Does better weight estimation mean better decisions? A Monte Carlo assessment of RANCOM-ST

Anna Shkurina[1,2][0009−0001−2114−9985]

[1] Institute of Management, University of Szczecin,
ul. Cukrowa 8, 71-004 Szczecin, Poland
[2] Doctoral School of University of Szczecin,
ul. Mickiewicza 16, 70-383 Szczecin, Poland
anna.shkurina@phd.usz.edu.pl

Abstract. Weighting of criteria is a critical step in multi-criteria decision analysis (MCDA), yet the usefulness of a weighting method should be assessed in terms of decision reliability rather than only numerical accuracy. This study investigates the decision-level consequences of the statistical refinement introduced by the RANCOM-ST weighting procedure in comparison with the original RANCOM method. A large-scale Monte Carlo simulation framework is employed, systematically varying the number of criteria, expert judgment error, and the aggregation model (SAW and TOPSIS). The analysis focuses on the ability to correctly identify the best alternative, shortlist consistency, and full ranking similarity, complemented by a decision transition analysis that measures repaired and deteriorated decisions. The results show that RANCOM-ST significantly improves the probability of selecting the correct alternative, particularly under higher expert noise, while rarely degrading already correct decisions. However, the magnitude and stability of improvement depend on the aggregation model, with more predictable gains under linear aggregation. The findings clarify the relationship between weight accuracy and decision reliability and indicate conditions under which statistical weight correction is beneficial.

Keywords: RANCOM method· Subjective Weighting · MCDA.

1 Introduction

Decision-support methods are commonly evaluated in terms of how accurately they estimate model parameters; however, in practical applications the primary objective is not parameter recovery but decision reliability. In Multi-Criteria Decision Analysis (MCDA), the final outcome of the process is the ranking of alternatives or the selection of a single preferred option [3]. Consequently, even small changes in model inputs may or may not alter the decision itself. From a decision-maker's perspective, a weighting procedure is useful only if it leads to stable and reliable choices rather than merely numerically accurate parameters.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

A key source of variability in MCDA arises from the elicitation of criteria importance. When preferences are derived from expert judgments, uncertainty, hesitation, and cognitive bias inevitably affect the obtained weights [2]. Previous studies have shown that experts are often consistent in general preference structure but imprecise in fine distinctions between similarly important criteria. As a result, weighting errors do not always translate proportionally into ranking errors: in some cases the best alternative remains unchanged, while in others minor perturbations may alter the decision outcome [7].

The RANking COMparison (RANCOM) method was proposed as a simple subjective weighting approach based on ordinal comparisons between criteria [8]. Instead of relying on precise numerical assessments, it derives weights from comparative preference information, which reduces sensitivity to small judgment inconsistencies. Due to its computational simplicity and robustness to minor inconsistencies, the method has attracted attention as a practical alternative to more cognitively demanding elicitation procedures. An extension of this approach, RANCOM-ST, later introduced a statistical correction mechanism that adjusts the estimated weights using threshold-based feedback [6].

While such corrections are intended to improve the agreement between estimated and underlying preferences, improved weight accuracy does not necessarily imply improved decisions. In MCDA, the mapping between weights and rankings is nonlinear and depends on the aggregation model and the structure of the decision problem. Therefore, evaluating weighting procedures solely by comparing weight vectors may be misleading: a method can reduce weight estimation error while leaving the selected alternative unchanged, or conversely alter the decision despite small numerical differences.

The aim of this paper is not to propose a new weighting method but to investigate the decision-level consequences of the statistical refinement introduced by RANCOM-ST [6]. Specifically, we analyze whether the correction improves, preserves, or occasionally degrades the final decision when compared with the original RANCOM procedure. To enable controlled analysis, a large-scale Monte Carlo simulation framework is used, allowing systematic variation of the number of criteria, the level of expert judgment error, and the aggregation model.

The study is guided by the following research questions:

(RQ1) To what extent does the statistical correction introduced by RANCOM-ST improve the accuracy of identifying the best alternative (Top-1 Hit Rate) compared to standard RANCOM, and how does this improvement vary with the number of criteria and the level of expert error?

(RQ2) How does the choice of aggregation method (SAW vs. TOPSIS) mediate the effectiveness of the RANCOM-ST correction, and does the distance-based nature of TOPSIS amplify or attenuate the benefits of weight refinement relative to the linear SAW model?

(RQ3) Under what conditions does RANCOM-ST introduce a risk of degrading an otherwise correct RANCOM decision, and what is the net trade-off between rescued and deteriorated rankings across varying problem sizes and noise levels?

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

By addressing these questions, the study provides a methodological assessment of weighting refinement in MCDA and clarifies the relationship between weight accuracy and decision accuracy. The findings indicate when statistical correction is beneficial and when it may be unnecessary or potentially detrimental in decision-support applications.

The remainder of the paper is organized as follows. Section 2 presents the methodological background. Section 3 describes the simulation framework and experimental design. Section 4 reports and discusses the results. Finally, Section 5 concludes the paper and outlines directions for future research.

