Journal of Urban Mobility 10 (2026) 100247

Contents lists available at ScienceDirect

Journal of Urban Mobility

journal homepage: www.elsevier.com/locate/urbmob

Citizen-based decision support for sustainable public transport system selection: Evidence from Vietnam

Roberto Ventura, Martina Carra *, Benedetto Barabino

Department of Civil Engineering, Architecture, Land, Environment and Mathematics (DICATAM), University of Brescia, via Branze 43, 25123, Brescia, Italy

ARTICLE INFO

Keywords:
Developing countries
Multi-criteria decision-making methods
Public transport

ABSTRACT

Over the last decade, Multi-Criteria Decision-Making Methods (MCDMMs) have been widely applied in public transport to support alternative selection. However, despite citizens being primary users, their opinions are rarely considered in choosing transit modes. Moreover, the variety of MCDMMs can make method selection challenging, and, to the best of the authors' knowledge, no prior research in transportation has quantitatively compared multiple MCDMMs under the noise effect, an essential factor for assessing decision robustness in participatory contexts.

This study addresses these gaps by proposing an integrated multi-method decision framework combining Analytic Hierarchy Process (AHP) for citizen-based weighting with Simple Additive Weighting (SAW), Techniques of Order of Preference by Similarity to Ideal Solution (TOPSIS), *VIekriterijumsko KOmpromisno Rangiranje* (VIKOR), and Preference Ranking Organization METHod for Enrichment of Evaluations II (PROMETHEE II) for alternative ranking. In addition, two quantitative stability metrics—score-based and ranking-based—are introduced to assess methodological robustness under stochastic weight perturbations generated through Monte Carlo simulations.

The framework reduces single-method bias by integrating multiple MCDMMs, while noise analysis ensures reliability under data uncertainty, particularly when citizen-based inputs inform urban mobility strategies. Its viability is tested in Ho Chi Minh City in Vietnam, a rapidly urbanizing megacity under pressure to promote sustainable mobility choices. Results show that assessing both metrics offers a fuller picture of robustness, as methods may be stable in scores or unstable in rankings.

The consistent patterns observed in the study support the methodological transferability of the framework, offering a structured benchmark for planners, practitioners, and academics assessing public transport system alternatives in comparable urban contexts.

1. Introduction

1.1. Background

Multiple Criteria Decision-Making Methods (MCDMMs) have been developed to support decision-making across many aspects (e.g., Zavadskas and Turskis, 2011; Beria et al., 2012; Macharis & Bernardini, 2015; Donais et al., 2019; Yannis, 2020; Keshavarz-Ghorabaee et al., 2022). They are techniques for analyzing problems involving a set of main criteria, sub-criteria, and related parameters, and to reflect stakeholders' subjective perspectives on a problem (e.g., Bozorg-Haddad et al., 2021). However, in public transport planning, decisions are not only technical but also social, as they directly affect citizens' daily lives. International guidelines, such as those of the Sustainable Urban Mobility Plans (SUMPs), emphasize participatory approaches to ensure that mobility solutions are widely accepted and sustainable (Rupprecht Consult et al., 2019). This growing focus on citizens' desires makes MCDMMs particularly relevant, as they offer rational, flexible tools for incorporating diverse priorities into transport planning.

From a theoretical standpoint, these developments echo with broader strands of decision theory that emphasize bounded rationality, pluralistic preference structures, and the need for decision support tools capable of accommodating heterogeneous value systems (Keeney & Raiffa, 1993; Simon, 1955). Likewise, participatory governance

* Corresponding author at: Department of Civil Engineering, Architecture, Land, Environment and Mathematics (DICATAM), University of Brescia, Via Branze, 43, 25123 Brescia, Italy.

E-mail addresses: roberto.ventura@unibs.it (R. Ventura), martina.carra@unibs.it (M. Carra), benedetto.barabino@unibs.it (B. Barabino).

https://doi.org/10.1016/j.urbmob.2026.100247
Received 19 January 2026; Received in revised form 12 May 2026; Accepted 16 June 2026
Available online 3 July 2026

research highlights that transparency, inclusiveness, and deliberation are essential components of legitimate mobility planning, and that analytical tools must be adapted to meaningfully integrate stakeholder diversity and support fair and transparent participatory processes, particularly when large and heterogeneous groups of stakeholders are involved in collective decision contexts (Fung, 2006; Innes & Booher, 2010; Huang et al., 2025).

1.2. Literature review

This literature review follows a structured progression. It first outlines the main methodological families of MCDMMs and their recent hybrid, fuzzy, and AI-enhanced extensions. It then examines how these approaches have been applied in public transport planning and evaluation, before discussing the emerging recognition of participatory and socio-technical dimensions in transport decision-making.

MCDMMs are traditionally categorized into three main approaches: i) value-based methods, ii) goal-based methods, and iii) outranking methods (e.g., Ozsahin et al., 2021). This tripartition represents the foundational structure of multi-criteria analysis and continues to guide most applications in transportation.

Value-based methods rely on the partial or complete compensation of contributing factors. For instance, in the context of a bus route, strong performance in user services (e.g., high frequency) may compensate for weak performance in operational costs. Broniewicz & Ogrodnik (2020) highlight Simple Additive Weighting (SAW), Analytic Hierarchy Process (AHP), and Multi-Attribute Utility Theory (MAUT) as among the most popular methods. The fundamental basis of SAW is to sum the weighted criterion performance of each alternative and select the best one (Ciardiello & Genovese, 2023). The AHP is also a widely known method (Saaty, 1980a, b). It reorganizes the problem into branches with a hierarchical structure and helps the decision-makers break down and address the decision issues systematically. It can manage both qualitative and quantitative factors, making it particularly useful in complex decision-making environments. Moreover, AHP was extended to Group AHP (GAHP) when multiple decision-makers were involved, ensuring a more collaborative approach to evaluating alternatives (Aczél & Saaty, 1983). Advanced models have been developed to aggregate individual preferences, approximate evaluators' intentions, and enhance scoring accuracy (e.g., Wu & Tu, 2021; Du & Shan, 2021). AHP's flexibility and robustness make it a key component of MCDMMs across various fields, including transport planning, urban planning, resource allocation, and strategic management (e.g., Duleba & Moslem, 2021).

Goal-based methods, such as goal or compromise programming, assess alternatives against a desired or ideal solution, aiming to minimize deviation or distance from this goal. Consequently, the compromise solution selected is the one "closest to the ideal". Among these methods, the Techniques of Order of Preference by Similarity to Ideal Solution (TOPSIS) and *VIekriterijumsko KOmpromisno Rangiranje* (VIKOR) are among the most widely adopted (e.g., Broniewicz & Ogrodnik, 2020). The TOPSIS method considers the furthest distance to the worst scenario and the closest distance to the best scenario, while the VIKOR method considers the maximum utility index and minimum regret index (Chatterjee & Chakraborty, 2016; Ciardiello & Genovese, 2023).

Outranking methods[1] compare pairs of alternatives for each criterion to establish a preference. Once these preferences are combined, these methods favour selecting one alternative over another. They can accommodate various preferences, such as linear, nonlinear, and threshold, without requiring a standard scale or unit for criteria. Additionally, they enable partial or vague comparisons, mirroring human judgment and decision-making processes. However, they have drawbacks, including the complexity of parameters, weights, and calculations that may not be straightforward, understandable, or communicable, especially to stakeholders or non-experts. Additionally, outranking techniques may only present a set of unrelated or undifferentiated options rather than a clear ranking of alternatives, necessitating further investigation or discretion. These methods vary based on how preferences are aggregated. The most applied methods are *ELimination Et Choix Traduisant la REalite* (ELECTRE) - (Roy, 1990) and Preference Ranking Organization METhod for Enrichment of Evaluations (PROMETHEE) methods (Brans et al., 1986; Brans & De Smet, 2016).

While the classical tripartition still defines the core families of MCDMMs, recent research has progressively broadened the methodological landscape to include Artificial Intelligence (AI)-enhanced, hybrid, and fuzzy frameworks (Li et al., 2021; Kundu et al., 2024; Ngossaha et al., 2024; Broniewicz & Ogrodnik, 2025; Kilic & Demirel, 2025; Kumar & Pamucar, 2025). These developments largely build on established MCDMM logics, extending them to better accommodate non-linearity, ambiguity, and data-intensive evaluation contexts. Within this expanded landscape, an emerging stream integrates machine-learning components with traditional MCDMM structures to enhance both performance assessment and interpretability. Notable examples include explainable Data Envelopment Analysis (DEA)-based approaches that combine efficiency analysis with transparent attribution of performance drivers, and hybrid architectures that couple ranking methods such as TOPSIS with learning models (e.g., Wu, 2023; Lee, 2024). In this vein, Lee (2024) has shown how machine-learning algorithms (e.g., XGBoost) can be embedded within MCDMM workflows and complemented by post-hoc explainability tools such as *SHapley Additive exPlanations* (SHAP) to support decision understanding and communication. Alongside AI-enhanced formulations, hybrid frameworks combining fuzzy modelling, risk-based reasoning, and conventional MCDMM logic have proliferated, particularly in transport applications characterised by subjective judgments and uncertain criteria (e.g., Koohathongsumrit & Chankham, 2022, 2023; Demir & Moslem, 2025). These approaches typically integrate components from multiple MCDMMs through sequential stages embedded within a single, coherent workflow. Representative examples include hybrid fuzzy risk-assessment models merging incenter-of-centroid formulations, fuzzy AHP, and VIKOR (Koohathongsumrit & Chankham, 2022), as well as subsequent extensions combining fuzzy risk assessment with the *Best-Worst Method* (BWM) and *Measurement of Alternatives and Ranking according to the COmpromise Solution* (MARCOS) (Koohathongsumrit & Chankham, 2023).

Building on this evolution, emerges that value-based methods remain central in many hybrid MCDMM designs, especially when paired with fuzzy or entropy-based weighting (Aydin et al., 2025). Goal-based methods have likewise expanded through fuzzy TOPSIS and VIKOR variants to address ambiguous or non-linear preferences (Kaya et al., 2020; Tian et al., 2021; Nguyen et al., 2026), while outranking approaches have progressed through multi-hesitant fuzzy ELECTRE models, simplified ELECTRE–CRITIC hybrids (Peng et al., 2019; Chen et al., 2022), and fuzzy-PROMETHEE adaptations for decisions requiring high robustness under subjective uncertainty (Moradi et al., 2023). Collectively, these advances show how hybrid and fuzzy-based extensions now span all three classical MCDMM families, further complicating the methodological landscape and reinforcing the need for systematic cross-method comparisons.

Taken together, these classical and hybrid formulations form the core methodological foundation of contemporary transport applications, as evidenced by their broad and growing empirical use.

Usually, MCDMMs are applied in public transport to (i) select project

[1] Note that the definition of outranking can be determined as "*Option A outranks Option B if there are enough arguments to decide that A is at least as good as B* (called "concordance" principle, i.e., majority of criteria support), *while there is no overwhelming reason to refute that statement* (called "non-discordance" principle, i.e., no criterion is strongly opposed to)" according to Yannis et al. (2020). Therefore, its approach is also the pairwise comparison among alternatives.

alternatives, (ii) evaluate the quality of services, and (iii) select routes. For instance, as for (i), Le Pira et al. (2015) used the AHP and the Pairwise Majority Rule (PMR) to determine preferences among transport alternatives for Catania University Students. Hamurcu & Eren (2020) ranked various types of electric buses using AHP and TOPSIS. Similarly, Borghetti et al. (2024) applied AHP, ELECTRE I, and SAW to evaluate the best alternative fuel options for urban and interurban bus fleets, with a focus on cost competitiveness. Awasthi et al. (2018) used fuzzy TOPSIS, VIKOR, and Grey Relational Analysis (GRA) to evaluate the sustainability of urban mobility projects, comparing a new tramway, bus line reorganization, and electric car-sharing in Luxembourg City, highlighting MCDMMs role in project selection. Manzolli et al. (2021) combined the PROMETHEE method with scenario-based models to prioritize three modes of public transport, including Metro, Light Rail Transit, and Bus Rapid Transit. MCDMMs have also been applied to assess the feasibility of non-traditional transit systems, such as Tram-Train via AHP and ELECTRE (Assefa et al., 2024). As for (ii), Moslem et al. (2020) used the AHP and BWM to evaluate public transport quality concerning four main criteria: service quality, transport quality, tractability, and fare. Shabani et al. (2022) employed a combination of the Delphi method, BWM, and fuzzy-TOPSIS method to assess public transport user satisfaction, based on a survey conducted among approximately 400 individuals. Similarly, Nassereddine & Eskandari (2017) investigated passenger satisfaction through surveys and applied Delphi, GAHP, and PROMETHEE to evaluate and improve customer satisfaction in Tehran's public transportation system. Recent Fuzzy-PROMETHEE applications enhanced robustness under uncertain user judgments, allowing more stable prioritization of service-quality attributes even when passenger preferences are vague or imprecise (Moradi et al., 2023).

Finally, as for (iii), Güner (2018) proposed an approach that includes the AHP and TOPSIS methods to rank the bus routes in Sakarya from the user's viewpoint. Noureddine & Ristic (2019) developed an integrated model of FUCOM-TOPSIS-MABAC to support route planning for hazardous materials transport. Morfoulaki & Papathanasiou (2021a) implemented the PROMETHEE method to rank sustainable criteria in urban mobility planning.

Beyond methodological and application-oriented developments, an expanding body of transport research has recognised that evaluation outcomes are influenced not only by the choice of decision models but also by whose preferences inform them. In this vein, citizen and stakeholder participation has increasingly been discussed as a core element of contemporary transport planning, particularly regarding transparency, procedural legitimacy, and public acceptance of proposed interventions (e.g., Rupprecht Consult et al., 2019; Morfoulaki & Papathanasiou, 2021b). Participatory and deliberative assessments have highlighted the broader value of public involvement beyond technical optimisation. Although citizen input is not necessarily expected to improve the technical quality of decisions, it can contribute contextual knowledge, experiential insights, and perceptions of fairness that complement expert-driven evaluations, thereby shaping how transport alternatives are perceived and accepted (Cain, Gerber & Hui, 2020; Linovski & Baker, 2023). Moreover, understanding customers' preferences can help transit operators attract more users and alleviate pressure on traffic systems in congested corridors (Taylor & Morris, 2015; Duleba & Moslem, 2021). Therefore, from a planning perspective, transport-related decisions are widely understood not merely as technical ranking exercises, but as socio-technical processes with implications for institutional credibility and the long-term sustainability of mobility strategies.

1.3. Literature gaps

MCDMMs continue to evolve, with each approach showing specific strengths and limitations and supporting an increasingly wide set of applications (Velasquez & Hester, 2013; Broniewicz & Ogrodnik, 2020). Yet, despite this methodological expansion, the theoretical landscape remains fragmented.

First, despite the growing recognition of citizens and stakeholder participation in transport planning, its integration into quantitative MCDMM frameworks remains limited. Experts are the primary source of criterion weighting, while citizen input is typically confined to consultative or qualitative stages rather than playing a central role in the formal evaluation process. As a result, expert-defined priorities are rarely validated against user-derived preferences, leaving open the question of whether resulting rankings adequately reflect lived mobility needs. Therefore, examining citizen-derived preferences represents an important and underexplored line of inquiry.

Second, while individual MCDMMs have been widely applied in transport project evaluation, a gap remains in addressing the multidimensional trade-offs involved in system selection. Most studies rely on a single method, or at most a pair, which may fail to capture the full range of compensatory, compromise-based, and outranking logics underlying complex mobility decisions. Since each MCDMM is based on distinct assumptions about how alternatives should be compared, relying on one technique risks introducing method-specific biases into rankings. Meanwhile, the proliferation of hybrid formulations has expanded modelling possibilities while increasing methodological heterogeneity. Importantly, these hybrids typically operate sequentially rather than applying full MCDMMs in parallel. Although this enhances flexibility, it does not provide a structured basis for contrasting distinct decision logics or for understanding how different architectures may influence outcomes. For these reasons, there is value in adopting a unified framework in which multiple well-established MCDMMs are applied transparently and in parallel. A multi-method perspective enables cross-validation across evaluative logics, reduces methodological artefacts, and offers a more comprehensive representation of trade-offs by combining value-based, distance-to-ideal, compromise, and outranking approaches (Velasquez & Hester, 2013; Yannis, 2020). Convergences across methods strengthen confidence that results reflect the structure of the decision problem, whereas divergences highlight dimensions that are particularly sensitive to specific decision logics or weighting patterns.

Third, an additional challenge arises when MCDMMs are applied in participatory contexts, where criterion weights may be affected by noise, intended as random fluctuations or small inconsistencies in respondents' judgments linked to heterogeneous backgrounds, uneven experience, or cognitive fatigue. Citizen-provided weights thus might exhibit greater variability than expert-based assessments, increasing the sensitivity of decisions to perturbations. Understanding how different MCDMMs respond to such noise is essential for ensuring credible and robust rankings in participatory settings. However, to the authors' knowledge, systematic comparisons of multiple MCDMMs under stochastic weight perturbations remain scarce in transport research. Existing studies typically focus on selecting the optimal alternative and represent uncertainty through discrete, scenario-based weight changes (e.g., best-case/worst-case settings, percentage shifts, expert-defined variations), rather than probabilistic deviations that reflect heterogeneous citizen inputs (Nassereddine & Eskandari, 2017; Awasthi et al., 2018; Ghorbanzadeh et al., 2018; Kalifa et al., 2021). Resulting ranking changes are usually assessed qualitatively, without formal stability metrics and through only a handful of scenarios. Although recent fuzzy, hesitant-fuzzy, and hybrid approaches address imprecision in linguistic or hesitant judgments (Peng et al., 2019; Kaya et al., 2020; Chen et al., 2022; Moradi et al., 2023), they model structured ambiguity rather than the random respondent-level variability characteristic of noise. Therefore, they do not provide a stochastic robustness assessment of how different MCDMMs behave when weights vary unpredictably. Evidence outside transport confirms this gap: Kolios et al. (2016) introduce stochasticity in alternative performances but not in weights or method comparisons, while Fengmin et al. (2025) analyse several MCDMMs using scenario-based, not simulation-based, weight perturbations. Altogether, the lack of stochastic perturbations, quantitative stability metrics, and explicit cross-method comparisons limits our understanding of how resilient different MCDMMs are to noisy input data.

