several Montecarlo simulations to determine the percentiles of the random variable $\xi(BO, OW, B, W, u)$. Since the smaller the value $\xi(BO, OW, B, W, u)$,

Table 3
Percentiles for the consistency.

n	20%	40%	60%	80%	100%
4	0.00781	0.01563	0.03125	0.04688	0.11719
5	0.0125	0.025	0.03125	0.05	0.13125
6	0.01563	0.02604	0.03646	0.05208	0.13542
7	0.01786	0.02679	0.04018	0.05804	0.14286
8	0.01953	0.03125	0.04297	0.05859	0.14363
9	0.02083	0.03125	0.04514	0.0625	0.14536

the more consistent the results of the L-BWM, we can analyse the consistency of the preferences (BO, OW, B, W) by checking the percentile corresponding to the value $\xi(BO, OW, B, W, u)$.

First, note that for $n \in \mathbb{N}$, BWM preferences may be randomly generated as follows:

1. Assume that $B = 1$ and $W = n$.
2. Generate the pairwise comparison of the best and worst criteria $BO_n = OW_1$ as a random label in $[8, 16] \cap \mathbb{N}$.
3. Define $BO_1 = 8$, $OW_n = 8$.
4. Complete $BO, OW \in \mathbb{R}^n$ with random labels in $[8, BO_n]$.

Under these conditions, for the most common values of $n \in \mathbb{N}$, i.e. natural numbers between 4 and 9, we have generated a sample of 50000 linguistic BWM preferences and computed the corresponding weights. Then, for each BWM preference (BO, OW, B, W) and resulting output u, the objective value has been computed $X = \xi(BO, OW, B, W, u)$. Subsequently, the cumulative density function for X has been approximated using the function `ecdf` in the package StatsBase for Julia (Bezanson *et al.*, 2017). The obtained values for the percentiles are shown in Table 3, and allow classifying a BW preference according to the probability of obtaining a more consistent preference.

For instance, let us consider 5 criteria whose importance is evaluated from some BW preferences provided by evaluators. After solving the LC-BWM optimization model, the value of the objective function is $\xi(BO, OW, B, W, u) = 0.016$. Then, to determine the consistency index, we need to check the row that corresponds to 5 criteria ($n = 5$) in Table 3. According to the table, the first value in that row that is greater than 0.016 is 0.025, which corresponds to the percentile 40%. Therefore, the given preferences are allocated in the top 40% of the most consistent preferences for 5 criteria. Note that the acceptable values of the consistency index must be chosen according to the needs of the MCDM problem, taking into account that the lower the value $\xi(BO, OW, B, W, u)$, the higher the reliability of the preferences (BO, OW, B, W). On the contrary, preferences whose consistency index is closer to the percentiles $80\% - 100\%$, will show a poor level of consistency and, consequently, are not reliable. As a general rule, those preferences whose consistency index is allocated in the percentiles $20 - 40\%$, are reliable enough to generate reasonable results.

Therefore, we define an individual consistency index that measures how far the individual preferences are from the ideal consistency scenario, i.e. $\xi(BO, OW, B, W, u) = 0$.

Since our proposal aims to deal with LiMCGDM problems, the group consistency index is computed from the average of the individuals' consistency indices. The LC-BWM derives the additive weights minimizing the group consistency, and thus keeping as much as possible the initial evaluators' view, and satisfying the consensus condition to obtain an agreed solution. In any case, instead of the arithmetic mean, the maximum operator could be used for computing the group's consistency index and avoiding compensatory effects without impacting the structure of the model (García-Zamora *et al.*, 2023b; García-Zamora *et al.*, 2025).

4. Case Study

The food industry is a major contributor to global environmental challenges such as climate change, deforestation, and plastic pollution (D'Adamo, 2023; Prasanna *et al.*, 2025). One area where food companies can make a significant impact on sustainability is in their choice of packaging materials.

Packaging plays an important role in protecting food and extending its shelf life, but it also generates a significant amount of waste, particularly in the form of single-use plastics. In recent years, there has been a growing interest in sustainable packaging options, such as those made from biodegradable materials or those that are easily recyclable.

In this case study, a food company aims at reducing its environmental impact and is exploring new packaging options. The company's sustainability team has studied some potential packaging options and is seeking input from a diverse group of evaluators to help make a decision.

This case study includes 5 representatives from the packaging industry, environmental organizations, and the company's internal departments $E = \{E_1, E_2, E_3, E_4, E_5\}$. Each evaluator brings their own perspectives, values, and priorities to the decision-making process, highlighting the importance of a structured and transparent decision-making approach.

Before considering the new packaging options to reduce the environmental impact, the evaluators also have to identify the criteria to consider for evaluating such options. They have identified 4 main criteria $C = \{C_1 : \textit{recyclability},\ C_2 : \textit{cost},\ C_3 : \textit{energy usage},\ C_4 : \textit{carbon emissions}\}$. However, the evaluators have different views on the importance of each criterion, which led to differences in their evaluations. To ensure a unified judgment, the panel of members has agreed to participate in a CRP supervised by a moderator to achieve an agreement on the importance of each criterion. In this regard, the LC-BWM method will be used to support the moderator during the consensus process.

Following the BWM approach, the members of the panel first have to choose the best and worst criteria according to their expertise. In the LC-BWM, the vectors B and W, which contain the best and worst criteria selected by the evaluators, respectively, are described below:

$$B = (C_4, C_4, C_4, C_4, C_1), \qquad W = (C_3, C_2, C_3, C_3, C_3).$$

Table 4
Evaluators' weights.

Opinion	C_1	C_2	C_3	C_4
E_1	0.133	0.101	0.077	0.69
E_2	0.204	0.068	0.118	0.611
E_3	0.149	0.196	0.065	0.589
E_4	0.203	0.267	0.068	0.462
E_5	0.485	0.093	0.054	0.368

Once the members have chosen the best and worst criteria, they provide linguistic pairwise comparisons. The evaluators use the fuzzy linguistic term set S with $G = 16$ described in Section 3 to assess their pairwise comparisons, which are collected in the LC-BWM through the matrices BO, which contains the comparisons between the best criteria and the remainder, and OW, which contains the comparisons between the remainder criteria and the worst ones as follows:

$$BO = \begin{array}{c} \\ B^1 \\ B^2 \\ B^3 \\ B^4 \\ B^5 \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} \\ \begin{pmatrix} (DMI,0) & (VSMI,0) & (EMI,0) & (EI,0) \\ (SMI,0) & (EMI,0) & (DMI,0) & (EI,0) \\ (SPMI,0) & (SMI,0) & (EMI,0) & (EI,0) \\ (MPMI,0) & (MMI,0) & (VSMI,0) & (EI,0) \\ (EI,0) & (SMI,0) & (EMI,0) & (WMI,0) \end{pmatrix} \end{array},$$

$$OW = \begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \\ C_4 \end{array} \begin{array}{c} \begin{array}{ccccc} W^1 & W^2 & W^3 & W^4 & W^5 \end{array} \\ \begin{pmatrix} (MMI,0) & (SMI,0) & (MPMI,0) & (SMI,0) & (EMI,0) \\ (WMI,0) & (EI,0) & (SMI,0) & (SPMI,0) & (MMI,0) \\ (EI,0) & (MPMI,0) & (EI,0) & (EI,0) & (EI,0) \\ (EMI,0) & (EMI,0) & (EMI,0) & (VSMI,0) & (VSMI,0) \end{pmatrix} \end{array}.$$

The moderator can also compute the consistency index of these preferences by using the LC-BWM to check if they are reasonable. This is key because if the preferences are not consistent (contradictory/random), the results are not reliable, and the whole process may be compromised. In this regard, all of them are allocated in the top 40% of the most consistent preferences. In addition, note that we have transformed the single linguistic terms provided by the evaluators into 2-tuple linguistic values by adding a symbolic translation equal to 0 (see Remark 1). This is necessary to apply the LC-BWM later on.

