Expert Systems With Applications 334 (2027) 134456

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Development and application of HELDA: a multi-criteria decision analysis software tailored for sustainable energy system transitions

Laura Sofia Mesa Estrada [a,*,1], Martina Haase [a,2], Tim Müller [b,3]

[a] *Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Karlsruhe 76021, Germany*
[b] *Institute for Thermal Energy Technology and Safety, Karlsruhe Institute of Technology, Karlsruhe 76021, Germany*

A R T I C L E I N F O

Keywords:
MCDA
Decision analysis tool
Stakeholder integration
Sustainability assessment
Energy transition

A B S T R A C T

Achieving sustainable energy system transitions necessitates the development of innovative technologies and systems, and their integration into complex and rapidly changing societal frameworks. In such a dynamic context, it is important to have a tool that accelerates decision-making processes, including stakeholder preferences, while maintaining scientific rigor, transparency and inclusiveness. To achieve this objective, sustainability assessments (SA) with multi-criteria decision analysis (MCDA) methods have been widely used. Despite the development of MCDA software tools to facilitate the implementation of MCDA methods, challenges often arise concerning the selection of a proper MCDA method and the integration of stakeholders, which often requires significant resources. This research presents HELDA, the Helmholtz Decision Analysis Tool developed following the concepts of MCDA for SA including the strong sustainability concept and stakeholders' integration. To our knowledge, HELDA is the only free MCDA software that integrates binary relations methods (ELECTRE III and PROMETHEE I and II) with online surveys for weights elicitation of stakeholder groups (synchronous and asynchronous). The HELDA software is the result of extending the functionality of an existing MCDA software based on a previous study on the capabilities of MCDA software to conduct SA. HELDA development process was based on theories from agile software development. The functionality of HELDA is illustrated in a case study on the prospective SA of individual motorized mobility options using ELECTRE III method. The preferences of 13 researchers were elicited as individuals and in group settings to reach consensus on criteria weights.

1. Introduction

In the context of sustainable energy system transitions, a range of priorities have been identified, including, e.g., decarbonization, decentralization, digitalization and diversification, environmental protection and human well-being. This is a complex matter that involves political, economic, environmental, and ethical considerations, as well as the perspectives of relevant societal groups (Grunwald, 2017). Sustainability assessment (SA) methodologies have been employed to address this issue, thereby introducing diverse dimensions to the aforementioned decision-making process (Ness, Urbel-Piirsalu, Anderberg, & Olsson, 2007) and integrating different stakeholder perspectives (Turkson, Acquaye, Liu, & Papadopoulos, 2020). In decision-making processes related to sustainability, the joint participation of experts with other types of stakeholders, e.g. industry, government, and civil society is desirable (Munda, 2004). Multi-criteria decision analysis (MCDA) methods have been demonstrated to be a powerful tool for conducting SA (Lindfors, 2021).

MCDA can be distinguished into multi-objective decision-making (MODM) and multi-attribute decision-making (MADM) (Kumar et al., 2017). MADM and its respective methods address decision problems that involve a finite number of explicit alternatives and criteria, while in MODM, the alternatives are only implicitly known, and the criteria are expressed as mathematical objective functions. MADM methods can be categorized using the type of aggregation into: (i) scoring functions (e.g., weighted sum method, TOPSIS, AHP), (ii) binary relations (e.g., PROMETHEE, ELECTRE), and (iii) decision rules (e.g., DEX, DRSA) (Cinelli, Kadziński, Miebs, Gonzalez, & Słowiński, 2022). These methods

* Corresponding author.
E-mail address: laura.mesa-estrada@kit.de (L.S. Mesa Estrada).
[1] https://orcid.org/0000-0001-5492-627X.
[2] https://orcid.org/0000-0002-9098-3930.
[3] https://orcid.org/0000-0002-1405-7409.

https://doi.org/10.1016/j.eswa.2026.134456
Received 11 March 2026; Received in revised form 10 September 2026; Accepted 15 September 2026
Available online 16 September 2026

Open access funding enabled and organized by Projekt DEAL.

have different strengths and weaknesses, and their application depends on the decision problem and type of information available (Cinelli, Burgherr, Kadziński, & Słowiński, 2022; Wątróbski, Jankowski, Ziemba, Karczmarczyk, & Ziolo, 2019b). This paper focuses exclusively on MADM, and henceforth, the term MCDA will be used to refer specifically to MADM. In order to select a suitable MCDA method, it is first necessary to consider the characteristics of each decision-making problem (Cinelli et al., 2022b). These characteristics include, but are not limited to: the type of aggregation, the type of decision recommendation, the type of scale, and the level of compensation between criteria. There is extensive literature available describing the characteristics of decision-making problems in SA and suitable MCDA methods for modeling such decision-making problems, including Munda (2005), Cinelli, Coles, and Kirwan (2014), and Lindfors (2021). For instance, the selection of MCDA methods with null or partial level of compensation between criteria to model the strong concept of sustainability, for which different natural capital, e.g. ecosystems, biodiversity or climate, is not substitutable with man-made capital (Ziemba, 2019).

Free MCDA software has been developed to support users with the selection and application of MCDA methods. For method selection, there are two important contributions that provide a framework on a dedicated website. The first one is the MCDA Method Selection Tool[4], developed as an exemplary decision support tool based on the research of Wątróbski, Jankowski, Ziemba, Karczmarczyk, and Ziolo (2019a). The study includes 56 MCDA methods, described in 9 decision-making problem characteristics. The second one is the MCDA-Methods Selection Software[5] (MCDA-MSS) developed by Cinelli et al. (2021). This software is a more comprehensive solution, including 205 MCDA methods and 156 decision-making characteristics grouped in four sections: problem typology, preference model, elicitation of preferences, and exploitation of the preference relation. For method application, several different MCDA software exist to support users in the implementation of their decision analysis process. These software solutions have been developed by experts in MCDA methods, recognizing the need to provide software solutions for their methods. Several sourcesdescribe the available MCDA software and their capabilities; for example, the ORMS Today survey (Beekman, 2020) and the book chapter on MCDA software by Weistroffer and Li (2016). Reviews of the application of MCDA software to different types of decision-making problems are also available, including health systems management (Moreno-Calderón, Tong, & Thokala, 2020), environmental planning (Mustajoki & Marttunen, 2017), and sustainability assessment (Mesa Estrada, Haase, Baumann, & Cinelli, 2026). Despite the significant efforts and contributions of the MCDA community to inform and guide users on the theoretical and mathematical background of MCDA methods for supporting different decision-making problems, there has been a vast amount of improper applications of MCDA methods in the energy context, as demonstrated by Cinelli et al. (2022a). For sustainable energy transitions, MCDA is increasingly used by non-MCDA-experts in order to assess energy technologies and systems, considering several criteria simultaneously. As mentioned above, non-compensatory aggregation methods are preferred following the strong sustainability concept. Another relevant consideration is the importance of stakeholders' integration in SA with MCDA (SA-MCDA), as well as the assessment of related uncertainties. However, it is often the case that stakeholders are not involved because of time and resource constraints (Wulf et al., 2025).

Observing this phenomenon, the hypothesis with which this research started was that it was not enough to provide MCDA software suitable for SA, but to provide additionally a framework, specifying step by step how the software could be implemented. Therefore, two steps are necessary: first, developing a software that meets the demands for SA, and second, proposing the framework. This contribution focuses on the first step, presenting the process of developing HELDA, the Helmholtz Decision Analysis Tool, its main features, and an exemplary application. Recognizing the relevant efforts from different developers to provide MCDA software, a free software was selected as a baseline and enhanced to strengthen its capabilities for SA. The MCDA ToolKIT (Müller, Bai, & Raskob, 2020), a Java desktop application developed by KIT, was selected as baseline software. The reasons for its selection were its modular design, a plug-in subsystem that gives the possibility to easily extend functionality and to connect to other applications, and the reachability of the developers to contribute in the enhancement of the software. The process of software enhancement was based on the systematic analysis of MCDA software requirements for SA of Mesa Estrada et al. (2026). This systematic analysis presents 29 criteria related to SA, grouped in 8 domains: applicability and accessibility; problem typology; problem structure; preference model; stakeholder involvement; output variability analysis; transparency; and utility. Every criterion is defined using a qualitative scale to indicate strong, moderate, and weak fulfilment. The main goal of developing HELDA was to provide a software solution that supports SA-MCDA, including the concept of strong sustainability and accelerating stakeholder integration. By doing so, this enables more efficient and transparent decision-making processes in complex socio-technical environments.

This paper makes the following main contributions:

- Present a methodology to assess and enhance an existing MCDA software with respect to its use for SA
- Develop the MCDA software HELDA, tailored for SA with a combination of established and unique features for criteria weighting, criteria aggregation, stakeholders' integration, and output variability analysis
- Illustrate the application of HELDA in a use case for SA-MCDA of individual motorized mobility

The paper is organized as follows: Section 2 presents the state of the art of MCDA software for SA, while Section 3 presents the HELDA software development process and its main features for SA. In Section 4, the functionality of HELDA is demonstrated in a use case related to the SA of energy technologies. Section 5 discusses HELDA software limitations and future work. Finally, Section 6 concludes the main contributions of this work.

2. MCDA software for SA: state of the art

This section presents the state of the art of free MCDA software using a selection of tools chosen by the authors of this study for their key contributions related to fields relevant to SA, i.e. group decision-making, environmental management, and public policy. The selection was based on the review of Mesa Estrada et al. (2026) which provides a comprehensive comparison of MCDA software for SA. Table 1 presents a comparison of the selected software in terms of features for preference modeling, stakeholder integration and output variability analysis. The following paragraphs describe the software tools using each item.

2.1. Preference modeling

In the theory of MCDA for SA, there are two concepts for modeling decision problems: the strong and weak sustainability concepts. When modeling the strong sustainability concept, methods that handle quantitative and qualitative scales, weights as importance coefficients, null or partial compensation between criteria, and discriminating thresholds are preferred (Mesa Estrada et al., 2026). These requirements are typically fulfilled with binary relations type of aggregation methods as shown for SOCRATES (Munda, 2022) in Table 1. Scoring methods could have different levels of compensation depending on the type of aggregation function implemented (Cinelli, Spada, Kim, Zhang, & Burgherr,

[4] http://www.mcda.it

[5] https://mcda.cs.put.poznan.pl/

Table 1
Comparison of selected MCDA software for applications in sustainability assessment. TO: trade-offs, RIC: relative importance coefficients.

		Software (authors)	Apollo-Live (Koasidis et al., 2025)	Entscheidungsnavi (Peters, Tönsfeuerborn, & von Nitzsch, 2024)	ValueDecisions (Haag, Aubert, & Lienert, 2022)	MAMCA (Huang, Burgherr, & and Macharis, 2025)	AHP-OS (Goepel, 2018)	SOCRATES (Munda, 2022)
Preference modeling		Aggregation method	2-tuple group TOPSIS	MAUT	MAVT (non-additive model)	AHP, SMART	AHP	Kemeny median ranking
		Type of aggregation	Scoring function	Scoring function	Scoring function	Scoring function	Scoring function	Binary relations
		Type of information (scale)	relative	quantitative; qualitative	quantitative; qualitative	relative	relative	quantitative; qualitative
		Compensation level between criteria	Full	Full	Full – partial	Full	Full	Null
		Type of weights	Imprecise: linguistic scale	Precise-TO	Precise-TO	Precise-RIC	Precise-RIC	Precise-RIC
Stakeholder integration		Approach for preference elicitation*	Online surveys	Group session	None	Online surveys	Group session	None
		Problem structuring	x	✓	x	✓	x	x
		Weights elicitation	✓	✓	x	✓	✓	x
Output variability analysis Yes (✓), No (x)	Sensitivity analysis	Performance values	x	✓	x	x	✓	✓
		Weights	x	✓	✓	✓	x	✓
		Model	x	x	✓	x	x	x
	Uncertainty analysis	Performance values	x	✓	✓	x	x	x
		Weights	x	✓	x	x	x	✓
		Model	x	x	x	x	x	x

* This refers only to a solution supported by the software different to entering criteria weights.

2021). The software ValueDecisions (Haag, Aubert, & Lienert, 2022) uses weighted power mean, a non-additive aggregation function suitable for partial compensation. However, the weights used by the method represent trade-offs, i.e. compensatory rates. The software tools Entscheidungsnavi (Peters, Tönsfeuerborn, & von Nitzsch, 2024), MAMCA (Huang, Burgherr, & and Macharis, 2025), AHP-OS (Goepel, 2018), and Apollo-Live (Koasidis et al., 2025) use additive aggregation functions that are suitable for full compensation levels, modeling the weak sustainability concept. Other MCDA software tools supporting the strong concept of sustainability with binary relations methods include PROMETHEE Cloud (Pohl & Geldermann, 2024), MCDA-ULaval (Couture-Grenier & Abi-Zeid, 2015), and Visual PROMETHEE (Mareschal & Smet, 2009). These were not included in this analysis for not providing support for stakeholder integration.

2.2. Stakeholder integration

The dynamic nature of SA and the need to elicit stakeholders' preferences in related decision-making processes require tools that can reach a broader audience of participants (Mesa Estrada et al., 2026). Table 1 compares the software capabilities for stakeholder integration, as determined by the approach for eliciting group preferences, and the extent to which the integration supports problem structuring and/or criteria weights. MAMCA and Apollo-Live use online surveys for synchronous, i.e. real-time preference elicitation, group decision-making to support weights elicitation. Furthermore, MAMCA supports problem structuring by enabling participants to make suggestions to add or delete criteria. The integration of stakeholders in MAMCA is characterized by the ability of handling distinct criteria sets for each stakeholder group. Apollo-Live differentiates itself by the calculation of consensus using individual responses. Entscheidungsnavi and AHP-OS use group sessions in which stakeholders are invited to provide criteria weights. Entscheidungsnavi has a robust module for problem structuring based on the value-focused thinking (VFT) approach (Keeney, 1996). ValueDecisions does not provide a functionality for the elicitation of group preferences; however, it does facilitate the analysis of such data from diverse stakeholders. Similarly, SOCRATES allows preferences from stakeholders to be manually entered directly into the software by the user. However, SOCRATES differs from the rest of the approaches assessed in that stakeholder or social actors' preferences are not directly used in the evaluation matrix, but instead in an "equity matrix" exclusive for such input.