To summarize, current research highlights three interconnected gaps: the limited use of citizen-derived preferences in MCDMMs, the lack of systematic comparisons across different decision logics, and the absence of rigorous noise-based robustness analyses. Despite the versatility of MCDMMs, the field still lacks a unified perspective clarifying how heterogeneous or noisy preferences interact with the diverse assumptions and decision rules of individual methods.

1.4. Paper objective

This paper addresses the gaps identified above by developing an integrated decision framework for selecting the most suitable public transport alternative in a given urban context. The framework combines five theoretically distinct and widely recognised MCDMMs, i.e., AHP, SAW, TOPSIS, VIKOR, and PROMETHEE II, to capture the full range of compensatory, compromise-based and outranking logics traditionally employed in multi-criteria evaluation (Mardani et al., 2015; Greco et al., 2016; Yannis et al., 2020; Ozsahin et al., 2021). These methods embody different assumptions regarding aggregation, compensability, and dominance, providing complementary analytical perspectives. Applying (at least) one representative method from each major MCDM family in parallel mitigates method-specific biases and can support a richer interpretation of the multidimensional trade-offs inherent in transport system selection. The use of well-established and widely documented MCDMMs enhances interpretability and communicability for planners (Macharis & Bernardini, 2015), while evidence from other fields underscores the value of joint multi-method evaluations (e.g., Ahmed et al., 2024).

Within this framework, alternatives are defined as system-level public transport options (e.g., Bus, Bus Rapid Transit, Light Rail, and Metro) that constitute the unit of analysis for the decision problem. Operational and service-related characteristics are instead treated as evaluation criteria that describe how each system performs across multiple dimensions. Accordingly, the “most suitable” alternative is identified by combining citizens’ criterion weights, elicited via a survey, with literature-based performance indicators for each option. The resulting evaluation, therefore, reflects a planning-level comparison of predefined system alternatives rather than a strategic policy-level choice.

A distinctive feature of the framework is its dual-source structure: user-driven weights capture societal priorities, whereas the performance matrix draws on technical, quantitative, and qualitative data on system characteristics. This ensures that the evaluation balances societal preferences with evidence-based assessments of transport alternatives. AHP derives the relative importance of criteria from user feedback, and these weights are then applied by other MCDMMs in their scoring and ranking procedures, where they are combined with each alternative's objective performance values.

Because survey-based weights inevitably reflect heterogeneity and noise, the framework includes a dedicated sensitivity analysis that evaluates the stability of each method under stochastic perturbations using novel score- and ranking-based metrics. This allows the most robust method to be identified and helps mitigate potential distortions arising from inconsistent user inputs. Notably, the framework is not designed to outperform individual techniques in accuracy or computational efficiency, but to offer a broader and more transparent evaluative perspective that supports examination of ranking stability under different decision logics and heterogeneous, potentially noisy citizen-derived weights.

The integrated framework is implemented to assess public transport alternatives, ranging from basic to high-performance, through a case study of a rapidly growing megacity (Ho Chi Minh City - HCMC) in a developing country (Vietnam). HCMC represents a highly relevant application setting, as urban mobility is still dominated by motorcycles, while public authorities are simultaneously pursuing the expansion of public transport and the development of a mass rapid transit network (Centre for Liveable Cities & Urban Land Institute, 2017; Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), 2024; Transformative Urban Mobility Initiative, 2024). In this transitional context, citizens' preferences play a critical role, making HCMC an appropriate setting to test the adaptability and robustness of a participatory multicriteria framework. The quantitative dimension of these challenges is illustrated in Section 3 through context-specific indicators that provide the empirical motivation for applying the proposed framework. Accordingly, the HCMC application should be interpreted as an illustrative and informative case study rather than as a prescriptive policy recommendation. The primary contribution of the paper lies in the proposed decision-support framework itself, while the case study demonstrates its applicability and behaviour under realistic planning conditions.

The remaining paper is structured as follows: Section 2 illustrates the methodological framework; Section 3 presents the application of the framework in HCMC; Section 4 discusses the results in the context of the literature; Section 5 provides conclusions and paves the way for future research.

2. Methodological framework

The integrated framework has been designed as a decision-making tool to help citizens choose the most suitable type of public transport for their contexts. The overall framework is organised into four stages, each subdivided into specific steps and sub-steps, as illustrated in Figure 1. Specifically, Figure 1 summarises the proposed methodological workflow, showing the integration of AHP, multiple MCDMMs, and Monte Carlo simulation to rank alternatives and assess the robustness of the results.

STAGE 1 defines the sets of criteria and alternatives. **STAGE 2** weights the relative importance of evaluation criteria with the AHP. **STAGE 3** applies SAW, TOPSIS, VIKOR, and PROMETHEE II methods to prioritise the alternatives. Finally, **STAGE 4** performs a sensitivity analysis to assess the stability of rankings under varying weight variations.

The content of each stage is detailed in the following sections.

2.1. Defining sets of criteria and alternatives

STAGE 1 is divided into two steps: the criteria definition (**STEP 1.1**) and the alternative definition (**STEP 1.2**).

Defining a set of criteria (**STEP 1.1**) is a fundamental step in the decision-making process, as they provide the basis for measuring, evaluating, and ranking alternatives. Criteria reflect decision-makers’ views on the problem by capturing the characteristics used to assess different options. To improve interpretability and manage complexity, criteria can be further divided into sub-criteria. This decomposition makes relevant performance aspects more manageable.

Once the evaluation criteria are defined, **STEP 1.2** specifies the set of alternatives to be assessed. Alternatives reflect the objective of the decision-making process and must be measurable with respect to the selected criteria. In this study, since the aim is to select public transport alternatives and compare results across different MCDMMs, the alternatives correspond to system-level public transport options characterised by distinct performance profiles.

2.2. Weighting criteria

STAGE 2 weighs the decisional criteria.

More generally, this stage assesses the relative importance of the evaluation criteria. Differences in individual opinions, shaped by

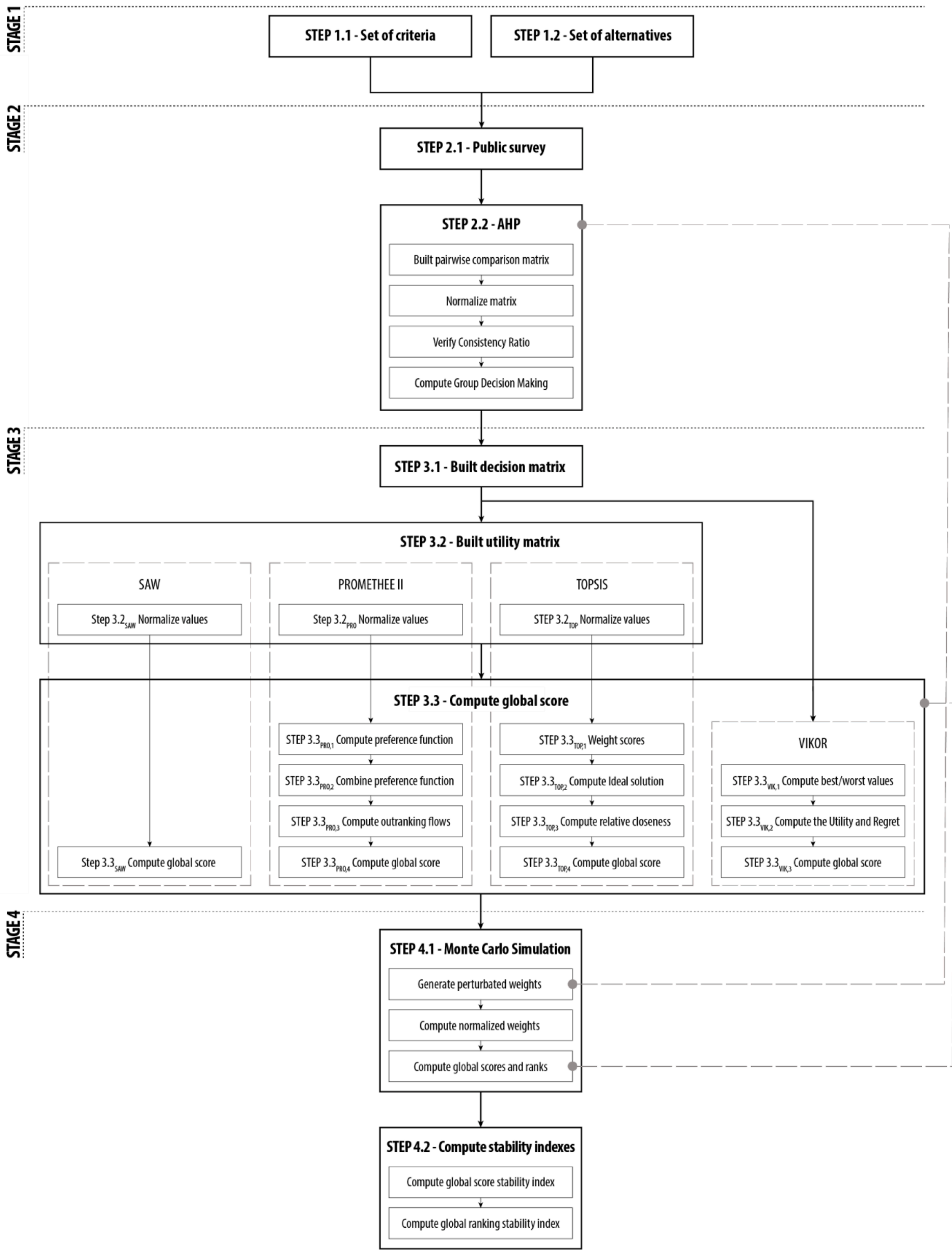

Figure 1. Overall methodological framework adopted in the study. The figure illustrates the four-stage process used to evaluate and rank the alternatives. Criteria weights are derived through the AHP, while the alternatives are assessed using four MCDMMs: SAW, PROMETHEE II, TOPSIS, and VIKOR. Finally, a Monte Carlo-based sensitivity analysis is performed to evaluate the robustness and stability of the obtained rankings.

varying expertise and understanding, may lead to heterogeneous judgments. For this reason, a structured weighting process is essential to establish comparable criterion priorities. The resulting weights reflect decision-makers' perspectives and play a crucial role in determining the final ranking of alternatives. For example, decision-makers who prefer the lowest-cost alternatives will have different choices from those who prefer the best-performing options. Weighing the criteria is a challenging task that requires a clear understanding of the decision problem and a consistent approach. Preferences can be used to assign weights, but respondents often struggle to evaluate many criteria simultaneously (e.g., Carrara et al., 2021; Carra et al., 2023). Various weighting approaches have been proposed to face this challenge (e.g., Hens & De Wit, 2003; Wang & Lee, 2009).

Among the available techniques, AHP is selected in this framework because it provides an explicit consistency ratio. This feature enables the observation and quantification of variability and noise inherent in citizen-derived judgments (Wind & Saaty, 1980; , Saaty, 1987, 1990, 1994. This aspect is well documented in prior work on preference instability and in empirical studies showing that direct-rating approaches often lead to scale misinterpretation and inconsistency (Zelany, 1974; Friman et al., 2001; Cantwell et al., 2009; Wang & Lee, 2009; Castillo and Pitfield, 2010; Hassan et al., 2013). In addition, AHP benefits from a solid mathematical foundation (e.g., eigenvector-based ratio-scale derivation). Despite noted critiques (e.g., Dyer, 1990), it has been extensively supported in the decision-analysis literature and remains one of the most rigorously formalised weighting approaches available for multi-criteria evaluation (Harker & Vargas, 1990; Saaty, 1990; Forman & Gass, 2001). Alternative techniques, such as the BWM, reduce cognitive load and typically yield smoother, more internally consistent weight vectors (Koohathongsumrit & Chankham, 2022, 2023; Kumar et al., 2024). However, this characteristic would limit the variance needed to meaningfully assess ranking robustness under realistic, noisy conditions (one of the paper's central analytical objectives). Finally, AHP has a long record of use in public-facing decision processes and is widely understood by planners and practitioners, enhancing the interpretability, communicability, and practical usability of the weighting results in applied transport-planning contexts (Macharis & Bernardini, 2015).

Usually, the weighting process is commonly completed by 'experts', with solid backgrounds and expertise in the field. In contrast, this framework collects weights from citizens, reflecting the growing interest in incorporating user perspectives into transport planning. Because citizen-based weighting is sensitive to response bias, it is essential to adopt a sampling strategy that ensures representativeness. For example, the demographic characteristics of survey participants (e.g., age and occupation) should be selected to mirror the population structure of the context at hand. This approach aims to reduce systematic distortions caused by the over- or underrepresentation of specific subgroups and to provide a weighting dataset as balanced and unbiased as possible.

To involve a broad group of users, a web survey could be designed to be processed using the AHP method, as it enables fast, low-cost data collection and is widely applied in research (Carra et al., 2025; Ferretto et al., 2026). The online format allows quick access, direct completion, and immediate data usability.

More specifically, **STAGE 2** is divided into two steps: the public survey **(STEP 2.1)** and the AHP (**STEP 2.2**), as detailed below.

2.2.1. Public survey

STEP 2.1 lays the foundation for the analysis by constructing a hierarchical structure using main criteria and sub-criteria (to mitigate high computational effort when there are many criteria) and by performing pairwise comparisons via a public survey. The survey pairs all main criteria with each other and all sub-criteria under each main criterion. Afterwards, participants rate the relative importance of each pairwise comparison according to the linguistic version of Saaty's scale, ranging from 1 (*"equal importance")* to, e.g., 9 ("*very important*").

A public survey could be conducted on different online platforms. This framework considered two options:

- *Google Form*, a popular platform, is easy to access and can be used by many people. However, this platform lacks analytical tools and notifications for consistency checks. Therefore, some pairings can be reversed and repeated to verify the consistency of the participants' responses during the questionnaires.
- *Analytic Hierarchy Process—Online System (AHP-OS)* is a free platform developed by Goepel (2017) to support complex decision-making problems. This web-based tool is built on the solution of an Eigenvalue problem. Participants can check their consistency in real time and adjust their choices to improve it during the survey. The platform provides analytical results (e.g., using Eqns. 1-5 in the next section) and can analyze data collected from a Google Form.

2.2.2. Theoretical framework of AHP

STEP 2.2 implements the mathematical process of the AHP method. Before comparing criteria, a pairwise comparison matrix is built for each participant. The rows and columns in this matrix represent the criteria. Each entry represents the weight (importance) assigned to a criterion in relation to another. From the matrix, a weight vector for each criterion or sub-criterion is computed and then normalized (Eqn. 1). Next, a consistency test is performed to verify the reliability of judgments in each matrix since some inconsistencies in judgments could occur (Eqns. 2-4). Several approaches can be followed. This study adopts the eigenvector-eigenvalue-based approach inside the AHP-OS platform. Specifically, the eigenvalues of the pairwise comparison matrix are used to determine whether the resulting weight meets the consistency standard. With it, a consistency measure is easily computed. The measure is a function of the largest eigenvalue and the number of criteria, illustrating the reliability of each answer. Finally, all the individual answers that meet the consistency threshold are aggregated into a 'group answer' using Weighed Geometric Mean Aggregation of Individual Judgments (WGM-AIJ; Eqn. 7).

To be more formal, let:

- J be the set of criteria, with $j \in J$ and $k \in J$ two generic criteria.
- D be the set of participants.
- $\boldsymbol{\omega}_{\mathbf{J,d}} = \{\omega_{j,d} : j \in J\}$ $(\boldsymbol{\omega}_{\mathbf{J,d}} \in \mathbb{R}^{|J|})$ be the 'judgement vector' for each $d \in D$. The generic $\omega_{j,d} \in \mathbb{R}$ is the importance of judgement that participant $d \in D$ assigns to the criterion $j \in J$.
- $\mathbf{A_{J,d}} = \{a_{jk,d} : j \in J;\ k \in J\}$ $(\mathbf{A_{J,d}} \in (\mathbb{R}^{|J|} \times \mathbb{R}^{|J|}))$ be the 'individual comparison matrix' of criteria J for each $d \in D$. The generic component $a_{jk,d} = \omega_{j,d}/\omega_{k,d}$ $(a_{jk,d} \in \mathbb{R})$ is the ratio between the importance of the criterion $j \in J$ (i.e., $\omega_{j,d}$) with respect to the importance of the criterion $k \in J$ (i.e., $\omega_{k,d}$) for participant $d \in D$. For instance, $\omega_{j,d}/\omega_{k,d} = 3/1$ means criterion $j \in J$ is three times more important than criterion $k \in J$, and *vice versa.*
- $w_{j,d} \in \mathbb{R}$ be the overall unnormalized weight for the criterion $j \in J$ and participant $d \in D$.
- $\widehat{w}_{j,d}$ be the normalised value of $w_{j,d}$.
- $\widehat{w}_j \in \mathbb{R}$ be the aggregated normalized weight associated with the criterion $j \in J$.
- $\lambda_{maxJ,d} \in \mathbb{R}$ be the largest eigenvalue of matrix $\mathbf{A_{J,d}}$.
- $\mathbf{v_{J,d}} \in \mathbb{R}^{|J|}$ be the eigenvector of $\mathbf{A_{J,d}}$ associated to $\lambda_{max,d}$.
- $\boldsymbol{Id}$ be the identity matrix.
- $CR_{J,d} \in \mathbb{R}$ be the Consistency Ratio, which expresses the consistency/ inconsistency of pairwise comparisons provided by each $d \in D$ on J. Precisely, $CR_{J,d}$ measures whether judgments of participants are logical and consistent with choices made throughout the survey.
- $RI \in \mathbb{R}$ be a parameter for calculating the $CR_{J,d}$. It is a tabled quantity related to the number of criteria considered in pairwise comparisons.
- $\alpha_d \in \mathbb{R}$ be the importance of the participant $d \in D,$ such that the condition $\sum_{d \in D} \alpha_d = 1$ is satisfied.

