Surrogate Weight Research Directions in MCDA

Mats Danielson[1,2]

[1] Dept. of Computer and Systems Sciences, Stockholm University
PO Box 7003, SE-164 07 Kista, Sweden
mats.danielson@su.se
[2] International Institute for Applied Systems Analysis, IIASA
Schlossplatz 1, AT-2361 Laxenburg, Austria

Abstract: Assigning meaningful weights to evaluation criteria remains one of the central challenges in Multi-Criteria Decision Analysis (MCDA). While precise preference elicitation is foundational to many MCDA methods, decision-makers often struggle to articulate exact trade-offs, particularly when facing uncertainty, cognitive burden, or a large number of criteria. This article surveys current directions in surrogate weighting. It presents a comprehensive examination of surrogate weighting techniques designed to generate plausible weight vectors from limited or indirect information. Drawing on developments spanning five decades, it categorises surrogate methods into four principal families: ordinal rank-based, cardinal extensions, dominance- and regression-based, and objective weighting approaches. Each category is analysed in terms of theoretical motivation, algorithmic formulation, robustness, and empirical performance. The study reviews classical ordinal methods such as Rank Sum and Rank Order Centroid in relation to contemporary enhancements like the SR weighting schemes. Dominance-based models and stochastic techniques such as SMAA are also discussed for their ability to support decisions without committing to a single weight vector. Objective weighting approaches, including entropy and CRITIC, are positioned as valuable complements when preference data are unavailable. Performance evaluation criteria, such as hit ratio, rank correlation, and stability radius, are used to compare methods systematically across simulated decision problems. While most articles take a narrow view, looking at the surrogate problem from a specific methodological standpoint, this article instead makes an overview of many current trends and research directions that aim at simplifying weight elicitation by automating the weight generation through representations and algorithms. The findings suggest that no single research direction dominates across all contexts; rather, the suitability depends on the available input structure and the desired goals with the analyses.

Keywords: Multi-Criteria Decision Analysis, Surrogate Weights, Preference Modelling, Weight Generation Techniques, Ordinal and Cardinal Weighting, Entropy Weighting

1. Introduction

Multi-criteria decision analysis (MCDA) constitutes an important part of decision theory as well as operational research, concerned with formulating, structuring, and solving problems that involve evaluating alternatives based on multiple, often conflicting, criteria or perspectives (Greco et al., 2016). At its core, MCDA provides a formal framework for integrating disparate pieces of information into coherent judgements, allowing decision-makers to move from conflicting data and values to defensible conclusions (Keeney and Raiffa, 1993). Its emergence has been motivated by the recognition that many real-world decisions resist reduction to a single objective function. A basic idea of MCDA is that decision alternatives are seldom evaluated on a single perspective. Real-world decision contexts frequently involve a mixture of economic, technical, environmental, and social criteria, each of which may be measured in different units and valued differently by stakeholders. MCDA methods provide a mathematical and conceptual toolkit for integrating these diverse dimensions into an overall assessment. Depending on the context, this integration may take the form of a complete ranking of alternatives, an identification of robustly non-dominated solutions, or the selection of a preferred compromise. The power of MCDA thus arises from its ability to accommodate conflicting objectives while maintaining logical consistency and traceability. Unlike single-criterion optimisation, which presumes a clearly defined goal function and scale of value, MCDA allows value pluralism. This feature positions MCDA as both a computational framework and an epistemological tool. It provides mechanisms not only for calculating rankings or scores but also for clarifying the value structures and preference assumptions underlying those results (Ishizaka and Nemery, 2013).

MCDA methods can be differentiated along several dimensions, reflecting varying assumptions about trade-offs, information availability, and decision-maker engagement. Some models are compensatory, allowing poor performance on one criterion to be offset by superior performance on another, as in weighted-sum models and multi-attribute utility theory (MAUT) (Keeney and Raiffa, 1993). Others are non-compensatory, using thresholds, outranking, or veto principles to model strict acceptability conditions, as seen in the ÉLECTRE and PROMÉTHÉE families (Roy, 1991; Brans and Mareschal, 2005). Further distinctions arise between constructive methods, which build a preference model interactively with the decision-maker, and descriptive or indirect methods, which infer preferences from holistic judgements or observed choices. The diversity of MCDA methods reflects the diversity of decision contexts. Some approaches are rooted in value theory and assume fully compensatory behaviour, where trade-offs among criteria are modelled explicitly using additive value functions or utility constructs. Others assume non-compensatory logic, introducing veto thresholds or outranking relations to represent decision-maker caution or intolerance for poor performance on certain dimensions (Bouyssou et al., 2006). These differing assumptions have led to the development of method families such as MAUT/SAW, AHP and outranking (ÉLECTRE, PROMÉTHÉE). Each offers distinct epistemological bases and computational mechanisms, and their selection often hinges as much on the decision-maker's comfort with a particular mode of expression as on formal properties of the models themselves. While most articles take a narrow view, looking at the surrogate problem from a specific methodological standpoint, this article instead makes an overview of many current trends and research directions that aim at simplifying weight elicitation by automating the weight generation through representations and algorithms.