2 Methods

This section briefly summarizes the methods required for the experimental evaluation: the RANCOM and RANCOM-ST weighting procedures and the SAW and TOPSIS aggregation models. Only the elements necessary to understand the experimental protocol are presented, while full methodological details can be found in the original references.

2.1 RANCOM Weighting Method

The RANking COMparison (RANCOM) method derives criteria weights from ordinal preference information provided by an expert. Instead of requesting precise numerical assessments, the decision-maker supplies a ranking of criteria according to their importance. The method transforms this ranking into a set of pairwise comparisons and estimates weights by aggregating the comparative relations.

Let $\boldsymbol{\pi}$ denote a ranking of n criteria, where a lower position index indicates higher importance. The ranking is converted into a pairwise comparison structure indicating whether criterion i is preferred to criterion j. For each criterion i, the number of favorable comparisons is counted and normalized:

$$w_i = \frac{s_i}{\sum_{k=1}^{n} s_k},$$

where s_i denotes the aggregated preference score obtained from the comparisons. The normalization ensures

$$\sum_{i=1}^{n} w_i = 1, \qquad w_i \geq 0.$$

The procedure reduces the cognitive burden on the expert and mitigates the effect of minor inconsistencies, as small ranking perturbations affect only local comparison relations rather than the entire weight structure.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

2.2 RANCOM-ST Statistical Refinement

RANCOM-ST extends the original RANCOM procedure by introducing a statistical refinement step applied after initial weight estimation. The method uses calibration parameters derived from the distribution of expected estimation errors.

Let $\mathbf{w}^{\mathrm{R}} = (w_1^{\mathrm{R}}, \ldots, w_n^{\mathrm{R}})$ denote the weights obtained from RANCOM. The refined weights $\mathbf{w}^{\mathrm{ST}}$ are computed by shifting each component in a direction indicated by expert feedback:

$$w_i^{\mathrm{ST}} = w_i^{\mathrm{R}} + d_i \cdot c_i,$$

where $d_i \in \{-1, 0, 1\}$ represents the suggested direction of change (decrease, no change, increase) and c_i is a correction magnitude derived from statistical thresholds based on the mean μ_n and standard deviation σ_n of estimation errors. After correction, weights are truncated to non-negative values and renormalized:

$$w_i^{\mathrm{ST}} \leftarrow \frac{\max(w_i^{\mathrm{ST}}, 0)}{\sum_{k=1}^{n} \max(w_k^{\mathrm{ST}}, 0)}.$$

The intention of the refinement is to compensate systematic bias in ordinal-based weight estimation without requiring precise numerical judgments from the expert.

2.3 SAW Aggregation Model

The Simple Additive Weighting (SAW) method evaluates each alternative by a weighted sum of normalized criterion values. Let $\mathbf{D} = [d_{ij}]$ denote the decision matrix. The score of alternative i is

$$S_i = \sum_{j=1}^{n} w_j d_{ij}.$$

Alternatives are ranked in descending order of S_i. Due to its additive structure, the method translates changes in weights directly into proportional score changes.

2.4 TOPSIS Aggregation Model

The Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) ranks alternatives according to their relative closeness to the ideal and anti-ideal solutions [1].

The decision matrix is first normalized:

$$r_{ij} = \frac{d_{ij}}{\sqrt{\sum_{k=1}^{m} d_{kj}^2}},$$

and weighted:

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

$$v_{ij} = w_j r_{ij}.$$

The positive and negative ideal solutions are defined as

$$A^+ = (\max_i v_{ij}), \qquad A^- = (\min_i v_{ij}).$$

Distances to the ideal and anti-ideal points are computed using Euclidean metrics:

$$D_i^+ = \sqrt{\sum_{j=1}^{n} (v_{ij} - A_j^+)^2}, \qquad D_i^- = \sqrt{\sum_{j=1}^{n} (v_{ij} - A_j^-)^2}.$$

The relative closeness is

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-},$$

and alternatives are ranked in descending order of C_i.

3 Experiments

This section presents a Monte Carlo simulation study comparing RANCOM and RANCOM-ST with respect to their ability to reproduce decision outcomes implied by the ground-truth model. In particular, we investigate whether the statistical weight correction affects (i) the identification of the best alternative (Top-1), (ii) the composition of shortlists (Top-k for $k \in \{3, 5\}$), and (iii) the overall similarity of alternative rankings under varying numbers of criteria and different levels of expert error. Monte Carlo simulation has been widely applied to evaluate the stability and accuracy of MCDA methods under varying problem configurations [4,9].

3.1 Experimental Setup

Since the ground-truth weights are known by construction, each simulation run produces a reference ranking that serves as an objective benchmark for evaluating both methods. The experimental design varies three primary factors:

- Number of criteria: $n \in \{3, 4, 5, 6, 7, 8, 9, 10\}$,
- Expert noise level: adjacent swap probability $p_{swap} \in \{0.10, 0.25, 0.40\}$, corresponding to low, medium, and high error,
- Aggregation method: SAW (Simple Additive Weighting) and TOPSIS.