For each participant $d \in D,$ the following procedure was used to compute the weights and check the consistency of judgments.

After rating the levels of importance, the procedure establishes a pairwise comparison matrix $\mathbf{A_{J,d}}$ as follows:

$$\mathbf{A_{J,d}} = [a_{jk,d}] = \begin{bmatrix} a_{11,d} & a_{12,d} & a_{13,d} & & a_{1|J|,d} \\ a_{21,d} & a_{22,d} & a_{23,d} & \cdots & a_{2|J|,d} \\ a_{31,d} & a_{32,d} & a_{33,d} & & a_{3|J|,d} \\ & \vdots & & \ddots & \vdots \\ a_{|J|1,d} & a_{|J|2,d} & a_{|J|3,d} & \cdots & a_{|J||J|,d} \end{bmatrix} \forall d \in D \quad (1)$$

Next, $\mathbf{A_{J,d}}$ is normalized. Specifically, an overall weight $\boldsymbol{w_{j,d}}$ is computed for each criterion $j \in J$ and then normalized by the average geometric method (among the several available approaches), as illustrated in Eqns. (2) and (3), respectively:

$$\mathbf{w_{j,d}} = \sqrt[|J|]{\prod_{k \in J} a_{jk,d}} \ \forall d \in D \quad (2)$$

$$\widehat{w}_{j,d} = \frac{w_{j,d}}{\sum_{j \in J} w_{j,d}} \ \forall d \in D \quad (3)$$

Next, the Consistency Ratio (CR) is verified since attaining a consistent pairwise comparison matrix is often unfeasible in practical scenarios. Thus, the aim is to derive a pairwise comparison matrix closely resembling consistency. Among the methods for evaluating consistency, Saaty (1977, 1980a) introduced the $CR_{J,d}$. It serves as a metric indicating the extent of deviation from a consistent matrix (Pant et al., 2022). $CR_{J,d}$ definition is provided in Eqn. (4) (Saaty, 1980a, b).

$$CR_{J,d} = \frac{(\lambda_{maxJ,d} - |J|)}{RI \cdot (|J| - 1)} \ \forall d \in D \quad (4)$$

If the paired comparisons exhibit perfect consistency (i.e., $a_{jk,d} = 1 / a_{kj,d} \ \forall (j,k) \in J^2$), then $\lambda_{maxJ,d}$ equals the size of the matrix (i.e., $\lambda_{maxJ,d} = |J|$), and $CR_{J,d}$ is zero. As the inconsistency between comparisons increases, both $\lambda_{maxJ,d}$ and $CR_{J,d}$ also increase. From a theoretical viewpoint, $\lambda_{maxJ,d}$ can be obtained by solving the following (equivalent) equations:

$$\mathbf{A_{J,d}} \boldsymbol{v}_{\mathbf{J,d}} = \lambda_{maxJ,d} \boldsymbol{v}_{\mathbf{J,d}} \ \forall d \in D \quad (5)$$

$$(\mathbf{A_{J,d}} - \lambda_{maxJ,d} \boldsymbol{Id}) \boldsymbol{v}_{\mathbf{J,d}} = 0 \ \forall d \in D \quad (6)$$

From a practical perspective, $\lambda_{maxJ,d}$ can be calculated using the power method for approximating eigenvalues (De Alwis, 1993). This process is programmed inside the AHP-OS platform.

An acceptable $CR_{J,d}$ is traditionally expected to be below 0.10 (10%), although values up to 0.20 (20%) may be tolerated in specific applications (e.g., Wedley, 1993; Carrara et al., 2021). However, the 0.10 threshold was originally designed for trained experts with substantial domain knowledge. In the present framework, respondents are ordinary citizens with heterogeneous backgrounds and limited technical expertise. Hence, slightly more permissive thresholds might help avoid excessive data exclusion and preserve sample representativeness (e.g., Sato & Tan, 2023). For this reason, a consistency standard of 0.15 is adopted here, reflecting a balance between methodological rigor and the practical need to retain a sufficiently broad and demographically representative sample.

Finally, the procedure aggregates the overall normalized weights assigned by the participants to each criterion $j \in J$ (i.e., $\widehat{w}_{j,d}$), considering their potential different "importance" (e.g., various levels of expertise). Therefore, $\widehat{w}_j$ represents the Group Decision Making. Among the several aggregation techniques available in the literature, this research adopts the weighted geometric mean - aggregation of individual judgments because it is the most widespread and the only one that satisfies several axiomatic conditions (Grošelj et al., 2015). More formally, assuming that the participant's opinions are not equally important, the aggregated normalized weight associated with the considered criterion $j \in J$ is computed as follows:

$$\widehat{w}_j = \prod_{d \in D} (\widehat{w}_{j,d})^{\alpha_d} \ \forall j \in J \quad (7)$$

2.3. Ranking alternatives

STAGE 3 prioritizes alternatives using SAW, TOPSIS, VIKOR, and PROMETHEE II methods. This stage involves three steps: building the decision matrix (**STEP 3.1**), converting it into a utility matrix (**STEP 3.2**), and computing the global scores (**STEP 3.3**), as detailed below.

STEP 3.1 builds a decision matrix by assigning a value to each criterion of each alternative. Unlike the criteria weights, these values are not based on subjective judgments from survey respondents. They are derived from technical, quantitative, and qualitative information obtained from literature, existing projects, or domain knowledge. This ensures that the evaluation of alternatives remains grounded in objective system characteristics. Qualitative values are converted to numerical values using the procedure outlined in Table 1. While this conversion process is influenced by how acceptable values are defined, the overall method does not require strict adherence to previous recommendations, as these values can be established through alternative, yet technically justified, approaches.

Next, **STEP 3.2** transforms the decision matrix into a utility matrix by normalising the performance values, since the criteria are expressed in different units. Generally, the best or worst normalized value of a criterion depends on whether it is a benefit or a cost to maximize utility. A higher value is preferred in the first case and a lower one in the second.

To be more formal, let:

- $M = \{SAW, \ TOP, \ VIK, PRO\}$ be the set of methods, encompassing SAW, TOPSIS, VIKOR, and PROMETHEE II, respectively.
- $GS_{m,i} \in \mathbb{R}$ be the global score of alternative $i \in I$, computed according to method $m \in M$.
- $RK_{m,i} \in \mathbb{N}$ be the rank of alternative $i \in I$, determined according to method $m \in M$.
- $J_B \subseteq J$ and $J_C \subseteq J$ be the sets of benefit and cost criteria, respectively.
- I be the set of alternatives, with $i \in I$ and $h \in I$ two generic alternatives.
- $\boldsymbol{X_{I \times J}} = \{x_{ij} : \ i \in I; \ j \in J\} \ (\boldsymbol{X_{I \times J}} \in (\mathbb{R}^{|I|} \times \mathbb{R}^{|J|}))$ be the decision matrix. The generic component $x_{ij} \in \mathbb{R}$ represents the performance value of the alternative $i \in I$ with respect to the criterion $j \in J$.
- $\widehat{x}_{ij} \in \mathbb{R}$ be the normalized value of $x_{ij} \in \mathbb{R}$.
- $y_{ij} \in \mathbb{R}$ be the weighted score for the alternative $i \in I$ with respect to the criterion $j \in J$.
- $i_j^+ \in I$ and $i_j^- \in I$ be the best-ideal solution and the worst-worst solution with respect to the criterion $j \in J$, respectively.
- $\Delta_i^+ \in \mathbb{R}$ and $\Delta_i^- \in \mathbb{R}$ be the (Euclidean) distances of the alternative $i \in I$ from the best-ideal solution and worst-ideal solution, respectively.
- $CC_i \in \mathbb{R}$ be the relative closeness of the alternative $i \in I$ to the ideal solution.
- $S_i \in \mathbb{R}$ be the value of the utility measure for the alternative $i \in I$.

Table 1
Conversion scale from qualitative value to quantitative value.

Benefit criteria		Cost criteria	
Qualitative value	*Judgment value*	*Qualitative value*	*Judgment value*
Average	1	Extremely low	9
Good	3	Low	7
Very good	5	Average	5
Excellent	7	High	3
Outstanding	9	Extremely high	1
Intermediate values	2, 4, 6, 8	Intermediate value	2, 4, 6, 8

- $R_i \in \mathbb{R}$ be the value of the regret (disutility) measure for the alternative $i \in I$.
- $x_j^+ \in \mathbb{R}$ and $x_j^- \in \mathbb{R}$ be the best and the worst performance values taken by the criterion $j \in J$, respectively.
- $\beta \in \mathbb{R}$ be a parameter (value between 0 and 1) which gives different weight to S_i and R_i ($\beta = 0.5$ indicates equal relevance; $\beta > 0.5$ favours utility measure values; $\beta < 0.5$ favours regret measure values). The value of β can be selected according to the preference of decision-makers.
- $P_j(i,h): I \times I \rightarrow [0,1]$ be the preference function for the pair of alternatives $(i,h) \in I \times I$ with respect of the criterion $j \in J$.
- $\pi(i, h): I \times I \rightarrow [0,1]$ be the combined (multicriteria) preference function for the pair of alternatives $(i,h) \in I \times I$.
- $\varnothing^+(i)$ and $\varnothing^-(i)$ be the positive and negative outranking flows of alternative $i \in I$, respectively.

Finally, **STEP 3.3** computes the global score $GS_{m,i}$ of each $i \in I$ for each $m \in M$. This step may be further subdivided into sub-steps, depending on the specific algorithm used.

Since both normalization and global score computation involve mathematical procedures tailored to each multi-criteria method, **STEP 3.2** and **STEP 3.3** are method dependent. Hence, they are denoted as **STEP 3.2$_m$** and **STEP 3.3$_m$** ($m \in M$), respectively. Further details regarding these steps are provided in the following sections.

2.3.1. SAW method

SAW is the simplest among MCDMMs. It sums up all the weighted values of $i \in I$ with respect to $j \in J$ to find the alternative with the highest score. First, **STEP 3.2$_{SAW}$** normalizes each term of $\boldsymbol{X}_{I\times J}$. For each $j \in J_B$ (i.e., for each criterion considered as a benefit), the normalized value $\hat{x}_{ij}$ is computed relatively to its maximum value (Eqn. 8). Conversely, for each $j \in J_C$ (i.e., for each criterion considered as a cost), the normalized value $\hat{x}_{ij}$ is computed relatively to its minimum value (Eqn. 9).

$$\hat{x}_{ij} = \frac{x_{ij}}{\max_{i\in I}(x_{ij})} \quad \forall j \in J_B \tag{8}$$

$$\hat{x}_{ij} = \frac{\min_{i\in I}(x_{ij})}{x_{ij}} \quad \forall j \in J_C \tag{9}$$

Next, **STEP 3.3$_{SAW}$** computes the global score $GS_{SAW,i}$ for each $i \in I$, as follows:

$$GS_{SAW,i} = \sum_{j\in J} \hat{w}_j \hat{x}_{ij} \tag{10}$$

The global score $G_{SAW,i}$ ranks alternatives. It is a positively oriented score, i.e., the highest score is associated with the best alternative. Therefore, alternatives are ranked in descending order of $GS_{SAW,i}$, and the corresponding rank $RK_{SAW,i}$ is assigned accordingly.

2.3.2. TOPSIS method

For each alternative $i \in I$, TOPSIS computes its relative closeness (CC_i) to the ideal solution by considering the distances Δ_i^+ and Δ_i^- from the best- and the worst- ideal scenarios, respectively. First, **STEP 3.2$_{TOP}$** normalizes criteria values and builds the decision matrix $\boldsymbol{X}_{I\times J}$. TOPSIS employs a vector normalization technique (Eqn. 11), eliminating the need for separate cost-benefit criteria normalization. Therefore, for each $j \in J$, the normalized value $\hat{x}_{ij}$ is computed as follows:

$$\hat{x}_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i\in I} x_{ij}^2}} \quad \forall j \in J \tag{11}$$

Next, **STEP 3.3$_{TOP}$** follows, and is divided into 4 sub-steps (**STEP 3.3$_{TOP,1}$**, **STEP 3.3$_{TOP,2}$**, **STEP 3.3$_{TOP,3}$**, and **STEP 3.3$_{TOP,4}$**). In **STEP 3.3$_{TOP,1}$**, normalized values are multiplied with their associated weights to obtain weighted scores y_{ij} of each $i \in I$ with respect to $j \in J$ (Eqn. 12).

$$y_{ij} = \hat{w}_j \hat{x}_{ij} \ \forall i \in I; \ \forall j \in J \tag{12}$$

Next, **STEP 3.3$_{TOP,2}$** determines i_j^+ (the ideal best solution) and i_j^- (the ideal worst solution) with respect to each criterion $j \in J_B$ considered a benefit (Eqn. 13) and for each criterion considered a cost $j \in J_C$ (Eqn. 14).

$$i_j^+ = \max_{i\in I}\left(y_{ij}\right) \ \forall j \in J_B \ i_j^- = \min_{i\in I}\left(y_{ij}\right) \ \forall j \in J_B \tag{13}$$

$$i_j^+ = \min_{i\in I}\left(y_{ij}\right) \ \forall j \in J_C \ i_j^- = \max_{i\in I}\left(y_{ij}\right) \ \forall j \in J_C \tag{14}$$

In **STEP 3.3$_{TOP,3}$**, distances from the best- and worst-ideal scenarios and their *relative closeness* are computed for each alternative $i \in I$ as follows:

$$\Delta_i^+ = \sqrt{\sum_{j\in J}\left(y_{ij} - y_{i_j^+ j}\right)^2} \ \forall \ i \in I \tag{15}$$

$$\Delta_i^- = \sqrt{\sum_{j\in J}\left(y_{ij} - y_{i_j^- j}\right)^2} \ \forall \ i \in I \tag{16}$$

Finally, **STEP 3.3$_{TOP,4}$** computes the global score $G_{TOP,i}$ for each $i \in I$, i.e., the relative closeness CC_i of each $i \in I$ to the ideal solution. CC_i is computed as in Eqn. (17), and $G_{TOP,i}$ is assumed to be equal to CC_i (Eqn. 18). Thus, alternatives are ranked accordingly with the highest closeness. $GS_{TOP,i}$ is a positively oriented score, i.e., the highest score is associated with the best alternative. Therefore, alternatives are ranked in descending order of $GS_{TOP,i}$, and the corresponding rank $RK_{TOP,i}$ is assigned accordingly.

$$CC_i = \frac{\Delta_i^-}{\Delta_i^- + \Delta_i^+} \ \forall \ i \in I \tag{17}$$

$$GS_{TOP,i} = CC_i \ \forall \ i \in I \tag{18}$$

2.3.3. VIKOR method

VIKOR provides a compromise between maximizing group utility (representing the "majority" rule) and minimizing individual regret (representing the "opponent" rule) (Opricović, 2009). A distinctive feature of VIKOR is that it processes performance values directly without an explicit preliminary normalization step. Hence, **STEP 3.2$_{VIK}$** does not exist. Conversely, **STEP 3.3$_{VIK}$** is divided into 3 sub-steps (**STEP 3.3$_{VIK,1}$**, **STEP 3.3$_{VIK,2}$**, and **STEP 3.3$_{VIK,3}$**). **STEP 3.3$_{VIK,1}$** computes the best value x_j^+ and the worst value x_j^- for each $j \in J_B$, i.e., for each criterion considered a benefit (Eqn. 19) and for each $j \in J_C$, i.e., for each criterion considered a cost (Eqn. 20).

$$\begin{cases} x_j^+ = \max\limits_{i\in I} x_{ij} \\ x_j^- = \min\limits_{i\in I} x_{ij} \end{cases} \ \forall \ j \in J^B \tag{19}$$

$$\begin{cases} x_j^+ = \min\limits_{i\in I} x_{ij} \\ x_j^- = \max\limits_{i\in I} x_{ij} \end{cases} \ \forall \ j \in J^C \tag{20}$$

Next, **STEP 3.3$_{VIK,2}$** computes the Utility S_i and Regret R_i measures for each $i \in I$ as follows:

$$S_i = \sum_{j\in J} \hat{w}_j \left[\frac{x_j^+ - x_{ij}}{x_j^+ - x_j^-}\right] \ \forall \ i \in I \tag{21}$$

$$R_i = \max_{j\in J}\left(\hat{w}_j \left[\frac{x_j^+ - x_{ij}}{x_j^+ - x_j^-}\right]\right) \ \forall \ i \in I \tag{22}$$

Finally, **STEP 3.3$_{VIK,3}$** aggregates into the global value $G_{VIK,i}$ the score of each alternative $i \in I$. It determines the final ranking of alternatives

(Eqn. 23).

$$GS_{VIK,i} = \beta \frac{S_i - \min_{i\epsilon I}(S_i)}{\max_{i\epsilon I}(S_i) - \min_{i\epsilon I}(S_i)} + (1-\beta)\frac{R_i - \min_{i\epsilon I}(R_i)}{\max_{i\epsilon I}(R_i) - \min_{i\epsilon I}(R_i)} \quad \forall\, i \in I \tag{23}$$

$GS_{VIK,i}$ is a negatively oriented score, i.e., the lowest score is associated with the best alternative. Therefore, alternatives are ranked in ascending order of $GS_{VIK,i}$, and the corresponding rank $RK_{VIK,i}$ is assigned accordingly. However, S_i and R_i are also directly considered in addition to $GS_{VIK,i}$ to find the best alternative. Specifically, an alternative $i \in I$ with the minimum $GS_{VIK,i}$ provides the best solution, and it is referred to as $i^{(1)}$ if it satisfies the following conditions:

1. Condition 1. Acceptable advantage of the first ranked alternative $i^{(1)}$ *vs* the next ranked alternative $i^{(2)}$.

$$GS_{VIK,i^{(2)}} - GS_{VIK,i^{(1)}} \geq \frac{1}{|I|-1} \ \forall\, i \in I \tag{24}$$

2. Condition 2. Acceptable stability in Decision Support of the best alternative ranked over the others by S_i and/or R_i values.

