Science, Engineering and Technology Vol. 6, No. S1, pp. 119-144
www.setjournal.com https://doi.org/10.54327/set2026/v6.is1.439

ECSW: An Integrated Model for Expert-Based Criteria Evaluation, Selection, and Weighting in Multi-Criteria Decision-Making

Ajdin Džananović, Aida Kalem, Adisa Medić, Alem Čolaković
University of Sarajevo, Faculty of Traffic and Communications, Zmaja od Bosne 8, Sarajevo 71000, Bosnia and Herzegovina.

Abstract

This paper proposes an integrated Expert Criteria Selection and Weighting (ECSW) model for expert-based criteria evaluation, selection, and weighting in multi-criteria decision-making (MCDM). The model is designed to preserve the expert-driven nature of the decision process while reducing the cognitive burden and potential inconsistency associated with pairwise comparisons and criteria ranking. It integrates four components: statistical criteria selection, assessment of the informativeness of expert ratings, determination of criteria weights based on their relative importance and rating stability, and sensitivity analysis. Candidate criteria with below-average ratings are evaluated by testing whether their expert-specific deviations from the experts' average rating patterns are significantly below zero, using a one-sided one-sample t-test followed by the Holm–Bonferroni correction for multiple comparisons. The final weights are obtained by adjusting equal base weights according to the relative importance of each criterion and the dispersion of expert ratings. Model stability is examined using a leave-one-expert-out procedure and Pearson and Spearman correlation coefficients. The weighting procedure was preliminarily verified against results obtained using the Ordinal Priority Approach. A strong agreement was observed for both criteria weights ($r = 0.956$) and rankings ($\rho = 0.929$), with $p < 0.001$ in both cases. The two models also identified the same ten highest-ranked criteria, although some differences occurred in their ordering. The ECSW model is particularly suitable for decision problems involving a large number of criteria for which conventional pairwise-comparison or ranking-based procedures may be difficult to apply. A supplementary ECSW implementation tool is provided to facilitate practical application and reproducibility.

Keywords: Multi-Criteria Decision-Making (MCDM), Criteria Weighting, Expert Ratings, Expert-Based Evaluation, Statistical Criteria Selection, Sensitivity Analysis.

1. Introduction

Criteria weighting represents one of the fundamental methodological issues in multi-criteria decision-making (MCDM), as the manner in which criteria are evaluated directly influences the interpretation of their importance, the final results of the analysis, and the selection of the optimal alternative. In many application domains, decision-makers are faced with a large number of criteria that do not contribute equally to the decision problem.

A wide range of models for determining criteria weights has been developed in the literature. These models generally rely on expert evaluations, mathematical and statistical characteristics of the available data, or a combination of subjective and objective information. Review studies indicate that the field of criteria weighting is methodologically well established and encompasses numerous classical and contemporary approaches, with the choice of weighting model having a direct impact on the reliability of decision

Corresponding author: Ajdin Džananović (ajdin.dzananovic@fsk.unsa.ba)
Received: 17 July 2026; Revised: 24 August 2026; Accepted: 1 September 2026; Published: 7 September 2026

outcomes [1], [2], [3]. Subjective weighting models [4] incorporate expert knowledge and contextual understanding of the problem through pairwise comparisons, criteria ranking, or other forms of relative evaluation. In contrast, objective weighting models [5] determine criteria weights based on the statistical properties of the decision data. Their application is particularly appropriate when a reliable decision matrix is available, alternatives are clearly defined, and measurable performance values exist for each criterion. However, in decision problems where quantitative performance data cannot be reliably established or a decision matrix cannot yet be constructed, objective weighting methods cannot be directly applied.

A particular challenge arises in decision problems involving a large number of criteria while preserving the expert-driven nature of the evaluation process [6]. Under these conditions, pairwise comparison of criteria becomes impractical, whereas complete ranking is often difficult, especially when criteria exhibit similar levels of importance or when experts do not share a clear consensus regarding their relative priorities.

Motivated by these challenges, this study proposes an integrated model for the expert-based evaluation, selection, and weighting of relevant criteria. Unlike approaches that rely on pairwise comparisons [7] or criteria ranking [8], the proposed model requires experts to evaluate each criterion independently with respect to the decision problem using a predefined numerical rating scale. This preserves the expert-driven nature of the evaluation process while simplifying data collection and reducing the cognitive burden and risk of inconsistent judgments, particularly when a large number of criteria are involved.

The objective of this study is to develop a methodologically straightforward, statistically supported, and computationally efficient model for expert-based criteria weighting. Its main contribution lies in integrating direct expert ratings, analysis of each criterion's deviation from the overall mean, statistical criteria selection, assessment of the informativeness of expert evaluations, weight determination based on both relative importance and rating stability, and sensitivity analysis of the final results. Expert ratings are therefore used not only to calculate criteria weights but also to determine whether the collected evaluations provide sufficient differentiation among criteria and whether the resulting weights and rankings remain stable. In this way, the model combines the simplicity of direct expert ratings with a level of statistical support that is generally absent from approaches based solely on pairwise comparisons, criteria ranking, or direct weight calculation. The model produces a statistically supported set of retained criteria, their corresponding weights, and an assessment of result stability. To support practical application and reproducibility, an implementation-ready tool for applying the Expert Criteria Selection and Weighting (ECSW) model is provided alongside the article.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature on criteria-weighting models. Section 3 presents the methodological framework of the proposed ECSW model. Section 4 provides its preliminary component-wise verification using numerical examples from the literature. Section 5 discusses the obtained results. Section 6 outlines the study's limitations and directions for future research, while Section 7 presents the main conclusions.

2. Literature Review

In response to the need for more reliable representation of criteria importance, a large number of criteria weighting models have been developed over the past decades, differing in terms of their information sources, the manner in which evaluations are collected, and the procedures used to derive criteria weights. In the literature, criteria weighting models are commonly classified into subjective, objective, and integrated approaches [1], [2]. Each of these groups exhibits specific advantages and limitations regarding ease of application, expert requirements, consistency of judgments, and the reliability of the resulting weights.

Subjective criteria weighting approaches are based on expert judgments regarding the relative importance of criteria. Information on decision-makers' preferences is obtained through direct rating, point allocation, criteria ranking, or pairwise comparisons. Among the most widely used methods in this group are the Analytic Hierarchy Process (AHP), Fuzzy Analytic Hierarchy

Process (FAHP), Best–Worst Method (BWM), and Full Consistency Method (FUCOM). Although all these methods rely on expert preference information, their elicitation requirements differ considerably. AHP and FAHP require a complete pairwise comparison matrix, which results in a rapidly increasing number of judgments as the number of criteria grows. BWM reduces this requirement by focusing on comparisons involving the best and worst criteria. FUCOM further simplifies the procedure by first ranking the criteria and then requiring only n−1 comparisons between adjacent criteria in the established ranking. Therefore, a large number of required judgments should not be considered a limitation of FUCOM. Its main elicitation requirement is the establishment of a complete criteria ranking and the assessment of comparative priorities between adjacent criteria. Nevertheless, establishing a complete ranking may become cognitively demanding when a large number of criteria with similar levels of importance is considered [9], [10], [11]. To reduce the number of required judgments and alleviate experts' cognitive burden, several approaches that do not rely on conventional pairwise comparisons have been proposed. One such model is the Ordinal Priority Approach (OPA), in which criteria weights are determined based on the ranking of criteria without constructing pairwise comparison matrices [12]. This simplifies the data collection process and reduces the cognitive effort required from experts. However, the model still requires experts to establish a complete ranking of criteria according to their importance, as well as to rank the participating experts.

A similar objective of simplifying the weighting process is pursued by the Fuzzy Weighted Zero Inconsistency Criterion (FWZIC) model. Unlike approaches based on pairwise comparisons or criteria ranking, FWZIC allows experts to evaluate the importance of each criterion independently, without directly comparing criteria with one another [13]. This eliminates the inconsistency problem characteristic of comparison-based models while simultaneously reducing experts' cognitive burden and simplifying the weighting procedure. Since FWZIC is a fuzzy-based approach, expert evaluations are transformed into fuzzy numbers, introducing an additional layer of methodological complexity into the weight calculation process.

Among subjective weighting models, the Criteria Importance through Intercriteria Correlation and Management by Assessment Scores (CIMAS) model also deserves attention. In this approach, criteria importance is determined using direct expert evaluations while additionally weighting experts according to their professional experience. Experts assess each criterion using a numerical scale from 1 to 10, whereas the weight assigned to each expert depends on the number of years of experience in the relevant domain. This approach preserves the simplicity of direct criteria evaluation while acknowledging the assumption that the judgments of more experienced experts should have a greater influence on the final criteria weights [14]. Overall, the review of subjective weighting models confirms the central role of expert knowledge in multi-criteria decision analysis while simultaneously highlighting the need for more robust, consistent, and methodologically simpler frameworks for collecting and processing expert evaluations [4].

Unlike subjective approaches, objective weighting models derive criteria weights directly from the decision matrix without relying on expert judgments. Classical objective models include Entropy, Criteria importance through intercriteria correlation (CRITIC), and the standard deviation model, whereas more recent developments include Method based on the Removal Effects of Criteria (MEREC), Centroidous, Integrated determination of objective criteria weights (IDOCRIW), Logarithmic normalization and standard deviation (LOGSTA), and various modified statistical approaches [5], [15], [16].

The MEREC introduces the original concept of measuring criterion importance according to the impact of removing a criterion on the performance of alternatives. Criteria weights are therefore determined from the changes observed in the overall performance of alternatives . Fuzzy MEREC extends this concept to uncertain and linguistic data by employing fuzzy numbers, enabling its application under conditions of imprecise information while demonstrating stable weighting results even in the presence of outliers [9], [17].

A more recent development in objective criteria weighting is the LOGSTA method. LOGSTA determines criteria weights directly from quantitative data by combining logarithmic normalization with the standard deviation of normalized criterion values. The method applies separate logarithmic normalization procedures for benefit and cost criteria, thereby explicitly accounting for criterion orientation. The resulting normalized values and their variability are subsequently used to derive the final objective criteria weights [18].

Several studies have critically examined the Entropy, CRITIC, and related objective weighting models, demonstrating that the resulting criteria weights may be highly sensitive to the normalization procedure, probability estimation, the structure of the decision matrix, and other methodological assumptions [5], [19]. Consequently, the literature increasingly proposes modified versions of existing models, integrated weighting approaches, and additional stability analyses. Comparative studies further indicate that different objective models, such as Entropy, MEREC, Logarithmic percentage change-driven objective weighting (LOPCOW), and CRITIC, often produce similar rankings of alternatives despite differences in the calculated criteria weights and optimal parameter settings. Therefore, many authors recommend validating the stability of the obtained results and examining the correlation between ranking outcomes [20], [21], [22], [23].

Integrated weighting approaches aim to combine the advantages of subjective and objective models by incorporating both expert knowledge and the mathematical and statistical characteristics of the available data. Such approaches may improve the robustness of the obtained results; however, they are generally more methodologically complex and primarily focus on combining multiple information sources to derive the final criteria weights. In many cases, integrated approaches do not fully address the issue of simplifying the collection of expert evaluations, nor do they necessarily incorporate statistical criteria selection, assessment of the statistical significance of differences between criteria, or evaluation of expert agreement.

