Gradus Vol 13, No 2 (2026)
ISSN 2064-8014
Article no. 2026.2.CSC.005

COMPUTATIONAL RECONSTRUCTION, AUDITABILITY AND DECISION STABILITY IN PUBLISHED MULTICRITERIA MODELS FOR ENTERPRISE RESOURCE PLANNING SELECTION: A CROSS-CASE BENCHMARK

Vasile Paul Bresfelean [1], Silviu Claudiu Popa [2], Zsolt Csaba Johanyák [3,4] and George Sebastian Chis [5]*

[1] Faculty of Economics and Business Administration, Babeș-Bolyai University, Romania, https://orcid.org/0000-0002-7240-6187
[2] Faculty of Economics and Business Administration, Babeș-Bolyai University, Romania, https://orcid.org/0009-0005-1062-3396
[3] Department of Information Technology, GAMF Faculty of Engineering and Computer Science, John von Neumann University, Hungary, https://orcid.org/0000-0001-9285-9178
[4] Institute of Mechatronics and Vehicle Engineering, Bánki Donát Faculty of Mechanical and Safety Engineering, Óbuda University, Hungary
[5] Faculty of Economics and Business Administration, Babeș-Bolyai University, Romania, https://orcid.org/0009-0006-4106-781X

https://doi.org/10.47833/2026.2.CSC.005

Keywords:
Enterprise resource planning
Multicriteria decision analysis
Computational reconstruction
Auditability
Decision stability

Article history:
Received 29.09.2026
Revised 28.10.2026
Accepted 29.10.2026

Abstract

We examined whether published enterprise resource planning (ERP) recommendations based on multicriteria decision analysis (MCDA) could be reconstructed from public evidence and how they responded to changes in weights and criteria. A purposive benchmark covered 12 distinct decision-problem lineages under common reconstruction and evaluability rules. Nine lineages had an evaluable terminal baseline; all nine reproduced the published winner or native recommendation, and eight reproduced the complete ranking. Terminal-weight perturbation was applicable to seven lineages, and single-criterion deletion yielded evaluable results for all eight applicable lineages. In NP11, five of six deletions retained the baseline winner; removing C6 changed it from S4 to S3. High retention under weight perturbation did not consistently extend to criterion deletion. Source-defined analyses provided additional evidence: one retained the baseline winner in 26 of 30 scenarios, while a separate audit found workbook inconsistencies that affected scores but not rankings or winners. Another case reproduced the published Q scores and recommendations at five parameter settings. These results distinguish output availability, reconstructability, robustness auditability and stability under a specified diagnostic.

* Corresponding author.
E-mail address: paul.bresfelean@econ.ubbcluj.ro

1 Introduction

Enterprise resource planning (ERP) package selection requires organizations to balance technical capability, organizational fit, implementation demands, cost, provider characteristics and other competing criteria [1] [2] [3]. Multicriteria decision analysis (MCDA) provides a natural framework for these tradeoffs. Comparable tradeoffs among competing objectives also arise in engineering design, for example when trajectory planning for a connected autonomous electric bus has to balance energy consumption, passenger comfort and tracking performance [4]. ERP applications have used hierarchy and network models [5] [6] [7] as well as software-evaluation and quality-based approaches [8] [9]. Yet public numerical records differ substantially in how well they allow an independent reader to recover the operative computation behind the final recommendation.

A published winner or ranking does not itself establish computational reconstructability. Research on reproducible computation emphasizes disclosure of inputs, transformations, software choices, provenance and implementation details and distinguishes independent reimplementation from merely rerunning an available artifact [10] [11] [12] [13]. In ERP-MCDA studies, even a complete score table can coexist with an underdetermined computational chain when the public record omits a direction map, tie rule, random state, reduced-model convention or required supplementary matrix. Conversely, recovering the published winner can still leave the complete ranking or terminal scores unmatched.

Decision stability is a separate question. MCDA research distinguishes sensitivity to weights or performance values [14] from broader robustness analysis [15], uncertainty modeling [16] and rank-reversal phenomena [17] [18]. ERP studies include simulation-based stability analysis [19], comparisons of decision models [20] and VIKOR parameter sweeps [21]. Related work examines intuitionistic fuzzy VIKOR [22] and worst-case ERP reanalysis [23]. These precedents establish sensitivity analysis as a standard tool; the unresolved issue is how reconstructability, auditability and stability relate across distinct published cases.

ERP-MCDA methods have continued to diversify through fuzzy, rough, intuitionistic, hesitant and group-decision formulations within a broader MCDA field that has moved from classical methods toward more robustness-oriented approaches [24] [25]. Recent ERP studies extend criteria evaluation and package selection with uncertainty-aware and hybrid models [26] [27] [28] [29], while work from 2025-2026 adds entropy-based weighting [30], hesitant or probabilistic formulations [31], Pythagorean-fuzzy models [32] and fuzzy or gray models [33]. Methodological sophistication, however, does not determine whether a public case can be reconstructed or whether its sensitivity claims can be audited.

We analyze a purposively selected, lineage-aware benchmark of 12 published ERP-MCDA decision problems. The benchmark is analytical rather than representative, so its fractions are not prevalence estimates. We address three research questions: how far terminal recommendations, rankings and reported numerical outputs can be reconstructed from public evidence (RQ1); how stable evaluable recommendations are under multiplicative terminal-weight perturbation and method-compatible single-criterion deletion (RQ2); and how source-defined sensitivity analyses and limits in robustness auditability qualify those stability conclusions (RQ3).

The comparison separates public computational depth from terminal evaluability and assesses winner, complete ranking and score agreement independently. We evaluate diagnostic applicability before reporting stability, retain each method's decision object and use case-specific denominators. Source-defined analyses complement the two common diagnostics. Together, these comparisons show which published results can be reconstructed and how their decisions respond to the changes examined.

2 Related work and positioning

2.1 ERP selection as a multicriteria decision problem and the evolution of method families

ERP selection draws on MCDA because organizations must reconcile diverse package attributes. Early work examined staged buying processes [1], selection criteria in manufacturing firms [2] and tradeoffs among enterprise-system attributes [3]. Applications then used the analytic hierarchy

process (AHP) [5], the analytic network process (ANP) with fuzzy judgments [6], ANP-based benchmarking [7] and fuzzy AHP with a balanced scorecard [34]. Software-evaluation models [8], ISO 9126-based fuzzy AHP [9] and AHP applications [35] broadened this foundation. AHP-based decision-support prototypes [36] and applications for small and medium-sized enterprises [37] [38] provide further context.

ERP-MCDA applications span several computational architectures. Examples include linguistic information processing [39], ANP combined with VIKOR [40], fuzzy quality function deployment [41], models with interacting criteria [42] and hybrid fuzzy methods [43]. Other approaches use behavioral decision rules [44], group sorting [45], fuzzy ANP with goal programming [46], causal and interdependence analysis [47] or integrated uncertainty modeling [25].

Ranking models combine fuzzy AHP with the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [48], ANP with the Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) [49] or fuzzy linguistic preference relations with TOPSIS [50]. Integrated fuzzy group-decision methods [51], AHP [52], fuzzy AHP-TOPSIS [53] and logarithmic fuzzy preference programming (LFPP) [54] provide further ERP applications. Other approaches produce a selected system [55], link ANP to an artificial neural network [56] or use the Simple Multi-Attribute Rating Technique (SMART) [57].

Recent applications include complex intuitionistic fuzzy Step-Wise Weight Assessment Ratio Analysis with Complex Proportional Assessment (SWARA-COPRAS) [58], intuitionistic-fuzzy evaluation [27], rough Best-Worst Method with Weighted Integrated Sum Product (BWM-WISP) [28] and fuzzy AHP-TOPSIS for open-source ERP selection [29]. Studies from 2025–2026 add entropy-based weighting [30], probabilistic-hesitant fuzzy COPRAS [31], two-tuple Pythagorean fuzzy Multi-Objective Optimization by Ratio Analysis (MOORA) [32] and fuzzy or gray knowledge-management formulations [33]. A recent synthesis reviews the broader development of MCDA [24]. Direct fuzzy-AHP criteria evaluation also remains in use [26]. This variety makes the operative computational chain more informative for reconstruction than a method's nominal complexity.

2.2 Sensitivity, robustness and rank-stability precedents

Sensitivity, robustness, uncertainty and rank stability are related but distinct concepts. Sensitivity analysis examines how changes in weights or performances affect outcomes [14]. Robustness is a broader decision-aiding concept [15], uncertainty can be represented through several structures [16] and rank reversal concerns changes in ordering when the decision context changes [17] [18]. Interval fuzzy modeling, for instance, describes uncertain system behavior through lower and upper bounding functions rather than a single crisp model [59]. Terminal-weight perturbation, single-criterion deletion and source-defined analyses therefore answer different questions and are reported separately.

