Chemometrics and Intelligent Laboratory Systems 270 (2026) 105624

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Smoothed Power-Weakness Ratio (sPWR): a new informative system for multi-criteria decision making

Viviana Consonni, Davide Ballabio, Enmanuel Cruz Muñoz, Veronica Termopoli, Roberto Todeschini *

Milano Chemometrics and QSAR Research Group, Department of Earth and Environmental Sciences, University of Milano-Bicocca, P.zza della Scienza 1, 20126, Milano, Italy

ARTICLE INFO

Keywords:
MCDM
PWR
TOPSIS
AHP
Tournament matrix

ABSTRACT

Nowadays, the large number of measurable variables has considerably increased the complexity of data. In the framework of the decision-making process, this leads to the need of adequate tools to set priorities and rank the available options. Ordering is one of the possible ways to analyse multivariate data, which provides an overview of the relationships among the elements of a system. The Multi-Criteria Decision Making (MCDM) encompasses a broad set of methods designed to set priority-based lists of alternatives based on multiple criteria, which support decision problems. Among the most widely adopted techniques, TOPSIS, dominance-based approaches, the Analytic Hierarchy Process (AHP), and Copeland scores represent some of the classical methodologies in both theoretical research and applied decision analysis.

Among the dominance-based approaches, an effective MCDM method is the Power-Weakness Ratio (PWR), which generates a tournament table (i.e., the pairwise comparison matrix) from a data matrix with a varying number of samples (i.e., alternatives to be compared) and variables (i.e., the criteria for pairwise comparisons), weighted according to their relative importance in determining the final ranking. In this study, a variant of the classical Power-Weakness Ratio is presented, significantly modifying the way the tournament table is obtained. The method, called smoothed Power-Weakness Ratio (sPWR), takes into account the dominance degree of the alternatives in each pairwise comparison exploiting the differences between the criterion values. The rationale behind the method is described by the aid of an illustrative example on a simple benchmark dataset with known reference ranking of the samples. The main advantage of the new method over PWR is that its tournament table is much more informative and sensitive to the original data values than the classical pairwise comparison matrix. A multivariate comparison with other classical MCDM methods, performed on several diverse datasets, demonstrated that the results obtained by sPWR were quite similar to those obtained by Copeland Score and TOPSIS with range scaling. However, sPWR showed a higher tendency toward generating full rankings with an enhanced ability to remove ties in the pairwise comparisons.

1. Introduction

The intrinsic complexity of systems analysed in scientific research, together with the significant grow of available data, requires suitable methodologies for multivariate statistical analysis and motivates the continuous development of new approaches. One way to analyse multivariate data is to order the elements of a system toward a specific goal, by exploiting the order relationships derived from the measured variables. This type of problem is very general and quite common, since many of our choices are frequently based on a set of preferences defined (consciously or unconsciously) according to multiple criteria. For example, when choosing a car, factors such as price, brand, model, size, colour, fuel consumption, power, and sustainability all play a role, though to varying degrees of importance. The final choice therefore represents a compromise among these variables, each reflecting the decision-maker's subjective preferences.

In recent years, ranking entities such as players, teams, scientists, universities, industrial materials, or countries, based on diverse indicators reflecting economic prosperity, public health, biodiversity, and other aspects, has become of greater importance. These rankings

* Corresponding author.
E-mail address: roberto.todeschini@unimib.it (R. Todeschini).

https://doi.org/10.1016/j.chemolab.2025.105624
Received 8 October 2025; Received in revised form 26 November 2025; Accepted 28 December 2025
Available online 29 December 2025

influence decision-making processes, leading to different actions being taken depending on the positions achieved. Due to an increasing number of measurable variables, the decision processes have become more challenging, thus requiring the support of adequate tools to set priorities and rank alternatives.

Ranking approaches can be divided into total- and partial-order methods, depending on the type of ordering they produce. These techniques support decision-making by establishing priority-based lists. Unlike the most common multivariate statistical methods, such as those used for data pre-processing, exploration, and modelling, ranking approaches typically rely on simpler procedures. Over the years, several Multi-Criteria Decision-Making (MCDM) techniques have been developed [1–3], with meaningful applications in diverse fields such as environmental management [4], risk assessment [5], Quantitative Structure-Activity Relationship (QSAR) modelling [6], energy planning [7], healthcare [8], food and agriculture [9], drug discovery [10], research impact evaluation [11], and industrial processes [12–14].

Ranking thus represents an essential component of multivariate data analysis when a problem requires to define priorities among objects (e.g., samples, cases, situations, molecules, events, options, etc.), which are ordered according to a score derived from multiple variables (or criteria) used for the pairwise comparisons. The resulting rank enables the solution of MCDM problems.

Dominance analysis, distance-based aggregation (TOPSIS), preference elicitation (AHP), and voting-based aggregation (Copeland scores) represent some of the most widely adopted methodologies, in both theoretical research and applied decision analysis [1,15]. Dominance-based methods rely on direct comparisons between alternatives [2]. An alternative is said to dominate another if it performs at least as well across all criteria and strictly better in at least one. Despite their simplicity and intuitive logic, these methods require subjective inputs, such as thresholds for indifference or preference and the assignment of criterion importance, and may produce ambiguous rankings in complex high-dimensional data systems, where several alternatives dominate each other in different ways or where highly correlated criteria can distort the resulting dominance relationships. TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) evaluates each alternative by measuring its relative distance to two hypothetical reference points (ideal and non-ideal alternatives) [16]. This method is one of the most frequently used ranking methods due to its conceptual simplicity and computational efficiency; however, the results can vary depending on how data are normalized and scaled (e.g., vector, linear, min–max). Moreover, it assumes that criteria are independent as correlated criteria can bias the results. Another shortcoming is that the method relies solely on distances to ideal and non-ideal points, ignoring how alternatives are distributed in the criteria space. The Analytic Hierarchy Process (AHP) is a structured method particularly suited for problems that combine quantitative and qualitative criteria [17]. AHP uses pairwise comparisons to assess both the importance of criteria and the relative performance of alternatives. These comparisons are then processed through an eigenvector-based weighting mechanism to derive a global ranking. AHP is widely appreciated for its transparency and ability to incorporate expert judgment, although its reliance on subjective evaluations has also been identified as a method shortcoming. Moreover, the use of the 1–9 preference scale can amplify small differences in judgments and distort priorities. Finally, Copeland scores represent a voting-theory-based approach for MCDM. In this method, alternatives are compared pairwise, and each pairwise outcome contributes to a global score: a win adds one point, a loss subtracts one, and a tie yields zero. Summing these results produces a Copeland score for each alternative, enabling a straightforward and robust ranking. This method is particularly useful in group decision-making scenarios where preferences must be aggregated across multiple evaluators.

Still in the framework of the dominance analysis approaches, the Power-Weakness Ratio (PWR) has been recently proposed [18] to introduce the possibility to derive a pairwise comparison matrix (i.e., the tournament table) from a data matrix **X** of n objects described by p variables (criteria) and to assign weights to these variables according to their predefined relevance in determining the final ranking. A tournament table is a square matrix (n x n) that is comprised of the outcomes of pairwise comparisons within a set of alternatives (e.g., players, teams, objects, etc.); each entry in the matrix results from a series of comparisons between two alternatives and represents the overall preference of one alternative over the other. This method provides a clear and complete ordering of alternatives, derived from the ratio of power over weakness, which directly measures the relative performance of alternatives. Moreover, it can be easily applied to multivariate datasets with varying numbers of objects and criteria. However, PWR is an ordinal comparison-based method that does not account for the magnitude of differences between alternatives. In other words, PWR evaluates which alternative is preferred and not how much an alternative is preferred over another. All differences are treated equally and this can lead to misleading rankings when numerical values strongly differ. Moreover, pairwise comparisons can easily become contradictory in the case of conflicting criteria.

In this paper, a variant of the classical Power-Weakness Ratio is presented, significantly modifying the way the tournament table is obtained. The method, called smoothed Power-Weakness Ratio (sPWR), takes into account the quantitative difference between criterion values for each pairwise comparison generating a tournament table that is much more informative and sensitive to the original data values than that of PWR. This variant specifically addresses the need to handle quantitative variables for which differences in their numerical values can be relevant to settle the final ranking.

In the following, the theoretical fundamentals of the proposed approach are first introduced. Then, the accuracy of sPWR is assessed on a benchmark dataset with a known reference raking of the alternatives. Finally, a multivariate comparison of sPWR with other classical multi-criteria decision-making methods is carried out on 12 diverse datasets to disclose similarities in their results and highlight their respective strengths and limitations.

2. Theory

The Power-Weakness Ratio is an effective multi-criteria decision method that involves three main steps: 1) constructing a weighted tournament table **T**, 2) performing the spectral decomposition of the tournament table **T** and its transpose $\mathbf{T}^{\mathrm{T}}$ to obtain the corresponding eigenvalues and eigenvectors, and 3) computing the object scores and ranks using the dominant eigenvectors corresponding to the maximum eigenvalues of **T** and $\mathbf{T}^{\mathrm{T}}$, respectively.

2.1. The weighted tournament table

The weighted tournament table **T** is built by following the classical rules of a round-robin tournament, where each player plays against all the other $n - 1$ players, once or more than once.

