Science, Engineering and Technology Vol. 6, No. S1, pp. 145-169
www.setjournal.com https://doi.org/10.54327/set2026/v6.is1.407

Comparing Robustness-Oriented Multi-Criteria Decision-Making Methods: A Cross-Domain Benchmark of MEGAN and Recent Aggregation-Based Approaches in Engineering Selection Problems

Ton Nguyen Trong Hien[1], Khoiriya Latifah[2]

[1]*Faculty of Business Administration, Van Lang University, Ho Chi Minh City, Vietnam.*
[2]*Faculty of Informatics and Engineering, Universitas PGRI Semarang, Semarang City, Indonesia.*

Abstract

Robustness-oriented multi-criteria decision-making (MCDM) methods aim to produce rankings that remain stable against the choice of normalisation and weighting. Yet, new members of this family are rarely validated independently. This study presents an independent cross-domain benchmark of MEGAN (Multi-criteria Evaluation via Gradual-weighting and Aggregation of Normalised distance matrices) against three established robustness-oriented methods, the Combined Compromise Solution (COCOSO), the Alternative Ranking Order Method Accounting for Two-Step Normalisation (AROMAN), and the Root Assessment Method (RAM), and four conventional reference methods (TOPSIS, VIKOR, WASPAS, EDAS). The benchmark pairs the synthetic renewable energy grid dataset of the original MEGAN study, used for verification, with a real-world flotation machine selection problem. Robustness claims were tested under systematic normalisation and weighting substitution, Monte Carlo weight perturbation, and leave-one-out rank reversal. Agreement proved strongly dataset-dependent: MEGAN aligned with the conventional methods on the flotation problem (Spearman near 0.70) but produced near-opposite rankings on the energy problem (−0.84), and the claimed normalisation and weighting insensitivity was not reproduced, while rank-reversal resistance held fully. A formal analysis shows that the second direction-unification step of MEGAN exactly cancels the normalisation's direction handling, so every cost criterion is effectively treated as a benefit criterion. Removing this redundant step yields the corrected variant MEGAN-S, which raised the mean rank correlation from −0.29 to +0.60 and from +0.61 to +0.97 on the two problems, while retaining full rank-reversal resistance. Robustness of an MCDM method is thus better understood as a property of the problem structure than a fixed guarantee.

Keywords: Benchmarking, Multi-Criteria Decision-Making (MCDM), MEGAN, Robustness, Normalisation sensitivity, Rank reversal.

1. Introduction

Decision-making in contemporary engineering rarely reduces to the optimisation of a single objective. Engineers must routinely weigh mechanical performance against cost, environmental footprint against manufacturability, and short-term feasibility against long-term reliability. Multi-Criteria Decision-Making (MCDM) provides a structured mathematical framework for ranking a finite set of discrete alternatives when several conflicting criteria must be considered simultaneously [1].

Corresponding author: Ton Nguyen Trong Hien (hien.tnt@vlu.edu.vn)
Received: 30 June 2026; Revised: 31 July 2026; Accepted: 27 August 2026; Published: 11 September 2026

Over the past five decades, MCDM has become a central tool across engineering design, operations research, energy planning, and environmental management, and the body of available methods has grown to encompass more than two hundred distinct algorithms, each resting on its own theoretical assumptions, computational procedures, and behavioural properties [2]. This proliferation of methods has produced a well-documented difficulty. When different MCDM algorithms are applied to the same decision problem, they frequently yield different rankings of the same alternatives, because each method embeds particular choices regarding normalisation, aggregation, and the treatment of beneficial and non-beneficial criteria [3]. Two sources of instability are especially consequential for the engineering practitioner. The first is sensitivity to the normalisation technique, whereby a method may reverse its ranking depending on whether linear, vector, or sum-based normalisation is applied to the decision matrix. The second is the rank reversal phenomenon, in which the relative ordering of alternatives changes when an alternative is added to or removed from the candidate set. Both phenomena have attracted renewed attention in the recent benchmarking literature: Sałabun et al. [4] showed through large-scale simulation that established methods disagree systematically as a function of problem characteristics, Shyur and Shih [5] traced rank reversal in the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) to the interaction of distance metrics with the normalisation technique, and Trung et al. [6] documented how strongly the choice of normalisation alone can reshape rankings in engineering applications. A further structural concern is that conventional methods rely on a single fixed weight vector, which captures only a static snapshot of the decision landscape and fails to reflect how ranking stability evolves as criteria priorities shift. Together, these limitations undermine the reliability of MCDM conclusions and leave practitioners with insufficient guidance for selecting an appropriate method.

In response to these concerns, a distinct family of recent MCDM methods has emerged whose defining objective is robustness rather than introducing a new ranking principle. These robustness-oriented, aggregation-based methods seek to produce stable rankings by combining multiple normalisation schemes, multiple weighting scenarios, or multiple intermediate utility measures into a single aggregated outcome. The Combined Compromise Solution method (COCOSO), proposed by Yazdani et al. [7], integrates three distinct aggregation strategies into a combined compromise score. The Alternative Ranking Order Method Accounting for Two-Step Normalisation (AROMAN), proposed by Bošković et al. [8], averages the results of linear and vector normalisation in a deliberate two-step procedure intended to yield more robust rankings. The Root Assessment Method (RAM), proposed by Sotoudeh-Anvari [9], aggregates the weighted scores of beneficial and non-beneficial criteria through a root-based function designed for reliability and computational simplicity. The most recent addition to this family is MEGAN, the Multi-Criteria Evaluation via Gradual-Weighting and Aggregation of Normalised Distance Matrices, introduced by Cebeci in April 2026 [10]. MEGAN perturbs the initial weight vector across a predefined range, ranks the alternatives under each resulting scenario, aggregates these rankings through the Borda Count method, and applies a dynamic threshold derived from the dispersion of superiority scores. Its proponents claim that this design confers robustness against the choice of normalisation technique, robustness against the choice of weighting method, and resistance to rank reversal.

Despite the shared robustness objective of these methods, two gaps in the literature motivate the present study. The first concerns independent and comparative validation. Newly proposed MCDM algorithms are typically validated by their own proponents, often on bespoke or synthetic datasets, with limited subsequent benchmarking by independent researchers and, in particular, limited direct comparison among methods of the same family [2]. In the specific case of MEGAN, the original article validated the algorithm on a synthetic renewable energy grid dataset and explicitly identified validation on real-world data as a direction for future work [10]. To the best of our knowledge, the robustness-oriented methods COCOSO, AROMAN, RAM, and MEGAN have not been benchmarked against one another under a common experimental protocol; consequently, it remains unknown whether they agree in their rankings and whether their respective robustness claims hold to a comparable degree. The second gap concerns cross-domain generality. Robustness properties demonstrated

on a single dataset, in a single application domain, cannot be assumed to transfer to problems exhibiting different scales, different criterion structures, and different degrees of unit heterogeneity. Whether the conclusions drawn for one engineering domain hold in another therefore remains an open empirical question.

The present study addresses these gaps through a cross-domain benchmark of MEGAN against three established robustness-oriented, aggregation-based methods, COCOSO, AROMAN, and RAM, with TOPSIS [11], VIseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR) [12], the Weighted Aggregated Sum Product Assessment (WASPAS) [13], and the Evaluation based on Distance from Average Solution (EDAS) [14] included as conventional reference methods to position the four robustness-oriented methods within the broader MCDM landscape. The benchmark is conducted on two engineering selection problems drawn from distinct domains: a published synthetic renewable energy grid selection benchmark, adopted from the original MEGAN study, and a real-world flotation machine selection problem from mineral processing practice. The specific objectives of the study are threefold. The first objective is to determine the degree of ranking agreement among MEGAN and the comparator robustness-oriented methods, and between this family and the conventional reference methods. The second objective is to evaluate, under a common protocol, the extent to which the robustness claims associated with these methods, specifically insensitivity to normalisation and weighting and resistance to rank reversal, hold empirically. The third objective is to assess whether the patterns observed for one engineering domain are reproduced in the other, thereby testing the cross-domain generality of the findings. Accordingly, the study is guided by three research questions. RQ1 asks to what extent the robustness-oriented methods produce mutually consistent rankings and how closely they align with the conventional reference methods. RQ2 asks whether the normalisation insensitivity, weighting insensitivity, and rank reversal resistance attributed to these methods are confirmed under systematic perturbation. RQ3 asks whether the answers to RQ1 and RQ2 remain stable when the analysis is transferred across engineering domains.

To remove any ambiguity about the nature and hierarchy of the contributions, the novelty of this study is stated explicitly. The primary contribution is an independent, domain-agnostic validation protocol that combines concordance-based agreement analysis, systematic normalisation and weighting perturbation, Monte Carlo weight perturbation, and leave-one-out rank reversal testing. It can be reused to audit the robustness claims of any aggregation-based MCDM method. The second contribution is the empirical and theoretical diagnosis obtained by applying this protocol to MEGAN: the study shows that the advertised insensitivity claims do not hold on the benchmark data and proves that the responsible mechanism is a redundant direction-unification step that silently discards the orientation of cost criteria (Proposition 1). The third contribution follows from the diagnosis: a minimally modified, direction-consistent variant, MEGAN-S, whose formal properties are established in Proposition 2 and which is evaluated under the same protocol as the methods it is meant to repair. The three contributions are ordered by dependence: the protocol yields the diagnosis, and the diagnosis yields the variant. The benchmark is therefore offered as a reusable instrument and a design principle for the robustness-oriented MCDM family, rather than as an assessment of a single method. Wherever a robustness claim is found not to hold, the analysis isolates the structural mechanism responsible instead of merely recording the failure, so that the finding generalises to the broader class of methods that share that mechanism. The variant realigns the divergent rankings with the established methods while, on the hardest problem, becoming more rather than less sensitive to perturbation.

The remainder of the paper is organised as follows. Section 2 describes the datasets, formalises the evaluated methods, introduces MEGAN-S together with its theoretical justification, and specifies the validation protocol. Section 3 reports and discusses the results. Section 4 concludes by stating the limitations of the study and outlining future work.

Nomenclature and Abbreviations

Consistent with the journal template, the main mathematical symbols and abbreviations used in this article are collected below; each is also defined at its first mention in the text. This section, like the Abstract and References, is excluded from the section numbering.