ERP studies have examined simulation-based AHP stability [19], comparisons of ERP decision models [20] and sensitivity and robustness in a logistics application [60]. A hybrid ERP selection model has also been evaluated in an industrial setting [61]. Other work examines VIKOR parameter sweeps [21], intuitionistic fuzzy extensions [22], worst-case ERP reanalysis [23] and uncertainty modeling [58]. Rough and fuzzy ERP studies also include scenario-based or source-defined sensitivity analyses [28] [29]. These studies cover individual elements of our design but do not resolve how reconstructability, robustness auditability and stability behave across independent decision lineages under a common reporting framework.

2.3 Computational reproducibility and the cross-case research gap

Computational reproducibility provides a second foundation for the cross-case audit. Public inputs, provenance, parameters, software versions, intermediate results, random states where relevant and executable artifacts determine whether a computation can be checked [10] [11] [12]. Optimization-based parameter tuning illustrates why such details matter: when fuzzy cognitive maps are adapted with particle swarm optimization, the result depends strongly on the initialization method [62]. Rerunning available code differs from independent reproduction and reuse [13]. Incomplete public model descriptions can obstruct independent reimplementation, as documented in another computational field [63].

Published ERP cases have been reused to demonstrate alternative MCDA formulations [64] [65]. An ANP-PROMETHEE case [49] was revisited using a robust worst-case formulation [23], and different decision models have been compared for ERP package selection [20]. Other studies examine simulation-based stability [19], VIKOR parameter sensitivity [21] and source-defined weight scenarios [28]. AHP applications [35] and decision-support prototypes [36] provide additional comparisons of computational scope. These studies inform our comparison of reconstructability, agreement and diagnostic-specific stability across independent decision lineages.

3 Methods

We constructed a computational benchmark of published ERP-MCDA decision problems. The unit of analysis is a decision-problem lineage: publications or artifacts representing the same underlying problem are treated as one case. Benchmark membership was fixed at 12 lineages before reconstruction and robustness outcomes were examined. The research questions address terminal reconstruction, robustness auditability and stability under specified diagnostics. The computational-depth and output classifications defined below were developed for this benchmark.

3.1 Benchmark construction and lineage rule

We selected cases to span different MCDA architectures and levels of public computational detail, while requiring ERP-selection relevance, independent decision lineages and feasible reconstruction from the available evidence. Cases with source-defined sensitivity analyses broadened the comparison. The resulting purposive sample supports case comparisons rather than estimates of reporting practices across the ERP-MCDA literature. Table 1 identifies the 12 lineages as NP01-NP12 (NP = numerical problem).

Table 1. Benchmark, lineage structure and public computational depth

Panel A. Identification, public depth and terminal evaluability					
NP	**Study**	**Method**	**D**	**O**	**Terminal evaluability**
NP01	Kılıç et al. [48]	Fuzzy AHP + TOPSIS	D3	O3	Evaluable
NP02	Motaki & Kamach [52]	AHP	D3	O3	Evaluable
NP03	Thanh [53]	Fuzzy AHP + TOPSIS	D2	O3	Evaluable from public checkpoint
NP04	Cao et al. [28]	Rough BWM + Rough WISP	D3	O3	Evaluable
NP05	Akpinar et al. [31]	Probabilistic/hesitant fuzzy COPRAS	D3	O3	Evaluable
NP06	Gürbüz et al. [42]	ANP + MACBETH + Choquet integral	D2	O3	Evaluable
NP07	Kılıç et al. [49]; Temur & Bolat [23]	ANP + PROMETHEE; linked CBDO reanalysis	D2/D3	O3/O2	Partially evaluable
NP08	Liao et al. [39]	2-tuple linguistic + LP/distance ranking	D3	O3	Not terminally evaluable
NP09	Mahapatra et al. [32]	Pythagorean fuzzy MOORA	D3	O3	Not evaluable: literal method non-finite
NP10	Deb et al. [27]	Intuitionistic-fuzzy IMARCOS	D0	O3	Not evaluable with examined inputs
NP11	Panjavongroj & Phruksaphanrat [54]	LFPP ERP ranking; DEMATEL separate	D3	O3	Evaluable; decision identified
NP12	Başar & Arslan [21]	VIKOR	D3	O3	Evaluable

Panel B. Lineage handling and public-source limitations		
NP	**Lineage handling**	**Primary public-source limitation**
NP01	One lineage	Score-level mismatch despite winner and complete ranking agreement
NP02	One lineage	Score-level mismatch despite winner and complete ranking agreement
NP03	One lineage; strict branch primary; plausible-direction interpretation secondary	Unresolved direction coding for ERP04 (implementation time) at public checkpoint
NP04	One lineage	Source-workbook implementation inconsistencies; rounding-sensitive tie in scenario S16
NP05	One lineage; 6 co-primary configurations	Six configurations; source criterion C6 is the sole cost criterion
NP06	One lineage	Reduced Choquet capacity not determined from examined inputs
NP07	One lineage; 2 linked publications	Exact repetition of the linked stochastic reanalysis not possible with examined inputs
NP08	One lineage	Non-unique LP; no source tie-break
NP09	One lineage	Literal chain reaches ln(0); no documented numerical convention
NP10	One lineage	Insufficient information obtained for terminal reconstruction
NP11	One lineage	Nonunique LFPP weights and scores; six criterion deletions evaluated
NP12	One lineage	Native recommendation-set rule governs diagnostics

Note. NP01-NP12 are benchmark identifiers. D and O are defined in Section 3.2. For NP10, the source reports O3 outputs [27], but the information available to us supported only D0. CBDO = cloud-based design optimization; DEMATEL = Decision-Making Trial and Evaluation Laboratory.

The lineage rule counts repeated publications on the same decision problem only once. NP07 comprises the ANP-PROMETHEE baseline [49] and its linked cloud-based design optimization (CBDO) reanalysis [23]. They form one lineage, while each retains its own public computational-depth classification. The two direction interpretations for NP03 and the six co-primary configurations for NP05 are likewise analytic branches within a lineage rather than additional benchmark members.

The 12 lineages span AHP; fuzzy AHP-TOPSIS; rough BWM-WISP; probabilistic-hesitant fuzzy COPRAS; ANP with Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) and a Choquet integral; ANP-PROMETHEE with a linked stochastic reanalysis; linguistic linear programming (LP); Pythagorean-fuzzy MOORA; intuitionistic-fuzzy improved Measurement Alternatives and Ranking according to Compromise Solution (IMARCOS); LFPP-based ranking; and VIKOR compromise ranking. This breadth tests the audit framework across computational conventions rather than comparing method-family performance.

3.2 Public computational depth and terminal evaluability

Two independent descriptors summarize what the public record exposes. Computational depth (D) records how far the public evidence supports the operative chain: D0, insufficient evidence to specify a complete terminal chain at a defensible computational point; D1, a near-terminal or intermediate point that supports final-chain reconstruction; D2, a complete alternatives-by-criteria matrix or equivalent standardized point with the operative weights and aggregation inputs; and D3, the earliest public numerical inputs used by the native method together with the method specification needed to attempt its chain. Published terminal output (O) records what the source reports: O0, no evaluable terminal output; O1, winner only; O2, a complete ranking without a complete comparable score vector; and O3, a complete ranking plus terminal numerical outputs. D3 describes exposure depth, not guaranteed uniqueness or finiteness.

D and O are independent of terminal evaluability. NP10 illustrates the distinction: published outputs are O3 [27], while independent reconstruction reached only D0 with the information available to us. NP08 and NP09, by contrast, retain D3/O3 depth descriptors but are not evaluable (NE): NP08's linear program has zero-objective optima that yield different decisions, without a source tie-break, while NP09's literal chain becomes non-finite at ln(0). NP11 remains D3/O3 and has an identified baseline winner and complete ranking despite nonunique LFPP weights and scores. NE

denotes a result that is not evaluable, not a negative agreement outcome. In diagnostic columns it also covers cases in which the diagnostic cannot be applied under the stated source-method rules.

Terminal evaluability requires a defensible public computational point and a finite, identified native decision. For optimization-based chains, nonunique weights or scores are compatible with an evaluable decision if the winner and complete ranking are invariant over the global optimum set. NP11 satisfies this criterion. Score agreement is assessed separately at the documented computational point.