Before carrying out a CRP, let us use the L-BWM model to compute the importance of the criteria according to each evaluator's opinion (see Table 4). The computation of the consensus degree of such preferences (0.93) evidences that there are some disagreements in the evaluators' opinions, a CRP is initiated to increase the level of agreement in the group. To easily identify such disagreements and provide precise suggestions to reach a consensus faster, the CRP is first theoretically computed by the LC-BWM.

Table 5
LC-BWM consensual weights.

Opinion	C_1	C_2	C_3	C_4
E_1	0.133	0.101	0.077	0.69
E_2	0.204	0.068	0.118	0.611
E_3	**0.168** (+0.019)	**0.192** (−0.004)	**0.064** (−0.001)	**0.576** (−0.013)
E_4	**0.233** (+0.03)	**0.16** (−0.107)	**0.078** (+0.01)	**0.53** (+0.068)
E_5	**0.38** (−0.105)	**0.127** (+0.034)	**0.063** (+0.009)	**0.43** (+0.062)
Group	0.216	0.126	0.08	0.578
Linguistic	$(AI, 0.177)$	$(AI, -0.246)$	$(U, 0.479)$	$(I, 0.482)$

To apply the LC-BWM, we transform the linguistic preferences into a linear scale in [0, 1] by using the mapping $\hat{\Delta}_S^{-1}$:

$$\hat{\Delta}_S^{-1}(BO) = \begin{array}{c} \\ B^1 \\ B^2 \\ B^3 \\ B^4 \\ B^5 \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} \\ \begin{pmatrix} 0.875 & 0.9375 & 1.0 & 0.5 \\ 0.75 & 1.0 & 0.875 & 0.5 \\ 0.8125 & 0.75 & 1.0 & 0.5 \\ 0.6875 & 0.625 & 0.9375 & 0.5 \\ 0.5 & 0.75 & 1.0 & 0.5625 \end{pmatrix} \end{array},$$

$$\hat{\Delta}_S^{-1}(OW) = \begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \\ C_4 \end{array} \begin{array}{c} \begin{array}{ccccc} W^1 & W^2 & W^3 & W^4 & W^5 \end{array} \\ \begin{pmatrix} 0.625 & 0.75 & 0.6875 & 0.75 & 1.0 \\ 0.5625 & 0.5 & 0.75 & 0.8125 & 0.625 \\ 0.5 & 0.6875 & 0.5 & 0.5 & 0.5 \\ 1.0 & 1.0 & 1.0 & 0.9375 & 0.9375 \end{pmatrix} \end{array}.$$

From the previous numerical matrices and the vectors B and W, we apply the LC-BWM optimization model considering $\varepsilon_0 = 0.05$ to obtain the agreed modified evaluators' opinions, and the agreed individual and group weights, which are expressed in a multiplicative scale by using Eq. (13) (see Table 5).

Note that the values of the modified preferences may be obtained from the modified individual weights by:

$$\begin{aligned} \overline{BO}_i^k &= \hat{\Delta}_S\left(\frac{1}{2} + u_{kB^k} - u_{ki}\right), \\ \overline{WO}_i^k &= \hat{\Delta}_S\left(\frac{1}{2} + u_{ki} - u_{kW^k}\right). \end{aligned} \tag{15}$$

Thus, we can obtain the feedback for the evaluators whose opinions are represented in $\overline{BO}$ and $\overline{OW}$, in which changes with respect to their initial opinions have been highlighted in bold.

$$\overline{BO} = \begin{array}{c} \\ B^1 \\ B^2 \\ B^3 \\ B^4 \\ B^5 \end{array} \begin{array}{c} \begin{array}{cccc} \quad C_1 \quad & \quad C_2 \quad & \quad C_3 \quad & \quad C_4 \quad \end{array} \\ \begin{pmatrix} (DMI, 0) & (VSMI, 0) & (EMI, 0) & (EI, 0) \\ (SMI, 0) & (EMI, 0) & (DMI, 0) & (EI, 0) \\ (\mathbf{SMI}, \mathbf{0.488}) & (SMI, 0) & (EMI, 0) & (EI, 0) \\ (MPMI, 0) & (\mathbf{SMI}, \mathbf{0.363}) & (VSMI, 0) & (EI, 0) \\ (\mathbf{EI}, \mathbf{0.45}) & (\mathbf{SMI}, \mathbf{0.45}) & (\mathbf{VSMI}, \mathbf{0}) & (\mathbf{EI}, \mathbf{0}) \end{pmatrix} \end{array},$$

$$\overline{OW} = \begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \\ C_4 \end{array} \begin{array}{c} \begin{array}{ccccc} \quad W^1 \quad & \quad W^2 \quad & \quad W^3 \quad & \quad W^4 \quad & \quad W^5 \quad \end{array} \\ \begin{pmatrix} (MMI, 0) & (SMI, 0) & (\mathbf{SMI}, -\mathbf{0.488}) & (SMI, 0) & (\mathbf{VSMI}, -\mathbf{0.45}) \\ (WMI, 0) & (EI, 0) & (SMI, 0) & (\mathbf{MPMI}, -\mathbf{0.362}) & (\mathbf{MPMI}, -\mathbf{0.45}) \\ (EI, 0) & (\mathbf{MMI}, \mathbf{0}) & (EI, 0) & (EI, 0) & (EI, 0) \\ (EMI, 0) & (EMI, 0) & (EMI, 0) & (VSMI, 0) & (VSMI, 0) \end{pmatrix} \end{array}.$$

Note that most of the evaluators' preferences remain the same. LC-BWM suggests that E_3 and E_4 should change their view regarding the importance of C_1 and C_2, respectively. The highest number of changes is recommended for E_5. It is suggested that all preferences of this evaluator's regarding the best-to-other comparison change slightly (no change involves more than one linguistic label away) since this evaluator's opinion is the furthest away from the rest. These changes are reflected in the resulting weights shown in Table 5.

Table 5 shows the optimal theoretical weights for the problem that allow achieving the desired level of consensus. In this table, we can see even more clearly that the changes suggested by the LC-BWM imply minimal changes in the weights obtained from the evaluators' initial preferences. The most significant change occurs in the weight given by E_4 to C_2 (-40%).