Alternative $i^{(1)}$ must have the best value(s) of the S_i and/or R_i among alternatives. If these conditions are not satisfied, the compromise solution is determined as follows:

- Alternative $i^{(1)}$ satisfied Condition 1 but not Condition 2. A set of compromise solutions is established, including first-ranked and second-ranked alternatives.
- Alternative $i^{(1)}$ met Condition 2 but not Condition 1. A set of compromise solutions is established, including $i^{(1)}$ and all alternatives where the first-ranked alternative has no “sufficient advantage”, indicating a condition of “closeness”.
- Alternative $i^{(1)}$ did not meet both Conditions 1 and 2. It is not “sufficiently” better than the second. As a result, a set of compromise solutions includes all alternatives (Opricović, 2009).

2.3.4. PROMETHEE II method

PROMETHEE II is a member of the PROMETHEE family of methods. It is an outranking method and consists of four steps. Its approach provides a visual comparison between each pair of criteria and alternatives. First, **STEP 3.2$_{\text{PRO}}$** normalizes criteria values and builds the decision matrix $\boldsymbol{X}_{I\times J}$ by a min-max normalization technique. For each $j \in J_B$ (i.e., for each criterion considered as a benefit), the normalized value $\widehat{x}_{ij}$ is computed by measuring the distance from its minimum value and scaling it relative to its variability range (i.e., the difference between the maximum and minimum values), as shown in Eqn. (25). Similarly, for each $j \in J_C$, i.e., for each criterion considered as a cost, the normalized value $\widehat{x}_{ij}$ is computed by measuring the distance from its maximum value and scaling it over the same variability range, as shown in Eqn. (26):

$$\widehat{x}_{ij} = \frac{x_{ij} - \min_{i\in I}(x_{ij})}{\max_{i\in I}(x_{ij}) - \min_{i\in I}(x_{ij})} \ \forall j \in J_B \tag{25}$$

$$\widehat{x}_{ij} = \frac{\max_{i\in I}(x_{ij}) - x_{ij}}{\max_{i\in I}(x_{ij}) - \min_{i\in I}(x_{ij})} \ \forall j \in J_C \tag{26}$$

Next, **STEP 3.3$_{\text{PRO}}$** follows, and is divided into 4 sub-steps (**STEP 3.3$_{\text{PRO,1}}$, STEP 3.3$_{\text{PRO,2}}$, STEP 3.3$_{\text{PRO,3}}$**, and **STEP 3.3$_{\text{PRO,4}}$**). **STEP 3.3$_{\text{PRO,1}}$** computes the preference function $P_j(i,h)$ between each pair of alternatives, i.e., measures the magnitude of preference of alternative $i \in I$ over the alternative $h \in I$ with respect to the criteria $j \in J$, such that:

- $P_j(i,h) = 0$ indicates no preference of $i \in I$ over $h \in I$;
- $P_j(i,h) = 1$ indicates a strict preference of $i \in I$ over $h \in I$;
- $0 < P_j(i,h) < 1$ indicates that $i \in I$ is preferred over $h \in I$ with a magnitude that increases with the distance of $P_j(i,h)$ from zero.

There are six types of preference functions (Brans et al., 1986). This study employs the linear function due to its straightforward interpretability.

$$P_j(i,h) = \begin{cases} 0; \ if\ \widehat{x}_{ij} \leq \widehat{x}_{hj} \\ \widehat{x}_{ij} - \widehat{x}_{hj}; \ if\ \widehat{x}_{ij} > \widehat{x}_{hj} \end{cases} \forall (i,h) \in I\times I;\ \forall j \in J \tag{27}$$

STEP 3.3$_{\text{PRO,2}}$ computes the combined preference function $\pi(i,\ h)$ for each pair of alternatives $i \in I$ and $h \in I$ by considering the weight w_j associated with each criterion $j \in J$. This function $\pi(i,\ h)$ is a preference index accounting for all the parameters (Eqn. 28).

$$\pi(i,\ h) = \sum_{j\ \epsilon\ J} w_j \left[p_j(i,h)\right] \ \forall (i,h) \in I\times I \tag{28}$$

Afterwards, **STEP 3.3$_{\text{PRO,3}}$** determines the *leaving* $\varnothing^+(i)$ and *incoming* $\varnothing^-(i)$ *outranking flows* for the alternative $i \in I$, representing the relationship of each pairwise (Eqns. 29 and 30). Intuitively, it measures the preference ($\varnothing^+(i)$) and non-preference ($\varnothing^-(i)$) of the alternative $i \in I$ over the other remaining ones. In other words, the positive outranking flow $\varnothing^+(i)$ shows how many alternatives $i \in I$ dominate the other alternatives. In contrast, the negative outranking flow $\varnothing^-(i)$ shows how others dominate the alternative $i \in I$.

$$\varnothing^+(i) = \frac{1}{|I|-1}\sum_{h\ \epsilon\ I:\ h\neq i} \pi(i,h) \ \forall i \in I \tag{29}$$

$$\varnothing^-(i) = \frac{1}{|I|-1}\sum_{h\ \epsilon\ I:\ h\neq i} \pi(h,i) \ \forall i \in I \tag{30}$$

Finally, **STEP 3.3$_{\text{PRO,4}}$** computes the global score $GS_{PRO,i}$ of each alternative $i \in I$ that is assumed to be equal to $\varnothing(i)$, the net outranking, as in Eqns. (31) and (32). $GS_{PRO,i}$ is a positively oriented score, i.e., the highest score is associated with the best alternative. Therefore, the alternatives are ranked in descending order of $GS_{PRO,i}$, and the corresponding rank $RK_{PRO,i}$ is assigned accordingly.

$$\varnothing(i) = \varnothing^+(i) - \varnothing^-(i) \ \forall i \in I \tag{31}$$

$$GS_{PRO,i} = \varnothing(i) \ \forall\ i \in I \tag{32}$$

2.4. Performing a sensitivity analysis

STAGE 4 performs a sensitivity analysis to assess the stability of scores and rankings under variations in citizen-derived weights. Citizen-derived weights naturally exhibit variability and may contain moderate inconsistencies. These issues are partly due to the cognitively demanding nature of pairwise comparisons and the more permissive consistency threshold adopted for non-expert respondents. For this reason, the sensitivity analysis plays a central methodological role. It quantifies how stochastic fluctuations in the weights propagate through the different MCDMMs and identifies which methods remain stable under such perturbations. By relying on both score- and ranking-based robustness metrics, this analysis enables the framework to mitigate the distortions that may arise from noisy or partially inconsistent inputs. Once the most stable method has been identified, its results can be used to anchor the final decision, thereby enhancing overall methodological robustness despite the unavoidable variability of citizen-elicited weights. **STAGE 4** comprises two **STEPS (4.1 and 4.2)**, which are formalized in what follows.

STEP 4.1 simulates multiple comparison instances among the alternatives using different weight sets. These weight sets are generated via a Monte Carlo Simulation (MCS), in which stochastic perturbations are applied to the base weights. More formally, let:

- N be the set of simulation instances.
- $r_{j,n} \in [0,1]$ a random number associated with criterion $j \in J$, generated at instance $n \in N$ according to a uniform distribution.
- $\theta_{j,n} \in \mathbb{R}$ be the perturbated weight for criterion $j \in J$, generated at instance $n \in N$.
- $\widehat{\theta}_{j,n} \in \mathbb{R}$ be the normalised value of $\theta_{j,n}$.
- $\Psi_j : [0,1] \rightarrow [0,1]$ be the cumulative distribution function of the weight associated with criterion $j \in J$.
- $GS_{m,i,n} \in \mathbb{R}$ and $RK_{m,i,n}$ be the global score and rank of alternative $i \in I$ in simulation instance $n \in N$, computed according to method $m \in M$, respectively.

In the MCS, each criterion weight is modelled as a random variable. A proper cumulative distribution function Ψ_j is selected for each criterion to reflect the variability observed in the survey-based weights. A possible approach could assume that the weights are normally distributed around their mean values derived from a survey. This provides a realistic representation of unimodal and approximately symmetric variability in judgments (Yaraghi et al., 2014; Dahri & Abida, 2017; Serrano-Gómez & Muñoz-Hernández, 2019; Wicaksono et al., 2020). Then, different uncertainty scenarios are generated by systematically adjusting the standard deviation (expressed as a percentage of each mean weight), enabling a controlled and gradual exploration of increasing noise levels. This stochastic perturbation scheme is essential because it explicitly reproduces the natural dispersion and inconsistency that might arise when weights are elicited from a heterogeneous pool of citizens rather than domain experts.

The perturbated weight $\theta_{j,n}$ is then obtained by inverting the cumulative distribution function using a uniformly distributed random variable $r_{j,n}$ (Eqn. 33). The resulting weights are subsequently normalized to ensure they sum to 1 (Eqn. 34).

$$\theta_{j,n} = \Psi_j^{-1}\left(r_{j,n}\right) \ \forall\, j \in J;\ \forall\, n \in N \tag{33}$$

$$\widehat{\theta}_{j,n} = \frac{\theta_{j,n}}{\sum_{j\in J}\theta_{j,n}} \ \forall\, j \in J;\ \forall\, n \in N \tag{34}$$

Using the perturbed and normalized weights, the alternatives are re-evaluated. The global scores $GS_{m,i,n}$ and ranks $RK_{m,i,n}$ for each method are calculated as described in Stage 3, with the original weights replaced by the simulated perturbed ones. Notably, to ensure a rigorous and comprehensive assessment of robustness, all criterion weights are perturbed in MCS. Varying the entire weight vector avoids artificially constraining the solution space and allows the stability metrics to capture the full behavior of each MCDMM under stochastic fluctuations.

Next, **STEP 4.2** quantifies the stability of each decision method underweight changes using two novel and complementary metrics: score-based stability and ranking-based stability. More formally, let:

- $SS_{m,i} \in \mathbb{R}$ be the score-based stability index of alternative $i \in I$, for method $m \in M$.
- $RS_{m,i} \in \mathbb{R}$ be the ranking-based stability index of alternative $i \in I$, for method $m \in M$.
- $\overline{SS_m} \in \mathbb{R}$ and $\overline{RS_m} \in \mathbb{R}$ be the global score-based and ranking-based stability index for method $m \in M$, respectively.
- $\mu_{m,i} \in \mathbb{R}$ and $\sigma_{m,i} \in \mathbb{R}$ be the mean and standard deviation of $GS_{m,i,n}$ over all $n \in N$.
- ε be a small constant to prevent division by zero (e.g., $\varepsilon = 10^{-9}$).
- $RS_{m,i}^{shift} \in \mathbb{N}$ be the total number of rank changes for alternative $i \in I$, under method $m \in M$, across simulation instances.

For each method $m \in M$ and alternative $i \in I$, the score-based stability index is computed as follows:

$$SS_{m,i} = \frac{\left|\mu_{m,i}\right|}{\sigma_{m,i} + \varepsilon} \ \forall\, i \in I;\ \forall\, m \in M \tag{35}$$

The global score stability index is then defined for each method $m \in M$ as the average across all alternatives:

$$\overline{SS_m} = \frac{1}{|I|}\sum_{i\in I} SS_{m,i} \ \forall\, m \in M \tag{36}$$

Similarly, the ranking-based stability index is computed for each method $m \in M$ and alternative $i \in I$, by tracking the cumulative rank changes across successive instances:

$$RS_{m,i}^{shift} = \sum_{n\in N}\left|RK_{m,i,n} - RK_{m,i,n-1}\right| \ \forall\, i \in I;\ \forall\, m \in M \tag{37}$$

$RS_{m,i}^{shift}$ is then normalized by the number of transitions ($|N| - 1$) and inverted to yield the stability index:

$$RS_{m,i} = \frac{1}{\frac{RS_{m,i}^{shift}}{|N|-1} + \varepsilon} \ \forall\, i \in I;\ \forall\, m \in M \tag{38}$$

The global ranking stability index is then obtained by averaging over all alternatives:

$$\overline{RS_m} = \frac{1}{|I|}\sum_{i\in I} RS_{m,i} \ \forall\, m \in M \tag{39}$$

Both $\overline{SS_m}$ and $\overline{RS_m}$ are positively oriented values, i.e., the highest value is associated with the most stable method. A higher score-based stability index indicates that the method produces consistent evaluations with low score variability across different weighting schemes. Similarly, a higher ranking-based stability index signifies that the method tends to preserve the relative ordering of alternatives, even as weights fluctuate. These indices enable the framework to identify the most reliable methods in the presence of noisy or partially inconsistent citizen-derived weights, ensuring that the final decision relies on the most robust MCDMMs.

3. Real world experiment

3.1. Research context

The integrated framework was applied to the case study of Ho Chi Minh City (HCMC), also known as Saigon, the largest and most economically dynamic urban center in Vietnam, with a population of approximately 10 million and an urbanised area of roughly 2,061 km^2 along the Saigon River. The city is administratively divided into 19 districts. For instance, district 1 serves both the historic core and the central business district, hosting the headquarters of major multinational corporations and attracting significant daily commuter and tourist traffic. This concentration of economic activities generates substantial pressure on the city's transport corridors and exacerbates longstanding mobility challenges. Part of the difficulty stems from the fact that HCMC's transport infrastructure originated during the French colonial period, when the city's population was only around 50,000, resulting in a network that is structurally inadequate for present-day demand (Phan, 2023). Rapid motorization has further stressed the system: according to the most recently available modal-share data, motorcycles represent 74% of total trips, while active mobility modes account for 19%, public transport for 4%, and private cars for 1%, underscoring the minimal role of formal transit services in HCMC's travel patterns (Centre for Liveable Cities & Urban Land Institute, 2017; Transformative Urban Mobility Initiative, 2024; Ngoc & Nishiuchi, 2025). This imbalance places severe strain on an already outdated network and exacerbates daily congestion, especially as the number of registered motorcycles exceeds 7.8 million, further challenging the city's capacity to manage increasing travel volumes (Transformative Urban Mobility Initiative, 2024). At the same time, HCMC has articulated ambitious long-term objectives, including achieving a 25% public transport modal share in 2030, coupled with the development of a mass rapid transit system comprising eight metro lines and supported by long-term investments in electric bus deployment and sustainable transport transitions. Recent strategic planning documents confirm that these initiatives align with Vietnam's broader climate commitments and decarbonisation strategies, with measures such as the introduction of new e-bus routes and plans for fleet electrification forming part of the city's official mobility transition agenda (Deutsche Gesellschaft für Internationale Zusammenarbeit, 2024). Against this backdrop, HCMC emerges as a relevant case in a developing-country context where high-capacity, advanced transport systems are urgently needed. Its rapid growth, heavy commuter inflows, constrained infrastructure, and shift toward multimodal planning make it an ideal setting for applying the integrated multicriteria decision-making framework developed in this research. As such, the outcomes of this case study provide not only context-specific insights but also a valuable real-world reference for understanding public transport mode selection in fast-growing megacities facing structural mobility challenges.

Finally, it is worth clarifying that this study does not evaluate specific alignments or corridor-level designs within Ho Chi Minh City. Instead, it compares alternative public transport typologies at a strategic level, assuming representative operational characteristics. Detailed spatial configuration and route planning constitute a subsequent stage of analysis that can be informed by this framework's results.

Figure 2 highlights the spatial heterogeneity of the study area, showing the distribution of the analysed zones and the main urban corridors considered in the assessment.

3.2. Results

3.2.1. STAGE 1. Defining sets of alternatives and criteria

3.2.1.1. Defining criteria and sub-criteria. This study draws on existing literature to identify and define criteria that reflect the characteristics most influential in attracting individual users to public transport (**STEP 1.1**). The selection process was guided by four principles: (i) empirical relevance, focusing on factors consistently reported as determinants of modal choice; (ii) measurability, ensuring that criteria can be quantified and compared across alternatives; (iii) planning alignment, guaranteeing consistency with local and international transport planning guidelines (National Academies of Sciences, Engineering, and Medicine, 2013; Barabino et al., 2013a; Barabino et al., 2020; De Aloe et al., 2022; Ventura et al., 2022; Zhang et al., 2022; Keshavarz-Ghorabaee et al., 2022; Broniewicz & Ogrodnik, 2025), and (iv) respondent feasibility, striking a balance between representativeness and cognitive ease to avoid excessive burden during pairwise comparisons, thereby reducing survey fatigue and potential losses in response reliability.