2. Background

In recent decades, MCDA has broadened its methodological base to include formal treatments of preference uncertainty and incomplete information. It is now common to distinguish between interactive models, in which the decision-maker refines preferences during the evaluation process, and non-interactive or a priori models, in which all preference information is specified in advance (Belton and Stewart, 2002). The former group includes techniques that give feedback iteratively to converge on a preferred solution, while the latter relies on constructing a model based on the information available at the outset. In both cases, increasing attention has been given to the robustness of recommendations, particularly in light of incomplete preference input or the presence of imprecise data. Robustness in MCDA has been conceptualised in various ways. It may refer to the stability of rankings under small perturbations in weight parameters, the consistency of conclusions across different normalisation or scaling procedures, or the extent to which alternative decisions remain optimal under a range of admissible inputs (Roy and Slowinski, 2013). A robust MCDA model is one that offers recommendations not just under idealised assumptions, but under real-world uncertainty and imperfection. As such, robustness analysis has become a standard component of serious MCDA applications, whether through sensitivity analysis, scenario analysis, or the identification of decision regions where multiple alternatives perform equivalently well.

In addition to modelling preference structures, MCDA also provides tools for supporting deliberative processes. It plays an increasingly important role in participatory decision-making, where multiple stakeholders must converge on an acceptable solution. In such contexts, MCDA serves not only as a computational apparatus but also as a communicative device, making explicit the trade-offs and value conflicts involved. It structures discussions around transparent criteria and provides a vocabulary for articulating compromise. Techniques such as multi-actor aggregation, value-focused thinking, and consensus models are now embedded in the MCDA literature, reinforcing its role in collaborative governance and multi-stakeholder negotiation (de Marchi et al., 2000).

3. Surrogate Weights

Among the central components of any MCDA model is the elicitation and representation of preferences. This contains not only the evaluations of alternatives under each criterion but also the relative importance assigned to the criteria themselves. These importance weights function as the value-generating structure for the aggregation process. Specifying them is known to be cognitively demanding, particularly in high-dimensional problems or under uncertainty. Individuals often lack the clarity or confidence to express sharp trade-offs, especially when alternatives are complex or criteria are unfamiliar (French and Xu, 2005). As a result, a significant portion of MCDA research is devoted to preference modelling strategies that either reduce the cognitive load or reconstruct preferences from partial or indirect inputs.

To address the cognitive burden and incomplete preference information issues, researchers have since the 1970s developed automatic weight generation techniques, often termed surrogate weights, which derive a weight vector from limited or ordinal information about preferences. Over almost five decades, but especially after 2010, a rich variety of theoretical and algorithmic approaches to surrogate weighting have emerged. These methods range from simple rank-based formulas to more complex dominance-driven models and objective data-driven schemes. This research overview spans the evolution of these approaches, highlighting major directions such as rank-based ordinal and cardinal methods, geometric techniques, dominance-based models and entropy-based schemes.

3.1 Ordinal Surrogates

Early recognition of the need for surrogate weighting can be seen in the work of Stillwell et al. (1981), who provided one of the first comparisons of weight approximation techniques in multi-attribute utility problems. They introduced the idea of simple rank-order weighting heuristics that require only an ordering of criteria by importance, not exact numerical trade-offs. The premise was that decision-makers find it easier to rank criteria than to assign exact weights, so one could convert the rank order into approximate weights by a formula. Stillwell et al. proposed classic methods like the Rank Sum (RS) and Rank Reciprocal (RR) schemes. In Rank Sum weighting, weights are assigned in proportion to rank positions. If there are N criteria, the most important receives the number N, the second N–1, etc., down to 1, and then the sum is normalised to 1 by division. In Rank Reciprocal, weights are instead proportional to the reciprocal of each criterion's rank (i.e. 1/N) and then normalised, hence its name. These simple formulas were motivated by the idea of maximising discrimination among alternatives using only ordinal importance information. Stillwell et al. also considered a Rank Exponent (RE) method as a generalisation, where weights decline according to a power function of the rank position. The Rank Exponent method essentially introduces a parameter p governing how steeply weights decrease: $p = 0$ yields equal weights, $p = 1$ reproduces Rank Sum, and p values in between moderate the weight function. In the original formulation, Stillwell's rank exponent scheme required knowing the weight of the top criterion to calibrate the exponent, which fell within $0 < p < 1$ (Barron and Barrett, 1996). Later, (Danielson and Ekenberg, 2020) generalised Rank Exponent to RX, allowing $p > 1$ as well, and showed that RX performed best with p around 1.5–1.6 for most decision models.