2.3. Output variability analysis

The output variability analysis focuses on assessing the effect in the model of varying performance values, weights and/or model parameters (i.e. discriminating thresholds, normalization method and aggregation function) (Mesa Estrada et al., 2026). The distinction between sensitivity analysis and uncertainty analysis lies in the scope of parameter variation: sensitivity analysis examines the effect of varying one parameter at a time, whereas uncertainty analysis considers simultaneous variations (Cinelli, Spada, et al., 2021). Apollo-Live uses fuzzy sets to incorporate uncertainty in performance values and weights elicited from stakeholders, and does not support output variability analysis. Most software tools support sensitivity analysis of criteria weights, with Entscheidungsnavi and SOCRATES extending this capability to include both weights and performance values. AHP-OS represents an exception, offering sensitivity analysis exclusively for performance values. Notably, ValueDecisions is the only software that supports sensitivity analysis of model parameters. Capabilities for uncertainty analysis are more limited across the reviewed software tools. Entscheidungsnavi offers the most comprehensive support, enabling uncertainty analysis for both performance values and weights through Monte Carlo simulation (MCS). ValueDecisions restricts this functionality to performance values only, whereas SOCRATES provides uncertainty analysis exclusively for weights.

2.4. Research gap

This section confirms the significant efforts from MCDA experts and software developers to improve the accessibility of MCDA methods for different types of users in terms of preference modeling, stakeholder integration and output variability analysis. However, it reveals that MCDA software tools that support the integration of stakeholders, with e.g. online surveys, are commonly based on scoring functions which support weak sustainability concepts through fully compensatory methods and/or weights as trade-offs. Additionally, users would benefit from more comprehensive and accessible features for output variability analyses that empower them to gain deeper insights from their models, fostering more reliable and interpretable outcomes. To address these gaps, we developed the HELDA software to support SA integrating three core dimensions: (1) strong sustainability concept with binary relations methods (ELECTRE III and PROMETHEE I/II) that accommodate qualitative and quantitative scales with null/partial compensation; (2) stakeholder integration with preferences elicited via an online survey plug-in using an n-point scale weighting method, deployed in synchronous (real-time integration during workshops) and asynchronous (integration occurs at different times) settings; and (3) output variability analysis of performance values, weights and model parameters.

3. HELDA software

3.1. Overview

HELDA, the Helmholtz Decision Analysis Tool, is a desktop application developed in Java with a strictly modular system design as presented in Fig. 1. The central components are the MCDA specific data management and the user interface integrating this data and connecting to the different modules. It is notable that the user interface provides an immediate response system to active modules visible on screen based on a notification and update management, i.e. changes applied anywhere are instantly reflected in other modules e.g. changing charts, tables, or reports on the fly while modifying weights, thus enhancing the understanding of the modification's impact. HELDA's notification and update management enables flexible, dynamic interaction with the MCDA workflow unlike a sequential approach for operational use.

Fig. 1 shows the most important modules: visualization of data and results in various ways, interaction with the evaluation matrix, weighting and aggregation methods for preference modelling, and output variability analysis (sensitivity and uncertainty analysis). HELDA supports uncertainty analysis using probability distributions of input data, i.e. performance values and weights, and MCS. Additionally, HELDA functionality can be extended using plug-ins. This was used to implement the online surveys for stakeholder integration. Further information relevant for the mentioned modules is presented in the following sections of this chapter.

3.2. Development process

Fig. 2 summarizes the five-steps methodology employed for the development of HELDA. The initial step, baseline software assessment, consisted in the assessment of the MCDA ToolKIT against the SA criteria defined in Mesa Estrada et al. (2026). The second step, enhancement effort, consisted in the elaboration of user requirements and expert-based estimation of effort to enhance the baseline software. Software enhancement refers to the modifications that extend, modify, and delete the functionality of an existing software for a specific use (Banker, Davis, & Slaughter, 1998). Effort estimation is a common practice among software developers to assess time and costs involved in an enhancement project. The estimation often depends on software complexity, maintenance team experience, application size, application quality and the software functionality modified (Banker et al., 1998). The third step, implementation of enhancements, entailed the prioritization of requirements and the execution of enhancements. The fourth step, enhanced software assessment, the HELDA software was assessed based on the SA criteria. Finally, in the fifth step, the HELDA software underwent iterative testing and application. The outcomes of this step refined the user requirements, as indicated in Fig. 2. The following paragraphs describe the development process, focusing on steps one to four and their respective outcomes.

3.2.1. Baseline software assessment

The first step consisted in the assessment of the baseline software (MCDA Tool KIT) against 29 criteria derived from the literature review of Mesa Estrada et al. (2026) on MCDA-software user requirements and theoretical capabilities of MCDA methods for SA. Table 2 presents the 29 criteria across 9 domains, hereafter referred to as the assessment criteria. The assessment revealed that 13 out of 29 criteria (45%) of the criteria had strong fulfilment, 4 out of 29 criteria (14%) moderate fulfilment, and 12 out of 29 criteria (41%) weak fulfilment. The weakest points of the baseline software in handling SA-MCDA were in the following domains:

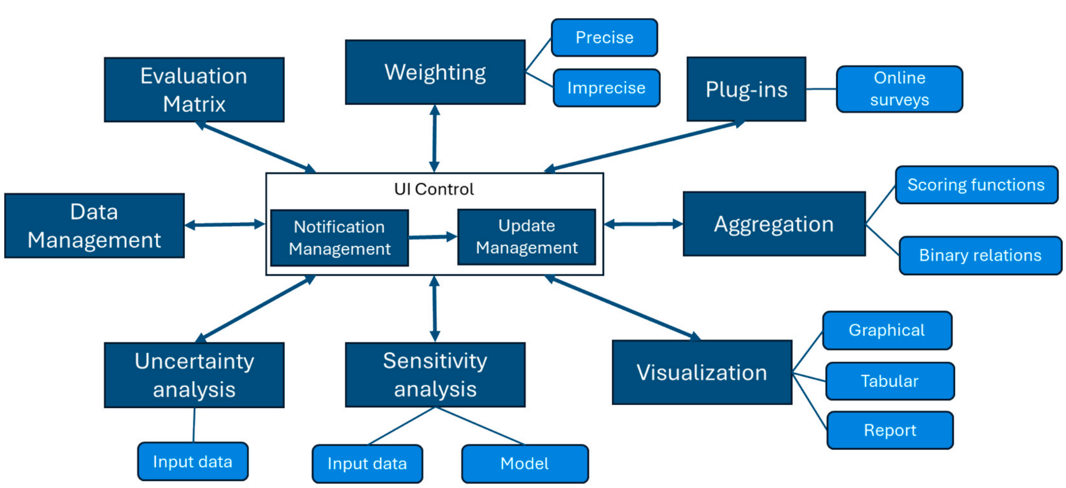

Fig. 1. Overview of the control flow of the HELDA user interface.

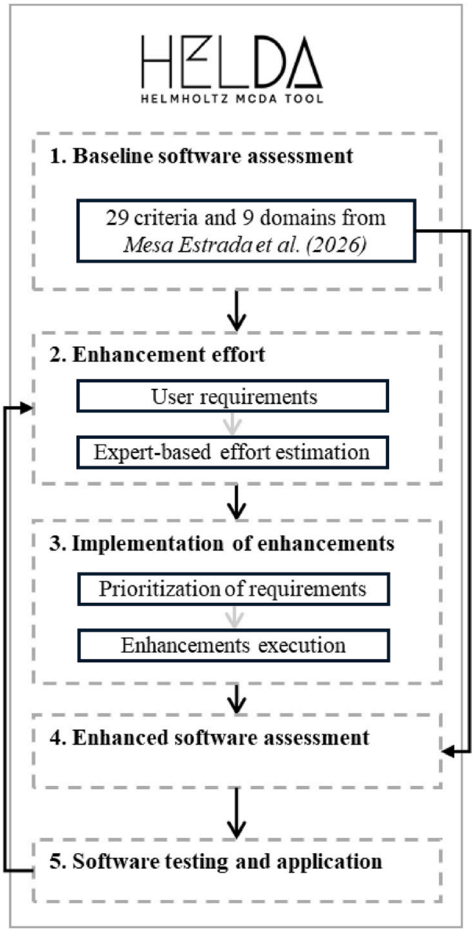

Fig. 2. HELDA development process.

- D4. Preference model: due to its lack of MCDA methods to model strong sustainability, i.e. aggregation methods with null or partial compensation of criteria and ability to handle qualitative and quantitative data (C4.1 Type of aggregation of multiple criteria evaluations, C4.3 Type of information (scale), C4.4 Compensation level between criteria, C4.5 Per-criterion pairwise comparison thresholds, C4.7 Interactions between criteria)
- D5. Stakeholders' involvement: due to its lack of functionality to handle stakeholders' preferences (C5.1 Problem structuring (groups), C5.2 Weights elicitation)
- D2. Problem typology: available MCDA methods only support ranking as decision recommendations (C2.1 Problem statement), i.e. sorting, clustering and choice are not supported.
- D3. Problem structure: lack of support to problem structuring (C3.1) with e.g. value-focused thinking (VFT) (Keeney, 1996), SWOT analysis (Livzeniece, Pubule, & Blumberga, 2021), and DPSIR (Moktadir & Ren, 2023).

3.2.2. Enhancement effort

The second step "Enhancement effort" consisted of two main parts. First, the elaboration of user requirements (see SI 1) for software extension based on the assessment criteria. Second, using these requirements, and the assessment of the baseline software from the previous stage, a technique for effort estimation based on expert judgement was conducted. This approach entailed a relative assessment of effort based on the story points (SP) methodology (Choudhari & Suman, 2012), using the Fibonacci sequence shown in Table 3. In the used sequence, the effort estimation is based on three factors: the amount of work (time to complete the task), the complexity of the task, and the

Table 2
29 criteria across 9 domains from Mesa Estrada et al. (2026) used for software assessment and to create user requirements.

	Criteria	Description
D1. Applicability and accessibility	C1.1 Ease of use	Degree of MADM knowledge required to use the software.
	C1.2 Software customizability	Type of permissions given to the user of the software, i.e. view, modify, and distribute.
	C1.3 Language inclusivity	Capability to support several languages
	C1.4 Personal information requirements	Type of personal data required to get access to the software.
	C1.5 Interoperability	Ability of the software to exchange information with e.g. external libraries, frameworks, or data sources.
D2. Problem typology	C2.1 Problem statement	Ability of the software to support several decision recommendations, among which ranking, sorting, clustering, choice
	C2.2 Criteria structure	Type of criteria structure supported by the software, i.e. flat and/or hierarchical.
	C2.3 Evaluation of alternatives on the criteria	Type of evaluation of alternatives on the criteria supported by the software, i.e. deterministic and/or uncertain
D3. Problem structure	C3.1 Problem structuring methods	Capability of the software to support and stimulate creative thinking in problem structuring with e.g. brainstorming, SWOT analysis, DPSIR modules
	C3.2 Dynamic alternatives management	Capability of the user interface to manage dynamic and static sets of alternatives.
	C3.3 Dynamic criteria management	Capability of the user interface to manage dynamic and static sets of criteria.
D4. Preference model	C4.1 Type of aggregation of multiple criteria evaluations	Capability of the software to deal with multiple types of aggregation, i.e. scoring function, binary relations, decision rules
	C4.2 Comparison of alternatives	Capability of the software to handle relative and absolute comparisons of alternatives performances.
	C4.3 Type of information (scale)	Capability of including information which is qualitative and quantitative in nature. Or the performances are not actually known from the performance table, and the DM make the relative assessment by making pairwise comparisons on a relative ratio scale.
	C4.4 Compensation level between criteria	Capability to model situations of full, partial and null compensation between criteria.
	C4.5 Per-criterion pairwise comparison thresholds	Capability to use per-criterion pairwise comparisons thresholds to model the decision problem, i.e. indifference, preference, veto thresholds.
	C4.6 Weights of criteria	Capability to support methods that include precise and imprecise weights of criteria.
	C4.7 Interactions between criteria	Capability to model interactions between criteria.

(continued on next page)

Table 2 (*continued*)

	Criteria	Description
D5. Stakeholders' involvement	C5.1 Problem structuring (groups)	Capability to supports collecting and processing input data from several participants for problem structuring (objectives, criteria, stakeholders) from e.g. online surveys
	C5.2 Weights elicitation	Capability to support collecting and processing preference information elicited from different stakeholders from e.g. online surveys
D6. Output variability analysis	C6.1 Sensitivity analysis: input data	Capability to carry out sensitivity analysis on weights and/or performance values
	C6.2 Sensitivity analysis: model	Capability to carry out sensitivity analysis on preference model, which include discriminating thresholds, type of aggregation and/or normalization.
	C6.3 Uncertainty: input data	Capability to carry out uncertainty analysis on weights and/or performance values (single preference model)
	C6.4 Uncertainty: model	Capability to carry out uncertainty analysis on multiple preference models
D7. Transparency	C7.1 Traceability of documentation	Availability of relevant and up-to-date documentation to support software users e.g. user manual, GitHub available
	C7.2 Transparency of documentation	Comprehensiveness of the documentation available including important sections for MADM: (1) problem formulation, (2) model (weights, aggregation functions, parameters, other features important for problem development, decision rules) (3) output variability analysis (on what it can be performed), and (4) visual representation (easy to understand what is visible)
D8. Utility	C8.1 Learning dimension	Capability to acknowledge and accept new information revealed during the evolution of the procedure
	C8.2 Interpretation of results	Capability to support interpretation of results with e.g. generate a report or summary with the main findings.
	C8.3 Graphical representation	Capability to represent results using figures and/or tables

uncertainty or risk to succeed in the task. A structured interview was conducted with one of the developers of the baseline software using a set of cards with the criteria names representing the user requirements (one card per criterion). The expert was asked to provide a relative comparison of the efforts to fulfil the requirements (strong fulfillment) by arranging the cards in order of effort and assigning SPs. For the interview with the software developer, the deck of cards was created using the Conceptboard web application (Conceptboard, 2025). Although this effort estimation technique may involve several experts, this study involved only one expert. The reason for this was that, of the former group of developers, only one person was actively maintaining the tool. Besides, the resulting effort estimation was used as a general guideline rather than a strict or absolute result.

Fig. 3 presents the results of the expert-based estimation of effort to enhance the baseline software in the direction of the user requirements.

Table 3
Fibonacci scale for assessment of efforts (Chauhan, Dakoliya, & Sharma, 2023).

Story point (SP)	Effort required	Time required	Complexity	Risk
0	No effort	A few minutes	None	None
1	Minimum effort	A few minutes	Little complexity	None
2	Minimum effort	A few hours	Little complexity	None
3	Mild effort	A day	Low complexity	Low
5	Moderate effort	A few days	Medium complexity	Moderate
8	Sever effort	A week	Medium complexity	Moderate
13	Maximum effort	A month	High complexity	High
21/34	Can not be completed in expected time frame	More than a month	–	–

7 (24.1%) of the implementations to fulfil the user requirements were assigned to no effort (0 SP). 18 (62.1%) were assigned low to moderate effort (1–5 SPs) while only 4 (13.8%) were assigned the highest levels of effort (8-21 SPs) (see Table 3). The following paragraphs describe the performance and estimation of efforts to enhance the baseline software.