Let **X** be a data matrix being comprised of n objects (individuals, alternatives, cases, events, scenarios, samples, etc.) described by p criteria (variables, attributes). The tournament table (**T**), also called dominance matrix [19], is a weighted count matrix of size n x n obtained by defining its elements as follows [5]:

$$[\mathbf{T}]_{it} = \sum_{j=1}^{p} w_j \cdot \delta_{it,j} \quad \text{where} \quad \delta_{it,j} = \begin{cases} 1 & \text{if } x_{ij} \rhd x_{tj} \\ 0.5 & \text{if } x_{ij} \triangleq x_{tj} \\ 0 & \text{if } x_{ij} \lhd x_{tj} \end{cases} \quad \text{and} \quad \sum_{j=1}^{p} w_j = 1 \tag{1}$$

where w_j is the weight (i.e., relevance) given by the decision maker to the j-th criterion, x_{ij} and x_{tj} are the values of the j-th criterion for the i-th and t-th object, respectively. In other words, for each variable j, the i-th object will dominate against the t-th object if its value is preferable to that of t (i.e., $x_{ij} \rhd x_{tj}$). If the contrary happens ($x_{ij} \lhd x_{tj}$), the t-th object will dominate over i. Finally, if the values are equal ($x_{ij} \triangleq x_{tj}$), the two objects "tie the comparison" for the j-th criterion and half a credit is given to both. This approach is a very simple and efficient way to compare objects taking also into account the variable importance. Note that all the entries of the matrix **T** range from 0 to 1 and the sum of symmetrical entries is equal to 1, that is:

$$[\mathbf{T}]_{it} + [\mathbf{T}]_{ti} = 1 \quad \text{and} \quad [\mathbf{T}]_{ii} = 0 \qquad i, t = 1, \ldots, n \tag{2}$$

2.2. The Power-Weakness Ratio (PWR)

The Power–Weakness Ratio (PWR) was proposed in 1964 by Ramanujacharyulu [20] as a method to find the winner of a round-robin tournament or the most influential person within a group. PWR aims to locate the most talented individual, defined as the one who won over the largest number of opponents (maximum "power") and was simultaneously defeated by the smallest number of opponents (minimum "weakness"). PWR was hence supposed to capture the balance between the power and the weakness of individuals (objects, alternatives, cases, etc.).

The PWR calculated on the tournament table **T** including n players encodes the power of any player over the remaining $n-1$ ones, that is, the row sum of this matrix represents the number of times a player has defeated the other players. On the other hand, the PWR calculated on the transpose of the tournament table $\mathbf{T}^{\mathrm{T}}$ encodes the weakness of any player over the remaining $n-1$, that is, the row sum in this case represents the number of times a player is defeated by the other players.

The properties of both these matrices have been studied extensively elsewhere [21]. The tournament matrix is asymmetrical, nonnegative and irreducible. The Perron–Frobenius theorem [22] ensures that the spectral decomposition of this kind of matrices always gives a positive eigenvalue. The eigenvector that corresponds to this eigenvalue was proposed by Kendall and Wei as an effective ranking criterion for the players [23–25].

Let ℓ and ℓ^* be the n-dimensional dominant eigenvectors corresponding to the highest positive eigenvalue of the tournament table **T** with n objects and its transpose matrix $\mathbf{T}^{\mathrm{T}}$, respectively. The PWR score of the i-th object is then defined as the ratio of the corresponding elements (absolute values) of these two eigenvectors:

$$\mathrm{PWR}_i = \frac{\ell_i}{\ell_i^*} \tag{3}$$

In some cases, zero or very small eigenvector entries can be obtained, thus generating singularities or spikes. To overcome this drawback, the original function can be corrected by a parameter α that allows to obtain more suitable PWR scales, as follows:

$$\mathrm{PWR}_i = \frac{\alpha + \ell_i}{\alpha + \ell_i^*} \qquad \alpha = \frac{1}{n} \tag{4}$$

n being the number of objects taking part in the tournament.

Moreover, a simple similarity measure between objects i and t can be defined as a function of the average Manhattan distance on the corresponding elements of the two eigenvectors ℓ and ℓ^*:

$$S_{it} = 1 - \frac{|\ell_i - \ell_t| + |\ell_i^* - \ell_t^*|}{2} \qquad 0 \le S_{it} \le 1 \tag{5}$$

It is noteworthy that the pairwise similarity calculated on the PWR eigenvectors differs from the similarity measured on the original data matrix **X**. In other words, data and ranking structures represent two distinct sources of information that can be compared to gain deeper insights into the relationships between objects.

The PWR algorithm is simple and the two eigenvectors (i.e., *power* and *weakness*) encode in a holistic way the relative behaviour of the objects, with spectral decomposition providing a powerful framework for capturing all their relationships.

2.3. The smoothed power-weakness ratio (sPWR)

The classical rules for constructing the tournament table fail to capture the dominance degree of an object over another, as they merely stablish whether one variable's value is preferable to another, without quantifying the extent of their difference. Indeed, in the case of p equally weighted criteria, the entries of the tournament matrix can assume only $2 \cdot p + 1$ distinct values. For example, having 3 criteria, only the following values can be obtained: 0, 0.167, 0.333, 0.5, 0.667, 0.833, 1. The extreme values 1 and 0 are obtained when one object dominates the other one for all criteria and *vice versa*; the other values are obtained for different combinations of wins/losses and ties.

In order to consider the quantitative information on the degree of one object's dominance over another, the smoothed Power-Weakness Ratio (sPWR) is proposed as a refined version of the PWR approach. Given p criteria, the sPWR algorithm involves the following steps:

1. For all the pairs of objects i and t and for each j-th criterion, the following quantity is calculated:

$$q_{it,j} = \frac{1}{2} \cdot \left[1 + \omega_j \cdot \frac{(x_{ij} - x_{tj})}{\Delta x_j}\right] \quad i \neq t$$
$$0 \le q_{it,j} \le 1 \quad \text{and} \quad q_{ti,j} = 1 - q_{it,j} \tag{6}$$

The range $\Delta x_j = x_j^{Max} - x_j^{Min}$ of the j-th criterion is used to obtain a scaled quantity $q_{it,j}$, which is oriented towards the optimality of each criterion according to the rule:

$$\omega_j = \begin{cases} +1 & \text{if optimality is towards high values} \\ -1 & \text{if optimality is towards low values} \end{cases}$$

2. Afterwards, the sPWR weighted tournament table **T** is built as:

$$[\mathbf{T}]_{it} = \begin{cases} \sum_{j=1}^{p} w_j \cdot q_{it,j} & t \neq i \\ 0 & t = i \end{cases} \qquad 0 \le [\mathbf{T}]_{it} \le 1 \tag{7}$$

where the symmetrical entries are complementary, i.e. $[\mathbf{T}]_{it} + [\mathbf{T}]_{ti} = 1$ and w_j is the criterion weight under the usual condition $\sum_{j=1}^{p} w_j = 1$.

3. The sPWR scores and corresponding ranks of the n objects are obtained as in PWR, that is, by performing the spectral decomposition of **T** and $\mathbf{T}^{\mathrm{T}}$ and by using the elements (absolute values) of the dominant eigenvectors ℓ (*power*) and ℓ^* (*weakness*) corresponding to the highest positive eigenvalue of the tournament table **T** and its transpose matrix $\mathbf{T}^{\mathrm{T}}$, respectively:

$$\mathrm{sPWR}_i = \frac{\alpha + \ell_i}{\alpha + \ell_i^*} \qquad \alpha = \frac{1}{n} \tag{8}$$

In MCDM comparative analyses, the sPWR scores can be range scaled or scaled with respect to the maximum score, thus preserving the proportions.

It can be noted that sPWR algorithm is unique and the definition of $q_{it,j}$ allows to rescale both positive and negative differences in the range [0, 1]. For example, for a criterion j with optimality towards high values ($\omega_j = +1$), the following three benchmarking cases can be observed:

$$\begin{aligned} &x_{ij} = x_{tj} && \to \quad q_{it,j} = 0.5 \\ &x_{ij} = x_j^{Max} \wedge x_{tj} = x_j^{Min} && \to \quad q_{it,j} = 1 \\ &x_{ij} = x_j^{Min} \wedge x_{tj} = x_j^{Max} && \to \quad q_{it,j} = 0 \end{aligned}$$

Thus, a value of 0.5 is assigned in the case of a tie as in PWR, while the values 1 and 0 are assigned only when one object is the absolute best and the other the absolute worst with respect to the considered criterion. In practice, $q_{it,j}$, which assumes any possible value in the range [0, 1], replaces $\delta_{it,j}$ of the original PWR algorithm, which can take only a finite set of three values (0, 0.5, 1). However, this small change has a big impact on the encoded information available in the tournament matrix. This means that, if data are reported with the appropriate number of significant figures, sPWR is able to take into account all the numerical information available.

