University of Tehran Press

Industrial Management Journal

Online ISSN: 3115 - 7386

Homepage: https://imj.ut.ac.ir

ED-ELEC: A Robust Multi-Criteria Decision-Making Framework for Evaluating Sustainable Urban Delivery Systems

Morteza Daripour [1], **Ezzatollah Asgharizadeh** ✉ [2], **and Mohammadreza Taghizadeh-Yazdi** [3]

1. PhD Candidate, Faculty of Industrial Management & Technology, College of Management, University of Tehran, Tehran, Iran. E-mail: m.daripour@ut.ac.ir
2. Corresponding author, Prof., Faculty of Industrial Management & Technology, College of Management, University of Tehran, Tehran, Iran. E-mail: asghari@ut.ac.ir
3. Prof., Faculty of Industrial Management & Technology, College of Management, University of Tehran, Tehran, Iran. E-mail: mrtaghizadeh@ut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received October 25, 2025
Received in revised form November 13, 2025
Accepted January 23, 2026
Available online February 20, 2026

Keywords:
Multi-Criteria Decision-Making (MCDM), D-numbers theory, ELECTRE III, sustainable logistics, online food delivery, incomplete information.

ABSTRACT

Objective: The rapid expansion of online food delivery platforms has intensified the need for sustainable and efficient logistics solutions. Evaluating delivery alternatives across economic, environmental, technical, and social dimensions requires robust multi-criteria decision-making (MCDM) tools capable of managing incomplete, uncertain, and conflicting expert judgments. This study develops and validates ED-ELEC, an enhanced MCDM framework that extends D-numbers theory and integrates it with the ELECTRE III outranking method to enable reliable and transparent decision-making under imperfect information.

Methodology: A new integration function for D-numbers is introduced to convert incomplete expert evaluations into crisp values while retaining the effects of incompleteness and conflict. A two-stage validation was performed: first, an illustrative example compared the proposed integration function with existing ones and benchmarked the complete ED-ELEC framework against another hybrid method; second, the framework was applied to evaluate sustainable delivery systems involving gasoline motorcycles, electric motorcycles, and drones.

Results: The proposed integration function achieved balanced sensitivity to uncertainty and better interpretability. ED-ELEC produced robust, transparent, and stable rankings, identifying electric motorcycles as the most sustainable option, followed by drones and gasoline motorcycles.

Conclusion: The ED-ELEC framework effectively combines uncertainty modeling and outranking analysis, providing an adaptive decision-support tool for sustainability evaluation. It enables decision-makers to address incomplete data while improving ranking clarity and robustness in sustainable urban logistics planning.

Cite this article: Daripour, M., Asgharizadeh, E. & Taghizadeh-Yazdi, M., (2026). ED-ELEC: A robust multi-criteria decision-making framework for evaluating sustainable urban delivery systems, ***Industrial Management Journal***, 18(2), 34-73. https://doi.org/10.22059/imj.2026.405611.1008271

Publisher: University of Tehran Press.

DOI: https://doi.org/10.22059/imj.2026.405611.1008271

Introduction

The rapid growth of online food delivery platforms has transformed urban consumption patterns, creating new opportunities and challenges for logistics systems. Recent studies highlight the exponential expansion of this sector, with market values reaching USD 150 billion in 2024 and projected to reach USD 320 billion by 2033 (Lohmann et al., 2024). While these services offer significant convenience to consumers, they also generate notable environmental and social impacts. In many cities, the dominant mode of delivery remains the gasoline-powered motorcycle, which, although cost-efficient and operationally flexible, contributes to air pollution, greenhouse gas emissions, noise, and traffic congestion. Balancing economic efficiency with environmental sustainability has therefore become a key priority, prompting researchers, policymakers, and practitioners to explore alternative delivery modes such as electric motorcycles and unmanned aerial vehicles (Garg et al., 2023; Stolaroff et al., 2018; Yaftiyan et al., 2025).

Evaluating such alternatives is complex because decision-makers must consider multiple, often conflicting criteria across economic, environmental, operational, and social dimensions. Multi-criteria decision-making (MCDM) methods provide a systematic framework for addressing such complexity and have been widely applied in sustainability and supply chain analysis (Mardani et al., 2015; Govindan et al., 2015; Safaei & Gholamrezatabar, 2014; Özaşkın & Görener, 2023). Among these, ELECTRE III, an outranking method introduced by Roy (1991), is particularly relevant due to its ability to incorporate thresholds of indifference, preference, and veto—features that make it suitable for non-compensatory decision contexts (Figueira et al., 2013). However, ELECTRE III faces two key limitations. First, it relies on crisp input data, reducing its effectiveness when expert evaluations are incomplete or uncertain. Second, its traditional distillation procedure may lead to unstable or ambiguous rankings when alternatives exhibit similar performance (Liang et al., 2023).

To address these challenges, this study proposes ED-ELEC (Enhanced D-numbers based ELECTRE), a hybrid framework that integrates D-numbers theory with ELECTRE III. D-numbers theory, introduced by Deng (2012), extends Dempster-Shafer evidence theory by allowing non-exclusiveness and incompleteness of information, making it well-suited for modeling uncertain expert judgments. Building on previous integration functions for D-numbers (Deng, 2012; Wang et al., 2018; Hou & Zhao, 2021), this study introduces a new integration function designed to retain incomplete and conflicting information while producing stable aggregated outcomes. Furthermore, ED-ELEC incorporates an improved ranking procedure from the literature to replace ELECTRE III's traditional distillation step, thereby enhancing ranking robustness and interpretability.

To ensure methodological rigor and demonstrate the advantages of the proposed approach, a two-stage validation process is employed. In the first stage, an illustrative example from the literature is used to compare four D-number integration functions in pure ranking form and to benchmark the complete ED-ELEC framework against another established hybrid method applied to the same problem. This step validates the performance, sensitivity, and stability of the proposed integration function and framework. In the second stage, ED-ELEC is applied to a real-world case study evaluating sustainable delivery systems for online food platforms, comparing gasoline motorcycles, electric motorcycles, and drones across economic, environmental, technical, and social criteria.

This study makes three main contributions. First, it proposes a new integration function for D-numbers that extends their applicability to multi-criteria decision-making problems under incomplete and uncertain information. Second, it develops the ED-ELEC framework, a hybrid approach that combines the enhanced D-numbers method with ELECTRE III and an improved ranking procedure to achieve more stable and discriminative decision outcomes. Third, the usefulness of the proposed framework is demonstrated through a two-stage validation process, consisting of an illustrative example and a real-world application to sustainable delivery system selection, thereby ensuring robustness, transparency, and practical relevance.

The remainder of this study is organized as follows: Chapter 2 reviews the literature on MCDM methods, D-numbers theory, and sustainable logistics. Chapter 3 presents the methodological foundation of ED-ELEC, including the new integration function and the modified outranking process. Chapter 4 applies the framework to the illustrative example and the case study, supported by comparative and sensitivity analyses. Chapter 5 concludes the study, summarizing the main contributions, limitations, and directions for future research.

Literature Background

Multi-Criteria Decision-Making and Uncertainty

Multi-criteria decision-making (MCDM) offers structured frameworks for evaluating alternatives based on multiple, often conflicting, criteria. Classical MCDM methods—such as the Analytic Hierarchy Process (AHP), Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), VIseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR), Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE), and the ELimination Et Choix Traduisant la REalité (ELECTRE) family—are widely applied across engineering, logistics, and policy domains (Figueira et al., 2005). Despite their popularity, these approaches generally assume that input information, such as criteria weights and performance evaluations, is precise and complete.

Recent reviews emphasize that uncertainty handling remains one of the main challenges in sustainability-oriented decision making (Das et al., 2024; French, 2023). Real-world problems are often characterized by vagueness, incompleteness, and subjectivity: experts may hesitate, provide imprecise judgments, or lack data for some criteria. To overcome these issues, several uncertainty modeling frameworks have been integrated with MCDM, including fuzzy set theory (Zadeh, 1965), intuitionistic fuzzy sets (Atanassov, 1986), and grey system theory (Liu & Lin, 2011). Dempster-Shafer evidence theory also provides a mechanism for expressing epistemic uncertainty (Dempster, 1967; Shafer, 1976), though its assumptions of mutual exclusiveness and complete belief assignment limit its applicability to complex decision contexts (Deng, 2012).

D-Numbers Theory in Decision-Making

To address these limitations, D-numbers theory generalizes Dempster-Shafer theory by allowing non-exclusive and incomplete belief assignments (Deng, 2012). This flexibility makes D-numbers well-suited for decision problems involving overlapping, hesitant, or partially informed expert judgments. A decade of research has confirmed their usefulness across various domains (Sotoudeh-Anvari, 2024).

In practice, once expert evaluations are expressed as D-numbers, they must be transformed into a crisp value through an integration function—a crucial step that connects uncertain evidence to conventional MCDM computation. Several integration mechanisms exist. The original function by Deng (2012) established the foundational approach, followed by a modified version by Wang et al. (2018) to manage conflicting evidence, and the SCRI (Stepwise Comparison and Replacement Integration) method by Hou and Zhao (2021) to improve interpretability. Yet, existing methods still struggle to balance incompleteness, uncertainty, and conflict, often leading to unrealistic or unstable outcomes.

D-numbers have since been incorporated into various decision-making frameworks. Deng et al. (2015) extended fuzzy preference relations into the D-CFPR model to handle incomplete information, while Deng and Jiang (2019) developed a D-numbers-based approach integrating belief intervals. More recently, Qi et al. (2025) proposed the Interval D-preference-based VIKOR method for multi-criteria group design evaluation under imprecise and unreliable information. Subsequent applications have spanned supplier evaluation (Deng et al., 2014a), risk assessment (Božanić et al., 2023), and environmental performance evaluation (Liu et al., 2021). However, most integrations remain confined to compensatory MCDM methods such as TOPSIS, VIKOR, and AHP, while limited attempts have been made to combine D-numbers with outranking approaches—leaving a methodological gap that this study seeks to address

ELECTRE III and Its Extensions

The ELECTRE family (Roy, 1991) represents a major class of outranking methods designed to model partial preference and non-compensatory decision behavior. ELECTRE III is particularly valued for its pseudo-criteria thresholds—indifference (q), preference (p), and veto (v)—which allow small performance differences to be treated as negligible and large ones as decisive (Figueira et al., 2013).

Nonetheless, ELECTRE III faces two limitations: (i) it assumes crisp input data, which is unrealistic when evaluations are uncertain; and (ii) its traditional distillation procedure may yield unstable or non-intuitive rankings (Liang et al., 2023). Li and Wang (2007) later proposed an improved ranking method offering enhanced discrimination and stability.

Recent studies continue to extend ELECTRE III under uncertain environments—such as fuzzy (Montazer et al., 2009), interval-valued intuitionistic fuzzy (Hashemi et al., 2016), and circular intuitionistic fuzzy versions (Yusoff et al., 2023). Emerging models also explore probabilistic and regret-based extensions (Liang et al., 2023; Fernández et al., 2025). However, no established integration currently exists between D-numbers theory and ELECTRE III, leaving open the opportunity for a hybrid model that can explicitly account for incompleteness and uncertainty while maintaining outranking logic.

Sustainable Delivery Systems and MCDM

Sustainability in logistics, particularly in last-mile delivery, has gained attention as urban transport systems face increasing environmental and social pressures. The expansion of online food delivery platforms, still dominated by gasoline-powered motorcycles, exacerbates carbon emissions, noise, and congestion. Sustainable alternatives, including electric motorcycles and drones, present opportunities but also raise challenges in terms of cost, reliability, regulation, and social acceptance (Cherry et al., 2009; Matsuyuki et al., 2024; World Economic Forum & Accenture, 2024).

MCDM approaches have been widely used to support sustainable transport decisions. Aljohani and Thompson (2018) applied fuzzy AHP and PROMETHEE for urban transport planning, while Nur et al. (2020) evaluated drones as logistics vehicles. Wang et al. (2021) used Two-Stage Fuzzy MCDM Approach to examine the sustainability of delivery companies in Vietnam. Although these studies highlight the importance of multi-criteria frameworks, they typically employ fuzzy or probabilistic techniques—none have leveraged D-numbers or outranking methods to address uncertainty in emerging delivery technologies.