Symbol	Definition
X, x_{ij}	decision matrix; performance of alternative i on criterion j
n, m	number of alternatives; number of criteria
v, v_j	benefit-cost vector, $v_j = 1$ (benefit) or -1 (cost)
w, w_j	criterion weight vector, $\sum_j w_j = 1$
N, n_{ij}	normalised decision matrix and its entries
Y	weighted normalised matrix
U	direction-adjusted matrix (MEGAN)
D, d_{ij}	distance matrix to the criterion-wise minimum
s_i, p_i	superiority score and its percentage form
τ	dynamic ranking threshold
r, r_l, L	gradual-weighting rate vector, one rate, number of rates
$R_i^{(l)}, b_i$	rank of alternative i in scenario l; Borda score
ρ	Spearman rank correlation coefficient
W	Kendall coefficient of concordance
χ_F^2, F_{ID}	Friedman statistic; Iman-Davenport statistic
u_j	Monte Carlo weight perturbation, $u_j \sim U(-0.25,0.25)$

Abbreviation	Meaning
AROMAN	Alternative Ranking Order Method Accounting for Two-Step Normalisation
COCOSO	Combined Compromise Solution
CRITIC	Criteria Importance Through Intercriteria Correlation
EDAS	Evaluation based on Distance from Average Solution
JSD, KL	Jensen-Shannon divergence; Kullback-Leibler divergence
MC, CI, SD	Monte Carlo; confidence interval; standard deviation
MCDM	Multi-Criteria Decision-Making
MEGAN	Multi-criteria Evaluation via Gradual-weighting and Aggregation of Normalised distance matrices
MEGAN-S	Proposed direction-consistent (single-unification) variant of MEGAN
MEREC	MEthod based on the Removal Effects of Criteria
RAM	Root Assessment Method
RRS	Rank Range Similarity
SPOTIS	Stable Preference Ordering Towards Ideal Solution
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
VIKOR	VIseKriterijumska Optimizacija I Kompromisno Resenje
WASPAS	Weighted Aggregated Sum Product Assessment

2. Methods

This section first describes the two decision problems and the rationale for their selection (Section 2.1), then formalises the four robustness-oriented methods under comparison (Section 2.2) and the conventional reference methods and weighting schemes (Section 2.3), introduces the proposed variant MEGAN-S together with its theoretical justification and computational complexity (Section 2.4), and finally specifies the validation protocol, including all parameter settings required to reproduce the analysis (Section 2.5).

2.1. Decision problems and datasets

The benchmark is conducted on two engineering selection problems drawn from distinct domains: one published synthetic benchmark and one real-world case study. The use of two datasets rather than one serves a deliberate methodological purpose. A single dataset can demonstrate whether a given method behaves consistently within a particular case; however, such evidence is insufficient to establish whether the observed behaviour generalises across datasets with different scales, criterion structures, and levels of unit heterogeneity. Employing two datasets from different engineering domains, therefore, allows the study to test the cross-domain stability of its conclusions, which is the subject of the third research question. Both datasets are publicly available in peer-reviewed sources, ensuring that the study can be reproduced exactly by independent researchers.

The first dataset addresses renewable energy grid selection and is adopted from the original MEGAN article by Cebeci [10]. It is a synthetic benchmark constructed by the method's proposer, not a field-collected dataset, and it is used here deliberately in that capacity. It comprises ten grid alternatives, denoted G1 through G10, evaluated against 14 criteria spanning four dimensions: social, environmental, technical, and economic. The criteria weights were derived from an expert study of renewable energy decision factors and sum to unity. Of the fourteen criteria, three are cost-type, for which lower values are preferred, and eleven are benefit-type. This dataset serves a verification role in the present study. Because the rankings produced by MEGAN on these data are reported in the original article, reproducing them confirms that the present implementation of MEGAN is faithful to its published specification before the algorithm is applied to the second, independent dataset. The complete decision matrix is presented in Table 1.

The second dataset addresses flotation machine selection in mineral processing and is the real-world case of the benchmark. It originates in the study by Štirbanović et al. [15], in which five flotation machines from two manufacturers were evaluated by three domain experts using the Delphi method against ten criteria organised into three groups: constructional, economic, and technical parameters. The present study uses this dataset in the systematised form reproduced by Stanujkić et al. [16], in which the ten criteria are presented as a standard decision matrix with explicit optimisation directions and weights. This dataset serves the cross-domain validation role. It originates in an engineering domain entirely distinct from renewable energy; it exhibits a substantially smaller problem scale, and it has been independently solved by several established MCDM methods, providing external reference rankings. The complete decision matrix is shown in Table 2, and the criterion definitions, optimisation directions, and weights are shown in Table 3.

The two datasets are complementary along several dimensions relevant to a robustness benchmark. They originate in different engineering domains and differ in provenance: one is synthetic, and one is real-world, allowing the study to check whether conclusions drawn from constructed data survive contact with field data. They differ markedly in scale, with ten and five alternatives respectively. They differ in the number of criteria: 14 and 10, respectively. Both, however, contain a mixture of benefit-type and cost-type criteria, exposing every method to the heterogeneity of optimisation direction that a robustness claim must withstand.

A clarification regarding the criterion type is warranted for Table 3. The sources specify the optimisation direction for each criterion; criteria C2, C5, C7, and C9 are cost-type, and the remaining six are benefit-type. The weights are those assigned by the three-expert panel and sum to unity, providing the published baseline weighting configuration for this dataset.

Table 1. Decision matrix for renewable energy grid selection, adopted from Cebeci [10] (synthetic benchmark). Rows G1 through G10 are alternatives; columns C1 through C14 are criteria. The type row indicates criterion type, where 1 denotes a benefit criterion and −1 a cost criterion.

Grid	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
G1	78	74	66	18	11	75	65	78	78	171	69	44	59	50
G2	69	74	80	11	10	76	68	78	85	195	62	40	48	63
G3	69	72	75	18	9	79	50	92	65	121	78	42	48	54
G4	61	74	69	20	14	78	69	87	72	135	68	46	41	60
G5	74	58	65	15	12	85	65	86	82	119	65	47	54	50
G6	61	58	64	15	7	79	50	83	75	100	72	33	47	58
G7	76	66	65	11	6	71	63	75	69	193	79	45	47	55
G8	75	67	73	12	7	73	69	94	80	173	71	30	52	62
G9	75	64	71	15	5	72	62	92	72	141	71	42	54	53
G10	78	74	65	17	14	77	63	94	82	158	76	32	55	57
Type	1	1	1	−1	−1	1	1	1	1	−1	1	1	1	1
Weight	0.10	0.05	0.08	0.08	0.09	0.13	0.05	0.06	0.08	0.09	0.08	0.04	0.03	0.04

Table 2. Decision matrix for flotation machine selection, adopted from Štirbanović et al. [15] as reproduced by Stanujkić et al. [16]. Rows A1 through A5 are flotation machine alternatives; columns C1 through C10 are criteria. Three experts assigned ratings on a one-to-nine scale.

Alt	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
A1	3	7	4	4	6	4	6	8	5	8
A2	4	6	5	5	5	5	5	8	6	9
A3	6	4	5	6	4	5	5	9	7	9
A4	5	6	6	5	3	6	4	7	8	9
A5	2	8	3	4	6	3	6	7	7	8

Table 3. Criteria definitions, optimisation directions, and weights for the flotation machine selection problem. The optimisation direction max denotes a benefit criterion, and min denotes a cost criterion.

Criteria	Definition	Group	Direction	Weight
C1	Size and shape of the machine	Constructional	max	0.07
C2	Volume or capacity of the machine	Constructional	min	0.07
C3	Construction of agitation and aeration system	Constructional	max	0.07
C4	Number of machines	Constructional	max	0.14
C5	Investments	Economical	min	0.2
C6	Terms of payment and maintenance	Economical	max	0.08
C7	Operating costs	Economical	min	0.12
C8	Warranty period	Technical	max	0.125
C9	Delivery time	Technical	min	0.05
C10	Maintenance conditions	Technical	max	0.075

2.2. Robustness-oriented methods under comparison

This subsection specifies the four robustness-oriented, aggregation-based methods that form the core of the benchmark. All four share a common design philosophy. Rather than producing a ranking from a single computational pass, each combines multiple intermediate results into an aggregated outcome intended to confer stability. Throughout, the decision matrix is denoted $X = [x_{ij}] \in \mathbb{R}^{n\times m}$ with n alternatives and m criteria, the benefit-cost vector is $v = (v_1, \dots, v_m)$ with $v_j = 1$ for benefit criteria and $v_j = -1$ for cost criteria, and the weight vector is $w = (w_1, \dots, w_m)$ with $\sum_j w_j = 1$.

Four normalisation techniques appear in this study, each in a direction-aware form that maps every criterion onto the unit interval with larger values preferable. The MaxMin (linear) normalisation is

$$n_{ij} = \begin{cases} \dfrac{x_{ij} - \min\limits_i x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}}, & v_j = 1 \\ \dfrac{\max\limits_i x_{ij} - x_{ij}}{\max\limits_i x_{ij} - \min\limits_i x_{ij}}, & v_j = -1 \end{cases} \tag{1}$$

The vector normalisation is

$$n_{ij} = \begin{cases} x_{ij} \Big/ \sqrt{\sum_i x_{ij}^2}, & v_j = 1 \\ 1 - x_{ij} \Big/ \sqrt{\sum_i x_{ij}^2}, & v_j = -1 \end{cases} \tag{2}$$

The sum normalisation is

$$n_{ij} = \begin{cases} x_{ij} \Big/ \sum_i x_{ij}, & v_j = 1 \\ \dfrac{1/x_{ij}}{\sum_i (1/x_{ij})}, & v_j = -1 \end{cases} \tag{3}$$

and the max normalisation is

$$n_{ij} = \begin{cases} x_{ij} / \max\limits_i x_{ij}, & v_j = 1 \\ 1 - x_{ij} / \max\limits_i x_{ij}, & v_j = -1 \end{cases} \tag{4}$$

MaxMin normalisation, Eq. (1), is the default throughout; Eqs. (2) to (4) are used in the normalisation sensitivity analysis.

The MEGAN method, introduced by Cebeci [10], is a function-free multi-attribute decision-making method based on outranking relations. The defining feature of MEGAN is the construction of a gradual-weighting matrix W, which contains multiple weight scenarios obtained by systematically perturbing the initial weights. Given a rate vector $r = (r_1, \dots, r_L)$ of perturbation magnitudes, each scenario reduces the weight of one criterion j at one rate r_l and rescales the remaining weights proportionally so that the total again sums to unity:

$$\begin{aligned} w_j' &= w_j(1 - r_l), \qquad w_k' \\ &= w_k \frac{1 - w_j'}{1 - w_j} \quad (k \neq j) \end{aligned} \tag{5}$$

The first row of W contains the baseline weights, and the subsequent $L \cdot m$ rows contain the perturbed scenarios, so that W has $1 + Lm$ rows in total. Following the original specification, this study uses $r = (0.05, 0.15, 0.20, 0.25)$, hence $L = 4$. For each weight scenario $w^{(l)}$, the decision matrix is normalised by Eq. (1), weighted, and multiplied a second time by the direction vector:

$$Y = N \operatorname{diag}\left(w^{(l)}\right) \tag{6}$$

$$U = Y \operatorname{diag}(v) \tag{7}$$

An L_1-type distance matrix is then formed against the criterion-wise minimum and normalised a second time by its global maximum:

$$d_{ij} = u_{ij} - \min_i u_{ij}, \qquad \tilde{d}_{ij} = \frac{d_{ij}}{\max\limits_{i,j} d_{ij}} \tag{8}$$

and the row sums of the twice-normalised matrix yield the superiority score of each alternative, expressed in percentage form relative to the smallest score:

$$s_i = \sum_{j=1}^{m} \tilde{d}_{ij}, \qquad p_i = 100 \cdot \frac{s_i - \min\limits_k s_k}{\min\limits_k s_k} \tag{9}$$

Rather than ranking by simple sorting, MEGAN applies a dynamic threshold τ, computed when unspecified as the fifth percentile of the successive differences of the sorted superiority percentages,

$$\tau = P_5\big(\, p_{(k)} - p_{(k+1)} \,\big) \tag{10}$$

which permits tied ranks when consecutive alternatives are not meaningfully distinguishable (in the implementation used here, the threshold parameter is set to its published default). The rankings $R^{(l)}$ obtained across all $1 + Lm$ weight scenarios are finally aggregated through the Borda Count:

$$b_i = \sum_{l} \left(n - R_i^{(l)}\right) \tag{11}$$

The alternative with the highest Borda score b_i is assigned the top rank. The complete procedure is summarised as Algorithm 1.