In addition, the agreed group solution identifies C_4 (carbon emissions) as the most important criterion and C_3 (energy usage) as the least important one (see the row Group in Table 5). To facilitate the understanding of the results from the evaluators, the numerically agreed weights can be retranslated into a linguistic representation (see the row Linguistic in Table 5). Notice that the most important criterion C_4 is labelled with "important", whereas the least important criterion C_3 is labelled as "unimportant". This linguistic representation can be useful for both moderators and evaluators, as it allows them to immediately identify the importance of a criterion by using linguistic terms that people often use to express degrees of importance.

Therefore, this information can be used by the moderator to provide suggestions in the following round of the CRP and achieve the desired level of consensus efficiently. Notice that the changes suggested by the LC-BWM are quite slight since the model finds a solution that changes the initial preferences of the evaluators as little as possible. This is key for the use of the LC-BWM in CRPs with feedback, since it is expected that the evaluators will be willing to modify their opinions if these changes are minimal.

5. Numerical Analysis

For all experiments reported in this manuscript, we have used JuMP (Julia for Mathematical Programming), a domain-specific modelling language for mathematical optimization embedded in Julia (Dunning *et al.*, 2017). Specifically, optimization experiments are conducted in Julia 1.12.6 on a laptop with Windows 11 Professional OS, 1.4 GHz Intel Core Ultra i7-155H CPU, and 32 GB RAM by invoking the Gurobi 9.0.3 optimizer.

5.1. *Computational Scalability Analysis*

The case study presented in the previous section was intentionally limited to five evaluators and four criteria in order to facilitate the explanation of the proposed LC-BWM model and allow the reader to clearly follow each stage of the decision-making process. Although such a configuration is representative of many real-world group decision-making problems, it does not fully illustrate the computational capabilities of the proposed approach when dealing with larger problem instances.

Therefore, an additional computational scalability analysis was conducted to evaluate the performance of the model under increasingly demanding conditions. Specifically, the number of evaluators was progressively increased from 5 to 100 evaluators. This allows us to assess whether the proposed formulation remains computationally efficient when applied to large-scale group decision-making scenarios involving a substantial number of participants.

Regarding the number of criteria, the experiments were restricted to scenarios involving 3, 5, 7, and 9 criteria. This decision is motivated by the cognitive limitations associated with pairwise comparison-based preference elicitation methods. According to Miller's well-known "seven plus or minus two" principle (Miller, 1956), individuals have a limited capacity to simultaneously process information, and several studies have highlighted that excessively large sets of elements to compare may impose a significant cognitive burden on evaluators (Goodwin and Wright, 2014; Rezaei, 2015; Saaty, 1990). Consequently, considering more than nine criteria would be less representative of realistic applications of the BWM framework and similar pairwise comparison approaches.

To perform this analysis, consistent synthetic BWM preference relations were automatically generated for each combination of evaluators and criteria. The execution time required to solve each instance was then recorded. Table 6 summarizes the obtained results. As expected, the execution time increases with both the number of evaluators and the number of criteria. Nevertheless, even for the largest tested scenario involving 100 evaluators and 9 criteria, the proposed LC-BWM model required less than one second to obtain the optimal solution. These results suggest that the computational burden introduced by the consensus mechanism remains moderate and that the proposed formulation is suitable for large-scale group decision-making problems.

5.2. *Comparative Analysis*

This section compares the LC-BWM approach with other existing group BWM approaches. The comparison focuses on G-BWM and MCC-BWM because these approaches

Table 6
Computational scalability analysis of the proposed LC-BWM model.

Evaluators	Criteria	Execution time (s)
5	3	0.0031
	5	0.0027
	7	0.0037
	9	0.0044
10	3	0.0022
	5	0.0062
	7	0.0087
	9	0.0135
20	3	0.0048
	5	0.0127
	7	0.0228
	9	0.0445
50	3	0.0166
	5	0.0377
	7	0.1258
	9	0.1324
100	3	0.0680
	5	0.1109
	7	0.2496
	9	0.5481

share the same objective of deriving criteria weights from BWM preference structures in group decision-making environments. Moreover, all three methods operate on the same type of preference information, which allows a direct and fair comparison of their weighting and consensus mechanisms. In contrast, many well-known MCDM methods either address different decision-making tasks or require alternative preference representations. Consequently, comparing the proposed LC-BWM with such methods would require transforming the original BWM preference information into different formats. This transformation process may introduce information loss or distort the original evaluators' assessments, potentially affecting the validity and fairness of the comparison. Among criteria-weighting approaches, AHP is one of the most widely adopted methods. However, AHP relies on a different preference elicitation process based on complete pairwise comparison matrices. In fact, one of the main motivations behind the development of the BWM was to reduce the number of required comparisons and the associated cognitive burden while maintaining satisfactory consistency levels. These advantages have already been extensively discussed and validated in the BWM literature. Therefore, the objective of this study is not to reassess BWM against AHP, but rather to analyse how consensus mechanisms can be incorporated into the BWM framework. For this reason, G-BWM and MCC-BWM constitute the most appropriate benchmark methods for evaluating the contribution of the proposed consensus-based BWM approach.

To do this comparison, we first define default BW preferences that are then used to run the different proposals based on the BWM. Note that the proposals sometimes present

Table 7
LC-BWM results.

Opinion	C_1	C_2	C_3	C_4	C_5	Consistency (ξ)
E_1	0.466723	0.143833	0.155574	0.182011	0.051858	0.066964
E_2	0.111957	0.121096	0.193915	0.209745	0.363288	0.100893
E_3	0.087721	0.066654	0.599885	0.158018	0.087721	0.048214
Group	0.19917	0.126015	0.314793	0.218265	0.141757	0.072024

different characteristics, so it will be necessary to perform some operations on the preferences and optimization models to make a fair comparison.

In this comparative analysis, for the sake of simplicity, we consider 3 evaluators that provide the following BW preferences $B = (C_1, C_5, C_3)$, $W = (C_5, C_1, C_2)$

$$BO = \begin{array}{c} \\ B^1 \\ B^2 \\ B^3 \end{array} \begin{array}{c} \begin{array}{ccccc} C_1 & C_2 & C_3 & C_4 & C_5 \end{array} \\ \begin{pmatrix} (EI,0) & (MMI,0) & (SMI,0) & (DMI,0) & (EMI,0) \\ (EMI,0) & (SMI,0) & (DMI,0) & (MMI,0) & (EI,0) \\ (VSMI,0) & (EMI,0) & (EI,0) & (WMI,0) & (VSMI,0) \end{pmatrix} \end{array},$$

$$OW = \begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \end{array} \begin{array}{c} \begin{array}{ccc} W^1 & W^2 & W^3 \end{array} \\ \begin{pmatrix} (EMI,0) & (EI,0) & (WMI,0) \\ (DMI,0) & (SMI,0) & (EI,0) \\ (SMI,0) & (MMI,0) & (EMI,0) \\ (WMI,0) & (WMI,0) & (VSMI,0) \\ (EI,0) & (EMI,0) & (WMI,0) \end{pmatrix} \end{array}.$$

Initially, we perform the LC-BWM, obtaining the results shown in Table 7.