By the 1990s, these rank-based surrogate weight methods became more refined and were systematically evaluated. Barron and Barrett used this work in the MCDA method SMARTER (Simple Multi-Attribute Rating Technique Exploiting Ranks). They used the now well-known Rank Order Centroid (ROC) method, attributed to Barron (1992). The ROC method sets each weight to the average of all possible weights it could take given its rank, effectively the centroid of the feasible weight simplex under the rank ordering constraints. If the criteria are ranked from 1 (most important) to N (least important), the ROC weight for the criterion ranked in position i is $w_i = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{j}$. This formula yields a sharply decreasing weight profile, with higher-ranked criteria getting proportionally more weight, sometimes overwhelmingly so. Intuitively, ROC assumes that a priori (in the absence of detailed information) all weight assignments consistent with the rank order are equally likely, and it uses the average (centroid) of that set. Early simulation studies thought that ROC weights often led to better decision outcomes (i.e. higher chance of selecting the true best alternative) than other rank-based schemes known at the time. For example, Barron and Barrett (1996) reported that ROC outperformed Equal Weights (EW), RS and RR in terms of decision quality across many random test problems. Surprisingly, before their measurements, many held the view that EW was a plausible method with relatively good performance, such that the additional effort required to elicit more precise weight information was often seen as unnecessary. Subsequent analyses indicated that RR and definitely even more so EW are generally less effective surrogates and are easily overshadowed by improved methods like ROC and others (Chergui and Jiménez-Martín, 2024).

By the beginning of the 2010s, it became common to view RS, RR and ROC as baselines to be improved upon. These older methods, while easy to apply, have limited expressiveness (they assume a fixed pattern of weight decline) and may perform poorly if the true criterion importance differences are irregular. The early rank-based methods also set a standard for understanding trade-offs between accuracy and discrimination. RS and ROC heavily discriminate between ranks (especially ROC, giving a big spread), which can misrepresent a decision-maker who actually saw criteria as nearly equal; RR is more "soft" in that it keeps weights tighter. A shortcoming, however, is that if the true preferences are very skewed (for instance, the top criterion is vastly more important than all others), RR will underestimate those differences, potentially leading to suboptimal choices in those cases.

Centroid-based approaches form a subset of rank-based methods, represented by ROC described above. The term centroid emphasises its geometric interpretation: given a partial order on weights (e.g. criteria $1 \geq 2 \geq 3 \ldots \geq N$ in importance), one can consider the convex polytope of all weight vectors satisfying those order constraints (and non-negativity, summing to 1). The centroid of this feasible region, i.e. the average of all extreme points or the uniform centroid of the polytope, is a seemingly natural candidate for a neutral or unbiased weight vector respecting the rank order. The ROC formula is exactly that when only the rank ordering (no further strength information) is known. One attractive property of the centroid weight is that it is invariant to any further ignorance. It does not arbitrarily favour any corner of the weight space and thus can be seen as a neutral compromise. This relates to the principle of insufficient reason or maximum entropy: if nothing beyond an order is known, one might assume weights are as "spread out" as possible (maximum entropy) under that order, yielding the centroid (Ezell et al., 2021). In practice, centroid weights (ROC) tend to have a bias toward the highest ranked criterion. Despite this, extensive empirical tests have affirmed ROC's performance across many decision distributions, making it "the most famous method in the state-of-the-art" of rank-based weighting in earlier times (Ahn, 2011). This was before the upsurge of surrogate weight research in the 2010s. It became a benchmark for new methods, and some recent studies have proposed tweaks to ROC. As an example, the Improved ROC (IROC) method adjusts the averaging formula to further enhance accuracy in certain decision problems (Hatefi, 2023). Overall, centroid approaches have the advantages of convexity (providing interior and balanced weight vectors) and often exhibit robustness to noisy inputs, at the cost of good performance when the decision-maker's weight distribution is more even.

3.2 Cardinal Surrogates

A limitation of ordinal rank-order methods is their lack of expressiveness. They treat all successive rank information equivalently, which may not reflect the decision-maker's view. For instance, a decision-maker might feel the first criterion is much more important than the second, while the second and 3rd are almost tied; a simple rank order 1 > 2 > 3 does not convey this nuance. To bridge this gap, researchers introduced methods that accept additional preference strength information beyond a simple ordering. One classic approach is the Simos method, proposed by Simos in 1990. In Simos' procedure, the decision-maker sorts criteria on a series of cards from least to most important, and is allowed to insert blank cards to signify larger "gaps" between importance levels. The analyst then computes weights by assigning equal increments within groups of equal importance and larger jumps between groups, effectively translating the ordinal grouping into cardinal weights. A refined version known as Revised Simos (Figueira and Roy, 2002) adjusted how blank cards are counted to obtain more consistent results. The Simos family of methods was one of the first to incorporate cardinal information, which can alternatively be seen as a mild form of ratio information while remaining fairly simple for elicitation. Algorithmically, Simos' method yields weights by linear interpolation: all criteria in the lowest importance category get the smallest weight, those in the next category get a higher weight determined by the number of blank cards (importance levels) in between, and so on. This method, being cardinal, lies in between ordinal and fully interactive weighting, and has been popular in conjunction with the ÉLECTRE method. However, despite its intuitive appeal, analyses have found that Simos' approach does not perform any better than other good rank-based cardinal methods, for example cardinal extensions to ordinal methods such as RS, SR and ROC, in terms of precision and robustness (Danielson and Ekenberg (2017). One major disadvantage of Simos' method is that it requires an exogenously determined parameter, thus not making it a fully automatic surrogate weight method.