Although integrated approaches combine the strengths of subjective and objective weighting models, they do not necessarily incorporate statistical criteria selection as an integral part of the weighting procedure, nor do they consistently assess the informativeness and stability of expert ratings. In a systematic review of 62 MCDM studies, Theunissen et al. [24] found that most studies do not employ a formalized procedure for criteria selection. Instead, criteria are typically adopted from previous literature or determined through expert judgment without appropriate validation. The authors specifically emphasize the need for integrated approaches that combine criteria selection and validation in order to improve the reliability of MCDM models. In existing research, this issue is most commonly addressed through separate preliminary criteria selection models, such as the Fuzzy Delphi Method, followed by the application of appropriate MCDM techniques to determine the weights of the selected criteria [25]. Statistical models constitute an important approach for analyzing and validating expert evaluations collected during the decision-making process. Their application depends on the characteristics of the collected data, the measurement scale used for expert evaluations, the number of experts involved, and the objectives of the statistical analysis. Accordingly, a variety of statistical procedures have been employed in the literature to assess expert agreement, validate expert judgments, examine the statistical significance of differences between criteria, and identify redundant criteria prior to applying MCDM models. For example, Chiclana et al. [26] applied the Wilcoxon matched-pairs signed-rank test to examine whether different similarity measures used to quantify consensus in group decision-making produce statistically different results. Mohammadi et al. [27] developed a Bayesian version of the Wilcoxon signed-rank test to determine whether differences between the weights of two criteria, derived from a group of expert evaluations, are statistically significant. In addition to the Bayesian Wilcoxon test, the authors also considered the application of the Bayesian t-test and the Bayesian Sign test (beta-binomial conjugate) for comparing criteria weights. Furthermore, Yurdakul and İç [28] proposed the use of a correlation test to identify and eliminate redundant criteria before applying MCDM models. These studies demonstrate that different statistical procedures serve different purposes depending on the characteristics of expert evaluations and the objectives of the analysis, including validation of expert judgments, assessment of expert agreement, comparison of criteria, and elimination of redundant criteria.

To better position the proposed Expert Criteria Selection and Weighting (ECSW) model with respect to existing weighting models, Table 1 presents a comparative overview of representative subjective and objective approaches. The comparison considers the type of model, the basis for determining criteria weights, the requirement for pairwise comparisons or criteria ranking, and their suitability for decision-making problems involving a large number of criteria. The comparative overview demonstrates that the proposed ECSW model occupies a distinct position among existing criteria weighting models. Compared with pairwise comparison-based approaches, the proposed model reduces the number of expert judgments required and consequently lowers the risk of inconsistent evaluations.

Table 1. Comparative overview of selected criteria weighting models.

Model	Model Type	Basis for Criteria Weight Determination	Pairwise Comparison/ Ranking	Suitability for a Large Number of Criteria
AHP	Subjective	Expert pairwise comparisons of criteria	Yes	Limited, as the number of pairwise comparisons increases rapidly with the number of criteria
FAHP	Subjective	Fuzzy expert pairwise comparisons of criteria	Yes	Limited due to the additional complexity introduced by fuzzy evaluations
BWM	Subjective	Expert comparisons of the best and worst criteria with all remaining criteria	Yes	Relatively suitable; requires fewer comparisons than AHP but still relies on pairwise comparisons
FUCOM	Subjective	Expert ranking of criteria and comparison of priorities between adjacent criteria	Yes	Suitable, as it requires only n-1 comparisons and follows a relatively simple procedure, although an initial ranking of criteria must be established
OPA	Subjective	Ranking of criteria, alternatives, and experts	Yes	Relatively suitable, as it does not require pairwise comparisons, but it requires a complete ranking
FWZIC	Subjective	Direct fuzzy expert evaluations of criteria	No	Suitable because it does not require pairwise comparisons or ranking, although it involves fuzzy transformation
CIMAS	Subjective	Direct expert evaluations of criteria with expert weighting based on experience	No	Suitable because it relies on direct evaluation of criteria
Entropy	Objective	Performance values of alternatives in the decision matrix	No	Depends on the availability of a complete and reliable decision matrix
CRITIC	Objective	Criteria variability and inter-criteria correlation within the decision matrix	No	Depends on the availability of a complete and reliable decision matrix
MEREC	Objective	Effect of criterion removal on the performance of alternatives	No	Depends on the availability of a complete and reliable decision matrix
LOGSTA	Objective	Logarithmic normalization of decision-matrix values combined with standard deviation	No	Depends on the availability of a complete and reliable decision matrix; accommodates both benefit and cost criteria
Centroidous	Objective	Statistical characteristics of alternative performance values	No	Depends on the availability of a complete and reliable decision matrix
IDOCRIW	Objective	Combination of objective information derived from the decision matrix	No	Depends on the availability of a complete and reliable decision matrix
Proposed ECSW Model	Subjective	Direct expert evaluations of criteria using an interval scale	No	Suitable for problems involving a large number of criteria

In contrast to ranking-based models, it does not require the establishment of a complete ranking of criteria, which is particularly advantageous when dealing with a large number of criteria or when the criteria exhibit similar levels of importance. Unlike objective weighting models, the proposed model can be applied even when the alternatives and the decision matrix have not yet been defined, as it relies solely on expert ratings of criteria importance.

Furthermore, its distinguishing feature lies in the fact that expert ratings are not used exclusively for calculating criteria weights but are also subjected to statistical processing through criteria selection, assessment of the informativeness of expert ratings, and analysis of the stability of the obtained results.

The review of existing models indicates that each group of criteria weighting models offers specific advantages but also suffers from inherent limitations. Despite the considerable development of weighting techniques, relatively few approaches simultaneously provide a simple expert ratings procedure without requiring pairwise comparisons or complete criteria ranking while also incorporating statistical evaluation of the collected expert ratings. In particular, comparatively few methodological models explicitly integrate statistical criteria selection, assessment of the informativeness of expert ratings, stability-adjusted weighting, and sensitivity analysis within a single expert-based procedure. This comparatively limited methodological integration constitutes the primary motivation for developing the Expert Criteria Selection and Weighting (ECSW) model proposed in this study.

3. The Proposed ECSW Model

This section presents an integrated model for the expert-based selection of relevant criteria and the determination of their corresponding weights. In this study, the term integrated refers to the combination of expert-based evaluation, statistical criteria selection, informativeness assessment, weight determination, and sensitivity analysis within a single procedural framework. It does not refer to the combination of subjective and objective weighting information, according to its source of input data, the ECSW model remains an expert-based subjective weighting model. The fundamental assumption of the proposed model is that experts evaluate each criterion independently with respect to the decision problem using a predefined rating scale, without the need to perform pairwise comparisons or establish a complete ranking of criteria according to their importance.

The proposed model is based on the overall mean of all expert ratings for the initial set of criteria. Each criterion is subsequently evaluated according to the relationship between its average score and the overall mean of the evaluated criteria set. Criteria with average scores above the overall mean are assigned greater relative importance, whereas criteria with average scores below the overall mean are assigned lower relative importance and are further subjected to statistical testing to assess whether they should be excluded from subsequent analysis. In this way, the proposed ECSW model provides a comprehensive framework for the statistical processing of expert ratings. The final output of the model consists of a statistically supported set of retained criteria with their corresponding weights, an interpretation of the differences between criteria, and an assessment of the stability of the obtained results.

The overall implementation procedure of the proposed model can be described through a sequence of interconnected steps. The process begins with the definition of the initial set of criteria and the reference rating scale, followed by expert ratings of the criteria. Subsequently, the relative position of each criterion with respect to the overall mean of the expert ratings is determined, statistical criteria selection is performed, the retained set of criteria is established, the informativeness of the retained criteria set is assessed, the final criteria weights are calculated, and a sensitivity analysis is conducted.

The complete workflow of the proposed ECSW model is illustrated in the flowchart shown in Figure 1, while a detailed description of each step is provided in the following subsections. To support practical implementation of the proposed procedure, an implementation-ready ECSW tool is provided alongside the article.

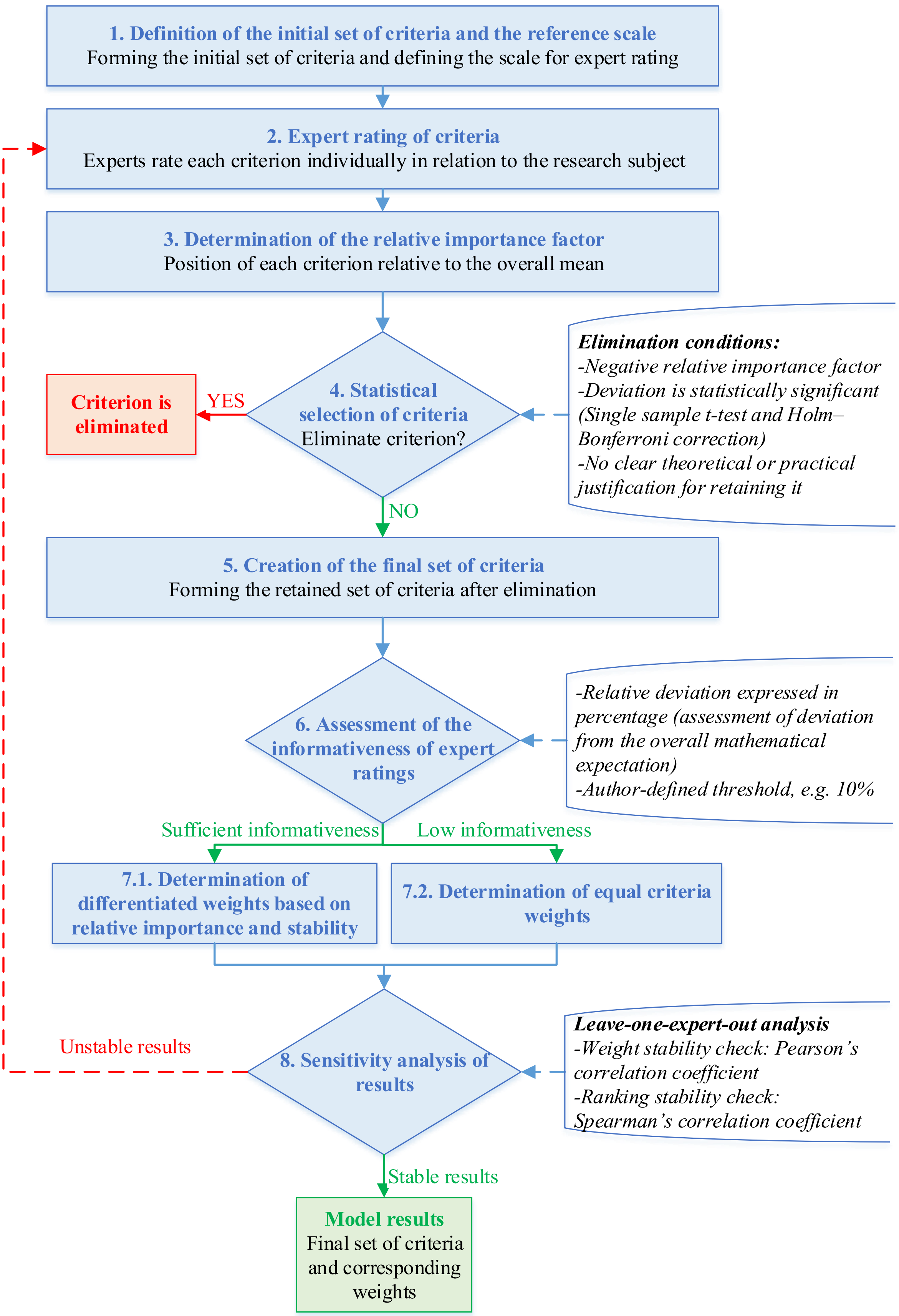

Figure 1. Flowchart of the proposed ECSW model.