The literature highlights a clear need for decision-making approaches that can reliably handle the incomplete and conflicting judgments common in sustainability evaluations. While D-numbers offer a promising way to model ignorance and non-exclusiveness, existing integration functions often yield unstable or unrealistic aggregations under imperfect information. In parallel, although ELECTRE III is well suited to complex, non-compensatory decision contexts, it has not yet been effectively combined with D-numbers to improve ranking reliability when evaluations are incomplete. This creates an opportunity to develop a more balanced D-number integration mechanism, embed it within an outranking framework, and demonstrate its usefulness in selecting sustainable urban delivery systems—an emerging domain where uncertainty is pervasive.

Materials and Methods

Building on the contributions outlined in Section 2, this section presents the methodological foundations of the proposed ED-ELEC framework. It is structured to provide both theoretical background and the specific enhancements introduced in this study.

D-Numbers Theory

The theory of D-numbers was developed by Deng (2012) as a generalization of Dempster-Shafer Theory (DST), which itself extends probability theory (Dempster, 1967; Shafer, 1976). In DST, information is represented through basic probability assignments (BPAs), which distribute belief mass across a frame of discernment. Two strong assumptions underlie this approach (Deng et al., 2014b):

- The elements in the frame of discernment must be mutually exclusive.
- The sum of all BPAs must equal one (completeness).

These assumptions are often unrealistic in real-world decision-making. For example, linguistic evaluations such as “good” and “very good” are not strictly exclusive, yet DST requires them to be disjoint. Similarly, experts may lack knowledge, leading to incomplete assessments that violate the completeness condition.

D-numbers theory relaxes both assumptions. First, it allows overlap among elements, making it possible to represent non-exclusive judgments. Second, it permits incompleteness, so that the sum of assigned masses may be less than one, with the residual representing ignorance or missing knowledge. These innovations make D-numbers particularly suitable for modeling uncertain, hesitant, and partially informed expert judgments, which are common in multi-criteria decision-making contexts (Deng et al., 2014a; Deng & Jiang, 2019). This theory can be defined as follows (Deng et al., 2014b; Deng, 2012):

Definition 1: Let Ω be a finite nonempty set. A D-number is a mapping:

$D: \Omega \rightarrow [0,1]$

such that

$\sum_{B\subseteq\Omega} D(B) \leq 1$ and $D(\emptyset) = 0$

Where $\emptyset$ is an empty set and B is a subset of Ω. If the inequality becomes equality, the information is considered complete; otherwise, it is incomplete. Here lies the initial distinction from DST.

For a discrete set $\Omega = \{b_1, b_2, \dots, b_i, \dots, b_n\}$, considering $b_i \in R$ and $b_i \neq b_j$ if $i \neq j$, a special form of D-numbers can be expressed as:

$D = \{(b_1, v_1), (b_2, v_2), \dots, (b_i, v_i), \dots, (b_n, v_n)\}$

Where each pair (b_i, v_i) denotes the support value b_i with an associated confidence v_i and we have:

$v_i > 0$ and $\sum_{i=1}^{n} v_i \leq 1$

Definition 2 (Permutation invariability): If two D-numbers differ only in the order of their elements, they are equivalent. These D-numbers can be shown as:

$D_1 = \{(b_1, v_1), \dots, (b_i, v_i), \dots, (b_n, v_n)\}$, and

$D_2 = \{(b_n, v_n), \dots, (b_i, v_i), \dots, (b_1, v_1)\}$, then

$D_1 \Leftrightarrow D_2$

Definition 3 (Integration): For $D = \{(b_1, v_1), (b_2, v_2), \dots, (b_i, v_i), \dots, (b_n, v_n)\}$ the integration representation of D is defined as:

$$I(D) = \sum_{i=1}^{n} b_i v_i$$

Where $I(D)$ is real number, $b_i \neq b_j$, $v_i > 0$ and $\sum_{i=1}^{n} v_i \leq 1$

Existing Integration Functions for D-Numbers

Since its introduction, several integration functions have been proposed to transform D-numbers into usable values for decision analysis. The strength of D-numbers lies in their ability to handle non-deterministic and incomplete information. In most decision-making problems, the decision matrix contains uncertain assessments provided by experts. In practice, however, experts may lack opinions on some criteria. D-numbers allow such responses to be represented as “do not know” or left blank, thereby preserving the incompleteness of information.

Nevertheless, the original integration method for D-numbers, proposed by Deng (2012), has limitations that can influence results. The integration is defined as:

$$I(D) = \sum_{i=1}^{n} b_i v_i \quad (1)$$

where b_i represents the evaluation value and v_i its associated support. Wang et al. (2018) pointed out that this approach has a defect: it assigns a value of zero to missing evaluations, which may distort the final integration result.

Example 1 (Wang et al, 2018): Consider an evaluation of two types of cars, where ten experts provide assessments on the performance scale $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. The evaluations are represented as:

$D_A = \{(7,0.6), (8,0.2)\}$ and $D_B = \{(6,1)\}$

This means that for Car A, 6 of the 10 experts assign a score of 7, 2 experts assign a score of 8, and the remaining 2 experts respond “do not know.” For Car B, all 10 experts assign a score of 6. The degree of incompleteness for Car A is $1 - 0.8 = 0.2$. Using Deng’s original integration function, the results are:

$I(D_A) = 7 \times 0.6 + 8 \times 0.2 = 5.8$ and $I(D_B) = 6 \times 1 = 6$

Although Car A receives higher evaluations from the knowledgeable experts, its integrated value is reduced due to missing information. This occurs because the original integration method implicitly treats absent evaluations as zero, which distorts the outcome and produces a less reasonable result.

To address this issue, Wang et al. (2018) proposed a modified integration function that incorporates the degree of incompleteness into the aggregation process. After simplification, the integration is expressed as:

$$I'(D) = \sum_{i=1}^{n} \left\{ b_i \times \left[v_i + \left(1 - \sum_{i=1} v_i\right) \times \frac{v_i}{\sum_{i=1} v_i} \right] \right\} \quad (2)$$

The results become as below:

$$I'(D_A) = 7 \times \left(0.6 + 0.2 \times \frac{0.6}{0.8}\right) + 8 \times \left(0.2 + 0.2 \times \frac{0.2}{0.8}\right) = \frac{29}{4} = 7.25 \text{ and } I'(D_B) = 6 \times 1 = 6$$

This method produces results that better reflect the underlying expert judgments. However, Hou and Zhao (2021) later observed that Wang's method still exhibits shortcomings.

Example 2 (Hou & Zhao, 2021): Suppose 10 experts evaluate two alternatives on the performance scale $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. For Alternative A, 7 experts assign a score of 8, while the remaining 3 give no response. For Alternative B, 8 experts assign a score of 8 and 2 experts assign a score of 7, those represented as:

$$D_A = \{(8,0.7)\} \qquad \text{and} \qquad D_B = \{(8,0.8).(7,0.2)\}$$

Using Wang's integration function, the results are:

$$I'(D_A) = 8 \times \left(0.7 + 0.3 \times \frac{0.7}{0.7}\right) = 8 \qquad \text{and} \qquad I'(D_B) = 8 \times 0.8 + 7 \times 0.2 = 7.8$$

Although more experts evaluated Alternative B with a score of 8, Wang's method incorrectly favors Alternative A. This occurs because the missing information for A is artificially replaced by the average of the complete values (equal to 8), which can distort the decision outcome. To address this issue, Hou and Zhao (2021) proposed the Stepwise Comparison and Replacement Integration (SCRI) method. The procedure is summarized as follows:

1. Construct a matrix R of size $N \times M$, where N is the number of alternatives and M the number of experts. Expert evaluations are placed in descending order, with 0 denoting missing responses.
2. Multiply each element of R by $1/M$ to obtain matrix R_{int}.
3. If no all-zero columns appear, proceed to Step 5. Otherwise, continue to Step 4.
4. Identify The first zero element, then the zero value of the column in step 3 is replaced with the average of the row values before that first zero.
5. Replace any remaining zeros with the minimum nonzero value in the corresponding column.

6. Aggregate the final adjusted values to obtain the integrated D-numbers for each alternative.

Applying this method to *Example 2,* the initial matrix R is:

$$R = \begin{pmatrix} 8 & 8 & 8 & 8 & 8 & 8 & 8 & 0 & 0 & 0 \\ 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 7 & 7 \end{pmatrix}$$

After normalization by dividing by 10 (the number of experts):

$$R_{Int} = \begin{pmatrix} 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 0 & 0 & 0 \\ 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 7/10 & 7/10 \end{pmatrix}$$

Because no column is entirely zero, Step 4 is omitted. Step 5 replaces the zero entries with the smallest values in each column, producing:

$$R_{Int^*} = \begin{pmatrix} 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 7/10 & 7/10 \\ 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 8/10 & 7/10 & 7/10 \end{pmatrix}$$

Summing each row, the integrated values are then:

$I''_{D(A)} = 7.8$ and $I''_{D(B)} = 7.8$

The integration results for the two cars are identical, which appears more reasonable than the outcome of Wang's method. However, obtaining exactly equal integration values in such an uncertain and incomplete setting raises concerns about potential shortcomings in this approach. Further analysis reveals two main limitations. First, the method produces different integration values when new alternatives are added or existing ones are removed. This instability reduces its reliability as a universal evaluation method and increases computational complexity. Second, similar to Wang's method, this approach cannot distinguish between complete and incomplete information when all expert evaluations are identical, which may lead to misleading conclusions.

Example 3: Consider again the situation in Example 2, but with a third alternative, C, evaluated by 10 experts assigning a score of 8, which can be expressed as follows:

$D(A) = \{(8,0.7)\},$ $D(B) = \{(8,0.8),(7,0.2)\}$ and $D(C) = \{(8,1)\}.$

Applying Hou's SCRI method, the integrated results of all the aforementioned methods are presented in Table 1.

Table 1. The integrated D-numbers of Deng, wang and Hou methods.

Original method (Deng)	Wang method	Hou's method considering objects A and B	Hou's method considering objects A and C	Hou's method considering objects A, B and C
$I_{D(A)} = 5.6$	$I'_{D(A)} = 8$	$I''_{D(A)} = 7.8$	$I''_{D(A)} = 8$	$I''_{D(A)} = 7.8$
$I_{D(B)} = 7.8$	$I'_{D(B)} = 7.8$	$I''_{D(B)} = 7.8$	-	$I''_{D(B)} = 7.8$
$I_{D(C)} = 8$	$I'_{D(C)} = 8$	-	$I''_{D(C)} = 8$	$I''_{D(C)} = 8$

Above results show that the aggregated score in Hou's method changes with the addition or removal of alternatives, and therefore is not stable. Moreover, both Wang's method and Hou's method (in the case of comparing two cars) assign the same score to cars A and C. In reality, however, car C received a score of 8 with greater certainty, and logically its score should not be equal to that of car A. These developments illustrate the evolution of D-number integration functions, each addressing specific shortcomings of its predecessors.

Proposed New Integration Function

The integration method proposed in this study adopts a balanced approach to handling incomplete information in D-numbers. Existing methods treat incompleteness in extreme ways: Deng's method overly penalizes missing values by assigning zeros, while Wang's method neutralizes their effect by replacing them with averages, both of which may distort decision outcomes. Hou's SCRI method was introduced as an alternative by replacing missing evaluations stepwise using averages or minimum values; however, this procedure makes the integration results dependent on the set of alternatives considered, leading to instability when alternatives are added or removed. Moreover, SCRI cannot distinguish between complete and incomplete information when expert evaluations are identical, which limits its ability to reflect uncertainty realistically.

The proposed method addresses these shortcomings by replacing missing values with the mean of Deng's and Wang's integration results. This strategy preserves the penalizing effect of incompleteness without exaggeration, while ensuring that the integrated value remains stable and independent of the number of alternatives. Unlike Hou's method, the proposed integration consistently differentiates between complete and incomplete information and converges to the same result as existing methods when data are complete. The outcome is a more realistic and balanced integration function, which better reflects both the quality of the available evaluations and the uncertainty introduced by missing information. The new integration function formulated as follows.

Let a D-number be $D = \{(b_1, v_1), (b_2, v_2), .., (b_i, v_i), \dots, (b_n, v_n)\}$, Where $0 \leq v_i \leq 1$ and $\sum_{i=1}^{n} v_i \leq 1$. The proposed integration value of D is:

$$I^3(D) = \sum_{i=1}^{n} b_i v_i \times \left(1 + \frac{1 - (\sum v_i)^2}{2 \sum v_i}\right) \quad (3)$$

Example 4: Suppose Example 3, the integration number of car A becomes:

$$I^3{}_{D(A)} = (8 \times 0.7) \times \left(1 + \frac{1 - 0.7^2}{2 \times 0.7}\right) = 7.64$$

Other objects integration number is obtained same as other methods as their D numbers are complete:

$I^3{}_{D(B)} = 7.8$ and $I^3{}_{D(C)} = 8$

The above results indicate a more rational evaluation, as a larger number of experts assigned a score of 8 to object B, making its outcome more certain compared to object A, which had three missing evaluations. Considering the objects in Example 3, the integrated D-numbers obtained from the aforementioned methods are presented in Table 2.