Modern developments have proposed alternative ways to encode preference strength on cardinal scales. For example, one can ask the decision-maker to rate the differences between successive ranks on a verbal or numeric semantic scale (e.g. "criterion 1 is of *much* greater importance than criterion 2, which is only *slightly* more important than criterion 3", etc.). This information can then be incorporated into weight-generation formulas. Cardinal RS (CRS) and Cardinal RR (CRR) are extensions of RS and RR that allow the decision-maker to specify not just an ordering but also an *importance level* for each rank position. One way to implement this is to imagine an "importance scale" with Q discrete levels (for instance, 10 levels from least to most important). Each criterion's ordinal rank is mapped to a specific position on this scale, which may not be equally spaced. The decision-maker can effectively compress or stretch gaps between ranks by assigning ranks to scale levels. Danielson and Ekenberg (2015) introduced a method called Cardinal Rank Order Centroid (CRC) which generalises ROC using the importance scale concept. In their formulation, one first computes the "ordinal ROC" weights as if the criteria were simply ranked, then treats those as baseline weights for each importance level on the scale, and finally recomputes a centroid weight given the criteria's assigned levels on the scale. The result is a Cardinal ROC (CRC) weight vector that reduces to ordinary ROC if all rank gaps are equal but deviates if the decision-maker indicates smaller (none) or larger gaps for some ranks. Likewise, they define a Cardinal SR (CSR) method that extends the ordinal SR concept to account for uneven rank intervals (Danielson and Ekenberg, 2015).

In an empirical investigation spanning many weighting scenarios, Danielson and Ekenberg (2014) found that this CSR method achieved very good precision and stability. It was more accurate in matching true underlying weights than earlier methods and showed good robustness under reasonable assumptions about preference strength variability. In fact, the ordinal SR approach from (Danielson and Ekenberg, 2014) and its cardinal counterpart CSR outperform not only RS/RR/ROC but also other, more recent proposals. These findings align with a broader insight: allowing *some* quantitative input (like strength ranks or a semantic differential) can significantly improve weight estimation if used judiciously. Not all such extensions are equal, however. The weighting method must integrate the extra information in a way that remains robust. For example, Chergui and Jiménez-Martín (2024) conclude in a comprehensive review that when the decision-maker provides additional information via a semantic scale, the best cardinal method is indeed CSR. On the other hand, if only a pure rank order is given (no strength information), or if the decision-maker provides other forms of inputs (precise numerical ratios, a ranking of differences, etc.), the literature has several promising methods but no single one has been universally recommended yet. This underscores that research is still ongoing to identify the most effective and reliable surrogate weighting techniques for each type of preference input.

3.3 Regression Surrogates

Another branch of methodologies treats the weight derivation problem as an inverse decision problem using dominance relations or ordinal regression. Instead of starting from criteria preferences, these approaches start from holistic judgments about alternatives and work backward to infer weights that would make those judgments consistent with an additive utility model. For instance, the old UTA/UTASTAR methods (Jacquet-Lagrèze and Słowiński, 1982; Siskos and Yannacopoulos, 1985) take as input a ranking or partial ranking of a set of alternatives and solve a linear programming problem to find an additive value function (criteria weights and value scores on each criterion) that best reproduces that ranking. Similarly, in ordinal regression approaches, the decision-maker might say "Alternative A should outrank B, given my preferences" (without explicitly stating any weights), and the algorithm searches for any weight vector (and perhaps criterion utility shapes) that satisfies these pairwise dominance statements. If multiple weight vectors can fit the preference ordering, the outcome might be a *set* of feasible weights rather than a unique surrogate. Early work in this direction includes the ÉLECTRE family's indirect weight inference and MACBETH (which uses qualitative pairwise comparisons to derive interval scales), though those involve some user-supplied judgments.

A later key contribution was the development of robust ordinal regression (ROR) for MCDA. ROR does not select a single "best" weight vector. Instead, it characterises all weight vectors compatible with the decision-maker's ordinal statements (on criteria or alternatives) and uses this to provide robust conclusions such as identifying alternatives that are optimal for all such weights or calculating the frequency an alternative is optimal (Sarabando et al., 2019). This approach gave rise to tools like SMAA (Stochastic Multi-criteria Acceptability Analysis), which treat uncertain or partially specified weights as random variables and compute statistics such as each alternative's probability of being top-ranked (Mazurek and Strzałka 2022). For example, SMAA can assume a uniform distribution over all weight vectors consistent with whatever partial information is available and estimate an acceptability index for each alternative (the percentage of weight vectors for which that alternative is the best. These dominance-based techniques shift the focus from picking a single surrogate weight vector to understanding the space of possible weights. They leverage pairwise dominance: if under all admissible weights alternative A has a higher total score than B, $A \succ B$ can be concluded without ever having exact weights.