Step 1 – Definition of the Initial Criteria Set and the Reference Rating Scale

The first step involves defining the initial set of criteria. The criteria are identified based on the theoretical framework, previous research, practical experience, regulatory or professional guidelines, and the specific characteristics of the decision problem under consideration. This step is particularly important because the initial set of criteria does not necessarily represent the final set of criteria. The proposed ECSW model allows expert ratings to confirm, weaken, or effectively eliminate the importance of individual criteria during the subsequent analysis. After defining the criteria, a reference numerical interval scale is established for expert ratings. The rating scale should be adapted to the specific decision problem, since experts do not assess criteria in an abstract sense but rather evaluate the magnitude of their influence on the problem being analyzed. The proposed model is not restricted to a predefined rating scale and supports the use of various rating scales, such as 1–5, 1–7, 1–9, 1–10, or any other interval scale, provided that the intervals between consecutive values are equal and that the values increase from the lowest to the highest level of importance. This assumption allows expert ratings to be treated as interval data, thereby justifying the application of descriptive statistical measures, including the arithmetic mean, standard deviation, and coefficient of variation, in the process of criteria analysis and evaluation.

Step 2 – Expert Ratings of Criteria

In the second step, experts evaluate each criterion independently according to the importance of its influence on the decision problem, without performing pairwise comparisons between criteria. The score assigned by expert j to criterion i is denoted as AE_{ij}. Based on these evaluations, the expert rating matrix is constructed as follows:

$$AE = \begin{bmatrix} AE_{11} & AE_{12} & \cdots & AE_{1m} \\ AE_{21} & AE_{22} & \cdots & AE_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ AE_{n1} & AE_{n2} & \cdots & AE_{nm} \end{bmatrix} \qquad (1)$$

where n denotes the number of criteria and m denotes the number of experts.

Step 3 – Determination of the Relative Importance Factor of Each Criterion with Respect to the Overall Mean of Expert Ratings

After constructing the expert ratings matrix, the next step is to determine the relative position of each criterion with respect to the overall mean of the expert ratings. The purpose of this step is to identify whether a given criterion has been evaluated as being above, below, or at the average level of importance within the entire set of criteria. First, the overall mean of all expert ratings is calculated as follows:

$$\overline{AE} = \frac{1}{n * m} \sum_{i=1}^{n} \sum_{j=1}^{m} AE_{ij} \qquad (2)$$

The value of $\overline{AE}$ represents the overall mean level of importance assigned by the experts to the criteria in the decision problem under consideration. Subsequently, the mean expert ratings for each criterion, denoted by $\overline{AE_i}$ is calculated as follows:

$$\overline{AE_i} = \frac{1}{m} \sum_{j=1}^{m} AE_{ij} \qquad (3)$$

Based on the relationship between the mean expert ratings of each criterion and the overall mean of all expert ratings, the relative importance factor of each criterion (MJ_i) is determined as follows:

$$MJ_i = \frac{\overline{AE_i}}{\overline{AE}} \qquad (4)$$

The relative importance factor indicates the position of each criterion with respect to the overall mean expert ratings of all criteria. If $MJ_i > 1$, the criterion has been evaluated above the overall mean, indicating that experts consider it to be more important than the average criterion within the evaluated set. Conversely, if $MJ_i < 1$, the criterion has been evaluated below the overall mean, indicating a lower level of relative importance. The overall mean is used as an internal, set-dependent reference point rather than as an absolute threshold of criterion relevance. Therefore, the relative position of a criterion depends on the composition of the initial criteria set, which should be theoretically justified, sufficiently comprehensive, and appropriately aligned with the decision problem.

To facilitate further interpretation, both the absolute deviation (D_i) and the relative percentage deviation (RD_i) from the overall mean are calculated as follows:

$$D_i = \overline{AE_i} - \overline{AE} \tag{5}$$

$$RD_i = (MJ_i - 1) * 100 \tag{6}$$

The absolute deviation indicates the extent to which the mean evaluation of an individual criterion differs from the overall mean of the evaluated criteria set, whereas the relative percentage deviation expresses the magnitude of this difference relative to the overall mean. This step provides the basis for the subsequent stages of the proposed ECSW model, as criteria selection, assessment of the informativeness of expert ratings, and criteria weighting are all derived from the deviation of each criterion's mean evaluation from the overall mean of the complete set of expert ratings.

Step 4 – Statistical Criteria Selection and Formation of the Final Criteria Set

Following the determination of the relative importance factor of each criterion, a statistical criteria selection procedure is performed. The purpose of this step is to exclude from further analysis those criteria that can be statistically and substantively demonstrated to exhibit a significantly lower level of expert-assessed importance.

The relative importance factor MJ_i serves as the initial criterion for the selection procedure. Criteria for which $MJ_i \geq 1$ are retained for further analysis, as their mean evaluations are equal to or greater than the overall mean of the expert ratings. In contrast, criteria with $MJ_i < 1$ are identified as candidate criteria for potential elimination. It should be emphasized that the relative deviation from the overall criterion-set mean is used only to identify candidate criteria for further assessment. These criteria are subjected to additional statistical evaluation because, although their mean evaluations indicate a negative deviation from the overall mean, being below the overall mean alone does not constitute sufficient evidence for their exclusion. For each candidate criterion, an expert-specific centred deviation is calculated by comparing the individual expert rating of that criterion with the same expert's mean rating of all remaining criteria. This procedure accounts for differences in experts' rating behavior, recognizing that some experts systematically assign higher scores, whereas others tend to provide lower evaluations. The deviation (δ_{ij}) is calculated as follows:

$$\delta_{ij} = AE_{ij} - \overline{AE}_{-ij} = AE_{ij} - \frac{1}{n-1}\sum_{\substack{k=1 \\ k\neq i}}^{n} AE_{kj} \tag{7}$$

where $\overline{AE}_{-ij}$ denotes the mean score assigned by expert j to all criteria except criterion i. The value of δ_{ij} indicates whether expert j evaluated the candidate criterion i above or below their own average evaluation pattern. Excluding criterion i from expert j's reference mean prevents the rating being tested from contributing to its own comparison value and provides an expert-specific centred deviation.

Since the reference scale adopted in the proposed ECSW model is defined as a numerical interval scale with equal intervals between consecutive values, expert ratings are treated as interval data. This assumption is consistent with the mathematical framework of the proposed model. Accordingly, the statistical selection of candidate criteria is performed using a one-sided one-sample t-test [29] applied to the values of δ_{ij}. For each candidate criterion, the test examines whether the mean deviation of its expert ratings is statistically significantly less than zero. The null hypothesis assumes that the mean deviation of the candidate criterion is not negative, indicating that the criterion has not been systematically evaluated below the experts' average evaluation pattern. The result of the test for each candidate criterion is a p-value. If $p_i < 0{,}05$, the negative deviation of the candidate criterion is considered statistically significant at the 5% significance level.

However, because the statistical test is performed for multiple candidate criteria, the obtained p-values must be adjusted to reduce the risk of incorrectly eliminating criteria due to multiple hypothesis testing. For this purpose, the Holm–Bonferroni correction is applied. This procedure orders the p-values from the smallest to the largest and adjusts the decision regarding statistical significance while controlling the family-wise Type I error rate across the entire set of conducted tests [30]. The correction yields an adjusted p-value, denoted as p_i^{adj}. If, after applying the Holm–Bonferroni correction, the adjusted p-value satisfies $p_i^{adj} \geq 0{,}05$ the candidate criterion is retained in the model because its negative deviation is no longer statistically significant. Otherwise, the criterion is identified as a strong candidate for elimination.

The final decision regarding the exclusion of a criterion is not based solely on statistical evidence but also considers its substantive importance within the specific decision problem. Consequently, a criterion identified as a candidate criterion for elimination may still be retained if there is a clear legal, theoretical, or practical justification for its inclusion. Such justification may exist, for example, when the criterion is prescribed by legal or regulatory frameworks, when its importance has been consistently confirmed in previous research, or when it represents an indispensable component of the analyzed problem. In such cases, the rationale for retaining the criterion should be explicitly stated and adequately justified. This procedure prevents the automatic elimination of statistically weaker criteria that may nevertheless possess substantial practical or theoretical relevance. Therefore, a criterion is excluded only if all three of the following conditions are simultaneously satisfied:

- it exhibits a negative relative deviation RD_i,
- the negative deviation remains statistically significant after the Holm–Bonferroni correction for multiple testing; and
- there is no clear legal, theoretical, or practical justification for retaining the criterion.

Within the proposed ECSW model, statistical criteria selection is performed when the number of experts (observations) falls within a range considered appropriate for the application of the one-sample t-test. Rather than prescribing universal thresholds applicable to all decision-making problems, the proposed model adopts a context-dependent approach. Although the one-sample t-test was originally developed for relatively small samples, previous studies have highlighted the limitations associated with very small sample sizes [31]. As a practical, context-dependent guideline rather than a universal minimum requirement, the statistical selection stage is generally more defensible when approximately ten or more experts are available. Before applying the one-sample t-test, the distribution of the expert-specific centred deviations and the presence of potentially influential observations should be examined. This is particularly important for small expert panels, in which individual ratings may have a substantial effect on the test result and non-parametric alternatives may be more appropriate. The observations used in the test are assumed to be independent across experts. However, the reference value for each expert-specific deviation is derived from that expert's ratings of the remaining criteria and therefore reflects the expert's overall rating pattern. Consequently, the statistical result should be interpreted within the context of the number of criteria, the distribution of ratings, and the composition of the expert panel. For larger samples, the t-test remains applicable, although other procedures may also be considered depending on the distributional characteristics of the data [32]. Depending on the measurement scale adopted for expert ratings, other statistical procedures may also be applied at this stage of the model. If expert ratings are regarded as ordinal data rather than interval data, non-parametric statistical tests [33], [34], such as the Wilcoxon signed-rank test or the one-sample sign test, should be considered. Regardless of the statistical procedure employed, the Holm–Bonferroni correction is still recommended whenever multiple candidate criteria are tested simultaneously in order to control the family-wise Type I error rate.

Step 5 – Formation of the Final Set of Retained Criteria

Following the completion of the statistical selection procedure, the retained set of criteria is established, and the criteria weights are subsequently determined only for this final set. All subsequent steps of the proposed ECSW model are performed exclusively on the retained criteria.

The overall mean, denoted by $\overline{AE}'$ is recalculated using only the retained set of criteria. This step represents a recalibration of the model, ensuring that the final weighting procedure is based solely on the criteria that remain in the analysis. Consequently, criteria eliminated during the statistical selection stage no longer influence the reference value used to determine the relative importance of the retained criteria.

If no criterion is eliminated during the statistical selection procedure, the initial and the final criteria sets are identical, and therefore $(\overline{AE}' = \overline{AE})$ In this case, no recalibration of the overall mean is required.

Step 6 – Assessment of the Informativeness of Expert Ratings for the Retained Criteria Set

After establishing the retained set of criteria, the informativeness of the expert ratings is assessed. The purpose of this step is to determine whether the expert ratings contain sufficiently pronounced differences to justify differentiated weighting of the criteria in the subsequent stage of the model. Specifically, this step examines whether the mean evaluations of the retained criteria deviate sufficiently from the overall mean or whether all criteria have been evaluated at approximately the same level of importance. The assessment of informativeness is based on the relative percentage deviation RD_i calculated for each criterion.