Strong performance of baseline software. The enhancement effort estimated by the developer with respect to criteria with strong performance of MCDA ToolKIT ranges from *No effort (0)* to *Moderate effort (5)*:

- *No effort (0)* was estimated for criteria representing functionality that would automatically work with, for example, new methods added (C2.2 Criteria structure, C2.3 Evaluation of alternatives on the criteria, C4.6 Weights of criteria, C6.1 Sensitivity analysis: input data, C6.3 Uncertainty: input data).
- *Minimum effort (1)* was estimated for C1.1 Ease of use, C1.3 Language inclusivity and C8.3 Graphical representation, since the addition of new MCDA methods requires to make some adaptations to the existing functionality e.g. different graphical representation (C8.3) or inserting the related contextual help frame (C1.1).
- *Moderate effort (5)* was estimated for enhancing transparency of documentation (C7.2), and adapting interoperability, e.g. importing and exporting options for new MCDA methods (C1.5). The developer argues that despite being simple and straight forward tasks, the time required for this should not be underestimated.

Moderate performance of baseline software. For the enhancement of criteria with moderate performance of the baseline software, *Minimum effort (1)* to *Moderate effort (5)* was estimated:

- *Minimum effort (1)* was assigned to enhance the software customizability (C1.2), since the baseline software was already in GitLab and only the settings and license would need to be changed to provide open access.
- *Minimum effort (2)* was assigned to enhance dynamic alternatives and criteria management (C3.2 and C3.3), since it would only take few hours considering the implementation of a simple solution with a user interface where this can be managed.
- *Moderate effort (5)* was assigned to ensuring the documentation was up to date and incorporated the enhancements (C7.1). Similarly to C7.2 and C1.5 in the previous paragraph, the time required to implement this is not insignificant.

Weak performance of baseline software. The effort assigned to enhance criteria with weak performance ranges from *Mild effort (3)* to *Maximum effort (21)*:

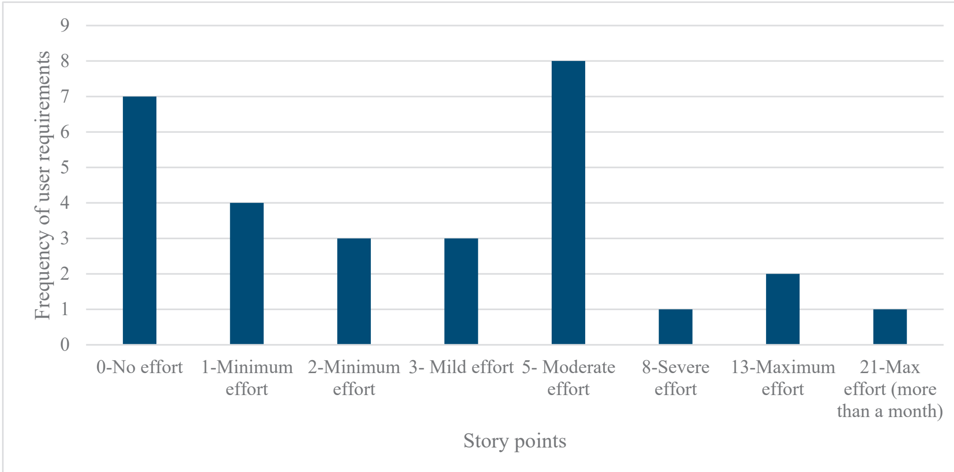

Fig. 3. Expert-based estimation of effort to enhance the baseline software.

- *Mild effort (3)* was assigned to requirements that could benefit from the existing functionality of the baseline software. For example, for C6.4 Uncertainty: model and C6.2 Sensitivity analysis: model, for the scoring methods, the software already has different normalization functions and aggregation methods that could be used in the implementation of this requirement. For the outranking methods, the implementation could use the existing interface for sensitivity analysis of input data and adapt it for thresholds.
- *Moderate effort (5)* was assigned to the requirements that explicitly involved adding MCDA methods to the software (C2.1 Problem statement, C4.1 Type of aggregation, C4.3 Type of information, C4.4 Compensation level between criteria, and C4.5 Per-criterion pairwise comparison thresholds). The baseline software had scoring functions methods, the addition of a different type of methods such as binary relations involve a different management.
- *Sever effort (8)* to *Maximum effort (21),* was assigned to criteria with high uncertainty in developing a solution because it was not clear how this could work in the software. The developer recognizes that the solutions could be simple or as well very complex, but definitely a completely new functionality is required to fulfil those requirements. The criterion C5.1 Problem structuring (groups) was assigned *Maximum effort (21)*, the developer emphasized on the need to add a completely new functionality to the software to fulfil this requirement, and acknowledging the high uncertainty on what could be a user-friendly implementation. Consequently, the second highest effort, *Maximum effort (13),* was assigned to C3.1 Problem structuring methods. Similarly, a new functionality is required to fulfil this requirement; however, relatively lower effort than C5.1 since it is not for groups. *Maximum effort (13)* was also assigned to C5.2 Weights elicitation, because it requires a new functionality to elicit weights from a group of stakeholders, but has relatively lower uncertainty as the baseline software already provides a module to handle weights elicited as histograms. The criterion C8.1 Learning dimension was assigned *Severe effort (8),* the developer argues this enhancement implies lower uncertainty but the amount of work to have a "comfortable" solution is high.

3.2.3. Implementation of enhancements

In the third step, "Implementation of enhancements", requirements were prioritized for execution based on research priorities and budget constraints using a qualitative scale (low, moderate, high). This research aimed to enhance an existing software to support SA-MCDA integrating three core goals: (1) strong sustainability concept with binary relations methods (2) stakeholder integration and (3) output variability analysis of weights and model parameters. All enhancements considered three sub-goals: improving usability, broadening analytical capabilities, and providing transparent documentation. The software targets sustainability practitioners and researchers from diverse fields (e.g., Life Cycle Assessment, energy systems) who are not necessarily MCDA experts. Therefore, development priorities were established based on the assessment of the baseline software (Section 3.2.1) and the enhancement effort estimation (Section 3.2.2). High priority was given to requirements in two cases. First, addressing capabilities in which the baseline software performed "Weak" and were considered insufficient to conduct SA. Second, enhancements requiring low to moderate effort (0–5 story points) in already strong or moderate criteria to maximize system performance. Low priority was assigned to requirements deemed less urgent, either because the baseline software already demonstrated strong or moderate fulfilment, or due to budget constraints. The following list presents selected requirements and enhancements for the different domains (D) as defined in Table 2 (in order of importance for SA). For a complete list of executed enhancements refer to Table 4 and SI 1.

- D4. Preference model: High priority was assigned to adding binary relations methods, i.e. PROMETHEE II and I and ELECTRE III, to model the concept of strong sustainability (C4.1 Type of aggregation of multiple criteria evaluations, C4.3 Type of information, C4.4 Compensation level of criteria, C4.5 Per-criterion pairwise comparison thresholds, C4.6 Weights of criteria). C4.2 Comparison of alternatives was assigned low priority; however, the capabilities of the software for relative assessments were strengthened by the addition of the mentioned methods. C4.7 Interactions between criteria was assigned low priority considering its relevance for the scope of the decision problems typically addressed by the authors of this study.
- D5. Stakeholders' involvement: High priority was assigned to adding a plug-in for online surveys to elicit and process preferences from stakeholders regarding criteria reflection (C5.1 Problem structuring - groups) and criteria weights (C5.2 Weights elicitation). This was complemented with the addition of the SRF method (Figueira and Roy, 2002) for criteria weighting.
- D6. Output variability analysis: High priority was assigned to complementing features for sensitivity analysis (C6.2) and uncertainty analysis (C6.3), including thresholds and stakeholder weights. C6.4 Uncertainty: model was given low priority considering its relevance for the scope of the decision problems addressed by the authors of this study.
- D1. Applicability and accessibility: although the baseline software was assessed as strong in regard to ease of use (C1.1) and interoperability (C1.5), some enhancements were executed. These included

Table 4
HELDA development process: Summary of Baseline software assessment, Enhancement effort and Implementation of enhancements (S – Strong performance, M – Medium performance, W – Weak performance).

Domains (D)	Criteria (C)	Baseline software assessment	Estimation of enhancement effort (Fibonacci sequence)	Prioritization	Enhancements implemented (yes✓/no x)	HELDA assessment
D1. Applicability and accessibility	C1.1 Ease of use	S	1	High	✓	S
	C1.2 Software customizability	M	1	Low	×	M
	C1.3 Language inclusivity	S	1	Low	×	S
	C1.4 Personal information requirements	S	0	High	×	M
	C1.5 Interoperability	S	5	High	✓	S
D2. Problem typology	C2.1 Problem statement	W	5	Low	×	W
	C2.2 Criteria structure	S	0	Low	×	S
	C2.3 Evaluation of alternatives on the criteria	S	0	Low	×	S
D3. Problem structure	C3.1 Problem structuring methods	W	13	Low	×	W
	C3.2 Dynamic alternatives management	M	2	High	✓	S
	C3.3 Dynamic criteria management	M	2	Low	×	M
D4. Preference model	C4.1 Type of aggregation of multiple criteria evaluations	W	5	High	✓	M
	C4.2 Comparison of alternatives	S	0	Low	×	S
	C4.3 Type of information (scale)	W	5	High	✓	S
	C4.4 Compensation level between criteria	M	5	High	✓	S
	C4.5 Per-criterion pairwise comparison thresholds	W	5	High	✓	S
	C4.6 Weights of criteria	S	0	High	✓	S
	C4.7 Interactions between criteria	W	3	Low	×	W
D5. Stakeholders Involvement	C5.1 Problem structuring (groups)	W	21	High	✓	S
	C5.2 Weights elicitation	W	13	High	✓	S
D6. Output variability analysis	C6.1 Sensitivity analysis: input data	S	0	Low	×	S
	C6.2 Sensitivity analysis: model	W	3	High	✓	S
	C6.3 Uncertainty: input data	S	0	High	✓	S
	C6.4 Uncertainty: model	W	3	Low	×	W
D7. Transparency	C7.1 Traceability of documentation	W	5	High	✓	S
	C7.2 Transparency of documentation	S	5	High	✓	S
D8. Utility	C8.1 Learning dimension	W	8	High	×	W
	C8.2 Interpretation of results	S	2	Low	×	S
	C8.3 Graphical representation	S	1	High	✓	S

adding a module for supporting aggregation method selection, thereby facilitating the ease of use (C1.1) for non-MCDA experts. Additionally, enhancing and adding export and import capabilities of data and figures to improve interoperability (C1.5), and making the software accessible in a dedicated website.

- D8. Utility: High priority was assigned to complementing visualization options for the added MCDA methods and the group preferences elicited with the plug-in developed enhancing C8.3 Graphical representation. C8.1 Learning dimension was given high priority but enhancements have not been executed yet due to less urgency and budget constraints.
- D7. Transparency: High priority was assigned to complementing and updating documentation (C7.1 Traceability of documentation and C7.2 Transparency of documentation) and making it accessible in a dedicated website.
- D3. Problem structure: High priority was assigned to adding a module to analyze the impact of specific alternatives in the decision model enhancing dynamic alternatives management (C3.2). Low priority was assigned to integrating problem structuring methods (C3.1), recognizing the great efforts done by the developers of the software Entscheidungsnavi (von Nitzsch, Tönsfeuerborn, & Siebert, 2020) in this direction, and bearing in mind potential future collaborations.
- D2. Problem typology was the only dimension in which all requirements were given a low priority and enhancements were not executed. Despite the weak performance of the baseline software in C2.1, ranking as a unique decision recommendation was sufficient for the scope of the decision problems addressed by the authors of this study.

3.2.4. Enhanced software assessment

Table 4 summarizes the development process that resulted in the HELDA software. The assessment of the HELDA software showed that 20 out of 29 criteria (69%) of the criteria have strong fulfilment, 4 out of 29 criteria (14%) have moderate fulfilment, and 5 out of 29 criteria (17%) have weak fulfilment. HELDA's weakest points in handling SA are its lack of functionality for handling different problem statements (C2.1), problem structuring methods (C3.1), and learning dimension (C8.1). It should be noted that HELDA remains in development by the time this document was prepared.

3.3. Software main features

3.3.1. Applicability and accessibility (D1.)

HELDA software aims to meet the requirements for SA-MCDA, while still being applicable to decision making problems in general. HELDA has a Creative Commons (CC) license (BY-NC-ND 4.0)[6] and is free to use for non-commercial usage. A dedicated website[7] provides access to the software and relevant information for the users, including: the user manual, latest versions released, and video tutorials. Regarding the user interface, the objective was to upgrade the image of the software, to make it look more modern and appealing to the users with simple changes. This implied changes to colors, fonts, and icons.

The conceptualization of HELDAs aims to serve both MCDA experts and users with no expertise or deep knowledge on MCDA, such as LCA experts. Therefore, a new functionality was implemented to enhance ease of use (C1.1), with the selection of a proper MCDA method using a series of questions or criteria. This implementation was based on the software MCDA-MSS[8] developed by Cinelli et al. (2022b). Fig. 4 shows the method selection dialogue implemented in HELDA, using an example of the recommendations that the software provides to users after answering a selection of questions. All questions offer question-specific answers and the answer "I don't know" which is a likely answer for novice users. The objective of this functionality is not only to provide a recommendation to users, but to promote reflection about relevant considerations when modeling decision problems with MCDA methods. To enhance interoperability (C1.5), export options for images (.svg) and data (.csv, .xls) were added. With respect to personal information requirements (C1.4), providing an email address is required to get access to the HELDA download link.

3.3.2. Preference modeling (D4.)

HELDA allows preference modeling using scoring functions and binary relations methods. Table 5 presents the software capabilities for SA in terms of type of aggregation (C4.1), comparison of alternatives (C4.2), type of information (scale) (C4.3), compensation level between criteria (C4.4), per-criterion pairwise comparison thresholds (C4.5), and weights of criteria (C4.6). The addition of binary relations methods strengthened the capabilities of the software for preference modeling. ELECTRE III and PROMETHEE I and II libraries were used and adapted to integrate these methods into the existing structures of the baseline software. For PROMETHEE I and II, the code from Krasucki (2012) was adapted. ELECTRE III - Net flow Scores was taken from the Python implementation in Sarmas, Xidonas, and Doukas (2020) and translated to Java. Similarly, the Python code for ELECTRE III-distillation in pyDecision library by Pereira, Basilio, and Santos (2025) was translated to Java.