It is also interesting to consider the difference between the new quantitative sPWR tournament table and the classical (proportional) PWR tournament table in terms of mean information content (entropy). For this purpose, only the upper off-diagonal elements $m = n\cdot (n-1)/2$ of the tournament table are considered since the symmetrical elements are complementary. Assuming that, given p criteria, the possible $z = 2\cdot p + 1$ values by PWR are equally distributed, then the maximum entropy is $log_2(z)$; while, by considering the same m elements obtained by sPWR, the maximum entropy is $log_2(m)$, assuming all the values are equally distributed in the case of unique distinct values. This means that the amount of theoretical available information depends on the number of objects for sPWR, while it generally depends on the number of criteria for PWR.

The sPWR algorithm is also applicable to binary and ordered categorical variables without modification. It can be considered as a smoothing procedure of the crisp PWR analysis, the latter based on the relative frequencies of the number of cases answering to questions as “is larger than”, “is better than”, “yes or not”, “win or lose”, “on or off”.

It can be observed that, in general, sPWR tournament matrices tend to show several quasi-symmetrical entries around 0.5, due to balanced outcomes in multi-criteria comparisons. On the contrary, extreme values (0 and 1), which indicate the presence of clearly dominant or clearly inferior players, are typically rare.

As the tournament matrix approaches a state of balance among players (no player emerges as a clear winner or loser), the relationship between the dominant eigenvectors of $\mathbf{T}$ and $\mathbf{T}^{\mathrm{T}}$ tends to be more and more linear. In the extreme case of perfect balance among all the alternatives, the two eigenvectors are collinear. For more asymmetrical matrices, inverse non-linear relationships are usually obtained.

A simple example can clarify the behaviour of the sPWR algorithm. Table 1 collects the data for 6 cases described by 4 criteria (C1 – C4). The optimality of each criterion is towards high values and all the weights are equal (0.25). Without loss of generality, it is assumed that the cases behave in the same way for all the criteria; for instance, the first object always wins against all the others and, on the contrary, the fifth object always loses; the second object is always the second best; the fourth and sixth objects tie the comparison. The corresponding tournament tables obtained by applying the rules of PWR and sPWR are reported in Table 2.

When looking at the PWR tournament table, it can be easily observed that object 1 dominates all the others in the same way, regardless of the quantitative difference from the other objects. On the other side, it is evident that the sPWR table provides more information, as it incorporates the quantitative differences in the values each object takes for the criteria. For instance, object 1 also in sPWR table wins against all the others with dominance values greater than 0.5; however, the dominance degree varies depending on the object being compared. For example, its dominance over object 2 is less marked than its dominance over the objects 4 and 6.

Considering the information content, the 15 entries of the upper PWR matrix include 13 times the value 1 and only once both the values 0.5 and 0; hence, the entropy is:

$$H_{PWR} = -\left(\frac{13}{15}\right)\cdot log_2\left(\frac{13}{15}\right) - 2\cdot\left(\frac{1}{15}\right)\cdot log_2\left(\frac{1}{15}\right) = 0.700$$

In the sPWR matrix, there is much diversity among the entries (i.e., once the values 1, 0.4 and 0.5; twice the values 0.6, 0.75 and 0.8; three times the values 0.65 and 0.9) and consequently the entropy is higher:

$$H_{sPWR} = -3\cdot\left(\frac{1}{15}\right)\cdot log_2\left(\frac{1}{15}\right) - 3\cdot\left(\frac{2}{15}\right)\cdot log_2\left(\frac{2}{15}\right) - 2\cdot\left(\frac{3}{15}\right)\cdot log_2\left(\frac{3}{15}\right) = 2.873$$

The entropy in the case of PWR is 0.700 against a maximum available entropy of 3.170 (i.e., $log_2(2\cdot p + 1) = log_2(9)$), while in the case of sPWR, the entropy is 2.873 against a maximum available entropy of 3.907 (i.e., $log_2(15)$).

The final PWR and sPWR scores are reported in Table 3 along with the corresponding scaled values on the maximum score and the final ranks of the six objects.

As expected in this case, the ranks of the objects are the same, but the scores give a remarkable different information as shown in Fig. 1, where the smoothing action of sPWR is well highlighted. In other words, sPWR acts directly on the original values of the variables, thus having a high sensitivity to the relative behaviour of the objects.

Exploiting the proposed similarity measure based on the elements of the dominant eigenvectors *power* and *weakness*, the similarities between the objects of the simulated example have been calculated for both PWR and sPWR approaches. These similarities are compared in Table 4. As it can be easily observed, the similarities perceived by sPWR are overall higher than those calculated by the classical PWR; for example, as expected considering the values of the four criteria (Table 1), the similarity between objects 1 and 2 increases from 0.686 with PWR to 0.912 with sPWR.

Table 1
Simulated dataset with 6 objects described by 4 criteria.

id	C1	C2	C3	C4
1	1	10	1.2	2
2	0.8	8	1	1.6
3	0.5	5	0.7	1
4	0.2	2	0.4	0.4
5	0	0	0.2	0
6	0.2	2	0.4	0.4

Table 2
Tournament tables **T** obtained by the PWR and sPWR algorithms for the simulated dataset (Table 1).

$$\mathbf{T}^{PWR} = \begin{array}{c|cccccc} id & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 2 & 0 & 0 & 1 & 1 & 1 & 1 \\ 3 & 0 & 0 & 0 & 1 & 1 & 1 \\ 4 & 0 & 0 & 0 & 0 & 1 & 0.5 \\ 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 & 0.5 & 1 & 0 \end{array} \qquad \mathbf{T}^{sPWR} = \begin{array}{c|cccccc} id & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 0.60 & 0.75 & 0.90 & 1 & 0.90 \\ 2 & 0.40 & 0 & 0.65 & 0.80 & 0.90 & 0.80 \\ 3 & 0.25 & 0.35 & 0 & 0.65 & 0.75 & 0.65 \\ 4 & 0.10 & 0.20 & 0.35 & 0 & 0.60 & 0.50 \\ 5 & 0 & 0.10 & 0.25 & 0.40 & 0 & 0.40 \\ 6 & 0.10 & 0.20 & 0.35 & 0.50 & 0.60 & 0 \end{array}$$

Table 3
Scores and ranks calculated by the spectral decomposition of PWR and sPWR tournament tables (Table 2) for the simulated dataset (Table 1).

id	scores		max scaled scores		ranks	
	PWR	sPWR	PWR	sPWR	PWR	sPWR
1	6.657	2.619	1.000	1.000	1	1
2	2.886	1.796	0.434	0.686	2	2
3	1.629	1.085	0.245	0.414	3	3
4	0.479	0.662	0.072	0.253	4.5	4.5
5	0.150	0.464	0.023	0.177	6	6
6	0.479	0.662	0.072	0.253	4.5	4.5

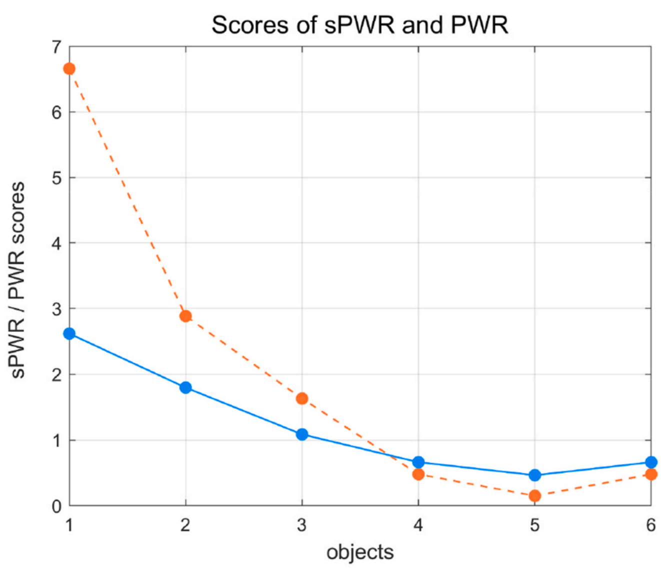

Fig. 1. Comparison of the unscaled scores obtained by PWR (dashed orange line) and sPWR (solid blue line).

2.4. Multi-criteria decision methods used for comparison

In order to deeply investigate the results of sPWR, the most classical ranking methods were considered for a comparative analysis. These methods are briefly described below, assuming to have a data matrix $\mathbf{X}(n \times p)$ of n objects described by p criteria to which weights w_j are assigned according to their relative importance in determining the decision process.

2.4.1. Simple Average Ranking (SAR)

This is a basic method that calculates a score as the weighted sum of the ranks generated by each individual criterion [26]:

$$SAR_i = \sum_{j=1}^{p} w_j \cdot r_{ij} \qquad 1 \leq SAR_i \leq n \qquad \sum_{j=1}^{p} w_j = 1 \tag{9}$$

where r_{ij} is the rank of the i-th object on the j–th criterion; preferable values of SAR_i are the lowest ones, corresponding to lowest average rank. Therefore, the final ranking of the alternatives is obtained according to ascending order of the SAR values.

2.4.2. Copeland Score (CS)

Based on pairwise comparisons, the Copeland Score was originally proposed as the difference between the number of wins and the number of losses of each alternative against every other alternative [27,28]. In this study, to contemporaneously considering all the criteria and the corresponding weighting schemes, the Copeland Score was calculated as the row sum of the tournament table $\mathbf{T}$, that is:

$$CS_i = \sum_{t=1}^{n} [\mathbf{T}]_{it} \tag{10}$$

The row sum of this matrix represents the overall net performance (*power*) of each alternative, as it reflects the relative number of times the alternative has been preferred. The alternatives are finally ordered from the highest to the lowest values of this score.