A value of 0% indicates that the mean evaluation of a criterion is equal to the overall mean of the retained criteria set. If most criteria exhibit only very small relative percentage deviations, this suggests that the criteria have been evaluated almost equally and that the expert ratings provide limited information for meaningful differentiated weighting. Conversely, if the criteria exhibit more pronounced percentage deviations from the overall mean, the expert ratings contain sufficient informativeness to distinguish between criteria according to their perceived importance.

The informativeness threshold is intended as a guideline rather than a universal or exclusion criterion, and its value depends on the nature of the decision problem and the researcher's methodological judgment. In practical applications, the researcher defines an appropriate informativeness threshold (e.g., 5% or 10%). Nevertheless, if at least one criterion exhibits a relative deviation greater than 10% from the overall mean, it would not be appropriate to conclude that all criteria have been evaluated almost identically. In such cases, the expert ratings can be considered sufficiently informative, and the model proceeds with differentiated criteria weighting.

Based on this assessment, a methodological decision is made regarding the continuation of the weighting procedure. If the relative deviations indicate an adequate level of informativeness, the model proceeds with the calculation of the final criteria weights. Conversely, if the deviations are small and concentrated around the overall mean, the application of equal weights to all retained criteria is recommended.

Step 7 – Determination of the Final Criteria Weights

Following the statistical selection and validation of the retained set of criteria, the central step of the model is performed, namely the determination of the final criteria weights. The initial assumption is that all criteria have an equal base weight, but that this weight increases or decreases depending on the relative position of the criterion in relation to the overall mean of the expert ratings. At the same time, the intensity of the adjustment also depends on the stability of the expert ratings for the criterion under consideration (dispersion of expert ratings). In this way, the simplicity of direct expert ratings is retained, while an additional calibration is introduced that allows the criteria to be differentiated according to their relative importance. The base weight of a criterion (BT_i) is determined on the basis of the number of retained criteria:

$$BT_i = \frac{1}{n'} \tag{8}$$

where n' is the number of criteria that remained in the final set after statistical selection. The base weight represents the initial equal weight of each criterion before adjustment.

Next, new values of the relative importance factor of the criterion $MJ_i{}'$. are determined for each criterion. To clearly represent the direction and intensity of the base-weight adjustment, the base-weight adjustment factor (FK_i) is introduced. This expression determines the direction of the adjustment based on the relative position of the criterion in relation to the overall mean of the set of criteria, while the intensity of the adjustment is adapted to the stability of the expert ratings. If the adjustment factor is positive, the base weight increases. If it is negative, the base weight decreases, and if it equals zero, the base weight remains unchanged. The adjustment factor is determined as follows:

$$FK_i = ({MJ_i}' - 1) * S_i \tag{9}$$

where S_i is the stability factor, which depends on the standard deviation of the expert ratings by criterion. The value of the stability factor is higher when the dispersion of the expert ratings is lower. Conversely, greater dispersion of the ratings reduces the stability factor and moderates the adjustment of the base weight. The stability factor is determined as:

$$S_i = \frac{1}{1 + CV_i} = \frac{1}{1 + \frac{s_i}{\overline{AE'}_i}} = \frac{\overline{AE'}_i}{\overline{AE'}_i + s_i} \tag{10}$$

where CV_i is the coefficient of variation, and s_i is the standard deviation of the expert ratings for the criterion. The coefficient of variation indicates the relative dispersion of the expert ratings in relation to the average rating of the criterion. A higher value indicates greater relative dispersion of the expert ratings, while a lower value indicates that the ratings are more concentrated around the criterion mean. Within the ECSW model, the coefficient of variation is therefore interpreted as an indicator of rating stability rather than as a formal measure of inter-rater agreement or reliability.

The final weight of a criterion is obtained by adjusting the base weight using the adjustment factor, after which final normalization is performed:

$$w_i = \frac{BT_i + BT_i * FK_i}{\sum_{i=1}^{n'} (BT_i + BT_i * FK_i)}$$

$$w_i = \frac{BT_i(1 + FK_i)}{\sum_{i=1}^{n'} BT_i(1 + FK_i)} \tag{11}$$

The application of the coefficient of variation is justified because the stability of the expert ratings is not considered through absolute dispersion, but through relative dispersion in relation to the average rating of the criterion. The stability factor is defined as a decreasing function of dispersion. Therefore, when expert agreement is greater, CV_i is lower and S_i is closer to 1; when disagreement is greater, CV_i increases, while S_i reduces the intensity of the base-weight adjustment. The multiplicative relationship $({MJ_i}' - 1) * S_i$ allows relative importance to determine the direction of the adjustment, while stability determines its intensity. In this way, the model increases the weights of criteria rated above average, but does so more strongly only when there is greater agreement among the experts, whereas in cases of greater disagreement, the adjustment is brought closer to the base weight. Since S_i ranges from 0 to 1 and the expert ratings are positive, the procedure does not produce negative weights, while the final normalization ensures that the sum of the final weights is equal to one.

The use of equal base weights reflects a neutral starting position in which no criterion is assigned an a priori advantage before the expert ratings are processed. The subsequent adjustment introduces differentiation only on the basis of the relative importance and relative dispersion observed in the expert ratings. The multiplicative form of the adjustment factor was selected because it allows the relative importance factor to determine the direction of the adjustment, while the stability factor controls its intensity.

Table 2. Interpretation of the ECSW weighting mechanism.

Criterion position and rating dispersion	Expected effect on the base weight
Mean rating equal to the overall mean	No directional adjustment; weight remains at or close to the base weight
Above-average mean and low dispersion	Stronger positive adjustment
Above-average mean and high dispersion	Moderated positive adjustment
Below-average mean and low dispersion	Stronger negative adjustment
Below-average mean and high dispersion	Negative adjustment is moderated and the weight remains closer to the base weight

The proposed weighting mechanism has several interpretable properties. If the mean rating of a criterion is equal to the overall mean, its relative importance factor produces no directional adjustment and its weight remains at or close to the equal base weight. For a criterion rated above the overall mean, lower relative dispersion produces a stronger positive adjustment, whereas greater dispersion moderates the increase. For a criterion rated below the overall mean, lower dispersion produces a stronger negative adjustment, whereas greater dispersion moves the adjusted weight closer to the equal base weight. Thus, highly dispersed expert ratings do not strongly increase or decrease a criterion's weight. Detail Interpretation of the ECSW weighting mechanism is shown in Table 2.

Step 8 – Sensitivity Analysis of the Final Criteria Weights and Rankings

After determining the final criteria weights, a sensitivity analysis is conducted. The purpose of this step is to examine the stability of the obtained criteria weights and rankings with respect to changes in the expert group. Since the model is based on expert ratings, it is necessary to determine whether the final results depend on an individual expert or represent a stable group result.

In the proposed model, the sensitivity analysis is conducted using the leave-one-expert-out approach, which serves as a reference approach to sensitivity analysis in criteria weighting [35]. The basic idea of this approach is that the weight calculation is repeated several times, with one expert omitted from the expert rating matrix in each iteration. In this way, it is examined how much the final criteria weights and rankings change when an individual expert is removed from the evaluation procedure. If the total number of experts is m, the sensitivity analysis is conducted through m iterations. In the first iteration, Expert 1 is omitted, in the second iteration Expert 2, and so on up to Expert m. For each iteration, the weight calculation steps are repeated using the remaining experts, resulting in new criteria weights $w_{i_{(-j)}}$.

The stability of the numerical weight values is examined by comparing the baseline weight vector, obtained on the basis of all experts, with the weight vector obtained after omitting an individual expert. For this purpose, Pearson's correlation coefficient is used to examine the linear agreement between the baseline and iteratively obtained criteria weights.

In addition to weight stability, the stability of the criteria rankings is also examined. After each iteration, the criteria are ranked again according to the calculated weights, and the resulting ranking is compared with the baseline ranking formed on the basis of all experts. Spearman's rank correlation coefficient is used for this examination, indicating the degree of agreement between the baseline and iteratively obtained ordering of the criteria.

If the values of Pearson's and Spearman's coefficients are high in all or most iterations, the final criteria weights and rankings are considered a stable result of group expert evaluation. However, if low values of the correlation coefficients occur in certain iterations, the results are interpreted as sensitive to the ratings of individual experts. In that case, it is necessary to further analyse the structure of the expert ratings, emphasize the limitation of the model or, where possible, include a larger number of experts and repeat the evaluation procedure.

In interpreting the results of the sensitivity analysis, similarity or agreement coefficient values from 0,80 to 0,89 may be considered an indication of high stability, while values of 0,90 and above indicate very high or excellent stability. These limits should be regarded as indicative thresholds.

In addition to the correlation-based examination of weight and ranking stability, the sensitivity analysis may be supplemented by examining changes in the weights of individual criteria. This indicator is used to identify criteria whose weights are particularly sensitive to the structure of the expert group. In this case, the maximum change in weight after omitting an individual expert may be calculated for each criterion:

$$\Delta w_i^{max} = max \left| w_i - w_{i_{(-j)}} \right| \quad (12)$$

where w_i is the criterion weight obtained on the basis of all experts, and $w_{i_{(-j)}}$ is the weight of the same criterion after omitting one expert. A higher value of this indicator shows that the assessment of the importance of the criterion under consideration is more dependent on individual expert ratings, which means that such a criterion requires more cautious interpretation or additional expert verification.

To facilitate the practical application of the proposed model, an implementation-ready Microsoft Excel-based ECSW tool was developed and is available as supplementary material on the article webpage. The user interface of the tool is shown in Figure 2.

ECSW IMPLEMENTATION TOOL

Setup, user instructions, and input model parameters

PURPOSE OF THE TOOL

This Microsoft Excel-based tool supports the implementation of the Expert Criteria Selection and Weighting (ECSW) model. It guides the user through expert-based criteria evaluation, statistical criteria selection, informativeness assessment, criteria weighting, and sensitivity analysis.

HOW TO USE THE TOOL

1. Configure the model – Enter the number of criteria, number of experts, rating scale, and review default model parameters on this sheet
2. Enter expert ratings – Define criterion labels and enter expert ratings on the RATINGS sheet
3. Review criteria selection – Examine criteria identified as potential candidates for elimination, review the statistical results, and indicate whether substantive reasons justify their retention
4. Review final weights – Examine the retained criteria set, informativeness assessment, and final normalized criteria weights
5. Perform sensitivity analysis – Review leave-one-expert-out weights, rankings, and Pearson and Spearman correlations

MODEL SETUP

Parameter	User input	Description
Number of criteria (n)	**24**	Initial number of criteria to be evaluated
Number of experts (m)	**10**	Number of experts providing ratings
Minimum rating	**1**	Lowest value of the numerical rating scale
Maximum rating	**6**	Highest value of the numerical rating scale
Minimum panel size	**10**	Minimum number of experts required for statistical criteria selection (recommended 10). If the expert panel is smaller, the statistical elimination is skipped (see Step 4)
Statistical significance level (α)	**0.05**	Used in the one-sided statistical test for candidate criteria, in the Holm–Bonferroni correction, and used in Sensitivity analysis (see Step 4 and Step 8) - (recommended 0.05)
Informativeness threshold (in %)	**5**	Indicative percentage threshold used to assess whether the retained expert ratings are sufficiently differentiated to justify unequal weighting (see Step 6) - (recommended 5)

MODEL STATUS

Criteria / Experts	24 / 10
Rating scale	1 - 6
Statistical selection	Available
Significance level	0.05
Inf. Threshold (%)	5

The expert panel meets the minimum recommended size for formal statistical criteria selection. Candidate criteria will be evaluated using the statistical procedure defined in the ECSW model.