The weighting methods available in HELDA are SWING, SMART, the SRF method (Figueira & Roy, 2002) and sliding bars. The SRF method was added in the enhancement process because of its suitability in contexts with multiple stakeholders, and for ELECTRE III and PROMETHEE methods. In the SRF method, participants are given a set of *n* cards with the names of the *n* decision criteria and a set of white cards. They are asked to order the criteria cards according to their relative importance, from the most important to the least important. If the participants consider that criteria have same weight, they can put them next to each other. Participants are asked to assign blank cards between criteria to represent the difference in the importance of successive criteria. Finally, participants should select a number *z,* which denotes the ratio of importance between the most and least important criteria. The calculation of the weights is then performed in two steps: first, determination of non-normalized weights, *k(r),* and then the normalized weights, *ki.* The baseline software had the functionality to input imprecise weights using distribution functions; however, it was never used, nor did it provide a systematic procedure to elicit this type of information. This issue is addressed in the present study, with a solution provided in the output variability analysis (see Section 3.3.4).

3.3.3. Stakeholders' integration (D.5)

A plug-in for online surveys was created to facilitate the interaction between stakeholders and decision models in HELDA. The objective was to enable stakeholders or survey participants to easily express their preferences regarding the decision criteria, using a low threshold approach. This can be done without downloading a specific software, but simply by opening a link or scanning a QR code with a mobile device, and afterwards working in a web browser. The information is organized by stakeholder categories, meaning participants should select a category that best represents their perspective or role in the decision-making process. The stakeholder categories are defined by the analyst when setting up the survey in the plug-in. The platform LimeSurvey was used to collect the preferences, not only because of its robust capabilities for handling online surveys, but also because it is an open-source software that allows to host and store the data in one's own server location (LimeSurveyGmbH, 2025). LimeSurvey was connected to HELDA to enable two-way communication for processing preference information elicited. Firstly, the plug-in can automatically create surveys using a (pre-)defined list of stakeholders' categories and the criteria available in the evaluation matrix in HELDA. Once the survey is created, the plug-in displays a QR code and a link that stakeholders can scan to access the survey. The stakeholders or survey participants submit their responses to the online survey platform, and these are in turn sent to HELDA in real-time which means that elicited preferences are directly visible in HELDA. The plug-in offers the possibility to export and import online surveys, so the data can be processed or verified at any time, even if the software is closed. Two types of preference information are collected using the online survey plug-in. Firstly, stakeholder categories and the acceptance and reflection of the decision criteria used in the model are collected. This information serves, for example, for problem structuring in groups (C5.1). Secondly, stakeholder categories and the relative importance of the decision criteria, which serves for weights elicitation (C5.2). In the following paragraphs, both are explained in more detailed.

Problem structuring. The first part of the plug-in provides a solution for eliciting stakeholder preferences regarding the decision criteria. The analyst view in HELDA, where this online survey is created (Set-up tab) is available in SI 2 - Figure 2-1. The *Criteria tab* displays the QR code, as well as the input received from the stakeholders that submitted their responses (SI 2 - Figure 2-2). Fig. 5 shows the two parts of the online survey that the stakeholder needs to complete. First, the selection of the stakeholder category (a). Second, reflecting on and adding missing criteria (b). The objective is to encourage reflection on the decision criteria and their meanings, and to evaluate how the set of criteria is perceived by stakeholders.

Weights elicitation. The second part of the plug-in is a solution for eliciting stakeholder criteria weights. The interface allows defining the settings of the survey, e.g. the range of scale and updating the decision model in HELDA (see SI 2 -Figure 2-3). The plug-in uses an n-point scale to assign relative weights of the criteria, where n can take values up to

[6] https://creativecommons.org/licenses/by-nc-nd/4.0/

[7] https://www.mcda-helmholtz.de/64.php

[8] https://mcda.cs.put.poznan.pl/

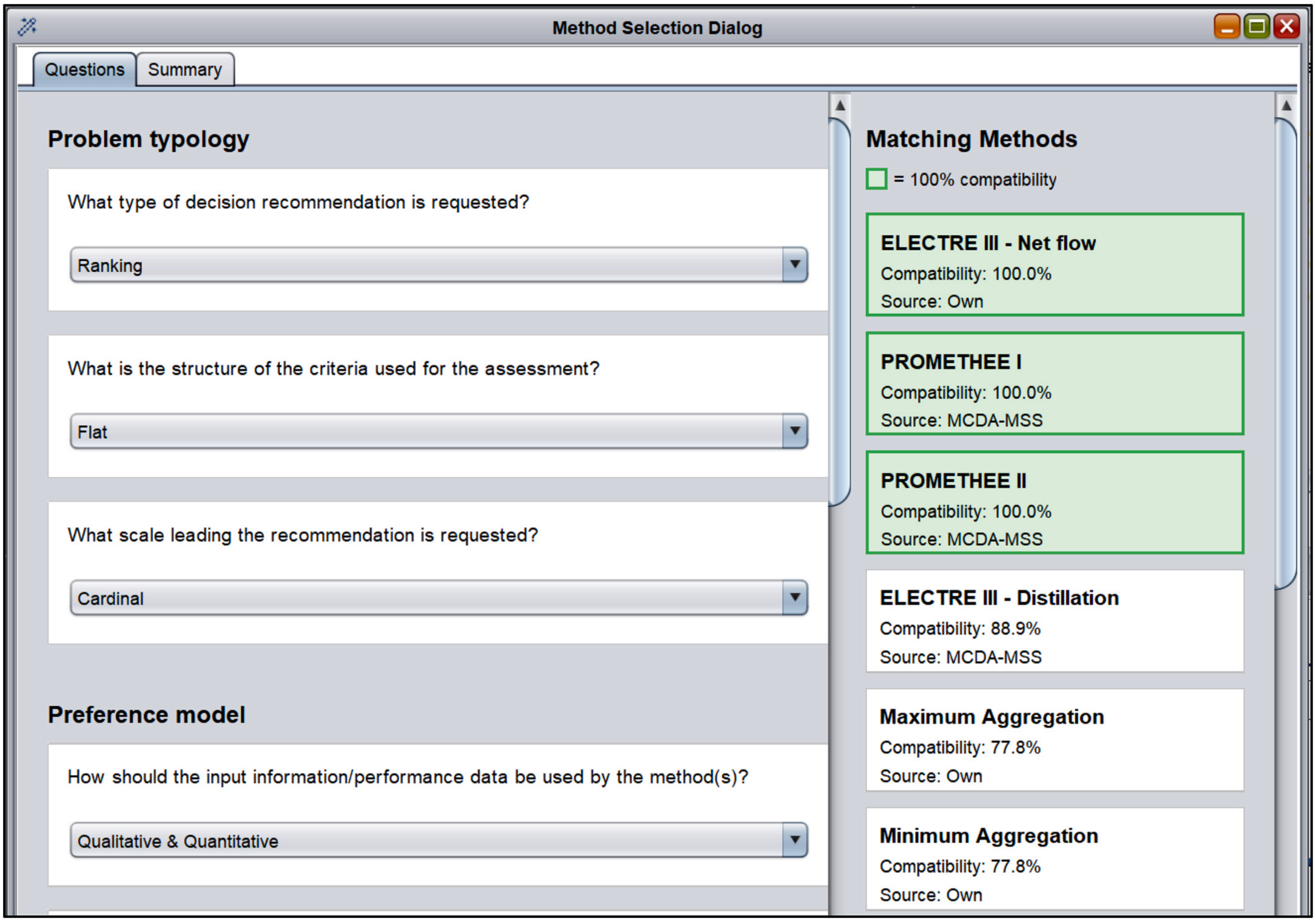

Fig. 4. Method selection dialogue in HELDA.

Table 5
Aggregation methods available in HELDA. Based on (Cinelli et al. 2022).

Method	Type of aggregation	Comparison of alternatives	Type of information (scale)	Compensation level between criteria	Pairwise comparison thresholds	Types of criteria weights
Weighted sum	Scoring function	Absolute, relative	Quantitative	Full	None	Precise-TO
Weighted product	Scoring function	Absolute, relative	Quantitative	Partial	None	Precise-TO
Weighted rank	Scoring function	Absolute, relative	Quantitative	Partial	None	Precise-TO
Minimum aggregation*	Scoring function	Absolute, relative	Quantitative	Partial	None	Precise-TO
Maximum aggregation*	Scoring function	Absolute, relative	Quantitative	Partial	None	Precise-TO
TOPSIS	Scoring function	Absolute, relative	Quantitative	Full	None	Precise-TO
VIKOR	Scoring function	Absolute, relative	Quantitative	Full	None	Precise-TO
PROMETHEE I	Binary relations	Relative	Quantitative; qualitative	Null, partial	p, q	Precise-RIC
PROMETHEE II	Binary relations	Relative	Quantitative; qualitative	Null, partial	p, q	Precise-RIC
ELECTRE III - Net flow**	Binary relations	Relative	Quantitative; qualitative	Null, partial	p, q, v	Precise-RIC
ELECTRE III - Distillation	Binary relations	Relative	Quantitative; qualitative	Null, Partial	p, q, v	Precise-RIC

P: preference, q: indifference, v: veto. RIC: relative importance coefficients, T:tradeoffs. *(Langhans, Reichert, & Schuwirth, 2014), **(Bouyssou, 1992).

10. The selection of this method was done based on: the use of a proper weighting method for binary relations methods, the reduction of development effort, and the provision of an easy-to-use, clear interface for stakeholders to submit their responses. The submitted relative weights are collected and processed by the plug-in. This results in the creation of weighting sets of the stakeholder categories participating in the online surveys, which can be used as precise and/or imprecise inputs in the decision model in HELDA. The precise weights are calculated using the average of the inputs of each criterion per stakeholder category, and then normalized to the interval (0,1). They are used for comparative analysis of the preferences among stakeholders' categories, using e.g. cobweb charts and weight sets analysis (bubble charts), and their respective rankings. The uncertain sets of weights refer to the histograms created with the preferences of the survey participants, and

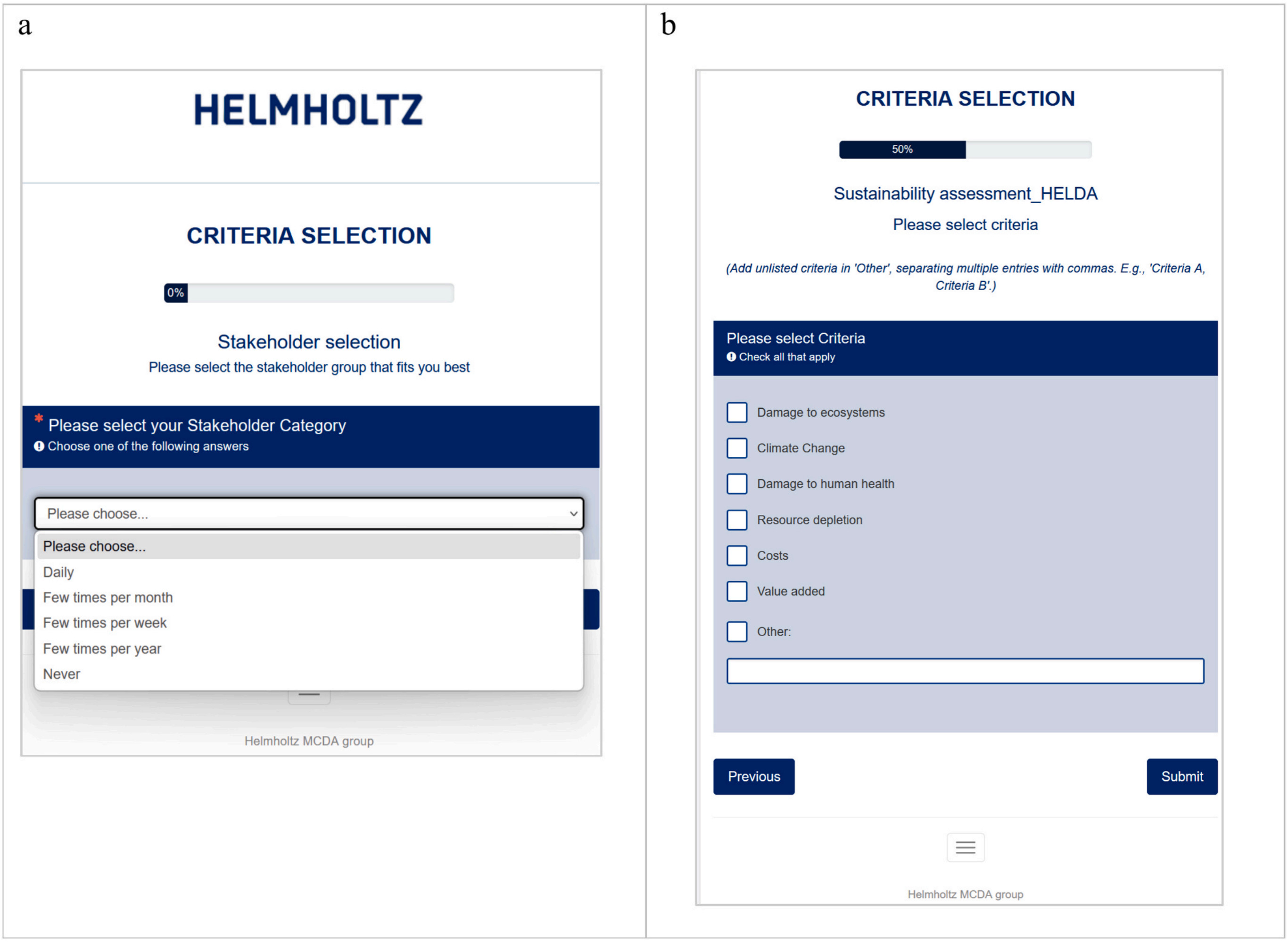

Fig. 5. Online survey for criteria selection (stakeholder view) with selection of stakeholder category (a) and reflection on criteria (b).

will be covered in the next section on output variability analysis. Fig. 6 shows the two parts of the online survey for weights elicitation that the stakeholder needs to complete. First, the selection of the stakeholder category (a). Second, the weighting of criteria, i.e. assigning relative importance (b).

3.3.4. Output variability analysis (D.6)

HELDA allows for sensitivity analysis of the input data (i.e. performance values and weights), as well as to define uncertain/imprecise input data as probability distributions, and the execution of MCS. As previously stated, the software's architecture ensures that the modules dedicated to sensitivity and uncertainty analysis are automatically linked to the added MCDA methods. The only exception was the ELECTRE III – distillation exploitation, since it has an ordinal type of ranking results. The following paragraphs describe relevant enhancements implemented in the respective modules.