5.2.1. *Classic BWM with Minimum Cost Consensus*

First, the LC-BWM is compared with adapted versions of the BWM extensions introduced in Licerán-Gutiérrez *et al.* (2022); Moreno-Albarracín *et al.* (2020). These MCC-BWM approaches also deal with MCGDM problems and the conflicts among evaluators, as the LC-BWM, by following a common process consisting of two steps:

1. A BWM optimization model is applied for each evaluator to obtain his/her individual weights.
2. An optimization model based on the notion of MCC (Ben-Arieh *et al.*, 2008) is accomplished over the individual evaluators' weights to obtain the collectively agreed weights. The MCC model aims at minimizing the cost of changing the individual evaluators' weights to obtain an agreed solution.

Therefore, the main difference between these proposals and the one presented here is that, whereas the LC-BWM optimization model returns both the evaluators' weights and the collectively agreed weights by using one single model, these proposals carry out two

Table 8
Individual weights obtained from L-BWM.

Opinion	C_1	C_2	C_3	C_4	C_5	Consistency (ξ)
E_1	0.451621	0.260744	0.15054	0.086915	0.05018	0.00625
E_2	0.05018	0.15054	0.086915	0.260744	0.451621	0.03125
E_3	0.067593	0.051359	0.462234	0.351222	0.067593	0.0
Group	0.156044	0.171004	0.246631	0.270276	0.156044	0.0125

Table 9
Consensual weights obtained from MCC.

Opinion	C_1	C_2	C_3	C_4	C_5	Consistency (ξ)
E_1	0.133218	0.287439	0.165953	0.358072	0.055318	0.21625
E_2	0.05018	0.15054	0.086915	0.260744	0.451621	0.03125
E_3	0.077114	0.086069	0.359007	0.400696	0.077114	0.0875
Group	0.092474	0.178769	0.199528	0.385724	0.143505	0.111667

separate processes. This difference will have a direct influence on the results, which will be discussed in Section 5.2.3.

To compare our proposal with these approaches, first, the L-BWM is applied to individual opinions to obtain individual weights. The results are shown in Table 8. However, the consensus degree of these values is $\varepsilon = 0.17$, which proves there are disagreements among the evaluators. To smooth such disagreements until an acceptable level of agreement is reached ($\varepsilon_0 = 0.1$) and following the process introduced in Licerán-Gutiérrez *et al.* (2022); Moreno-Albarracín *et al.* (2020), an MCC optimization model is applied to the individual weights to obtain collectively agreed weights. The results are shown in Table 9.

5.2.2. *Group Best-Worst Method*

This section compares the LC-BWM approach with the so-called G-BWM introduced in Safarzadeh *et al.* (2018). Such BWM extension deals with MCGDM problems and provides collective weights for the group of evaluators. To do so, the G-BWM initially assumes common best and worst criteria for all the evaluators. Afterward, the authors provide two different nonlinear optimization models to derive the weights. Whereas the former minimizes the summation of the consistency deviation for all the evaluators, and according to the original BWM, the latter uses a min-max objective (Safarzadeh *et al.*, 2018).

Below, we evaluate the performance of the G-BWM. To do a fair comparison, we adapt the G-BWM to deal with additive preferences, as the LC-BWM, obtaining the results shown in Table 10. Notice the individual evaluators' weights are not derived by the G-BWM, and they have been noted in Table 10 as "–". Furthermore, we have computed the consistency indexes of the individual and collective preferences by using the measure of the LC-BWM to properly compare both approaches.

Finally, we have also derived the level of agreement of the G-BWM solution. Despite it not being possible to know the individual weights from the G-BWM, we have computed

Table 10
Group weights obtained from G-BWM.

Opinion	C_1	C_2	C_3	C_4	C_5	Consistency (ξ)
E_1	–	–	–	–	–	0.24375
E_2	–	–	–	–	–	0.24375
E_3	–	–	–	–	–	0.1125
Group	0.1625	0.0375	0.4125	0.225	0.1625	0.2

Table 11
Comparison among BWM approaches.

Features	LC-BWM	MCC-BWM	G-BWM
Consensus degree (ε)	0.1	0.1	0.15
Group consistency (ξ)	0.072024	0.111667	0.2
Multiple preference structures	✓	✗	✗
No need to fix best and worst criteria	✓	✓	✗
Provides individual modified weights	✓	✓	✗
Linear optimization	✓	✗	✗

the consensus degree by comparing the individual weights obtained from the LC-BWM (see Table 8) and the collective weights obtained from the G-BWM, resulting in a consensus degree $\varepsilon = 0.15$ (as in the previous section, we have considered as acceptable level of agreement $\varepsilon_0 = 0.1$). Therefore, some disagreements are ignored by the G-BWM and, consequently, some evaluators may not be satisfied with the solution.

5.2.3. *Comparison*

This section provides a quantitative and qualitative analysis of the performance of the three proposals mentioned above, namely, our proposed LC-BWM, the MCC-BWM (Licerán-Gutiérrez *et al.*, 2022; Moreno-Albarracín *et al.*, 2020), and the G-BWM (Safarzadeh *et al.*, 2018) (see Table 11).

Regarding the consensus of the group, the G-BWM does not consider any consensus threshold in the optimization model, which leads to a low consensus degree in the final solution ($\varepsilon = 0.15$). On the contrary, both LC-BWM and MCC-BWM reach the consensus degree $\varepsilon = 0.1$, resulting in an acceptable consensus level for the final weights.

Another remarkable difference among the analysed approaches is related to the consistency of the obtained solutions. Recall that the function ξ measures the deviation between the evaluators' original preferences and the weights returned by the optimization model. As shown in Table 11, the proposed LC-BWM achieves the lowest consistency deviation ($\xi = 0.072024$), outperforming both MCC-BWM ($\xi = 0.111667$) and G-BWM ($\xi = 0.2$). This indicates that the collective solution obtained by the LC-BWM remains closer to the evaluators' original opinions while still satisfying the desired consensus requirements. These results suggest that incorporating consensus requirements directly into the weighting process may provide a more effective balance between consensus attainment and preference preservation than the sequential strategies adopted by existing approaches.

Table 12
Preferences simulation experiments.