The COCOSO method, the Combined Compromise Solution proposed by Yazdani et al. [7], aggregates three distinct appraisal strategies into a single combined score. After normalising the decision matrix by linear normalisation, the method computes for each alternative a weighted sum S_i of comparability sequences and a weighted power P_i of comparability sequences:

$$S_i = \sum_{j=1}^{m} w_j \; n_{ij}, \qquad P_i = \sum_{j=1}^{m} n_{ij}^{\,w_j} \tag{12}$$

These two quantities are combined into three appraisal scores,

$$k_{ia} = \frac{P_i + S_i}{\sum_k (P_k + S_k)}, \qquad k_{ib} = \frac{S_i}{\min\limits_k S_k} + \frac{P_i}{\min\limits_k} \tag{13}$$

$$k_{ic} = \frac{\lambda S_i + (1-\lambda) P_i}{\lambda \max\limits_k S_k + (1-\lambda) \max\limits_k P_k}$$

with $\lambda = 0.5$, and the three are aggregated into the final compromise score

$$k_i = (k_{ia}\, k_{ib}\, k_{ic})^{1/3} + 1/3\,(k_{ia} + k_{ib} + k_{ic}) \tag{14}$$

The aggregation of multiple appraisal strategies is the source of the method's claimed stability.

The AROMAN method, the Alternative Ranking Order Method Accounting for Two-Step Normalisation proposed by Bošković et al. [8], is built upon the deliberate combination of two normalisation techniques. The decision matrix is first normalised using linear MaxMin normalisation and then vector normalisation, and the two normalised matrices are averaged:

$$n_{ij}^{A} = 1/2\left(n_{ij}^{\mathrm{MaxMin}} + n_{ij}^{\mathrm{Vector}}\right) \tag{15}$$

The averaged matrix is weighted, and the weighted sums of benefit-type criteria (A_i) and cost-type criteria (L_i) are combined into the final ranking score with balance parameter $\lambda = 0.5$:

$$F_i = \lambda A_i - (1-\lambda) L_i + (1-\lambda) \max_k L_k \tag{16}$$

By averaging across two normalisation schemes rather than committing to one, AROMAN is designed to reduce the ranking's dependence on the normalisation choice.

The RAM (Root Assessment Method) proposed by Sotoudeh-Anvari [9] aggregates the weighted, sum-normalised scores of beneficial and non-beneficial criteria using a root-based function. With S_{+i} the sum of weighted normalised scores over benefit criteria and S_{-i} the corresponding sum over cost criteria, the evaluation score is

$$RI_i = (2 + S_{+i})^{1/(2+S_{-i})} \tag{17}$$

designed to provide a controlled level of compensation between criteria while preserving computational simplicity and reliability.

2.3. Conventional reference methods and weighting schemes

To place the four robustness-oriented methods against a familiar backdrop, four conventional methods are included as reference points: TOPSIS [11], VIKOR [12], WASPAS [13], and EDAS [14]. They span the principal MCDM families, distance-based, compromise-based, and aggregation-based, and their behaviour is well understood, which makes them a useful baseline. Each is

implemented according to its original specification, as formalised in the respective source publications; because these four methods are standard, their equations are not reproduced here, but the exact implementations are available in the public code accompanying this article (see the Data Availability and Code Availability Statements). The VIKOR strategy weight and the WASPAS combination parameter are both set to 0.5.

Because the criterion weights affect the ranking, and because one of the claims under test concerns insensitivity to the weighting scheme, the benchmark uses the published weights of each dataset as the principal setting and adds five objective weighting methods for the sensitivity analysis: equal weighting, the entropy method, the standard deviation method, the Criteria Importance Through Intercriteria Correlation method (CRITIC) [17], and the MEthod based on the Removal Effects of Criteria (MEREC) [18]. These range from the neutral allocation of equal weighting to the correlation-aware allocation of CRITIC, so that each method is tested against weight vectors that differ substantially from one another.

2.4. MEGAN-S: a direction-consistent variant, its theoretical justification, and computational complexity

The divergence of MEGAN on the renewable-energy problem, reported below, can be traced to a single redundant operation in how the optimisation direction is handled. In the implementation that reproduces the published rankings, the decision matrix is first normalised with the benefit-and-cost-aware technique of Eq. (1), so that after normalisation every column is already oriented with larger values preferable and bounded to the unit interval. The method then multiplies the weighted normalised matrix by the optimisation-direction vector a second time, Eq. (7). An informal appeal to the numerical scale of the cost criteria does not account for this distortion; the following analysis gives its exact mechanism.

Proposition 1 (direction cancellation). *Let the decision matrix be normalised by the direction-aware MaxMin technique of Eq. (1) and let the MEGAN superiority scores be computed through Eqs. (6) to (9). Then, for every weight scenario, the scores are identical to those obtained when every criterion is declared benefit-type, i.e., the second multiplication by* $\mathrm{diag}(v)$ *in Eq. (7) exactly cancels the direction handling performed during normalisation, and the MEGAN ranking is independent of the benefit-cost vector* v.

Proof. For a benefit criterion j, $u_{ij} = w_j n_{ij} \geq 0$ and $\min_i u_{ij} = 0$ because the minimum of Eq. (1) is attained; hence $d_{ij} = w_j n_{ij}$. For a cost criterion j, $u_{ij} = -w_j n_{ij}$ and $\min_i u_{ij} = -w_j$ because the maximum $n_{ij} = 1$ is attained; hence $d_{ij} = w_j(1 - n_{ij})$. Substituting Eq. (1) for a cost criterion gives $1 - n_{ij} = (x_{ij} - \min_i x_{ij})/(\max_i x_{ij} - \min_i x_{ij})$, which is precisely the benefit-oriented MaxMin normalisation of the raw column. Therefore, every column of D, benefit or cost, equals the benefit-oriented normalised column scaled by its weight, which is the matrix that Eqs. (6) to (9) produce when $v_j = 1$ for all j. The scores, and hence the threshold ranking and the Borda aggregation built on them, coincide.

Corollary 1. *Under its default normalisation, MEGAN maximises cost criteria: alternatives with higher raw values on a cost criterion receive higher superiority contributions from that criterion. The information carried by* v *is silently discarded.* This property was verified computationally on both datasets: declaring all fourteen, respectively all ten, criteria as benefit-type reproduces the MEGAN rankings of Tables 4 and 5 exactly. The same cancellation holds under the vector and max normalisations, Eqs. (2) and (4), and was likewise verified computationally; under the sum normalisation, Eq. (3), the cancellation is no longer an exact identity, but the effective orientation of the cost criteria is still reversed, since d_{ij} remains increasing in the raw cost values.

Proposition 1 also settles the role of scale: because Eq. (1) is invariant under positive affine transformations $x_{ij} \mapsto a_j x_{ij} + c_j$ $(a_j > 0)$ of any criterion column, the raw numerical range of a cost criterion does not by itself influence the MEGAN ranking; this affine invariance was also confirmed computationally for both MEGAN and MEGAN-S. What drives the divergence is not the scale of the implementation-cost column but the fact that its orientation is reversed: the benchmark problem rewards inexpensive grids, while MEGAN, by Corollary 1, rewards expensive ones. How severely the cancellation distorts the final ranking depends on how strongly the cost

criteria discriminate among the alternatives and how they correlate with the benefit criteria, which is a property of the problem structure, and this is why the distortion is nearly invisible on one dataset and decisive on the other.

The proposed variant, MEGAN-S, removes this second multiplication and leaves every other step of MEGAN unchanged: Eq. (7) is replaced by

$$U = Y \tag{18}$$

so that the optimisation direction is unified exactly once, during normalisation, while the gradual-weighting scheme, the distance-to-minimum scoring, the threshold-based rank assignment, and the Borda-count aggregation are all identical to the original. In the name MEGAN-S, the S denotes single-direction unification.

Proposition 2 (properties of MEGAN-S). *With Eqs. (6), (18), (8) and (9): (i) the contribution of criterion j to the superiority score of alternative i equals $w_j n_{ij} / \max_{k,l}(w_l n_{kl})$, is bounded by $w_j / \max_{k,l}(w_l n_{kl})$, and is monotonically increasing in the normalised performance n_{ij}, so every criterion acts in the direction specified by v and its influence is governed by its weight; (ii) the ranking is invariant under positive affine transformations of any criterion column, inherited from Eq. (1); (iii) resistance to rank reversal is retained, because the Borda aggregation of Eq. (11), which the removal trials identify as the source of that resistance, is unchanged.* The proof of (i) is immediate from substituting Eq. (18) into Eqs. (8) and (9), using $\min_i w_j n_{ij} = 0$; (ii) follows from the affine invariance of Eq. (1); (iii) is confirmed empirically by the rank-reversal trials of Table 6.

Computational complexity. For a problem with n alternatives and m criteria, every single-pass comparator (COCOSO, AROMAN, RAM, TOPSIS, VIKOR, WASPAS, EDAS) runs in $\Theta(nm)$ time. MEGAN and MEGAN-S evaluate $1 + Lm$ weight scenarios, each requiring $\Theta(nm)$ work plus an $\Theta(n\log n)$ sort for the threshold ranking, giving

$$\Theta\big((1 + Lm)(nm + n\log n)\big) \tag{19}$$

i.e., a factor of roughly Lm more work than a single-pass method: 57 scenario evaluations for the energy problem ($m = 14$) and 41 for the flotation problem ($m = 10$) at $L = 4$. The variant adds no cost of its own, since removing one elementwise multiplication is negligible. Measured runtimes are reported in Table 9.

Algorithm 1. MEGAN and MEGAN-S. The two methods differ only in line 5.

```
Input:  X (n x m decision matrix), v (benefit-cost vector),
        w (initial weights, sum = 1), rates r, threshold t
Output: final ranking of the n alternatives

1  build gradual-weighting matrix W (1 + L*m rows)   [Eq. (5)]
2  for each weight scenario w' in W do
3      N <- direction-aware MaxMin normalisation [Eq. (1)]
4      Y <- N * diag(w')                             [Eq. (6)]
5      U <- Y * diag(v) (MEGAN)   |   U <- Y (MEGAN-S)
6      D <- U - column minima; D <- D / max(D)       [Eq. (8)]
7      s <- row sums of D; p <- percentage form      [Eq. (9)]
8      R(w') <- threshold-based ranks from p with t  [Eq. (10)]
9  end for
10 b_i <- sum over scenarios of (n - R_i)            [Eq. (11)]
11 return alternatives sorted by decreasing b_i
```

2.5. Validation protocol

The validation protocol comprises four components, namely ranking similarity assessment, statistical testing, sensitivity analysis, and rank reversal testing. Each is applied identically to both datasets, so that the cross-domain comparison rests on a common methodological basis. All stochastic elements use fixed seeds, and every parameter of the protocol is stated below and in the public code, so that the entire analysis is bit-for-bit reproducible.