For each configuration, $T = 100{,}000$ Monte Carlo repetitions are performed using $m = 10$ alternatives. All runs share a fixed random seed to ensure reproducibility. The full factorial design yields $8 \times 3 \times 2 = 48$ experimental configurations and 4,800,000 simulation runs in total.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

Algorithm 1 Simulated Expert Ranking with Noise

Require: True weights $\mathbf{w}^* = (w_1^*, \ldots, w_n^*)$, swap probability p_{swap}
Ensure: Noisy expert ranking $\hat{\boldsymbol{\pi}}$

$\boldsymbol{\pi}^* \leftarrow \text{argsort}(-\mathbf{w}^*)$ ▷ Ideal ranking (descending)
$\hat{\boldsymbol{\pi}} \leftarrow \boldsymbol{\pi}^*$ ▷ Copy
for $i = 1$ **to** $n - 1$ **do**
  $a \leftarrow \hat{\pi}_i, \quad b \leftarrow \hat{\pi}_{i+1}$
  $\delta \leftarrow |w_a^* - w_b^*|$
  $p_{eff} \leftarrow \begin{cases} \min(p_{swap} + 0.15,\ 0.85) & \text{if } \delta < 0.05 \\ p_{swap} & \text{otherwise} \end{cases}$
  if $U(0,1) < p_{eff}$ **then**
    Swap $\hat{\pi}_i \leftrightarrow \hat{\pi}_{i+1}$
  end if
end for
return $\hat{\boldsymbol{\pi}}$

3.2 Data Generation Process

Each Monte Carlo iteration constructs a synthetic decision problem consisting of true criteria weights, a decision matrix, and an imperfect expert ranking.

First, a ground-truth weight vector $\mathbf{w}^* \in \mathbb{R}^n$ is sampled from a symmetric Dirichlet distribution, $\mathbf{w}^* \sim \text{Dir}(\alpha, \ldots, \alpha)$ with concentration parameter $\alpha = 1.0$. This produces a uniform distribution over the probability simplex, meaning that no particular importance structure is favored a priori. Consequently, the simulation spans a wide range of decision-maker preference structures, from nearly equal weights to strongly differentiated criteria.

Next, a decision matrix $\mathbf{D} \in \mathbb{R}^{m \times n}$ is generated with entries independently drawn from the uniform distribution $d_{ij} \sim U(0,1)$. All criteria are treated as benefit-type attributes (larger values preferred). This controlled setting isolates the influence of weight estimation errors on the final ranking without introducing additional effects associated with cost transformations.

Finally, expert judgment errors are introduced at the preference elicitation stage. The ideal ranking $\boldsymbol{\pi}^*$ is obtained by sorting criteria in descending order of their true weights. The observed expert ranking $\hat{\boldsymbol{\pi}}$ is then generated by stochastic adjacent transpositions, as specified in Algorithm 1. This noise model reflects bounded rationality: decision-makers are more likely to confuse criteria of similar importance than criteria with clearly different significance levels, thereby preserving the overall preference structure while introducing local inconsistencies.

3.3 RANCOM-ST Correction (Oracle Mode)

For each generated expert ranking, criteria weights are first estimated using the RANCOM method, yielding the weight vector $\mathbf{w}^{\mathrm{R}}$. An additional refinement step may then be applied using the RANCOM-ST statistical correction.

The correction is based on calibration parameters (μ_n, σ_n) obtained in Section 2. In the oracle feedback mode, the simulated expert perfectly identifies the

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

direction of deviation between the estimated and true weights. This configuration provides an upper bound on the achievable performance of RANCOM-ST and isolates the influence of the correction magnitude from potential feedback interpretation errors. The correction mechanism is specified in Algorithm 2.

Algorithm 2 RANCOM-ST Weight Correction (Oracle Mode)

Require: RANCOM weights $\mathbf{w}^{\mathrm{R}}$, true weights $\mathbf{w}^{*}$, calibration parameters (μ_n, σ_n)
Ensure: Corrected weights $\mathbf{w}^{\mathrm{ST}}$