Thus, six primary criteria and fifteen associated sub-criteria were selected. Specifically, the primary Individual Criteria, their corresponding sub-criteria, and the rationale are listed in the following:

- **Vehicle.** It refers to on-board facilities that influence passenger comfort and capacity, such as seating, climate control, and ventilation systems. This criterion is subdivided into:
  - (i) *Comfort*, including the availability of seating, air conditioning, and vehicle cleanliness.
  - (ii) *Capacity*, referring to the number of passengers the vehicle can accommodate simultaneously.

Comfort and capacity are among the most recognized determinants of user satisfaction and willingness to use public transport (Al Suleiman et al., 2023; Göransson & Andersson, 2023). Accurate measurement of passenger volumes is also fundamental for planning service capacity and operational adjustments in public transport systems (Olivo et al., 2019).

- **Security and Safety.** They encompass facilities and design elements related to emergencies and perceived risk, such as emergency buttons, surveillance cameras, staff presence, and crash-response tools (e.g., glass-breaking hammers). Safety and security concerns (both real and perceived) are major barriers to public transport use, influencing accessibility and trust (Barabino, 2018; Friman et al., 2020; Ingvardson & Nielsen, 2022). This criterion is divided into:
  - (i) *Security at stops*, i.e., perceived security while waiting or exiting.
  - (ii) *Security on board*, i.e., perceived security during the journey.
  - (iii) *Safety of the vehicle*, i.e., the likelihood of crashes or mechanical failure.

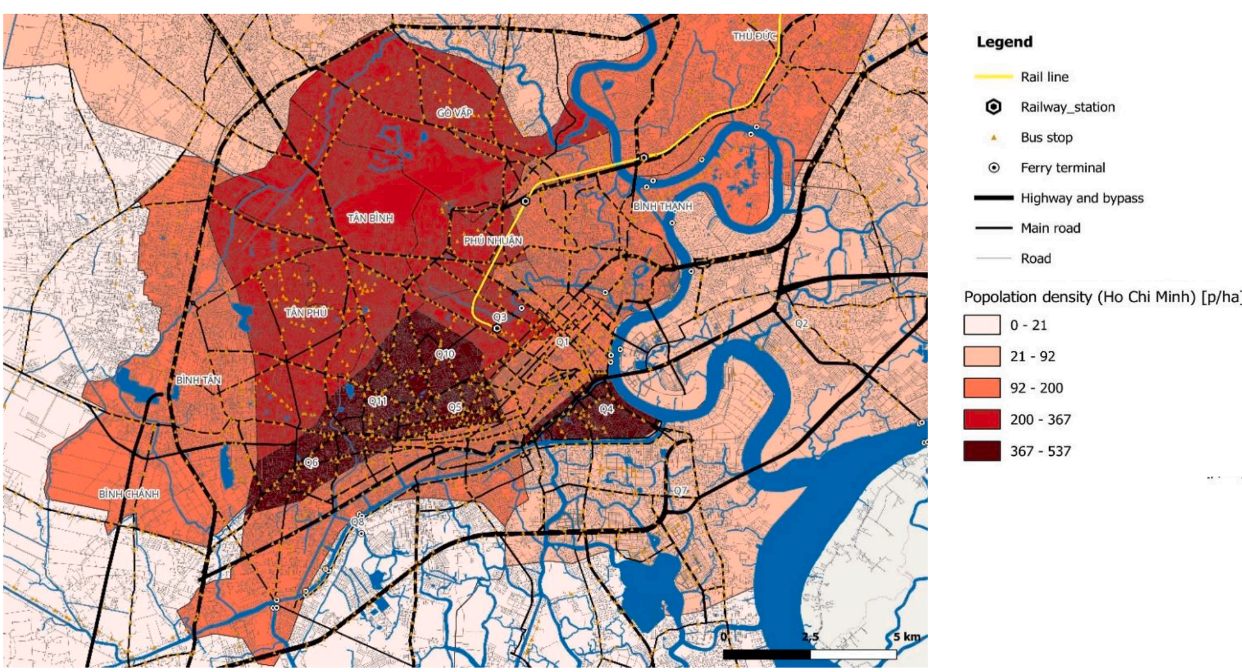

Figure 2. Ho Chi Minh City's transport systems and population densities. The figure shows the geographic context considered in the analysis, highlighting the main districts, road network, and spatial concentration of the evaluated alternatives and criteria within the urban area.

- **Travel time.** It refers to the total time required for a journey from an origin to a destination. Time competitiveness with private modes is a primary determinant of modal choice; reducing access, waiting, and in-vehicle time substantially increases attractiveness (Fiagborlo et al., 2024; Kühnel et al., 2025). It is divided into:
  - (i) *Access time*, i.e., the time needed to reach the first stop or to arrive at the destination from the last stop.
  - (ii) *Waiting time*, i.e., time spent waiting at the stop before boarding.
  - (iii) *In-vehicle time*, i.e., the duration of the actual ride.
- **Availability.** It refers to service features that define how accessible and usable the system is. This includes:
  - (i) *Operating hours*, i.e., the hours of service per day and the frequency of operation,
  - (ii) *Ticketing*, i.e., the duration of validity and/or the price of a single ticket.

Extended operating hours and flexible ticketing options enhance inclusivity and affordability, which are essential for equitable access although fare evasion issue should be also recognised (Barabino et al., 2023; Göransson & Andersson, 2023).

- **Reliability**. It is understood as the dependability of transit services across multiple dimensions, including timing, crowding, safety, comfort, and information availability (e.g., Barabino et al., 2013b; 2017). Predictability and real-time information reduce uncertainty and foster trust in the system, strongly influencing user satisfaction (Soza-Parra et al., 2021). It is categorized into:
  - (i) *Schedule adherence*, i.e., punctuality (on-time performance) and regularity (even spacing between vehicles).
  - (ii) *Tracking availability*, i.e., the possibility of accessing real-time information on vehicle locations.
- **Directness.** It represents the efficiency of travel between origin and destination, including connectivity to other transport options and stop spacing. Efficient routes and integration with other modes minimize transfers and cognitive effort, improving convenience and reducing perceived complexity (Liu, Cats, & Gkiotsalitis, 2021; Sipetas et al., 2024). It includes:
  - (i) *Integration*, i.e., connectivity with other vehicles or routes at stops.
  - (ii) *Need for transfer*, i.e., whether route changes are required to reach the destination.
  - (iii) *Distance*, i.e., the spacing between consecutive stops.

In addition to user-driven factors, the evaluation included investment cost—a key financial criterion representing expenses such as per-kilometer construction—since it is a major, tangible constraint on project feasibility and is easily understood by citizens and commonly cited in the literature (e.g., Lee, 2018). Conversely, operating costs were not considered, as citizens tend to focus more on visible service aspects, such as convenience or coverage, than on long-term financial sustainability.

Overall, the number of criteria and sub-criteria adopted reflects a deliberate trade-off between capturing the aspects most emphasized in the literature and maintaining an assessment structure that remains manageable for non-expert respondents engaged in pairwise comparisons. Since the weighting procedure relies on citizens rather than specialists, the inclusion of an excessive number of criteria would have resulted in lengthy, cognitively demanding questionnaires, potentially compromising response quality and consistency and increasing the risk of non-completion. Moreover, one is not required to adopt the previous list of criteria for using the overall framework: other criteria could be adopted.

Finally, it is worth noting that the case study is intended as an illustrative application of the proposed framework rather than as an exhaustive representation of all policy-relevant dimensions influencing public transport system selection. The framework itself remains fully flexible by design and can accommodate expanded or alternative sets of criteria in future applications, provided that appropriate data availability and elicitation conditions are met.

3.2.1.2. Defining alternatives and specifying characteristics. According to their performance, transport modes can be subdivided into three main categories (Vuchic, 2007): (i) basic modes, i.e., regular bus (RB), tram, streetcar; (ii) Medium-performance modes, i.e., light rail transit (LRT), bus rapid transit (BRT), trolleybus system; (iii) high-performance modes, i.e., metro, light rail rapid transit (LRRT), and automatic light rail transit (ALRT). In **STEP 1.2**, four technological alternatives are considered, aligning with high demand and infrastructural context: RB, BRT, LRT, and Metro, and their features are described in Table 2.

Each alternative is characterised by a set of performance values (Table 3) that comprises qualitative and quantitative criteria. Qualitative criteria capture user-perceived service attributes and are assessed through a structured linguistic scale (i.e., average, good, very good, excellent, outstanding), with evaluations grounded in recognised technical specifications (Vuchic, 2007). For example, a Metro station

Table 2
Alternative features (Adapted from Vuchic, 2007).

Public transport alternatives	Features description
Regular Bus	RB is the most widely used type of public transport, with the ability to operate on most streets and affordable investment costs. However, it has a limited capacity. A standard bus has 28 to 35 seats and can fit a total capacity of 100 passengers. An RB transit could provide all-day operating hours with fixed schedules and affordable prices. However, sticking to the schedule is not easy, especially during rush hours, because it shares the same lanes as private vehicles. RB transit is indispensable to public transport due to its primary role and good efficiency. The station distance ranges from 150m to 300m, with a transported capacity of 3,000-6,000 passengers/hour.
Bus Rapid Transit	BRT is developed by RB but has an independent infrastructure from other traffic. BRT can be designed larger than RB, thanks to its separate lanes, with 80-120 spaces per vehicle. Its own traffic lanes guarantee higher speed, greater reliability, and better safety than RB. BRT may have isolated stops or stations with good passenger protection and information, offering faster passenger exchange. The station distance ranges from 300m to 600m in the city centre and can be greater in suburban areas, with a transported capacity of 6,000-24,000 passengers/hour.
Light Rail Transit	LRT is an electrically powered rail transit mode with 110-350 spaces per vehicle. It can cover comparative urban area networks and provides high capacity, quiet operation, and excellent riding quality. Sometimes, LRT utilizes tunnel sections to bypass congested areas, thereby significantly enhancing its service level. Station spacing in urban areas ranges from 400m to 800m, while in the suburbs, it is up to 1200m, with a transport capacity of 10,000-24,000 passengers/hour.
Metro (or Heavy Rail Transit or Rapid Rail Transit)	Metro represents the ideal mode of transport for high-performance transport services, with 720-2500 spaces per transit unit. Metro has an entirely separate right-of-way, free from external interference. Therefore, it offers excellent comfort, high efficiency, and virtually absolute safety. Metro requires the most investment of all modes (primarily because of its fully grade-separated right of way and large stations). Hence, its implementation is mainly in high-density corridors. The distance between the two stations ranges from 500m to 2000m, with a transport capacity of 40,000-75,000 passengers/hour.

Table 3
Input characteristics for the case study in Ho Chi Minh City.

Primary criteria	Sub-criteria	Typology	Unit*	Category	Bus	BRT	LRT	Metro
Vehicle	Comfortability	Qualitative	-	Benefit	Average	Good	Excellent	Outstanding
	Capacity	Quantitative	Space/h/route	Benefit	3000	6000	10000	50000
Security and Safety	At stops	Qualitative	-	Benefit	Average	Very good	Very good	Outstanding
	On board	Qualitative	-	Benefit	Average	Very good	Very good	Outstanding
	Of vehicle	Qualitative	-	Benefit	Average	Good	Very good	Outstanding
Travel time	Access time	Quantitative	Mins	Cost	5	10	10	10
	Waiting time	Quantitative	Mins	Cost	20	10	10	5
	In-vehicle time	Quantitative	Mins	Cost	42	29	18	10
Availability	Operating hours	Quantitative	Hours	Benefit	16	17	19	19
	Ticket	Quantitative	$	Cost	0.5	1	2	2
Reliability	Schedule Adherence	Qualitative	-	Benefit	Good	Very good	Very good	Outstanding
	Tracking availability	Qualitative	-	Benefit	Average	Very good	Very good	Outstanding
Directness	Integration	Quantitative	Times	Benefit	Outstanding	Good	Average	Outstanding
	Need for Transfer	Qualitative	-	Cost	2	1	2	1
	Distance	Quantitative	Km	Benefit	0.2	0.3	0.5	1.2
Investment cost	-	Quantitative	Mil-$/km	Cost	1	11.5	45	100

*Qualitative sub-criteria do not report a unit of measurement because their performance is evaluated through a structured linguistic scale rather than numerical indicators.

equipped with cameras, controllers, and security personnel is assigned an "outstanding" safety value, whereas a standard bus stop receives an "average" value due to its limited security features. These qualitative criteria include Comfortability, Safety at stops, on board and of the vehicle, Punctuality (scheduled adherence), Tracking availability, and the Need for Transfer. Quantitative criteria instead represent objectively measurable system characteristics expressed in physical or monetary units and derived from established literature and technical reports (ECORYS Nederland BV, 2005; Vuchic, 2007; Levinson et al., 2009; Management Authority for Urban Railways, 2016; Transit Costs Project, 2023; Institute for Transportation and Development Policy, 2024). These include Capacity (space/h/route), Access time, Waiting time and In-vehicle time (minutes), Operating hours (hours), Ticket cost (USD), Integration (number of connected services or systems), Stop Distance (km), and Investment Cost (million USD per kilometre). Unlike qualitative service attributes directly perceived by users, Investment Cost is treated as an infrastructure-level economic constraint reflecting project feasibility rather than operational experience.

Additional operational assumptions are also grounded in empirical evidence. Since RB represents a basic transit mode, it is assumed that at least two transfers are required to reach the destination, whereas higher-performance modes are expected to operate through transport hubs requiring only one transfer. Users are generally willing to walk longer distances to access faster systems (O'Sullivan & Morrall, 1996); therefore, access time is set to 5 minutes for RB and 10 minutes for the other modes. Waiting time corresponds to the headway between consecutive trips, and in-vehicle time is estimated as the travel time for 5 km, including a 1-minute dwell time at each station.

3.2.2. STAGE 2. Weighting criteria

STAGE 2 assessed the criteria weights through a public survey and AHP, engaging actual public transport users rather than experts to verify whether meaningful priorities could emerge from the user perspective. The online, two-phase survey, conducted via Google Forms and AHP-OS, maximized participation and geographical reach while remaining cost-effective and accessible across devices, helping reduce the exclusion of less tech-savvy respondents (**STEP 2.1**).

The sampling strategy was designed to ensure a representative and unbiased dataset. A total of 252 citizens participated in the survey. The demographic characteristics of participants (age and occupation) were selected to reflect the structure of Vietnamese society, based on official NSO (2023) statistics. Their distributions are reported in Table 4.

In **STEP 2.2**, the AHP was applied (Table 5). The Individual Criteria were explicitly included in the weighting process to enable citizens to actively participate in evaluating public transport alternatives. Thus, the normalized weights for each criterion and sub-criterion of the Individual

Table 4
Survey participants' demographics.

Socio-demographic characteristics	Sample size
Age group	
<20	28
20-24	20
25-29	19
30-34	27
35-39	21
40-44	28
45-49	23
50-54	23
55-59	18
60-64	14
65-69	11
70-74	5
≥75	15
Occupation	
Household	99
Student	9
Worker	144
Total	252

Table 5
Normalized weights (and relative ranking) for Ho Chi Minh City.

Level 1		Level 2		Terminal node	Rank
Primary Criteria	*Weight*	*Sub-Criteria*	*Weight*		
Vehicle	0.1808	Comfortability	0.3776	0.0683	6
		Passenger capacity	0.6224	0.1125	2
Security and Safety	0.1652	Security on board	0.2793	0.0461	11
		Security at stops	0.4169	0.0689	5
		Safety of Vehicle	0.3038	0.0502	10
Speed	0.1262	Reaching time	0.2577	0.0325	14
		Awaiting time	0.4113	0.0519	9
		Travel time	0.3310	0.0418	12
Time Available	0.1608	Working hours	0.4078	0.0656	7
		Ticket	0.5922	0.0952	3
Reliability	0.1297	Punctuality	0.4652	0.0604	8
		Tracking availability	0.5348	0.0694	4
Directness	0.0944	Transfer	0.3190	0.0301	15
		Integration	0.2891	0.0273	16
		Distance	0.3920	0.0370	13
Investment Costs	0.1429	Investment costs	1.0000	0.1429	1

Criteria were determined using Eqns. 1-7. Conversely, the Investment Cost criterion was intentionally excluded from the AHP process, as the average user might lack sufficient awareness of this technical and politically sensitive parameter. Therefore, a cautious approach was adopted: the weight of this criterion was set to the average of all other primary criteria, resulting in a neutral normalized weight equal to the reciprocal of the total number of criteria considered (i.e., 1/7 in this case).

As shown in Table 5, Investment Cost (despite not being directly user-weighted) is among the top-ranked criteria. Moreover, it represents the top-ranked criterion at the terminal node level. This confirms its strategic relevance in broader planning evaluation, even if it lies outside the scope of average user awareness.

Moreover, user preferences tend to focus on the foundational aspects of the public transport system. Vehicle and Safety emerge as the two most influential primary criteria, indicating a strong public demand for systems that can handle high ridership and guarantee user protection, both on board and at stops. This reflects the city's status as a rapidly expanding metropolis in a developing country, where ensuring adequate basic services remains a top concern. Criteria related to Speed, Reliability, and Time Availability follow with moderate weights, suggesting that while efficiency and availability are valued, they remain secondary to more pressing structural and safety issues. Directness, encompassing integration and transfer needs, receives the lowest weight, possibly indicating a public expectation of network fragmentation or a tolerance for more complex travel patterns in exchange for improvements in comfort and safety. This outcome may also reflect the scale and spatial complexity of the urban context, in which large distances and fragmented urban structures tend to normalize indirect and transfer-based travel. When examining terminal nodes, the dominance of sub-criteria related to passenger capacity, ticket affordability, and information availability becomes evident. These reflect the practical, daily concerns of urban dwellers in a congested, economically diverse environment. Lower rankings for aspects such as travel time or distance reinforce the idea that users are willing to trade travel efficiency for gains in capacity and accessibility. Overall, the results indicate a user base primarily focused on upgrading core system functionality rather than fine-tuning performance.