Table 2. The integrated D-numbers of the proposed and other methods.

Original method (Deng)	Wang method	Hou method considering objects A and B	Hou method considering objects A and C	Proposed method
$I_{D(A)} = 5.6$	$I'_{D(A)} = 8$	$I''_{D(A)} = 7.8$	$I''_{D(A)} = 8$	$I^3{}_{D(A)} = 7.64$
$I_{D(B)} = 7.8$	$I'_{D(B)} = 7.8$	$I''_{D(B)} = 7.8$	-	$I^3{}_{D(A)} = 7.8$
$I_{D(C)} = 8$	$I'_{D(C)} = 8$	-	$I''_{D(C)} = 8$	$I^3{}_{D(A)} = 8$

Overview of ELECTRE III

ELECTRE III, developed by Roy (1991), is an outranking method designed to support complex multi-criteria decision-making (MCDM) problems where preferences are not necessarily compensatory. The method is based on the construction of a valued outranking relation among alternatives, combining concordance (support for one alternative over another), discordance (opposition), and predefined thresholds that capture decision-maker hesitation. Its key strength lies in the ability to model situations where small differences in performance do not indicate strict preference, while large differences may trigger strong opposition (Figueira et al., 2013).

(a) Pseudo-criteria and Thresholds

For each criterion, ELECTRE III introduces three thresholds:

- Indifference threshold (q): below which differences are negligible.
- Preference threshold (p): above which a strict preference is considered.
- Veto threshold (v): beyond which the preference of one alternative over another is strongly rejected.

These thresholds allow ELECTRE III to model gradual transitions between indifference, weak preference, and strong preference.

(b) Partial Concordance Index

Given two alternatives a and b, the partial concordance index $c_j(a,b)$ measures the degree to which criterion j supports the statement "a is at least as good as b." It is defined using the indifference and preference thresholds. The global concordance index $C(a,b)$ is then obtained as the weighted sum of the partial concordance indices across all criteria.

(c) Discordance Index

The discordance index captures the extent to which criterion j opposes the assertion that a outranks b. It uses the preference and veto thresholds to determine whether criterion j blocks the outranking relation. If the difference between b and a exceeds the veto threshold, the criterion effectively rejects the outranking.

(d) Credibility Index

The credibility index $\sigma(a,b)$ aggregates concordance and discordance, representing the overall strength of the outranking relation "a is at least as good as b." It is computed by reducing the global concordance when one or more criteria exhibit strong discordance.

(e) Ranking Procedure

Traditionally, ELECTRE III applies a distillation process to generate a partial preorder of alternatives. This involves iterative descending and ascending distillations based on credibility indices, which are then combined into a final ranking. However, this process has been criticized for producing unstable or counterintuitive results when credibility values are close. To address these issues, Li and Wang (2007) proposed an improved ranking method that directly uses credibility indices to yield more stable and discriminative outcomes.

The ED-ELEC Framework

Building on D-numbers theory and the ELECTRE III method, this study proposes the ED-ELEC (Enhanced D-numbers-based ELECTRE) framework to support outranking analysis under incomplete and uncertain expert evaluations. ED-ELEC addresses key limitations of classical ELECTRE III by explicitly modeling uncertainty and improving ranking stability.

The framework evaluates a predefined set of alternatives and criteria using expert judgments expressed on a discrete Saaty 1–9 scale (Saaty, 1987). Experts may leave evaluations blank when information is insufficient, and each expert is assigned a weight reflecting their expertise or decision authority. These weighted assessments are aggregated into D-numbers, where missing evaluations are explicitly represented as incompleteness.

To obtain a usable decision matrix, the aggregated D-numbers are converted into crisp values using the proposed balanced integration function, which moderates the effect of missing information. The resulting matrix is analyzed using ELECTRE III to compute concordance, discordance, and credibility indices. To avoid instability associated with the classical distillation procedure, ED-ELEC employs the improved ranking method proposed by Li and Wang (2007).

Stepwise Flow of ED-ELEC

Input:

- Set of alternatives $A = \{A_1, A_2, \dots, A_m\}$.
- Set of criteria $C = \{C_1, C_2, \dots, C_n\}$ with predefined weights w_j.
- Group of experts $E = \{E_1, E_2, \dots, E_k\}$, each assigned a weight λ_k (reflecting expertise, power, or experience).
- Performance scale $\Omega = \{1, 3, 5, 7, 9\}$.
- Indifference, preference, and veto thresholds (q_j, p_j, v_j) for each criterion.

Stage 1: Data Preparation

1- Expert evaluation: Each expert $\mathrm{E_k}$ evaluates every alternative $\mathrm{A_i}$ under each criterion $\mathrm{C_j}$ using the Saaty's 1–9 scale with odd-numbered increments (1, 3, 5, 7, 9). If the expert has no information, they respond *"do not know"* or leave the entry blank.

2- Construct expert decision matrices: For each expert, form a decision matrix $\mathrm{M^k} = [\mathrm{m_{ij}^k}]$, where $\mathrm{m_{ij}^k} \in \Omega \cup \{\emptyset\}$.

3- Aggregate into weighted D-numbers: For each element (i, j):

- For each distinct score $b \in \Omega$, define the support value v_b as the **sum of the weights of experts who assigned score** b:

$$v_b = \sum_{\{k|m_{ij}^k = b\}} \lambda_k \tag{4}$$

- If some experts leave blanks, their weights contribute to the incompleteness degree.
- The resulting D-number is:

$$D_{ij} = \{(b, v_b | b \in \Omega, v_b > 0\}, \qquad \sum v_b \leq 1$$

4- Apply integration function: Convert each D_{ij} into a crisp value using the proposed integration function (eq. (3)) Which is:

$$x_{ij} = \sum_{l=1}^{n} b_l v_l \times (1 + \frac{1 - (\sum v_l)^2}{2 \sum v_l})$$

5- Form crisp decision matrix: Construct the integrated matrix $M^* = [x_{ij}]$, where each entry x_{ij} is the crisp value of D_{ij}.

Stage 2: ELECTRE III (Hashemi et al., 2016)

6- Compute pairwise concordance indices: The concordance index expresses the degree to which the performance of alternative A_r is at least as good as that of A_s, considering all criteria.

- For each criterion C_j**,** define the difference:

$$d_{rs}^j = x_{rj} - x_{sj} \tag{5}$$

- Then compute the partial concordance index $c_j(A_r, A_s)$ **as:**

$$c_j(A_r, A_s) = \begin{cases} 1, & d_{rs}^j \geq -q_j \\ 0, & d_{rs}^j \leq -p_j \\ \dfrac{d_{rs}^j + p_j}{p_j - q_j}, & \text{otherwise} \end{cases} \tag{6}$$

Where q_j is the indifference threshold and p_j the preference threshold.

- The global concordance index is a weighted sum of partial concordance indices:

$$C(A_r, A_s) = \sum_{j=1}^{n} w_j c_j(A_r, A_s) \tag{7}$$

7- Compute discordance indices: The discordance index measures the degree to which a large disadvantage on one criterion may veto the outranking of A_r over A_s. For each criterion C_j we have:

$$d_j(A_r, A_s) = \begin{cases} 0, & d_{rs}^j \geq -p_j \\ 1, & d_{rs}^j \leq -v_j \\ \dfrac{d_{rs}^j + p_j}{p_j - v_j}, & \text{otherwise} \end{cases} \tag{8}$$

Where v_j is the veto threshold.

8- Build credibility matrix: The credibility index combines concordance and discordance into a single measure of how strongly "A_r outranks A_s". it is calculated as:

$$S(A_r, A_s) = \begin{cases} C(A_r, A_s) & \text{if } d_j(A_r, A_s) \leq C(A_r, A_s)\ \forall j, \\ C(A_r, A_s) \times \prod_{j \in J(A_r, A_s)} \frac{1 - d_j(A_r, A_s)}{1 - C(A_r, A_s)} & \text{otherwise.} \end{cases} \tag{9}$$

This ensures that when a criterion strongly disagrees (discordance above concordance), the credibility index is penalized. The result is a **credibility matrix** $S = [S(A_r, A_s)]$, where each entry shows the outranking strength between two alternatives.

9- Compute concordance credibility degree: For each alternative A_i, compute its concordance credibility degree:

$$\Phi^+(A_r) = \sum_{A_s \in A} S(A_r, A_s), \qquad \forall A_r \in A \tag{10}$$

where $S(A_r, A_s)$ is the credibility index of A_r outranking A_s. This measures the **dominance strength** of A_r, i.e., how strongly it outranks all other alternatives.

10- Compute discordance credibility degree: For each alternative A_r, compute its discordance credibility degree:

$$\Phi^{-}(A_r) = \sum_{A_s \in A} S(A_s, A_r), \qquad \forall A_r \in A \tag{11}$$

which measures the extent to which A_r is dominated by other alternatives.

11- Compute net credibility degree and ranking: For each alternative A_r, compute its net credibility degree:

$$\Phi(A_r) = \Phi^{+}(A_r) - \Phi^{-}(A_r), \qquad \forall A_r \in A \tag{12}$$

A higher $\Phi(A_r)$ indicates a more attractive alternative. Finally, all alternatives are ranked in descending order of $\Phi(A_r)$, yielding the complete and stable final ranking.

Results

This chapter presents the results obtained using the proposed ED-ELEC framework for evaluating alternative delivery modes under uncertainty. The analysis is divided into two main stages; First, an illustrative example from the literature is used to validate the proposed approach and compare it with existing D-number and hybrid decision-making methods. Second, the framework is applied to a real-world case study involving food delivery vehicles, incorporating expert judgments and performance criteria relevant to sustainable urban logistics. Both stages include comparative and sensitivity analyses to evaluate the framework's stability, consistency, and discriminative ability.

Illustrative Example from the Literature

To ensure comparability and validation, an illustrative multi-criteria decision problem was adapted from work of Hou and Zhao (2021, 2022), where the same problem was solved using the SCRI method integration in two frameworks:

- a pure D-number ranking method, and
- a D-SCRI-EDAS hybrid model.

The problem involves three alternatives (A_1, A_2, A_3) evaluated across 16 criteria $(C_1, C_2, \dots, C_{16})$. The D-number evaluations and normalized criteria weights are summarized in Table 3.

Table 3. Decision matrix and criteria weights from the literature example.