$\mathbf{w}^{\mathrm{ST}} \leftarrow \mathbf{w}^{\mathrm{R}}$ ▷ Initialize
for $i = 1$ **to** n **do**
  $\delta_i \leftarrow w_i^{*} - w_i^{\mathrm{R}}$ ▷ True deviation
  if $|\delta_i| < 0.001$ **then**
    continue ▷ No correction needed
  end if
  $d_i \leftarrow \mathrm{sign}(\delta_i)$ ▷ Correction direction
  if $|\delta_i| < \mu_n - 0.5\sigma_n$ **then**
    $c_i \leftarrow \max(\mu_n - \sigma_n,\ 0)$ ▷ Low correction
  else if $|\delta_i| < \mu_n + 0.5\sigma_n$ **then**
    $c_i \leftarrow \mu_n$ ▷ Medium correction
  else
    $c_i \leftarrow \mu_n + \sigma_n$ ▷ High correction
  end if
  $w_i^{\mathrm{ST}} \leftarrow w_i^{\mathrm{ST}} + d_i \cdot c_i$
end for
$\mathbf{w}^{\mathrm{ST}} \leftarrow \max(\mathbf{w}^{\mathrm{ST}},\ \epsilon)$ where $\epsilon = 10^{-6}$ ▷ Non-negativity
$\mathbf{w}^{\mathrm{ST}} \leftarrow \mathbf{w}^{\mathrm{ST}} / \|\mathbf{w}^{\mathrm{ST}}\|_1$ ▷ Normalize
return $\mathbf{w}^{\mathrm{ST}}$

The corrected weights are subsequently used in the aggregation stage to compute the final ranking of alternatives.

3.4 Aggregation Methods

To evaluate how weight refinement affects final decision outcomes, two aggregation models with different structural properties are considered: SAW and TOPSIS (both introduced in Section 2).

SAW represents an additive decision model in which the influence of criteria weights on the final scores is direct and proportional. Consequently, any change in the weight vector translates linearly into score changes.

In contrast, TOPSIS determines preferences based on distances to ideal and anti-ideal solutions in the weighted space. Because the ranking depends on relative distances rather than a simple weighted sum, weight perturbations propagate through normalization and distance computation.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

3.5 Evaluation Metrics

For each iteration, the ideal ranking $\mathbf{R}^*$ (computed using the true weights $\mathbf{w}^*$) is compared with the rankings obtained using RANCOM and RANCOM-ST. Several complementary metrics are employed because different MCDA tasks emphasize different decision objectives: selecting a single best alternative, forming a shortlist, or preserving the overall ranking structure.

Top-1 Hit Rate measures how often the highest-ranked alternative coincides with the ideal choice,

$$\mathbf{1}[R_1^* = R_1^{\text{test}}].$$

This metric reflects choice problems, where the decision-maker must select exactly one option.

Top-k Overlap ($k \in \{3, 5\}$) evaluates agreement within the leading subset of alternatives,

$$\frac{|\text{Top}_k(\mathbf{R}^*) \cap \text{Top}_k(\mathbf{R}^{\text{test}})|}{k},$$

corresponding to screening scenarios in which a shortlist of promising candidates is required.

WS Coefficient [5] measures similarity of the full rankings while assigning greater importance to higher positions. Unlike Top-k, it evaluates the entire ordering but still prioritizes top-ranked alternatives.

Mean Rank Distance measures the average positional displacement across the ranking,

$$\frac{1}{m}\sum_{i=1}^{m} |\text{pos}^*(i) - \text{pos}^{\text{test}}(i)|,$$

capturing global ranking distortion independently of whether the best alternative is preserved.

Weight MAE is the mean absolute error between estimated and true weights,

$$\frac{1}{n}\sum_{j=1}^{n} |w_j^{\text{est}} - w_j^*|.$$

3.6 Simulation Procedure

The experiment follows a Monte Carlo protocol in which each iteration represents an independent decision-making instance. For every configuration of the number of criteria, noise level, and aggregation model, synthetic decision problems are repeatedly generated and evaluated using both RANCOM and RANCOM-ST. The overall performance measures are obtained by aggregating the results over all repetitions. The procedure is summarized in Algorithm 3.

The reported performance metrics correspond to averages over all Monte Carlo repetitions for each experimental configuration. Differences between RANCOM and RANCOM-ST are evaluated across identical simulation instances, ensuring paired comparisons under the same decision scenarios.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

Algorithm 3 Monte Carlo Simulation: RANCOM vs RANCOM-ST

Require: Parameter sets $\mathcal{N}$, $\mathcal{L}$, $\mathcal{A}$; iterations T; alternatives m
Ensure: Aggregated metric results for all configurations

for each $n \in \mathcal{N}$, $\ell \in \mathcal{L}$, agg $\in \mathcal{A}$ **do**
  for $t = 1$ **to** T **do**
    $\mathbf{w}^* \sim \mathrm{Dir}(1, \ldots, 1)$ ▷ Sample true weights
    $\mathbf{D} \sim U(0,1)^{m \times n}$ ▷ Generate decision matrix
    $\mathbf{R}^* \leftarrow \mathrm{agg}(\mathbf{D}, \mathbf{w}^*)$ ▷ Ideal ranking
    $\hat{\boldsymbol{\pi}} \leftarrow \mathrm{ExpertRanking}(\mathbf{w}^*, p_\ell)$ ▷ Alg. 1
    $\mathbf{MAC} \leftarrow \mathrm{RankingToMAC}(\hat{\boldsymbol{\pi}}, n)$
    $\mathbf{w}^{\mathrm{R}} \leftarrow \mathrm{RANCOM}(\mathbf{MAC})$
    $\mathbf{w}^{\mathrm{ST}} \leftarrow \mathrm{RANCOM\text{-}ST}(\mathbf{w}^{\mathrm{R}}, \mathbf{w}^*, n)$ ▷ Alg. 2
    $\mathbf{R}^{\mathrm{R}} \leftarrow \mathrm{agg}(\mathbf{D}, \mathbf{w}^{\mathrm{R}})$
    $\mathbf{R}^{\mathrm{ST}} \leftarrow \mathrm{agg}(\mathbf{D}, \mathbf{w}^{\mathrm{ST}})$
    Record metrics: $\mathrm{Compare}(\mathbf{R}^*, \mathbf{R}^{\mathrm{R}})$ and $\mathrm{Compare}(\mathbf{R}^*, \mathbf{R}^{\mathrm{ST}})$
  end for
end for