Because the terms are used with distinct meanings in this paper, they are defined here explicitly. Normalisation (or weighting) insensitivity denotes the strict property that the ranking remains identical when the normalisation technique (or weighting scheme) is exchanged, quantified as the percentage of ranking positions that agree with the baseline. Normalisation (or weighting) robustness denotes the graded property that rankings remain highly correlated rather than identical, quantified by the mean Spearman and Kendall correlations between the baseline and perturbed rankings. A method can be robust in the graded sense while failing strict insensitivity, and the two notions are reported separately throughout.

Ranking similarity between each pair of methods is quantified using three complementary measures. The Spearman rank correlation coefficient assesses the overall monotonic agreement between two complete rankings, with values approaching unity indicating strong agreement; for rankings r and r' of n alternatives,

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(r_i - r_i')^2}{n(n^2-1)} \tag{20}$$

The Rank Range Similarity metric, introduced by Cebeci [10] and evaluated here for the top three alternatives, quantifies the degree of overlap in the identification of the best-performing alternatives, a focus of particular practical interest to decision-makers. Denoting by $T_k(r)$ the set of alternatives occupying the first k positions of ranking r, the metric is defined as

$$\mathrm{RRS}_k(r, r') = \frac{|T_k(r) \cap T_k(r')|}{k} \times 100\,\% \tag{21}$$

with $k = 3$ throughout this study. The Jensen-Shannon divergence (JSD) [19], a symmetric and bounded measure derived from information theory, quantifies the dissimilarity between the ranking distributions of method pairs:

$$\mathrm{JSD}(P \parallel Q) = 1/2\ \mathrm{KL}(P \parallel M) + 1/2\ \mathrm{KL}(Q \parallel M) \qquad M = 1/2\,(P+Q) \tag{22}$$

where P and Q are the rank vectors normalised to unit sum, and KL denotes the Kullback-Leibler divergence.

Statistical testing assesses whether the agreement among the method rankings is greater than would be expected by chance. Concordance among the eight benchmarked methods, computed without the proposed MEGAN-S variant so that it measures agreement within the existing field, is quantified by Kendall's coefficient of concordance W [20], defined for k methods ranking n alternatives as

$$W = \frac{12\sum_{i=1}^{n}(R_i - \bar{R})^2}{k^2(n^3-n)} \tag{23}$$

where R_i is the rank sum of alternative i across the k methods. Its significance is obtained from a permutation distribution generated by randomly relabelling the alternatives, using 10,000 permutations with seed 42. The concordance analysis is complemented by the Friedman statistic, which for rank data equals $\chi_F^2 = k(n-1)W$ with $n-1$ degrees of freedom, and by the Iman-Davenport correction

$$F_{ID} = \frac{(k-1)\,\chi_F^2}{k(n-1) - \chi_F^2} \tag{24}$$

with $(n-1)$ and $(n-1)(k-1)$ degrees of freedom. Pairwise Wilcoxon signed-rank tests between MEGAN (and MEGAN-S) and each comparator, with Holm correction for multiplicity, are also reported. Pairwise agreement is evaluated with Spearman's ρ and Kendall's τ, the significance of each coefficient being obtained from an exact permutation test, which is appropriate given the small number of alternatives in each problem: for $n = 5$ the full set of $5! = 120$ orderings is enumerated, and for $n = 10$ the p-value is estimated from 100,000 random permutations with seed 42. Because these statistics operate directly on the ordering of the alternatives, they are sensitive to exactly the ranking differences with which the benchmark is concerned, and they establish whether

the observed inter-method agreement is statistically meaningful.

Two caveats on statistical power are recorded here and taken up again in Sections 3.1 and 3.7. First, with $n = 5$ alternatives only 120 distinct orderings exist, so the smallest attainable two-sided permutation p-value is $2/120 \approx 0.017$; even a perfect rank agreement cannot produce stronger evidence, and non-significant results on the flotation problem must be read in that light. Similarly, the Wilcoxon signed-rank tests are underpowered at $n = 5$ and $n = 10$ and are reported for completeness rather than inference. Second, several modern post-hoc instruments, in particular critical difference diagrams based on the Nemenyi test, are designed for comparisons across many datasets; with two datasets they would convey no information beyond the per-dataset statistics above, and they are therefore deliberately not reported. Extending the benchmark to a dataset collection large enough to support such analyses is identified as future work.

Sensitivity analysis evaluates the robustness claims directly. To test insensitivity to normalisation, MEGAN and the comparator methods are re-executed under four normalisation techniques, namely MaxMin, Vector, Sum, and Max normalisation, Eqs. (1) to (4), while the decision matrix and weights are held fixed, and the resulting rankings are compared. To test insensitivity to the weighting scheme, the methods are re-executed under the five objective weighting methods introduced above, and the stability of the rankings across these configurations is recorded.

The proportion of configurations under which the ranking remains unchanged provides a quantitative measure of strict insensitivity. In addition, the graded robustness measures defined above are reported: the mean Spearman and Kendall correlations between the baseline and the perturbed rankings, and the mean per-alternative standard deviation of ranks across the weighting schemes. Finally, a Monte Carlo experiment perturbs the published weights randomly rather than replacing them by design: each weight is multiplied by $(1 + u_j)$ with u_j drawn independently from the uniform distribution

$$u_j \sim U(-0.25, +0.25)$$
$$w_j^{MC} = w_j(1 + u_j) / \sum_k w_k\,(1 + u_k) \tag{25}$$

and the ranking is recomputed; 1,000 samples are drawn with seed 42, and the mean Spearman correlation with the baseline ranking, its 2.5th to 97.5th percentile interval, and the share of samples preserving the top-ranked alternative are reported. The perturbation half-width of 25 per cent matches the largest rate in MEGAN's own gradual-weighting vector, so that the external perturbation is commensurate with the internal one. Note that the gradual weighting inside MEGAN, Eq. (5), and the perturbations applied by this protocol are two independent layers: the former is part of the algorithm under test, the latter is part of the test itself.

Rank reversal testing examines whether the ordering of alternatives is preserved when the candidate set is modified. Each non-optimal alternative is removed in turn, the methods are re-executed on the reduced decision matrix, and the proportion of removals that alter the top-ranked alternative is reported as the rank reversal rate. A lower rate indicates greater resistance to rank reversal.

The full benchmark, including every table and figure reported below, can be regenerated with a single command from the Python implementation released with this article; the implementation relies only on NumPy, SciPy, and Matplotlib, in line with recent open MCDM software practice [21].

3. Results and Discussion

The eight benchmarked methods, together with the proposed MEGAN-S variant, were applied to both datasets using the published weights and then re-evaluated under the perturbations of the validation protocol. Before the comparison, the MEGAN implementation was checked against the renewable-energy rankings reported by Cebeci [10]; it reproduced them exactly, which confirms that the implementation follows the published algorithm. The TOPSIS and COCOSO implementations were verified in the same way against the published flotation-machine results [15], [16] and also matched.

3.1. Inter-method ranking agreement

Table 4 and Table 5 report the rankings produced under the published weights, and Figure 1 displays them as heat maps. The tables additionally report the mean and standard deviation of each alternative's rank across the eight benchmarked methods, and the top-ranked position of each method is set in bold. The two datasets give very different pictures.

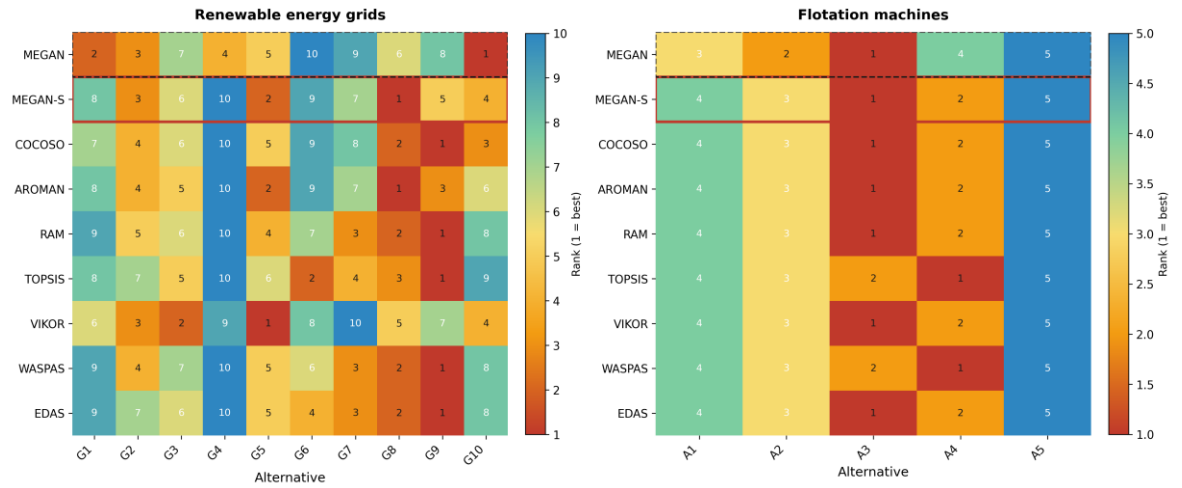

Figure 1. Rankings produced by the eight benchmarked methods and the proposed MEGAN-S variant (rank 1 = best). The MEGAN row carries a dashed outline and MEGAN-S a solid one.

Table 4. Rankings of the ten renewable-energy grid alternatives produced by the eight benchmarked methods and the proposed MEGAN-S variant under the published weights (rank 1 = best; best rank per method in bold). The final two rows report the mean and standard deviation of each alternative's rank across the eight benchmarked methods (MEGAN-S excluded).

Method	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
MEGAN	2	3	7	4	5	10	9	6	8	**1**
MEGAN-S	8	3	6	10	2	9	7	**1**	5	4
COCOSO	7	4	6	10	5	9	8	2	**1**	3
AROMAN	8	4	5	10	2	9	7	**1**	3	6
RAM	9	5	6	10	4	7	3	2	**1**	8
TOPSIS	8	7	5	10	6	2	4	3	**1**	9
VIKOR	6	3	2	9	**1**	8	10	5	7	4
WASPAS	9	4	7	10	5	6	3	2	**1**	8
EDAS	9	7	6	10	5	4	3	2	**1**	8
Mean	7.2	4.6	5.5	9.1	4.1	6.9	5.9	2.9	2.9	5.9
SD	2.4	1.6	1.6	2.1	1.7	2.7	2.9	1.7	2.9	2.9

Table 5. Rankings of the five flotation-machine alternatives produced by the eight benchmarked methods and the proposed MEGAN-S variant under the published weights (rank 1 = best; best rank per method in bold). The final two rows report the mean and standard deviation of each alternative's rank across the eight benchmarked methods (MEGAN-S excluded).

Method	A1	A2	A3	A4	A5
MEGAN	3	2	**1**	4	5
MEGAN-S	4	3	**1**	2	5
COCOSO	4	3	**1**	2	5
AROMAN	4	3	**1**	2	5
RAM	4	3	**1**	2	5
TOPSIS	4	3	2	**1**	5
VIKOR	4	3	**1**	2	5
WASPAS	4	3	2	**1**	5
EDAS	4	3	**1**	2	5
Mean	3.9	2.9	1.2	2.0	5.0
SD	0.4	0.4	0.5	0.9	0.0

On the flotation problem the methods largely agree. Among the nine rankings shown, seven place A3 first and two place A4 first; all place A5 last. MEGAN ranks A3 first, A2 second, A1 third, A4 fourth, and A5 fifth, which is close to the consensus. The Spearman coefficient between MEGAN and COCOSO, AROMAN, RAM, VIKOR and EDAS is 0.70, and 0.40 against TOPSIS and WASPAS (Figure 2). Measured as Rank Range Similarity, the overlap among the top three alternatives is 66.7 per cent against every other method (Figure 3). At this problem size, however, none of the MEGAN correlations reaches significance under the exact permutation test (two-sided $p = 0.233$ for $\rho = 0.70$), which reflects the power limitation of a five-alternative problem rather than an absence of agreement: with five alternatives even a perfect correlation cannot yield a p-value below 0.017. The MEGAN-S correlations of 1.00 with five of the seven comparators attain exactly that attainable minimum.