Criteria	Weight	A_1	A_2	A_3
C_1	0.1466	{(9,0.2),(8,0.5),(7,0.3)}	{(8,0.5),(7,0.2),(6,0.2),(5,0.1)}	{(7,0.5),(6,0.2),(5,0.2),(4,0.1)}
C_2	0.1162	{(9,0.2),(8,0.5),(6,0.2)}	{(8,0.2),(7,0.2),(6,0.2),(5,0.2)}	{(8,0.2),(7,0.3),(6,0.1),(5,0.1)}
C_3	0.0339	{(9,0.2),(8,0.2),(7,0.3),(6,0.1)}	{(8,0.4),(7,0.3),(6,0.2)}	{(8,0.2),(7,0.4),(5,0.1)}
C_4	0.0967	{(8,0.2),(7,0.2),(6,0.4)}	{(8,0.2),(7,0.2),(6,0.6)}	{(7,0.2),(6,0.4),(5,0.2),(4,0.1)}
C_5	0.0371	{(9,0.6),(8,0.2),(7,0.2)}	{(8,0.2),(7,0.2),(6,0.1),(5,0.4)}	{(8,0.4),(7,0.2),(5,0.2)}
C_6	0.0308	{(7,0.2),(6,0.2),(5,0.1),(4,0.2)}	{(7,0.2),(5,0.4),(4,0.2)}	{(9,0.2),(8,0.2),(7,0.3),(5,0.1)}
C_7	0.0185	{(8,0.2),(6,0.2),(5,0.4)}	{(9,0.2),(7,0.2),(6,0.4)}	{(7,0.4),(6,0.2),(5,0.2)}
C_8	0.0077	{(9,0.4),(8,0.4),(6,0.2)}	{(9,0.2),(8,0.3),(6,0.2),(5,0.2)}	{(8,0.4),(7,0.1),(5,0.2),(4,0.1)}
C_9	0.0149	{(9,0.1),(8,0.4),(6,0.3)}	{(8,0.1),(7,0.6),(6,0.1)}	{(8,0.4),(7,0.3),(4,0.2)}
C_{10}	0.0756	{(9,0.3),(7,0.2),(6,0.4)}	{(8,0.5),(7,0.2),(6,0.2)}	{(9,0.2),(8,0.6)}
C_{11}	0.0416	{(6,0.6),(5,0.2)}	{(8,0.2),(7,0.2),(6,0.2),(4,0.1),(3,0.1)}	{(8,0.4),(6,0.2),(4,0.1)}
C_{12}	0.0238	{(8,0.3),(7,0.4),(5,0.1)}	{(9,0.2),(8,0.2),(6,0.3),(5,0.1)}	{(8,0.3),(7,0.2),(6,0.2),(5,0.1)}
C_{13}	0.0803	{(9,0.4),(8,0.2),(7,0.2),(5,0.2)}	{(8,0.4),(7,0.2),(6,0.2),(5,0.2)}	{(9,0.2),(7,0.4),(6,0.2)}
C_{14}	0.1345	{(9,0.2),(8,0.4),(6,0.3)}	{(7,0.4),(6,0.4),(5,0.2)}	{(8,0.2),(7,0.2),(6,0.2),(5,0.1),(4,0.1)}
C_{15}	0.0358	{(9,0.2),(8,0.4),(7,0.2),(5,0.2)}	{(8,0.2),(7,0.2),(6,0.6)}	{(8,0.2),(7,0.2),(6,0.2),(5,0.4)}
C_{16}	0.1058	{(8,0.4),(7,0.1),(6,0.4)}	{(8,0.3),(7,0.2),(6,0.4)}	{(7,0.5),(6,0.1),(5,0.3)}

Comparison of D-Number Integration Functions

The first analysis investigates how different D-number integration methods influence the resulting crisp values and rankings. Four integration methods were compared:

1. Original integration function (Deng, 2012), referred to here as the Deng method.
2. Modified integration function (Wang et al., 2018), referred to here as the Wang method.
3. SCRI integration function (Hou & Zhao, 2021), referred to here as the Hou method.
4. Proposed integration function (this study).

Step 1: Conversion to Crisp Values

Each D-number $D = \{(b_i, v_i)\}$ was transformed into a crisp equivalent value using the corresponding methods. The resulting crisp decision matrices are presented in Tables 4.

Step 2: Ranking Using the WSM

The weighted sum model (WSM) ranking procedure was applied to each set of crisp values. The computed scores, normalized scores, resulting alternative rankings, and the normalized max-min difference (Δ_0) are shown in Table 5.

Tables 4. Crisp decision matrices obtained using Deng, Wang, Hou, and the proposed integration methods

Deng Method						Wang Method				
C	A_1	A_2	A_3	Partial Rank		C	A_1	A_2	A_3	Partial Rank
C_1	7.9	7.1	6.1	$A_1 > A_2 > A_3$		C_1	7.9	7.1	6.1	$A_1 > A_2 > A_3$
C_2	7	5.2	4.8	$A_1 > A_2 > A_3$		C_2	7.7778	6.5	6.8571	$A_1 > A_3 > A_2$
C_3	6.1	6.5	4.9	$A_2 > A_1 > A_3$		C_3	7.625	7.2222	7	$A_1 > A_2 > A_3$
C_4	5.4	6.6	5.2	$A_2 > A_1 > A_3$		C_4	6.75	6.6	5.7778	$A_1 > A_2 > A_3$
C_5	8.4	5.6	5.6	$A_1 > A_2 = A_3$		C_5	8.4	6.2222	7	$A_1 > A_3 > A_2$
C_6	3.9	4.2	6	$A_3 > A_2 > A_1$		C_6	5.5714	5.25	7.5	$A_3 > A_1 > A_2$
C_7	4.8	5.6	5	$A_2 > A_3 > A_1$		C_7	6	7	6.25	$A_2 > A_3 > A_1$
C_8	8	6.4	5.3	$A_1 > A_2 > A_3$		C_8	8	7.1111	6.625	$A_1 > A_2 > A_3$
C_9	5.9	5.6	6.1	$A_3 > A_1 > A_2$		C_9	7.375	7	6.7778	$A_1 > A_2 > A_3$
C_{10}	6.5	6.6	6.6	$A_2 = A_3 > A_1$		C_{10}	7.2222	7.3333	8.25	$A_3 > A_2 > A_1$
C_{11}	4.6	4.9	4.8	$A_2 > A_3 > A_1$		C_{11}	5.75	6.125	6.8571	$A_3 > A_2 > A_1$
C_{12}	5.7	5.7	5.5	$A_1 = A_2 > A_3$		C_{12}	7.125	7.125	6.875	$A_1 = A_2 > A_3$
C_{13}	7.6	6.8	5.8	$A_1 > A_2 > A_3$		C_{13}	7.6	6.8	7.25	$A_1 > A_3 > A_2$
C_{14}	6.8	6.2	5.1	$A_1 > A_2 > A_3$		C_{14}	7.5556	6.2	6.375	$A_1 > A_3 > A_2$
C_{15}	7.4	6.6	6.2	$A_1 > A_2 > A_3$		C_{15}	7.4	6.6	6.2	$A_1 > A_2 > A_3$
C_{16}	6.3	6.2	5.6	$A_1 > A_2 > A_3$		C_{16}	7	6.8889	6.2222	$A_1 > A_2 > A_3$
Hou Method						The Proposed Method				
C	A_1	A_2	A_3	Partial Rank		C	A_1	A_2	A_3	Partial Rank
C_1	7.9	7.1	6.1	$A_1 > A_2 > A_3$		C_1	7.9	7.1	6.1	$A_1 > A_2 > A_3$
C_2	7.8286	6.4714	6.5857	$A_1 > A_3 > A_2$		C_2	7.7389	6.37	6.5486	$A_1 > A_3 > A_2$
C_3	7.4857	7.2571	6.8	$A_1 > A_2 > A_3$		C_3	7.4725	7.1861	6.685	$A_1 > A_2 > A_3$
C_4	6.4	6.6	5.8	$A_2 > A_1 > A_3$		C_4	6.615	6.6	5.7489	$A_1 > A_2 > A_3$
C_5	8.4	6.3	6.8	$A_1 > A_3 > A_2$		C_5	8.4	6.1911	6.86	$A_1 > A_3 > A_2$
C_6	5.4143	5.2857	7.5714	$A_3 > A_1 > A_2$		C_6	5.3207	5.145	7.35	$A_3 > A_1 > A_2$
C_7	6	7	6.25	$A_2 > A_3 > A_1$		C_7	5.88	6.86	6.125	$A_2 > A_3 > A_1$
C_8	8	7	6.4	$A_1 > A_2 > A_3$		C_8	8	7.0756	6.4925	$A_1 > A_2 > A_3$
C_9	7.0375	6.7	6.8125	$A_1 > A_3 > A_2$		C_9	7.2275	6.86	6.7439	$A_1 > A_2 > A_3$
C_{10}	7.2375	7.35	8.025	$A_3 > A_2 > A_1$		C_{10}	7.1861	7.2967	8.085	$A_3 > A_2 > A_1$
C_{11}	5.7714	6.2143	6.4714	$A_3 > A_2 > A_1$		C_{11}	5.635	6.0025	6.5486	$A_3 > A_2 > A_1$
C_{12}	7.125	7.125	6.875	$A_1 = A_2 > A_3$		C_{12}	6.9825	6.9825	6.7375	$A_1 = A_2 > A_3$
C_{13}	7.6	6.8	6.8	$A_1 > A_2 = A_3$		C_{13}	7.6	6.8	7.105	$A_1 > A_3 > A_2$
C_{14}	7.3	6.2	6.1	$A_1 > A_2 > A_3$		C_{14}	7.5178	6.2	6.2475	$A_1 > A_3 > A_2$
C_{15}	7.4	6.6	6.2	$A_1 > A_2 > A_3$		C_{15}	7.4	6.6	6.2	$A_1 > A_2 > A_3$
C_{16}	7	6.8889	6.2222	$A_1 > A_2 > A_3$		C_{16}	6.965	6.8544	6.1911	$A_1 > A_2 > A_3$

Table 5. Rankings and scores information of alternatives under different D-number integration functions

Method		A_1	A_2	A_3	Δ_0
Deng Method	Scores	6.6453	6.1878	5.5334	0.1673
	Normalized Score	1	0.9311	0.8327	
	ranking	1	2	3	
Wang Method	Scores	7.3257	6.6842	6.6379	0.0939
	Normalized Score	1	0.9124	0.9061	
	ranking	1	2	3	
Hou Method	Scores	7.2508	6.6856	6.4891	0.1051
	Normalized Score	1	0.9221	0.8949	
	ranking	1	2	3	
The Proposed Method	Scores	7.2711	6.6436	6.5142	0.1041
	Normalized Score	1	0.9137	0.8959	
	ranking	1	2	3	

Step 3: Comparative Robustness and Sensitivity Analysis

Given the high completeness of the original data (approximately 85% on average), all integration functions produced identical ordinal rankings, as expected. However, differences were observed in preference intensity and discriminatory power. To further assess the robustness and sensitivity of the integration functions, three complementary analyses were conducted as Scenarios 1, 2, and 3.

SC1: Data completeness reduction

To assess the performance of different methods in handling uncertain data, the information completeness of the dataset was reduced by 20% by applying this reduction to the components v_i of all D-numbers. The comparative analysis of various D-number integration methods reveals distinct behaviors in addressing missing information, as quantified by the delta index (Δ), defined as the difference between the maximum and minimum normalized scores. Table 6 presents the original Δ values (SC_0), the Δ values after a 20% reduction in data completeness (SC_1), and the resulting change (Δ_1-Δ_0) for each method.

Table 6. Comparative results of SC1 and SC2 through different methods

Method		SC0	SC1	$\Delta_1 - \Delta_0$
Deng Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_2 > A_3$	0
	Δ	0.1673	0.1673	
Wang Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_2 > A_3$	0
	Δ	0.0939	0.0939	
Hou Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_2 > A_3$	0.001
	Δ	0.1051	0.1061	
The Proposed Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_2 > A_3$	0.0076
	Δ	0.1041	0.1117	

Industrial Management Journal, Volume 18, Issue 2, 2026

The initial discriminatory power, measured by the delta index, differs notably across methods. The proposed method exhibits an intermediate initial Δ (0.1041), balancing the high sensitivity of the Deng method (Δ = 0.1673) and the lower discrimination of the Wang method (Δ = 0.0939), consistent with recommended robustness–sensitivity trade-offs in MCDM. When data completeness is reduced by 20%, the Deng and Wang methods show no change in Δ (Δ_1–Δ_0 = 0.0000), indicating insensitivity to increased uncertainty, while the Hou method exhibits only a marginal increase (Δ_1–Δ_0 = 0.0010). In contrast, the proposed method's Δ rises to 0.1117 (Δ_1–Δ_0 = 0.0076), demonstrating an uncertainty-aware response. By balancing zero-replacement and mean-imputation strategies, the proposed method penalizes missing information while preserving data structure, dynamically adjusting discriminatory strength and making it well suited for real-world decision problems with incomplete data.

SC2: Addition of a new criterion

To further assess the robustness and validity of the proposed method, an additional criterion, Reliability (C_{17}), was introduced into the decision matrix. This new criterion was assigned a higher-than-average weight ($w_{17} = 0.15$) and designed with varying completeness ratios to test the method's responsiveness under increased uncertainty. The assumed expert evaluations for C_{17} were then converted into D-numbers as follows:

- A_1: {(9, 0.5), (8, 0.2)}; 50% assigned 9, 20% assigned 8, and 30% provided no evaluation.
- A_2: {(9, 0.4)}; 40% of experts assigned score 9, while 60% provided no evaluation.
- A_3: {(9, 0.5), (8, 0.4)}; 50% assigned 9, 40% assigned 8, and 10% with no evaluation.

The crisp scores obtained using the different integration functions are reported in Table 7. The Deng method excessively penalizes incompleteness by assigning zero to missing evaluations, while the Wang method overcompensates by assigning a full score of 9 to alternative A_2 based on only 40% data availability and by ranking A_1 above A_3, which is logically inconsistent. The Hou method yields identical scores for all alternatives, failing to reflect actual performance differences. Both the Wang and Hou methods rely heavily on averaging available data to offset missing information, resulting in distorted evaluations. In contrast, the proposed integration function accounts for incompleteness in a balanced manner, producing more realistic, logically consistent, and interpretable results under partial information.