3.3 Reconstruction, agreement and source conventions

For each evaluable lineage, we independently reconstructed the selected public computational point through the terminal recommendation. Agreement was assessed at three increasingly demanding levels: winner agreement for the best alternative or native recommendation, complete ranking agreement for the entire published ordering and score agreement for comparable terminal numerical outputs at the defined computational precision. A matching winner does not establish numerical reconstruction of the full model; a score discrepancy need not change the decision.

NP05 illustrates the distinction at the score level: all six co-primary configurations reproduced the winner and complete ranking, but only two matched the terminal scores at the published precision. NP03 illustrates the distinction at the ranking level: the winner was retained although the complete ranking was not under the stated public-checkpoint reconstruction.

NP08 and NP09 were NE under their stated mathematical models, while NP10 was NE on the information available to us. All three were excluded from agreement denominators.

For NP03, the primary branch followed the explicitly reported criterion directions, while a secondary branch treated ERP04 (implementation time) as a cost criterion. For NP04, we evaluated the 30 scenarios defined in [28] separately from the workbook implementation. The source labels the scenarios S1-S30; S16 contains a rounding-sensitive tie. These are within-lineage comparisons, so each case contributes once to the benchmark.

3.4 Applicability and non-evaluable robustness cases

We assessed applicability from the diagnostic's compatibility with the published method, the inputs available for recomputation and the native decision rule. Execution status, mathematical identification and decision retention were recorded separately. For models with nonunique optimal weights, deletion outcomes were evaluated by checking whether the decision was invariant over the global optimum set.

Single-criterion deletion was applicable to eight lineages (NP01, NP02, NP03, NP04, NP05, NP07, NP11 and NP12), all of which yielded evaluable deletion results.

Other exclusions were method-specific. NP04's rough interval weights did not support the common scalar terminal-weight perturbation. For NP06, the inputs examined did not determine a unique reduced Choquet capacity or a model for the interaction-off comparison. In NP05, source criterion C6 was the sole cost criterion, so only five criteria were eligible for deletion in each configuration.

3.5 Terminal-weight perturbation analysis

Terminal-weight perturbation was applicable to seven lineages: NP01, NP02, NP03, NP05, NP07, NP11 and NP12. Terminal weights are the scalar weights entering the final aggregation stage. At each relative amplitude $\delta \in \{0.05, 0.10, 0.20, 0.30, 0.40\}$, we multiplied each weight by an independent factor drawn from Uniform$[1-\delta, 1+\delta]$ and renormalized the vector. Each branch or configuration used 10,000 perturbed vectors per δ, with the downstream model and native decision rule preserved. Retention is the fraction of trials that preserved the baseline winner or native recommendation. It describes this stress test, not the probability that an alternative is truly optimal. For NP11, the fractions are conditional on the documented criterion-weight vector and local ERP weights; the perturbations do not sample all LFPP optima.

NP03 used the explicit-direction implementation as the primary branch and a secondary interpretation treating ERP04, the source's implementation-time criterion, as a cost criterion. This

comparison quantifies the effect of unresolved direction coding. Both NP03 branches and all six NP05 configurations were evaluated separately within their respective lineages.

3.6 Single-criterion deletion analysis

Single-criterion deletion tests whether the baseline decision survives removal of one eligible criterion followed by method-compatible recomputation. The exact operation depended on the selected computational point: native local reweighting for NP01-NP02; deletion from the normalized TOPSIS point with weight renormalization and terminal recomputation for both NP03 direction branches; the literal Rough-WISP chain for NP04; configuration-specific entropy-distance weighting and COPRAS for NP05; method-compatible recomputation of the NP07 baseline [49]; reduced LFPP for NP11; and VIKOR at q/v = 0.5 for NP12.

The NP12 source [21] uses v in the VIKOR equation and q in the sensitivity table for the same compromise parameter, denoted q/v here. At 0.5, group utility and individual regret receive equal weight.

For NP11, all six eligible criterion deletions were evaluated in a post hoc sensitivity analysis. For each deletion, we removed the corresponding row and column from the reassessed criteria-comparison matrix, solved the five-criterion LFPP with M = 1000 and held the baseline local ERP alternative weights fixed for the remaining criteria. We established each ranking by bounding pairwise terminal-score differences over the global optimum set, using independent convex solvers and global optimality certificates. The six rankings were unchanged when the published rounded local weights were used.

We retained case-specific denominators because the criteria sets and recomputation operations differ. Winner retention is the primary outcome except for NP12, which uses its native recommendation set. NP05 also reports top-two retention to capture movement immediately below the winner. No pooled deletion proportion is calculated.

3.7 Source-defined sensitivity analyses

Source-defined analyses follow the sensitivity designs in the original publications. For NP04, we evaluated the 30 scenarios generated by the weight-adjustment rule in Equation (31) of [28]. For NP12, we evaluated the five q/v settings reported in [21]: 0, 0.25, 0.5, 0.75 and 1. We compared Q scores with Table 6 of that study and rankings and acceptability conditions with its Table 7. NP06 reported an interaction-off comparison, but we could not reconstruct it from the materials we examined; we therefore retained it as source-reported evidence.

The three diagnostics have different designs and denominators. Terminal-weight perturbation samples weight vectors, deletion examines every eligible one-at-a-time removal under the case-specific recomputation rule, and source-defined analyses reproduce the scenarios or parameter settings specified in the original studies. Their retention fractions are therefore reported separately.

3.8 Computational implementation and reproducibility

We reconstructed baseline decisions and robustness diagnostics from standardized inputs extracted from the published sources. For NP04, reconstruction began with the managers' ratings and published global rough weights; the upstream BWM optimization was not re-estimated. Terminal-weight perturbation used a seeded random-number generator. The code generates the table and figure data from computational results and documented case classifications. Software versions and the NP11 global optimality and ranking certificates are available in the repository.

4 Results

Tables 2 and 3 report reconstruction and robustness results. Figure 1 summarizes evaluability, agreement and diagnostic coverage across the 12 lineages.

4.1 Reconstruction and evaluability

Nine of 12 lineages contained at least one evaluable terminal baseline. The published winner or native recommendation was retained in all nine, and the complete ranking agreed in eight; NP03 was the exception. Score agreement was more restrictive. NP12 matched the published Q scores at

the source precision, while NP05 matched scores at the published precision in two of six configurations despite winner and complete ranking agreement in all six.

NP03 was evaluable from a D2 public computational point. Its winner matched the source, but its complete ranking differed. Because the direction of ERP04 (implementation time) remained unresolved, we compared robustness under the explicit-direction branch and the secondary interpretation treating it as a cost criterion.

The three terminally non-evaluable lineages were NE for different reasons. NP08 had D3/O3 public depth but optimal linear-program solutions yielding different decisions without a source tie-break; NP09 had D3/O3 depth but a literal chain that became non-finite at ln(0); and NP10 had O3 terminal outputs but D0 because the source-referenced supplementary tables and the operative direction mapping required for the terminal chain were unavailable to us.

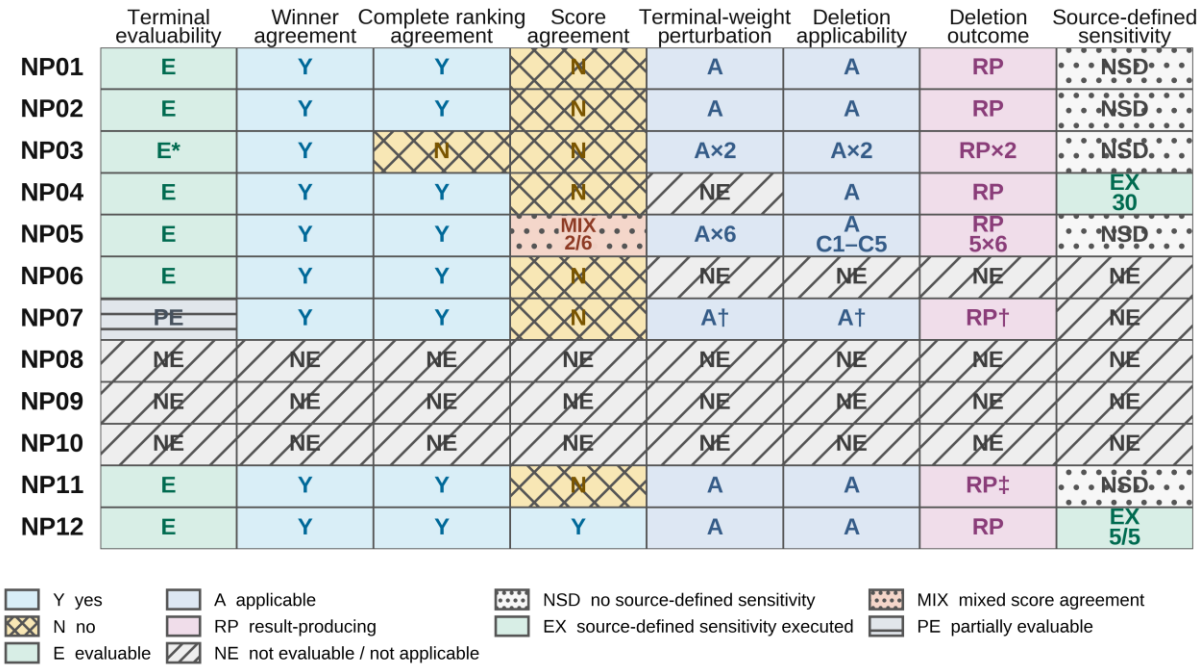
NP07 comprises two linked publications with different reconstruction outcomes. The baseline study [49] provided a D2/O3 computational point from which we reproduced the winner and complete ranking, but not the score vector. The linked CBDO reanalysis [23] had D3/O2 depth. We could not reproduce its stochastic results exactly from the information available to us.