Scenario	Opinion	C_1	C_2	C_3	C_4	C_5	Consistency (ξ)
	E_1	0.466723	0.143833	0.155574	0.182011	0.051858	0.066964
	E_2	0.111957	0.121096	0.193915	0.209745	0.363288	0.100893
Original	E_3	0.087721	0.066654	0.599885	0.158018	0.087721	0.048214
	Group	0.199170	0.126015	0.314793	0.218265	0.141757	0.072024
	Importance	3	5	1	2	4	–
	E_1	0.502946	0.127385	0.167649	0.146137	0.055883	0.025000
	E_2	0.123705	0.115496	0.214264	0.200046	0.346489	0.110938
1	E_3	0.086316	0.065586	0.590272	0.171511	0.086316	0.043750
	Group	0.208340	0.117563	0.329286	0.203625	0.141186	0.059896
	Importance	2	5	1	3	4	–
	E_1	0.34202	0.197466	0.25988	0.114007	0.086627	0.068750
	E_2	0.087813	0.142655	0.152096	0.189471	0.427965	0.073125
2	E_3	0.096050	0.072982	0.39699	0.301647	0.13233	0.053125
	Group	0.162401	0.145061	0.285611	0.213080	0.193846	0.065000
	Importance	4	5	1	2	3	–
	E_1	0.414518	0.19669	0.138173	0.204561	0.046058	0.058036
	E_2	0.093169	0.130051	0.161374	0.225254	0.390152	0.083482
3	E_3	0.122086	0.070486	0.482021	0.203322	0.122086	0.045536
	Group	0.197084	0.143051	0.259377	0.247771	0.152718	0.062351
	Importance	3	5	1	2	4	–

5.3. *Sensitivity Analysis*

This section provides a sensitivity analysis to provide the robustness and validity of the proposal. To do so, we carry out two different analyses. The former is regarding the LC-BWM behaviour when the initial evaluators' preferences are changed to a greater or lesser extent. The latter analyses the results obtained from the LC-BWM, considering different values of ε_0.

5.3.1. *Dependency with Respect to Original Opinions*

Here, we present several heterogeneous scenarios that introduce slight modifications to the evaluators' preferences, and afterward, analyse the impact of such changes on the final recommendations. Keeping in mind the preferences introduced in Section 5.2, we consider 3 scenarios:

- Scenario 1: E_1 changes his/her preferences over C_2 so that $a_{B^12} = SPMI$ and $a_{2W^1} = MPMI$.
- Scenario 2: E_1 and E_3 change their preferences over C_5 so that $a_{B^15} = SPMI$ and $a_{B^35} = SMI$.
- Scenario 3: E_3 changes all his/her preferences over all the criteria so that $BO^3 = (DMI, VSMI, EI, EI, DMI)$ and $OW^3 = (MMI, EI, VSMI, DMI, MMI)$.

Table 12 shows the results obtained for each scenario, considering $\varepsilon_0 = 0.1$ and the ones obtained from the original preferences shown in Section 5.2.

Scenario 1 evaluates a slight change in the E_1's preferences. In this scenario, the evaluator decreases the relative importance of C_2. Both the original scenario and Scenario 1 set C_2 as the least important, but with different weights. In Scenario 1, the C_2 group weight, originally 0.126015, decreases to 0.117563, which seems logical, taking into account the changes. Notice that, due to the modifications in E_1's preferences, the relative importance of C_2 regarding the best criterion (C_1) has been changed. Consequently, the latter has increased its relative importance, and it is also reflected in the group's weight, which has increased to 0.20834. Certainly, such a change has provoked an exchange of positions in the ranking between C_1 and C_4. However, they originally were very close to each other (0.19917 and 0.218265), and a small variation could lead to such an exchange of positions. This situation happens in this scenario, despite the resulting weights of both criteria being quite similar to the original one (C_1:0.199170/0.208340 and C_4:0.218265/0.203625), which is the expected behaviour since a slight modification in the evaluators' preferences should not provoke significant changes in the criteria weights.

Scenario 2 simulates two changes over the BO preferences associated with the criterion C_5 for all the evaluators. E_1 and E_3 increase the C_5 relative importance. As in the previous scenario, such an increment can be appreciated in the resulting weights. C_5 goes from having a weight of 0.141757 to 0.193846. Notice that the consistency index of E_2 has decreased. The reason behind this is that, in the initial preferences, E_1 and E_3 evaluate C_5 with quite low importance. On the contrary, E_2 considers C_5 to be the most important criterion. By increasing the relative importance of C_5 from E_1 and E_3, the opinions between all the evaluators are closer to each other, which implies that E_2 does not need to change that much his/her initial opinions to reach a consensus. In addition, such variation over the relative importance of C_5 has led to a change in the ranking, moving up one position. Again, the LC-BWM presents a logical behaviour regarding the applied changes.

The last scenario changes all the E_3's preferences. They are not extreme changes. For BO^3, E_3' preference has been changed to the label immediately preceding the one used in the original scenario, except for those criteria that were evaluated with an s_8. For the OW^3, the preferences have been changed by the labels immediately following the ones used in the original scenario, except for those criteria that were evaluated with an s_8. The changes have been applied in this way to obtain meaningful and non-random preferences. In this scenario, the evaluator has increased the relative importance of all criteria with respect to the most important one, in this case, C_3. If we look at E_3's original preferences, the relative importance of C_3 with respect to the others is quite strong, except with C_4. This can also be appreciated in the original weights, in which C_3 has been assigned a weight close to 0.6. However, Table 12 shows that such importance is now a little more balanced. The weight of C_3 has decreased to 0.482021, and the remaining weights have increased. In addition, the consistency index associated with E_3' preferences has not been significantly affected, since although the preferences have undergone several changes, they have not implied a substantial modification in the relative importance assigned to the criteria in the original scenario. We see again how the model behaves logically to changes in evaluators' preferences.

Therefore, the LC-BWM presents a reliable and robust solution according to the results obtained from the different scenarios. Furthermore, we would like to remark that the values

Table 13
Consensus degree simulation.

ε_0	Opinion	C_1	C_2	C_3	C_4	C_5	Consistency (ξ)
	E_1	0.466723	0.143833	0.155574	0.182011	0.051858	0.066964
	E_2	0.111957	0.121096	0.193915	0.209745	0.363288	0.100893
0.1	E_3	0.087721	0.066654	0.599885	0.158018	0.087721	0.048214
	Group	0.199170	0.126015	0.314793	0.218265	0.141757	0.072024
	Importance	3	5	1	2	4	–
	E_1	0.408551	0.113398	0.235877	0.163548	0.078626	0.110417
	E_2	0.138119	0.126035	0.239229	0.181774	0.314843	0.125000
0.05	E_3	0.162993	0.123848	0.371545	0.17862	0.162993	0.145833
	Group	0.222919	0.128702	0.293377	0.185621	0.169381	0.127083
	Importance	2	5	1	3	4	–
	E_1	0.145066	0.083754	0.435197	0.190917	0.145066	0.243750
	E_2	0.145066	0.083754	0.435197	0.190917	0.145066	0.243750
0	E_3	0.145066	0.083754	0.435197	0.190917	0.145066	0.112500
	Group	0.145066	0.083754	0.435197	0.190917	0.145066	0.200000
	Importance	3	5	1	2	4	–

of the consistency indexes are generally low, which is even more evidence of the model's good performance.

5.3.2. *Dependency with Respect to the Consensus Degree*

This section analyses the results obtained from the LC-BWM for different values of ε_0. Taking as an example the problem presented in Section 5, where the initial group consensus is $\varepsilon = 0.17$, Table 13 shows the results obtained for the different values of ε_0.