TOOL ACTIVATION REQUIREMENT

This tool uses VBA macros and requires macros to be enabled in Microsoft Excel. If prompted, select Enable Content or Enable Macros when opening the file (if macros remain blocked, close Excel, right-click the file in Windows Explorer, select Properties, check Unblock, and reopen the file again)

Figure 2. ECSW implementation tool interface.

4. Verification of the Proposed ECSW Model

To examine the applicability of the proposed model, this chapter uses a numerical example based on real empirical expert-rating data from the study by Berković et al. [36]. The verification is component-wise and preliminary. The published example with five experts is used to examine the weighting procedure and sensitivity to expert-panel composition, while the criteria-selection component is additionally examined using a separate example with a larger expert panel. The comparison with the Ordinal Priority Approach (OPA) results is interpreted as evidence of convergence between the obtained weights and rankings rather than as comprehensive validation of the ECSW model.

The cited study considered the problem of determining the importance of criteria in the context of introducing energy-efficient electric vehicles on optimal routes within the transport chain for municipal waste collection and removal. In the adopted numerical example, the criteria weights were determined using the subjective OPA model [12], which allows comparison with the results obtained using the proposed model. In this numerical example, the steps of the proposed model are carried out using the adopted expert rating matrix. However, due to the limited number of experts in the available example, the statistical elimination of criteria from Step 4 is not conducted. For this reason, the statistical criteria selection step in this example has an exclusively interpretative role, while all criteria are retained in the subsequent calculation procedure. The results of the model application are presented below through the eight previously defined steps.

Step 1 and **Step 2** were fully adopted from study [36]. In the original study, a set of 24 criteria relevant to the research problem was defined. Five experts participated in the evaluation procedure and assigned ratings to each criterion according to the defined scale. The expert-rating matrix used for the application of the ECSW model is presented in Table A1 in the appendix.

In **Step 3** the average level of importance assigned by the experts to the criteria in the observed problem was calculated. The overall mean value of the expert ratings $\overline{AE}$, calculated according to Equation (2):

$$\overline{AE} = \frac{1}{n*m}\sum_{i=1}^{n}\sum_{j=1}^{m} AE_{ij} = \frac{3+5+3+\cdots+2+3+1}{24*5} = \frac{333}{120} = 2{,}775$$

After that, the average expert rating, $\overline{AE_i}$ was calculated separately for each criterion according to Equation (3) (Criterion C1 is used as an illustrative example):

$$\overline{AE_1} = \frac{1}{m}\sum_{j=1}^{m} AE_{1j} = \frac{3+5+3+1+5}{5} = \frac{17}{5} = 3{,}4$$

The relative importance factor of C1, was then determined according to Equation (4):

$$MJ_1 = \frac{\overline{AE_1}}{\overline{AE}} = \frac{3{,}4}{2{,}775} = 1{,}225$$

The relative percentage deviation of C1 from the overall mean was calculated according to Equation (6):

$$RD_1 = (MJ_1 - 1) * 100 = (1.225 - 1) * 100 = 22{,}5\ \%$$

The same calculation procedure was applied to all remaining criteria, and the resulting values are presented in Table 3.

Step 4 was not conducted in full in this numerical example. Since five experts participated in the adopted example, the number of expert ratings is not sufficient for the methodologically reliable application of the statistical selection of criteria. Accordingly, in **Step 5,** the initial set of criteria is equal to the final set of criteria (n = n' and $\overline{AE} = \overline{AE'}$). Although the statistical criteria-selection procedure could not be fully applied in the main numerical example because of the limited number of experts, an additional verification of the statistical criteria-selection component was conducted using real empirical expert-rating data from study [25], in which 14 experts evaluated a set of eight criteria. Expert evaluation is provided in Table A2 in the appendix.

In the cited study, the criteria were selected using the fuzzy Delphi model. In the additional validation, three criteria (C4, C5 and C6) had a negative relative deviation (-12,1, -6 and -1,3) and were therefore treated as candidate criteria for potential elimination. Criterion C4 is used to illustrate the calculation of the expert-specific deviation defined in Equation (7). For Expert E1, the rating of C4 was $AE_{41} = 4$. The mean rating assigned by E1 to all remaining seven criteria was $\overline{AE}_{-41} = 4{,}714$. The deviation for C4 and Expert E1 δ_{41} was therefore:

$$\delta_{41} = AE_{41} - \overline{AE}_{-41} = AE_{41} - \frac{1}{n-1}\sum_{\substack{k=1 \\ k\neq i}}^{n} AE_{k1} = 4 - 4{,}714 = -0{,}714$$

After identifying the criteria with a negative relative importance factor, the procedure proceeds to the statistical testing stage. The statistical analysis was conducted in Minitab using a one-sided one-sample t-test applied to the expert-specific deviations calculated for each candidate criterion. The one-sided one-sample t-test applied to the deviations in expert ratings showed that C4 had a mean deviation of −0,643, with t = −3,202 and p = 0,003, while C5 had a mean deviation of −0,316, with t = −2,003 and p = 0,033. Both criteria were statistically significant before correction. In contrast, C6 had a mean deviation of −0,071, with t = −0,422 and p = 0,340, indicating that its negative deviation was not statistically significant.

Since three candidate criteria were tested simultaneously, the obtained p-values were subsequently evaluated using the Holm–Bonferroni correction in order to control the family-wise Type I error rate. The p-values were first ordered from the smallest to the largest: 0,003 for C4, 0,033 for C5, and 0,340 for C6. The smallest p-value was compared with the first Holm threshold, calculated as (0,05/3=0,0167). Since (0,003<0,0167), C4 remained statistically significant after correction. The second p-value was then compared with the next threshold, (0,05/2=0,0250). Since (0,033>0,0250), C5 did not remain statistically significant after correction. Accordingly, C6 also remained non-significant. Therefore, after the Holm–Bonferroni procedure, only C4 satisfied the statistical condition for potential elimination.

However, statistical significance alone does not automatically result in criterion removal within the ECSW model. As defined in the methodological framework, the final decision must also consider whether there is a clear theoretical, legal, regulatory, or practical justification for retaining the criterion. Therefore, in an actual application, criterion C4 should additionally be examined in relation to the substantive context of the decision problem. If no such justification for retaining C4 exists, all three elimination conditions are satisfied and the criterion may be removed from the final set. In the present verification example, this conclusion is also consistent with the original fuzzy Delphi analysis, in which C4 was eliminated.

Table 3. Calculation of the relative importance, stability, and final criteria weights according to the proposed ECSW model.

Criterion	$\overline{AE_i}$	MJ_i	RD_i (%)	BT_i	s_i	CV_i	S_i	FK_i	w_i	Rank
C1	3,4	1,225	22,5	0,041667	1,673	0,492	0,670	0,151	0,0476	3
C2	2,4	0,865	-13,5	0,041667	1,517	0,632	0,613	-0,083	0,0380	18
C3	3	1,081	8,1	0,041667	1,414	0,471	0,680	0,055	0,0437	8
C4	2,4	0,865	-13,5	0,041667	0,894	0,373	0,729	-0,098	0,0373	22
C5	4	1,441	44,1	0,041667	0,707	0,177	0,850	0,375	0,0569	1
C6	2,8	1,009	0,9	0,041667	1,304	0,466	0,682	0,006	0,0416	10
C7	2,6	0,937	-6,3	0,041667	1,140	0,439	0,695	-0,044	0,0396	14
C8	3,2	1,153	15,3	0,041667	1,483	0,464	0,683	0,105	0,0457	4
C9	2,6	0,937	-6,3	0,041667	1,140	0,439	0,695	-0,044	0,0396	14
C10	3,4	1,225	22,5	0,041667	1,517	0,446	0,692	0,156	0,0478	2
C11	2,4	0,865	-13,5	0,041667	1,140	0,475	0,678	-0,092	0,0376	21
C12	3	1,081	8,1	0,041667	1,581	0,527	0,655	0,053	0,0436	9
C13	2	0,721	-27,9	0,041667	1,225	0,612	0,620	-0,173	0,0342	24
C14	2,4	0,865	-13,5	0,041667	1,673	0,697	0,589	-0,080	0,0381	16
C15	2,8	1,009	0,9	0,041667	2,049	0,732	0,577	0,005	0,0416	13
C16	2,4	0,865	-13,5	0,041667	1,342	0,559	0,641	-0,087	0,0378	19
C17	2,4	0,865	-13,5	0,041667	1,342	0,559	0,641	-0,087	0,0378	19
C18	3,2	1,153	15,3	0,041667	1,483	0,464	0,683	0,105	0,0457	4
C19	2,8	1,009	0,9	0,041667	1,789	0,639	0,610	0,005	0,0416	12
C20	2,8	1,009	0,9	0,041667	1,304	0,466	0,682	0,006	0,0416	10
C21	3	1,081	8,1	0,041667	1,225	0,408	0,710	0,058	0,0438	7
C22	3,2	1,153	15,3	0,041667	2,049	0,640	0,610	0,093	0,0453	6
C23	2	0,721	-27,9	0,041667	1,732	0,866	0,536	-0,150	0,0352	23
C24	2,4	0,865	-13,5	0,041667	1,673	0,697	0,589	-0,080	0,0381	16

Note: Ranks were determined using the unrounded weight values, therefore, criteria displaying identical rounded weights may have different ranks.

In **Step 6**, it was assessed whether the average expert ratings contained sufficiently pronounced differences to continue the differentiated weighting procedure. The assessment was based on the relative deviation RD_i, determined according to Equation (6), with the values of this indicator presented in Table 3. For example, the relative deviation of C1 was 22,5%, C5 exhibited a relative deviation of 44,1%, C13 and C23 exhibited deviations of −27,9%, while C10 exhibited a deviation of 22,5%. Overall, 15 of the 24 criteria exhibited an absolute relative deviation greater than the indicative threshold of 10%. Therefore, the deviations were not concentrated around zero, and the expert ratings were considered sufficiently informative to justify differentiated criteria weighting.

In **Step 7**, the base weight BT_i, was first calculated for all criteria according to Equation (8):

$$BT_i = \frac{1}{n'} = \frac{1}{24} = 0{,}041667$$

Criterion C1 is again used as an illustrative example of the complete weight calculation. The sample standard deviation of the ratings 3, 5, 3, 1, and 5 was s_1 = 1,673, and the coefficient of variation was CV_1 = 0,492. After that, the stability factor S_i, was determined for each criterion according to Equation (10):

$$S_1 = \frac{1}{1 + CV_1} = \frac{1}{1 + \frac{s_i}{\overline{AE}'_i}} = \frac{1}{1 + 0{,}492} = 0{,}67$$

The base-weight adjustment factor FK_i was calculated according to Equation (9) by combining the relative importance factor and the stability factor:

$$FK_1 = ({MJ_1}' - 1) * S_1 = (1{,}225 - 1) * 0{,}67 = 0{,}151$$

The final normalized criterion weight w_i was calculated according to Equation (11), ensuring that the sum of all final weights was equal to 1:

$$w_1 = \frac{BT_i(1 + FK_i)}{\sum_{i=1}^{n'} BT_i(1 + FK_i)}$$
$$= \frac{0{,}041667(1 + 0{,}151)}{0{,}041667(1 + 0{,}151) + \cdots + 0{,}041667(1 - 0{,}08)}$$
$$= 0{,}0476$$

The same calculation procedure was applied to all remaining criteria, and the resulting values are presented in Table 3.