Sensitivity analysis. With the addition of the binary relations methods to HELDA, the preference model includes additional information in the form of pairwise comparison thresholds (Roy, Figueira, & Almeida-Dias, 2014): indifference (q), preference (p), and veto (v) thresholds. Therefore, the sensitivity analysis module was extended (C6.2) to allow the analysis of variations of q and p thresholds. This enhancement resulted in HELDA being capable of conducting sensitivity analysis of performance values, weights, and thresholds (if applicable). The HELDA software generates graphs for sensitivity analysis of weights, performance values and discriminating thresholds as a function of the ranking position and/or the alternative score (depending on the aggregation method).

Uncertainty analysis. The baseline software allowed to define uncertain input data, i.e., the performance values and weights, using probability functions (Müller et al., 2020). However, a systematic procedure for collecting these types of weights was not implemented. This issue was addressed in HELDA for groups of stakeholders (C6.3), by building histograms with the responses collected with the online survey's plug-in, as outlined in Section 3.3.3. Such histograms have an equidistant distribution as a default setting, meaning that each interval has an equal width (Fig. 7). When feasible, the distribution can be determined using the graphical representation. MCS is used to generate an ensemble of deterministic MCDAs. The members of the ensemble are evaluated individually, and the results are then combined into a single result which is inherently uncertain. Two approaches are available to obtain the single result in HELDA: firstly, aggregating the ranking values, and secondly, aggregating the actual ranking, i.e., determining the rank winners (Müller et al., 2020).

3.3.5. Utility (D.8)

HELDA allows for tabular and graphical representation of results, including the newly implemented binary relation methods. However, these features were enhanced to represent specific results from the binary relation methods and the plug-in for online surveys. In the case of the binary relations methods, these include, for example, adding an outranking matrix table to display preference relations between alternatives: preference, indifference, incomparability. Furthermore, tables displaying flows for PROMETHEE I and II, and concordance,

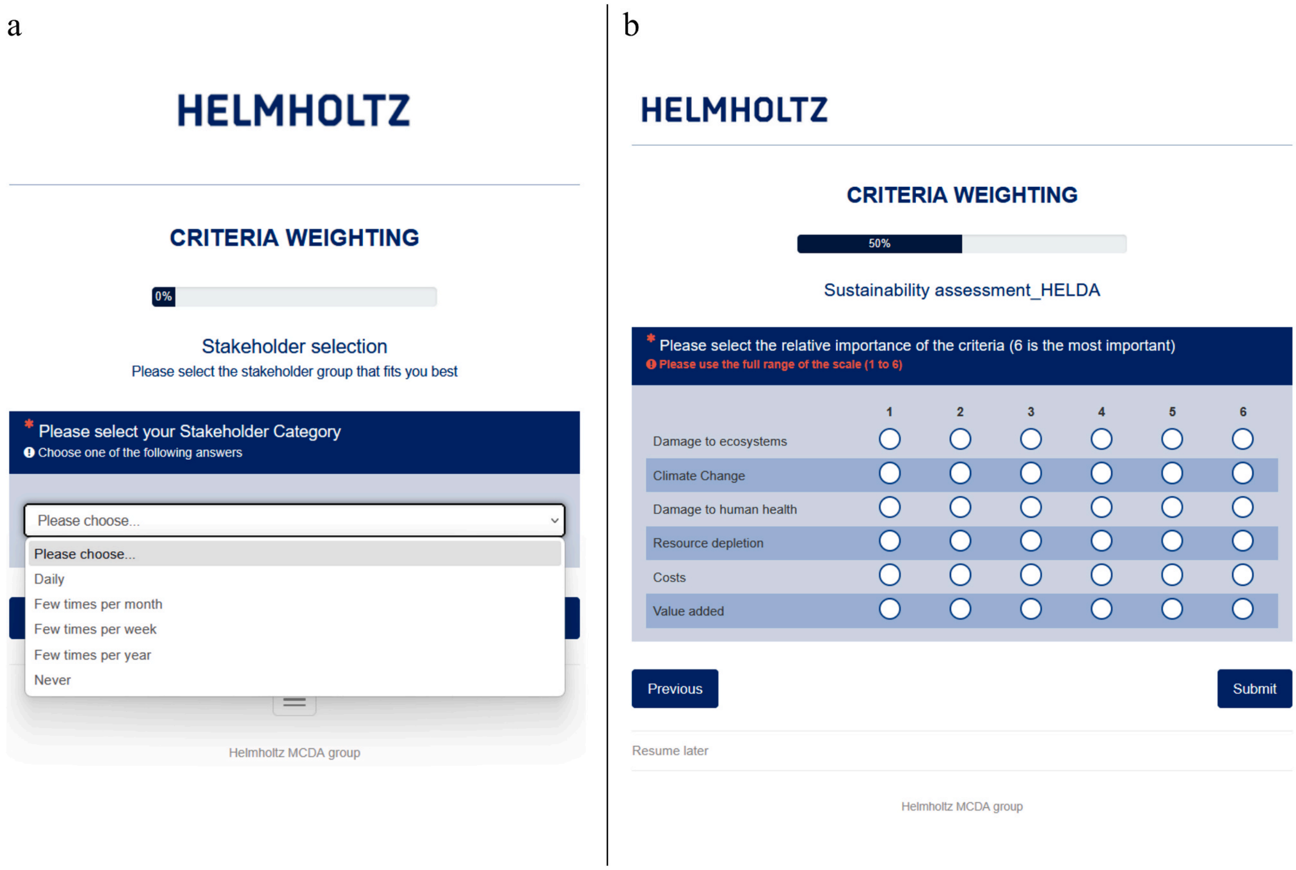

Fig. 6. Online survey for weights elicitation (stakeholder view) with selection of stakeholder category (a) and weighting of criteria (b)

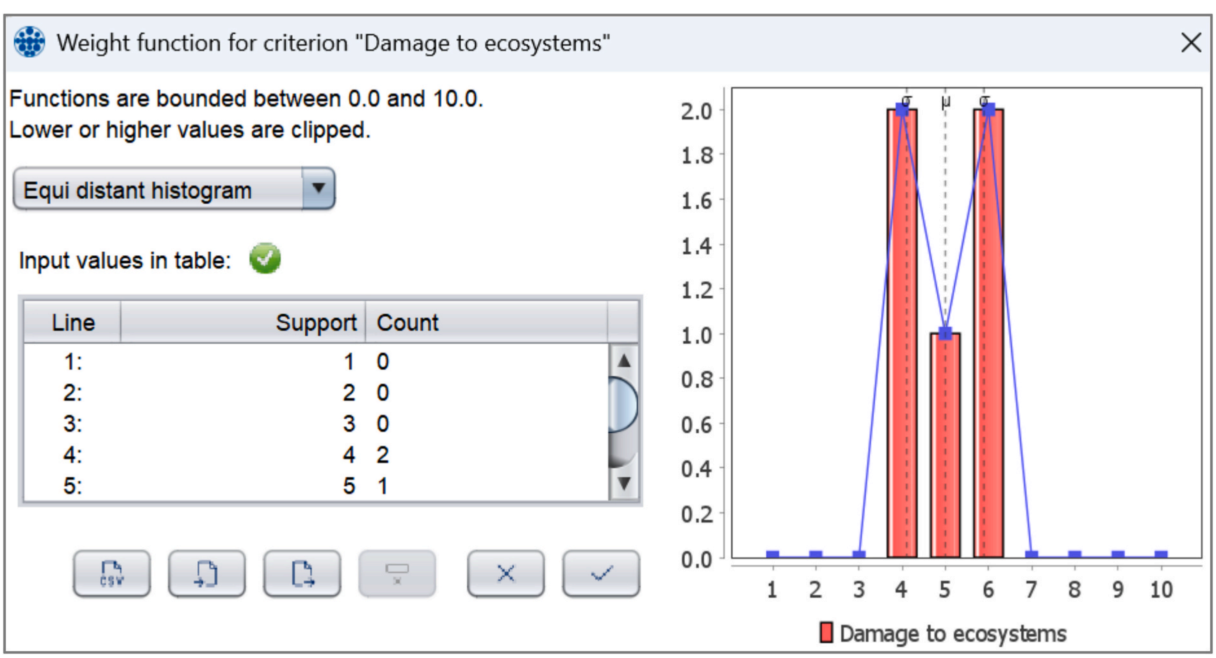

Fig. 7. Example of a histogram created in HELDA with the responses of an online survey.

discordance, and credibility matrix for ELECTRE III were added. In terms of graphs, a directed graph window was added to display the rankings and related preference relations between the alternatives (preference, indifference, and incomparability). In HELDA software, connections between aggregation methods and the relevant tables and graphs were implemented, meaning that when a specific aggregation method is selected, irrelevant graphs or tables are disabled. The extensions to present the results of the online survey plug-in include a table to manage stakeholder weight sets, and different types of graphs to compare stakeholder categories/groups simultaneously, including cobweb and bubble charts. A selection of these graphs is presented in Section 4.

4. Software illustration: SA of individual motorized mobility options

In this section, the application of HELDA is shown in a use case for prospective SA of individual motorized mobility options, i.e. five different alternatives concerning propulsion technologies and fuels, using MCDA and including stakeholder preferences for criteria weighting. The use case outlined in this section is exclusively intended for illustrative purposes, to showcase the functionalities of the HELDA software. The elicited preferences and displayed results are not designed for substantive analysis or the derivation of actionable insights.

4.1. Use case description

This use case involves five alternatives and six decision criteria. The selection of alternatives and criteria, as well as related performance values are based on Haase et al. (2022). The following five different alternatives were considered: (1) Internal combustion engine vehicle with low sulphur gasoline (ICEV-fossil), (2) Internal combustion engine vehicle with synthetic biofuel from straw (ICEV-straw), (3) Battery electric vehicle with electricity from wind power (BEV-wind), (4) Battery electric vehicle with electricity from production mix Germany (BEV-mix_DE) and (5) Fuel cell electric vehicle with hydrogen from wind power (FCEV-wind). For technical specifications, please refer to Haase et al. (2022). In order to achieve low complexity for stakeholders' involvement, from the original 15 criteria, five criteria have been aggregated to the criterion damage to ecosystems (Acidification, Ecotoxicity-freshwater, Eutrophication-freshwater, Eutrophication-marine, Eutrophication-terrestrial) and six criteria have been aggregated to the criterion human health (Human toxicity-carcinogenic, Human toxicity-non-carcinogenic, Ionizing radiation-human health, Ozone depletion, Particulate matter/Respiratory inorganics, Photochemical ozone formation (POF)), resulting in six sustainability criteria considered in this use case: Damage to ecosystems, Climate Change, Damage to human health, Resource depletion, Costs, and Valued added. For Damage to ecosystems and Damage to human health, performance values have been estimated on a qualitative scale from 1–5 for this study based on the original LCA-based estimations of underlying sub-criteria. Table 6 displays the evaluation matrix with criteria, preference directions, scales, units, alternatives, and respective performance values.

4.2. Preference modeling

For this use case, we aimed for cardinal ranking of alternatives based on a flat structure of criteria. We had qualitative and quantitative criteria and aimed to elicit stakeholder preferences for criteria weights. In order to account for the concept of strong sustainability, we aimed for no compensation of criteria, i.e. good performance in one criterion should not compensate for bad performance in another criterion. Furthermore, we wanted to use pairwise comparison thresholds to model imperfect knowledge or uncertainty of the data (Roy et al., 2014). Using the HELDA method selection dialogue (see Fig. 4), we selected ELECTRE III net flow aggregation method from the three suitable aggregation methods. For quantitative criteria, indifference threshold q was defined as 5% of minimum criterion value and preference threshold p as 10 % of minimum criterion value. For qualitative criteria, q was defined as 0 and p as 1. For the current use case, no veto was defined. Resulting discriminating thresholds are listed in Table 7.

4.3. Stakeholders' integration

A two-hour workshop was conducted with the objective of eliciting the preferences of 13 researchers with different academic backgrounds, including sociology, engineering, political science, and meteorology. The HELDA software was used to collect individual and group preferences, for the selection and weighting of criteria, to visualize and compare weights sets (from the online survey plug-in as well as from the SRF method) and resulting rankings, and to conduct output variability analysis. The workshop exercise started with a brief introduction to MCDA methods and the HELDA software, and a description of the use case (20 min). Relevant information about the decision-making problem, including a description of the available alternatives and the respective criteria, was sent in advance to the participants. Fig. 8 presents the three-stage preference elicitation process used to demonstrate HELDA capabilities for stakeholders integration, including the type of information elicited and the time used for it. The objective of these stages was two-fold. First, to exemplify how HELDA can support preference elicitation in individual and group settings. And second, to check the impact of group deliberation in individual preferences. The following paragraphs provide more details on these stages.

The first stage, Individual Preference (part 1), consisted of the elicitation of preferences for the selection and weighting of criteria using the HELDA's online survey plug-in, as described in Section 3.3.3. Five stakeholder categories were proposed according to the frequency of car usage: daily, few times per week, few times per month, few times per year, never (see Fig. 5a). In total, eight participants chose "few times per month" and five participants chose "few times per year". Afterwards, participants were asked to select the criteria that they considered relevant for the SA of mobility options from the pre-defined six criteria (Section 4.1), by clicking on the respective field next to the criterion and to suggest criteria that from their point of view were missing in a text box. In the next step of the online survey, participants were asked to assign criteria weights, i.e. relative importance to the pre-defined six criteria by clicking the respective number on a point scale from 1 (least

Table 7
Discriminating thresholds (ELECTRE III).

Criterion	q	p	veto
Damage to ecosystems	0	1	–
Climate change	2,84 E-03	5,68 E-03	–
Damage to human health	0	1	–
Resource depletion	9,96 E-07	1,99 E-06	–
Costs	1,19	2,38	–
Value added	3,094	6,188	–

Table 6
Evaluation matrix.