4 Results

4.1 SAW Aggregation Model

The results for the additive SAW model are presented in Table 1. Across all experimental configurations, the statistical refinement consistently improves both the estimated weights and the resulting decision rankings.

The most immediate effect of the correction appears at the parameter level. The Weight MAE is substantially reduced for every number of criteria and every noise level. For instance, for $n = 3$ under low noise the error decreases from 0.1131 to 0.0355. Similar reductions are observed throughout the table, typically by a factor of about three. This confirms that the calibration mechanism effectively moves the RANCOM estimates toward the true weight vector.

More importantly, the improvement in weight estimation translates into improved decision outcomes. The Top-1 hit rate increases in every configuration. Under low noise, the probability of selecting the optimal alternative rises from approximately 75–76% for RANCOM to about 86–92% for RANCOM-ST depending on the number of criteria. Even under high noise, where preference information is strongly distorted, the correction provides a substantial benefit (e.g., for $n = 3$ from 63.9% to 82.6%). This indicates that the statistical refinement is capable of recovering decision-relevant information even when the elicited ranking contains considerable local inconsistencies.

A similar pattern is visible for the Top-3 overlap. The composition of the shortlist becomes significantly more stable, with improvements typically exceeding 10 percentage points. Since many practical MCDA applications involve screening rather than strict ranking, this suggests that the correction primarily enhances the reliability of identifying promising alternatives.

The global ranking similarity, measured by the WS coefficient, also increases systematically. For small numbers of criteria the improvement is particularly

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

strong (e.g., from 0.8617 to 0.9576 for $n = 3$ under low noise). As the number of criteria increases, the gain remains positive but gradually decreases. This behavior reflects a structural property of additive models: when more criteria are present, individual weight errors have a diluted influence on the aggregated score because each criterion contributes a smaller portion of the total evaluation.

The magnitude of improvement depends jointly on the noise level and the number of criteria. The largest gains occur for small n, where inaccuracies in weights strongly affect the final scores. As n increases, the baseline performance of RANCOM improves and the relative advantage of RANCOM-ST becomes smaller. This indicates that decision sensitivity to weight estimation error is structurally dependent on the dimensionality of the decision problem.

Overall, the results reported in Table 1 demonstrate a monotonic relationship: reducing the weight estimation error leads to higher decision accuracy. However, the strength of this relationship is not constant. The benefit of statistical correction is greatest in low-dimensional problems and remains meaningful even under substantial expert noise, showing that decision reliability is considerably more sensitive to weight errors than suggested by weight-level metrics alone.

4.2 TOPSIS Aggregation Model

The results for the TOPSIS aggregation model are reported in Table 2. As in the additive case, the statistical refinement consistently reduces the weight estimation error and improves all decision-quality metrics. However, the magnitude and structure of the improvements differ from those observed for the SAW model.

At the parameter level, the behavior remains unchanged. The Weight MAE is reduced across all configurations by approximately a factor of three, similarly to the SAW results. For example, for $n = 3$ under low noise the error decreases from 0.1134 to 0.0356. This confirms that the calibration step improves weight estimation independently of the aggregation procedure.

At the decision level, the improvement in the Top-1 hit rate is again systematic. Under low noise, the probability of selecting the optimal alternative increases from approximately 70–74% for RANCOM to about 84–91% for RANCOM-ST depending on the number of criteria. Even under high noise, the correction provides a substantial improvement (e.g., for $n = 3$ from 59.5% to 80.3%). Therefore, the statistical correction remains beneficial even when the preference ranking is strongly perturbed.

The Top-3 overlap follows the same pattern, with improvements typically exceeding 10 percentage points. This indicates that the correction not only affects the exact ordering but also stabilizes the identification of promising alternatives.