Table 7. New Criterion Evaluation through different integration methods

	A_1	A_2	A_3
D-number Score	{(9, 0.5), (8, 0.2)}	{(9, 0.4)}	{(9, 0.5), (8, 0.4)}
Deng Score	6.1	3.60	7.70
Wang Score	8.7143	9	8.5556
Hou Score	8.6	8.6	8.6
The Proposed Method Score	8.3221	7.38	8.5128

The ranking results after introducing the new criterion are presented in Table 8. Since this criterion was assigned a higher-than-average weight and contained substantial incompleteness, its inclusion tested the robustness of each integration function. The results show that rankings changed inconsistently across methods. Both the Deng and proposed methods resulted in a rank swap between A2 and A3; however, the Deng method exhibited a larger Δ difference ($\Delta_2 - \Delta_0 = -0.0509$) due to its excessive penalization of missing information. In contrast, the proposed method produced a smaller and more moderate change ($\Delta_2 - \Delta_0 = -0.0133$), reflecting greater stability. The rankings obtained from the Wang and Hou methods remained unchanged, indicating their insensitivity to the new criterion and their unrealistic handling of incomplete data.

Table 8. The ranking results after introducing the new criterion

Method		SC0	SC2	$\Delta_2 - \Delta_0$
Deng Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_3 > A_2$	
	Δ	0.1673	0.1164	-0.0509
Wang Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_2 > A_3$	
	Δ	0.0939	0.0808	-0.0131
Hou Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_2 > A_3$	
	Δ	0.1051	0.0869	-0.0182
The Proposed Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_3 > A_2$	
	Δ	0.1041	0.0908	-0.0133

SC3: Criteria weight modification

Following the previous two comparative analyses on data completeness and the addition of a new criterion, a third sensitivity test was conducted to examine the robustness and stability of each MCDM method under varying preference structures. In this analysis, the criteria weights were modified according to the adjustments applied in Hou and Zhao (2022) for the same illustrative example. This test assesses whether the methods maintain consistent discriminatory performance and ranking stability when decision-makers' preferences shift; an essential attribute of reliable MCDM frameworks. The original and adjusted weights are presented in Table 9, while the resulting rankings and delta indices after weight modification are summarized in Table 10.

Table 9. The original and new weights for comparison analysis

Criteria	Initial weights	Modified weights
C_1	0.1466	0.1027
C_2	0.1162	0.1511
C_3	0.0339	0.0719
C_4	0.0967	0.0222
C_5	0.0371	0.0227
C_6	0.0308	0.0289
C_7	0.0185	0.0017
C_8	0.0077	0.1213
C_9	0.0149	0.0797

C_{10}	0.0756	0.0823
C_{11}	0.0416	0.0476
C_{12}	0.0238	0.0285
C_{13}	0.0803	0.1237
C_{14}	0.1345	0.0317
C_{15}	0.0358	0.0571
C_{16}	0.1058	0.0269

Table 10. The ranking results after weight adjustment

Method		SC0	SC3	$\Delta_3 - \Delta_0$
Deng Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_2 > A_3$	
	Δ	0.1673	0.1861	0.0188
Wang Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_3 > A_2$	
	Δ	0.0939	0.0892	-0.0047
Hou Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_2 > A_3$	
	Δ	0.1051	0.1019	-0.0032
The Proposed Method	Ranking	$A_1 > A_2 > A_3$	$A_1 > A_2 > A_3$	
	Δ	0.1041	0.0969	-0.0072

The weight modification test reveals clear differences in the stability of the evaluated methods. The Deng method exhibits an increase in discriminatory power ($\Delta_3 - \Delta_0 = +0.0188$), making it the only approach with a positive change; although the ranking order remains unchanged, this heightened sensitivity to weight adjustments indicates potential instability when criteria weights are uncertain or contested. In contrast, the Wang method shows a ranking reversal between alternatives A_2 and A_3 together with a slight decrease in delta ($\Delta_3 - \Delta_0 = -0.0047$), raising concerns about its robustness under changing weights, particularly in light of its previously noted insensitivity to data incompleteness. The Hou method demonstrates a conservative response, maintaining ranking consistency with only a minimal reduction in delta ($\Delta_3 - \Delta_0 = -0.0032$), which reflects strong resistance to weight variations and aligns with its stable performance in earlier analyses. Finally, the proposed method preserves the original ranking while exhibiting a moderate decrease in discriminatory power ($\Delta_3 - \Delta_0 = -0.0072$), indicating a balanced and appropriate sensitivity to preference changes without compromising overall ranking stability.

Comparison of Full Frameworks

To assess the validity and effectiveness of the full ED-ELEC framework, the same illustrative example was solved, and the results were compared with those obtained using the D-SCRI-EDAS approach (Hou & Zhao, 2022), which integrates the Hou-based D-number aggregation with the EDAS ranking method.

Step 1: Application of ED-ELEC

The crisp values obtained with the proposed integration function were used to compute partial concordance, global concordance, and discordance indices following Eqs. (6)-(8). The indifference, preference, and veto thresholds were uniformly set to $q_j = 0.2, p_j = 0.6, v_j = 7$ for all criteria, Following common practice in the literature and consistent with the applied scoring scale. The resulting credibility matrix is presented in Table 11.

Table 11. Credibility matrix for ED-ELEC

Credibility Matrix	A_1	A_2	A_3
A_1	-	0.9641	0.8499
A_2	0.4281	-	0.7994
A_3	0.2131	0.5788	-

Step 2: Derivation of Final Ranking

The Li and Wang (2007) ranking procedure (Eqs. (10)-(12)) was applied to obtain the net credibility degree for each alternative. The results are presented in Table 12.

Table 12. Final ranking and credibility degrees using ED-ELEC

Alternative	Φ^+	Φ^-	Φ	Rank
A_1	1.814	0.6412	1.1728	1
A_2	1.2275	1.5429	-0.3154	2
A_3	0.7919	1.6493	-0.8574	3

Step 3: Comparative Discussion

A comparative summary of the D-SCRI-EDAS and ED-ELEC frameworks is presented in Table 13, showing the alternatives' raw scores, normalized scores, and corresponding delta indices. Additionally, the same weight modification analysis was performed, and the resulting outcomes are reported in Table 14.

Table 13. Comparative performance of D–SCRI–EDAS and ED-ELEC

Method		A_1	A_2	A_3	Δ_0
D-SCRI-EDAS	Scores	0.9163	0.3385	0.0979	0.8932
	Normalized Score	1	0.3694	0.1068	
	ranking	1	2	3	
ED-ELEC	Scores	1.1728	-0.3154	-0.8574	1.7311
	Normalized Score	1	-0.2689	-0.7311	
	ranking	1	2	3	

Table 14. Comparative performance of the D–SCRI–EDAS and ED-ELEC frameworks underweight modification

Method		A_1	A_2	A_3	Δ_1
D-SCRI-EDAS	Scores	0.9137	0.2529	0.0974	0.8936
	Normalized Score	1	0.2763	0.1064	
	ranking	1	2	3	
ED-ELEC	Scores	1.2371	-0.4897	-0.7474	1.6042
	Normalized Score	1	-0.3958	-0.6042	
	ranking	1	2	3	

The normalized performance of the alternatives under each method is shown in Figure 1. Comparison with D-SCRI-EDAS highlights the superior discriminatory capability and adaptive behavior of the proposed ED-ELEC framework. Although both methods produced the same ranking ($A_1 > A_2 > A_3$), confirming the validity of ED-ELEC, the proposed method achieved a higher delta index ($\Delta_0 = 1.7311$ versus 0.8932), indicating clearer differentiation among alternatives. Sensitivity analysis under weight modification showed stable rankings for both methods; however, their delta responses differed. While D-SCRI-EDAS exhibited only a negligible increase ($\Delta_1 - \Delta_0 = 0.0004$), ED-ELEC showed a moderate reduction ($\Delta_1 - \Delta_0 = -0.1269$), reflecting adaptive uncertainty awareness. Because ED-ELEC relies on the proposed balanced D-number integration function, it handles incomplete data more realistically by explicitly accounting for incompleteness rather than ignoring or overcompensating for it, resulting in controlled confidence adjustment and greater theoretical consistency in complex decision environments.

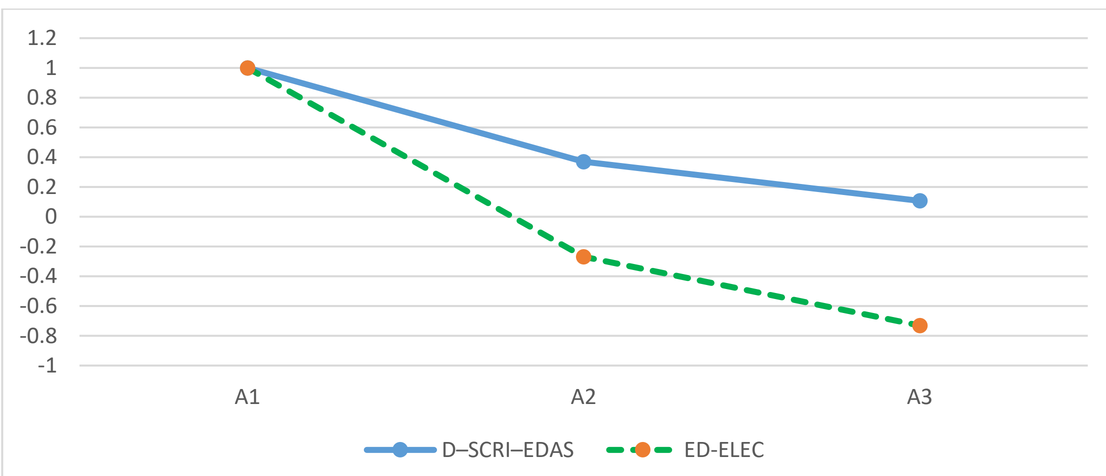

Figure 1. Performance profiles of alternatives across D–SCRI–EDAS and ED-ELEC

Summary of Illustrative Example

In summary, the comparative analyses confirm the effectiveness of both the proposed D-number integration function and the complete ED-ELEC framework. The integration function demonstrated balanced sensitivity to uncertainty and realistic evaluation of incomplete information, ensuring reliable D-number aggregation across varying data quality levels. Building upon this, the full ED-ELEC framework exhibited superior discriminatory capability, stable ranking performance underweight variation, and adaptive behavior when preference structures changed. Together, these results validate the robustness, interpretability, and practical applicability of the proposed methodological developments. Having established their validity through the illustrative example, the next section applies the ED-ELEC framework to the main case study, focusing on the evaluation of sustainable delivery systems for online food platforms.

Case Study – Evaluation of Sustainable Food Delivery Vehicles

Problem Context

Online food delivery platforms have grown rapidly in recent years, driven by increasing consumer demand for convenience and the expansion of digital marketplaces. However, this growth has also intensified sustainability concerns, particularly regarding the environmental and social impacts of delivery operations. In many cities, including the capital of Iran, Tehran, gasoline-powered motorcycles are the dominant delivery mode. They serve as the baseline scenario in this study, reflecting the current operational reality of food delivery services. Despite their ubiquity, gasoline motorcycles contribute substantially to greenhouse gas emissions, air pollution, and noise, raising concerns about their long-term sustainability (Weiss et al., 2015; Matsuyuki et al., 2024; World Economic Forum & Accenture, 2024).

To capture emerging alternatives, two additional delivery modes were selected based on both literature evidence and practical relevance:

- Electric motorcycles: Increasingly recognized as a short- to medium-term sustainable option, electric motorcycles offer significant reductions in emissions and noise while retaining compatibility with existing road infrastructure. Their feasibility has been supported in recent logistics and sustainability studies, as well as in policy initiatives promoting electric mobility in urban transport (Weiss et. al, 2015; Hollingsworth et al., 2019; Matsuyuki et al., 2024).
- Drones: Representing a more disruptive and forward-looking alternative, drones have gained considerable attention in the logistics and operations research literature as a potential solution for last-mile delivery, especially for lightweight, short-distance orders (Goodchild & toy, 2018; Stolaroff et al., 2018; Rodrigues et al., 2022; Schmidt & Saraceni, 2024). While their widespread adoption is limited by regulatory, technical, and social acceptance challenges, they are nonetheless an important benchmark for evaluating future-oriented delivery innovations (Seidakhmetov et al., 2022).