In this study, CS1 and CS2 refer to the Copeland Scores obtained by sPWR and PWR tournament table, respectively.

2.4.3. Dominance function (DOM)

This ranking approach is based on the calculation of the benefit W^+ and the cost of each pairwise comparison as [2]:

$$W_{it}^{+} = \sum_{j=1}^{p} w_j \cdot \delta_{it,j} \quad \text{where} \quad \delta_{it,j} = \begin{cases} 1 & \text{if } x_{ij} \rhd x_{tj} \\ 0.5 & \text{if } x_{ij} \triangleq x_{tj} \\ 0 & \text{if } x_{ij} \lhd x_{tj} \end{cases} \tag{11}$$

and $W_{it}^{-} = 1 - W_{it}^{+}$, as $\sum_{j=1}^{p} w_j = 1$.

The result of the comparison between each pair of objects i and t is calculated as:

$$C_{it} = \frac{1 + W_{it}^{+}}{1 + W_{it}^{-}} \qquad 0.5 \leq C_{it} \leq 2 \tag{12}$$

and the final score of each object is obtained as the arithmetic mean:

$$DOM_i = \frac{\sum_{t=1}^{n} C_{it}}{n-1} \qquad t \neq i \qquad 0.5 \leq DOM_i \leq 2 \tag{13}$$

As it can be easily observed, each pairwise comparison is obtained in the same way as in the PWR algorithm. The two methods differ in that, while in PWR the scores are derived from a double spectral decomposition of the tournament matrix, in DOM each object's score is simply calculated as the average of the n – 1 pairwise comparisons.

2.4.4. Technique for order of preference by similarity to ideal solution (TOPSIS)

The TOPSIS method is widely applied in industrial processes; however, several versions have been proposed, particularly concerning the initial step of data scaling [12,16,29]. The method relies on the principle that the optimal alternative is the one closest to the ideal solution (best for all criteria) and farthest from the non-ideal solution (worst for all criteria).

Table 4
The similarity values between the objects of the simulated dataset (Table 1) from the PWR/sPWR eigenvectors obtained by the two corresponding tournament tables (Table 2).

id	Similarity from PWR						Similarity from sPWR					
	1	2	3	4	5	6	1	2	3	4	5	6
1	1	0.686	0.581	0.424	0.057	0.424	1	0.912	0.779	0.646	0.558	0.646
2	0.686	1	0.895	0.738	0.371	0.738	0.912	1	0.867	0.735	0.646	0.735
3	0.581	0.895	1	0.843	0.476	0.843	0.779	0.867	1	0.867	0.779	0.867
4	0.424	0.738	0.843	1	0.633	1.000	0.646	0.735	0.867	1	0.912	1.000
5	0.057	0.371	0.476	0.633	1	0.633	0.558	0.646	0.779	0.912	1	0.912
6	0.424	0.738	0.843	1.000	0.633	1	0.646	0.735	0.867	1.000	0.912	1

The most common TOPSIS method (TOPSIS1) involves the following scaling scheme:

$$a)\quad x'_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} \qquad b)\quad x'_{ij} = 1 - \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} \tag{14}$$

where a) applies when high values are preferable and b) when low values are preferable.

In order to evaluate the influence of the scaling procedure on the TOPSIS results, two TOPSIS variants have been also considered. The TOPSIS2 method is a variant based on the max scaling of each criterion:

$$a)\quad x'_{ij} = \frac{0.1 + x_{ij}}{0.1 + x_j^{Max}} \qquad b)\quad x'_{ij} = \frac{0.1 + x_j^{Min}}{0.1 + x_{ij}} \qquad x'_{ij} \le 1 \tag{15}$$

where x_j^{Min} and x_j^{Max} are the minimum and maximum values of the j-th criterion, respectively; the offset 0.1 is added to avoid singularities.

The TOPSIS3 method is another variant based on the range scaling of each criterion:

$$a)\quad x'_{ij} = \frac{x_{ij} - x_j^{Min}}{x_j^{Max} - x_j^{Min}} \qquad b)\quad x'_{ij} = \frac{x_j^{Max} - x_{ij}}{x_j^{Max} - x_j^{Min}} \qquad 0 \le x'_{ij} \le 1 \tag{16}$$

After the variable scaling, the worst reference vector ($\mathbf{x}^-$) and the best reference vector ($\mathbf{x}^+$) are defined as:

$$\mathbf{x}^- = \left\{ min_i(x'_{i1}), min_i(x'_{i2}), \ldots, min_i(x'_{ip}) \right\} \quad i = 1, \ldots, n$$

$$\mathbf{x}^+ = \left\{ max_i(x'_{i1}), max_i(x'_{i2}), \ldots, max_i(x'_{ip}) \right\} \quad i = 1, \ldots, n$$

Finally, the weighted Euclidean distances of each i-th object from the worst and best reference vectors are calculated, respectively, as:

$$D_i^- = \sqrt{\sum_{j=1}^{p} w_j \cdot \left(x'_{ij} - x_j^-\right)^2} \quad \text{and} \quad D_i^+ = \sqrt{\sum_{j=1}^{p} w_j \cdot \left(x'_{ij} - x_j^+\right)^2} \tag{17}$$

with the usual constraint $\sum_{j=1}^{p} w_j = 1$; then, the final score is obtained as:

$$TOPSIS_i = \frac{D_i^-}{D_i^- + D_i^+} \qquad 0 \le TOPSIS_i \le 1 \tag{18}$$

where $TOPSIS_i = 1$ if the i-th alternative is the best solution ($D_i^+ = 0$) and $TOPSIS_i = 0$ if the i-th alternative is the worst solution ($D_i^- = 0$). The alternatives are finally ranked in descending order of their TOPSIS scores.

2.4.5. Analytic Hierarchy Process (AHP)

This method is still based on the spectral decomposition of a tournament table, which in this context is called *pairwise comparison matrix* [17]. Unlike the PWR and sPWR approaches, this tournament table does not account for all the considered criteria but it is defined for each criterion one at a time. Then, the final score is calculated as a weighted sum of the coefficients of the dominant eigenvector of each tournament matrix.

The first step of AHP involves assigning a relative preference on a 1–9 scale to each pairwise comparison. The assigned preferences are collected into the tournament matrix, which is a positive and reciprocal square matrix of order n obtained by pairwise comparing the n alternatives on the basis of a selected criterion.

The element $\left[\mathbf{T}_j^{\mathrm{AHP}}\right]_{it}$ of this matrix is thus an estimation of the degree of preference of x_{ij} over x_{tj} x_{tj}, with the constraint $\left[\mathbf{T}_j^{\mathrm{AHP}}\right]_{it} \cdot \left[\mathbf{T}_j^{\mathrm{AHP}}\right]_{ti} = 1$ or, alternatively, $\left[\mathbf{T}_j^{\mathrm{AHP}}\right]_{ti} = 1/\left[\mathbf{T}_j^{\mathrm{AHP}}\right]_{it}$.

The second step of AHP involves the calculation of the dominant eigenvector of each tournament matrix (i.e., the eigenvector corresponding to the highest eigenvalue); then, the final score of each i-th alternative is obtained as a weighted sum of the corresponding coefficients ℓ_{ij} of the dominant eigenvectors obtained by the spectral decomposition of each j-th tournament table, i.e.

$$AHP_i = \sum_{j=1}^{p} w_j \cdot \ell_{ij} \tag{19}$$

where w_j are the normalized weights defining the relative importance of the criteria in the pairwise comparisons.

3. Software

All the algorithms and calculations were carried out in MATLAB (version 2020b) by means of in-house software written by the Authors, the sPWR toolbox, which is available for free download at https://michem.unimib.it/download/matlab-toolboxes/power-weakness-ratio-toolbox/ together with some of the datasets used in this study.

4. Datasets

The rationale of the novel multicriteria ranking method was described by means of an illustrative example on the Fruit dataset reported in Table 5. Fruits are characterized by four variables related to their nutritional composition: water (g), protein (g), fat (g) and carbohydrates (g); in addition, the energy (Kcal) is also known and used as a benchmark ranking.

The whole comparative analysis of the new sPWR with nine classical ranking methods was carried out on other 12 benchmark datasets taken from the literature and briefly described in Table 6.

5. Results and discussion

As a first step, the proposed sPWR method was tested on a dataset of 37 fresh and dried fruits characterized by their nutritional composition and ordered in descending energy value (Kcal). For this dataset, the known reference ranking of the fruits was used to evaluate the accuracy of the final rankings produced by sPWR in comparison with the other methods considered.

Next, in order to investigate the similarities among the results of sPWR and the other ranking methods, a multivariate comparison was carried out on 12 small-to large-sized datasets with a number of criteria ranging from 3 to 32 (Table 6). The ranking methods selected for this comparison are: smoothed Power-Weakness Ratio (sPWR), Power-Weakness Ratio (PWR), Simple Average Ranking (SAR), Dominance functions (DOM), three TOPSIS versions (TOPSIS1, TOPSIS2 and TOPSIS3), the Analytic Hierarchy Process (AHP) and two Copeland scores (CS1 and CS2) calculated on the tournament tables derived by sPWR and PWR, respectively.