In **Step 8**, the stability of the results was examined using the leave-one-expert-out approach. Starting from the baseline solution obtained using all five experts, the complete ECSW weighting procedure was repeated five times, with one expert omitted in each iteration. This produced five additional weight vectors and their corresponding criterion rankings. For illustration, when Expert E1 was omitted, the ECSW model was recalculated using only Experts E2–E5. For criterion C1, the baseline weight was w_1 = 0,0476, while after omitting E1 the recalculated weight was w_{1-1} = 0,0497. Its ranking consequently changed from rank 3 in the baseline solution to rank 2. The same recalculation was performed for all 24 criteria. The complete baseline and leave-one-expert-out weight and rank vectors are provided in Table A3 in the appendix.

The numerical stability of the criteria weights was assessed by calculating Pearson's correlation coefficient and ranking stability was assessed using Spearman's rank correlation coefficient. The results presented in Table 4 show that all correlations are positive and statistically significant, but that weight stability and ranking stability cannot be interpreted in the same way.

Table 4. Results of the sensitivity analysis using the leave-one-expert-out approach.

Omitted expert	Pearson (r)	p	Spearman (ρ)	p
E1	0,913	<0,001	0,868	<0,001
E2	0,894	<0,001	0,854	<0,001
E3	0,825	<0,001	0,806	<0,001
E4	0,833	<0,001	0,821	<0,001
E5	0,732	<0,001	0,619	0,001

Pearson's correlation coefficient for the criteria weights ranges from 0,732 to 0,913, indicating moderate to very high stability of the numerical weight values after individual experts were omitted. On the other hand, Spearman's rank correlation coefficient ranges from 0,619 to 0,868, showing that the rankings are more sensitive to changes in the composition of the expert group than the weights themselves. The lowest values of both coefficients were obtained when Expert E5 was omitted, with a particularly pronounced decrease in Spearman's rank correlation coefficient. This indicates that Expert E5 had the strongest individual influence on the resulting criteria weights and rankings within the analyzed expert panel.

Overall, the numerical weights demonstrate moderate to very high stability, whereas the rankings are more sensitive to the omission of individual experts. In particular, the omission of Expert E5 produces a noticeable reduction in both weight and rank agreement.

5. Discussion of the Results

The application of the proposed model to the adopted numerical example showed that the input expert ratings do not form a homogeneous set in which all criteria are rated at approximately the same level of importance. The relative deviations of the average ratings from the overall mean show that 15 of the 24 criteria exceed the indicative threshold of 10%, either in a positive or negative direction. This structure of ratings indicates that the differences between the criteria are sufficiently pronounced to continue with differentiated weighting rather than applying equal weights to all criteria.

The results obtained from the numerical example show that the application of the proposed model produced a differentiated set of criteria weights, with the highest weights assigned to criteria C5, C10, C1, C8, and C18. Criterion C5 has the highest final weight (0,0569) and occupies first place in the final ranking. The lowest weights were assigned to criteria C13, C23, C4, and C11. These results show that the proposed model clearly distinguishes between criteria that are above and below the average estimated level of importance.

For the comparative verification of the weighting component, the weights and rankings obtained using the proposed ECSW model were compared with the results of the OPA model presented in study [36]. The comparative results are presented in Table 5.

Table 5. Comparison of criteria weights and rankings according to the OPA model and the proposed ECSW model.

Position	OPA criterion	OPA weight	OPA rank	ECSW criterion	ECSW weight	ECSW rank
1	C5	0,0732	1	C5	0,0569	1
2	C1	0,0563	2	C10	0,0478	2
3	C10	0,0505	3	C1	0,0476	3
4	C3	0,0488	4	C8	0,0457	4
5	C18	0,0469	5	C18	0,0457	4
6	C8	0,0469	5	C22	0,0453	6
7	C22	0,0466	7	C21	0,0438	7
8	C21	0,0437	8	C3	0,0437	8
9	C12	0,0437	8	C12	0,0436	9
10	C6	0,0410	10	C6	0,0416	10
11	C20	0,0410	10	C20	0,0416	10
12	C19	0,0410	10	C19	0,0416	12
13	C15	0,0410	10	C15	0,0416	13
14	C2	0,0407	14	C7	0,0396	14
15	C4	0,0407	14	C9	0,0396	14
16	C7	0,0386	16	C14	0,0381	16
17	C9	0,0386	17	C24	0,0381	16
18	C16	0,0364	18	C2	0,0380	18
19	C11	0,0364	19	C16	0,0378	19
20	C17	0,0364	20	C17	0,0378	19
21	C24	0,0361	21	C11	0,0376	21
22	C14	0,0361	22	C4	0,0373	22
23	C23	0,0328	23	C23	0,0352	23
24	C13	0,0328	23	C13	0,0342	24

Note: Ranks were determined using the unrounded weight values, therefore, criteria displaying identical rounded weights may have different ranks.

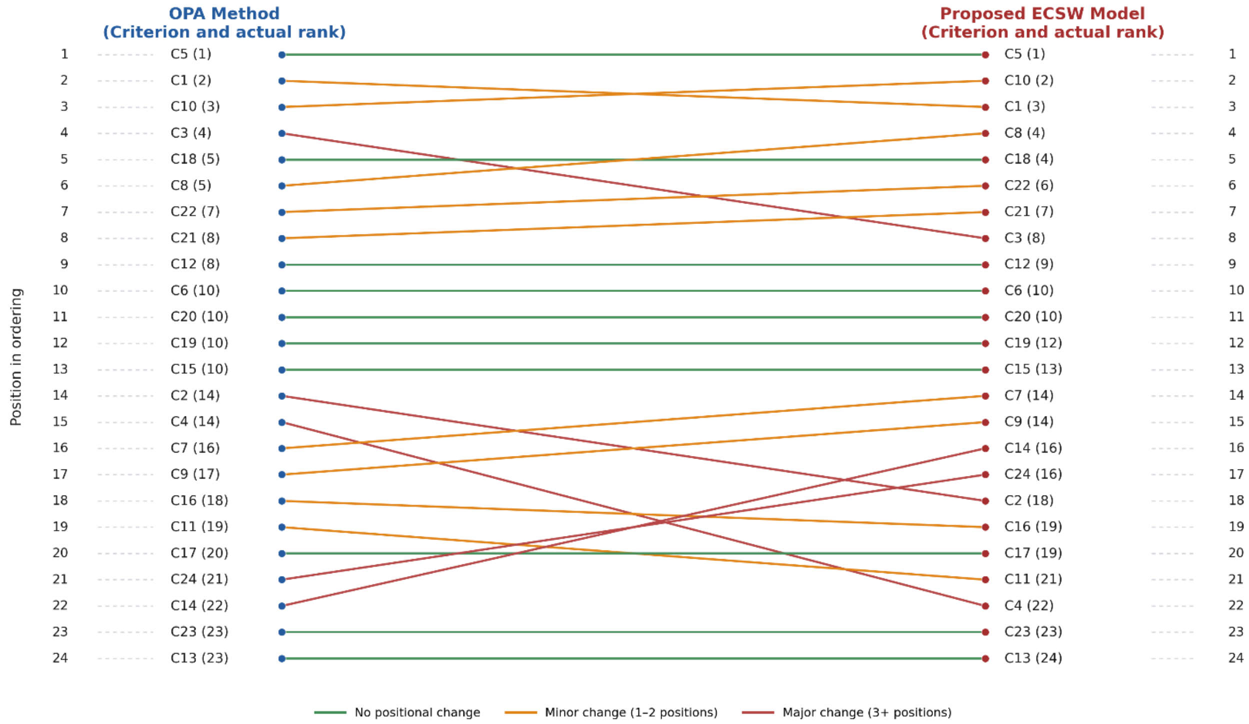

Figure 3. Comparison of criterion rankings obtained using the OPA method and the ECSW model.

The table provides a direct comparison of criteria weights and rankings according to both models. The column "Position" denotes the sequential display order from 1 to 24, whereas the OPA rank and ECSW rank columns show the actual rank values obtained within each model. These values are not always identical because criteria with equal weights are assigned the same rank. For example, in the OPA method, criteria C6, C20, C19, and C15 share rank 10, although they are displayed in consecutive positions from 10 to 13. An additional graphical comparison of the ranking results is provided in Figure 3. The figure should be interpreted as a comparison of the ordered lists of criteria, while the values shown in parentheses indicate the actual method-specific ranks. The results show that identical rank values were obtained for 10 of the 24 criteria. The most important criterion, C5, occupies first place according to both models, while criteria C1 and C10 only exchange the second and third positions. The largest ranking deviations were recorded for a smaller number of criteria, particularly those located in the middle part of the ranking. Such changes are expected because their weights are very close to one another, meaning that even a small change in the calculation procedure may lead to a change in ranking position. On the other hand, the criteria with the lowest weights remain stably positioned.

In addition to the comparison of ranking positions, Figure 4 provides a visual representation of the agreement between the criteria weights obtained using the original OPA model and the proposed ECSW model. The criteria weights are very close throughout most of the ranking. This confirms that the differences in individual ranking positions generally do not arise from large differences in the assigned weights, but from the fact that a larger number of criteria have similar weight values.

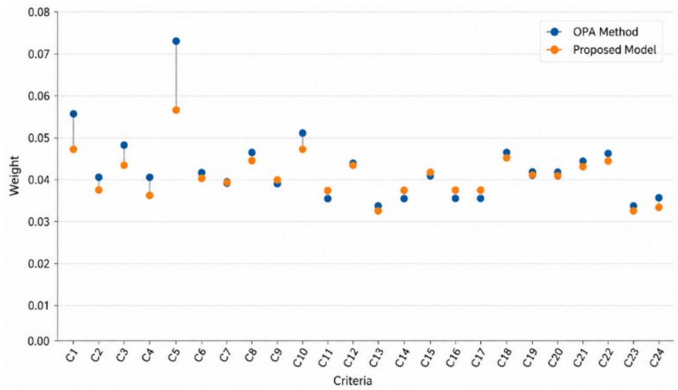

Figure 4. Criteria weights according to the original OPA model and the proposed ECSW model.

For an additional examination of the agreement between the results, Pearson's correlation coefficient for the weights and Spearman's rank correlation coefficient were calculated between the original OPA model and the proposed model. Pearson's correlation coefficient for the weights was $r = 0{,}956$, showing a very strong linear relationship between the weight values obtained using the two models. Spearman's rank correlation coefficient was $\rho = 0{,}929$, confirming a high degree of agreement between the criteria orderings. Both correlations were statistically significant at $p < 0{,}001$, although their interpretation in this context is based primarily on the magnitude of the coefficients and the practical degree of agreement between the results, rather than solely on statistical significance.

Although the two correlation coefficients indicate strong overall agreement, they do not imply identical weight values or ranking positions. Pearson's coefficient primarily reflects the similarity of the overall weight pattern, whereas Spearman's coefficient reflects the similarity of the ordering. Therefore, the correlation results are interpreted together with the overlap of the ten highest-ranked criteria and the observed criterion-level differences.

The strong agreement between the ECSW and OPA results indicates convergence of the obtained weights and rankings, but it does not imply that the two methods are methodologically equivalent or intended for the same decision-making context. OPA derives criteria weights from ordinal preference information and requires the criteria, as well as the participating experts, to be ranked. In contrast, ECSW is based on independent numerical evaluations of individual criteria and does not require experts to establish a complete criteria ranking. This distinction becomes particularly relevant when a large number of criteria is considered.

Furthermore, criteria weighting represents only one component of the ECSW framework. Therefore, the high agreement observed with OPA supports the convergence of the weighting results, while the methodological contribution of ECSW lies in providing a broader expert-based framework for criteria evaluation, selection, weighting, and stability assessment.