Some recent methods explicitly compute dominance intensity indices. For instance, Mateos et al. (2014) proposed to calculate a dominance matrix where each element indicates how strongly one alternative dominates another across the weight space. By aggregating these pair-wise dominance values, they derive a ranking of alternatives without needing to commit to one weight vector. Such methods are particularly useful when criteria weights are highly contentious or hard to pin down: they produce a ranking that is *robust* to weight uncertainty and high-light cases of close trade-offs (where dominance intensity is low or mutual). An illustrative example is minimax regret weighting: one can choose the alternative that minimises the maximum regret over all weight vectors consistent with the rank order of criteria. This essentially finds a decision that is safest against weight ambiguity, a concept explored by Sarabando and Dias (2009) who compared decision rules like maximax, maximin (optimist/pessimist criteria), and central weight heuristics under ordinal weight information and measured how often each rule selected the truly best alternative (their "hit rate"). Interestingly, they found that using the ROC weights as a representative point estimate was quite competitive with more elaborate rules across many test problems.

In general, dominance-based models contribute important theoretical insights: for instance, they highlight that if a particular alternative is never top-ranked under any weight vector satisfying the decision-maker's constraints, then no surrogate weighting (short of violating the decision-maker's input) could make it optimal. They also introduce concepts like weight stability regions, the set of weight vectors for which a given alternative remains optimal. Recent work by Mazurek and Strzałka (2022) defined a notion of central weights for each alternative: essentially the weight vector within that alternative's optimality region that is farthest from the boundaries (in some metric), along with a robustness radius that indicates how much weights can change before the alternative loses its top position. These concepts provide quantitative robustness indices: e.g. an alternative with a large stability radius is robustly optimal (requires a big weight perturbation to lose its status. While dominance and regression approaches are not "surrogate weighting" in the sense of producing a single weight vector, they often underpin the *validation* of surrogate methods and guide the design of new ones. By understanding the structure of weight space and dominance, researchers can craft surrogate weight formulas that, for example, maximise the probability of identifying the correct decision or maximise the stability of the choice.

3.4 Entropy Surrogates

Apart from subjective preference-based methods, a distinct category of automatic weight generation consists of objective weighting methods, which determine criteria weights from the data characteristics of the decision matrix alone (without any direct input from the decision-maker about importance). These approaches, often described as geometric or entropy-based, stem from the idea that the relative importance of a criterion can sometimes be inferred from how much information or variability that criterion contributes to the overall evaluation of alternatives. One of the best-known objective techniques is the Shannon entropy weight method. In this method, each criterion's performances across the set of alternatives are analysed to compute an entropy value E_j which reflects the amount of uncertainty or "disorder" in that criterion. If criterion j has identical values for all alternatives, its entropy is maximal (indicating it provides no useful discrimination) and it should receive a low weight. Conversely, if one alternative is very high and another very low on that criterion (high contrast), entropy is lower and the criterion is considered more informative for making distinctions (Zakeri et al., 2025). The normalised complement of entropy (often $d_j = 1 - E_j$) is taken as a measure of criterion j's information content, and weights are assigned proportionally to d_j. This entropy weight scheme (dating back to the 1980s) has been widely used in engineering and selection problems where the criteria are objective indicators (Dwivedi, and Sharma 2022). It automatically emphasises criteria that show more divergence among alternatives, on the rationale that such criteria carry more decision-relevant information.

A closely related method is the CRITIC method (Criteria Importance Through Inter-criteria Correlation) proposed by Diakoulaki et al. (1995). CRITIC goes beyond individual criterion variability by also accounting for conflict or redundancy between criteria. For each criterion j, it computes a contrast intensity $C_j = \sum_{k=1}^{N}(1 - r_{jk})$ where σ_j is the standard deviation of criterion j (measuring its intra-criterion variability) and r_{jk} is the Pearson correlation coefficient between criteria j and k. The term $\sum_k(1 - \mathrm{r_{jk}})$ effectively gauges how independent criterion j is from the others. It sums to a larger value if j is not strongly correlated with any other criterion, meaning j provides unique information. Thus, C_j will be high for criteria that both vary greatly across alternatives and are not duplicating information provided by other criteria. These C_j values are then normalised into weights. The philosophy behind CRITIC is to extract all the information included in the criteria investigated. Criteria with higher inherent information (variance) and less overlap with others get more weight. In a sense, entropy and CRITIC weights are geometric in that they depend on the geometry of the data cloud in the criteria space (entropy relates to the distribution shape, CRITIC to data spread and correlations). They require no input from the decision-maker, which makes them useful as default or supporting analyses.