Global ranking similarity measured by the WS coefficient also increases in every configuration. However, compared to SAW, the absolute WS values are consistently lower for both methods. This difference reflects the structural properties of TOPSIS: because rankings depend on distances to ideal and anti-ideal solutions, perturbations in weights propagate through normalization and distance computation in a nonlinear manner. Consequently, identical improvements

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

Table 1. Decision quality metrics for the SAW aggregation model

n	Noise	Top-1 Hit Rate		WS coefficient		Weight MAE		Top-3 overlap	
		R	ST	R	ST	R	ST	R	ST
3	low	76.2	92.0	0.8617	0.9576	0.1131	0.0355	84.3	95.0
3	medium	69.7	87.5	0.8141	0.9275	0.1449	0.0578	79.6	91.7
3	high	63.9	82.6	0.7662	0.8950	0.1778	0.0824	75.2	88.2
4	low	76.0	91.5	0.8603	0.9545	0.0847	0.0287	84.0	94.6
4	medium	70.5	87.5	0.8219	0.9299	0.1036	0.0418	80.1	91.8
4	high	65.1	83.1	0.7814	0.9006	0.1240	0.0575	76.1	88.5
5	low	76.0	90.8	0.8585	0.9490	0.0680	0.0257	83.7	93.9
5	medium	71.2	87.4	0.8287	0.9291	0.0795	0.0335	80.6	91.6
5	high	66.6	83.5	0.7942	0.9040	0.0931	0.0439	77.2	88.8
6	low	75.7	89.8	0.8562	0.9427	0.0563	0.0230	83.4	93.1
6	medium	72.1	87.3	0.8332	0.9271	0.0641	0.0282	81.0	91.2
6	high	68.0	83.8	0.8048	0.9055	0.0733	0.0351	78.0	88.7
7	low	75.0	88.7	0.8536	0.9362	0.0482	0.0211	83.0	92.2
7	medium	72.3	86.6	0.8347	0.9230	0.0537	0.0247	81.1	90.7
7	high	68.7	83.6	0.8110	0.9043	0.0604	0.0296	78.7	88.6
8	low	75.2	88.1	0.8523	0.9306	0.0421	0.0195	82.8	91.5
8	medium	72.3	86.0	0.8354	0.9184	0.0462	0.0222	81.2	90.1
8	high	69.5	83.3	0.8157	0.9022	0.0513	0.0259	79.1	88.3
9	low	74.7	87.1	0.8499	0.9249	0.0372	0.0180	82.5	90.8
9	medium	72.3	85.4	0.8360	0.9148	0.0404	0.0200	81.2	89.7
9	high	69.5	82.9	0.8181	0.9003	0.0444	0.0230	79.3	88.0
10	low	74.5	86.5	0.8490	0.9204	0.0333	0.0167	82.5	90.2
10	medium	72.5	84.6	0.8357	0.9103	0.0358	0.0183	81.1	89.1
10	high	70.0	82.6	0.8201	0.8979	0.0390	0.0206	79.5	87.8

R – original RANCOM weights; ST – statistically refined weights (RANCOM-ST). Noise levels correspond to adjacent-swap probabilities defined in Section 3.1.

in weight accuracy do not translate into equally large improvements in ranking similarity.

An important observation concerns the effect of the number of criteria. As n increases, the relative advantage of RANCOM-ST gradually decreases, similarly to the SAW case, but the decrease is more pronounced. This indicates that distance-based aggregation attenuates the influence of individual weight errors more strongly than additive aggregation. In other words, the decision sensitivity to weight estimation error is model-dependent.

Overall, the results in Table 2 confirm that improving weight estimation accuracy leads to better decision outcomes also in nonlinear aggregation models. Nevertheless, the relationship is weaker than in the additive case. This demonstrates that the impact of weight errors on decisions is not universal but depends on the decision mechanism itself. The statistical correction therefore improves

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

Table 2. Decision quality metrics for the TOPSIS aggregation model

n	Noise	Top-1 Hit Rate		WS coefficient		Weight MAE		Top-3 overlap	
		R	ST	R	ST	R	ST	R	ST
3	low	74.2	91.3	0.8475	0.9546	0.1134	0.0356	83.0	94.8
3	medium	67.0	86.1	0.7873	0.9189	0.1440	0.0573	77.3	90.9
3	high	59.5	80.3	0.7237	0.8757	0.1781	0.0827	71.3	86.2
4	low	73.6	90.8	0.8459	0.9526	0.0849	0.0289	82.6	94.5
4	medium	67.0	86.2	0.7957	0.9225	0.1034	0.0416	77.5	91.1
4	high	60.0	80.7	0.7401	0.8834	0.1238	0.0573	72.2	86.8
5	low	72.9	90.0	0.8429	0.9478	0.0677	0.0255	82.0	93.8
5	medium	67.5	85.9	0.8009	0.9221	0.0794	0.0334	77.7	90.7
5	high	61.5	81.0	0.7534	0.8871	0.0928	0.0437	73.0	86.9
6	low	72.3	88.8	0.8376	0.9401	0.0565	0.0231	81.2	92.8
6	medium	67.6	85.5	0.8022	0.9185	0.0643	0.0284	77.6	90.3
6	high	62.4	81.0	0.7630	0.8888	0.0735	0.0353	73.9	87.0
7	low	72.2	87.9	0.8335	0.9324	0.0481	0.0210	80.6	91.7
7	medium	67.7	84.8	0.8033	0.9134	0.0537	0.0247	77.7	89.6
7	high	63.2	80.9	0.7682	0.8876	0.0604	0.0296	74.2	86.6
8	low	71.1	86.6	0.8269	0.9235	0.0419	0.0193	79.9	90.6
8	medium	67.7	84.1	0.8022	0.9074	0.0462	0.0222	77.5	88.6
8	high	63.6	80.6	0.7708	0.8837	0.0511	0.0258	74.2	86.0
9	low	70.8	85.4	0.8222	0.9148	0.0372	0.0180	79.3	89.4
9	medium	67.5	83.3	0.8003	0.9005	0.0403	0.0200	77.1	87.8
9	high	64.1	80.0	0.7734	0.8798	0.0443	0.0228	74.4	85.5
10	low	70.0	84.3	0.8170	0.9062	0.0335	0.0168	78.7	88.3
10	medium	67.1	82.2	0.7966	0.8931	0.0358	0.0183	76.6	86.8
10	high	63.9	79.2	0.7732	0.8748	0.0390	0.0205	74.4	84.8