6. Limitations and Future Research

The limitations of this study primarily concern the scope of the preliminary verification and the conditions required for applying individual components of the ECSW model. The statistical criteria-selection component is particularly dependent on the size of the expert panel, since an adequately sized panel is required for a more reliable application of statistical tests. With smaller expert panels, this component should have a supporting and indicative role, and the formal statistical exclusion of criteria is not recommended. Moreover, the one-sample t-test assumes independence across experts and requires examination of the distribution of the expert-specific centred deviations and potentially influential observations, particularly when the expert panel is small.

The application of the model also requires a clearly defined rating scale and a justified treatment of numerical expert ratings as interval data. Equal distances between adjacent scale values must therefore be reasonably assumed in each application. In addition, the coefficient of variation used in the stability factor represents the relative dispersion of ratings and should not be interpreted as a formal measure of inter-rater agreement or reliability. Future applications may supplement it with dedicated agreement measures when such an assessment is required.

A further limitation concerns potential scale dependence. Although the ECSW model can be applied using different interval rating scales, equivalent judgments represented on scales with different origins may produce changes in relative deviations, coefficients of variation, and final weight adjustments. The robustness of the results to alternative scale specifications should therefore be examined in future applications.

The overall mean used by the model represents an internal, set-dependent reference point rather than an absolute threshold of criterion relevance. Consequently, the relative position of each criterion depends partly on the composition of the initial criteria set, which must be theoretically justified, sufficiently comprehensive, and appropriately aligned with the decision problem.

Another limitation concerns the researcher-defined informativeness threshold. Its value may affect the interpretation of whether the retained expert ratings provide sufficient differentiation for unequal weighting.

The threshold should therefore be treated as an indicative and context-dependent guideline rather than as a universal statistical cut-off, and its interpretation should consider the overall distribution of relative deviations rather than the deviation of a single criterion.

Finally, the numerical examination was based on a limited number of examples from the literature. The weighting component and its sensitivity to expert-panel composition were examined using one published example, while the statistical criteria-selection component was additionally examined using a separate example with a larger expert panel. These results should therefore be interpreted as preliminary component-wise verification rather than as comprehensive validation across different MCDM domains.

Future research should focus on the following directions:

- further evaluating the ECSW model across different MCDM application domains and on a larger number of empirical and numerical datasets in order to assess its robustness, transferability, and stability under different decision contexts;
- applying the model with larger and more diverse expert panels, which would enable a more reliable implementation of the statistical criteria selection procedure and a more detailed examination of the influence of panel size and composition on the final results;
- comparing the proposed model with a broader range of subjective, objective, and integrated criteria-weighting methods, rather than only with the OPA model, in order to clarify the conditions under which ECSW produces convergent or divergent weights and rankings and to define its methodological position more precisely;
- examining alternative statistical procedures for criteria assessment and selection, including non-parametric and robust methods, to determine whether different testing approaches lead to comparable sets of retained criteria;
- introducing expert weights based on experience, competencies, field-specific expertise, or the demonstrated reliability of previous assessments, thereby enabling a more refined aggregation of group judgments;
- investigating alternative approaches for defining the informativeness threshold and adapting it to the decision context, the number of criteria, the size of the expert panel, and the properties of the rating scale.

Addressing these directions would provide a stronger basis for evaluating the general applicability, robustness, and methodological advantages of the proposed model.

7. Conclusion

This paper proposes the Expert Criteria Selection and Weighting (ECSW) model, an integrated framework for the expert-based evaluation, selection, and weighting of criteria in multi-criteria decision-making. Unlike approaches that require pairwise comparisons or prior criteria ranking, the proposed model is based on independent numerical ratings assigned by experts to each criterion with respect to the decision problem. This simplifies the collection of expert ratings, particularly in cases involving a large number of criteria, where conventional comparisons or rankings may impose a considerable cognitive burden on experts. The model is particularly suitable for decision problems in which a broader set of criteria must be defined and evaluated on the basis of expert knowledge before the alternatives or a complete decision matrix required for objective weighting methods are available.

The methodological contribution of the study lies in integrating several components within a single procedure: determining the relative position of criteria with respect to the overall mean of expert ratings, statistically selecting criteria, assessing the informativeness of the retained ratings, calculating final weights based on both relative importance and rating stability, and analysing the sensitivity of the resulting weights and rankings. In this way, the model evaluates criteria within the overall structure of expert ratings while also accounting for the relative dispersion of ratings across experts. To support practical application and reproducibility, the ECSW model is accompanied by a dedicated implementation tool, available as supplementary material on the article webpage.

The main components of the proposed model were preliminarily examined using numerical examples from the literature. The weighting procedure and its sensitivity to expert-panel composition were examined using a sustainable transport decision problem involving 24 criteria and five experts. The ratings provided sufficient differentiation for unequal weighting, as 15 of the 24 criteria exhibited absolute relative deviations exceeding the indicative threshold of 10%, and the resulting weights clearly differentiated the criteria. The statistical criteria-selection component was additionally examined using a separate example involving 14 experts, in which the ECSW procedure produced the same criterion-exclusion decision as the original fuzzy Delphi approach.

The leave-one-expert-out sensitivity analysis indicated moderate to very high stability of the numerical weights, while the criteria rankings were more sensitive to changes in the expert panel. Pearson's correlation coefficients between the baseline weights and the weights obtained after omitting individual experts ranged from 0.732 to 0.913, whereas Spearman's rank correlation coefficients ranged from 0.619 to 0.868. The lowest values of both coefficients were obtained when Expert E5 was omitted, indicating that this expert had the strongest individual influence on the resulting weights and rankings within the analysed panel. Nevertheless, all correlations were positive and statistically significant.

The weighting component was also comparatively examined against the original OPA model. Pearson's correlation coefficient for the criteria weights was 0.956, while Spearman's rank correlation coefficient was 0.929, with both correlations being statistically significant. The two models also identified the same ten highest-ranked criteria, although some differences occurred in their internal ordering.

Overall, the results indicate that the ECSW model provides a transparent and computationally efficient framework for expert-based criteria evaluation, selection, and weighting, particularly in decision problems involving a large number of criteria and no complete decision matrix. As discussed in the preceding section, the current findings should be interpreted in light of the model's dependence on the expert-panel size, rating-scale assumptions, the composition of the initial criteria set, and researcher-defined informativeness thresholds. Broader empirical examination across different decision domains, rating scales, and expert-panel configurations is therefore required before the model's general applicability and comparative advantages can be established more comprehensively.

Competing Interest Statement

Alem Čolaković, one of the authors of this manuscript, currently serves as Editor-in-Chief of Science, Engineering and Technology. He was not involved in the peer-review process, editorial evaluation, or final decision regarding this article. The other authors declare no competing interests.

Data Availability Statement

All data used in this study are included in the article and its Appendix. The expert-rating datasets used for the numerical verification of the ECSW model were adopted from previously published studies cited in the manuscript.

Software Availability Statement

An implementation-ready ECSW tool is available as supplementary material on the article webpage. The tool supports reproduction of the calculations presented in this study and enables the complete ECSW procedure to be readily applied to new expert-rating datasets.

Statement on the Ethical Use of AI Tools

During the preparation of this article, the authors used Grammarly solely for language editing, grammar correction, and improving overall readability. All content has been reviewed and revised by the authors. The authors take full responsibility for the accuracy and originality of the work.

References

[1] G. O. Odu, "Weighting methods for multi-criteria decision making technique," *J. Appl. Sci. Environ. Manag.*, vol. 23, no. 8, pp. 1449–1457, Sep. 2019, doi: 10.4314/jasem.v23i8.7.

[2] B. Ayan, S. Abacıoğlu, and M. P. Basilio, "A Comprehensive Review of the Novel Weighting Methods for Multi-Criteria Decision-Making," *Information*, vol. 14, no. 5, Art. no. 285, May 2023, doi: 10.3390/info14050285.

[3] A. Sarkar and S. S. Goswami, "A Comprehensive Review: The Novel Weighting Methods For Multi-Criteria Decision-Making (MCDM)," *Spectr. Oper. Res.*, pp. 1–26, 2026, doi: 10.31181/sor202781.

[4] B. Kizielewicz, T. Tomczyk, M. Gandor, and W. Sałabun, "Subjective weight determination methods in multi-criteria decision-making: a systematic review," *Procedia Comput. Sci.*, vol. 246, pp. 5396–5407, Jan. 2024, doi: 10.1016/j.procs.2024.09.673.

[5] I. Mukhametzyanov, "Specific character of objective methods for determining weights of criteria in MCDM problems: Entropy, CRITIC and SD," *Decis. Mak. Appl. Manag. Eng.*, vol. 4, no. 2, pp. 76–105, Jun. 2021, doi: 10.31181/dmame210402076i.

[6] S. Zakeri, D. Konstantas, P. Chatterjee, and E. K. Zavadskas, "Soft cluster-rectangle method for eliciting criteria weights in multi-criteria decision-making," *Sci. Rep.*, vol. 15, no. 1, Art. no. 284, Jan. 2025, doi: 10.1038/s41598-024-81027-4.

[7] P. Grošelj and G. Dolinar, "A comparative analysis of MCDM methods based on pairwise comparison: AHP, BWM and FUCOM," in *XXI International May Conference on Strategic Management – IMCSM25 Proceedings - zbornik radova*, University of Belgrade - Technical Faculty in Bor, Bor, 2025, pp. 256–266. doi: 10.5937/IMCSM25256G.

[8] O. Uzhga-Rebrov and G. Kuļešova, "A Review and Comparative Analysis of Methods for Determining Criteria Weights in MCDM Tasks," *Inf. Technol. Manag. Sci.*, vol. 26, no. 1, pp. 35–40, Nov. 2023, doi: 10.7250/itms-2023-0005.

[9] M. Tavana, M. Soltanifar, and F. J. Santos-Arteaga, "Analytical hierarchy process: revolution and evolution," *Ann. Oper. Res.*, vol. 326, no. 2, pp. 879–907, Jul. 2023, doi: 10.1007/s10479-021-04432-2.

[10] J. Rezaei, "Best-worst multi-criteria decision-making method," *Omega*, vol. 53, pp. 49–57, Jun. 2015, doi: 10.1016/j.omega.2014.11.009.

[11] D. Pamučar, Ž. Stević, and S. Sremac, "A New Model for Determining Weight Coefficients of Criteria in MCDM Models: Full Consistency Method (FUCOM)," *Symmetry*, vol. 10, no. 9, Art. no. 393, Sep. 2018, doi: 10.3390/sym10090393.

[12] Y. Ataei, A. Mahmoudi, M. R. Feylizadeh, and D.-F. Li, "Ordinal Priority Approach (OPA) in Multiple Attribute Decision-Making," *Appl. Soft Comput.*, vol. 86, Art. no. 105893, Jan. 2020, doi: 10.1016/j.asoc.2019.105893.

[13] R. T. Mohammed *et al.*, "Determining Importance of Many-Objective Optimisation Competitive Algorithms Evaluation Criteria Based on a Novel Fuzzy-Weighted Zero-Inconsistency Method," *Int. J. Inf. Technol. Decis. Mak.*, vol. 21, no. 01, pp. 195–241, Jan. 2022, doi: 10.1142/S0219622021500140.