Criterion	Preference direction (min/max)	Scale	Unit	Alternatives				
				ICEV-straw	BEV-Mix_DE	FCEV-wind	ICEV-fossil	BEV-wind
Damage to ecosystems	Min	Qualitative	–	5	4	3	2	1
Climate change	Min	Quantitative	kg CO_2-eq./km	0.06	0.08	0.08	0.22	0.06
Damage to human health	Min	Qualitative	-	3	4	2	5	1
Resource depletion	Min	Quantitative	kg Sb eq./km	5.1 E-05	2.3 E-05	2.0E-05	4.3E-05	2.2E-05
Costs	Min	Quantitative	Cent/km	30	25,74	23,86	28,6	24,82
Value added	Max	Quantitative	%	69.2	61.88	62.9	67.7	69.1

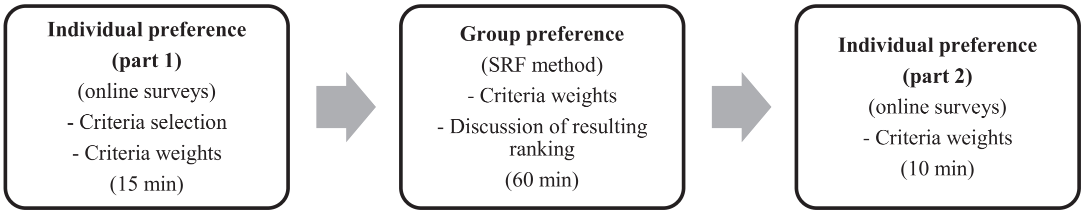

Fig. 8. Three-stages preference elicitation process conducted during the workshop.

important) to 6 (most important) (see Fig. 6b). Participants were informed about the procedure and the meaning of the weights, e.g. that they need to use the whole scale and that the importance factor is relative i.e. if two criteria are defined with importance factors 1.0 and 2.0 respectively, the second criterion is twice as important as the first one, same number means same importance.

The second stage, Group preferences, consisted of forming two groups of 6 and 7 researchers which used the SRF method as described in Section 3.3.2. The results were then entered for each group to the HELDA software in order to obtain the respective relative and absolute weights. Besides eliciting preferences, the groups discussed resulting ranking of alternatives using various graphical representations in HELDA, e.g. sliders for relative weights and sensitivity analysis of weights. Participants were given the opportunity to reflect on the MCDA model and change the weights.

In the third stage, Individual Preference (part 2), another online survey was conducted to elicit criteria weights individually. The implementation of this elicitation process resulted in a total of six sets of weights, with two sets from each stage.

4.4. Results analysis

4.4.1. Stakeholder preferences

The results for stakeholder preferences are presented in the following order: first, the results from the criteria selection survey elicited in the first stage (Individual preference - part 1). Second, a comparison of the weight sets that resulted from the second stage (Group preference). Finally, a comparison of the weight sets that resulted from stage one (Individual preference - part 1) and stage three (Individual preference - part 2), representing preferences before and after group deliberation.

Fig. 9 presents the results from the criteria selection survey elicited in the stage one (Individual preference - part 1). From the pre-defined six criteria, all 13 participants selected Climate Change and Damage to human health, while 12 participants selected Damage to ecosystems as relevant criterion. 10 participants selected Resource depletion, 9 participants selected Costs and 6 participants Value added. Furthermore, participants could suggest criteria, which were missing in the list of pre-defined criteria from their point of view (see Fig. 5b). The stakeholders suggested 11 criteria. Stakeholders using a car few times per year proposed 7 additional criteria: water use, materials, ease of use, availability of fuel stations, transport capacity, social impact of production, availability on demand. Stakeholders using a car few times per month suggested four additional criteria: operational safety, material extraction, lifetime expectancy, disposal process.

Fig. 10 shows a comparative cobweb chart of the weight sets that resulted from the second stage (Group preference) and equal weights. Both groups assigned the highest importance to the criterion Climate change, followed by Damage to ecosystems and Damage to human health. The lowest importance was given to the criterion Value added, although in a more significant way from Group 1. After seeing the resulting ranking in HELDA, both groups agreed to keep the weighting set obtained using the SRF method without modification.

Fig. 11 presents the weight sets resulting from the individual online surveys for the stakeholder categories few times per month (a) and few times per year (b) before and after the group preference stage. Fig. 11a shows that the average weight of the criterion Costs reduced after the group preference, while the weights of the criteria Damage to ecosystems, Damage to human health and Resource depletion increased. Similarly, Fig. 11b shows reductions of in the weight of the criteria Costs and Climate Change, respectively. Instead, the average weight increased for the criteria Value added and Damage to Human Health after group preference.

4.4.2. Indicative ranking

Fig. 12 presents the resulting net flow scores (NFS) of the five mobility options using the weight sets from the individual survey part 1 and part 2, and group preferences using the SRF method. The ranking is hardly affected by the different weight elicitation processes, i.e.

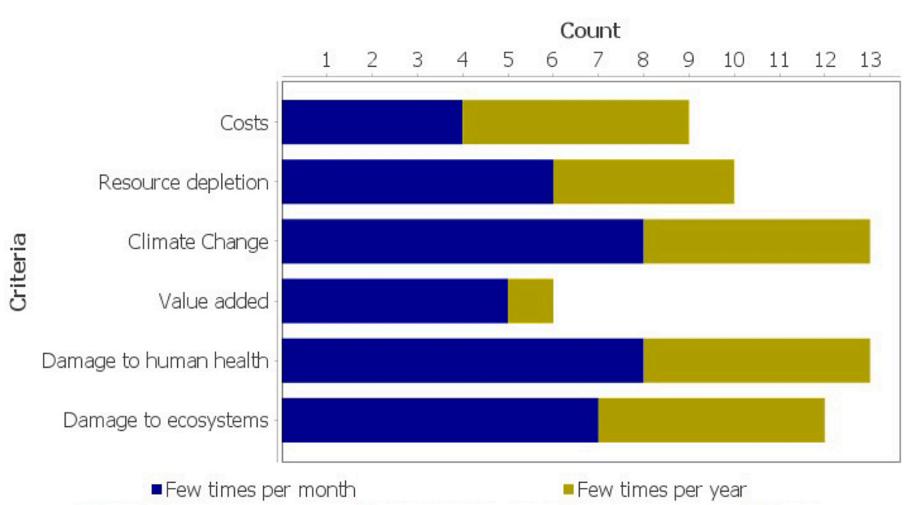

Fig. 9. Total votes of criteria (criteria selection survey).

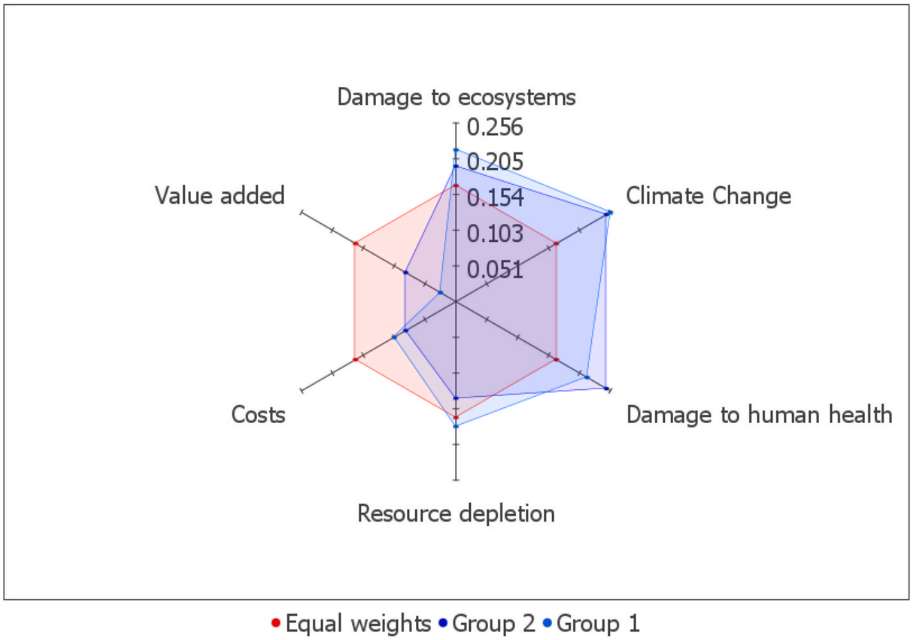

Fig. 10. Resulting weight sets from the group preference stage using the SRF method.

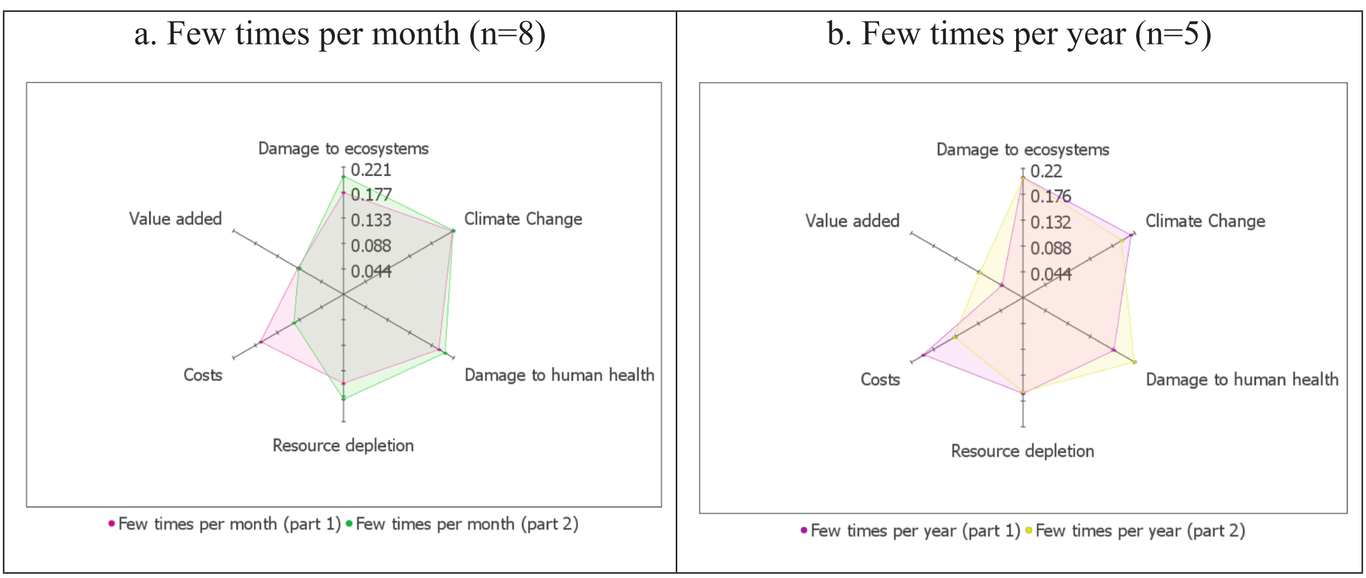

Fig. 11. Resulting weight sets from individual online surveys before (part 1) and after (part 2) group discussion for each stakeholder category.

individual preferences via online surveys and group preferences via SRF method, and information status (individual weights before and after group work). BEV-wind ranks first (highest NFS), followed by FCEV-wind in second place and ICEV-fossil ranks last (lowest NFS) for all mentioned weight sets. BEV Mix_DE performs better than ICEV-straw, except for Group 2, for which ICEV-straw (rank 3) has higher NFS than BEV-Mix_DE (rank 4).

4.4.3. Output variability analysis

4.4.3.1. Sensitivity analysis. This section illustrates the module for sensitivity analysis of thresholds in HELDA using the weight set from Group 1. Fig. 13 presents the sensitivity of the NFS to single variations of p and q thresholds in the criterion Climate change. The variations of q and p have no significant effect on the resulting NFS. The graph indicates that an exchange of positions in the ranking between BEV-MIX_DE and FCEV-wind would occur if the p threshold value is 6.9 times higher than the one used in the model (see Table 7).

4.4.3.2. Uncertainty analysis. MCS with 10.000 samples was run in HELDA to quantify the uncertainties related to the different weights elicited via individual surveys. The preference information used is presented as histograms of criteria weights in SI 3. Fig. 14 shows the results of the MCS using the individual survey after group discussion (Individual preference – part 2). Uncertainty analysis shows that the first two positions in the ranking and the last position are rather robust: BEV-wind ranks first with 100% probability, FCEV-wind ranks second with 99.8% probability, ICEV-fossil ranks last with 89.7% probability, respectively. Results for third and fourth rank are less stable: BEV-mix_DE ranks third with 75.3% probability, ICEV-straw ranks fourth with 67.5% probability.

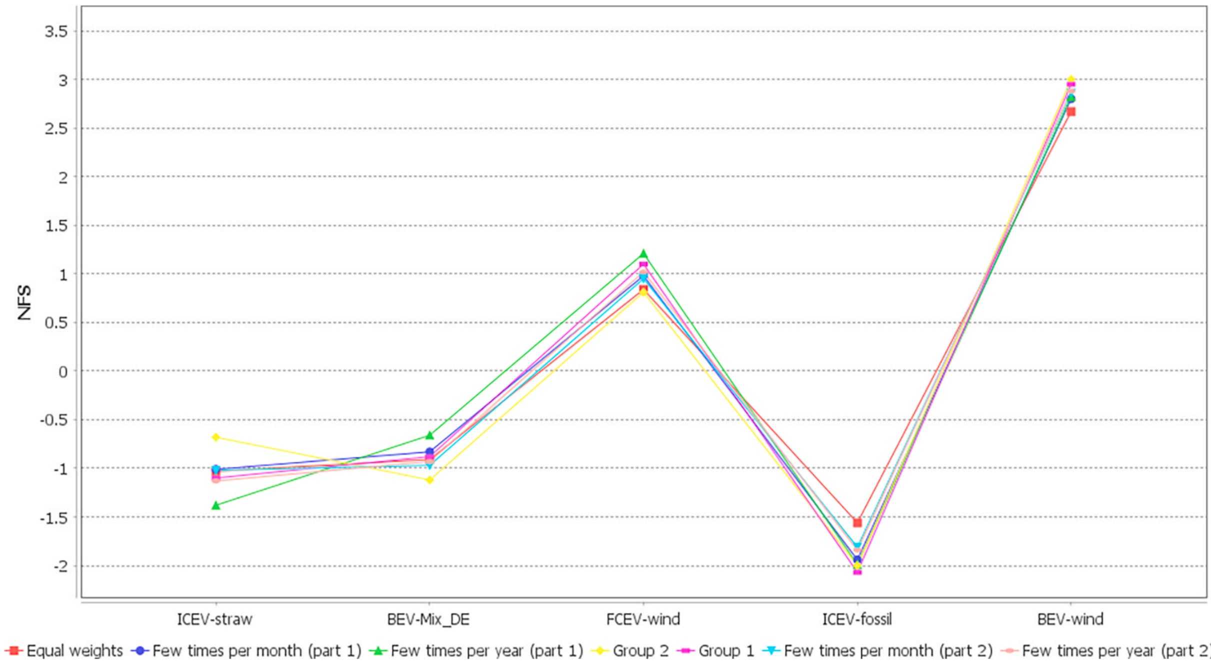

Fig. 12. Calculated net flow scores (NFS) using the weight sets obtained in the preference elicitation processes.