By including these three alternatives, the case study reflects a spectrum of options: the status quo (gasoline motorcycles), a feasible green transition (electric motorcycles), and a radically innovative future technology (drones). This ensures that the evaluation not only addresses current operational realities but also informs decision makers about the strategic implications of adopting sustainable and emerging delivery modes under conditions of uncertainty.

Evaluation Criteria

The selection of appropriate evaluation criteria is a crucial step in multi-criteria decision-making, as it directly shapes the relevance and validity of the outcomes. In sustainable logistics and urban mobility studies, criteria are typically derived from the triple bottom line (economic, environmental, and social dimensions), complemented by technical and operational considerations (Dekker et al., 2012; Kraus & Proff, 2021; Weiss et al., 2015; Corti et al., 2024). Following this approach, and informed by the literature on green transport and last-mile delivery, this study employs four categories of criteria.

- **Economic criteria:** purchase cost, operating cost, maintenance cost, and infrastructure cost. These capture the financial implications of adopting each delivery system, which remain central to feasibility in cost-sensitive logistics operations.
- **Environmental criteria:** CO_2 emissions, air pollution, noise pollution, and energy efficiency. These reflect the growing emphasis on reducing the ecological footprint of last-mile delivery.
- **Technical and operational criteria:** delivery speed, reliability, payload capacity, weather resistance, and range/coverage. This category balances performance attributes such as speed with operational limitations, e.g., drones' payload restrictions compared to motorcycles.
- **Social and regulatory criteria:** safety, public acceptance, and regulatory feasibility. These address broader societal and institutional concerns, including risks, acceptance, and integration within existing policy frameworks.

The sixteen criteria were derived from recent MCDM research in sustainable logistics and urban transport (Corti et al., 2024; Saha et al., 2023; Shekhovtsov et al., 2020; Weiss et al., 2015). Table 15 provides the full list of criteria together with their weights. Their relative importance is assumed a priori, based on evidence from the literature and relevant policy documents, which allows the case study to focus on the methodological contribution of handling incomplete and uncertain expert evaluations rather than on weight elicitation.

Table 15. The case study problem's criteria and their relative weights

Category	Criterion	Weights	Short explanation
Economic	C11: Purchase cost	0.08	Initial investment required to acquire the delivery system.
	C12: Operating cost	0.08	Day-to-day expenses for running the delivery system, such as fuel, electricity, or labor.
	C13: Maintenance cost	0.10	Costs associated with upkeep, repairs, and replacement of parts.
	C14: Infrastructure cost	0.06	Expenses related to supporting infrastructure, such as charging stations or docking areas.
Environmental	C21: CO_2 emissions	0.09	Amount of carbon dioxide released by the delivery system.
	C22: Air pollution	0.08	Emissions of pollutants affecting air quality, such as NOx and PM.
	C23: Noise pollution	0.02	Noise generated during delivery operations, impacting urban environments.
	C24: Energy efficiency	0.09	Effectiveness in using energy to perform delivery operations.
Technical & Operational	C31: Delivery speed	0.04	Time required to complete a delivery.
	C32: Reliability	0.08	Ability to consistently deliver goods on schedule under varying conditions.
	C33: Payload capacity	0.07	Maximum weight or volume the delivery system can carry per trip.
	C34: Weather resistance	0.05	Ability to operate safely and effectively under different weather conditions.
	C35: Range/coverage	0.04	Maximum operational distance or area covered by the delivery system.
Social & Regulatory	C41: Safety	0.06	Risk of accidents or harm to people and property during operation.
	C42: Public acceptance	0.03	Degree of societal approval or support for using the delivery system.
	C43: Regulatory feasibility	0.03	Alignment with existing laws, regulations, and policies for urban logistics.

Expert Evaluation Process

To apply the ED-ELEC framework, expert judgments were collected regarding the performance of the three delivery alternatives across the selected criteria. The expert panel was composed of specialists with backgrounds in urban logistics, aerospace engineering, and transport operations, ensuring a diversity of perspectives relevant to online food delivery systems. The weight of each expert, reflecting their relative influence in the aggregation process, is presented in Table 16.

Table 16. The Experts weights

Experts	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}
Weights (λ_k)	0.15	0.09	0.1	0.07	0.1	0.05	0.12	0.08	0.14	0.1

Experts evaluated each alternative using Saaty's 1–9 scale with odd-numbered increments, where higher values indicated better performance. This scale, shown in Table 17, is widely used in

MCDM studies because it balances usability with sufficient granularity for meaningful comparison (Saaty, 1987). In some cases, experts were unable to provide scores for certain criteria due to unavailable or inconclusive data, resulting in incomplete judgments.

Table 17. The evaluation scale used by the experts.

Score	Qualitative Meaning	Description
1	Very Poor	The alternative performs very poorly with respect to the criterion.
3	Poor	The alternative performs below average with respect to the criterion.
5	Moderate	The alternative has an average or moderate performance.
7	Good	The alternative performs well with respect to the criterion.
9	Very Good	The alternative performs exceptionally well with respect to the criterion.

To capture this heterogeneity, all expert scores for a given criterion of an alternative were aggregated into a D-number. If all experts provided scores, the D-number is complete; if even one expert left a score blank, the D-number is incomplete, formally representing uncertainty or missing information. This approach preserves the integrity of expert opinions while enabling systematic analysis within the ED-ELEC framework. To illustrate the evaluation process, an expert evaluation matrix with missing entries is presented in Table 18, reflecting cases in which the expert lacked sufficient information to assess certain criteria.

Table 18. An expert evaluation with incomplete information

Criterion	A1:Gasoline Motorcycle	A2:Electric Motorcycle	A3:Drone
C11: Purchase cost	7	--	--
C12: Operating cost	3	9	--
C13: Maintenance cost	5	5	--
C14: Infrastructure cost	7	5	1
C21: CO_2 emissions	1	7	7
C22: Air pollution	3	9	9
C23: Noise pollution	3	7	7
C24: Energy efficiency	3	7	7
C31: Delivery speed	--	--	--
C32: Reliability	9	7	--
C33: Payload capacity	7	7	1
C34: Weather resistance	7	7	3
C35: Range/coverage	7	5	--
C41: Safety	5	7	3
C42: Public acceptance	--	9	3
C43: Regulatory feasibility	7	7	1

Application of ED-ELEC

Once expert evaluations were collected, they were aggregated into a single decision matrix with D-number entries. For each alternative–criterion pair, the distinct scores assigned by the experts form the b elements of the D-number, while the corresponding v values represent the weighted proportion of experts who assigned that score. In this way, both the precise evaluations and any missing information are formally captured, producing a single D-number matrix suitable as input for the subsequent ED-ELEC analysis. The aggregated D-number matrix is presented in Table 19.

Table 19. The aggregated D-numbers of experts' evaluation

Criterion	A1:Gasoline Motorcycle	A2:Electric Motorcycle	A3:Drone
C11	{(9,0.56)(7,0.24)}	{(5,0.71)(3,0.12)}	{(5,0.23)(3,0.3)}
C12	{(5,0.29)(3,0.63)}	{(9,0.09)(7,0.83)(5,0.08)}	{(9,0.08)(7,0.1)(5,0.52)}
C13	{(7,0.15)(5,0.51)}	{(9,0.05)(7,0.63)(5,0.17)(3,0.15)}	{(9,0.08)(5,0.69)}
C14	{(7,0.66)}	{(7,0.05)(5,0.95)}	{(7,0.08)(5,0.05)(3,0.78)(1,0.09)}
C21	{(3,0.71)(1,0.21)}	{(9,0.12)(7,0.8)}	{(9,0.91)(7,0.09)}
C22	{(3,1)}	{(9,0.36)(7,0.64)}	{(9,1)}
C23	{(5,0.39)(3,0.39)}	{(9,0.51)(7,0.49)}	{(9,0.13)(7,0.73)}
C24	{(5,0.54)(3,0.24)}	{(9,0.47)(7,0.53)}	{(9,0.13)(7,0.73)}
C31	{(7,0.56)(5,0.23)}	{(7,0.62)(5,0.29)}	{(9,0.79)(7,0.12)}
C32	{(9,0.51)(7,0.41)(5,0.08)}	{(7,0.79)(5,0.21)}	{(9,0.08)(7,0.14)(5,0.47)(3,0.15)}
C33	{(9,0.71)(7,0.21)(5,0.08)}	{(9,0.61)(7,0.39)}	{(9,0.08)(5,0.24)(3,0.59)(1,0.09)}
C34	{(9,0.61)(7,0.32)}	{(9,0.13)(7,0.87)}	{(9,0.08)(5,0.61)(3,0.16)(1,0.15)}
C35	{(9,0.91)(7,0.09)}	{(9,0.08)(7,0.61)(5,0.31)}	{(9,0.08)(5,0.61)(3,0.15)}
C41	{(9,0.14)(7,0.7)(5,0.16)}	{(9,0.19)(7,0.81)}	{(9,0.08)(7,0.14)(5,0.69)(3,0.09)}
C42	{(9,0.51)(7,0.15)(5,0.07)}	{(9,1)}	{(9,0.08)(7,0.68)(5,0.15)(3,0.09)}
C43	{(9,0.64)(7,0.24)}	{(9,0.68)(7,0.24)}	{(9,0.08)(7,0.37)(5,0.21)(3,0.15)(1,0.09)}

In Table 19, consider the D-number for criterion C_{12} of alternative A_1, given as $D = \{(5{,}0.29),(3{,}0.63)\}$. This indicates that the total weight of experts assigning a score of 5 is 0.29, while the total weight of those assigning a score of 3 is 0.63. The remaining weight, corresponding to experts who did not provide a score, is $1-(0.29+0.63)=0.08$, reflecting the degree of incompleteness in the evaluation.

Once all D-numbers for the alternative–criterion pairs were established, the proposed D-numbers integration function was applied to convert each D-number into a crisp equivalent value. For the example above, the crisp value is computed as the weighted average of the scores present in the D-number, ignoring the unassigned weight:

$$x_{2,12} = \sum_{l=1}^{n} b_l v_l \times \left(1 + \frac{1-(\sum v_l)^2}{2\sum v_l}\right) = [(5\times 0.29)+(3\times 0.63)]\times\left(1+\frac{1-0.92^2}{2\times 0.92}\right) = 3.6188$$

The resulting crisp values for each alternative and criterion were then used as direct inputs to the outranking analysis. The complete decision matrix with crisp values for all alternatives and criteria is presented in Table 20.

Table 20. The final decision matrix with crisp values

Criterion	A1:Gasoline Motorcycle	A2:Electric Motorcycle	A3:Drone
C11	8.2320	4.6428	3.4407
C12	3.6188	7.0200	5.4844
C13	5.1393	6.1600	5.2723
C14	6.5954	5.1000	3.2400
C21	2.5353	7.2376	8.8200
C22	3.0000	7.7200	9.0000
C23	3.9032	8.0200	7.2308
C24	4.2785	7.9400	7.2308
C31	6.2762	6.3369	8.7009
C32	7.8600	6.5800	5.2886
C33	8.2600	8.2200	3.7800
C34	8.2915	7.2600	4.4000
C35	8.8200	6.5400	4.9595
C41	6.9600	7.3800	5.4200
C42	7.9064	9.0000	6.5000
C43	8.3937	8.4511	5.4172

The ELECTRE III procedure was then applied using the obtained crisp values. For each pair of alternatives, partial concordance and discordance indices were computed based on the predefined indifference, preference, and veto thresholds. Following common practice in the literature and in line with the applied scoring scale, these thresholds were uniformly set as $q_j = 0.2, p_j = 0.8, v_j = 8$ for all criteria. The partial and global concordance indices, as well as the discordance indices, were calculated using Eqs. (6), (7) and (8) respectively. For brevity, the partial concordance index for the Energy Efficiency criterion ($c_{24}(A_r, A_s)$), the global concordance index ($C(A_r, A_s)$), and the discordance index for the same criterion ($d_{24}(A_r, A_s)$) are presented in Tables 21, 22, and 23, respectively.

Table 21. Partial concordance

c_{24}	A_1	A_2	A_3
A_1	-	0	0
A_2	1	-	1
A_3	1	0.1513	-

Table 22. Global concordance

C	A_1	A_2	A_3
A_1	-	0.488	0.6
A_2	0.69	-	0.79
A_3	0.5	0.224	-

Table 23. Discordance

d_{24}	A_1	A_2	A_3
A_1	0	0.3974	0.2989
A_2	0	0	0
A_3	0	0	0

These indices were then combined, according to Eq. (9), to obtain the credibility index, which quantifies the extent to which one alternative can be considered to outrank another across all criteria. The resulting credibility indices are presented in Table 24.