For sake of simplicity, all the variable weights were assumed to be equal ($w_j = 1/p$), thus obtaining an equal contribution of all the considered criteria to the final ranking. Moreover, the results were obtained by rounding the object scores with 3 decimal digits for all the considered methods. Obviously, the fewer the decimal digits, the higher the number of ties in the obtained ranks.

In the following, the ranking methods are compared: 5.1) for their accuracy, on the benchmark dataset Fruit, for which a known ranking of fruit is provided; 5.2) through a general pairwise comparison based on the Kendall rank correlation coefficient, and 5.3) by evaluating their full ranking tendencies. The complete set of the results for each dataset is provided in the supplementary material of the article.

5.1. Illustrative example on the dataset fruit

The first analysis of the proposed sPWR approach was performed on

Table 5
Nutrient composition of some fresh and dried fruits listed in descending order of their energy values (Kcal). Nutrient values (per 100 g edible portion) are taken from the Italian CREA (Centro di Ricerca per gli Alimenti e la Nutrizione) database [30].

id	Fruit	Water (g)	Protein (g)	Fat (g)	Carbohydrate (g)	Energy (Kcal)	r_{KCAL}
1	Walnut dried (*Juglans regia*)	3.5	14.3	68.1	5.1	2936	1
2	Hazelnut (*Corylus avellana*)	4.5	13.8	64.1	6.1	2808	2
3	Pine nut (*Pinus pinea*)	4.3	31.9	50.3	4.0	2528	3
4	Chestnut dried (*Castanea vulgaris*)	10.1	6.0	3.4	62.0	1317	4
5	Raisin (*Vitis vinifera*)	17.1	1.9	0.6	72.0	1228	5
6	Fig dried (*Ficus carica*)	19.4	3.5	2.7	58.0	1179	6
7	Date (*Phoenix dactylifera*)	22.3	2.7	0.6	63.1	1131	7
8	Prune (*Prunus domestica*)	29.3	2.2	0.5	55.0	989	8
9	Peanut toasted (*Arachis hypogaea*)	2.3	29.0	50.0	8.5	620	9
10	Walnut (*Juglans regia*)	19.2	10.5	57.7	5.5	589	10
11	Chestnut (*Castanea vulgaris*)	55.8	2.9	1.7	36.7	174	11
12	Mandarin (*Citrus reticulata*)	81.4	0.9	0.3	17.6	76	12.5
13	Banana (*Musa sapientium*)	78.8	1.2	0.3	17.4	76	12.5
14	Persimmon (*Diospyros kaki*)	82.0	0.6	0.3	16.0	70	14
15	Pomegranate (*Punica granatum*)	80.5	0.5	0.2	15.9	68	15
16	Grape (*Vitis vinifera*)	81.3	0.5	0.1	15.6	64	16
17	Prickly pear (*Opuntia ficus-indica*)	83.2	0.8	0.1	13.0	63	17.5
18	Fig (*Ficus carica*)	81.9	0.9	0.2	14.2	63	17.5
19	Tangerine (*Citrus aurantium var. nobilis*)	85.3	0.8	0.2	12.8	57	19
20	Pear (*Pyrus communis var. William*)	85.0	0.3	0.5	10.2	52	20
21	Raspberry (*Rubus*)*idaeus*	84.6	1.0	0.6	6.5	49	21
22	Cherry (*Prunus avium*)	86.2	0.8	0.1	11.0	48	22
23	Apple (*Malus domestica var. Golden*)	86.9	0.4	0.1	10.7	46	23
24	Plum (*Prunus domestica*)	87.5	0.5	0.1	10.5	45	24
25	Sour cherry (*Prunus cerasus*)	85.2	0.8	0	10.2	44	25
26	Pineapple (*Ananas sativa*)	86.4	0.5	0	10.0	42	26.5
27	Apricot (*Prunus armeniaca*)	86.3	0.4	0.1	9.8	42	26.5
28	Quince (*Cydonia oblonga*)	84.3	0.3	0.1	6.3	38	28
29	Orange (*Citrus sinensis*)	87.2	0.7	0.2	7.8	37	29
30	Summer melon (*Cucumis melo*)	90.1	0.8	0.2	7.4	34	30
31	Loquat (*Eriobotrya japonica*)	88.3	0.4	0.4	6.1	32	31
32	Strawberry (*Fragaria vesca*)	90.5	0.9	0.4	5.3	30	32
33	Grapefruit (*Citrus paradise*)	91.2	0.6	0	6.2	29	33
34	Peach *(Prunus persica)*	88.2	0.7	0	5.8	28	34
35	Winter melon *(Cucumis melo var. Inodorus)*	94.1	0.5	0.2	4.9	24	35
36	Watermelon (*Citrusllus vulgaris*)	95.3	0.4	0	3.7	16	36.5
37	Lemon (*Citrus limon*)	91.5	0.6	0	2.5	16	36.5

Table 6
Characteristics of the 12 datasets used for the comparisons, showing the number of objects, the number of criteria and the optimality values. The term *all*+ indicates that all the variables have the optimality towards high values, while the term *all*– indicates that all the variables have the optimality towards low values.

id	Dataset	Objects	Criteria	Optimality	Ref.
1	Pesticides	17	6	*all* +	[31]
2	Winesmet	38	17	*all* +	[32]
3	Anilines	45	4	+ − + −	[6]
4	Refrigerants	40	3	*all* +	[33]
5	Baden pollution	59	4	*all* +	[34]
6	Bioaccumulation	168	3	*all* +	[35]
7	Countries	154	21	*all* +	[32]
8	Poly-pharmacology	55	7	*all* –	[36]
9	CMethods	10	32	*all* +	[18]
10	Centrality	6	3	*all* +	[37]
11	Air monitoring	15	5	*all* +	[38]
12	Endocrine disruptors	30	7	*all* +	[39]

a simple and intuitive dataset (Table 5) that was generated by selecting 37 common fresh and dried fruits whose basic nutritional composition and corresponding energy content (Kcal) were retrieved from the Italian CREA database [30]. The fruits are in descending order from the highest to the lowest energy value (reference ranking, r_{KCAL}). The contents of water, protein, fat, and carbohydrates, expressed in grams per 100 g of edible portion, were used as the four criteria for ranking.

Similarity relationships among the fruits were visualized using a Minimum Spanning Tree (MST) computed on the four nutritional criteria (Fig. 2). The MST was derived from a Euclidean distance matrix after range-scaling the variables. In the tree plot, each node represents a

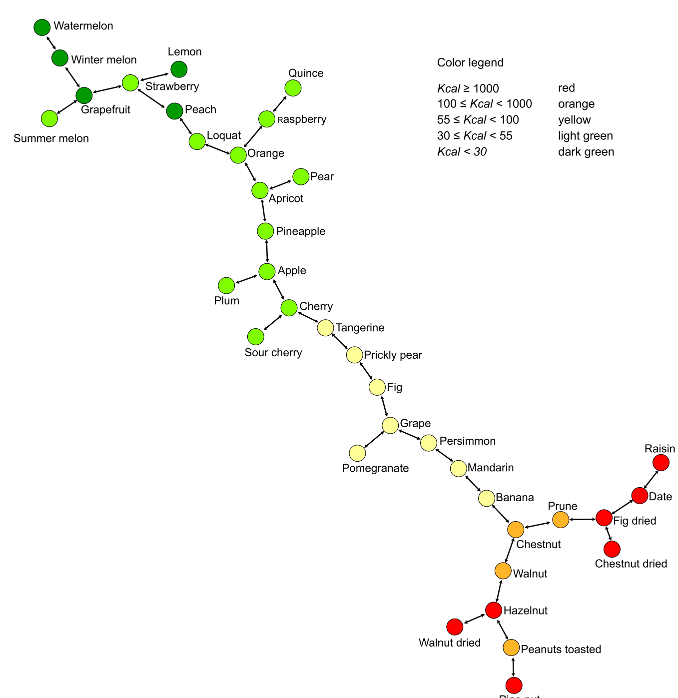

Fig. 2. Minimum Spanning Tree (MST) calculated on the 37 fruits (Table 5) characterized by content of water, protein, fat and carbohydrates. The tree nodes are coloured according to the energy value scale.

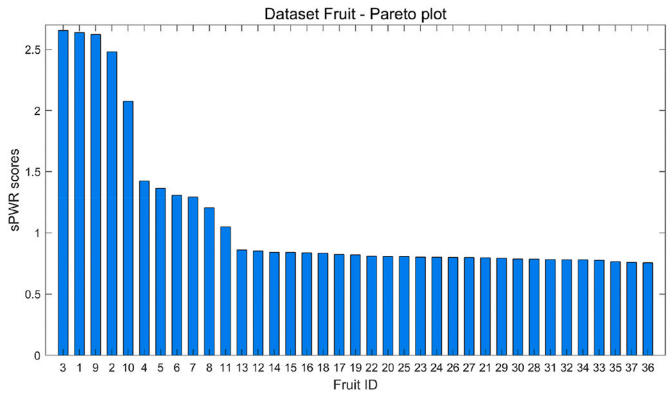

Fig. 3. Pareto plot of the 37 fruits (Table 5) ordered by the sPWR scores.

fruit and is coloured according to the respective energy content. From this plot, the relationship between nutritional composition (i.e., the selected criteria) and energy value (i.e., fruit ranking) is clearly reflected in the colour gradient along the tree structure.