Table 2. Native reconstruction outcomes

NP	Terminal evaluability	Winner agreement	Complete ranking agreement	Score agreement	Primary caveat
NP01	Evaluable	Yes	Yes	No	Complete ranking was reconstructed despite score mismatch; stability differed by diagnostic
NP02	Evaluable	Yes	Yes	No	Near-invariant to terminal-weight perturbation; one single-criterion deletion changes the winner
NP03	Evaluable from public checkpoint	Yes	No	No	Published winner reproduced; complete ranking differs; stability depends on criterion direction
NP04	Evaluable	Yes	Yes	No	Winner retained in 26/30 source-defined scenarios; workbook inconsistencies affected scores but not rankings or winner sets
NP05	Evaluable	Yes (6/6)	Yes (6/6)	Match 2/6; differ 4/6	Six co-primary configurations remain one lineage
NP06	Evaluable	Yes	Yes	No	Source-reported interaction-off comparison; examined inputs did not permit a unique reconstruction
NP07	Partially evaluable	Yes (baseline [49])	Yes (baseline [49])	No (baseline [49]); linked exact reanalysis NE	Linked studies have different computational depths; exact stochastic replication not possible from examined inputs
NP08	Not terminally evaluable	NE	NE	NE	Non-unique LP; no source tie-break or solver-specific terminal ranking
NP09	Not evaluable: literal method non-finite	NE	NE	NE	Literal chain reaches ln(0); no documented epsilon or clipping convention
NP10	Not evaluable with examined inputs	NE	NE	NE	Source reports O3 outputs [27]; examined inputs insufficient for reconstruction
NP11	Evaluable; decision identified	Yes	Yes	No	LFPP weights and scores are nonunique, but the baseline winner and complete ranking are invariant across global optima. Score agreement refers to the documented numerical solution.
NP12	Evaluable	Yes	Yes	Yes	Native recommendation-set rule governs agreement and robustness

Note. Yes and No denote evaluable agreement outcomes. NE denotes a result that is not evaluable under the stated source-method rules and is not a negative agreement outcome. NP05 reports six co-primary configurations within one lineage; NP07 agreement refers to its baseline study [49]. NP11 score agreement

compares the documented computational point with the published values; its weights and scores are not mathematically unique.



Nested branches/configurations remain within lineage rows: NP03 ×2; NP05 ×6.
E* = evaluable from public checkpoint; † = Kılıç et al. (2015) baseline only.
NE differs from No; ‡ = NP11 deletions evaluated post hoc.

Figure 1. Evidence matrix for the 12 ERP-MCDA lineages. Cells show terminal evaluability, reconstruction agreement, diagnostic applicability and outcomes, and source-defined sensitivity. NE denotes a calculation that is not evaluable or not applicable under the stated source-method rules. RP‡ marks the NP11 deletion results (Table 3).

4.2 Terminal-weight perturbation results

The seven applicable lineages showed markedly different terminal-weight perturbation profiles. At δ = 0.40, winner retention was 0.7207 for NP01, 0.9999 for NP02, 0.5295 for the NP07 baseline and 1.0000 for NP11. NP01 remained at 1.0000 through δ = 0.10 before declining, whereas NP07 declined from 0.7028 at δ = 0.05 to 0.5295 at δ = 0.40. NP05 retained the winner in all draws (1.0000) for every co-primary configuration at every tested δ. NP12 had a recommendation-set retention fraction of 0.9803 at δ = 0.40, while NP12 alternative A2 was retained in all draws (1.0000). Figure 2 shows the trajectories across all five perturbation amplitudes.

NP03 showed the strongest within-lineage effect of unresolved direction coding. Under the primary strict branch, retention remained 1.0000 through $\delta = 0.20$ and was 0.9902 at $\delta = 0.40$. Under the secondary plausible-direction branch, it fell from 0.7262 at $\delta = 0.05$ to 0.4115 at $\delta = 0.40$. Because the two trajectories belong to the same lineage and use the same perturbation design, their divergence isolates sensitivity to the direction convention rather than between-study variation.

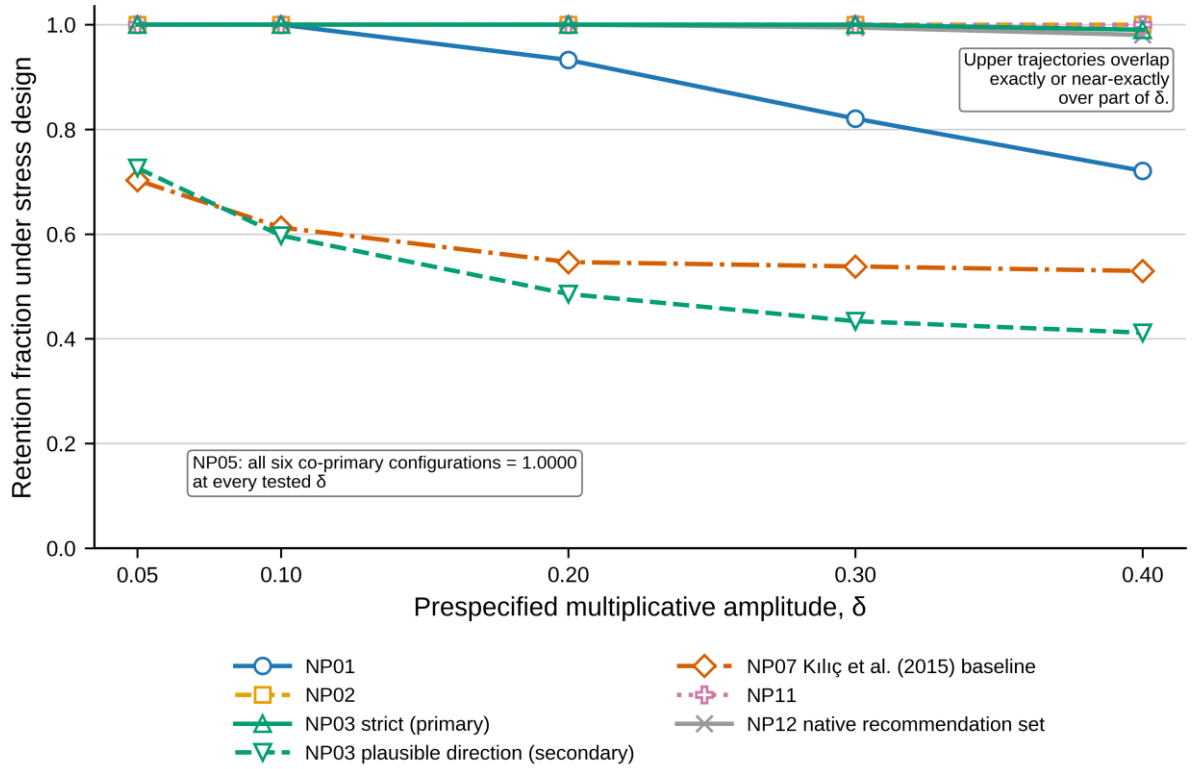


Figure 2. Retention under terminal-weight perturbation. At each δ, 10,000 independent perturbation vectors were generated, with one factor per criterion drawn from Uniform[1−δ, 1+δ]. Each vector was applied to the terminal weights and renormalized. Retention is the fraction of trials preserving the baseline decision. NP03 strict is the explicit-direction branch; the secondary branch treats implementation time as a cost criterion. NP05 retained the winner in all trials for all six configurations at every δ. NP11 uses the documented LFPP criterion-weight and local ERP vectors.