From this table, we can draw several conclusions. The smaller the ε_0 value, the higher the values obtained from the consistency index. This is logical, taking into account that a lower value of ε_0 implies reaching a higher consensus and, consequently, modifying to a greater extent the initial evaluations' opinions. This can also be appreciated in the corresponding individual weights. The higher the degree of consensus to be reached, the more similar the opinions of the evaluators should be, and thus the more similar the individual weights of the criteria.

Notice that, despite the scenario with $\varepsilon_0 = 0$, which represents unanimity, requiring many changes in the evaluators' original opinions, the importance of the criteria varies slightly in all the cases. The only difference is regarding C_1 and C_4, which exchange their positions in the ranking with $\varepsilon_0 = 0.05$ in comparison with the remaining scenarios. This is possible because the optimization model obtains a solution that involves as few changes as possible in the initial preferences while minimizing the consistency measure ξ, thus the modifications provided by the model do not involve abrupt variations in the ranking of importance of the criteria, which is, in fact, a desired behaviour. In this sense, the LC-BWM presents a robust consensus solution, where in all cases the selection of the most and least important criterion is clear, which can make it easier for the moderator or analyst of a CRP with a feedback process to provide suggestions to the evaluators and reach an agreement.

5.3.3. *Insights from the Sensitivity Analysis*

The results obtained from the different sensitivity analyses provide additional evidence of the robustness of the proposed LC-BWM approach. Regarding the modifications introduced in the evaluators' preferences, the resulting collective weights evolved coherently with the changes applied to the input information. In all the analysed scenarios, the model exhibited a predictable behaviour, assigning higher importance to those criteria whose relative relevance was increased by the evaluators and reducing the importance of those criteria whose assessments became less favourable.

Another remarkable aspect is the stability of the obtained rankings. Although some exchanges of positions were observed among criteria with very similar importance values, the overall structure of the rankings remained largely unchanged across the analysed scenarios. In particular, the most and least important criteria were consistently identified in almost all cases, which suggests that the proposed methodology is capable of producing reliable recommendations despite moderate variations in the input preferences.

The analysis of different consensus thresholds also highlights an important property of the LC-BWM. As expected, lower values of ε_0 require greater modifications of the evaluators' original opinions to achieve higher levels of agreement. Nevertheless, even under strict consensus requirements, the resulting collective weights remained reasonably close to those obtained in the original scenario. This behaviour indicates that the proposed optimization model successfully balances the two conflicting objectives of preserving the evaluators' initial views and achieving a desired level of consensus.

Therefore, the sensitivity analyses confirm that the proposed methodology behaves in a stable and consistent manner under different conditions, supporting its applicability as a consensus support tool in LiMCGDM problems.

5.4. *Discussion*

5.4.1. *Comparison with Existing Approaches*

The comparative analysis reveals several conceptual and practical differences between the proposed LC-BWM and existing BWM-based approaches. Although all the analysed methods aim to derive collective criteria weights in group decision-making environments, they follow substantially different strategies to address disagreements among evaluators.

The most important distinction concerns the way consensus is incorporated into the weighting process. Traditional MCC-BWM approaches obtain individual weights in a first stage and subsequently apply a consensus optimization model to modify those weights until an acceptable agreement level is achieved. In contrast, the proposed LC-BWM integrates both objectives into a single optimization framework. As a consequence, consensus requirements are considered from the beginning of the weighting process, allowing the model to simultaneously derive individual and collective consensual weights. This integrated formulation contributes to obtaining collective solutions that remain closer to the evaluators' original preferences while satisfying the desired consensus requirements.

Another relevant advantage of the LC-BWM is its flexibility regarding preference representation. Unlike G-BWM, which assumes common best and worst criteria for all evaluators, the proposed approach allows each evaluator to independently identify the criteria that best reflect his or her personal perspective. This characteristic increases the applicability of the model in real-world group decision-making problems, where evaluators frequently have heterogeneous viewpoints and priorities. Furthermore, LC-BWM can operate with linguistic, additive, and multiplicative preference structures, providing additional flexibility for preference elicitation.

Finally, the proposed methodology benefits from a linear optimization formulation. This property simplifies the computational resolution of the problem and contributes to the scalability of the approach, as evidenced by the computational experiments presented in Section 5.1. The ability to efficiently manage large numbers of evaluators while preserving consensus and consistency requirements makes the proposed approach particularly suitable for complex group decision-making environments.

Therefore, the obtained results suggest that the proposed LC-BWM provides a more flexible and integrated framework for consensus-based criteria weighting than existing BWM extensions, while maintaining high levels of consistency, interpretability, and computational efficiency.

5.4.2. *Practical Implications*

Beyond its methodological contributions, the proposed LC-BWM may provide practical benefits for moderators and decision analysts involved in CRPs. One of its main advantages is its ability to identify the minimum modifications required in the evaluators' preferences to achieve a desired level of agreement. This information can be used to generate targeted feedback recommendations, helping moderators focus their efforts on the most critical disagreements instead of manually analysing all the assessments provided by the evaluators.

Another relevant implication concerns the interpretability of the results. The use of the 2-tuple linguistic model allows evaluators to express their preferences through linguistic terms that are closer to natural human reasoning than purely numerical assessments. Furthermore, the resulting criteria weights can also be represented linguistically, facilitating the communication of the final collective solution and improving its transparency and understandability for non-technical participants.

The proposed methodology may be particularly useful in decision-making contexts involving heterogeneous stakeholders with different backgrounds, interests, and priorities. Examples include sustainability assessment, public policy design, strategic planning, supplier selection, and other group decision-making scenarios in which reaching consensus is often as important as obtaining an accurate weighting of the criteria.

Finally, the computational efficiency observed in the scalability analysis suggests that the proposed approach can be applied not only to small decision panels but also to larger groups of evaluators without introducing significant computational burdens. This characteristic increases the practical applicability of the model in real-world consensus-based decision-making environments.

5.4.3. *Limitations*

Despite the promising results obtained throughout the experimental analyses, several limitations of the proposed LC-BWM approach should be acknowledged. First, the methodology provides a theoretically optimal consensual solution according to the specified consensus requirements and optimization objectives. However, the model cannot guarantee that evaluators will accept the suggested modifications to their preferences during a real CRP. Consequently, LC-BWM should be viewed as a decision support tool that assists moderators and evaluators rather than as a mechanism capable of automatically enforcing consensus.

A second limitation concerns the representation of linguistic information. The proposed approach relies on predefined linguistic semantics derived from the 2-tuple linguistic model. Although this framework has been widely adopted in the literature, the correspondence between linguistic expressions and their computational representation may not always perfectly reflect the individual interpretation of each evaluator. Therefore, some degree of semantic mismatch may exist between the intended meaning of a linguistic assessment and its mathematical representation.

Finally, although the proposal has been validated through a case study, comparative analyses, sensitivity analyses, and computational scalability experiments, additional empirical applications involving different domains and decision-making contexts would provide further evidence regarding its practical applicability and generalizability. Future studies involving real CRPs could offer valuable insights into the interaction between the theoretical recommendations generated by the model and the actual behaviour of evaluators.