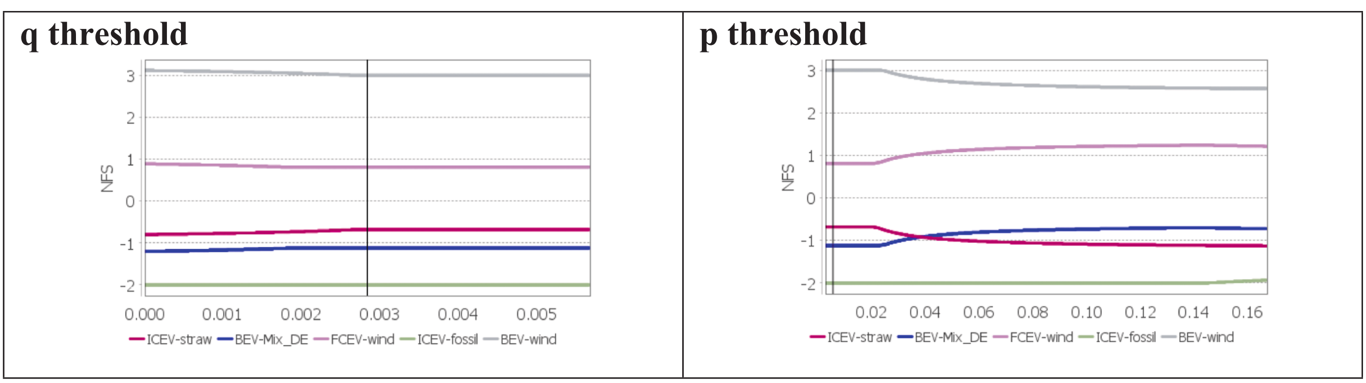

Fig. 13. Sensitivity analysis of discriminating thresholds from criterion Climate change using the weight sets from Group 1.

4.4.3.3. Stakeholders feedback on HELDA. A key objective of developing HELDA was to accelerate stakeholder integration for the energy transition by reducing the level of expertise and time required to elicit stakeholder preferences. During the two-hour workshop, participants were able to provide their individual preferences, collaborate with other stakeholders to provide group preferences, and observe the impact of these preferences on the decision model within HELDA. Participants used their own mobile devices to access the online surveys. Data gathering and processing with the online survey plug-in in HELDA ran smoothly. The same applied to displaying the results of group weighting with the SRF method, and the visualization of the resulting weight sets and ranking of alternatives. As final feedback for the exercise, participants were asked to fill in a survey answering five questions, two related to prior knowledge of participants (knowledge in MCDA, previous participation in similar formats), four related to their experiences during the workshop (ease of use of the six-point-scale for weighting of criteria, suitability of the SRF method to reach a group consensus on criteria weights, possible future utilization of HELDA). The exact wording of the questions as well as related answers (pre-defined categories as well as free text fields), can be found in SI 4.

Of the 13 participants, 11 filled in the final feedback completely. The group of participants was diverse in terms of their familiarity with MCDA methods. Of the pre-defined categories to express familiarity with MCDA methods, two participants stated they were “extremely familiar”, one participant “very familiar”, four participants “moderately familiar”, two participants “slightly familiar”, and two participants “not at all familiar”. All participants who stated to be “extremely” or “moderately” familiar already participated in similar exercises.

When asking participants if interacting with HELDA helped them to improve their understanding of the decision-making process, six participants stated “very much”, while three participants stated “moderately” and two participants stated “slightly”. Participants reported overall confidence in expressing their preferences with the six-point-scale for criteria weighting with nine participants choosing “Yes, to some extent” and two participants choosing “Yes, definitely”.

The suitability of the SRF method for finding a group consensus was assessed by one participant with “Very easy”, by six participants with “Somewhat easy”, by two participants with “Somewhat difficult”, and by two participants with “Neutral”. From 4 participants giving additional feedback on the SRF method exercise using the free text fields, two were stating that “loudest voices” or “heavy participants” overtake discussions. It is important to emphasize that this is a natural behavior in group dynamics (Dean, 2022); however, following a consensus reaching process, it is a necessary step of decision-making processes related to

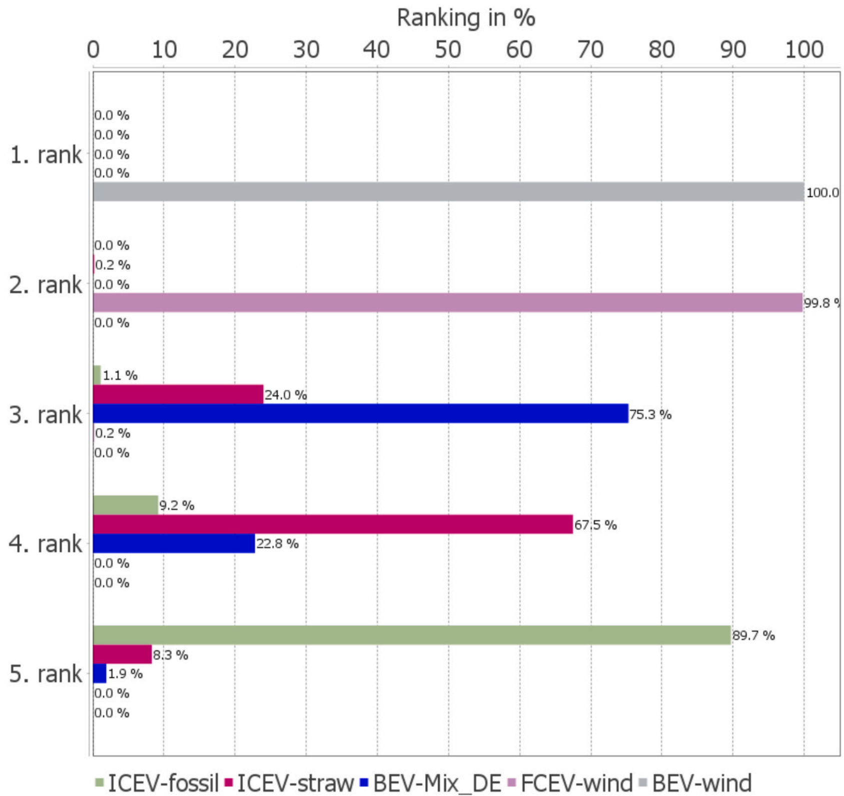

Fig. 14. MCS based on histograms of individual survey after group discussion (Individual preference – part 2).

energy system transitions (and other fields). Acknowledging this situation, we designed the individual elicitation of preferences with the plug-in in HELDA.

Concerning future utilization or recommendation of HELDA, eight participants stated "Yes" and three participants "Maybe". From those stating yes, two gave additional comments, emphasizing the possibility to use a different aggregation method and the real time analysis of preferences and visualization capabilities in HELDA respectively. Using this feedback, we could validate that interacting with HELDA supports participants in enhancing their understanding of decision-making processes. Furthermore, the participants felt confident expressing their preferences and acknowledging how these were considered in the MCDA process and the final ranking of alternatives.

5. Discussion

5.1. HELDA for sustainable energy system transformations

Sustainability assessments within the energy transition occur across multiple contexts, including energy planning, environmental management, and science-based policy advice (Wulf et al., 2025). HELDA was designed to support these processes by combining expert or technical knowledge and stakeholder preferences in decision-making (Mesa Estrada et al., 2026). HELDA software operationalizes the integration of qualitative and quantitative information from diverse methods, such as LCA, optimization models, and acceptance surveys, with societal values and political positions. Through its methodological features, HELDA facilitates comparisons and rankings of alternatives under diverse preferences, empowering stakeholders to understand their own preferences, as well as those of others. By transparently managing preferences, HELDA fosters mutual learning and the development of transformational knowledge essential for energy system transformations (Grunwald, 2017). The robust capabilities of HELDA to conduct output variability analysis enable users to consider the effect of inherent uncertainties in SA and energy system decision problems, e.g. resources availability and geopolitical risks (Ivanovski & Marinucci, 2021). These considerations position HELDA as a practical instrument for translating scientific evidence into actionable policy and decision recommendations, while maintaining transparency in decision-making processes.

5.2. HELDA for sustainability practitioners and energy researchers

HELDA was developed to provide sustainability practitioners and energy researchers with a MCDA software for SA in the energy transition context, combining scientific rigor with practical applicability. Its capabilities were tested in various applications and are illustrated in this paper through a use case on sustainability assessment of individual motorized mobility options. HELDA offers three core functions: 1) preference modeling using binary relations and scoring function methods; 2) stakeholder integration through an accessible online survey interface for preference elicitation and a SRF method module, and 3) output variability analysis for performance values, weights and thresholds. The application workflow follows a structured sequence: first, the user collaborates with decision makers and/or experts to structure the decision problem (elaborate evaluation matrix and identify stakeholder categories) before entering HELDA. Second, users access the *Method Selection Dialogue* to choose proper aggregation methods, bearing in mind that while HELDA development focused on binary relation methods aligned with strong sustainability principles, scoring functions might be preferable when only quantitative data is available. Third, create online surveys for individual preference elicitation. HELDA software design supports iterative analysis through multiple preference elicitation rounds with different stakeholder categories. HELDA is positioned as a decision support tool rather than a solution generator, facilitating analysis of complex decision-making processes through transparent, reproducible workflows while leaving final decisions to decision makers.

5.3. HELDA development process

HELDA is the result of a collaboration between sustainability experts, MCDA specialists, and software developers, demonstrating how interdisciplinary cooperation can produce solutions tailored to specific contexts. This is in the context of exploring solutions for the sustainability of academic software beyond the provision of open-source software tools (e.g., Haag et al. (2022)). Three features are used to discuss the HELDA development process and its long-term usability (Mesa Estrada et al., 2026): licensing models, software maintenance, and software users community. The development process of HELDA was facilitated by two key features: the use of the plug-in subsystem to add capabilities for stakeholder integration, and open-source Python libraries to extend preference modelling within the core code. The use of libraries represented a challenge as the codes were translated to Java manually. Current AI-driven tools for translating Python into Java code could support and improve this process, e.g. (Rgaguena, Chlioui, & Radgui, 2025). The use of AI-driven tools for enhancing HELDA is not considered at the moment, as the intellectual property rights are restricted with the type of license. Beyond the programming capabilities, having the support of the software developer was key in the testing phase of the software enhancements, to facilitate the solution of bugs and errors. In terms of software maintenance and the user community, we see a strong codependency. Active maintenance requires resources and funding, which could be provided by various projects in which HELDA could be applied. These collaborative and networking efforts could result in the establishment of user communities that continuously learn, apply, and enhance HELDA capabilities. We encourage collaboration between MCDA users (researchers from diverse fields) and MCDA software developers to adapt existing software for specific research objectives, thereby strengthening software user communities and ensuring long-term usability. In this study, we demonstrated how through iterative enhancement, HELDA evolved into a robust platform capable of addressing a wide range of decision problems using both scoring and binary relations methods, integrating stakeholders and conducting output variability analysis.

5.4. Limitations and outlook

HELDA functionality is intuitive and robust. Nevertheless, this is an ongoing project, and as such, certain features may have not been thoroughly tested or may require refinement or extension. This section mentions these features and provides some strategies that would be required to fulfil this.

HELDA is currently available exclusively as a desktop application. While desktop applications provide enhanced local processing capabilities and data security, particularly advantageous for handling sensitive information, web-based alternatives offer greater accessibility and cross-platform compatibility across devices, e.g. computers, smartphones, and tablets. To bridge this gap, a web-based version of HELDA is currently under development.

The utilization of HELDA necessitates a fundamental comprehension of problem structuring methods (PSM). Users need to structure the problem independently, i.e. define alternatives, criteria and stakeholders, before using HELDA. There are several approaches for problem structuring in MCDA that could be implemented (Marttunen, Lienert, & Belton, 2017). The implementation of features for PSM is out of the scope in HELDA and the user is free to employ the most convenient method for this.

Future work on HELDA will focus on enhancing its practical applicability by integrating .mcda files and related descriptions of exemplary use cases relevant to the energy transition. These use cases will serve as reference for users, enabling them to evaluate how other researchers define decision problems and to explore HELDA's analytical capabilities. Additionally, future work may include interoperability with specialized software, such as LCA tools, to ensure seamless data exchange and integration.

The primary objective of developing the HELDA online survey plug-in was the integration of stakeholders. However, the solution developed has potential for the integration of citizens in SA-MCDA. Further development of HELDA involves extending the plug-in for online surveys to additionally request socio-demographic and/or socio-economic data from participants. The online survey plug-in is designed for eliciting weights as importance coefficients for binary relations type of aggregation methods. Future work will include integrating a solution suitable for scoring functions (trade-offs) with e.g. sliding bars. Moreover, the HELDA online survey plug-in has the capacity to create surveys for both, flat and hierarchical structures of criteria. Despite the stability of the functionality for hierarchies, this implementation has not yet been tested.

HELDA supports one type of decision recommendation: ranking. Within the current project it has not been considered to add other types of decision recommendations, i.e. sorting, choice, clustering, yet other types of users could benefit from it.

When conducting output variability analysis, the MCS module is not applicable in its current form to the ELECTRE III-distillation method, as it does not provide cardinal ranking. Further effort is required to adapt this module so that it can work with the preference relation incomparability.

Finally, the current version of HELDA can only process one decision model at a time. This means that it is not possible to carry out simultaneous comparison of the results with different information, such as new alternatives or variations of data. Future work in HELDA will include the implementation of a functionality that will allow several configurations of the decision model to be run simultaneously, such as scenario analysis.

6. Conclusions

Sustainable energy system transitions require systematic approaches and software tools that balance rapid, science-based decision-making with the integration of diverse stakeholders, such as those involved in policy advice, energy planning, and environmental management. This paper presents the conceptualization of HELDA, a software tool developed based on a systematic review on MCDA software for sustainability assessment. HELDA has been developed by extending the functionality of the MCDA ToolKIT, a java desktop software developed by KIT. The development process of HELDA primarily involved extending stakeholder integration capabilities through its plug-in subsystem, and enhancing preference modeling by integrating aggregation methods from open-source libraries. The process of enhancing the functionality of the software was described, including the definition of requirements, estimation of effort for enhancements, prioritization of implementations and execution. The effort to enhance the software focused on two main priorities. First, the addition of MCDA methods suitable for conducting sustainability assessment using the concept of strong sustainability, i.e. methods that handle qualitative and quantitative information, allow null/partial compensation between criteria, and consider weights as importance coefficients. This resulted in the addition of ELECTRE III and PROMETHEE I and II aggregation methods. Second, to enable the integration of stakeholders by collecting and processing individual and group preferences. For individual preferences, a plug-in for online surveys was developed using an n-point scale. With this solution, stakeholders submit their preferences using an accessible and user-friendly interface. The SRF method was implemented in HELDA and used for eliciting group preferences. Unlike an ultimate solution, HELDA was developed as a specialized tool that integrates three core aspects of sustainability assessment: modeling strong sustainability, incorporating stakeholders' preferences, and addressing associated uncertainties. Its design prioritizes flexibility and practical applicability over universality. To our knowledge, HELDA is the only free MCDA software that integrates binary relations methods (ELECTRE III and PROMETHEE I and

II) with online surveys for individual weights elicitation of stakeholder groups (synchronous and asynchronous). With respect to output variability analysis, sensitivity analysis of preference thresholds has been implemented applicable to the newly implemented binary relation methods. Further enhancements involved the adaptation of the functionality of the software mainly with respect to these new extensions, including graphical representation and tables to present different types of information. Considering the importance on selecting a suitable method for a specific decision-making problem, a module for method selection was implemented based on the MCDA-Method Selection Software (MCDA-MSS) (Cinelli et al., 2022b). The highest effort corresponded to the conceptualization and implementation of the plug-in for involvement of stakeholders, followed by the addition of binary relation methods, i.e. PROMETHEE I and II and ELECTRE III.