4.3 Single-criterion deletion results and auditability

Single-criterion deletion yielded evaluable results for all eight applicable lineages. In NP11, with the original local ERP weights held fixed, deleting C1, C3 or C5 preserved S4 > S3 > S1 > S2. Deleting C2 gave S4 > S1 > S3 > S2, deleting C4 gave S4 > S3 > S2 > S1, and deleting C6 gave S3 > S1 > S2 > S4. Each ranking is invariant over the corresponding global criterion-weight optimum set. The baseline winner was retained in 5/6 deletions (0.8333) and the complete ranking in 3/6 (0.5000). The same six rankings were obtained using the published rounded local weights.

Among evaluable deletion results, NP01 retained the winner in 11/12 deletions, NP02 in 4/5, NP04 in 11/12 and NP07 in 7/11. NP03 retained the winner in 15/15 deletions under the strict branch and 10/15 under the secondary plausible-direction branch.

NP05 retained the winner in all five eligible deletions in each of six configurations, but both members of the baseline top two remained together in only 1/5 deletions. C6 was excluded because it is the sole cost criterion. The case shows that a stable winner can coexist with substantial movement immediately below it.

NP12 retained its native VIKOR recommendation set in all 18 eligible deletions. Because VIKOR's acceptance rule yields a set-valued decision object in this case, we evaluated stability at the recommendation-set level. For NP06, the examined inputs did not determine a reduced Choquet capacity for this diagnostic. Table 3 integrates these deletion results with the terminal-weight and source-defined diagnostics.

Table 3. Integrated robustness diagnostics

Panel A. Terminal-weight perturbation and deletion analysis				
NP	**Terminal-weight perturbation: key outcome**	**Deletion applicability**	**Deletion result**	**Deletion recomputation**
NP01	Applicable; winner retention 0.7207 at δ = 0.40	Applicable	Evaluable	Native reweighting after deletion
NP02	Applicable; winner retention 0.9999 at δ = 0.40	Applicable	Evaluable	Native reweighting after deletion
NP03	Applicable (2 direction branches); winner retention at δ = 0.40: strict 0.9902; plausible 0.4115	Applicable (2 branches)	Evaluable (2 branches)	TOPSIS recomputation under preserved direction branch
NP04	NE	Applicable	Evaluable	Literal Rough-WISP deletion chain
NP05	Applicable (6 configurations); winner retained in all draws (1.0000) in all configurations at every δ	Applicable (source criteria C1-C5); C6 not deletion-eligible because it is the sole cost criterion	Evaluable (5 eligible deletions/configuration)	Terminal/native recomputation; C1-C5 only
NP06	NE	NE	NE	NE
NP07	Applicable (baseline [49] only); winner retention 0.5295 at δ = 0.40	Applicable (baseline [49] only)	Evaluable (baseline [49] only)	Baseline [49]; method-compatible deletion
NP08	NE	NE	NE	NE
NP09	NE	NE	NE	NE
NP10	NE	NE	NE	NE
NP11	Applicable; winner retention 1.0000 at δ = 0.40	Applicable	Evaluable (6/6 deletions)	Reduced LFPP; original local ERP weights held fixed; rankings identified across global criterion optima
NP12	Applicable; recommendation-set retention 0.9803 at δ = 0.40; A2 retained 1.0000	Applicable	Evaluable	VIKOR q/v = 0.5; native recommendation-set rule

Panel B. Deletion outcomes, source-defined sensitivity and interpretation			
NP	**Deletion denominator and result**	**Source-defined sensitivity**	**Interpretation**
NP01	Winner retained 11/12 deletions (0.9167)	None	Complete ranking was reconstructed despite score mismatch; stability differed by diagnostic
NP02	Winner retained 4/5 deletions (0.8000)	None	Near-invariant to terminal-weight perturbation; one deletion changes the winner
NP03	Winner retained: strict 15/15; plausible-direction branch 10/15	None	Published winner reproduced; complete ranking differs; stability depends on criterion direction
NP04	Winner retained 11/12 deletions (0.9167)	Equation (31) in [28]: 30 scenarios; winner retained in 26	Source-workbook inconsistencies: 0 complete ranking changes, 0 winner-set changes; source scenario S16 had a rounding-sensitive SFT-5/SFT-6 tie with no winner effect
NP05	6 configurations: winner 5/5 each; source criterion C6 not deletion-eligible; baseline top two retained together 1/5	None	Winner stable; ordering below the winner is deletion-sensitive
NP06	NE: reduced Choquet capacity not identified from examined inputs	Publication-reported interaction-off comparison; reconstruction NE	Interaction-off result reported by source; independent reconstruction not possible from examined inputs
NP07	Winner retained 7/11 deletions (0.6364)	NE for exact linked CBDO replication	Lower retention under both diagnostics; exact linked stochastic replication not possible from examined inputs
NP08	NE	NE (method/source)	Non-unique LP with no source tie-break; no solver-specific terminal ranking retained
NP09	NE	NE (method/source)	Literal chain reaches ln(0); no epsilon or clipping convention is documented
NP10	NE	NE (method/source)	Published O3 outputs; reconstruction depth D0 with examined inputs
NP11	Winner retained 5/6 deletions (0.8333); complete ranking retained 3/6 (0.5000); C6 changes S4 to S3	None	Complete retention under terminal-weight perturbation; C6 deletion changes the winner
NP12	Native recommendation set retained 18/18 deletions	Q scores, rankings and acceptability agree at all five q/v settings	Recommendation set retained in all deletions and q/v settings; set retention under weight perturbation was 0.9803 at δ = 0.40

Note. Terminal-weight perturbation uses 10,000 vectors per δ for each branch or configuration. Deletion denominators count the eligible criteria in each lineage. NP11 holds the original local ERP weights fixed. NP03 has two direction branches and NP05 six configurations. For NP04, S1-S30 label the source scenarios and SFT-1–SFT-6 the ERP alternatives. NP05 uses source criterion labels C1-C6. For NP12, A1-A3 are the source alternatives and q/v is the VIKOR compromise parameter.

4.4 Source-defined sensitivity and source-workbook findings

NP04 retained its baseline winner in 26 of the 30 weight-sensitivity scenarios defined by Equation (31) in [28], compared with 11 of 12 single-criterion deletions.

The NP04 workbook contained 20 inconsistent lower or upper endpoint records for scenarios S21-S30. These inconsistencies affected terminal scores without changing any complete ranking or winner set. In S16, the full-precision reconstruction placed SFT-5 above SFT-6, while the source workbook's rounded computational checkpoint yielded a tie; SFT-4 remained the winner. The source [28] uses SFT-1–SFT-6 to identify the six ERP alternatives.

For NP12, the reconstructed Q scores matched Table 6 of the source study [21] at its reported precision for all five q/v settings. The rankings and acceptability conditions agreed with its Table 7, but its printed advantage differences were inconsistent with Table 6. This discrepancy did not change

the recommendation of A2. The NP06 interaction-off comparison is reported by the source, but we could not reconstruct it independently from the inputs examined.

4.5 Integrated cross-diagnostic patterns

All nine evaluable baseline decisions matched the published winner or native recommendation. Eight reproduced complete rankings, while score agreement was absent in seven lineages, mixed across six NP05 configurations and complete for NP12. NP11 shows why nonunique optimal weights do not necessarily make a decision unidentifiable: every reduced ranking is identified, and deleting C6 changes the winner. Reconstructability did not guarantee robustness auditability in every case; NP06 reconstructed at baseline but lacked a defensible reduced Choquet deletion model.

High retention under terminal-weight perturbation likewise did not guarantee invariance to single-criterion deletion. NP02 had a winner-retention fraction of 0.9999 at $\delta = 0.40$ but lost the winner after one of five deletions. NP05 retained the winner under both diagnostics, yet its baseline top two remained together in only 1/5 eligible deletions. NP03 further showed that an unresolved source convention could materially change both diagnostic profiles.

5 Discussion

The benchmark separates decision agreement from numerical reconstruction. All 9 evaluable baselines retained the published winner, but complete ranking agreement was 8/9 and score agreement was less uniform. A matching winner may be sufficient for a narrow package-choice question, yet scientific reuse can depend on near-ties, second-best alternatives, scenario comparisons and terminal values. NP03 preserved the recommendation while changing the complete ranking; NP05 preserved winner agreement and complete ranking agreement in all six configurations while four score vectors remained non-identical. ERP models can also terminate in different decision objects, including a selected system [55], a full ordering or a set-valued recommendation, so agreement should be reported at the level the method actually produces.