R – original RANCOM weights; ST – statistically refined weights (RANCOM-ST). Noise levels correspond to adjacent-swap probabilities defined in Section 3.1.

decision reliability, but the scale of the improvement is determined jointly by the level of expert noise and the structural properties of the aggregation model.

4.3 Decision transition analysis

To assess whether the statistical correction improves only average accuracy or also the reliability of individual decisions, a transition analysis between RANCOM and RANCOM-ST was performed. For each Monte Carlo run, the Top-1 alternative selected using RANCOM was compared with the alternative obtained after applying the RANCOM-ST correction. Three outcomes were distinguished: (i) the decision remained unchanged, (ii) an incorrect decision produced by RANCOM became correct after correction (rescued decision), and (iii) a previously correct RANCOM decision became incorrect after correction (deteriorated decision).

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

Table 3. Decision transitions between RANCOM and RANCOM-ST

Aggregation	Noise	RANCOM error rate	Rescued decisions	Deteriorated decisions
SAW	Low	24.7%	59.8%	1.2%
SAW	Medium	28.4%	55.2%	1.1%
SAW	High	32.2%	50.4%	1.1%
TOPSIS	Low	27.8%	60.6%	1.4%
TOPSIS	Medium	32.6%	56.3%	1.3%
TOPSIS	High	37.6%	50.6%	1.4%

Table 3 presents aggregated results for both aggregation models and all noise levels. A pronounced asymmetry between improvement and degradation can be observed. Across all conditions, RANCOM-ST repairs a substantial portion of erroneous RANCOM decisions, while only rarely deteriorating correct ones. Depending on the noise level, approximately 45%–70% of incorrect decisions are corrected, whereas only about 1%–2% of correct decisions become incorrect after the correction.

These results indicate that the statistical correction behaves as a conservative refinement rather than an aggressive modification of the decision model. The correction frequently improves incorrect outcomes but only exceptionally disrupts already correct decisions.

A systematic influence of the number of criteria is also observed. As the number of criteria increases, the fraction of rescued decisions gradually decreases. This behavior is consistent with the mechanism of the method: with a larger number of criteria, individual weights become smaller and closer to each other, reducing the relative impact of a threshold-based adjustment. At the same time, the deterioration rate remains nearly constant, indicating stable behavior of the correction.

A difference between aggregation models can also be identified. The deterioration rate is consistently higher for TOPSIS than for SAW, although the fraction of rescued decisions remains comparable. This is explained by the structural properties of the models. SAW is linear with respect to the weights, so local weight corrections produce proportional score changes. In contrast, TOPSIS is distance-based, and weight changes also affect the positions of the ideal and anti-ideal solutions, which may induce a cascade effect in the ranking. Consequently, the correction acts as a stable refinement under linear aggregation but as a more sensitive intervention under distance-based aggregation.

5 Discussion and Conclusions

This paper investigated the decision-level effects of the statistical refinement introduced by the RANCOM-ST procedure. Instead of evaluating the method solely in terms of weight estimation accuracy, the study focused on its influence on final decision outcomes, including the identification of the best alterna-

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

tive, shortlist stability, and overall ranking similarity. A large-scale Monte Carlo framework enabled controlled analysis across different numbers of criteria, levels of expert judgment error, and aggregation models.

The results show that the statistical correction generally improves decision performance, but its effectiveness is strongly context-dependent. With respect to RQ1, RANCOM-ST increases the Top-1 Hit Rate compared to the original RANCOM method, particularly for problems with a small number of criteria. In such cases, the applied correction represents a substantial portion of the weight magnitude and therefore meaningfully affects the ranking. For larger numbers of criteria, where weights are smaller and closer to each other, the relative influence of the correction decreases. The benefit of the correction grows with increasing expert noise: when expert judgments are already accurate, the original RANCOM performs well and the potential for improvement is limited, whereas under higher uncertainty the correction becomes more valuable.