However, these methods implicitly assume that *all variation is valuable*, which may not align with a decision-maker's values. A criterion might show high variability, but the decision-maker could still judge it as unimportant (e.g. perhaps cost varies a lot, but for some reason, cost is not a key criterion in a particular decision). Therefore, objective weights are often used in combination with subjective judgment rather than a replacement. They do, nonetheless, satisfy certain formal properties: typically, the objective weight formulas define a unique convex weight vector given the data, and these weights often maximise some entropy or dispersion criterion. For example, equal weights can be seen as the weights that maximise the entropy of the weighted sum since they treat all criteria uniformly (Ezell et al., 2021), whereas CRITIC weights could be seen as solving a certain variance allocation problem. In terms of robustness, objective weights can be sensitive to outliers or changes in the dataset (since a change in alternatives' values alters the computed weights). This raises an interesting point about *stability*: a robust surrogate weighting method ideally should not be too volatile to minor changes.

Some extensions in recent years, like MEREC (Method based on Removal Effects of Criteria), emphasise robustness by measuring how the overall evaluation would change if a criterion was removed (Keshavarz-Ghorabaee et al., 2021). In MEREC, the importance of a criterion is judged by the difference in some aggregate score (e.g. sum of performances or a composite utility) with and without that criterion; larger differences imply the criterion has a greater impact, thus a higher weight. This method is another data-driven weighting heuristic that tends to align with CRITIC/entropy in highlighting criteria that significantly affect outcomes.

It is important to point out that objective methods by design ignore the decision-maker's subjective value trade-offs since they measure information content, not value. Therefore, in theoretical discussions, they are often recommended as a starting point or a complementary analysis rather than a final say on weights. They work well in contexts where no decision-maker is available to provide detailed preferences, or where one wants to test how much the structure of the data alone would dictate the weighting. In the broader landscape of surrogate weighting, objective methods represent an extreme end of information requirement (requiring zero preference input), opposite the other extreme of full preference articulation (AHP, swing weights, etc.). Many real-world MCDA applications use a mix, for example using entropy weights to set an initial weighting which is then adjusted based on stakeholder feedback.

4. Performance Evaluation

After deriving weights through any of the above methods, researchers evaluate their performance using various validation methodologies and metrics. Since the ground truth weights in a real decision are rarely known, a common approach is simulation experiments: one generates synthetic decision problems with known "true" weights, applies the surrogate weighting methods using limited information (like only the rank order of those true weights), and checks how well each method approximates the true decision outcome (Sarabando et al., 2019). One key metric is the hit ratio, which measures the frequency that a method correctly identifies the top-ranked alternative (or the correct full ranking) as would be obtained with the true weights (Chergui and Jiménez-Martín, 2024). For instance, if method X yields a weight vector that leads to the same best choice as the true weight vector in 90 out of 100 simulated cases, its hit ratio is 90%. A high hit ratio indicates a method is effective at preserving the decision ordering. Another common metric is the rank correlation (such as Kendall's tau or Spearman's rho) between the alternative ranking produced by the surrogate weights and the ranking under the true weights (Chergui and Jiménez-Martín, 2024). This assesses overall ordering accuracy, not just the top choice. A high Kendall's tau (close to 1) means the surrogate weighting closely mirrors the true ordering of all alternatives (Danielson and Ekenberg, 2017).

Beyond alternative ordering, one can evaluate weight vector accuracy directly. This might involve computing the distance between the surrogate weight vector and the true weight vector (e.g. Euclidean or Manhattan distance in the weight space). However, a small error in weights does not always translate to a wrong decision if the decision problem is insensitive to those differences. Therefore, outcome-based measures (such as hit rate, rank correlation or regret) are usually more relevant. In some studies, researchers use the mean absolute error or mean squared error of weights as a descriptive metric, but focus interpretation on whether those errors matter for the decision.

5. Robustness

Another performance aspect is robustness: how stable the method's recommendations are under uncertainty or varying assumptions. Robustness can refer to *internal* robustness (sensitivity to noise in the provided input) or *external* robustness (stability of the decision with respect to weight perturbations). Internal robustness is evaluated by, for example, introducing slight errors or inconsistencies in the decision-maker's input (like a small mistake in the rank ordering) and seeing if the method still yields similar weights. A robust method should not drastically change weights for minor input changes. External robustness is often quantified by metrics like the stability radius or robustness index mentioned earlier. For a given alternative that is top-ranked by the surrogate weights, one can compute the minimum change in the weight vector needed to make a different alternative become the best. Methods that yield a larger such radius (meaning their recommended decision remains best for a larger neighbourhood of weight perturbations) are considered more robust.