This distinction matters because published ERP cases have been reused or reformulated in later MCDA demonstrations [64] [65]. Analysts may reuse an AHP application [35] or another published ERP model with different criteria, alternatives or organizational requirements. Discrepancies below the top alternative can then matter. Reporting intermediate computational points, criterion directions, normalization choices, tie rules and terminal precision supports such reuse even when the recommended system is unchanged.

Reconstructability and robustness auditability are distinct. Reconstruction depends on the information and conventions available to the researcher, not merely on whether the original authors performed a computation [10] [13]. In NP11, every reduced LFPP model has an identified complete ranking when the original local ERP weights are held fixed, although optimal weights and terminal scores are nonunique. Removing C6 changes the winner from S4 to S3. NP06 supported baseline reconstruction, but the examined inputs did not identify a unique model for criterion deletion or the interaction-off comparison.

The two common diagnostics also show why a stability claim must name the perturbation. Terminal-weight perturbation changes terminal weights while holding the criteria set fixed; single-criterion deletion changes model structure and then applies method-compatible recomputation. NP02 was nearly invariant under the former (winner-retention fraction 0.9999 at $\delta = 0.40$) but changed winner in one of five deletions. NP05 preserved the winner under both diagnostics but retained its top two together in only 1/5 deletions. These results are consistent with broader distinctions among sensitivity, robustness, uncertainty and rank reversal [14] [15] [16] [17] [18].

NP03 demonstrates the effect of unresolved criterion direction. At $\delta = 0.40$, winner retention was 0.9902 under the strict branch and 0.4115 when ERP04 (implementation time) was treated as a cost criterion; deletion retention was 15/15 versus 10/15. Because both branches use the same diagnostics, the contrast isolates the consequence of that convention. Criterion direction must therefore be documented as part of the executable model.

Source-defined analyses provide complementary evidence. NP04 retained the baseline winner in 26/30 source-defined sensitivity scenarios [28] and in 11/12 deletions in our single-criterion analysis. Its workbook had 20 inconsistent endpoint records that affected scores without changing

rankings or winner sets. The separate S16 rounding effect also left the winner unchanged. These discrepancies matter for numerical reuse even when the recommended alternative is preserved. NP12 retained its recommendation set in all deletions and at all five q/v settings. Under terminal-weight perturbation, set retention was 0.9803 at δ = 0.40, while A2 remained recommended in every trial.

These findings support practical reporting requirements. Reproducibility guidance calls for inputs, provenance, software versions, parameters, intermediate outputs, random states and code [11] [12] [13]. ERP-MCDA studies should also report criterion types and directions, weights at adequate precision, rules for regenerating weights after criterion removal, tie and solver conventions and the native decision object. Sensitivity reports should identify what changes, what remains fixed and whether retention is calculated from sampled perturbations, eligible deletions or source-defined settings. This information makes both the baseline result and its stability analysis independently auditable.

The contribution is a common evaluability framework for comparing independent decision-problem lineages. Earlier studies examined ERP stability [19], comparisons of decision models [20], robust reanalysis [23] and source-defined scenarios [28] [21]. Published cases have also been reused to develop alternative MCDA formulations [64] [65]. AHP applications [35] and decision-support models [36] offer reporting comparisons, while independent reimplementation has a broader computational foundation [13]. Applying these ideas across cases reveals where reconstruction agreement and stability diverge, while retaining the decision rules and denominators of each method.

The benchmark is purposive rather than representative. Its 12 lineages were selected for computational depth, method diversity, ERP relevance, lineage independence and feasible public-source reconstruction, so the denominators describe this benchmark rather than the ERP-MCDA literature. The selection favors cases that expose enough evidence for cross-case analysis, which limits inference about less documented studies and the frequency of reporting practices in the wider literature. Method heterogeneity also precludes inferring from these cases that any family is intrinsically more auditable or stable [24].

The common diagnostics examine a limited range of uncertainty [15] [16]. Multiplicative terminal-weight perturbation is a designed stress test whose retention fractions do not estimate probabilities of true optimality. Single-criterion deletion changes one structural element at a time under a method-compatible recomputation rule and differs from classical rank reversal [17] [18]. Performance values, alternative sets, fuzzy parameters and elicitation uncertainty were not varied unless covered by a source-defined analysis.

Evaluability here concerns the evidence examined for each case. NP10 reports terminal results, but the information available to us was insufficient for independent reconstruction. The stated methods did not identify a unique decision in NP08 and did not yield finite scores in NP09. Score comparisons also depend on the precision and computational point reported by each source; NP04's S16 tie illustrates how rounding in the source workbook can change a local ordering without changing the winner.

6 Conclusions

This benchmark separates four distinct properties: output availability, public-source reconstructability, robustness auditability and stability under a stated diagnostic. Nine of 12 ERP-MCDA lineages contained an evaluable terminal baseline; all 9 retained the published winner or native recommendation, while 8 reproduced the complete ranking and score agreement was more restrictive.

Terminal-weight perturbation was applicable to seven lineages, and deletion yielded evaluable results for all eight applicable lineages. NP11 retained its baseline winner in 5/6 deletions and its complete ranking in 3/6; removing C6 changed the winner from S4 to S3 despite complete retention under terminal-weight perturbation. NP03 showed that uncertain criterion directions can change both diagnostic profiles. Source-defined analyses in NP04 and NP12 provided further case-specific evidence.

ERP-MCDA studies should publish the inputs and computational steps needed to reproduce the baseline recommendation, together with criterion directions, weight-generation rules and tie and

solver conventions. Each stability analysis should specify its perturbations, fixed assumptions and decision rule. This documentation makes published decisions easier to examine and reuse.

Data and code availability

The benchmark data and analysis code are available at https://github.com/myresearchbvp/ERP-MCDA-Benchmark.

References

[1] J. Verville and A. Halingten, "A six-stage model of the buying process for ERP software," *Industrial Marketing Management*, vol. 32, no. 7, pp. 585–594, 2003. DOI: 10.1016/S0019-8501(03)00007-5.

[2] B. Baki and K. Çakar, "Determining the ERP package-selecting criteria: The case of Turkish manufacturing companies," *Business Process Management Journal*, vol. 11, no. 1, pp. 75–86, 2005. DOI: 10.1108/14637150510578746.

[3] M. Keil and A. Tiwana, "Relative importance of evaluation criteria for enterprise systems: a conjoint study," *Information Systems Journal*, vol. 16, no. 3, pp. 237–262, 2006. DOI: 10.1111/j.1365-2575.2006.00218.x.

[4] R.-E. Precup, C.-A. Bojan-Dragos, K. Gao, and S. Cui, "Problem Setting for Trajectory Planning and Cruise Control of a Connected Autonomous Electric Bus in Intersection Scenarios with Human-Driven Vehicles to Optimize Energy, Comfort and Tracking," *Romanian Journal of Information Science and Technology*, vol. 28, no. 3, pp. 299–312, 2025. DOI: 10.59277/ROMJIST.2025.3.05.

[5] C.-C. Wei, C.-F. Chien, and M.-J. J. Wang, "An AHP-based approach to ERP system selection," *International Journal of Production Economics*, vol. 96, no. 1, pp. 47–62, 2005. DOI: 10.1016/j.ijpe.2004.03.004.

[6] Z. Ayağ and R. G. Özdemir, "An intelligent approach to ERP software selection through fuzzy ANP," *International Journal of Production Research*, vol. 45, no. 10, pp. 2169–2194, 2007. DOI: 10.1080/00207540600724849.

[7] S. Perçin, "Using the ANP approach in selecting and benchmarking ERP systems," *Benchmarking: An International Journal*, vol. 15, no. 5, pp. 630–649, 2008. DOI: 10.1108/14635770810903196.

[8] D. M. Fisher, M. Y. Kiang, S. A. Fisher, and R. T. Chi, "Evaluating Mid-Level ERP Software," *Journal of Computer Information Systems*, vol. 45, no. 1, pp. 38–46, 2004. DOI: 10.1080/08874417.2004.11645815.

[9] S.-K. Liang and C.-T. Lien, "Selecting the Optimal ERP Software by Combining the ISO 9126 Standard and Fuzzy AHP Approach," *Contemporary Management Research*, vol. 3, no. 1, pp. 23–44, 2006. DOI: 10.7903/cmr.10.