HELDA's capabilities were illustrated in this article with a use case for the sustainability assessment of individual motorized mobility options in a workshop. The exercise highlighted the functionality of HELDA, the advantages in terms of time and resources needed to collect preferences from stakeholders in individual and group settings, as well as the challenges related to group discussions and finding consensus. The steps presented in this paper for using HELDA for stakeholder integration are only an example; adaptations are required when working with e.g. larger groups or different objectives. HELDA is a flexible software that can be used in synchronous and asynchronous settings.

Although HELDA's conceptualization is based on MCDA capabilities for SA it can be applied to any context. Further work on HELDA will include (among others) developing a web-based interface, extending the usability of the online survey plug-in to collect socio-demographic and/or socio-economic data from participants, and the addition of methods with different types of decision recommendations (e.g. scoring, sorting). HELDA is still under development and open for collaboration and application in new projects.

CRediT authorship contribution statement

Laura Sofia Mesa Estrada: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Martina Haase:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Tim Müller:** Software, Validation, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2026.134456

Data availability

No data was used for the research described in the article.

References

Banker, R. D., Davis, G. B., & Slaughter, S. A. (1998). Software Development Practices, Software Complexity, and Software Maintenance Performance: A Field Study. *Management Science, 44*(4), 433–450. https://doi.org/10.1287/mnsc.44.4.433

Beekman, J. (2020, December 14). Decision Analysis Software Survey. Retrieved from https://pubsonline.informs.org/do/10.1287/orms.2020.06.04/full/.

Bouyssou, D. (1992). Ranking methods based on valued preference relations: A characterization of the net flow method. *European Journal of Operational Research, 60* (1), 61–67. https://doi.org/10.1016/0377-2217(92)90333-5

Chauhan, C., Dakoliya, M., & Sharma, R. (2023). Use of Fibonacci Sequence in Project Estimation. *Indian Journal of Science and Technology, 16*(33), 2649–2652.

Choudhari, J., & Suman, U. (2012). Story points based effort estimation model for software maintenance. *Procedia Technology, 4*, 761–765. https://doi.org/10.1016/j.protcy.2012.05.124

Cinelli, M., Burgherr, P., Kadziński, M., & Słowiński, R. (2022a). Proper and improper uses of MCDA methods in energy systems analysis. *Decision Support Systems, 163*, Article 113848. https://doi.org/10.1016/j.dss.2022.113848

Cinelli, M., Coles, S. R., & Kirwan, K. (2014). Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. *Ecological Indicators, 46*, 138–148. https://doi.org/10.1016/j.ecolind.2014.06.011

Cinelli, M., Kadzinski, M., Gonzalez, M., & Roman, S. (2021). Recommending Multiple Criteria Decision Analysis Methods with A New Taxonomy-based Decision Support System. *arXiv preprint arXiv:2106.07378*. Retrieved from https://arxiv.org/abs/2106.07378.

Cinelli, M., Kadziński, M., Miebs, G., Gonzalez, M., & Słowiński, R. (2022b). Recommending multiple criteria decision analysis methods with a new taxonomy-based decision support system. *European Journal of Operational Research, 302*(2), 633–651. https://doi.org/10.1016/j.ejor.2022.01.011

Cinelli, M., Spada, M., Kim, W., Zhang, Y., & Burgherr, P. (2021b). MCDA Index Tool: An interactive software to develop indices and rankings. *Environment Systems and Decisions, 41*(1), 82–109. https://doi.org/10.1007/s10669-020-09784-x

Conceptboard. (2025). Conceptboard Retrieved from https://conceptboard.com/about-us/.

Couture-Grenier, N., & Abi-Zeid, I. (2015). MCDA ULaval: A Multi-Platform Multicriteria Software. Retrieved from *Newsletter of EURO Working Group on Multicriteria Decision Aiding, Serie, 3*(31) http://www.cs.put.poznan.pl/ewgmcda/newsletter/Spr15.pdf.

Dean, M. (2022). Including multiple perspectives in participatory multi-criteria analysis: A framework for investigation. *Evaluation, 28*(4), 505–539. https://doi.org/10.1177/13563890221123822

Figueira, J. R., & Roy, B. (2002). Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. *Eur. J. Oper. Res., 139*, 317–326.

Goepel, K. D. (2018). Implementation of an online software tool for the analytic hierarchy process (AHP-OS). *International Journal of the Analytic Hierarchy Process*. https://doi.org/10.13033/ijahp.v10i3.590

Grunwald, A. (2017). Technology assessment and policy advice in the field of sustainable development. In L. W. Zacher (Ed.), *Technology*. Society and Sustainability: Springer, Cham.

Haag, F., Aubert, A. H., & Lienert, J. (2022). ValueDecisions, a web app to support decisions with conflicting objectives, multiple stakeholders, and uncertainty. *Environmental Modelling and Software, 150*. https://doi.org/10.1016/j.envsoft.2022.105361

Haase, M., Wulf, C., Baumann, M., Ersoy, H., Koj, J. C., Harzendorf, F., & Mesa Estrada, L. S. (2022). Multi-criteria decision analysis for prospective sustainability assessment of alternative technologies and fuels for individual motorized transport. *Clean Technologies and Environmental Policy, 24*(10), 3171–3197. https://doi.org/10.1007/s10098-022-02407-w

Huang, H., Burgherr, P., & Macharis, C. (2025). A collaborative group decision-support system: The survey based multi-actor multi-criteria analysis (MAMCA) software. *Journal of the Operational Research Society, 76*(5), 844–865. https://doi.org/10.1080/01605682.2024.2398114

Ivanovski, K., & Marinucci, N. (2021). Policy uncertainty and renewable energy: Exploring the implications for global energy transitions, energy security, and environmental risk management. *Energy Research & Social Science, 82*, Article 102415. https://doi.org/10.1016/j.erss.2021.102415

Keeney, R. L. (1996). Value-focused thinking: Identifying decision opportunities and creating alternatives. *European Journal of Operational Research, 92*(3), 537–549. https://doi.org/10.1016/0377-2217(96)00004-5

Koasidis, K., Soursos, A., Xexakis, G., Labella, Á., Karamaneas, A., & Nikas, A. (2025). APOLLO-Live: A multi-criteria-based webtool for synchronous group decision making and consensus support in energy and climate policy deliberations [version 3; peer review: 2 approved, 2 approved with reservations]. *Open Research Europe, 5* (88). https://doi.org/10.12688/openreseurope.19614.3.

Krasucki, M. (2012). Java Library of Multicriteria Optimization Methods Retrieved from https://github.com/mateuszkrasucki/MCO_Java_Library.git.

Kumar, A., Sah, B., Singh, A. R., Deng, Y., He, X., Kumar, P., & Bansal, R. C. (2017). A review of multi criteria decision making (MCDM) towards sustainable renewable energy development. *Renewable and Sustainable Energy Reviews, 69*, 596–609. https://doi.org/10.1016/j.rser.2016.11.191

Langhans, S. D., Reichert, P., & Schuwirth, N. (2014). The method matters: A guide for indicator aggregation in ecological assessments. *Ecological Indicators, 45*, 494–507. https://doi.org/10.1016/j.ecolind.2014.05.014

LimeSurveyGmbH. (2025). LimeSurvey: An open source survey tool. https://www.limesurvey.org.

Lindfors, A. (2021). Assessing sustainability with multi-criteria methods: A methodologically focused literature review. *Environmental and Sustainability Indicators, 12*, Article 100149. https://doi.org/10.1016/j.indic.2021.100149

Livzeniece, L., Pubule, J., & Blumberga, D. (2021). Sustainability assessment of wind energy in latvia: Sustainability SWOT and multi-criteria analysis. *Environmental and Climate Technologies, 25*(1), 1253–1269. https://doi.org/10.2478/rtuect-2021-0095

Mareschal, B., & Smet, Y. D. (2009, 8-11 Dec. 2009). *Visual PROMETHEE: Developments of the PROMETHEE & GAIA multicriteria decision aid methods.* Paper presented at the 2009 IEEE International Conference on Industrial Engineering and Engineering Management.

Marttunen, M., Lienert, J., & Belton, V. (2017). Structuring problems for Multi-Criteria Decision Analysis in practice: A literature review of method combinations. *European*

Journal of Operational Research, 263(1), 1–17. https://doi.org/10.1016/j.ejor.2017.04.041

Mesa Estrada, L. S., Haase, M., Baumann, M., & Cinelli, M. (2026). Decision support for energy system transformation – A systematic analysis of MADM software for sustainability assessment. *Energy Strategy Reviews, 63*, Article 102016. https://doi.org/10.1016/j.esr.2025.102016

Moktadir, M. A., & Ren, J. (2023). Promoting sustainable management of hazardous waste-to-wealth practices: An innovative integrated DPSIR and decision-making framework. *Journal of Environmental Management, 344*, Article 118470. https://doi.org/10.1016/j.jenvman.2023.118470

Moreno-Calderón, A., Tong, T. S., & Thokala, P. (2020). Multi-criteria decision analysis software in healthcare priority setting: A systematic review. *PharmacoEconomics, 38*(3), 269–283. https://doi.org/10.1007/s40273-019-00863-9

Müller, T., Bai, S., & Raskob, W. (2020). MCDA handling uncertainties. *Radioprotection, 55*, S181–S185. https://doi.org/10.1051/radiopro/2020030

Munda, G. (2004). Social multi-criteria evaluation: Methodological foundations and operational consequences. *European Journal of Operational Research, 158*(3), 662–677. https://doi.org/10.1016/S0377-2217(03)00369-2

Munda, G. (2005). Multiple criteria decision analysis and sustainable development. In J. Figueira, S. Greco, & M. Ehrogott (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys* (pp. 953–986). New York, NY: Springer, New York.

Munda, G. (2022). SOCRATES (SOcial multi-CRiteria AssessmenT of European policieS). https://doi.org/10.13140/RG.2.2.17252.86401.

Mustajoki, J., & Marttunen, M. (2017). Comparison of multi-criteria decision analytical software for supporting environmental planning processes. *Environmental Modelling & Software, 93*, 78–91. https://doi.org/10.1016/j.envsoft.2017.02.026

Ness, B., Urbel-Piirsalu, E., Anderberg, S., & Olsson, L. (2007). Categorising tools for sustainability assessment. *Ecological Economics, 60*(3), 498–508. https://doi.org/10.1016/j.ecolecon.2006.07.023

Pereira, V., Basilio, M. P., & Santos, C. H. T. (2025). Enhancing decision analysis with a large language model: pyDecision a comprehensive library of MCDA methods in Python. *Journal of Modelling in Management*. https://doi.org/10.1108/jm2-04-2024-0118

Peters, S., Tönsfeuerborn, M., & von Nitzsch, R. (2024). Integrating uncertainties in a multi-criteria decision analysis with the entscheidungsnavi. Retrieved from *Mathematics, 12*(11), 1746 https://www.mdpi.com/2227-7390/12/11/1746.

Pohl, E., & Geldermann, J. (2024). PROMETHEE-Cloud: A web app to support multi-criteria decisions. *EURO Journal on Decision Processes, 12*, Article 100053. https://doi.org/10.1016/j.ejdp.2024.100053

Rgaguena, A., Chlioui, I., & Radgui, M. (2025). Generative AI for code translation: a systematic mapping study. Retrieved from *Engineering Proceedings, 112*(1), 33 https://www.mdpi.com/2673-4591/112/1/33.

Roy, B., Figueira, J. R., & Almeida-Dias, J. (2014). Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: Theoretical results and practical issues. *Omega, 43*, 9–20. https://doi.org/10.1016/j.omega.2013.05.003

Sarmas, E., Xidonas, P., & Doukas, H. (2020). *Multicriteria portfolio construction with python*. Springer.

Turkson, C., Acquaye, A., Liu, W., & Papadopoulos, T. (2020). Sustainability assessment of energy production: A critical review of methods, measures and issues. *Journal of Environmental Management, 264*, Article 110464. https://doi.org/10.1016/j.jenvman.2020.110464

von Nitzsch, R., Tönsfeuerborn, M., & Siebert, J. U. (2020) Decision Skill Training with the Entscheidungsnavi. In*: Vol. 405. Lecture Notes in Business Information Processing* (pp. 15-30).

Wątróbski, J., Jankowski, J., Ziemba, P., Karczmarczyk, A., & Zioło, M. (2019a). Generalised framework for multi-criteria method selection. *Omega, 86*, 107–124. https://doi.org/10.1016/j.omega.2018.07.004

Wątróbski, J., Jankowski, J., Ziemba, P., Karczmarczyk, A., & Zioło, M. (2019b). Generalised framework for multi-criteria method selection: Rule set database and exemplary decision support system implementation blueprints. *Data in Brief, 22*, 639–642. https://doi.org/10.1016/j.dib.2018.12.015

Weistroffer, H., & Li, Y. (2016). Multiple criteria decision analysis software. In M. E. Salvatore Greco, & J. R. Figueira (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys* (Vol. 233, pp. 1301–1341). New York, NY: Springer.

Wulf, C., Mesa Estrada, L. S., Haase, M., Tippe, M., Wigger, H., & Brand-Daniels, U. (2025). MCDA for the sustainability assessment of energy technologies and systems: Identifying challenges and opportunities. *Energy, Sustainability and Society, 15*(1), 45. https://doi.org/10.1186/s13705-025-00546-8

Ziemba, P. (2019). Towards Strong Sustainability Management—A Generalized PROSA Method. *Sustainability, 11*(6), 1555. Retrieved from https://www.mdpi.com/2071-1050/11/6/1555.