To enable group-wise comparison of preferences and support a more refined interpretation of citizen heterogeneity, a segmented analysis has been performed, with results reported in the Supplementary Materials (Tables S1–S3), presenting normalized weights and rankings by age class for each occupation group. Despite the sample was limited and the results should be interpreted with care, the segmented findings indicate that heterogeneity mainly affects the relative ranking of sub-criteria rather than producing radical shifts in overarching priorities, with affordability, capacity, safety, and information consistently remaining central across groups. Age and occupation influence the emphasis placed on specific attributes, such as tracking availability, ticket cost, and passenger capacity, but without overturning the dominance of core system requirements. This confirms that aggregate results are suitable for high-level strategic interpretation, while segmentation adds interpretive depth by revealing how different demographic groups nuance their preferences within a broadly shared evaluative structure.

3.2.3. STAGE 3. Ranking alternatives

STAGE 3 applied SAW, TOPSIS, VIKOR, and PROMETHEE II to assess and rank public transport alternatives. **STEP 3.1** constructed the decision matrix by assigning a value to each sub-criterion for each alternative. Qualitative values, such as reliability and safety sub-criteria, were converted into numerical values using the linguistic version of Saaty's scale. Next, in **STEP 3.2,** the decision matrix was converted into a utility matrix. Finally, in **STEP 3.3**, the global score $GS_{m,i}$ and the rank $RK_{m,i}$ of each $i \in I$ was computed according to each method $m \in M$. Therefore, the comparison concerns system-level alternatives, with rankings reflecting differences in their overall performance profiles rather than comparisons between individual service attributes. The results are shown in Tables 6-8.

The global scores of the SAW method ($GS_{SAW,i}$) were computed using Eqns. 8–10, and the corresponding rankings ($RK_{SAW,i}$) were derived by ordering the alternatives in descending order by score. The results show that Metro stands out clearly as the top-ranked option. The remaining systems cluster more closely together in the global scores. This indicates a strong preference for heavy rail infrastructure in such a fast-growing urban environment.

The TOPSIS method ($GS_{TOP,i}$, $RK_{TOP,i}$), computed through Eqns. 11–18, presents some shifts in rankings. BRT slightly overtakes Metro, although the overall preference distribution remains relatively balanced. This reflects the method's sensitivity to proximity to the ideal solution.

VIKOR ($GS_{VIK,i}$, $RK_{VIK,i}$), calculated via Eqns. 19–23 with equal weight on utility and regret ($\beta = 0.5$), places LRT as the top-ranked option. However, the required compromise conditions (Eq. 24 and Table 8) are not fully met, indicating that no option satisfies both the principles of maximizing utility and minimizing regret.

Finally, the PROMETHEE II method ($GS_{PRO,i}$, $RK_{PRO,i}$), using Eqns. 25–32, confirms the strong dominance of Metro, highlighting its superior net flow over competitors. This consistency with SAW underlines Metro's robust perceived value in a developing megacity context.

Overall, the results reveal a recurring and consistent pattern: Metro dominates more, reflecting the city's pressing infrastructure and safety needs.

3.2.4. STAGE 4. Performing sensitivity analysis

STAGE 4 assessed the stability of the rankings through a sensitivity analysis by examining how variations in weights affected the results.

According to **STEP 4.1**, three scenarios were generated by assuming that the weights follow a normal distribution centered around their mean values. The standard deviation of the distributions was progressively increased—expressed as a percentage of the mean weight—to model different levels of noise: 0.1 (Low Noise → minor changes → high stability), 0.5 (Medium Noise → moderate changes → moderate sensitivity), and 1.0 (High Noise → major perturbations → extreme robustness testing). These thresholds were chosen to test the model's robustness across a wide range of perturbations, rather than derived from a specific prior study. Although noise levels vary depending on how variability ranges are defined, strict adherence to previously suggested thresholds is not mandatory. The method remains general, allowing these ranges to be determined using alternative approaches.

For each noise scenario, 10^6 instances (N) were simulated, and normalized weights were generated according to Eqns. (33) and (34). Although the baseline evaluation adopted a fixed neutral weight for the Investment Cost criterion, the sensitivity analysis required a broader perspective. Since this stage aims to test how each MCDMM responds to stochastic deviations around the user-derived mean weights, all criteria were perturbed, including the one (i.e., the investment cost) fixed in the baseline scenario.

Next, following **STEP 4.2**, metrics for global score-based stability ($\overline{SS_m}$) and global ranking-based stability ($\overline{RK_m}$) were computed, as defined in Eqns. (35) to (39). The results show that score-based stability under low noise is highest for the SAW method, followed by TOPSIS (Table 9). VIKOR and PROMETHEE II show notably lower score stability in this scenario, suggesting that even small perturbations in weights can remarkably affect their final scores. As noise increases, all methods experience a reduction in score stability. However, TOPSIS and VIKOR show more gradual declines under medium and high noise compared to SAW, and especially PROMETHEE II, which exhibits the lowest score stability at high noise levels.

In contrast, ranking-based stability presents a more nuanced picture (Table 10) Under low noise, PROMETHEE II exhibits outstanding ranking stability, indicating that minor variations in weights have

Table 6
Global scores and ranking of SAW, TOPSIS, VIKOR, and PROMETHEE II methods (the top-performing alternative for each method is highlighted in bold).

Alternative	SAW Score ($GS_{SAW,i}$)	SAW Rank ($RK_{SAW,i}$)	TOPSIS Score ($GS_{TOP,i}$)	TOPSIS Rank ($RK_{TOP,i}$)	VIKOR Score ($GS_{VIK,i}$)	VIKOR Rank ($RK_{VIK,i}$)	PROMETHEE II Score ($GS_{PRO,i}$)	PROMETHEE II Rank ($RK_{PRO,i}$)
Bus	0.489	2	0.485	3	0.678	4	−0.205	4
BRT	0.428	4	**0.525**	**1**	0.393	2	−0.006	2
LRT	0.430	3	0.435	4	**0.350**	**1**	−0.062	3
Metro	**0.740**	**1**	0.519	2	0.500	3	**0.272**	**1**

Table 7
Method-specific indicators underlying global scores and rankings.

Method	Indicator	Bus	BRT	LRT	Metro
TOPSIS	Δ_i^+	0.15	0.119	0.127	0.14
	Δ_i^-	0.141	0.131	0.098	0.151
VIKOR	S_i	0.665	0.516	0.558	0.308
	R_i	0.113	0.105	0.096	0.143
PROMETHEE II	$\varnothing^+(i)$	0.214	0.211	0.185	0.484
	$\varnothing^-(i)$	0.418	0.217	0.247	0.212

almost no impact on the final rankings. SAW also performs well in this respect. However, as the noise level increases, PROMETHEE II's ranking stability declines sharply—although it still outperforms all other methods, even under high-noise conditions.

The contrasting behavior between score-based and ranking-based stability might seem counterintuitive at first. However, a possible explanation lies in the different nature of each method: score-stable methods, such as SAW, tend to generate minor variations in absolute scores; yet these small changes may still lead to remarkable shifts in ranking when alternatives are closely clustered. Conversely, rank-stable methods like PROMETHEE II may experience greater variation in score magnitudes, but maintain stable rankings because the alternatives are more clearly differentiated, reducing the likelihood of ranking reversals even under noise.

4. Discussion

Interpreting the results from a planning and decision-making perspective highlights the interaction among citizen participation, methodological diversity, and robustness under noisy inputs. Together, these dimensions shape the credibility and transparency of public transport planning outcomes. By examining the ranking results alongside the stability analysis, several substantive insights emerge that address the research gaps identified in Section 1.2.

Summarizing and comparing the rankings obtained from all methods, a clear pattern emerges: rail-based modes dominate across three of the four decision-making methods, with Metro ranking first in SAW and PROMETHEE II, and LRT ranking first in VIKOR. Only TOPSIS points to a different outcome, placing BRT slightly ahead.

Specifically, Figure 3 shows that the ranking of the transport alternatives varies with the adopted MCDMM, although some alternatives maintain relatively stable positions across methods.

This coherence across three structurally different methods reinforces a well-established principle in the literature: mass-transit systems generally represent the most appropriate solution for fast-growing, high-density cities where infrastructure, crowding, and congestion justify high-capacity rail investments. Prior research mirrors this conclusion, with Metro-like modes consistently emerging as the preferred option across contexts, as evidenced in Tehran and South Korea (Nassereddine & Eskandari, 2017; Lee, 2018). The recurrent preference for Metro or LRT across methods therefore reflects user-driven priorities and broader empirical insights about the needs of megacities such as Ho Chi Minh City.

From a planning perspective, the predominance of Metro and LRT reflects evidence that high-capacity rail reduces congestion and emissions in rapidly motorizing cities. Recent studies show that urban rail systems shift trips from private vehicles, improving air quality and reducing peak-hour traffic and emissions (Cai et al., 2025; Tao et al., 2025). These benefits underscore Metro/LRT as strategic long-term investments, although their high coordination, financial, and implementation requirements may limit short-term feasibility compared to bus-based systems.

Beyond identifying Metro and LRT as consistently strong performers, the results directly address the three major gaps highlighted in Section 1.2.

First, the study provides empirical evidence on the feasibility and value of citizen participation in MCDMM-based transport evaluation (Gap 1). The criteria's weights used in all methods originate from a large, diverse sample of actual users, not experts. This allows the final rankings to genuinely reflect people's priorities. Interestingly, the dominance of Metro and LRT might be heavily influenced by citizens' strong emphasis on Vehicle and Safety (Table 5), suggesting that user preferences align with long-term strategic needs of dense or rapidly urbanizing cities (e.g., Ingvardson & Nielsen, 2022; Göransson & Andersson, 2023). At the same time, the survey phase also exposed the practical challenges of involving citizens in weighting processes: many respondents struggled with the cognitive load of AHP pairwise comparisons. This confirms prior concerns noted in the literature and demonstrates that while citizen participation broadens the decision-making perspective, it inevitably introduces judgment variability and noise, precisely the conditions under which robust analytical tools are needed (e.g., Pant et al., 2022; Wu & Tu, 2021).

Second, the study demonstrates why integrated MCDMM-based approaches are increasingly advocated (Gap 2). By applying SAW, TOPSIS, VIKOR, and PROMETHEE II in parallel, the evaluation benefits from complementary decision logics (compensatory, distance-to-ideal, and outranking), thus offering a fuller analytical perspective than any single method could provide (Velasquez & Hester, 2013; Yannis et al., 2020). Within this integrated view, a notable insight emerges: three decision frameworks, grounded in distinct conceptual principles, independently point toward rail-based solutions as the most appropriate option for Ho

Table 8
VIKOR method. Condition Checking.

Condition 1: Acceptance advantage	$G_{VIK,i^{(2)}} - G_{VIK,i^{(1)}}$	$1/(\lvert I\rvert - 1)$	**Outcome**
$GS_{VIK,i^{(2)}} - GS_{VIK,i^{(1)}} \geq 1/(\lvert I\rvert - 1)$	0.043	0.333	*Not satisfied*
Condition 2: Acceptable stability	S_i Or R_i	$\min_{i\in I}(S_i)$ Or $\min_{i\in I}(R_i)$	
$S(i^{(1)}) = \min_{i\in I}(S_i)$	0.558	0.308	*Not satisfied*
$R(i^{(1)}) = \min_{i\in I}(R_i)$	0.096	0.096	*Satisfied*

Table 9
Sensitivity analysis for Ho Chi Minh City. Score-Based Stability (colour coding, ranging from green to orange, is used to improve visual distinction across increasing noise levels; the most stable method for each noise level is highlighted in bold).

Noise Level	Method	Score-Based Stability
Low Noise	**SAW**	**52.32**
Low Noise	TOPSIS	22.99
Low Noise	VIKOR	6.43
Low Noise	PROMETHEE II	8.40
Medium Noise	**SAW**	**10.75**
Medium Noise	TOPSIS	5.41
Medium Noise	VIKOR	2.53
Medium Noise	PROMETHEE II	1.74
High Noise	**SAW**	**7.66**
High Noise	TOPSIS	4.15
High Noise	VIKOR	2.18
High Noise	PROMETHEE II	1.24

Score-based stability measures the average robustness of global scores to simulated noise perturbations. It is computed following Eqns. (35)–(36); a higher value denotes a more stable method.

Table 10
Sensitivity analysis for Ho Chi Minh City. Ranking-Based Stability (colour coding, ranging from green to orange, is used to improve visual distinction across increasing noise levels; the most stable method for each noise level is highlighted in bold).

Noise Level	Method	Ranking-Based Stability
Low Noise	SAW	2.50*10^8
Low Noise	TOPSIS	9.08
Low Noise	VIKOR	4.81
Low Noise	**PROMETHEE II**	**5.00*10^8**
Medium Noise	SAW	4.11
Medium Noise	TOPSIS	0.93
Medium Noise	VIKOR	1.13
Medium Noise	**PROMETHEE II**	**5.07**
High Noise	SAW	1.65
High Noise	TOPSIS	0.93
High Noise	VIKOR	1.05
High Noise	**PROMETHEE II**	**2.02**

Ranking-based stability captures the degree of rank preservation among alternatives under simulated noise perturbations. It is computed following Eqns. (37)–(39); a higher value denotes a more stable method.

Chi Minh City. This methodological concordance is important not only because of the specific alternatives it highlights, but because it shows how different evaluative logics tend to reinforce similar strategic conclusions, an effect that would remain invisible under a single-method assessment. Within this broader convergence, TOPSIS stands out as the only method identifying BRT as the preferred option. However, this divergence does not weaken the overall interpretability of results. On the contrary, it provides an instructive example of why parallel-method designs are needed for multidimensional problems: TOPSIS is the least stable method in ranking-based stability under medium and high noise, meaning its rankings are highly sensitive to the fluctuations typical of citizen-derived weights. As such, its preference for BRT might be interpreted not as a genuinely competing preference structure, but as the output of a method whose decision surface is more volatile when operating with noisy inputs.

Third, this observation connects to Gap 3, as the sensitivity analysis reveals that SAW and PROMETHEE II, i.e., the methods that consistently rank Metro first, are also the most robust under stochastic perturbations of weights (SAW for score-based stability and PROMETHEE II for ranking-based stability). Their ability to preserve rankings even when judgment variability is high indicates that convergence on rail-based systems is not an artifact of specific methodological assumptions but rather a stable, resilient outcome across noisy participatory conditions. Conversely, the instability of TOPSIS provides an empirical demonstration of why quantitative robustness assessment is crucial in citizen-based decision-making contexts.

These three dimensions are deeply interdependent. Citizen-derived weights introduce variability that makes robustness essential; robustness becomes meaningful only within a multi-method framework that reveals convergent or divergent patterns; and comparing multiple methods gains value precisely because the goal is to support inclusive, citizen-centered decision-making.

5. Conclusions and research perspectives

Multi-Criteria Decision-Making Methods (MCDMMs) are widely recognized as essential tools for supporting decision-makers, particularly in complex domains such as public transport planning. Most prior studies have relied predominantly on expert-driven evaluations based on single MCDMMs, almost disregarding the perspectives of everyday users. Moreover, while the existing literature has largely demonstrated the value of individual MCDMMs across various contexts, to the authors' knowledge, limited attention has been devoted to their integrated application for evaluating the influence of public opinion on the selection of transport alternatives. In addition, the comparison of multiple MCDMMs with respect to noise effects has received inadequate scholarly attention, despite its importance for assessing the robustness of decision outcomes in participatory contexts.

The main contributions of this study are:

- Incorporating citizen-derived input, introducing a novel participatory perspective to MCDMMs in urban transport planning.
- Explicitly combining user-driven preference weights with evidence-based performance data, ensuring that the evaluation combines societal priorities with objective system characteristics.
- Proposing an integrated framework for evaluating public transport alternatives through the combined use of multiple MCDMMs.
- Introducing two novel stability measures, i.e., score-based and ranking-based metrics, to quantitatively evaluate the robustness of multiple MCDMMs to noise according to a systematic sensitivity analysis.
- Applying for the first time this framework in Vietnam for selecting public transportation. This study is one of the first attempts to bring the public and experts together to select transit alternatives with a representative case study.

Incorporating citizen opinions offers planners an additional source of insight into user needs, helping align evaluations with everyday travel priorities and supporting more transparent and socially legitimate decision-making. Conducting a multidimensional trade-off analysis of transport alternatives using multiple MCDMMs and assessing their performance in a relevant case study could contribute to a more resilient and comprehensive decision-making framework. The joint consideration of score-based and ranking-based stabilities represents an underexplored dimension and shows that MCDMMs can behave differently across these two axes of robustness. Therefore, they provide practical guidance for selecting the most suitable MCDMM in uncertain decision environments and, hence, offer transport planners and practitioners a valuable tool for enhancing transparency and reliability in participatory decision-making. For example, when different MCDMMs yield conflicting scores and rankings, planners could rely on the stability indices introduced in this study to inform their final decision. Giving priority to the most stable method ensures that the resulting choice is less affected by random variability in user preferences and therefore more resilient in real-world applications.