Mazurek and Strzałka's (2022) concept of central weights and robustness radius provides a concrete way to measure this. If surrogate method A often produces weight vectors that are near the centre of the true optimality region (hence large radius), and method B produces weights near the boundary (small perturbation could flip the decision), then A would be judged more robust in a practical sense. In the context of rank-based surrogates, ROC has been lauded for its robustness: it tends to pick weight vectors that are not extreme, thereby avoiding borderline cases. On the other hand, a method like pure Rank Reciprocal yields more extreme weights (favouring the top-ranked criterion heavily), which could be risky. If the top two criteria were actually closer in importance than assumed, such a weighting might misrank alternatives.

Researchers also consider "efficacy" metrics that combine multiple aspects, for example a composite score that penalises both large weight estimation error and wrong alternative selection. Some studies use a loss function approach: define a loss if the chosen alternative is not the true best (perhaps weighted by how far down it is), or use the concept of regret (difference in value between the alternative chosen by surrogate weights and the value of the ideal alternative under true weights). The maximum regret across scenarios or the average regret can indicate how well a method performs in the worst case or on average. Sarabando and Dias (2009) examined such measures by looking at how close each method's chosen alternative's utility was to the optimal utility, on average, a measure also used in (Danielson and Ekenberg, 2016).

6. Recommendations

This has led to guidelines and recommendations, for instance: "if the decision-maker can provide a rough quantitative sense of differences (e.g. via a semantic scale of importance), then use method X"; "if only a rank order is available with no strength information, method Y (like ROC or CSR) is a safe choice"; if data characteristics are the only guide (no preference input), use entropy to at least reflect criterion variability" and so on. A recent review by Chergui and Jiménez-Martín (2024) attempts to identify which weighting methods are best suited for which type of input information, summarising decades of comparison studies. They concluded that for purely ordinal inputs, several methods (not just one) are "outstanding" and more comparative analysis is needed to pinpoint a winner, whereas for cardinal input, a clear winner CSR) emerged. A guide for choosing between ordinal and cardinal surrogate methods is provided in (Danielson and Ekenberg, 2021).

7. Conclusion

Automatic weight generation in MCDA has evolved from simple ad-hoc rules to a sophisticated toolkit of methods grounded in both normative theory and empirical testing. Rank-based methods like RS, RR, and ROC provided the first viable surrogates, with ROC generally offering an excellent balance of discrimination and robustness. Centroid/average approaches formalised the idea of neutrality under ignorance, giving a principled rationale for methods like ROC. Enhancements incorporating strength information (Simos, ROL, CSR, etc.) have increased the expressiveness of surrogate weights, allowing decision-makers to convey more nuance and thereby improving accuracy when used appropriately. Dominance-based models and ordinal regression shifted focus to what can be definitively concluded without precise weights, enriching our understanding of weight robustness and guiding the creation of surrogates that perform well across many weight scenarios. Geometric and entropy-based objective weighting offered an orthogonal perspective by deriving weights from data patterns, which is useful in the absence of preference information and ensures criteria with greater impact (variance or information) are not overlooked.

Each category comes with trade-offs in convexity (extreme vs. balanced weighting), robustness (sensitivity to input or data), information needs, and ability to represent the decision-maker's true value structure. The state-of-the-art today does not point to a one-size-fits-all solution; rather, the best surrogate weighting method depends on the context of information and the desired properties. If a decision-maker can only supply a rank ordering, methods like SR or other advanced ordinal techniques are recommended for their proven efficacy. If some strength distinctions are available, newer methods like CSR should be employed. And if no subjective input is available, entropy or CRITIC weights can serve as a starting point, albeit remaining cautious of their assumptions. The literature since 2010 has greatly expanded the toolbox and provided systematic comparisons, but it also highlights open questions. As noted recently by researchers, further comparative analyses are needed, especially for cases with only ordinal inputs, to decisively recommend a single method. The ongoing development of validation frameworks, from simulation testbeds to analytical measures of robustness, will continue to shape the understanding of which surrogate weights perform the best and under which circumstances.

Acknowledgement

This work was partially funded by the European Commission research programme Horizon Europe, grant agreement number 101074075.

Conflicts of Interest

The author declares no conflicts of interest.

References

Ahn, B. S. (2011). Compatible weighting method with rank order centroid: Maximum entropy ordered weighted averaging approach. *European Journal of Operational Research, 212*(3), 552–559.

Barron, F. H. (1992). Selecting a best multiattribute alternative with partial information about attribute weights. *Acta Psychologica, 80*(1–3), 91–103.

Barron, F. H., & Barrett, B. E. (1996). Decision quality using ranked attribute weights. *Management Science, 42*(11), 1515–1523.

Belton, V., & Stewart, T. J. (2002). *Multiple criteria decision analysis: An integrated approach*. Springer.

Bouyssou, D., Marchant, T., Pirlot, M., Tsoukias, A., & Vincke, P. (2006). *Evaluation and decision models: A critical perspective*. Springer.

Brans, J. P., & Mareschal, B. (2005). PROMÉTHÉE methods. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 163–195). Springer.