[10] R. D. Peng, "Reproducible Research in Computational Science," *Science*, vol. 334, no. 6060, pp. 1226–1227, 2011. DOI: 10.1126/science.1213847.

[11] G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, "Ten Simple Rules for Reproducible Computational Research," *PLOS Computational Biology*, vol. 9, no. 10, Art. no. e1003285, 2013. DOI: 10.1371/journal.pcbi.1003285.

[12] V. Stodden, M. McNutt, D. H. Bailey, E. Deelman, Y. Gil, B. Hanson, M. A. Heroux, J. P. A. Ioannidis, and M. Taufer, "Enhancing reproducibility for computational methods," *Science*, vol. 354, no. 6317, pp. 1240–1241, 2016. DOI: 10.1126/science.aah6168.

[13] F. C. Y. Benureau and N. P. Rougier, "Re-run, Repeat, Reproduce, Reuse, Replicate: Transforming Code into Scientific Contributions," *Frontiers in Neuroinformatics*, vol. 11, Art. no. 69, 2018. DOI: 10.3389/fninf.2017.00069.

[14] E. Triantaphyllou and A. Sánchez, "A Sensitivity Analysis Approach for Some Deterministic Multi-Criteria Decision-Making Methods," *Decision Sciences*, vol. 28, no. 1, pp. 151–194, 1997. DOI: 10.1111/j.1540-5915.1997.tb01306.x.

[15] B. Roy, "Robustness in operational research and decision aiding: A multi-faceted issue," *European Journal of Operational Research*, vol. 200, no. 3, pp. 629–638, 2010. DOI: 10.1016/j.ejor.2008.12.036.

[16] I. N. Durbach and T. J. Stewart, "Modeling uncertainty in multi-criteria decision analysis," *European Journal of Operational Research*, vol. 223, no. 1, pp. 1–14, 2012. DOI: 10.1016/j.ejor.2012.04.038.

[17] R. F. de Farias Aires and L. Ferreira, "The rank reversal problem in multi-criteria decision making: A literature review," *Pesquisa Operacional*, vol. 38, no. 2, pp. 331–362, 2018. DOI: 10.1590/0101-7438.2018.038.02.0331.

[18] B. Li and A. Abbas, "A Trend Analysis of Rank Reversal in Widely Used Decision-Making Methods," *Journal of Multi-Criteria Decision Analysis*, vol. 33, no. 1, Art. no. e70027, 2026. DOI: 10.1002/mcda.70027.

[19] B. S. Ahn and S. H. Choi, "ERP system selection using a simulation-based AHP approach: a case of Korean homeshopping company," *Journal of the Operational Research Society*, vol. 59, no. 3, pp. 322–330, 2008. DOI: 10.1057/palgrave.jors.2602365.

[20] I. Ghosh and S. Biswas, "A Comparative Analysis of Multi-criteria Decision Models for ERP Package Selection for Improving Supply Chain Performance," *Asia-Pacific Journal of Management Research and Innovation*, vol. 12, no. 3-4, pp. 250–270, 2016. DOI: 10.1177/2319510X16688988.

[21] R. Basar and H. M. Arslan, "Selection of the Enterprise Resource Planning (ERP) Software with the Most Appropriate Solution (VIKOR)," *Suleyman Demirel University Journal of Faculty of Economics and Administrative Sciences*, vol. 22, no. 4, pp. 1065–1080, 2017 (in Turkish). [Online]. Available: https://dergipark.org.tr/en/pub/sduiibfd/issue/52995/704367

[22] X. Luo and X. Wang, "Extended VIKOR Method for Intuitionistic Fuzzy Multiattribute Decision-Making Based on a New Distance Measure," *Mathematical Problems in Engineering*, vol. 2017, Art. no. 4072486, 16 pp., 2017. DOI: 10.1155/2017/4072486.

[23] G. T. Temur and B. Bolat, “A robust MCDM approach for ERP system selection under uncertain environment based on worst case scenario,” *Journal of Enterprise Information Management*, vol. 31, no. 3, pp. 405–425, 2018. DOI: 10.1108/JEIM-12-2017-0175.
[24] S. Greco, R. Słowiński, and J. Wallenius, “Fifty years of multiple criteria decision analysis: From classical methods to robust ordinal regression,” *European Journal of Operational Research*, vol. 323, no. 2, pp. 351–377, 2025. DOI: 10.1016/j.ejor.2024.07.038.
[25] Y.-R. Zeng, L. Wang, and X.-H. Xu, “An integrated model to select an ERP system for Chinese small- and medium-sized enterprise under uncertainty,” *Technological and Economic Development of Economy*, vol. 23, no. 1, pp. 38–58, 2017. DOI: 10.3846/20294913.2015.1072748.
[26] B. Dağci Yüksel and F. Ersöz, “Evaluation of ERP software selection criteria with fuzzy AHP approach: an application in the metal production enterprises in the aviation industry,” *International Journal of System Assurance Engineering and Management*, vol. 15, no. 6, pp. 2656–2667, 2024. DOI: 10.1007/s13198-024-02287-x.
[27] P. P. Deb, D. Bhattacharya, I. Chatterjee, A. Saha, A. R. Mishra, and S. H. Ahammad, “A Decision-Making Model With Intuitionistic Fuzzy Information for Selection of Enterprise Resource Planning Systems,” *IEEE Transactions on Engineering Management*, vol. 71, pp. 4820–4834, 2024. DOI: 10.1109/TEM.2022.3215608.
[28] B. Cao, Y. Jin, A. Ulutaş, A. Topal, Z. Stević, D. Karabasevic, and C. Sava, “A new integrated rough multi-criteria decision-making model for enterprise resource planning software selection,” *PeerJ Computer Science*, vol. 10, Art. no. e2096, 2024. DOI: 10.7717/peerj-cs.2096.
[29] D. M. Utama, M. F. Ibrahim, and A. N. A. A. K. Jabari, “A Fuzzy Multi-Criteria Approach for Selecting Open-Source ERP Systems in SMEs Using Fuzzy AHP and TOPSIS,” *Jurnal Optimasi Sistem Industri*, vol. 23, no. 2, pp. 167–187, 2025. DOI: 10.25077/josi.v23.n2.p167-187.2024.
[30] H. Akpınar, “An Application on Enterprise Resource Planning (ERP) Selection with Multi-Criteria Decision-Making Methods,” *Celal Bayar University Journal of Science*, vol. 21, no. 2, pp. 28–34, 2025. DOI: 10.18466/cbayarfbe.1513901.
[31] I. Akpinar, B. Kececi, K. Atalay, and M. Boran, “ERP software selection problem with a novel Hybrid hesitant fuzzy cross-entropy and probabilistic hesitant fuzzy MCDM approach,” *Journal of Industrial and Management Optimization*, vol. 21, no. 9, pp. 5662–5686, 2025. DOI: 10.3934/jimo.2025109.
[32] B. S. Mahapatra, D. Ghosh, D. Pamucar, and G. S. Mahapatra, “Dynamic group decision-making for enterprise resource planning selection using two-tuples Pythagorean fuzzy MOORA approach,” *Expert Systems with Applications*, vol. 263, Art. no. 125675, 2025. DOI: 10.1016/j.eswa.2024.125675.
[33] D. S. S. Esmail Pour and A. Faghfouri, “Integrating Knowledge Management Into ERP Selection: A Fuzzy Multi-Criteria Decision-Making Model Based on Gray Theory in the Automotive Industry,” *Knowledge and Process Management*, Early View, pp. 1–18, 2026. DOI: 10.1002/kpm.70105.
[34] U. Cebeci, “Fuzzy AHP-based decision support system for selecting ERP systems in textile industry by using balanced scorecard,” *Expert Systems with Applications*, vol. 36, no. 5, pp. 8900–8909, 2009. DOI: 10.1016/j.eswa.2008.11.046.
[35] R. M. Czekster, T. Webber, A. H. Jandrey, and C. A. M. Marcon, “Selection of enterprise resource planning software using analytic hierarchy process,” *Enterprise Information Systems*, vol. 13, no. 6, pp. 895–915, 2019. DOI: 10.1080/17517575.2019.1606285.
[36] M. M. Cruz-Cunha, J. P. Silva, J. J. Gonçalves, J. A. Fernandes, and P. S. Ávila, “ERP Selection using an AHP-based Decision Support System,” *Information Resources Management Journal*, vol. 29, no. 4, pp. 65–81, 2016. DOI: 10.4018/IRMJ.2016100105.
[37] R. Lacurezeanu, A. Chis, and V. P. Bresfelean, “Integrated Management Solution for a Sustainable SME - Selection Proposal Using AHP,” *Sustainability*, vol. 13, no. 19, Art. no. 10616, 2021. DOI: 10.3390/su131910616.
[38] R. Lacurezeanu and V. P. Bresfelean, “Making a Multi-criteria Analysis Model for Choosing an ERP for SMEs in a KM World,” in *Intelligent Methods in Computing, Communications and Control (ICCCC 2020)*, Advances in Intelligent Systems and Computing, vol. 1243, pp. 257–273, 2021. DOI: 10.1007/978-3-030-53651-0_22.
[39] X. Liao, Y. Li, and B. Lu, “A model for selecting an ERP system based on linguistic information processing,” *Information Systems*, vol. 32, no. 7, pp. 1005–1017, 2007. DOI: 10.1016/j.is.2006.10.005.
[40] A. Görener, “ERP Software Selection Using a Combined ANP and VIKOR Approach,” *Journal of Aeronautics and Space Technologies (Havacılık ve Uzay Teknolojileri Dergisi)*, vol. 5, no. 1, pp. 97–110, 2011. [Online]. Available: https://jast.msu.edu.tr/index.php/JAST/article/view/274
[41] M. R. Zahedi, S. Yousefi, and M. Cheshmberah, “A Fuzzy Quality Function Deployment Approach to Enterprise Resource Planning Software Selection,” *Asian Journal of Scientific Research*, vol. 4, no. 2, pp. 114–128, 2011. DOI: 10.3923/ajsr.2011.114.128.
[42] T. Gürbüz, S. E. Alptekin, and G. Işıklar Alptekin, “A hybrid MCDM methodology for ERP selection problem with interacting criteria,” *Decision Support Systems*, vol. 54, no. 1, pp. 206–214, 2012. DOI: 10.1016/j.dss.2012.05.006.
[43] A. Jafarnejad, M. Ansari, H. R. Youshanlouei, and M. M. Mood, “A Hybrid MCDM Approach for Solving the ERP System Selection Problem with Application to Steel Industry,” *International Journal of Enterprise Information Systems*, vol. 8, no. 3, pp. 54–73, 2012. DOI: 10.4018/jeis.2012070104.
[44] Y. Kazancoglu and S. Burmaoglu, “ERP software selection with MCDM: application of TODIM method,” *International Journal of Business Information Systems*, vol. 13, no. 4, pp. 435–452, 2013. DOI: 10.1504/IJBIS.2013.055300.
[45] C. López and A. Ishizaka, “GAHPSort: A new group multi-criteria decision method for sorting a large number of the cloud-based ERP solutions,” *Computers in Industry*, vol. 92–93, pp. 12–25, 2017. DOI: 10.1016/j.compind.2017.06.007.