6. Conclusions

This study proposes a novel extension of the BWM based on the fuzzy linguistic approach that uses the 2-tuple linguistic model to derive agreed criteria weights in LiMCGDM problems. Evaluators express pairwise comparisons through a linguistic scale, enabling the modelling of uncertainty in their assessments. A novel BWM optimization model has been presented to derive consensual individual and group weights by preserving as much as possible the initial evaluators' views. The resulting weights are represented both numerically and linguistically to enhance their interpretability. In addition, we have provided a novel consistency index based on a cumulative density function that evaluates the consistency in evaluators' preferences according to the probability of finding more consistent values. Finally, we have proved the feasibility of the model by solving a real-world case study and comparing the performance of the proposal with other versions of BWM for groups.

Beyond extending the BWM to a linguistic group decision-making environment, the main novelty of the proposed LC-BWM lies in the integration of consensus requirements directly into the weighting process. Unlike existing approaches that first derive individual weights and subsequently apply a consensus model, LC-BWM simultaneously obtains individual and collective consensual weights within a single optimization framework. This allows the method to preserve the evaluators' original opinions as much as possible while ensuring a predefined level of agreement among the participants.

The key advantages of the proposal are listed below:

- The proposed approach is capable of considering several evaluators' views at the same time, facing MCGDM problems.
- The linguistic representation of the evaluators' opinions and the resulting weights facilitates the elicitation task, modelling the uncertainty in such opinions, and improves the interpretability of the results.
- The detection of disagreements between evaluators and the capacity to smooth them provides solutions in which evaluators agree, which is key in several real-world MCGDM problems.
- The stochastic consistency index here defined allows classifying the consistency of the BW preferences according to the probability of finding a more consistent preference.

From a practical perspective, the proposed LC-BWM provides moderators and decision analysts with a support tool for CRPs. By simultaneously deriving individual and collective consensual weights while preserving the evaluators' original views as much as possible, the approach facilitates the identification of disagreement sources and supports the generation of targeted recommendations. Furthermore, the linguistic representation of preferences and results enhances the interpretability of the obtained solutions.

Despite these advantages, several limitations should be acknowledged. The proposed methodology cannot guarantee that evaluators will accept the suggested preference modifications during a real CRP. In addition, the computational representation of linguistic assessments relies on predefined semantics that may not perfectly match the interpretation of all evaluators. Finally, further empirical validation in additional decision-making domains would strengthen the practical evidence supporting the proposal.

Regarding future research, several directions deserve further investigation. First, although the proposed LC-BWM can theoretically handle large groups of evaluators, its performance in large-scale group decision-making scenarios should be analysed in greater detail, particularly from a computational and consensus-management perspective. Second, future studies could investigate alternative consensus measures and aggregation operators to analyse their impact on the resulting collective weights. Third, it would be interesting to extend the proposal to more sophisticated uncertainty modelling frameworks, such as hesitant, probabilistic, or multi-granular linguistic information.

Furthermore, future work will explore the simulation of different evaluators' behaviours during CRPs, including situations in which some evaluators partially accept or reject the suggested modifications. The integration of intelligent support mechanisms to automatically generate personalized feedback recommendations for evaluators also represents a promising research direction.

Finally, we will consider the use of co-constructive preference modelling approaches to validate and refine the conversion of evaluators' linguistic assessments into computational representations (Corrente *et al.*, 2021). Such approaches could provide a more faithful representation of the semantics associated with linguistic expressions and further improve the interpretability and acceptance of the generated recommendations.

Statements and Declarations

Conflict of Interest

The authors have no relevant financial or non-financial interests to disclose.

Ethical Approval

This study does not involve human participants and/or animals.

Author Contribution

Álvaro Labella: Writing – review & editing, Writing original draft, Validation, Software, Formal analysis, Conceptualization. Diego García-Zamora: Writing – review & editing, Writing original draft, Validation, Software, Formal analysis, Conceptualization. Bapi Dutta: Writing – review & editing, Writing original draft, Validation, Software, Formal analysis, Conceptualization. Luis Martínez: Writing – review & editing, Writing original draft, Supervision, Formal analysis, Conceptualization.

Funding

Bapi Dutta gratefully acknowledges the support of the Spanish Ministry of Science, Innovation and Universities, and the Spanish State Research Agency through the Ramón y Cajal Research Grant (RYC2023-045020-I).

References

Adali, E.A., Tuş, A. (2026). Integration of analytic hierarchy process and multi attributive border approximation area comparison for the hybrid vehicle selection problem in intuitionistic fuzzy environment. *Informatica*, 37(1), 25–60. https://doi.org/10.15388/25-INFOR596.

Asakereh, A., Soleymani, M., Ardebili, S.M.S. (2022). Multi-criteria evaluation of renewable energy technologies for electricity generation: a case study in Khuzestan province, Iran. *Sustainable Energy Technologies and Assessments*, 52, 102220.

Ben-Arieh, D., Easton, T., Evans, B. (2008). Minimum cost consensus with quadratic cost functions. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(1), 210–217.

Bezanson, J., Edelman, A., Karpinski, S., Shah, V. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1), 65–98.

Bouraima, M.B., Jovčić, S., Dobrodolac, M., Pamucar, D., Badi, I., Maraka, N.D. (2024). Sustainable healthcare system devolution strategy selection using the AROMAN MCDM approach. *Spectrum of Decision Making and Applications*, 1(1), 46–63.

Brunelli, M., Rezaei, J. (2019). A multiplicative best–worst method for multi-criteria decision making. *Operations Research Letters*, 47(1), 12–15.

Butler, C.T.L., Rothstein, A. (2006). *On Conflict and Consensus: A Handbook on Formal Consensus Decision Making*. Takoma Park.

Cheng, C. (1998). A new approach for ranking fuzzy numbers by distance method. *Fuzzy Sets and Systems*, 95(3), 307–317.

Corrente, S., Figueira, J.R., Greco, S. (2021). Pairwise comparison tables within the deck of cards method in multiple criteria decision aiding. *European Journal of Operational Research*, 291(2), 738–756. https://doi.org/10.1016/j.ejor.2020.09.036.

Du, J., Xie, N., Liu, S., Goh, M. (2023). Grey linguistic term sets for decision-making. *Annals of Operations Research*, 348, 489–509. https://doi.org/10.1007/s10479-023-05319-0.

Dunning, I., Huchette, J., Lubin, M. (2017). JuMP: a modeling language for mathematical optimization. *SIAM Review*, 59(2), 295–320.

D'Adamo, I. (2023). The analytic hierarchy process as an innovative way to enable stakeholder engagement for sustainability reporting in the food industry. *Environment, Development and Sustainability*, 25(12), 15025–15042.

García-Zamora, D., Dutta, B., Labella, Á., Martínez, L. (2025). OWA operators in large-scale group decision-making: an analysis based on comprehensive minimum cost consensus. *Informatica*, 36(3), 557–588. https://doi.org/10.15388/25-INFOR599.

García-Zamora, D., Dutta, B., Labella, Á., Martínez, L. (2023a). A Fuzzy-set based formulation for minimum cost consensus models. *Computers & Industrial Engineering*, 181, 109295. https://doi.org/10.1016/j.cie.2023.109295.