Despite its robustness and comprehensiveness, this study opens several avenues for future research. In this respect, citizen input in the

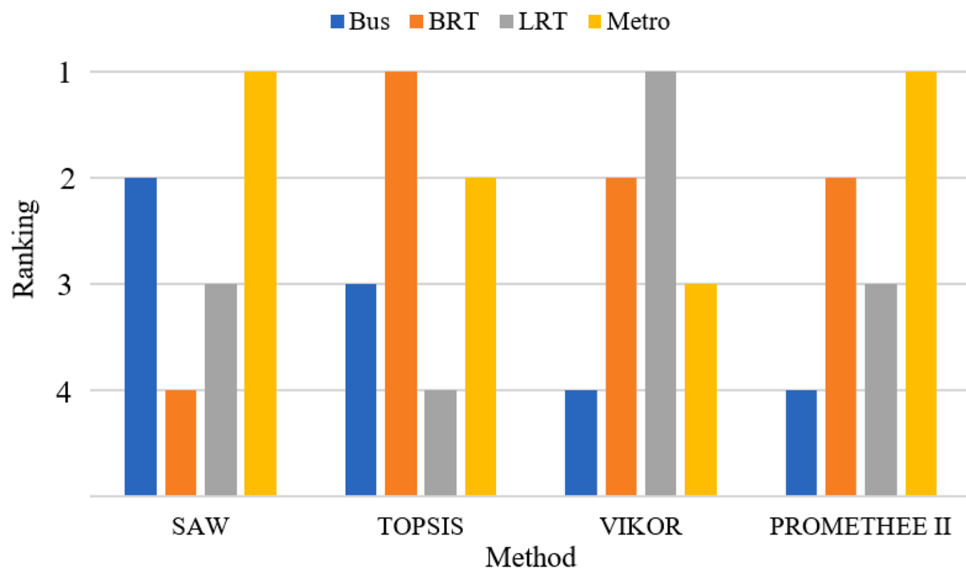

Figure 3. Ranking results obtained for the analysed transport alternatives (Bus, BRT, LRT, and Metro) using the four adopted MCDMMs: SAW, TOPSIS, VIKOR, and PROMETHEE II. The figure highlights the variability in rankings across methods and how some alternatives exhibit more stable performance than others.

proposed framework offers contextual and experiential insights that complement technical evaluation, but it is not interpreted as a driver of improved system performance. The heterogeneity observed in citizen-derived weights underscores the need to better understand how participatory inputs interact with formal evaluation methods.

First, future research could examine more explicitly whether, and under what conditions, user involvement increases the robustness of decision outcomes or contributes indirectly to planning effectiveness, moving toward a deeper assessment of potential links between participation and real-world impacts. Second, subsequent studies may compare alternative weighting techniques, such as the Best–Worst Method, with the AHP-based elicitation adopted here, to assess how different preference-derivation mechanisms affect variability, consistency, and the robustness of transport-mode rankings under citizen-generated inputs (e.g., Koohathongsumrit & Chankham, 2023; Kumar et al., 2024). Third, an additional research line would be to expand the set of evaluation criteria to incorporate additional dimensions, such as environmental impact and social equity. However, such extensions must be carefully balanced with the need to keep citizen-based pairwise comparisons cognitively manageable. Addressing these dimensions would also help mitigate potential omitted-variable bias present in the current specification. Fourth, group-based analyses of criterion weights (e.g., by age, income, trip purpose, or transport mode usage) could be explored to examine more explicitly how respondents' socio-demographic and behavioral characteristics influence the weighting process. Such group-level comparisons could reveal heterogeneities that remain hidden in aggregate results and provide a more nuanced understanding of how different population segments prioritize public transport attributes. Fifth, future research could extend the current analysis by incorporating scenario-based sensitivity analyses aimed at evaluating how variations in criteria weights influence the ranking results. Specifically, testing alternative weighting configurations for economic, environmental, and operational criteria could provide additional insights into the structural robustness and stability of the decision-making outcomes.

Finally, the multicriteria evaluation could be complemented by impact-oriented transport modelling to examine how each alternative may affect congestion, travel times, and modal shares. Such analyses, while valuable for reinforcing the link between ranking outcomes and planning implications, require dedicated modelling frameworks and therefore fall beyond the scope of this study, which is intended as an early-stage decision-support tool.

Overall, this study contributes to research on sustainable urban mobility planning by strengthening the methodological bases of participatory decision-making and by providing robust analytical tools to support planning choices in rapidly changing urban contexts.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work, the authors used ChatGPT (OpenAI) to improve English language and readability. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

CRediT authorship contribution statement

Roberto Ventura: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Martina Carra:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation. **Benedetto Barabino:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors report there are no competing interests to declare.

Acknowledgments

This research was partially funded by the University of Brescia & MUR through the Grant "CUP: D75F21002920001" PON R&I 2014-2020 (FSE REACT-EU). The authors also wish to acknowledge that the first version of this work was developed as part of the Master's thesis of Dr. Phan Nguyen Minh Tan.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.urbmob.2026.100247.

References

Aczél, J., & Saaty, T. L. (1983). Procedures for synthesizing ratio judgements. *Journal of Mathematical Psychology, 27*(1), 93–102. https://doi.org/10.1016/0022-2496(83)90028-7

Ahmed, T., Ahsan, A., Khan, M. H. R. B., Nahian, T. K., Antar, R. H., Hasan, A., … Imteaz, M. (2024). Comprehensive study on the selection and performance of the best electrode pair for electrocoagulation of textile wastewater using multi-criteria decision-making methods (TOPSIS, VIKOR and PROMETHEE II). *Journal of Environmental Management, 363*, Article 121337. https://doi.org/10.1016/j.jenvman.2024.121337

Al Suleiman, S., et al. (2023). How to improve public transport usage in a medium-sized city: key factors for a successful bus system. *European Transport Research Review, 15*(16). https://doi.org/10.1186/s12544-023-00616-y

Assefa, N. G., Carra, M., Ventura, R., Richiedei, A., & Barabino, B. (2024). Assessing tram-Train feasibility via multicriteria approach: the case of Brescia (Italy). *Research in Transportation Business & Management.* , Article 101178. https://doi.org/10.1016/j.rtbm.2024.101178

Awasthi, A., Omrani, H., & Gerber, P. (2018). Investigating ideal-solution based multicriteria decision making techniques for sustainability evaluation of urban mobility projects. *Transportation Research Part A: Policy and Practice, 116*, 247–259. https://doi.org/10.1016/j.tra.2018.06.007

Aydin, N., Cari, M., Kara, B., & Ayyildiz, E. (2025). A review of artificial intelligence-enhanced fuzzy multi-criteria decision-making approaches for sustainable transportation planning. *Computers, Materials & Continua, 85*(2), 1785–1805. https://doi.org/10.32604/cmc.2025.067290

Barabino, B., Lai, C., Casari, C., Demontis, R., & Mozzoni, S. (2017). Rethinking transit time reliability by integrating automated vehicle location data, passenger patterns, and web tools. *IEEE Transactions on Intelligent Transportation Systems, 18*(4), 756–766. https://doi.org/10.1109/TITS.2016.2585342

Barabino, B. (2018). Automatic recognition of "low-quality" vehicles and bus stops in bus services. *Public Transport, 10*(2), 257–289. https://doi.org/10.1007/s12469-018-0180-8

Barabino, B., Cabras, N. A., Conversano, C., & Olivo, A. (2020). An integrated approach to select key quality indicators in transit services. *Social Indicators Research, 149*(3), 1045–1080. https://doi.org/10.1007/s11205-020-02284-0

Barabino, B., Deiana, E., & Mozzoni, S. (2013a). The quality of public transport service: the 13816 standard and a methodological approach to an Italian case. *Ingegneria Ferroviaria, 68*(5), 475–499.

Barabino, B., Di Francesco, M., & Mozzoni, S. (2013b). Regularity diagnosis by automatic vehicle location raw data. *Public transport, 4*(3), 187–208. https://doi.org/10.1007/s12469-012-0059-z

Barabino, B., Di Francesco, M., & Ventura, R. (2023). Evaluating fare evasion risk in bus transit networks. *Transportation Research Interdisciplinary Perspectives, 20*, Article 100854. https://doi.org/10.1016/j.trip.2023.100854

Beria, P., Maltese, I., & Mariotti, I. (2012). Multicriteria versus cost benefit analysis: a comparative perspective in the assessment of sustainable mobility. *European Transport Research Review, 4*(3), 137–152. https://doi.org/10.1007/s12544-012-0074-9

Borghetti, F., Carra, M., Besson, C., Matarrese, E., Maja, R., & Barabino, B. (2024). Evaluating alternative fuels for a bus fleet: an Italian case. *Transport Policy, 154*, 1–15. https://doi.org/10.1016/j.tranpol.2024.05.025

Bozorg-Haddad, O., Zolghadr-Asli, B., & Loáiciga, H. A. (2021). A handbook on multi-attribute decision-making methods. *A Handbook on Multi-Attribute Decision-Making Methods*. https://doi.org/10.1002/9781119563501

Brans, J. P., & De Smet, Y. (2016). PROMETHEE methods. *International Series in Operations Research and Management Science*, 233. https://doi.org/10.1007/978-1-4939-3094-4_6

Brans, J. P., Vincke, P., & Mareschal, B. (1986). How to select and how to rank projects: the Promethee method. *European Journal of Operational Research, 24*(2). https://doi.org/10.1016/0377-2217(86)90044-5

Broniewicz, E., & Ogrodnik, K. (2020). Multi-criteria analysis of transport infrastructure projects. *Transportation Research Part D, 83*, Article 102351. https://doi.org/10.1016/j.trd.2020.102351

Broniewicz, E., & Ogrodnik, K. (2025). Application Potential of MCDM/MCDA methods in transport—Literature review and case study. *Sustainability, 17*(17), 7671. https://doi.org/10.3390/su17177671

Cai, M., Cui, L., & Zhang, Y. (2025). Beyond metros: Pollution mitigation and environmental benefits in diverse transit systems. *Transportation.* https://doi.org/10.1007/s11116-025-, 10598–6.

Cain, B. E., Gerber, E. R., & Hui, I. (2020). Public perceptions of collaborative governance in transportation policy. *Political Research Quarterly, 73*(4), 1–14. https://doi.org/10.1177/1065912920943954

Cantwell, M., Caulfield, B., & O'Mahony, M. (2009). Examining the factors that impact public transport commuting satisfaction. *Journal of Public Transportation, 12*(2), 1–21.

Carra, M., Pavesi, F. C., & Barabino, B. (2023). Sustainable cycle-tourism for society: integrating multi-criteria decision-making and land use approaches for route selection. *Sustainable Cities and Society, 99*, Article 104905. https://doi.org/10.1016/j.scs.2023.104905

Carra, M., Ventura, R., & Barabino, B. (2025). Engaging citizens in public transit choice: insights from Vietnam and Italy. In *International Conference on Computational Science and Its Applications* (pp. 294–305). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-97663-6_26.

Carrara, E., Ciavarella, R., Boglietti, S., Carra, M., Maternini, G., & Barabino, B. (2021). Identifying and selecting key sustainable parameters for the monitoring of e-powered micro personal mobility vehicles. Evidence from italy. *Sustainability (Switzerland), 13*(16). https://doi.org/10.3390/su13169226

Castillo, H., & Pitfield, D. (2010). ELASTIC – A methodological framework for identifying and selecting sustainable transport indicators. *Transportation Research Part D, 15*(4), 179–188.

Centre for Liveable Cities, & Urban Land Institute (2017). Urban mobility: 10 cities leading the way in asia-Pacific.

Chatterjee, P., & Chakraborty, S. (2016). A comparative analysis of VIKOR method and its variants. *Decision Science Letters, 5*(4), 469–486. https://doi.org/10.5267/j.dsl.2016.5.004

Chen, F., Zhu, Y., Zu, J., Lyu, J., & Yang, J. (2022). Appraising road safety attainment by CRITIC-ELECTRE-FCM: A policymaking support for Southeast Asia. *Transport policy, 122*, 104–118. https://doi.org/10.1016/j.tranpol.2022.04.014

Ciardiello, F., & Genovese, A. (2023). A comparison between TOPSIS and SAW methods. *Annals of Operations Research, 325*(2), 967–994. https://doi.org/10.1007/s10479-023-05339-w

Dahri, Noura, & Abida, Habib (2017). Monte Carlo simulation-aided analytical hierarchy process (AHP) for flood susceptibility mapping in Gabes Basin (southeastern Tunisia). *Environmental earth sciences, 76*(7), 302. https://doi.org/10.1007/s12665-017-6619-4

De Aloe, M., Ventura, R., Bonera, M., Barabino, B., & Maternini, G. (2022). Applying cost–benefit analysis to the economic evaluation of a tram-train system: evidence from Brescia (Italy). *Research in Transportation Business and Management, 47*, Article 100916. https://doi.org/10.1016/j.rtbm.2022.100916

De Alwis, T. (1993). Mathematica and the power method. *International Journal of Mathematical Education in Science and Technology*, (6), 24. https://doi.org/10.1080/0020739930240605

Demir, G., & Moslem, S. (2025). Fuzzy multicriteria decision making for assessing public transportation systems for sustainable transportation. *Intelligent Urban Mobility* (pp. 229–258). Academic Press.

Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). (2024). June 2024 summary Report: developing an e-mobility action plan in Ho Chi Minh City. *NDC Transport Initiative for Asia*. Retrieved from https://changing-transport.org.

Donais, F. M., Abi-Zeid, I., Waygood, E. O. D., & Lavoie, R. (2019). A review of cost–benefit analysis and multicriteria decision analysis from the perspective of sustainable transport in project evaluation. *EURO Journal on Decision Processes, 7*(3), 327–358. https://doi.org/10.1007/s40070-019-00098-1

Du, YW., & Shan, YK. (2021). A dynamic intelligent recommendation method based on the analytical ER rule for evaluating product ideas in large-scale group decision-making. *Group Decis Negot, 30*, 1373–1393. https://doi.org/10.1007/s10726-020-09687-x

Duleba, S., & Moslem, S. (2021). User satisfaction survey on public transport by a new PAHP based model. *Applied Sciences, 11*(21), Article 10256. https://doi.org/10.3390/app112110256

Dyer, J. S. (1990). Remarks on the analytic hierarchy process. *Management Science, 36*(3), 249–258.

ECORYS Nederland BV. (2005). *Infrastructure expenditures and costs: Final report.* Rotterdam, Netherlands: ECORYS Transport.

Fengmin, L., Dengfeng, W., Ying, X., Zihao, M., & Jing, C. (2025). Classification, selection of MCDM methods and robust decision-making in multidisciplinary design optimization of automotive structures. *Structural and Multidisciplinary Optimization, 68*(11), 232. https://doi.org/10.1007/s00158-025-04150-4

Ferretto, L., Carra, M., & Barabino, B. (2026). Enlarging the concept of quality of service: from a sectoral to a multimodal perspective. *Transportation Letters*, 1–16. https://doi.org/10.1080/19427867.2026.2636006

Fiagborlo, J. D., et al. (2024). Effects of travel time on mode choice for different locations and genders in Ghana. *SN Social Sciences, 4*(172). https://doi.org/10.1007/s43545-024-00974-x

Forman, EH., & Gass, S. I. (2001). The analytical hierarchy process-an exposition. *Operations Research, 49*(4), 469–487.

Friman, M., Advardsson, B., & Gärling, T. (2001). Frequency of negative critical incidents and satisfaction with public transport services. *Journal of Retailing and Consumer Services, 8*, 95–104.

Friman, M., Lättman, K., & Olsson, L. E. (2020). Public transport quality, safety, and perceived accessibility. *Sustainability, 12*(9), 3563. https://doi.org/10.3390/su12093563

Fung, A. (2006). Varieties of participation in complex governance. *Public administration review, 66*, 66–75. https://doi.org/10.1111/j.1540-6210.2006.00667.x

Göransson, J., & Andersson, H. (2023). Factors that make public transport systems attractive: a review of travel preferences and travel mode choices. *European Transport Research Review, 15*(32). https://doi.org/10.1186/s12544-023-00609-x

Güner, S. (2018). Measuring the quality of public transportation systems and ranking the bus transit routes using multi-criteria decision making techniques. *Case Studies on Transport Policy, 6*(2). https://doi.org/10.1016/j.cstp.2018.05.005

Ghorbanzadeh, O., Moslem, S., Blaschke, T., & Duleba, S. (2018). Sustainable urban transport planning considering different stakeholder groups by an interval-AHP decision support model. *Sustainability, 11*(1), 9. https://doi.org/10.3390/su11010009

Goepel, K. D. (2017). Implementation of an online software tool for the analytic hierarchy process-challenges and practical experiences. *International Journal of the Analytic Hierarchy Process, 10*(3). https://doi.org/10.13033/ijahp.v10i3.590

Greco, S., Ehrgott, M., & Figueira, J. R. (2016). *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer. https://doi.org/10.1007/978-1-4939-3094-4

Grošelj, P., Zadnik Stirn, L., Ayrilmis, N., & Kuzman, M. K. (2015). Comparison of some aggregation techniques using group analytic hierarchy process. *Expert Systems with Applications, 42*(4). https://doi.org/10.1016/j.eswa.2014.09.060

Hamurcu, M., & Eren, T. (2020). Electric bus selection with multicriteria decision analysis for green transportation. *Sustainability (Switzerland)*, (7), 12. https://doi.org/10.3390/su12072777

Harker, P. T., & Vargas, L. G. (1990). Reply to "remarks on the analytic hierarchy process" by JS Dyer. *Management Science, 36*(3), 269–273.

Hassan, M. N., Hawas, J. E., & Ahmed, K. (2013). A multi-dimensional framework for evaluating the transit service performance. *Transportation Research Part A, 50*, 47–61.