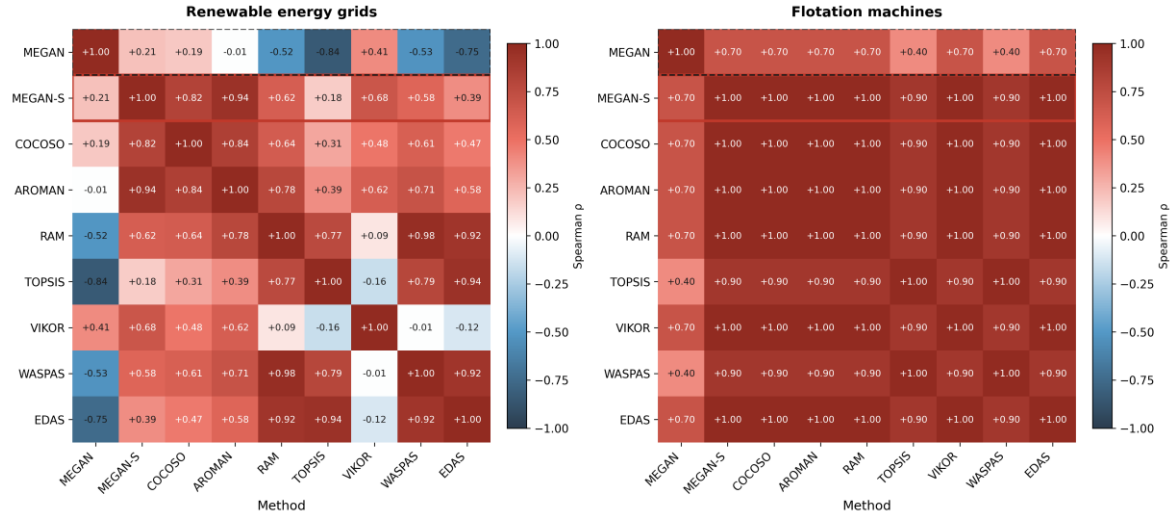

Figure 2. Pairwise Spearman rank correlation between the methods, with MEGAN (dashed outline) and the proposed MEGAN-S variant (solid outline) highlighted.

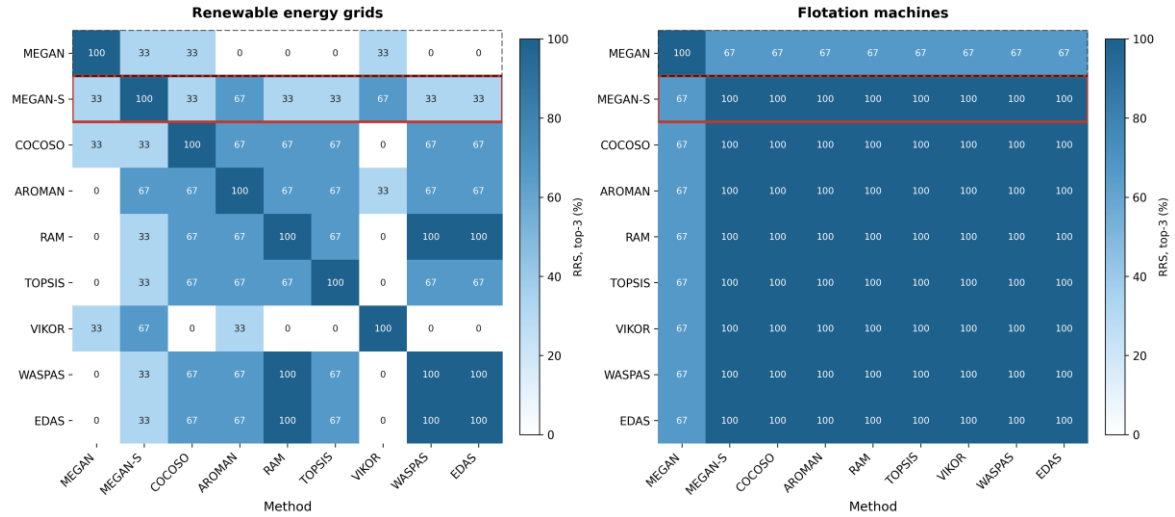

Figure 3. Rank Range Similarity of the top three alternatives across the methods, including the proposed MEGAN-S variant.

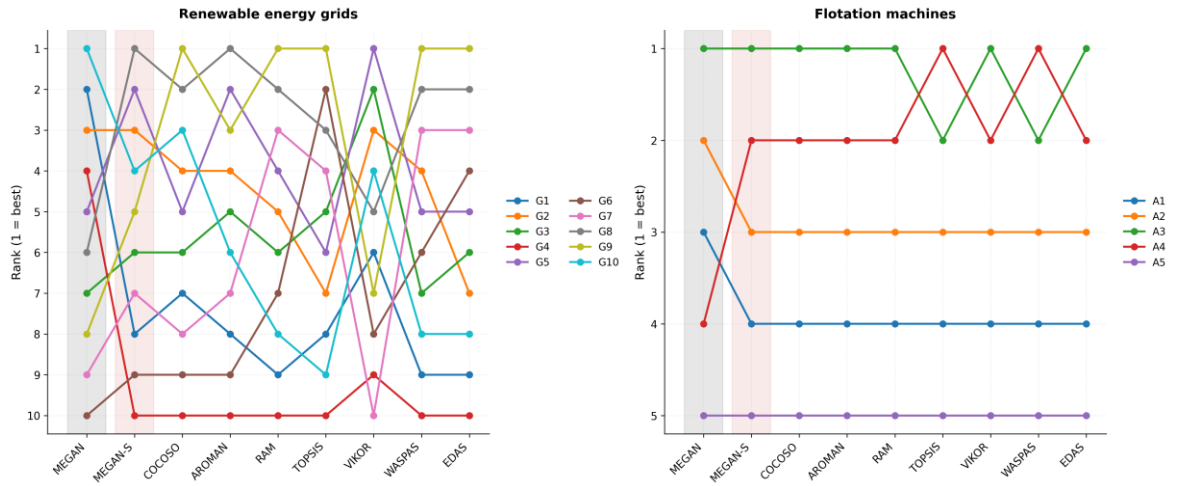

Figure 4. Rank trajectory of each alternative across the methods; the MEGAN and MEGAN-S columns are shaded.

The renewable-energy problem behaves differently. MEGAN places G10 first, G1 second, and G2 third, while the conventional methods mostly place G9 first and push G10 into the lower half. The Spearman correlation between MEGAN and the other methods ranges from +0.41 for VIKOR down to −0.84 for TOPSIS, −0.75 for EDAS, −0.53 for WASPAS and −0.52 for RAM. Under the permutation test with 100,000 resamples, the negative correlations with TOPSIS and EDAS are individually significant ($p = 0.004$ and $p = 0.017$), so the disagreement is not attributable to chance. For five of the seven pairs, the top-three overlap is 0 per cent. Figure 4 traces each alternative across the methods and shows how far the MEGAN column departs from the rest.

To establish whether the agreement among the eight methods exceeds what random ordering would produce, Kendall's coefficient of concordance W, Eq. (23), was computed across the full panel of methods and its significance obtained from a permutation distribution generated by randomly relabelling the alternatives, using 10,000 permutations. The concordance was statistically significant on both problems but differed markedly in degree: $W = 0.884$ on the flotation data (permutation $p = 0.0001$; Friedman $\chi^2 = 28.3$ on four degrees of freedom; Iman-Davenport $F = 53.5$ with $p = 1.0 \times 10^{-12}$) and $W = 0.422$ on the renewable-energy data (permutation $p = 0.0002$; Friedman $\chi^2 = 30.4$ on nine degrees of freedom; Iman-Davenport $F = 5.11$ with $p = 3.5 \times 10^{-5}$). The high value on the flotation problem reflects the panel's close agreement. In contrast, the moderate value on the renewable-energy problem is depressed by MEGAN's divergence from the remaining methods rather than by disagreement among those methods themselves. Because the coefficient of concordance is defined directly on the rank agreement among methods, it answers the panel-level question of interest here. In contrast, the pairwise Spearman coefficient and the Rank Range Similarity metric identify which specific method pairs drive the agreement or its absence. Pairwise Wilcoxon signed-rank tests with Holm correction, included in the released code for completeness, detect no significant location differences at either problem size, as expected for tests of this type applied to five or ten paired ranks; they are therefore not tabulated.

3.2. Robustness claims under perturbation

The original MEGAN article states that the method is insensitive to both the normalisation technique and the weighting scheme. Neither claim held up on the two datasets. Figure 5 shows the normalisation test. With MaxMin as the baseline, the MEGAN ranking on the renewable-energy data was preserved in only 30 per cent of positions under Vector normalisation and 40 per cent under Sum and Max normalisation. On the flotation data, the figure was 60 per cent for all three. A method described as normalisation-insensitive would be expected to stay close to 100 per cent, so the claim is not supported in the strict sense defined above. The graded picture is more nuanced: the mean Spearman correlation between the MaxMin baseline and the three alternative normalisations remains high for MEGAN, at +0.956 on

the energy data and +0.900 on the flotation data, indicating that the positional changes are mostly local swaps rather than wholesale reordering. Strict insensitivity, which is what the original claim advertises, is nevertheless clearly not attained.

The weighting test, shown in Figure 5 and summarised in Table 6, points in the same direction. Averaged across the five weighting schemes, the MEGAN ranking matched its published-weight baseline at 30 per cent of positions on the renewable-energy data and 44 per cent on the flotation data. On the flotation problem, MEGAN was the least stable of the four robustness-oriented methods: RAM held at 100 per cent, AROMAN at 92 per cent, and COCOSO at 84 per cent. On the renewable-energy problem, all four were unstable, ranging from 26 to 38 per cent.