[46] M. Sağnak and Y. Kazançoğlu, “Integrated Fuzzy Analytic Network Process And 0-1 Goal Programming Technique For Enterprise Resource Planning (Erp) Software Selection,” *Ege Academic Review*, vol. 19, no. 1, pp. 75–88, 2019. DOI: 10.21121/eab.2019148776.
[47] A. H. I. Lee, S.-C. Chen, and H.-Y. Kang, “A decision-making framework for evaluating enterprise resource planning systems in a high-tech industry,” *Quality Technology & Quantitative Management*, vol. 17, no. 3, pp. 319–336, 2020. DOI: 10.1080/16843703.2019.1626073.
[48] H. S. Kilic, S. Zaim, and D. Delen, “Development of a hybrid methodology for ERP system selection: The case of Turkish Airlines,” *Decision Support Systems*, vol. 66, pp. 82–92, 2014. DOI: 10.1016/j.dss.2014.06.011.
[49] H. S. Kilic, S. Zaim, and D. Delen, “Selecting ‘The Best’ ERP system for SMEs using a combination of ANP and PROMETHEE methods,” *Expert Systems with Applications*, vol. 42, no. 5, pp. 2343–2352, 2015. DOI: 10.1016/j.eswa.2014.10.034.
[50] S. Çakır, “Selecting appropriate ERP software using integrated fuzzy linguistic preference relations - fuzzy TOPSIS method,” *International Journal of Computational Intelligence Systems*, vol. 9, no. 3, pp. 433–449, 2016. DOI: 10.1080/18756891.2016.1175810.
[51] B. Efe, “An integrated fuzzy multi criteria group decision making approach for ERP system selection,” *Applied Soft Computing*, vol. 38, pp. 106–117, 2016. DOI: 10.1016/j.asoc.2015.09.037.
[52] N. Motaki and O. Kamach, “ERP selection: A step-by-step application of AHP Method,” *International Journal of Computer Applications*, vol. 176, no. 7, pp. 15–21, 2017. DOI: 10.5120/ijca2017915636.
[53] N. V. Thanh, “Designing a MCDM Model for Selection of an Optimal ERP Software in Organization,” *Systems*, vol. 10, no. 4, Art. no. 95, 2022. DOI: 10.3390/systems10040095.
[54] S. Panjavongroj and B. Phruksaphanrat, “Selection of ERP system and the best practice by hybrid method: A case study of Thai automotive supply chain network,” *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 6, pp. 7617–7631, 2022. DOI: 10.3233/JIFS-221476.
[55] E. E. Karsak and C. O. Özogul, “An integrated decision making approach for ERP system selection,” *Expert Systems with Applications*, vol. 36, no. 1, pp. 660–667, 2009. DOI: 10.1016/j.eswa.2007.09.016.
[56] H. R. Yazgan, S. Boran, and K. Goztepe, “An ERP software selection process with using artificial neural network based on analytic network process approach,” *Expert Systems with Applications*, vol. 36, no. 5, pp. 9214–9222, 2009. DOI: 10.1016/j.eswa.2008.12.022.
[57] M. Haddara, “ERP Selection: The SMART Way,” *Procedia Technology (CENTERIS 2014)*, vol. 16, pp. 394–403, 2014. DOI: 10.1016/j.protcy.2014.10.105.
[58] H. Garg, J. Vimala, S. Rajareega, D. Preethi, and L. Perez-Dominguez, “Complex intuitionistic fuzzy soft SWARA - COPRAS approach: An application of ERP software selection,” *AIMS Mathematics*, vol. 7, no. 4, pp. 5895–5909, 2022. DOI: 10.3934/math.2022327.
[59] I. Škrjanc, S. Blažič, and O. E. Agamennoni, “Interval Fuzzy Modeling Applied to Wiener Models With Uncertainties,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 35, no. 5, pp. 1092–1095, 2005. DOI: 10.1109/TSMCB.2005.850166.
[60] T. Pereira and F. A. Ferreira, “A multicriteria decision making model for assessment and selection of an ERP in a logistics context,” in *AIP Conference Proceedings*, vol. 1863, no. 1, Art. no. 050004, 2017. DOI: 10.1063/1.4992201.
[61] T. Pereira and F. A. Ferreira, “An ERP selection using a hybrid methodology,” *Journal of Computational Methods in Sciences and Engineering*, vol. 20, no. 3, pp. 689–695, 2020. DOI: 10.3233/JCM-194011.
[62] J. Vaščák, I. Zolotová, and E. Kajáti, “Navigation Fuzzy Cognitive Maps Adjusted by PSO,” in *2019 23rd International Conference on System Theory, Control and Computing (ICSTCC)*, Sinaia, Romania, 2019, pp. 107–112. DOI: 10.1109/ICSTCC.2019.8886149.
[63] T. Manninen, J. Aćimović, R. Havela, H. Teppola, and M.-L. Linne, “Challenges in Reproducibility, Replicability, and Comparability of Computational Models and Tools for Neuronal and Glial Networks, Cells, and Subcellular Structures,” *Frontiers in Neuroinformatics*, vol. 12, Art. no. 20, 2018. DOI: 10.3389/fninf.2018.00020.
[64] P. Karande and S. Chakraborty, “A Fuzzy-MOORA approach for ERP system selection,” *Decision Science Letters*, vol. 1, no. 1, pp. 11–22, 2012. DOI: 10.5267/j.dsl.2012.07.001.
[65] M. A. Nikjoo, M. Mosadegh Khah, and A. Moghimi, “Fuzzy TOPSIS and GP Application for Evaluation And Selection of a Suitable ERP,” *Australian Journal of Basic and Applied Sciences*, vol. 5, no. 11, pp. 1358–1367, 2011. [Online]. Available: https://ajbasweb.com/old/ajbas/2011/November-2011/1358-1367.pdf