García-Zamora, D., Dutta, B., Massanet, S., Riera, J.V., Martínez, L. (2023b). Relationship between the distance consensus and the consensus degree in comprehensive minimum cost consensus models: a polytope-based analysis. *European Journal of Operational Research*, 306(2), 764–776. https://doi.org/10.1016/j.ejor.2022.08.015.

González, J.A.T., Lantigua, M.E. (2024). Based on computing with words. *Computational Intelligence in Engineering and Project Management*, 1134, 339.

Goodwin, P., Wright, G. (2014). *Decision Analysis for Management Judgment*, 5th ed., Wiley.

Herrera, F., Herrera-Viedma, E., Martínez, L. (2008). A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Transactions on Fuzzy Systems*, 16(2), 354–370.

Ishizaka, A., Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software*. John Wiley & Sons.

Jameel, T., Yasin, Y., Riaz, M. (2026). An integrated hybrid MCDM framework for renewable energy prioritization in sustainable development. *Spectrum of Decision Making and Applications*, 3(1), 124–150.

Kabadayi, N., Dehghanimohammadabadi, M. (2022). Multi-objective supplier selection process: a simulation–optimization framework integrated with MCDM. *Annals of Operations Research*, 319(3), 1–23.

Kacprzyk, J., Fedrizzi, M. (1988). A 'soft' measure of consensus in the setting of partial (fuzzy) preferences. *European Journal of Operational Research*, 34(3), 316–325.

Kendall, M. (1970). Rank Correlation Methods 4th edition Charles Griffin. *High Wycombe, Bucks*.

Keršuliene, V., Zavadskas, E.K., Turskis, Z. (2010). Selection of rational dispute resolution method by applying new step-wise weight assessment ratio analysis (SWARA). *Journal of Business Economics and Management*, 11(2), 243–258.

Li, H., Dai, X., Zhou, L., Wu, Q. (2023). Encoding words into interval type-2 fuzzy sets: the retained region approach. *Information Sciences*, 629, 760–777.

Licerán-Gutiérrez, A., Ortega-Rodríguez, C., Moreno-Albarracín, A.L., Labella, Á., Rodríguez, R.M., Martínez, L. (2022). A linguistic group best–worst method for measuring good governance in the third sector: a Spanish case study. *International Journal of Fuzzy Systems*, 24(5), 2133–2156.

Majumder, P., Banik, S., Smarandache, F. (2026). An integrated AHP-SVNS-TOPSIS approach and its application to efficiency analysis of hydropower plant. *Informatica*, 1–17. https://doi.org/10.15388/26-INFOR628.

Martínez, L., Rodríguez, R.M., Herrera, F. (2015). *The 2-tuple Linguistic Model*. Springer International Publishing.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81.

Moreno-Albarracín, A.L., Licerán-Gutierrez, A., Ortega-Rodríguez, C., Labella, Á., Rodríguez, R.M. (2020). Measuring what is not seen—transparency and good governance nonprofit indicators to overcome the limitations of accounting models. *Sustainability*, 12(18), 7275.

Orlovski, S.A. (1978). Decision-making with fuzzy preference relations. *Fuzzy Sets and Systems*, 1, 155–167.

Ortega-Rodríguez, C., Labella, Á. (2026). Fuzzy logic meets accountability and management: linguistic modeling of the compliance officer functions in nonprofit cooperative societies. *Fuzzy Optimization and Decision Making*, 25(2), 225–256.

Palomares, I., Estrella, F.J., Martínez, L., Herrera, F. (2014). Consensus under a fuzzy context: taxonomy, analysis framework AFRYCA and experimental case of study. *Information Fusion*, 20, 252–271.

Parsons, S. (1996). Current approaches to handling imperfect information in data and knowledge bases. *IEEE Transactions on Knowledge and Data Engineering*, 8(3), 353–372.

Pingle, M. (1997). Submitting to authority: its effect on decision-making. *Journal of Economic Psychology*, 18(1), 45–68.

Prasanna, S., Verma, P., Bodh, S. (2025). The role of food industries in sustainability transition: a review. *Environment, Development and Sustainability*, 27(7), 15113–15133.

Puška, A., Stević, Ž., Pamučar, D. (2022). Evaluation and selection of healthcare waste incinerators using extended sustainability criteria and multi-criteria analysis methods. *Environment, Development and Sustainability*, 24, 11195–11225.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, 53, 49–57. https://doi.org/10.1016/J.OMEGA.2014.11.009.

Saaty, T. (1990). How to make a decision: the analytic hierarchy process. *European Journal of Operational Research*, 48(1), 9–26.

Saaty, T.L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3), 234–281. https://doi.org/10.1016/0022-2496(77)90033-5.

Safarzadeh, S., Khansefid, S., Rasti-Barzoki, M. (2018). A group multi-criteria decision-making based on best-worst method. *Computers & Industrial Engineering*, 126, 111–121.

Sahoo, S.K., Goswami, S.S., Halder, R. (2024). Supplier selection in the age of industry 4.0: a review on MCDM applications and trends. *Decision Making Advances*, 2(1), 32–47.

Srivastava, S., Tripathi, A., Arora, N. (2024). Multi-criteria decision making (MCDM) in diverse domains of education: a comprehensive bibliometric analysis for research directions. *International Journal of System Assurance Engineering and Management*, 1–18. https://doi.org/10.1007/s13198-024-02332-9.

Straffin Jr, P.D. (1977). Majority rule and general decision rules. *Theory and Decision*, 8(4), 351–360.

Tüysüz, N., Kahraman, C. (2023). A novel Z-fuzzy AHP&EDAS methodology and its application to wind turbine selection. *Informatica*, 34(4), 847–880. https://doi.org/10.15388/23-INFOR515.

Wang, J.H., Hao, J. (2006). A new version of 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 14(3), 435–445.

Zadeh, L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning, Part I, II, III. *Information Sciences*, 8(3), 199–249.

Zadeh, L.A. (1996). Fuzzy logic = Computing with words. *IEEE Transactions on Fuzzy Systems*, 4(2), 103–111.

Á. Labella holds a PhD in computer science from the University of Jaén where he currently works as an associate professor and researcher in the SINBAD2 group. His research focuses on decision-making, computing with words, and software application development.

D. García-Zamora received a bachelor's degree in mathematics from the University of Granada and he is doctor in information and communication technologies from the University of Jaén. Currently, he is an assistant professor in the Department of Mathematics at the University of Jaén and a researcher in the SINBAD2 research group. His work focuses on decision-making, aggregation operators, fuzzy logic, and optimization.

B. Dutta is a Ramón y Cajal researcher at the University of Jaén. His research focuses on decision-making, soft computing, simulation, optimization, and machine learning, with particular emphasis on developing advanced computational models and methodologies to address complex real-world problems.

L. Martínez is currently full professor with the Computer Science Department of the University of Jaén. His current research interests include multi-criteria decision-making, fuzzy logic-based systems, computing with words, and recommender systems.