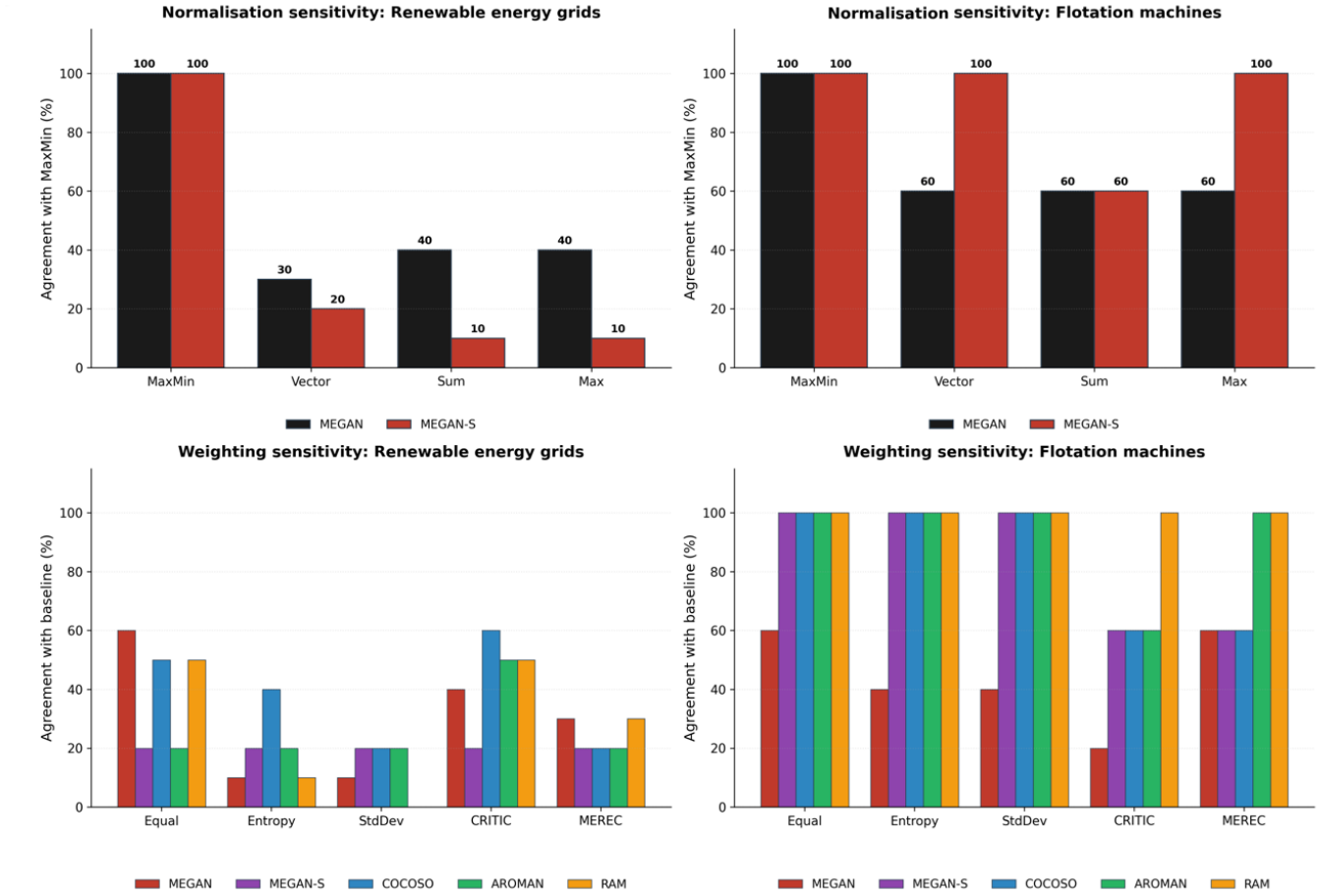

Figure 5. Robustness of MEGAN, the proposed MEGAN-S variant and the comparator methods to the choice of normalisation and weighting.

Table 6. Weighting-sensitivity and rank-reversal summary for the four robustness-oriented methods and the proposed MEGAN-S variant. Weighting agreement is the mean share of alternatives keeping their rank across the five weighting schemes; the rank-reversal rate is the share of alternative removals that change the top-ranked alternative.

Method	Weighting agree energy (%)	Weighting agree flotation (%)	Rank reversal energy (%)	Rank reversal flotation (%)
MEGAN	30	44	0.0	0.0
MEGAN-S	20	84	0.0	0.0
COCOSO	38	84	11.1	25.0
AROMAN	26	92	0.0	0.0
RAM	28	100	0.0	0.0

The strictness measures are reported in Table 7 and Figure 6. Two observations qualify the strict-insensitivity results above. First, the graded measures show that even where positional agreement is low, the orderings remain positively and often strongly correlated with the baseline: on the energy problem, the mean Spearman correlation across the five weighting schemes ranges from +0.585 (MEGAN-S) to +0.881 (MEGAN), with mean per-alternative rank standard deviations between 1.05 and 2.05 positions. Second, the Monte Carlo experiment, which perturbs the published weights randomly by up to ±25 per cent rather than replacing them by design, finds all five methods highly stable in the graded sense, with mean correlations of +0.93 or higher on both problems. The instability recorded against the five objective weighting schemes therefore stems from those schemes producing weight vectors far from the published ones, not from erratic behaviour under small weight changes. The Monte Carlo top-1 retention rates add a caution of their own: RAM, the strictest performer in Table 6 on the flotation data, preserves its top-ranked energy alternative in only 67.4 per cent of the random samples, while MEGAN preserves its top alternative in every single sample on both problems. As Section 3.4 shows, that perfect retention is itself a symptom rather than a virtue: the ranking is anchored by the direction-cancellation mechanism of Proposition 1.

Rank reversal was the one area where MEGAN matched its claim; COCOSO reversed its top alternative in 11 per cent of removals on the renewable-energy data and 25 per cent on the flotation data, and EDAS reversed in 50 per cent of removals on the flotation data. The gradual weighting and Borda aggregation in MEGAN therefore do deliver resistance to rank reversal, even though they do not deliver the normalisation and weighting insensitivity claimed alongside them.

Table 7. Extended weighting-robustness measures for the four robustness-oriented methods and the proposed MEGAN-S variant: mean Spearman ρ and Kendall τ between the published-weight baseline and the five objective weighting schemes, mean per-alternative rank standard deviation across those schemes, and Monte Carlo results over 1,000 random weight perturbations of ±25 per cent (mean ρ, 95 per cent percentile interval, and share of samples preserving the top-ranked alternative). Percentile intervals can be degenerate when fewer than 2.5 per cent of samples deviate from the baseline.

Dataset	Method	Mean ρ	Mean τ	Rank SD	MC mean ρ	MC 95% interval	MC top-1 kept (%)
Energy	MEGAN	+0.881	+0.760	1.05	+0.965	[+0.891, +1.000]	100.0
Energy	MEGAN-S	+0.585	+0.467	2.05	+0.931	[+0.806, +1.000]	93.5
Energy	COCOSO	+0.835	+0.698	1.33	+0.980	[+0.927, +1.000]	95.6
Energy	AROMAN	+0.629	+0.502	1.98	+0.942	[+0.855, +0.988]	96.5
Energy	RAM	+0.695	+0.582	1.90	+0.939	[+0.842, +0.988]	67.4
Flotation	MEGAN	+0.740	+0.640	0.48	+0.954	[+0.900, +1.000]	100.0
Flotation	MEGAN-S	+0.960	+0.920	0.21	+1.000	[+1.000, +1.000]	100.0
Flotation	COCOSO	+0.960	+0.920	0.21	+1.000	[+1.000, +1.000]	100.0
Flotation	AROMAN	+0.980	+0.960	0.16	+1.000	[+1.000, +1.000]	100.0
Flotation	RAM	+1.000	+1.000	0.00	+0.998	[+1.000, +1.000]	98.0

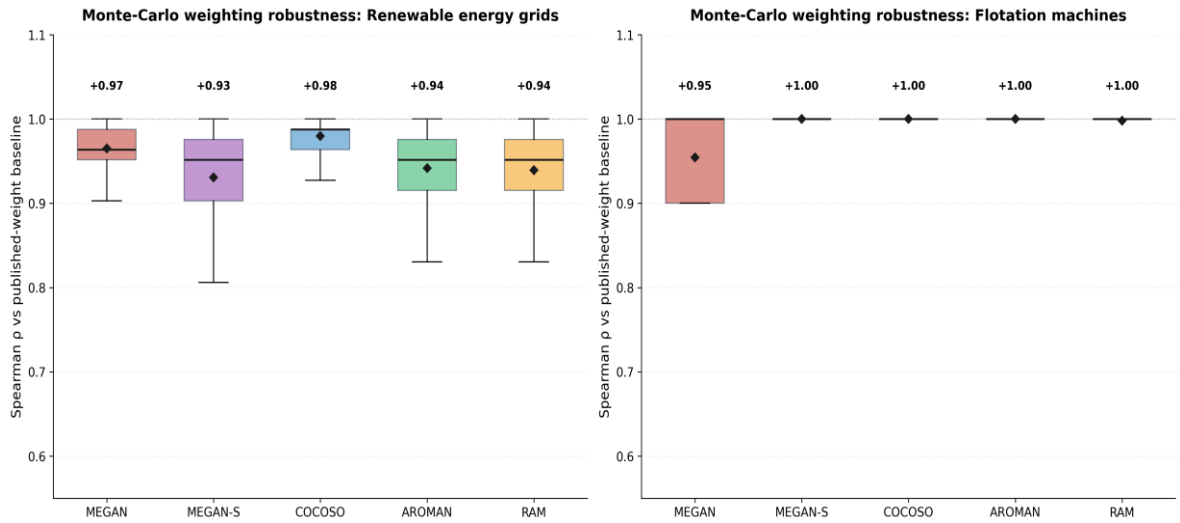

Figure 6. Monte Carlo weighting robustness of the four robustness-oriented methods and the proposed MEGAN-S variant: distribution of the Spearman correlation between the published-weight baseline ranking and the rankings under 1,000 random weight perturbations of ±25 per cent.

3.3. Cross-domain comparison

Reading the two datasets together answers the third research question. Of the three properties examined, only one was stable across both domains. Resistance to rank reversal held for MEGAN on the renewable-energy and the flotation problem alike. The other two properties did not transfer. The agreement between MEGAN and the remaining methods swung from clearly positive on the flotation data to strongly negative on the renewable-energy data, and the normalisation and weighting behaviour shifted accordingly. A study based solely on the flotation problem would have concluded that MEGAN sits comfortably within the MCDM mainstream; a study based solely on the renewable-energy problem would have concluded the opposite. The reason for this divergence is taken up next.

3.4. Effect of the direction-consistent correction

Re-running the benchmark with MEGAN-S changes the picture decisively on the problem where MEGAN had failed. Table 8 reports the pairwise Spearman correlation of MEGAN and of MEGAN-S with the seven comparator methods. On the renewable-energy problem, the correlation of MEGAN with the comparators averages −0.29 and ranges from −0.84 to +0.41; for MEGAN-S, the same average rises to +0.60, and every individual correlation becomes positive, ranging from +0.18 to +0.94. The variant therefore turns a method that contradicted the established approaches into one that broadly agrees with them, through the removal of a single redundant step, exactly as Proposition 1 predicts: what the removal restores is the orientation of the cost criteria, which the original formulation had silently discarded. On the flotation problem, where MEGAN already agreed with the comparators, MEGAN-S strengthens that agreement further, from an average of +0.61 to +0.97, reaching perfect rank correlation with five of the seven methods.

This realignment does not translate into uniformly greater stability under perturbation, and the pattern of where it does and does not is itself informative. On the flotation problem, MEGAN-S is markedly more stable than MEGAN: its mean agreement across the five weighting schemes rises from 44 to 84 per cent (Table 6), and its rankings are preserved under three of the four normalisation techniques rather than under none. On the renewable-energy problem, the opposite holds, with weighting agreement falling from 30 to 20 per cent. The explanation lies in what produced MEGAN's stability there in the first place.

Table 8. Pairwise Spearman rank correlation of MEGAN and of the proposed MEGAN-S variant with the seven comparator methods on both problems. Positive values denote agreement; the final row reports the mean across the seven comparators.