Table 24. The credibility indices.

	A_1	A_2	A_3
A_1	-	0.3884	0.2482
A_2	0.69	-	0.79
A_3	0.4358	0.1149	-

The concordance and discordance credibility degrees were computed using Eqs. (10) and (11), and the net credibility degree was then derived using Eq. (12) to determine the final ranking. The overall ranking and corresponding credibility values are summarized in Table 25.

Table 25. The overall ranking and credibility values

Alternatives	Φ^+	Φ^-	Φ	Rank
A_1	0.6366	1.1258	-0.4892	3
A_2	1.48	0.5033	0.9767	1
A_3	0.5507	1.0382	-0.4875	2

Through this process, the ED-ELEC framework produced both a ranking of the delivery alternatives (gasoline motorcycles, electric motorcycles, drones) and diagnostic insights into the role of each criterion and the relative strengths and weaknesses of the alternatives.

Results and Discussion

The application of the ED-ELEC framework produced a ranking of the three delivery system alternatives; gasoline motorcycles, electric motorcycles, and drones; with respect to the predefined criteria. The results highlight both the relative performance of each option and the capacity of the framework to handle incomplete and uncertain evaluations.

(a) Overall Ranking

Applying the validated ED-ELEC framework to the main case study yielded the ranking results summarized in Table E. Among the three evaluated alternatives, gasoline motorcycle ($\boldsymbol{A_1}$), electric motorcycle ($\boldsymbol{A_2}$), and drone ($\boldsymbol{A_3}$), the electric motorcycle achieved the highest overall score (0.9767), ranking first. The drone followed with a score of -0.4875, ranking second, while the gasoline motorcycle ranked third with a score of -0.4892.

These results indicate that electric motorcycles represent the most sustainable and balanced option for online food delivery operations. Their strong performance reflects superior environmental and long-term economic advantages, combined with operational feasibility under existing regulatory conditions. Drones, while slightly outperforming gasoline motorcycles, remain constrained by current limitations in payload capacity, weather sensitivity, and regulatory uncertainty. Gasoline motorcycles, the current dominant mode, perform poorly in environmental and social dimensions, emphasizing the need for a transition toward cleaner delivery technologies.

(b) Performance by Criteria

To better understand the ranking outcomes, Figure 2 illustrates the performance profiles of the alternatives across the four categories of criteria (economic, environmental, technical/operational, and social/regulatory).

- Economic Criteria

  Electric motorcycles achieved the highest score under economic criteria, reflecting a balanced trade-off between purchase cost, operational efficiency, and maintenance savings. Gasoline motorcycles, while still benefiting from lower initial investment and well-established maintenance networks, present uncertain long-term cost performance due to rising fuel prices and potential emission-related restrictions. Drones, on the other hand, remain economically less favorable because of their high unit cost, limited delivery capacity, and short battery lifespan, which collectively increase operational expenses.

- Environmental Criteria

  Both electric motorcycles and drones outperform gasoline motorcycles environmentally, as they operate on battery power and produce no direct tailpipe emissions. However, their environmental impacts differ in how they influence urban traffic and related emissions. Drones show stronger performance in eliminating local pollutants but face higher embodied energy and battery waste issues relative to electric motorcycles, which have more mature battery recycling pathways and grid-charging efficiency. Gasoline motorcycles scored lowest due to fuel combustion emissions, noise, and particulate pollution.

- Technical and Operational Criteria

  Gasoline motorcycles outperformed other alternatives in technical and operational dimensions, benefiting from mature infrastructure, reliability, and weather resilience. In contrast, drones suffered from payload limitations, restricted range, operational regulations, and high sensitivity to weather conditions. Electric motorcycles provided a balanced yet moderate performance, offering acceptable reliability while still depending on evolving charging infrastructure.

- Social and Regulatory Criteria

  Electric motorcycles achieved the highest score in social and regulatory dimensions, driven by growing public acceptance, municipal incentives, and alignment with decarbonization policies. Gasoline motorcycles maintain moderate performance, benefiting from familiarity but facing declining social favorability due to noise and emissions. Drones scored lowest, hindered by safety, privacy, and public acceptance concerns, as well as stringent aviation regulations.

Overall, the criterion-level analysis reinforces the final ranking outcomes ($A_2 > A_3 > A_1$). As shown in Figure 2, electric motorcycles represent the most balanced alternative, excelling across economic, environmental, and social dimensions while maintaining operational feasibility. Drones exhibit strong environmental potential but remain constrained by technical and regulatory barriers, whereas gasoline motorcycles continue to dominate current practice primarily due to legacy advantages rather than sustainability performance.

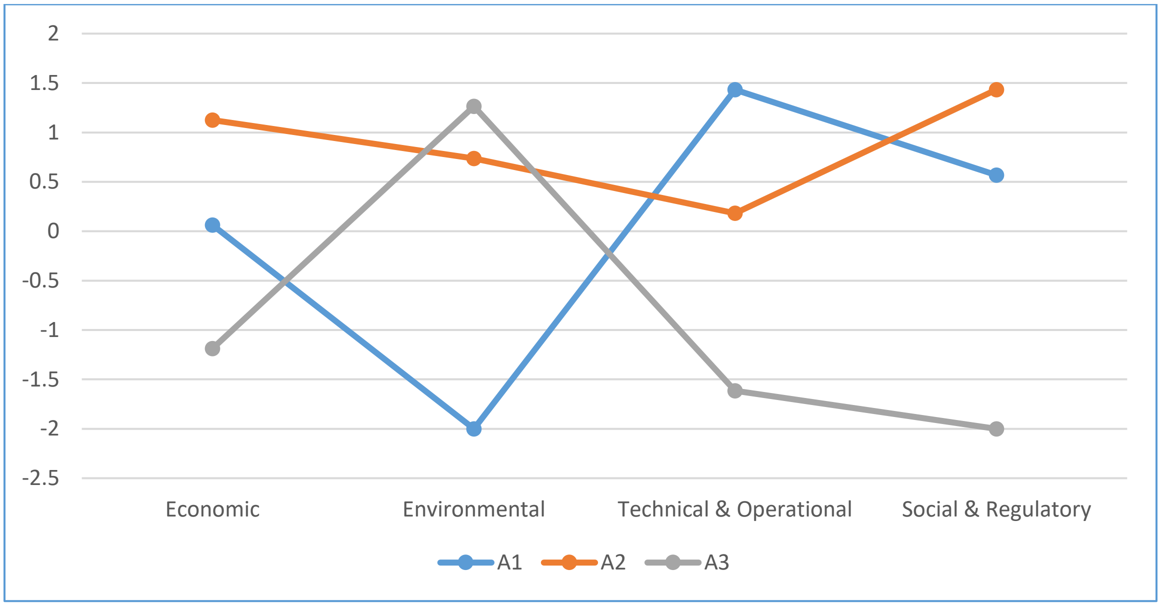

Figure 2. Performance profiles of alternatives across criteria categories.

(c) Managerial and Policy Implications

The ED-ELEC results provide several insights for logistics managers and policymakers aiming to advance sustainability in urban delivery systems.

1- Strategic Acceleration of Electric Motorcycle Adoption: Electric motorcycles ranked first, confirming their position as the most balanced and feasible solution for sustainable last-mile delivery. Their strong environmental and social performance, coupled with operational reliability, makes them the most suitable short- to medium-term alternative to conventional fleets. Policymakers can facilitate this transition by expanding charging infrastructure, offering purchase incentives, and supporting fleet electrification programs in collaboration with delivery platforms.

2- Targeted Development of Drone-Based Logistics: Although drones ranked second, their score was only marginally higher than that of gasoline motorcycles, indicating that their advantages remain limited under current technological and regulatory conditions.

Nevertheless, their zero tailpipe emissions and ability to bypass urban traffic congestion make them promising for specialized applications, such as lightweight, urgent, or remote-area deliveries. Policymakers should therefore invest in pilot projects, refine drone regulations, and develop safe low-altitude corridors to prepare for future integration as battery efficiency and payload capacity improve.

3- Managed Transition from Gasoline Fleets: Gasoline motorcycles, while operationally mature and widely available, are increasingly disadvantaged by environmental concerns and rising fossil fuel dependency. Although their short-term cost efficiency remains competitive, any increase in fossil fuel prices or carbon taxation would quickly erode their economic advantage, accelerating the shift toward electric or hybrid alternatives. Logistics managers should plan progressive fleet conversion strategies, prioritizing electrification in dense urban centers where environmental impacts are most severe.

4- Use of ED-ELEC for Evidence-Based Policy and Planning: The considerable performance gap between the top-ranked electric motorcycle and the other alternatives highlights the effectiveness of structured, data-driven evaluation methods. The ED-ELEC framework demonstrates strong capacity to manage uncertainty in expert judgment while still producing distinct and reliable rankings. This makes it particularly valuable for policy evaluation and strategic planning, where decision-makers must balance qualitative assessments with incomplete or evolving data. By integrating uncertainty-aware analysis into transport policy design, governments and logistics operators can prioritize technologies that deliver both quantifiable performance gains and long-term sustainability benefits.

In summary, the findings underscore that electric motorcycles represent the most immediate and effective pathway toward sustainable urban delivery, while drones hold significant long-term potential pending technological and regulatory progress. The ED-ELEC framework, by reliably differentiating among alternatives even under uncertainty, provides a powerful analytical foundation for both corporate strategy and public policy formulation.

Conclusion

This study introduced ED-ELEC, an enhanced multi-criteria decision-making (MCDM) framework that integrates D-numbers theory with the ELECTRE III outranking method to evaluate alternatives under incomplete and uncertain information. The framework builds upon a newly proposed D-number integration function, specifically designed to handle varying levels of incompleteness and conflicting expert judgments while preserving the structure of available knowledge.

The ED-ELEC framework was validated through a two-stage analysis. In the first stage, an illustrative example was employed to compare both (i) the proposed D-number integration function against other existing integration approaches and (ii) the full ED-ELEC framework against another D-number based MCDM method. These comparisons confirmed the superior balance, adaptability, and discriminatory capability of the proposed method. In the second stage, the validated ED-ELEC framework was applied to a real-world case study assessing sustainable delivery systems for online food platforms, where gasoline motorcycles, electric motorcycles, and drones were evaluated under economic, environmental, technical, and social criteria.

Results indicated that electric motorcycles achieved the highest overall performance, representing the most balanced and practical solution across environmental, social, and operational dimensions. Drones followed closely, showing strong environmental and congestion-reduction benefits but remaining constrained by technological maturity and regulatory challenges. Gasoline motorcycles, while reliable and cost-effective in the short term, performed weakest overall due to their environmental footprint and dependence on fossil fuels.

Overall, the findings confirm that the proposed D-number integration function effectively manages incomplete and uncertain data, and that the ED-ELEC framework delivers robust, interpretable, and policy-relevant results, offering decision-makers a reliable analytical tool for sustainable transport planning under uncertainty. Future research may extend this work by applying ED-ELEC to other sustainability-oriented domains to examine its generalizability, exploring its integration with alternative outranking methods to compare ranking behaviors, and developing software or web-based decision-support implementations to enhance the practical accessibility and adoption of the framework by practitioners and policymakers.

Data Availability Statement

Data available on request from the authors.

Acknowledgements

The authors would like to thank all participants of the present study.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding

The authors received no financial support for this article's research, authorship, and/or publication.

Ethical considerations

The authors declare no potential conflict of interest regarding the publication of this work.