Regarding RQ2, the aggregation model significantly mediates the effect of weight refinement. In the additive SAW model, weight errors translate directly and proportionally into score deviations; therefore, improvements in weights consistently lead to improved rankings. In contrast, TOPSIS exhibits more complex behavior. Due to the distance-based evaluation in the normalized space, small perturbations of weights may be absorbed, leading to greater robustness to minor inaccuracies. However, larger deviations can alter the relative positions of the ideal and anti-ideal solutions, causing non-local ranking changes. Consequently, RANCOM-ST produces more predictable and stable improvements when used with SAW, while in TOPSIS the effect is less regular but can be substantial under higher levels of noise.

In relation to RQ3, the correction mechanism introduces a measurable risk of deteriorating decisions. Because RANCOM-ST adjusts weights by a fixed statistically derived magnitude rather than the true individual error, a weight that is already close to the correct value may be shifted away from it. Such cases occur particularly when the expert input contains little noise or when the number of criteria is large and weights are relatively small. Nevertheless, the overall balance remains positive: the number of corrected ("rescued") decisions exceeds the number of degraded ones. This is expected, as the statistical thresholds are calibrated to the average population error rather than to individual instances.

Overall, the study demonstrates that improving weight estimates does not automatically guarantee improved decisions, but statistical refinement can increase decision reliability under appropriate conditions. The RANCOM-ST correction is most beneficial in problems characterized by moderate or high expert uncertainty and a limited number of criteria, whereas in low-noise settings its application may be unnecessary and occasionally detrimental.

The present work has several limitations. The analysis was conducted using synthetic decision problems with benefit-type criteria and simulated expert behavior. Although this approach enables controlled evaluation, real decision environments may involve heterogeneous criteria types and more complex cognitive effects. Future research should therefore include empirical case studies with

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35

human decision-makers and investigate adaptive correction magnitudes that depend on estimated uncertainty rather than fixed statistical thresholds.

In summary, this paper provides a methodological assessment of weighting refinement in MCDA and clarifies the relationship between weight accuracy and decision accuracy. The findings contribute practical guidance on when statistical correction should be applied and highlight the importance of evaluating decision-support methods at the level of decisions rather than parameters alone.

Acknowledgments. Publication funded by the Minister of Science under the "Regional Excellence Initiative" Program RID/SP/0046/2024/01.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Behzadian, M., Otaghsara, S.K., Yazdani, M., Ignatius, J.: A state-of the-art survey of TOPSIS applications. Expert Systems with Applications **39**(17), 13051–13069 (2012). https://doi.org/10.1016/j.eswa.2012.05.056
2. Brunelli, M.: A survey of inconsistency indices for pairwise comparisons. International Journal of General Systems **47**(8), 751–771 (Nov 2018). https://doi.org/10.1080/03081079.2018.1523156
3. Cinelli, M., Kadziński, M., Gonzalez, M., Słowiński, R.: How to support the application of multiple criteria decision analysis? let us start with a comprehensive taxonomy. Omega **96**, 102261 (Oct 2020). https://doi.org/10.1016/j.omega.2020.102261
4. Kosareva, N., Krylovas, A., Zavadskas, E.K.: Statistical analysis of MCDM data normalization methods using Monte Carlo approach. The case of ternary estimates matrix. Economic Computation and Economic Cybernetics Studies and Research **52**(4), 159–175 (2018). https://doi.org/10.24818/18423264/52.4.18.11
5. Sałabun, W., Urbaniak, K.: A new coefficient of rankings similarity in decision-making problems. In: Krzhizhanovskaya, V.V., Závodszky, G., Lees, M.H., Dongarra, J.J., Sloot, P.M.A., Brissos, S., Teixeira, J. (eds.) Computational Science – ICCS 2020. pp. 632–645. Springer International Publishing, Cham (2020)
6. Shkurina, A.: An adaptive rancom-st method for bias reduction using statistical thresholds. In: International Conference on Computational Science. pp. 281–295. Springer (2025). https://doi.org/10.1007/978-3-031-97567-7_22
7. Singh, M., Pant, M.: A review of selected weighing methods in mcdm with a case study. International Journal of System Assurance Engineering and Management **12**(1), 126–144 (Feb 2021). https://doi.org/10.1007/s13198-020-01033-3
8. Więckowski, J., Kizielewicz, B., Shekhovtsov, A., Sałabun, W.: RANCOM: A novel approach to identifying criteria relevance based on inaccuracy expert judgments. Engineering Applications of Artificial Intelligence **122**, 106114 (Jun 2023). https://doi.org/10.1016/j.engappai.2023.106114
9. Żak, J., Kruszyński, M.: Application of AHP and ELECTRE III/IV methods to multiple level, multiple criteria evaluation of urban transportation projects. Transportation Research Procedia **10**, 820–830 (2015). https://doi.org/10.1016/j.trpro.2015.09.035

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_35