Method	MEGAN (energy)	MEGAN-S (energy)	MEGAN (flotation)	MEGAN-S (flotation)
COCOSO	+0.19	+0.82	+0.70	+1.00
AROMAN	−0.01	+0.94	+0.70	+1.00
RAM	−0.52	+0.62	+0.70	+1.00
TOPSIS	−0.84	+0.18	+0.40	+0.90
VIKOR	+0.41	+0.68	+0.70	+1.00
WASPAS	−0.53	+0.58	+0.40	+0.90
EDAS	−0.75	+0.39	+0.70	+1.00
Mean	−0.29	+0.60	+0.61	+0.97

By Corollary 1, the original method scored the energy problem with every criterion treated as benefit-type, and the reversed implementation-cost criterion, on which the alternatives differ sharply, both drove the ranking away from the consensus and anchored it, so that changing the normalisation or the weights moved the ranking very little. Removing that domination makes the ranking responsive again to the criteria that the weights say should matter, which is the intended behaviour. Still, it also removes the artificial steadiness that the single dominant criterion had imposed. The increased sensitivity therefore reflects the genuine difficulty of the energy problem rather than a defect of the correction.

The two datasets together make a distinction that the robustness literature often leaves implicit. Stability, the tendency of a ranking to remain unchanged under perturbation, is not the same as validity, the tendency of a ranking to agree with the broader body of established methods. MEGAN on the energy problem was stable but divergent: steady under perturbation yet opposed to every conventional method. MEGAN-S is the reverse, where the data are hard, being more responsive to perturbation but aligned with the consensus. A method can therefore be stable for the wrong reason, when a misoriented criterion dominates the aggregation, and an apparent loss of stability can accompany a gain in validity. Resistance to rank reversal, which MEGAN-S retains in full on both datasets, is unaffected by the correction, since it derives from the Borda aggregation rather than from the direction-unification step, as stated in Proposition 2(iii).

3.5. Computational cost

The complexity analysis of Eq. (19) predicts that MEGAN and MEGAN-S require roughly Lm times the work of a single-pass method, and the measured runtimes in Table 9 confirm it. On the energy problem ($m = 14$, 57 weight scenarios), the median evaluation takes 6.4 ms for MEGAN against roughly 0.09 ms for the comparators, a factor of about 70; on the flotation problem (m = 10, 41 scenarios), the factor is about 50. MEGAN-S is indistinguishable from MEGAN in cost, since the correction removes one elementwise multiplication. Two practical remarks follow.

First, in absolute terms, even the most expensive method evaluates a full problem in a few milliseconds on a single core of a commodity machine, so computational cost is immaterial for one-off engineering selections. Second, the factor of Lm becomes relevant when a method is embedded in loops, as in the sensitivity protocols of this study, in Monte Carlo analyses, or in large-scale simulation benchmarks, where MEGAN-class methods multiply the total cost accordingly. Runtimes were measured as the median of 300 repeated evaluations per method (Python 3.11, NumPy/SciPy implementation, single x86-64 core); absolute values are indicative and machine-dependent, while the ratios track the operation counts of Eq. (19).

Table 9. Median runtime per full evaluation in milliseconds, measured over 300 repetitions per method on a single processor core, and the corresponding scenario counts. Absolute times are machine-dependent; ratios reflect Eq. (19).

Method	Energy (57 scenarios)	Flotation (41 scenarios)
MEGAN	6.35	4.27
MEGAN-S	6.27	4.43
COCOSO	0.10	0.10
AROMAN	0.09	0.09
RAM	0.08	0.08
TOPSIS	0.09	0.08
VIKOR	0.09	0.09
WASPAS	0.09	0.08
EDAS	0.10	0.09

3.6. Robustness depends on the structure of the problem

The clearest result of the benchmark is the gap between the two datasets, which can be traced to the way cost criteria are handled. The implementation that reproduces the published MEGAN rankings reorients the cost criteria during normalisation and then applies the benefit-cost vector a second time, which by Proposition 1 cancels the reorientation entirely: cost columns are not merely treated much like benefit columns; they are treated exactly as benefit columns. When the cost criteria of a problem discriminate only weakly among the alternatives, or point in directions that the benefit criteria echo, as in the flotation dataset with its four cost criteria on a one-to-nine rating scale, this has little visible effect. When a cost criterion sharply separates the alternatives and conflicts with the benefit criteria, as the implementation-cost criterion does in the renewable-energy dataset, where the reported cost reaches 195 while most other scores stay below 100, the same step rewards precisely the expensive alternatives that the remaining criteria penalise and pushes the ranking away from methods that conventionally treat cost.

This fits a theme that runs through recent benchmarking work. Baydaş et al. [22] compared nine methods on a large financial dataset and argued that the suitability of a method must be judged against the specific problem rather than assumed in advance. Cinelli et al. [2], in their taxonomy of MCDM methods, make a related point: method selection is itself a decision that depends on the features of the problem. The present results add a concrete case. A method built specifically for stable behaviour can still produce sharply different rankings on two problems, and the trigger is an ordinary structural feature, the discriminative power and orientation of the cost criteria, rather than anything unusual in the data.

3.7. Insensitivity claims need independent, multi-dataset testing

The normalisation and weighting insensitivity reported for MEGAN was not reproduced here, and recent studies make clear that this is the harder of the two claims to satisfy. Stević et al. [23] examined seven normalisation techniques on a vehicle-selection problem and reported that the choice of technique altered rankings in ways that single-normalisation studies had missed. Malefaki et al. [24] perturbed engineering sustainability data with one thousand Monte Carlo samples and found that the normalisation technique affected the stability of the results as much as the weighting did. Tien et al. [3] reached a comparable conclusion when they held the standardisation method fixed and still observed ranking differences across methods. Against this background, the present finding is not that MEGAN is unusually fragile, but that a method advertised as normalisation-insensitive behaved on real data much like the methods it was meant to improve on. This is the kind of claim that a single validation by the proposing author on a single synthetic dataset cannot settle, and it is why independent testing across multiple real datasets is needed.

A related interpretive caution concerns the role of agreement itself. Throughout this study, agreement with the panel of established methods is used as an external reference point because, in the absence of ground truth, it is the only available anchor; but consensus among methods does not establish which ranking is substantively correct. High congruence with existing methods does not prove that a ranking is right, and divergence does not by itself prove that an algorithm is wrong: a genuinely superior method would, by definition, disagree with its predecessors on some problems. What elevates the divergence observed here above that ambiguity is the mechanism established in Proposition 1: MEGAN's disagreement on the energy problem is not an alternative reading of the trade-offs but the arithmetic consequence of discarding the orientation of the cost criteria, a property no decision-maker would choose deliberately. Where no such mechanism can be identified, disagreement between methods should be treated as a signal for further analysis rather than as evidence of error, and agreement statistics should be read as measures of consistency with current practice, not of correctness.

3.8. Rank-reversal resistance was the property that held

The one claim that survived the benchmark was resistance to rank reversal. MEGAN, MEGAN-S, AROMAN, and RAM all left the top alternative unchanged across every removal trial, which explains the identical 0.0 per cent entries in Table 6: the four methods share the same structural property rather than coinciding by accident. This is consistent with how the four methods are built, since each aggregates scores computed against fixed, removal-invariant references rather than relying on comparisons between an alternative and a moving reference point: AROMAN and RAM sum weighted normalised scores directly, and MEGAN and MEGAN-S convert scenario scores into ranks and aggregate them by Borda count, so the removal of a non-optimal alternative cannot displace the leader. The same logic underlies methods designed to be rank-reversal-free, such as the Stable Preference Ordering Towards Ideal Solution (SPOTIS) [25], which avoids reversal by scoring each alternative only against fixed bounds. The contrast with COCOSO and EDAS, which did reverse, fits the same explanation from the opposite side: COCOSO divides by set-dependent minima and maxima in Eqs. (13) and (14), and EDAS scores against the set average, so removing an alternative shifts the reference frame itself. Resistance to rank reversal therefore comes from specific design choices rather than from the robustness-oriented label as a whole.

3.9. Practical guidance for method selection

The benchmark supports several concrete recommendations for practitioners, offered with the caveat that they derive from two datasets and should be re-examined as the evidence base grows. When the candidate set is dynamic, that is, when alternatives may enter or leave the shortlist during the decision process, rank-reversal resistance matters most, and score-aggregating designs with fixed references, such as AROMAN, RAM, and MEGAN-S, are preferable to reference-dependent designs such as COCOSO and EDAS, which reversed in up to half of the removal trials here. When the criterion weights are trusted, but the choice of normalisation is contested, AROMAN's built-in two-step normalisation and MEGAN-S on well-conditioned data showed the highest strict stability. At the same time, every method should be checked for the strictness gap documented in Table 7, since high rank correlation can coexist with frequent positional swaps. When the weights themselves are uncertain, the Monte Carlo results indicate that all five robustness-oriented methods tolerate moderate random weight noise well; The top-1 retention rate is nevertheless the statistic to inspect, and it varied from 67 to 100 per cent across methods on the harder problem. MEGAN in its published form should not be applied to any problem that contains cost criteria, because by Corollary 1 the method maximises them; users who value its gradual-weighting and Borda machinery should use the direction-consistent variant instead. Finally, the full protocol of this study runs in well under a minute on a commodity machine for problems of the size studied here, so an analyst can and should subject any candidate method to the same normalisation, weighting, Monte Carlo, and removal tests on the actual decision matrix at hand before relying on its ranking; the released code makes this a single command.

3.10. Implications and limitations

For practice, the main message is cautionary. A method that performs well in a published study, on one type of problem, should not be adopted for a different problem on that basis alone. The cost-criterion effect observed here is not evident in the original MEGAN article and emerged only when the method was applied to a dataset with a different criterion structure. Before relying on any MCDM method, an analyst is well advised to check its behaviour on the specific problem at hand, particularly when that problem involves cost criteria that strongly discriminate among the alternatives.

The study has limits, which are stated here explicitly. It uses two datasets, one synthetic and one real-world, and the flotation problem is small, with five alternatives, which caps the attainable significance of the pairwise permutation tests at $p = 0.017$ and renders signed-rank tests uninformative. The conclusions about cross-domain behaviour would be firmer with more datasets covering more domains, and the MEGAN-S correction in particular has been validated on only these two problems; testing it on problems containing several cost criteria that discriminate strongly among the alternatives, and with varying shares of total weight assigned to cost criteria, is required before its behaviour can be considered generally characterised. The analysis also rests on the implementation of MEGAN that reproduces the published rankings; Proposition 1 delimits exactly which implementations are affected, namely those that combine direction-aware normalisation with a second multiplication by the direction vector, and an implementation handling the cost criteria differently would behave differently. A fuller characterisation of that design space would be a useful contribution from the authors of the method; in that spirit, we note that this critique is offered as constructive scientific exchange, that the original author was not contacted before submission, and that a response or further specification from the method's proposer would be a welcome part of the ensuing scholarly discussion. Finally, agreement with established methods, used here as the external reference, measures consistency with current practice rather than substantive correctness.