All calculations were performed by assigning different weights to the four criteria according to their relative importance as suggested by nutritional experts: 1 for water, 2 for protein and carbohydrates, and 4 for fat. In addition, to ensure that the most energetic fruits ranked highest, lower water content and higher amounts of protein, fat, and carbohydrates were defined as the optimal values.

The resulting sPWR scores are visualized in the Pareto plot (Fig. 3). The top 11 fruits, although separated into two distinct groups, correspond to those with the highest energy content.

The sPWR ranking was also compared with those obtained by the other methods, which are shown in Table 7, together with the reference ranking (r_{KCAL}) derived from the energy values (Kcal) of fruits. The first row of Table 7 includes the Kendall correlations between each method ranking and the reference ranking.

The Kendall τ between two series j and k of ranks is defined by the comparison of all the pairs of a series with the corresponding pairs of the other series:

$$\tau_{jk} = \frac{2 \cdot F_{jk}}{n \cdot (n-1)} \qquad -1 \leq \tau_{jk} \leq +1 \tag{20}$$

where F_{jk} is the algebraic sum of ± 1 calculated as:

$$F_{jk} = \sum_{i=1}^{n-1} \sum_{t=i+1}^{n} sign(r_{ij} - r_{tj}) \cdot sign(r_{ik} - r_{tk}) \tag{21}$$

where r_{ij} and r_{tj} are the ranks of the objects i and t in the j-th series and r_{ik} and r_{tk} are the ranks of the same objects in the k-th series; the function *sign* takes a value of $+1$ if its argument is positive and a value -1 if its argument is negative.

The high correlation of sPWR (i.e., 0.919) indicates a remarkable accuracy. A similar accuracy was also obtained by CS1 and TOPSIS3. Although, DOM, PWR, SAR and CS2 show lower correlations, it is worth noting that all the ranking methods were able to rank the most energetic

Table 7
Ranking of 37 fruits on their nutrient composition (i.e., content of water, protein, fat, carbohydrates) provided by ten diverse MCDMs. Criteria were weighted according to their relative importance: 1 for water, 2 for protein and carbohydrate, 4 for fat. For each method, the Spearman correlation coefficient (τ) between the calculated and the reference ranking (r_{KCAL}) is reported in the first row of the table.

Fruit	τ	0.919	0.741	0.744	0.734	0.902	0.898	0.910	0.919	0.744	0.895
	r_{KCAL}	*sPWR*	*PWR*	*SAR*	*DOM*	*TOPSIS1*	*TOPSIS2*	*TOPSIS3*	*CS1*	*CS2*	*AHP*
Walnut dried	1	2	9	8	9	3	2	3	2	8	3
Hazelnut	2	4	8	7	8	4	4	4	4	7	4
Pine nut	3	1	11	10.5	12	1	3	2	1	10.5	1
Chestnut dried	4	6	1	1	1	7	7	6	6	1	6
Raisin	5	7	3	4	5	6	6	7	7	4	7
Fig dried	6	8	2	2	2	9	9	9	8	2	9
Date	7	9	5	5.5	4	8	8	8	9	5.5	8
Prune	8	10	7	9	6	10	10	10	10	9	10
Peanut toasted	9	3	6	3	7	2	1	1	3	3	2
Walnut	10	5	10	10.5	10	5	5	5	5	10.5	5
Chestnut	11	11	4	5.5	3	11	11	11	11	5.5	11
Mandarin	12.5	13	13	13	13	13	12	13	13	13	13
Banana	12.5	12	12	12	11	12	13	12	12	12	12
Persimmon	14	14.5	15	15	16	14.5	14	15	14.5	15	15
Pomegranate	15	14.5	18	18	18	14.5	15	14	14.5	18	15
Grape	16	16	24	25	25	16	16	16	16	25	15
Prickly pear	17.5	18	21	21	19	18	18	18	18	21	18
Fig	17.5	17	16	16	15	17	17	17	17	16	17
Tangerine	19	19	17	17	17	19	19	19	19	17	19
Pear	20	21.5	19	19	21	24	24	21	21.5	19	23.5
Raspberry	21	27	14	14	14	28	29	27	26.5	14	30
Cherry	22	20	22	22.5	22	20	20	20	20	22.5	20
Apple	23	23	29	29	29	21	21	23	24	29	23.5
Plum	24	24	27	27	27	22.5	22	24	24	27	23.5
Sour cherry	25	21.5	28	28	28	22.5	23	22	21.5	28	23.5
Pineapple	26.5	25	33	33	33	25	25	25	24	33	23.5
Apricot	26.5	26	30	30	30	26	26	26	26.5	30	23.5
Quince	28	30	32	32	32	30	30	28	30	32	30
Orange	29	28	25	24	23.5	27	27	29	28	24	27
Summer melon	30	29	23	22.5	23.5	29	28	30	29	22.5	28
Loquat	31	31	26	26	26	31	32	31	31	26	30
Strawberry	32	32.5	20	20	20	34	34	34	32.5	20	33.5
Grapefruit	33	34	35	35	35	32.5	31	33	34	35	33.5
Peach	34	32.5	34	34	34	32.5	33	32	32.5	34	32
Winter melon	35	35	31	31	31	35	35	35	35	31	35.5
Watermelon	36.5	37	37	37	37	36.5	36	37	37	37	37
Lemon	36.5	36	36	36	36	36.5	37	36	36	36	35.5

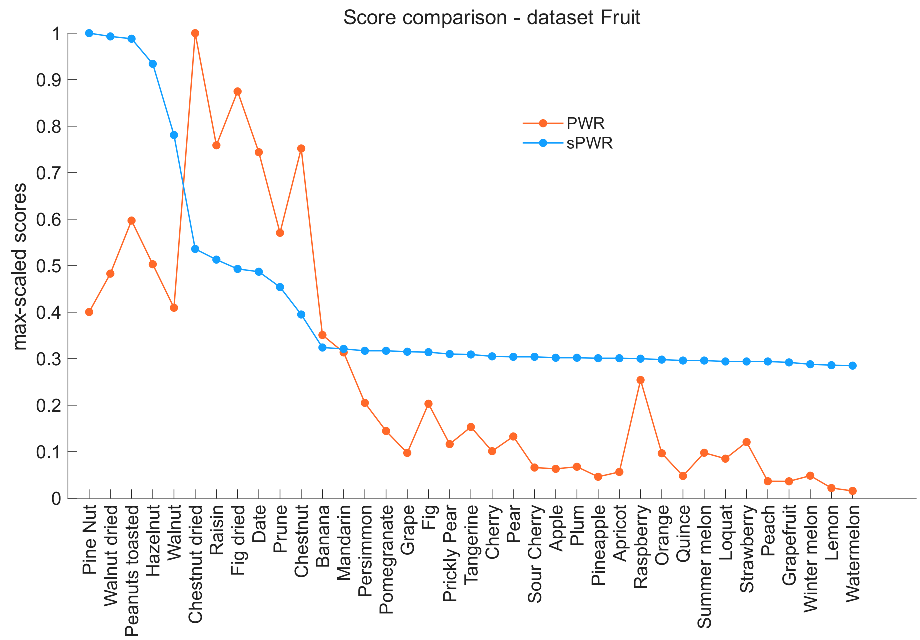

Fig. 4. Comparison of the max-scaled scores obtained by PWR (orange) and sPWR (blue) methods for the dataset *Fruit*. Objects are ordered on the sPWR ranks, assumed as the reference.

fruits in the top positions of the list.

The specific differences between PWR and sPWR are highlighted by the max scaled scores in Fig. 4 where results are ordered according to the ranking provided by sPWR. It can be easily observed that there is an order inversion of some fruits ranked in the first positions by the two methods; indeed, sPWR ranks first the group of dried non-sweet fruits (such as pine nuts, walnuts, hazelnuts, peanuts), while PWR ranks first the group of dried sweet fruits (such as chestnuts, figs dried, raisin, dates). This inversion can be explained by the greater importance given by sPWR to the differences in fat content than to those in carbohydrate contents.

Moreover, the effects of the smoothing action of sPWR with respect to the original proportional PWR approach are largely evident in the last portion of the ranking where the score differences are negligible for sPWR.

5.2. Multivariate comparison of MCDM methods

The general comparison of the 10 different MCDM methods studied in this work was performed on the 12 selected datasets (Table 6). The first simple pairwise comparison was performed by using the Kendall τ correlation between the final rankings provided by each method for each dataset.

Table 8 shows the average Kendall correlation between all the pairs of considered MCDM methods evaluated on the 12 datasets. Two main groups of highly correlated methods can be observed: a first group includes sPWR, CS1, TOPSIS3, AHP, TOPSIS2 and TOPSIS1, while a second group includes PWR, DOM, SAR and CS2. It is worth noting that the average correlation of sPWR with PWR (0.777) is among the lowest ones in the whole comparison.