References

Aljohani, K., & Thompson, R. G. (2018). A Stakeholder-Based Evaluation of the Most Suitable and Sustainable Delivery Fleet for Freight Consolidation Policies in the Inner-City Area. *Sustainability*, *11*(1), 124. https://doi.org/10.3390/su11010124

Atanassov, K. T. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, *20*(1), 87–96. https://doi.org/10.1016/S0165-0114(86)80034-3

Božanić, D., Pamučar, D., & Komazec, N. (2023). Applying D numbers in risk assessment process: General approach. *Journal of Decision Analytics and Intelligent Computing*, *3*(1), 286–295. https://doi.org/10.31181/jdaic10025122023b

Cherry, C. R., Weinert, J. X., & Xinmiao, Y. (2009). Comparative environmental impacts of electric bikes in China. *Transportation Research Part D: Transport and Environment*, *14*(5), 281–290. https://doi.org/10.1016/j.trd.2008.11.003

Corti, F., D'Alpaos, C., & Greco, L. (2024). Multicriteria evaluation of sustainability in last-mile logistics: a review. *Valori e Valutazioni*, *36*(36), 125–147. https://doi.org/10.48264/VVSIEV-20243608

Das, S., Dutta, R., De, S., & De, S. (2024). Review of multi-criteria decision-making for sustainable decentralized hybrid energy systems. *Renewable and Sustainable Energy Reviews*, *202*, 114676. https://doi.org/10.1016/j.rser.2024.114676

Dekker, R., Bloemhof, J., & Mallidis, I. (2012). Operations Research for green logistics – An overview of aspects, issues, contributions and challenges. *European Journal of Operational Research*, *219*(3), 671–679. https://doi.org/10.1016/j.ejor.2011.11.010

Dempster, A. P. (1967). Upper and Lower Probabilities Induced by a Multivalued Mapping. *The Annals of Mathematical Statistics*, *38*(2), 325–339. https://doi.org/10.1214/aoms/1177698950

Deng, X., & Jiang, W. (2019). A total uncertainty measure for D numbers based on belief intervals. *International Journal of Intelligent Systems*, *34*(12), 3302–3316. https://doi.org/10.1002/int.22195

Deng, X., Hu, Y., Deng, Y., & Mahadevan, S. (2014). Environmental impact assessment based on D numbers. *Expert Systems with Applications*, *41*(2), 635–643. https://doi.org/10.1016/j.eswa.2013.07.088

Deng, X., Hu, Y., Deng, Y., & Mahadevan, S. (2014). Supplier selection using AHP methodology extended by D numbers. *Expert Systems with Applications*, *41*(1), 156–167. https://doi.org/10.1016/j.eswa.2013.07.018

Deng, X., Lu, X., Chan, F. T. S., Sadiq, R., Mahadevan, S., & Deng, Y. (2015). D-CFPR: D numbers extended consistent fuzzy preference relations. *Knowledge-Based Systems*, *73*, 61–68. https://doi.org/10.1016/j.knosys.2014.09.007

Deng, Y. (2012). D numbers: theory and applications. *Journal of Information and Computational Science*, *9*, 2421–2428.

Fernández, E., Figueira, J. R., Navarro, J., Solares, E., & Díaz, R. (2025). Integrating second-order effects in ELECTRE methods with an interval-based approach. *Omega*, *137*, 103353. https://doi.org/10.1016/j.omega.2025.103353

Figueira, J. R., Greco, S., Roy, B., & Słowiński, R. (2013). An Overview of ELECTRE Methods and their Recent Extensions. *Journal of Multi-Criteria Decision Analysis*, *20*(1–2), 61–85. https://doi.org/10.1002/mcda.1482

Figueira, J., Greco, S., & Ehrogott, M. (2005). *Multiple Criteria Decision Analysis: State of the Art Surveys* (Vol. 78). Springer New York. https://doi.org/10.1007/b100605

French, S. (2023). Reflections on 50 years of MCDM: Issues and future research needs. *EURO Journal on Decision Processes*, *11*, 100030. https://doi.org/10.1016/j.ejdp.2023.100030

Garg, V., Niranjan, S., Prybutok, V., Pohlen, T., & Gligor, D. (2023). Drones in last-mile delivery: A systematic review on Efficiency, Accessibility, and Sustainability. *Transportation Research Part D: Transport and Environment*, *123*(April), 103831. https://doi.org/10.1016/j.trd.2023.103831

Goodchild, A., & Toy, J. (2018). Delivery by drone: An evaluation of unmanned aerial vehicle technology in reducing CO 2 emissions in the delivery service industry. *Transportation Research Part D: Transport and Environment*, *61*, 58–67. https://doi.org/10.1016/j.trd.2017.02.017

Govindan, K., Soleimani, H., & Kannan, D. (2015). Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. *European Journal of Operational Research*, *240*(3), 603–626. https://doi.org/10.1016/j.ejor.2014.07.012

Hashemi, S. S., Hajiagha, S. H. R., Zavadskas, E. K., & Mahdiraji, H. A. (2016). Multicriteria group decision making with ELECTRE III method based on interval-valued intuitionistic fuzzy information. *Applied Mathematical Modelling*, *40*(2), 1554–1564. https://doi.org/10.1016/j.apm.2015.08.011

Hollingsworth, J., Copeland, B., & Johnson, J. X. (2019). Are e-scooters polluters? The environmental impacts of shared dockless electric scooters. *Environmental Research Letters*, *14*(8), 084031. https://doi.org/10.1088/1748-9326/ab2da8

Hou, H., & Zhao, C. (2021). A Novel Integration Method for D Numbers Based on Horizontal Comparison. *Axioms*, *10*(4), 312. https://doi.org/10.3390/axioms10040312

Hou, H., & Zhao, C. (2022). A Novel D–SCRI–EDAS Method and Its Application to the Evaluation of an Online Live Course Platform. *Systems*, *10*(5), 157. https://doi.org/10.3390/systems10050157

Kraus, L., & Proff, H. (2021). Sustainable Urban Transportation Criteria and Measurement—A Systematic Literature Review. *Sustainability*, *13*(13), 7113. https://doi.org/10.3390/su13137113

Li, H. F., & Wang, J.J. (2007). An Improved Ranking Method for ELECTRE III. *In Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing*, 6659–6662. IEEE. https://doi.org/10.1109/WICOM.2007.1634

Liang, W., Rodríguez, R. M., Wang, Y.-M., Goh, M., & Ye, F. (2023). The extended ELECTRE III group decision making method based on regret theory under probabilistic interval-valued hesitant fuzzy environments. *Expert Systems with Applications*, *231*, 120618. https://doi.org/10.1016/j.eswa.2023.120618

Liu, P., Zhu, B., & Wang, P. (2021). A weighting model based on best–worst method and its application for environmental performance evaluation. *Applied Soft Computing*, *103*, 107168. https://doi.org/10.1016/j.asoc.2021.107168

Liu, S., & Lin, Y. (2011). *Grey systems: Theory and applications* (Vol. 68). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-16158-2

Lohmann, P. M., Gsottbauer, E., Farrington, J., Human, S., & Reisch, L. A. (2024). Choice architecture promotes sustainable choices in online food-delivery apps. *PNAS Nexus*, *3*(10). https://doi.org/10.1093/pnasnexus/pgae422

Mardani, A., Jusoh, A., & Zavadskas, E. K. (2015). Fuzzy multiple criteria decision-making techniques and applications – Two decades review from 1994 to 2014. *Expert Systems with Applications*, *42*(8), 4126–4148. https://doi.org/10.1016/j.eswa.2015.01.003

Matsuyuki, M., Chotib, Nurhasana, R., Antokida, Y., Fitrinitia, I. S., Made Shellasih, N., & Rizky Ningtyas, F. (2024). Assessment of the sustainability of online food delivery from the perspective of CO2 emissions and transport intensity: A case study in Jakarta. *Transportation Research Interdisciplinary Perspectives*, *27*, 101200. https://doi.org/10.1016/j.trip.2024.101200

Montazer, G. A., Saremi, H. Q., & Ramezani, M. (2009). Design a new mixed expert decision aiding system using fuzzy ELECTRE III method for vendor selection. *Expert Systems with Applications*, *36*(8), 10837–10847. https://doi.org/10.1016/j.eswa.2009.01.019

Nur, F., Alrahahleh, A., Burch, R., Babski-Reeves, K., & Marufuzzaman, M. (2020). Last mile delivery drone selection and evaluation using the interval-valued inferential fuzzy TOPSIS. *Journal of Computational Design and Engineering*, *7*(4), 397–411. https://doi.org/10.1093/jcde/qwaa033

Özaşkın, A., & Görener, A. (2023). An integrated multi-criteria decision-making approach for overcoming barriers to green supply chain management and prioritizing alternative solutions. *Supply Chain Analytics*, *3*, 100027. https://doi.org/10.1016/j.sca.2023.100027

Qi, J., Hu, J., & Peng, Y. (2025). Interval D-preference-based VIKOR for multiple-criteria group design evaluation with imprecise and unreliable information. *Advanced Engineering Informatics*, *64*, 103058. https://doi.org/10.1016/j.aei.2024.103058

Rodrigues, T. A., Patrikar, J., Oliveira, N. L., Matthews, H. S., Scherer, S., & Samaras, C. (2022). Drone flight data reveal energy and greenhouse gas emissions savings for very small package delivery. *Patterns*, *3*(8), 100569. https://doi.org/10.1016/j.patter.2022.100569

Roy, B. (1991). The outranking approach and the foundations of ELECTRE methods. *Theory and Decision*, *31*(1), 49–73. https://doi.org/10.1007/BF00134132

Saaty, R. W. (1987). The analytic hierarchy process-what it is and how it is used. In *Mathematical Modelling*, *9*(3-5), 161-179. https://doi.org/10.1016/0270-0255(87)90473-8

Safaei Ghadikolaei, A., & Gholamrezatabar Divkolaei, Z. (2014). Determinants of framework for assessing the sustainability of food supply chains using fuzzy analytic network process (Case study: Selected meat production companies of Mazandaran). *Industrial Management Journal*, *6*(3), 535-554. https://doi.org/10.22059/imj.2014.50704

Saha, A., Simic, V., Senapati, T., Dabic-Miletic, S., & Ala, A. (2023). A Dual Hesitant Fuzzy Sets-Based Methodology for Advantage Prioritization of Zero-Emission Last-Mile Delivery Solutions for Sustainable City Logistics. *IEEE Transactions on Fuzzy Systems*, *31*(2), 407–420. https://doi.org/10.1109/TFUZZ.2022.3164053

Schmidt, S., & Saraceni, A. (2024). Consumer acceptance of drone-based technology for last mile delivery. *Research in Transportation Economics*, *103*, 101404. https://doi.org/10.1016/j.retrec.2023.101404

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press. https://doi.org/10.2307/j.ctv10vm1qb

Shekhovtsov, A., Kozlov, V., Nosov, V., & Sałabun, W. (2020). Efficiency of Methods for Determining the Relevance of Criteria in Sustainable Transport Problems: A Comparative Case Study. *Sustainability*, *12*(19), 7915. https://doi.org/10.3390/su12197915

Sotoudeh-Anvari, A. (2024). A state-of-the-art review on D number (2012-2022): A scientometric analysis. *Engineering Applications of Artificial Intelligence*, *127*, 107309. https://doi.org/10.1016/j.engappai.2023.107309

Stolaroff, J. K., Samaras, C., O'Neill, E. R., Lubers, A., Mitchell, A. S., & Ceperley, D. (2018). Energy use and life cycle greenhouse gas emissions of drones for commercial package delivery. *Nature Communications*, *9*(1), 409. https://doi.org/10.1038/s41467-017-02411-5

Wang, C.-N., Nguyen, N.-A.-T., Dang, T.-T., & Hsu, H.-P. (2021). Evaluating Sustainable Last-Mile Delivery (LMD) in B2C E-Commerce Using Two-Stage Fuzzy MCDM Approach: A Case Study from Vietnam. *IEEE Access*, *9*, 146050–146067. https://doi.org/10.1109/ACCESS.2021.3121607

Wang, N., Liu, X., & Wei, D. (2018). A Modified D Numbers' Integration for Multiple Attributes Decision Making. *International Journal of Fuzzy Systems*, *20*(1), 104–115. https://doi.org/10.1007/s40815-017-0323-0

Weiss, M., Dekker, P., Moro, A., Scholz, H., & Patel, M. K. (2015). On the electrification of road transportation – A review of the environmental, economic, and social performance of electric two-wheelers. *Transportation Research Part D: Transport and Environment*, *41*, 348–366. https://doi.org/10.1016/j.trd.2015.09.007

World Economic Forum, & Accenture. (2024, December). *Transforming urban logistics: Sustainable and efficient last-mile delivery in cities* [White paper]. World Economic Forum. https://reports.weforum.org/docs/WEF_Transforming_Urban_Logistics_2024.pdf

Yaftiyan, F., Saghafi, F., & Hosseinzadeh, M. (2025). Analyzing Key Variables in Recurrent Carbon Reduction Policies Using a Hybrid Approach: A Focus on Pharmaceutical Distributors in Iran. *Industrial Management Journal*, *17*(2), 1-26. https://doi.org/10.22059/imj.2025.392285.1008233

Yusoff, B., Pratama, D., Kilicman, A., & Abdullah, L. (2023). Circular Intuitionistic Fuzzy ELECTRE III Model for Group Decision Analysis. *Informatica*, 881–908. https://doi.org/10.15388/23-INFOR536

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X