Hens, L., & De Wit, J. (2003). The development of indicators and core indicators for sustainable development: A state of the art review. *International Journal of Sustainable Development, 6*(4), 436–459. https://doi.org/10.1504/IJSD.2003.004188

Huang, H., Brusselaers, N., De Smet, Y., & Macharis, C. (2025). Engaging stakeholders in construction transport policy: A mass-participation framework. *Case Studies on Transport Policy, 19*, Article 101359. https://doi.org/10.1016/j.cstp.2024.101359

Ingvardson, J. B., & Nielsen, O. A. (2022). The influence of vicinity to stations, station characteristics and perceived safety on public transport mode choice. *Public Transport, 14*, 285–309. https://doi.org/10.1007/s12469-021-00285-x

Innes, J. E., & Booher, D. E. (2010). *Planning with complexity: An introduction to collaborative rationality for public policy*. Routledge. https://doi.org/10.1080/14649357.2010.525376

Institute for Transportation and Development Policy. (2024). *The BRT Standard: 2024 edition*. New York, NY: Institute for Transportation and Development Policy.

Kühnel, F., et al. (2025). Public transport accessibility and its effect on mode choice. *Urban Science, 9*(2), 49. https://doi.org/10.3390/urbansci9020049

Kalifa, M., Özdemir, A., Özkan, A., & Banar, M. (2021). Application of Multi-criteria decision analysis including sustainable indicators for prioritization of public transport system. *Integrated Environmental Assessment and Management, 18*(1), 25–38. https://doi.org/10.1002/ieam.4486

Kaya, Ö., Tortum, A., Alemdar, K. D., & Çodur, M. Y. (2020). Site selection for EVCS in Istanbul by GIS and multi-criteria decision-making. *Transportation Research Part D: Transport and Environment, 80*, Article 102271. https://doi.org/10.1016/j.trd.2020.102271

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press. https://doi.org/10.1017/CBO9781139174084

Keshavarz-Ghorabaee, M., et al. (2022). MCDM Approaches for evaluating urban and public transportation systems: A short review of recent studies. *Transport, 37*(6), 411–425. https://doi.org/10.3846/transport.2022.18376

Kilic, M., & Demirel, T. (2025). Fuzzy decision systems for sustainable public transport: A literature review. *International Journal of Computational Intelligence Systems, 18*, 164. https://doi.org/10.1007/s44196-025-00904-6

Kolios, A., Mytilinou, V., Lozano-Minguez, E., & Salonitis, K. (2016). A comparative study of multiple-criteria decision-making methods under stochastic inputs. *Energies, 9*(7), 566. https://doi.org/10.3390/en9070566

Koohathongsumrit, N., & Chankham, W. (2022). A hybrid approach of fuzzy risk assessment-based incenter of centroid and MCDM methods for multimodal transportation route selection. *Cogent Engineering, 9*(1), Article 2091672. https://doi.org/10.1080/23311916.2022.2091672

Koohathongsumrit, N., & Chankham, W. (2023). Route selection in multimodal supply chains: A fuzzy risk assessment model-BWM-MARCOS framework. *Applied Soft Computing, 137*, Article 110167. https://doi.org/10.1016/j.asoc.2023.110167

Kumar, R., & Pamucar, D. (2025). A comprehensive and systematic review of multi-criteria decision-making (MCDM) methods to solve decision-making problems: two decades from 2004 to 2024. *Spectrum of Decision Making and Applications, 2*(1), 24. https://doi.org/10.31181/sdmap21202524

Kumar, A., Nadeem, M., & Shameem, M. (2024). Assessment of DevOps lifecycle phases and their role in DevOps implementation using best–worst MCDM. *International Journal of Information Technology, 16*(4), 2139–2147. https://doi.org/10.1007/s41870-023-01566-3

Kundu, P., Görçün, Ö. F., Garg, C. P., Küçükönder, H., & Çanakçıoğlu, M. (2024). Evaluation of public transportation systems for sustainable cities using an integrated fuzzy multi-criteria group decision-making model. *Environment, Development and Sustainability, 26*, 27655–27684. https://doi.org/10.1007/s10668-023-03776-y

Le Pira, M., Inturri, G., Ignaccolo, M., & Pluchino, A. (2015). Analysis of AHP methods and the Pairwise Majority rule (PMR) for collective preference rankings of sustainable mobility solutions. *Transportation Research Procedia, 10*. https://doi.org/10.1016/j.trpro.2015.09.031

Lee, D. J. (2018). A multi-criteria approach for prioritizing advanced public transport modes (APTM) considering urban types in Korea. *Transportation Research Part A, 111*. https://doi.org/10.1016/j.tra.2018.02.005

Lee, E. H. (2024). eXplainable DEA approach for evaluating performance of public transport origin-destination pairs. *Research in Transportation Economics, 108*, Article 101491.

Levinson, H. S., Zimmerman, S., Clinger, J., Gast, J., Rutherford, S. C., & Bruhn, E. (2009). Bus versus rail: meta-analysis of cost characteristics, carrying capacities, and land use impacts. *Transportation Research Record, 2110*, 11–21. https://doi.org/10.3141/2110-11

Li, G., Kou, G., & Peng, Y. (2021). Heterogeneous large-scale group decision making using fuzzy cluster analysis and its application to emergency response plan selection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52*(6), 3391–3403. https://doi.org/10.1109/TSMC.2021.3068759

Linovski, O., & Baker, D. M. (2023). Community-designed participation: lessons for equitable engagement in transportation planning. *Transportation Research Record, 2677*(6), 172–181. https://doi.org/10.1177/03611981221145131

Liu, T., Cats, O., & Gkiotsalitis, K. (2021). A- review of public transport transfer coordination at the tactical planning phase. *Transportation Research Part C, 133*, Article 103450. https://doi.org/10.1016/j.trc.2021.103450

Macharis, C., & Bernardini, A. (2015). Reviewing the use of multi-criteria decision analysis for the evaluation of transport projects: time for a multi-actor approach. *Transport Policy, 37*, 177–186. https://doi.org/10.1016/j.tranpol.2014.11.002

Management Authority for Urban Railways. (2016). *Ho Chi Minh City Urban Railway Construction Project – Line 1 (Ben Thanh–Suoi Tien): Special Assistance for Project Implementation (SAPI) – Final Report*. Ho Chi Minh City, Vietnam: Management Authority for Urban Railways & Japan International Cooperation Agency.

Manzolli, J. A., Trovão, J. P., & Henggeler Antunes, C. (2021). Scenario-based multi-criteria decision analysis for rapid transit systems implementation in an urban context. *ETransportation, 7*. https://doi.org/10.1016/j.etran.2020.100101

Mardani, A., Jusoh, A., Nor, K., Khalifah, Z., Zakwan, N., & Valipour, A. (2015). Multiple criteria decision-making techniques and their applications–a review of the literature from 2000 to 2014. *Economic research-Ekonomska istraživanja, 28*(1), 516–571. https://doi.org/10.1080/1331677X.2015.1075139

Moradi, S., Sierpiński, G., & Masoumi, H. (2023). Selection of possible scenarios for improving the quality of public transport services through the use of hybrid Fuzzy-Mcdm models. *Scientific Journal of Silesian University of Technology, 119*, 189–204. https://doi.org/10.20858/sjsutst.2023.119.11. Series Transport.

Morfoulaki, M., & Papathanasiou, J. (2021a). Use of promethee mcda method for ranking alternative measures of sustainable urban mobility planning. *Mathematics*, (6), 9. https://doi.org/10.3390/math9060602

Moslem, S., Alkharabsheh, A., Ismael, K., & Duleba, S. (2020). An integrated decision support model for evaluating public transport quality. *Applied Sciences*, (12), 10. https://doi.org/10.3390/APP10124158

Nassereddine, M., & Eskandari, H. (2017). An integrated MCDM approach to evaluate public transportation systems in Tehran. *Transportation Research Part A, 106*. https://doi.org/10.1016/j.tra.2017.10.013

National Academies of Sciences, Engineering, and Medicine. (2013). *Transit Capacity and Quality of Service Manual* (Third Edition). Washington, DC: The National Academies Press. https://doi.org/10.17226/24766

Ngoc, A. M., & Nishiuchi, H. (2025). Understanding travel patterns of ride-hailing service sub-population groups and effects of transit investment on ride-hailing users' potential mode switching: A case study of Ho Chi Minh City, Vietnam. *Transport Policy, 162*, 456–476.

Ngossaha, J. M., Tatang, K. N. F., Ngamby, L. B. N., Mfopou, A., & Tsakou, S. B. (2024). Integrating fuzzy multicriteria decision-making approach for improving the quality of urban mobility services in developing countries. *Journal of Infrastructure, Policy and Development, 8*(8), 6183. https://doi.org/10.24294/jipd.v8i8.6183

Nguyen, T. N., Hung, K. V., Mai, N. T. H., Shiomi, Y., Nishiuchi, H., Nguyen-Ngoc, D., & Ngoc, A. M (2026). Multi-criteria decision-making framework for evaluating urban freight transport measures toward a sustainable transport system. *Innovative Infrastructure Solutions, 11*, 43. https://doi.org/10.1007/s41062-025-02391-2

Noureddine, M., & Ristic, M. (2019). Route planning for hazardous materials transportation: multi-criteria decision-making approach. *Decision Making: Applications in Management and Engineering, 2*(1). https://doi.org/10.31181/dmame1901066n

NSO (National Statistics Office). (2023). *Population and employment*. Retrieved from https://www.gso.gov.vn/.

O'Sullivan, S., & Morrall, J. (1996). Walking distances to and from light-rail transit stations. *Transportation Research Record, 1538*(1), 19–26. https://doi.org/10.3141/1538-03

Olivo, A., Maternini, G., & Barabino, B. (2019). Empirical study on the accuracy and precision of automatic passenger counting in European bus services. *The Open Transportation Journal, 13*(1), 250–260. https://doi.org/10.2174/1874447801913010250

Opricović, S. (2009). Compromise in cooperative game and the VIKOR method. *Yugoslav Journal of Operations Research, 19*(2), 225–238. https://doi.org/10.2298/yjor0902225o

Ozsahin, D. U., Gökcekus, H., Uzun, B., & LaMoreaux, J. W. (2021). *Application of multi-criteria decision analysis in environmental and civil engineering* (pp. 37–41). Cham, Switzerland: Springer.

Pant, S., Kumar, A., Ram, M., Klochkov, Y., & Sharma, H. K. (2022). Consistency indices in analytic hierarchy process: A review. *Mathematics, 10*(8), 1206. https://doi.org/10.3390/math10081206

Peng, X., Wang, H., & Wu, Z. (2019). A multi-hesitant fuzzy ELECTRE I method for choosing third-party logistics services. *Applied Soft Computing, 74*, 667–682. https://doi.org/10.1016/j.asoc.2018.10.051

Phan, N. M. T. (2023). Prioritizing modes of public transportation using multi-criteria decision-making methods: the two case studies in Ho Chi Minh City (Vietnam) and Brescia (Italy). *Master's thesis*. University of Brescia.

Roy, B. (1990). The outranking approach and the foundations of electre methods. *Theory and decision, 31*(1), 49–73. https://doi.org/10.1007/978-3-642-75935-2_8

Rupprecht Consult. (2019). *Guidelines for Developing and Implementing a Sustainable Urban Mobility Plan* (Second Edition). Retrieved from:. https://urban-mobility-observatory.transport.ec.europa.eu/system/files/2023-09/sump_guidelines_2019_second%20edition.pdf.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology, 15*(3), 234–281. https://doi.org/10.1016/0022-2496(77)90033-5

Saaty, T. L. (1980a). *The Analytic Hierarchy Process*. McGraw Hill.

Saaty, T. L. (1980b). The analytical hierarchy process, planning, priority. *Priority Setting. Resource Allocation*. MacGraw-Hill, New York International Book Company.

Saaty, R. W. (1987). The analytic hierarchy process-what it is and how it is used. *Mathematical Modelling, 9*(3-5), 161–176.
Saaty, T. L. (1990). An exposition of the AHP in reply to the paper "remarks on the analytic hierarchy process. *Management Science, 36*(3), 259–268.
Saaty, T. L. (1994). Highlights and critical points in the theory and application of the analytic hierarchy process. *European Journal of Operational Research, 74*(3), 426–447.
Sato, Y., & Tan, K. H. (2023). Inconsistency indices in pairwise comparisons: an improvement of the Consistency Index. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04431-3
Serrano-Gomez, L., & Munoz-Hernandez, J. I. (2019). Monte Carlo approach to fuzzy AHP risk analysis in renewable energy construction projects. *PLoS ONE, 14*(6), Article e0215943. https://doi.org/10.1371/journal.pone.0215943
Shabani, A., Shabani, A., Ahmadinejad, B., & Salmasnia, A. (2022). Measuring the customer satisfaction of public transportation in Tehran during the COVID-19 pandemic using MCDM techniques. *Case Studies on Transport Policy, 10*(3). https://doi.org/10.1016/j.cstp.2022.05.009
Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 99–118. https://doi.org/10.2307/1884852
Sipetas, C., Huang, Z., & Espinosa Mireles de Villafranca, A. (2024). Evaluation framework for multi-modal public transport systems based on connectivity and transfers at stop level. *Transportation Research Record, 2678*(10), 147–162. https://doi.org/10.1177/03611981241231963
Soza-Parra, J., Raveau, S., & Muñoz, J. C. (2021). Public transport reliability across preferences, modes, and space. *Transportation, 49*, 621–640. https://doi.org/10.1007/s11116-021-10188-2
Tao, Z., Feng, X., & Wang, H. (2025). The spatiotemporal and multilevel impacts of metro on urban environmental protection. Scientific Reports, 15, Article 98855-7. https://www.nature.com/articles/s41598-025-98855-7.
Taylor, B. D., & Morris, E. A. (2015). Public transportation objectives and rider demographics: are transit's priorities poor public policy? *Transportation, 42*, 347–367. https://doi.org/10.1007/s11116-014-9547-0
Tian, C., Peng, J. J., Zhang, S., Wang, J. Q., & Goh, M. (2021). A sustainability evaluation framework for WET-PPP projects based on a picture fuzzy similarity-based VIKOR method. *Journal of Cleaner Production, 289*, Article 125130. https://doi.org/10.1016/j.jclepro.2020.125130
Transformative Urban Mobility Initiative. (2024). Deep Dive City: Ho Chi Minh City. Retrieved from https://transformative-mobility.org.
Transit Costs Project. (2023). *Database of urban rail construction costs*. Transport Policy Lab, Massachusetts Institute of Technology & NYU Marron Institute. Retrieved in 2023 from https://transitcosts.com.
Velasquez, M., & Hester, P. T. (2013). An Analysis of Multi-Criteria Decision Making Methods. In. *International Journal of Operations Research, 10*(2).
Ventura, R., Bonera, M., Carra, M., Barabino, B., & Maternini, G. (2022). Evaluating the viability of a Tram-Train system. A case study from Salento (Italy). *Case Studies on Transport Policy, 10*(3), 1945–1963. https://doi.org/10.1016/j.cstp.2022.08.009
Vuchic, V. R. (2007). Urban Transit Systems and Technology. *Urban Transit Systems and Technology*. https://doi.org/10.1002/9780470168066
Wang, T., & Lee, H. (2009). Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. *Expert Systems with Applications, 36*(5), 8980–8985. https://doi.org/10.1016/j.eswa.2008.11.035
Wedley, W. C. (1993). Consistency prediction for incomplete AHP matrices. *Mathematical and Computer Modelling, 17*(4–5), 151–161. https://doi.org/10.1016/0895-7177(93)90183-Y
Wicaksono, F. D., Arshad, Y. B., & Sihombing, H. (2020). Norm-dist Monte-Carlo integrative method for the improvement of fuzzy analytic hierarchy process. *Heliyon, 6*(4). https://doi.org/10.1016/j.heliyon.2020.e03607
Wind, Y., & Saaty, T. L. (1980). Marketing applications of the analytic hierarchy process. *Management Science, 26*(7), 641–658.
Wu, Z., & Tu, J. (2021). Managing transitivity and consistency of preferences in AHP group decision making based on minimum modifications. *Information Fusion, 67*, 125–135. https://doi.org/10.1016/j.inffus.2020.10.012
Wu, Z. (2023). Evaluation of provincial economic resilience in China based on the TOPSIS-XGBoost-SHAP model. *Journal of Mathematics, 2023*(1), Article 6652800.
Yannis, G., Kopsacheili, A., Dragomanovits, A., & Petraki, V. (2020). State-of-the-art review on multi-criteria decision-making in the transport sector. *Journal of Traffic and Transportation Engineering (English Edition), 7*(4), 413–431. https://doi.org/10.1016/j.jtte.2020.05.005
Yaraghi, N., Tabesh, P., Guan, P., & Zhuang, J. (2014). Comparison of AHP and Monte Carlo AHP under different levels of uncertainty. *IEEE Transactions on Engineering Management, 62*(1), 122–132. https://doi.org/10.1109/TEM.2014.2360082
Zavadskas, E. K., & Turskis, Z. (2011). Multiple criteria decision making (MCDM) methods in economics: an overview. *Technological and Economic Development of Economy, 17*(2), 397–427. https://doi.org/10.3846/20294913.2011.593291
Zelany, M. (1974). A concept of compromise solutions and the method of the displaced ideal. *Computers & Operations Research, 1*(3-4), 479–496.
Zhang, C., Wang, M., Dong, J., Lu, W., Liu, Y., Ni, A., & Yu, X. (2022). Factors and mechanism affecting the attractiveness of public transport: macroscopic and microscopic perspectives. *Journal of Advanced Transportation, 2022*(1), 1–16. https://doi.org/10.1155/2022/5048678