To obtain a deeper insight into the relationships of sPWR with the other methods, Principal Component Analysis (PCA) was performed on the autoscaled data matrix (10 × 12) being comprised of the pairwise rank correlations of sPWR with the other methods on the 12 selected datasets. Obviously, sPWR, which is described by a 12-dimensional vector of 1, represents the reference for this analysis. The obtained results are shown in Fig. 5. The following colours have been assigned to the different methods: green colour to sPWR, PWR and AHP; orange to the TOPSIS methods; blue to SAR and DOM methods; black to CS1 and CS2 Copeland scores.

The first principal component PC1 measures the similarity degree of the MCDM methods to sPWR. As it can be observed in Fig. 5a, the ranks obtained by CS1 and TOPSIS3 are the most similar to those obtained by sPWR, even the AHP ranks are quite similar; the other approaches are significantly different from the first ones and constitute two different clusters. The lower cluster is constituted by the methods that perform proportional pairwise comparisons (i.e. DOM, SAR and the original PWR method) together with the Copeland score CS2 obtained on the PWR tournament table. Indeed, these methods are based on the same philosophy, whereby each object is assigned a score based on the weights of each criterion, simply following the rule “is better than”, without considering the numerical difference of the criterion between objects. The other cluster includes TOPSIS1 and TOPSIS2, which significantly differ from TOPSIS3 in spite they differ only by the initial scaling procedures. Fig. 5b shows the datasets that are more responsible for the differences in the similarities with sPWR of TOPSIS1 and TOPSIS2 on one side and PWR, SAR, DOM and CS2 on the other side.

5.3. Comparison of the full ranking tendency of the methods

An additional aspect of the ranking approach to be evaluated is its tendency to produce a complete ordering of the alternatives, that is, to rank all objects without ties. To assess this tendency across all the considered MCDM methods, the standardized Shannon entropy (H^*) was calculated on the ranks provided by each method for all the 12 datasets in analysis. The results of this analysis are shown as box-plots in Fig. 6.

The standardized Shannon entropy measures the mean information content of a ranking over the maximum available information content (i.e. $log_2\, n$):

$$H^* = \frac{-\sum_{g=1}^{G} p_g \cdot log_2\left(p_g\right)}{log_2\, n} \qquad \sum_{g=1}^{G} p_g = 1 \qquad 0 \le H^* \le 1$$

where G is the number of distinct elements (i.e., distinct rank values), p_g is the relative frequency of each element and n is the number of observations (i.e., the objects). In the case of a tie-free ranking, the relative frequency for each rank value is $p_g = 1/n \quad (g = 1, \ldots, n)$ and the entropy

Table 8
Average Kendall τ correlation coefficients for the pairwise comparison of the ten methods. All the correlation values greater than 0.800 are highlighted in bold.

	sPWR	PWR	SAR	DOM	TOPSIS1	TOPSIS2	TOPSIS3	CS1	CS2	AHP
sPWR	1	0.777	0.783	0.771	**0.811**	**0.819**	**0.956**	**0.996**	0.788	**0.915**
PWR	0.777	1	**0.963**	**0.953**	0.736	0.716	0.766	0.776	**0.987**	0.787
SAR	0.783	**0.963**	1	**0.973**	0.720	0.703	0.770	0.781	**0.975**	0.789
DOM	0.771	**0.953**	**0.973**	1	0.714	0.698	0.758	0.771	**0.949**	0.772
TOPSIS1	**0.811**	0.736	0.720	0.714	1	**0.851**	**0.813**	**0.810**	0.737	**0.808**
TOPSIS2	**0.819**	0.716	0.703	0.698	**0.851**	1	**0.826**	**0.819**	0.719	0.799
TOPSIS3	**0.956**	0.766	0.770	0.758	**0.813**	**0.826**	1	**0.955**	0.777	**0.918**
CS1	**0.996**	0.776	0.781	0.771	**0.810**	**0.819**	**0.955**	1	0.787	**0.919**
CS2	0.788	**0.987**	**0.975**	**0.949**	0.737	0.719	0.777	0.787	1	0.796
AHP	**0.915**	0.787	0.789	0.772	**0.808**	0.799	**0.918**	**0.919**	0.796	1

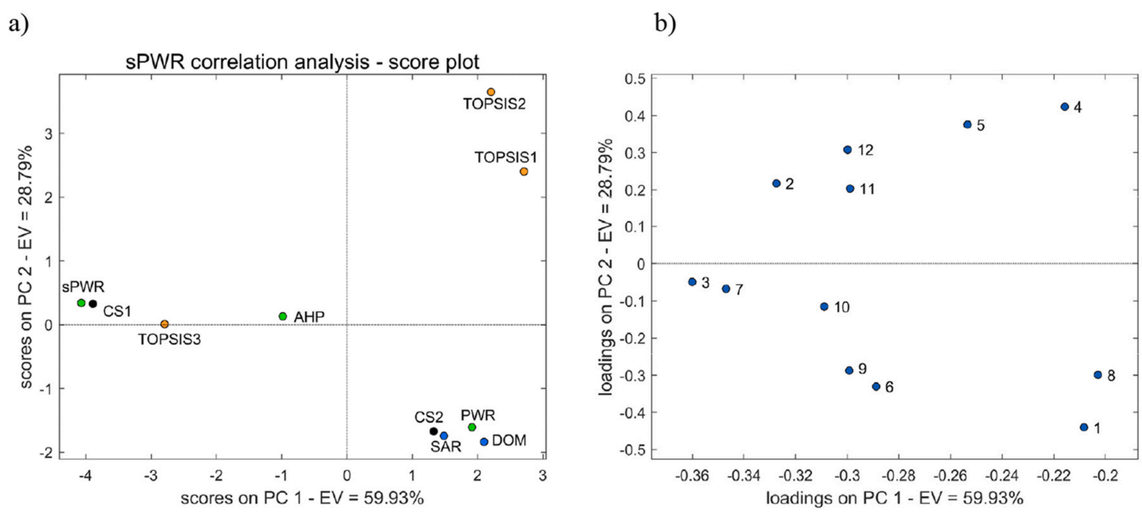

Fig. 5. The scores (a) and loadings (b) of the first two principal components on the Kendall correlations between sPWR and the other approaches. The total explained variance EV is 85.6 %.

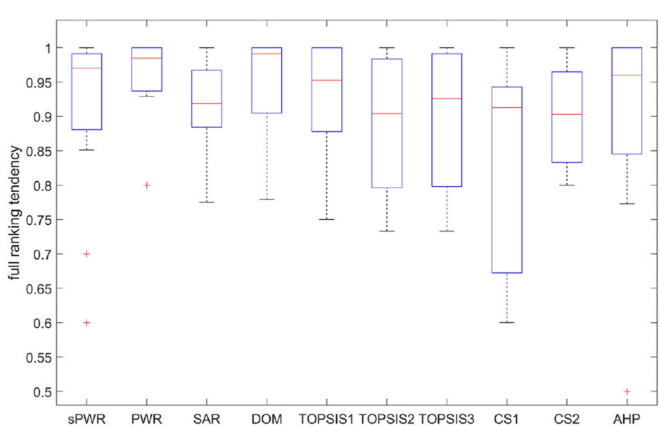

Fig. 6. Boxplot of the standardized Shannon entropy of the rankings produced by the MCDM methods over the 12 evaluated datasets. The red line denotes the median entropy, the box its inter-dataset variability, and the whiskers and outlier points below the box highlight datasets with a greater occurrence of ties (e.g., datasets 4 and 11 for sPWR).

$H^* = 1$ indicates a full ranking, while $H^* = 0$ indicates that all the objects have the same unique rank, that is, $G = 1$ and $p_g = n/n = 1$.

Among the considered methods, DOM, PWR and sPWR show the strongest tendency to produce full rankings, assigning a distinct rank to each object in several cases. On the opposite side, SAR tends to assign more ties among objects. A similar, though less pronounced, tendency is observed for Copeland scores and TOPSIS variants, which also show greater variability across the datasets. It can be noted that both sPWR and PWR show a full ranking tendency greater than the two corresponding Copeland scores (CS1 and CS2): indeed, the latter are based on the simply row sums of the corresponding tournament tables, while their spectral decompositions allow to remove some draws.

It is important to note that the tendency to produce a full ranking is not necessarily a desirable characteristic of the method. High entropy, which indicates a strong inclination toward full rankings, may arise either from a method that effectively captures also very small differences between alternatives, or from one that fails to identify ties where they are warranted. Therefore, the tendency to produce full rankings should be viewed more as a characteristic of a method than as an indicator of its quality. Obviously, a very low full ranking tendency is not a suitable property for a ranking method.

6. Conclusions

The smoothed Power-Weakness Ratio (sPWR) is a new informative method for multi-criteria decision making; it is based on a unique well-defined algorithm for the construction of the tournament table in such a way the values of the table are sensitive to the quantitative differences of the criteria. The sPWR score is obtained by applying the double spectral decomposition of the tournament table, as in the PWR approach. However, sPWR provides a higher information content for the subsequent ranking of objects than methods that do not account for quantitative differences between alternatives. The sPWR performance on the Fruit dataset with a known reference ranking was quite satisfactory with high accuracy.