[14] S. Bošković, S. Jovčić, V. Simic, L. Švadlenka, M. Dobrodolac, and N. Bacanin, "A NEW CRITERIA IMPORTANCE ASSESSMENT (CIMAS) METHOD IN MULTI-CRITERIA GROUP DECISION-MAKING: CRITERIA EVALUATION FOR SUPPLIER SELECTION," *Facta Univ. Ser. Mech. Eng.*, vol. 23, no. 2, pp. 335–349, Aug. 2025, doi: 10.22190/FUME230730050B.

[15] I. Vinogradova-Zinkevič, "Centroidous Method for Determining Objective Weights," *Mathematics*, vol. 12, no. 14, Art. no. 2269, Jul. 2024, doi: 10.3390/math12142269.

[16] Z. Stevic, A. Ulutas, A. Topal, D. Marinkovic, and S. Cavoski, "A New Objective Method for Determining Criteria Weights in MCDM Models – LOGSTA," *Int. J. Simul. Model.*, vol. 24, no. 4, pp. 589–600, Dec. 2025, doi: 10.2507/IJSIMM24-4-738.

[17] M. S. Saidin, L. S. Lee, S. M. Marjugi, M. Z. Ahmad, and H.-V. Seow, "Fuzzy Method Based on the Removal Effects of Criteria (MEREC) for Determining Objective Weights in Multi-Criteria Decision-Making Problems," *Mathematics*, vol. 11, no. 6, Art. no. 1544, Mar. 2023, doi: 10.3390/math11061544.

[18] H. Sari, Ž. Stević, A. Ulutaş, A. Topal, A. A. Yürüyen, and A. O. Bayrakçıl, "Development of a Novel Integrated MCDM Framework as a Decision-Support Tool for Lathe Selection," *Sci. Eng. Technol.*, vol. 6, no. S1, pp. 56–74, Aug. 2026, doi: 10.54327/set2026/v6.iS1.381.

[19] M. Žižović, B. Miljković, and D. Marinković, "Objective methods for determining criteria weight coefficients: A modification of the CRITIC method," *Decis. Mak. Appl. Manag. Eng.*, vol. 3, no. 2, pp. 149–161, Oct. 2020, doi: 10.31181/dmame2003149z.

[20] T. V. Dua, D. V. Duc, N. C. Bao, and D. D. Trung, "Integration of objective weighting methods for criteria and MCDM methods: application in material selection," *EUREKA Phys. Eng.*, no. 2, pp. 131–148, Mar. 2024, doi: 10.21303/2461-4262.2024.003171.

[21] N. T. D. Linh, N. H. Son, and D. X. Thao, "Evaluating the Impact of Weighting Methods on the Stability of Scores for Alternatives in Multi-Criteria Decision-Making Problems," *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 1, pp. 19998–20004, Feb. 2025, doi: 10.48084/etasr.9518.

[22] B. Paradowski, A. Shekhovtsov, A. Bączkiewicz, B. Kizielewicz, and W. Sałabun, "Similarity Analysis of Methods for Objective Determination of Weights in Multi-Criteria Decision Support Systems," *Symmetry*, vol. 13, no. 10, Art. no. 1874, Oct. 2021, doi: 10.3390/sym13101874.

[23] J. Wang, S. Setiawansyah, and S. Sumanto, "Comparison of Objective Weighting Methods in SAW and Their Effect on Alternative Ranking Results," *J. Masy. Inform.*, vol. 17, no. 1, pp. 89–112, Feb. 2026, doi: 10.14710/jmasif.17.1.78414.

[24] F. M. Theunissen, C. N. Bezuidenhout, and S. Alam, “Exploring the shortcomings in formal criteria selection for multicriteria decision making based inventory classification models: a systematic review and future directions,” *Int. J. Prod. Res.*, vol. 62, no. 19, pp. 7279–7299, Oct. 2024, doi: 10.1080/00207543.2024.2320680.

[25] A. Kalem, S. Tadić, M. Krstić, N. Čabrić, and N. Branković, “Performance Evaluation of Railway Infrastructure Managers: A Novel Hybrid Fuzzy MCDM Model,” *Mathematics*, vol. 12, no. 10, Art. no. 1590, May 2024, doi: 10.3390/math12101590.

[26] F. Chiclana, J. M. Tapia García, M. J. Del Moral, and E. Herrera-Viedma, “A statistical comparative study of different similarity measures of consensus in group decision making,” *Inf. Sci.*, vol. 221, pp. 110–123, Feb. 2013, doi: 10.1016/j.ins.2012.09.014.

[27] M. Mohammadi, D. A. Tamburri, and J. Rezaei, “Unveiling and Unraveling Aggregation and Dispersion Fallacies in Group MCDM,” *Group Decis. Negot.*, vol. 32, no. 4, pp. 779–806, Aug. 2023, doi: 10.1007/s10726-023-09822-4.

[28] M. Yurdakul and Y. Tansel İÇ, “Application of correlation test to criteria selection for multi criteria decision making (MCDM) models,” *Int. J. Adv. Manuf. Technol.*, vol. 40, no. 3, pp. 403–412, Jan. 2009, doi: 10.1007/s00170-007-1324-1.

[29] A. Ross and V. L. Willson, “One-Sample T-Test,” in *Basic and Advanced Statistical Tests: Writing Results Sections and Creating Tables and Figures*, A. Ross and V. L. Willson, Eds., Rotterdam: SensePublishers, 2017, pp. 9–12. doi: 10.1007/978-94-6351-086-8_2.

[30] O. Menyhart, B. Weltz, and B. Győrffy, “MultipleTesting.com: A tool for life science researchers for multiple hypothesis testing correction,” *PLOS ONE*, vol. 16, no. 6, Art. no. e0245824, Jun. 2021, doi: 10.1371/journal.pone.0245824.

[31] T. K. Kim and J. H. Park, “More about the basic assumptions of t-test: normality and sample size,” *Korean J. Anesthesiol.*, vol. 72, no. 4, pp. 331–335, Aug. 2019, doi: 10.4097/kja.d.18.00292.

[32] “Ultimate Guide to T Tests.” Accessed: Jul. 10, 2026. [Online]. Available: https://www.graphpad.com/guides/the-ultimate-guide-to-t-tests

[33] M. Krzywinski and N. Altman, “Nonparametric tests,” *Nat. Methods*, vol. 11, no. 5, pp. 467–468, May 2014, doi: 10.1038/nmeth.2937.

[34] A. P. King and R. J. Eckersley, “Chapter 6 - Inferential Statistics III: Nonparametric Hypothesis Testing,” in *Statistics for Biomedical Engineers and Scientists*, Academic Press, 2019, pp. 119–145. doi: 10.1016/B978-0-08-102939-8.00015-3.

[35] J. Więckowski and W. Sałabun, “Sensitivity analysis approaches in multi-criteria decision analysis: A systematic review,” *Appl. Soft Comput.*, vol. 148, Art. no. 110915, Nov. 2023, doi: 10.1016/j.asoc.2023.110915.

[36] M. Berković, L. Nikšić, A. Kalem, and S. Džaferović, “Methodological Framework for the Introduction of Energy Efficient Electric Vehicles on Optimal Routes in the Transport Chain of Municipal Waste Collection and Removal,” in *New Technologies, Development and Application IX*, I. Karabegović, A. Kovačević, and S. Mandžuka, Eds., Cham: Springer Nature Switzerland, 2026, pp. 109–130. doi: 10.1007/978-3-032-30107-9_11.

Appendix

Table A1. Expert-rating matrix adopted from study.

E/C	Criteria (C_i)																							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
E1	3	4	3	3	3	3	3	3	3	4	2	2	2	2	2	3	3	3	3	2	5	5	5	5
E2	5	4	3	3	4	4	3	3	3	3	3	3	1	3	1	3	3	3	1	1	3	4	1	1
E3	3	1	3	3	4	4	4	4	4	4	4	1	1	1	1	1	1	1	1	3	3	5	2	2
E4	1	1	1	1	4	1	1	1	1	1	1	4	4	5	5	4	4	4	4	4	2	1	1	3
E5	5	2	5	2	5	2	2	5	2	5	2	5	2	1	5	1	1	5	5	4	2	1	1	1

Table A2. Expert-rating matrix used for verification of the statistical criteria-selection procedure.

E/C	C1	C2	C3	C4	C5	C6	C7	C8
E1	5	5	5	4	4	4	5	5
E2	5	5	5	5	4	4	5	5
E3	5	5	5	3	4	5	5	5
E4	5	5	4	4	5	5	5	4
E5	5	5	4	4	4	5	5	5
E6	5	5	5	3	4	4	5	5
E7	5	5	5	5	5	4	4	5
E8	5	5	5	5	4	5	5	5
E9	5	4	4	4	5	5	5	4
E10	5	4	4	4	4	5	5	5
E11	5	5	5	4	4	4	5	5
E12	5	5	5	5	5	4	4	5
E13	5	5	5	3	4	5	5	5
E14	4	4	4	4	5	5	5	5

Table A3. Leave-one-expert-out weights and rankings.

Criterion	w_{-1}	w_{-2}	w_{-3}	w_{-4}	w_{-5}	R_{-1}	R_{-2}	R_{-3}	R_{-4}	R_{-5}
C1	0,0497	0,0435	0,0477	0,0538	0,0440	2	9	4	3	6
C2	0,0353	0,0346	0,0406	0,0401	0,0393	21	24	13	13	15
C3	0,0447	0,0435	0,0430	0,0483	0,0389	8	9	9	6	19
C4	0,0368	0,0358	0,0354	0,0398	0,0389	20	23	20	15	19
C5	0,0632	0,0564	0,0556	0,0547	0,0550	1	1	1	1	1
C6	0,0422	0,0386	0,0382	0,0454	0,0442	10	14	18	7	4
C7	0,0396	0,0386	0,0354	0,0427	0,0416	14	14	20	9	7
C8	0,0472	0,0460	0,0430	0,0513	0,0416	3	4	9	4	7
C9	0,0396	0,0386	0,0354	0,0427	0,0416	14	14	20	9	7
C10	0,0472	0,0485	0,0454	0,0547	0,0442	3	2	5	1	4
C11	0,0396	0,0363	0,0328	0,0399	0,0391	14	21	24	14	16
C12	0,0472	0,0434	0,0485	0,0401	0,0391	3	11	3	12	16
C13	0,0353	0,0363	0,0359	0,0270	0,0349	21	21	19	24	24
C14	0,0399	0,0371	0,0407	0,0308	0,0416	13	17	12	23	13

Table A3 continued on next page.

Table A3. (continued): Leave-one-expert-out weights and rankings.

Criterion	w_{-1}	w_{-2}	w_{-3}	w_{-4}	w_{-5}	R_{-1}	R_{-2}	R_{-3}	R_{-4}	R_{-5}
C15	0,0443	0,0457	0,0451	0,0363	0,0374	9	7	7	18	21
C16	0,0375	0,0366	0,0406	0,0333	0,0416	18	19	15	21	7
C17	0,0375	0,0366	0,0406	0,0333	0,0416	18	19	15	21	7
C18	0,0472	0,0460	0,0521	0,0424	0,0416	3	4	2	11	7
C19	0,0421	0,0460	0,0454	0,0382	0,0370	11	4	5	16	23
C20	0,0448	0,0468	0,0406	0,0377	0,0391	7	3	13	17	16
C21	0,0392	0,0436	0,0431	0,0452	0,0471	17	8	8	8	3
C22	0,0421	0,0432	0,0407	0,0496	0,0517	11	12	11	5	2
C23	0,0257	0,0371	0,0353	0,0363	0,0374	24	17	23	18	21
C24	0,0322	0,0411	0,0387	0,0363	0,0416	23	13	17	18	13