4. Conclusions

This paper reported an independent, cross-domain benchmark of MEGAN against three other robustness-oriented, aggregation-based methods, COCOSO, AROMAN and RAM, with TOPSIS, VIKOR, WASPAS and EDAS as conventional reference methods. The methods were compared on a synthetic renewable-energy grid selection benchmark and a real-world flotation-machine selection problem. They were assessed for inter-method agreement, sensitivity to normalisation and weighting, and resistance to rank reversal. Three findings stand out. First, the agreement between MEGAN and the other methods depended heavily on the dataset: it was clearly positive on the flotation problem and strongly negative on the renewable-energy problem, and the difference was traced to the handling of the cost criteria, both empirically and by formal proof: the second direction-unification step of the published method exactly cancels the direction handling performed during normalisation, so that every cost criterion is in effect maximised (Proposition 1), a result verified computationally on both datasets. Second, the insensitivity to normalisation and weighting claimed for MEGAN was not reproduced on either dataset, where rankings were preserved in only 30 to 60 per cent of the perturbed settings; on the flotation problem MEGAN was in fact the least stable of the four robustness-oriented methods under changes of weight, although the graded analysis shows that all methods, MEGAN included, remain highly rank-correlated with their baselines under moderate random weight noise. Third, MEGAN showed full resistance to rank reversal across both datasets, confirming a key design feature that it shares with AROMAN, RAM, and the proposed variant by virtue of their common fixed-reference, score-aggregating construction.

Taken together, the results suggest that the robustness of an MCDM method is better understood as a property conditioned on the structure of the problem than as a fixed guarantee. For the renewable-energy and mineral-processing problems studied here, that conditioning was strong enough to reverse the apparent standing of MEGAN relative to the established methods. The study also proposed and evaluated a remedy. A direction-consistent variant of MEGAN, MEGAN-S, obtained by removing the single redundant direction-unification step

so that optimisation direction is settled once during normalisation, was benchmarked under the same protocol, and its formal properties, weight-bounded direction-consistent contributions, affine scale invariance, and retained rank-reversal resistance were established in Proposition 2. On the renewable-energy problem, where MEGAN had diverged, MEGAN-S realigned the rankings with the established methods, lifting the mean rank correlation with the seven comparators from −0.29 to +0.60 and turning every individual correlation positive; on the flotation problem, it raised an already-positive agreement from +0.61 to +0.97. The correction did not, however, increase stability everywhere: on the harder energy problem, MEGAN-S became more sensitive to perturbation because the mis-oriented cost criterion it removed had also been an artificial anchor. This contrast carries the study's broader methodological message: stability under perturbation and agreement with established methods are distinct properties, and a ranking can be steady for a reason that has nothing to do with its validity.

The limitations stated above directly chart the future work. The benchmark should be broadened to further datasets and engineering domains, in particular to problems with several strongly discriminating cost criteria and varying cost-weight shares, so that the conditions under which the correction helps can be mapped precisely; a dataset collection of that size would also support the cross-dataset statistical instruments, such as critical difference diagrams, that two datasets cannot. Beyond MEGAN, the protocol released with this article (agreement analysis, normalisation and weighting substitution, Monte Carlo weight perturbation, and leave-one-out removal) is directly applicable to any newly proposed aggregation-based method, and its routine application by independent researchers before adoption is the practice this study argues for.

Author Contribution Roles

Ton Nguyen Trong Hien: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing original draft, Review & Editing. Khoiriya Latifah: Formal analysis, Validation, Writing, Review & Editing. All authors have read and approved the final version of the manuscript.

Competing Interest Statement

The authors declare that they have no known financial or non-financial competing interests in any material discussed in this paper.

Data Availability Statement

All data analysed in this study are publicly available in the cited sources. The renewable energy grid decision matrix is taken from the original MEGAN article [10]; the flotation machine selection dataset originates in Štirbanović et al. [15] and is reproduced in the systematised form by Stanujkić et al. [16]. Both decision matrices are additionally reproduced in full in Tables 1 to 3.

Code Availability Statement

The complete Python implementation of MEGAN, MEGAN-S, all comparator methods, and the entire validation protocol, which regenerates every table and figure of this article from a single command, is openly available at

https://github.com/tonnguyentronghien/megan-s-benchmark.

Statement on the Ethical Use of AI Tools

During the preparation of this work, the authors used an AI assistant to improve language and formatting. The authors reviewed and edited the content as necessary and take full responsibility for the final version.

Funding Statement

No funding was received from any financial organisation to conduct this research.

References

[1] H. Taherdoost and M. Madanchian, "Multi-criteria decision making (MCDM) methods and concepts," *Encyclopedia*, vol. 3, no. 1, pp. 77-87, 2023, https://doi.org/10.3390/encyclopedia3010006.

[2] M. Cinelli, M. Kadziński, M. Gonzalez, and R. Słowiński, "How to support the application of multiple criteria decision analysis? Let us start with a comprehensive taxonomy," *Omega*, vol. 96, p. 102261, 2020, https://doi.org/10.1016/j.omega.2020.102261.

[3] D. H. Tien, D. D. Trung, and N. V. Thien, "Comparison of multi-criteria decision-making methods using the same data standardisation method," *Strojnícky časopis - Journal of Mechanical Engineering*, vol. 72, no. 2, pp. 57-72, 2022, https://doi.org/10.2478/scjme-2022-0016.

[4] W. Sałabun, J. Wątróbski, and A. Shekhovtsov, "Are MCDA methods benchmarkable? A comparative study of TOPSIS, VIKOR, COPRAS, and PROMETHEE II methods," *Symmetry*, vol. 12, no. 9, p. 1549, 2020, https://doi.org/10.3390/sym12091549.

[5] H.-J. Shyur and H.-S. Shih, "Resolving rank reversal in TOPSIS: a comprehensive analysis of distance metrics and normalization methods," *Informatica*, vol. 35, no. 4, pp. 837-858, 2024, https://doi.org/10.15388/24-INFOR576.

[6] D. D. Trung, N. Ersoy, T. V. Dua, and H. X. Thinh, "A comparative evaluation of data normalization techniques using different metrics: practical application to a MCDM method," *Manufacturing Review*, vol. 12, p. 19, 2025, https://doi.org/10.1051/mfreview/2025013.

[7] M. Yazdani, P. Zarate, E. K. Zavadskas, and Z. Turskis, "A combined compromise solution (CoCoSo) method for multi-criteria decision-making problems," *Management Decision*, vol. 57, no. 9, pp. 2501-2519, 2019, https://doi.org/10.1108/MD-05-2017-0458.

[8] S. Bošković, L. Švadlenka, S. Jovčić, M. Dobrodolac, V. Simić, and N. Bacanin, "An alternative ranking order method accounting for two-step normalization (AROMAN): a case study of the electric vehicle selection problem," *IEEE Access*, vol. 11, pp. 39496-39507, 2023, https://doi.org/10.1109/ACCESS.2023.3265818.

[9] A. Sotoudeh-Anvari, "Root Assessment Method (RAM): a novel multi-criteria decision making method and its applications in sustainability challenges," *Journal of Cleaner Production*, vol. 423, p. 138695, 2023, https://doi.org/10.1016/j.jclepro.2023.138695.

[10] C. Cebeci, "Multi-criteria evaluation with gradual-weighting and aggregation of normalised distance matrices: a case study in renewable energy grid selection," *PeerJ Computer Science*, vol. 12, p. e3819, 2026, https://doi.org/10.7717/peerj-cs.3819.

[11] C. L. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications*. Berlin, Germany: Springer-Verlag, 1981, https://doi.org/10.1007/978-3-642-48318-9.

[12] S. Opricovic and G. H. Tzeng, "Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS," *European Journal of Operational Research*, vol. 156, no. 2, pp. 445-455, 2004, https://doi.org/10.1016/S0377-2217(03)00020-1.

[13] E. K. Zavadskas, Z. Turskis, J. Antucheviciene, and A. Zakarevicius, "Optimization of weighted aggregated sum product assessment," *Elektronika ir Elektrotechnika*, vol. 122, no. 6, pp. 3-6, 2012, https://doi.org/10.5755/j01.eee.122.6.1810.

[14] M. Keshavarz Ghorabaee, E. K. Zavadskas, L. Olfat, and Z. Turskis, "Multi-criteria inventory classification using a new method of evaluation based on distance from average solution (EDAS)," *Informatica*, vol. 26, no. 3, pp. 435-451, 2015, https://doi.org/10.15388/Informatica.2015.57.

[15] Z. Štirbanović, D. Stanujkić, I. Miljanović, and D. Milanović, "Application of MCDM methods for flotation machine selection," *Minerals Engineering*, vol. 137, pp. 140-146, 2019, https://doi.org/10.1016/j.mineng.2019.04.014.

[16] D. Stanujkić, D. Karabašević, G. Popović, E. K. Zavadskas, M. Saračević, P. S. Stanimirović, A. Ulutaş, V. N. Katsikis, and I. Meidute-Kavaliauskiene, "Comparative analysis of the Simple WISP and some prominent MCDM methods: a Python approach," *Axioms*, vol. 10, no. 4, p. 347, 2021, https://doi.org/10.3390/axioms10040347.

[17] D. Diakoulaki, G. Mavrotas, and L. Papayannakis, "Determining objective weights in multiple criteria problems: the CRITIC method," *Computers & Operations Research*, vol. 22, no. 7, pp. 763-770, 1995, https://doi.org/10.1016/0305-0548(94)00059-H.

[18] M. Keshavarz-Ghorabaee, M. Amiri, E. K. Zavadskas, Z. Turskis, and J. Antucheviciene, "Determination of objective weights using a new method based on the removal effects of criteria (MEREC)," *Symmetry*, vol. 13, no. 4, p. 525, 2021, https://doi.org/10.3390/sym13040525.

[19] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145-151, 1991, https://doi.org/10.1109/18.61115.

[20] M. G. Kendall and B. Babington Smith, "The problem of m rankings," *The Annals of Mathematical Statistics*, vol. 10, no. 3, pp. 275-287, 1939, https://doi.org/10.1214/aoms/1177732186.

[21] B. Kizielewicz, A. Shekhovtsov, and W. Sałabun, "pymcdm-The universal library for solving multi-criteria decision-making problems," *SoftwareX*, vol. 22, p. 101368, 2023, https://doi.org/10.1016/j.softx.2023.101368.

[22] M. Baydaş, T. Eren, Ž. Stević, V. Starčević, and R. Parlakkaya, "Proposal for an objective binary benchmarking framework that validates each other for comparing MCDM methods through data analytics,"

PeerJ Computer Science, vol. 9, p. e1350, 2023, https://doi.org/10.7717/peerj-cs.1350.

[23] Ž. Stević, M. Baydaş, M. Kavacık, E. Ayhan, and D. Marinković, "Selection of data conversion technique via sensitivity-performance matching: ranking of small e-vans with the PROBID method," *Facta Universitatis, Series: Mechanical Engineering*, vol. 22, no. 4, pp. 643-671, 2024, https://doi.org/10.22190/FUME240305023S.

[24] S. Malefaki, D. Markatos, A. Filippatos, and S. G. Pantelakis, "A comparative analysis of multi-criteria decision-making methods and normalization techniques in holistic sustainability assessment for engineering applications," *Aerospace*, vol. 12, no. 2, p. 100, 2025, https://doi.org/10.3390/aerospace12020100.

[25] J. Dezert, A. Tchamova, D. Han, and J.-M. Tacnet, "The SPOTIS rank reversal free method for multi-criteria decision-making support," in *Proc. 2020 IEEE 23rd Int. Conf. Information Fusion (FUSION)*, 2020, pp. 1-8, https://doi.org/10.23919/FUSION45008.2020.9190347.