Chergui, Z., & Jiménez-Martín, A. (2024). On Ordinal Information-Based Weighting Methods and Comparison Analyses. *Information*, *15*(9), 527. **NEW**

Danielson, M., & Ekenberg, L. (2015). Using surrogate weights for handling preference strength in multi-criteria decisions. In B. Kaminski, G. E. Kersten, & T. Szapiro (Eds.), *Outlooks and Insights on Group Decision and Negotiation* (pp. 107–118). Springer.

Danielson, M., & Ekenberg, L. (2016). The CAR method for using preference strength in multi-criteria decision making. *Group Decision and Negotiation*, *25*(4), 775–797.

Danielson, M., & Ekenberg, L. (2017). A robustness study of state-of-the-art surrogate weights for MCDM. *Group Decision and Negotiation*, *26*, 677–691.

Danielson, M., & Ekenberg, L. (2020). Automatic criteria weight generation for multi-criteria decision making under uncertainty. In A. T. de Almeida & D. C. Morais (Eds.), *Innovation for systems information and decision* (pp. 1–14). Springer.

Danielson, M., & Ekenberg, L. (2021). The worth of cardinal information in MCDM – A guide to selecting weight-generating functions. In A. T. de Almeida & D. C. Morais (Eds.), *Innovation for systems information and decision, INSID 2021, Proceedings* (pp. 20–35). Springer.

Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The CRITIC method. *Computers & Operations Research*, *22*(7), 763–770.

Dwivedi, P. P., & Sharma, D. K. (2022). Application of Shannon entropy and CoCoSo methods in selection of the most appropriate engineering sustainability components. *Cleaner Materials*, *5*, 100118.

Ezell, B., Lynch, C. J., & Hester, P. T. (2021). Methods for Weighting Decisions to Assist Modelers and Decision Analysts: A Review of Ratio Assignment and Approximate Techniques. *Applied Sciences*, *11*(21), 10397.

Figueira, J., & Roy, B. (2002). Determining the weights of criteria in the ELECTRE type methods with a revised Simos procedure. *European Journal of Operational Research*, *139*(2), 317–326.

French, S., & Xu, D. L. (2005). Comparison study of MCDA algorithms: Testing on the data generated by random weighting. *European Journal of Operational Research*, *160*(2), 564–582.

Greco, S., Ehrgott, M., & Figueira, J. R. (Eds.). (2016). *Multiple criteria decision analysis: State of the art surveys*. Springer.

Hatefi, M. A. (2023). An improved rank order centroid method (IROC) for criteria weight estimation. *Informatica*, *34*(2), 249–270.

Ishizaka, A., & Nemery, P. (2013). *Multi-criteria decision analysis: Methods and software*. Wiley.

Jacquet-Lagrèze, E., & Siskos, Y. (1982). Assessing a set of additive utility functions for multicriteria decision making: The UTA method. *European Journal of Operational Research*, *10*(2), 151–164.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value trade-offs*. Cambridge University Press.

Keshavarz-Ghorabaee, M., Amiri, M., Zavadskas, E. K., Turskis, Z., & Antucheviciene, J. (2021). Determination of Objective Weights Using a New Method Based on the Removal Effects of Criteria (MEREC). *Symmetry*, *13*(4), 525.

de Marchi, B., Lucertini, G., & Tsoukias, A. (2000). From evidence-based policy making to policy analytics. *Annals of Operations Research*, *275*(1), 165–184.

Mateos, A., Aguayo, F., & Siskos, Y. (2014). A new approach to criteria weighting in multicriteria decision-making. *Journal of the Operational Research Society*, *65*(1), 1–12.

Mazurek, J., & Strzałka, D. (2022). On the Monte Carlo weights in multiple criteria decision analysis. *PLOS ONE*, *17*(10), e0268950.

Roy, B. (1991). The outranking approach and the foundations of ELECTRE methods. *Theory and Decision*, *31*(1), 49–73.

Roy, B., & Slowinski, R. (2013). Questions guiding the choice of a multicriteria decision aiding method. *European Journal of Operational Research*, *231*(2), 261–271.

Sarabando, P., & Dias, L. C. (2009). Multiattribute choice with ordinal information: A comparison of different decision rules. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *39*(3), 545–554.

Sarabando, P., Dias, L. C., & Vetschera, R. (2019). Group decision making with incomplete information: A dominance and quasi-optimality volume-based approach using Monte-Carlo simulation. *International Transactions in Operational Research*, *26*(1), 318–339.

Siskos, Y., & Yannacopoulos, D. (1985). UTASTAR: An ordinal regression method for building additive value functions. *Investigação Operacional*, *5*(1), 39–53.

Stillwell, W. G., Seaver, D. A., & Edwards, W. (1981). A comparison of weight approximation techniques in multiattribute utility decision making. *Organizational Behavior and Human Performance*, *28*(1), 62–77.

Zakeri, S., Konstantas, D., Chatterjee, P., & Zavadskas, E. K. (2025). Soft cluster-rectangle method for eliciting criteria weights in multi-criteria decision-making. *Scientific Reports*, *15*, 284. **NEW**