We compared sPWR with nine benchmark MCDM methods, which were calculated over 13 different datasets taken from real-world problems. The Copeland score CS1 also served as a fairly good ranking estimator, although it showed a lower full ranking tendency than sPWR, characteristic which is partially shared with CS2, TOPSIS2, TOPSIS3 and SAR. The similar behaviour of CS1 and sPWR is due to their common origin from the same tournament table.

The greater similarity between sPWR and TOPSIS3 rankings, compared with the other TOPSIS variants, is likely due to the common scaling procedure adopted by these two algorithms. The DOM and PWR methods produced similar results, as both rely on the same pairwise-comparison rules and differ only in how the final scores are calculated. SAR, which is based on average ranks, also achieved results similar to those of DOM and PWR and different from the results obtained by the remaining methods. In contrast, the results produced by the various TOPSIS approaches were quite divergent, highlighting the strong dependence of their results on the choice of scaling method.

In conclusion, the novel sPWR method was specifically designed to handle quantitative variables, for which numerical differences can be relevant to determining final preferences. As in the original PWR framework, the concepts of power and weakness are intuitive and their ratio provides a direct measure of how strongly an alternative outperforms the other alternatives.

The sPWR score allows a finer differentiation among alternatives, thus reducing the number of ties. Correlations among criteria are taken into account by the spectral decomposition of the pairwise comparison matrix. Moreover, unlike the TOPSIS method, which relies solely on distances to ideal and non-ideal solutions, sPWR more effectively captures how alternatives are distributed within the criteria space, and its results are invariant to variable scaling. The major shortcoming of sPWR is that it can not handle qualitative/nominal criteria.

CRediT authorship contribution statement

Viviana Consonni: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Davide Ballabio:** Writing – review & editing, Visualization, Validation. **Enmanuel Cruz Muñoz:** Writing – review & editing, Validation, Data curation. **Veronica Termopoli:** Writing – review & editing, Visualization, Data curation. **Roberto Todeschini:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors have declared no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2025.105624.

Data availability

I have shared data and software code by a link in the article

References

[1] M. Dean, A Practical Guide to Multi-Criteria Analysis, University College of, London (UK), 2022.
[2] M. Pavan, R. Todeschini (Eds.), Scientific Data Ranking Methods: Theory and Applications, Elsevier, Amsterdam (The Netherlands), 2008.
[3] M.M.W.B. Hendriks, J.H. de Boer, A. Smilde, D.A. Doornbos, Multicriteria decision making, Chemometr. Intell. Lab. Syst. 16 (1992) 175–191.
[4] A. Kumar, B. Sah, R. Singh, Y. Deng, X. He, P. Kumar, R.C. Bansal, A review of multi criteria decision making (MCDM) towards sustainable rebewable energy development, Renew. Sustain. Energy Rev. 69 (2017) 596–609.
[5] F. Grisoni, V. Consonni, S. Nembri, R. Todeschini, How to weight Hasse matrices and reduce incomparabilities, Chemometr. Intell. Lab. Syst. 147 (2015) 95–104.
[6] L. Carlsen, A combined QSAR and partial order ranking approach to risk assessment, SAR QSAR Environ. Res. 17 (2006) 133–146.
[7] E. Nwokoaghara, A.K. Olaleye, M. Wang, Biodisel from microalgae: the use of multi-criteria decision analysis for strain selection, Fuel 159 (2015) 241–249.
[8] I. Ivlev, J. Vacek, P. Kneppo, Multi-criteria decision analysis for supporting the selection of medical devices under uncertainty, Eur. J. Oper. Res. 247 (2015) 216–228.
[9] G. Ferla, B. Mura, S. Falasco, P. Caputo, A. Matarazzo, Multi-criteria decision analysis (MCDA) for sustainability assessment in food sector. A systematic literature review on methods, indicators and tools, Sci. Tot. Environ. 946 (2024) 17423.
[10] M. Cassotti, F. Grisoni, S. Nembri, R. Todeschini, Application of the weighted power-weakness ratio (wPWR) as a fusion rule in ligand-based virtual screening, MATCH Commun, Math. Comput. Chem. 76 (2016) 359–376.
[11] L. Leydesdorff, W. De Nooy, L. Bornmann, G. Prathap, The "Tournaments" metaphor in citation impact studies: power-weakness ratios (PWR) as a journal indicator, ArXiv 1411 (2015) 0906v2.
[12] Z. Turskis, E.K. Zavadskas, F. Peldshus, Multi-criteria optimization system for decision making in construction design and management, Eng. Econ. 1 (2009) 7–17.
[13] G. Jimenez, G. Santos, M. Félix, H. Hernandez, C. Rondon, Good practices and trends in reverse logistics in the plastic products manufacturing industry, Procedia Manuf. 41 (2019) 367–374.
[14] S. Carpitella, V. Kratochvil, M. Pištěk, Multi-criteria decision making beyond consistency: an alternative to AHP for real-world industrial problems, Comput. Ind. Engin. 198 (2024) 110661.
[15] S. Greco, M. Ehrgott, J. Figueira (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, second ed., Springer, New York (USA), 2016, p. 1347.
[16] M. Madanchian, H. Taherdoost, A comprehensive guide to the TOPSIS method for multi-criteria decision making, Sustain. Soc. Develop. 1 (2023) 1–6.
[17] T.L. Saaty, The Analytical Hierarchy Process, McGraw-Hill, New York (USA), 1980.
[18] R. Todeschini, F. Grisoni, S. Nembri, Weighted power-weakness ratio for multi-criteria decision making, Chemometr. Intell. Lab. Syst. 146 (2015) 329–336.
[19] H.A. David, Ranking the players in a round robin tournament, Rev. Int. Stat. Inst. 39 (1971) 137–147.
[20] C. Ramanujacharyulu, Analysis of preferential experiments, Psychometrica 29 (1964) 257–261.
[21] I. Chen, Tournament matrices, document on web at, http://buzzard.ups.edu/courses/2014spring/420s2014.html, 2014 (accessed September 2025).
[22] J.P. Keenr, The Perron-Frobenius theorem and the ranking of footbal teams, SIAM Rev. 35 (1993) 80–93.
[23] M.G. Kendall, On the method of paired comparisons, Biometrika 33 (1940) 239–251.
[24] M.G. Kendall, Further contributions to the theory of paired comparisons, Biometrics 11 (1955) 43–62.
[25] T.-H. Wei, The Algebraic Foundation of Ranking Theory, University of, Cambridge (UK), 1952.
[26] H.J. Zimmermann, L. Gutsche, Multi-Criteria Analyse, Springer, Berlin (GER), 1991.
[27] A.H. Copeland, A reasonable social welfare function, in: Seminar on Applications of Mathematics on the Social Sciences, University of Michigan, USA, 1951.
[28] W. Han, A. Van Deemen, The copeland ratio ranking for abstract decision problems, Eur. J. Oper. Res. 323 (2025) 966–973.
[29] H.-S. Shih, H. Shyur, E.S. Lee, An extension of TOPSIS for group decision making, Math. Comput. Model. 45 (2007) 801–813.
[30] CREA database. https://www.alimentinutrizione.it/tabelle-nutrizionali/ricerca-per-alimento (accessed November 2025).
[31] A. Manganaro, D. Ballabio, V. Consonni, A. Mauri, M. Pavan, R. Todeschini, The DART (decision analysis by ranking techniques) software, in: R. Todeschini, M. Pavan (Eds.), Scientific Data Ranking Methods: Theory and Applications, Elsevier, Amsterdam (The Netherlands), 2008.
[32] R. Todeschini, F. Grisoni, D. Ballabio, Deep ranking analysis by power eigenvectors (DRAPE): a wizard for ranking and multi-criteria decision making, Chemometr. Intell. Lab. Syst. 191 (2019) 129–137.
[33] G. Restrepo, M. Weckert, R. Bruggemann, S. Gerstmann, H. Frank, Ranking of refrigerants, Environ. Sci. Technol. 42 (2008) 2925–2930.
[34] G. Restrepo, R. Bruggemann, Ranking regions using cluster analysis, Hasse diagram technique and topology, Int. Congr. Environ. Model. Softw. (2006) 3–7.
[35] F. Grisoni, V. Consonni, M. Vighi, Detecting the bioaccumulation patterns of chemicals through data-driven approaches, Chemosphere 208 (2018) 273–284.
[36] C. Valsecchi, D. Ballabio, V. Consonni, R. Todeschini, Deep ranking analysis by power eigenvectors (DRAPE): a polypharmacology case study, Chemometr. Intell. Lab. Syst. 203 (2020) 104001.
[37] J. Hu, Y. Du, H. Mo, D. Wei, Y. Deng, A modified weighted TOPSIS to identify influential nodes in complex networks, Physica A 444 (2016) 73–85.
[38] K. Voigt, G. Welz, R. Bruggemann, Data analysis of environmental air pollutant monitoring systems in Europe, Environmetrics 15 (2004) 577–596.
[39] L. Carlsen, R. Bruggemann, Y. Sailaukhanuli, Application of selected partial order tools to analyze fate and toxicity indicators of environmentally hazardous chemicals, Ecol. Indic. 29 (2013) 191–202.