Energy Strategy Reviews 63 (2026) 102016

Contents lists available at ScienceDirect

Energy Strategy Reviews

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

Decision support for energy system transformation – A systematic analysis of MADM software for sustainability assessment

Laura Sofia Mesa Estrada [a,*], Martina Haase [a], Manuel Baumann [a], Marco Cinelli [b,c]

[a] Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Karlsruhe, 76021, Germany

[b] Decision Engineering for Sustainability and Resilience Laboratory, Leiden University College, Faculty Governance and Global Affairs, Leiden University, Anna van Buerenplein 301, The Hague, DG, 2595, the Netherlands

[c] Institute of Environmental Sciences (CML), Faculty of Science, Leiden University, P.O. Box 9518, Leiden, 2300 RA, the Netherlands

ARTICLE INFO

Handling Editor: Dr Xi Lu

Keywords:
Sustainability assessment
Multiple criteria decision analysis
Software assessment
Energy system transformation
Decision support

ABSTRACT

Multi-criteria decision analysis (MCDA) methods for sustainability assessment (SA) have been used in various decision-making processes for energy system transformation. However, concerns remain about the complexity and resource intensity of their application. The literature on MCDA and SA involves methodological advances and case studies, but lacks of systematic analysis of MCDA software for decision support. This study conducts a systematic analysis on multi-attribute decision making (MADM) software and their functionality for SA from a theoretical and user perspective. The users are SA practitioners with different levels of knowledge and expertise of MADM methods. The functionality is determined by a set of features gathered from a literature review of studies conducting SA with MADM software and from dedicated literature exploring the capabilities of MADM methods for SA. These features and capabilities are translated into assessment criteria and domains that encompass the entire MADM process, from supporting problem formulation and preference elicitation to robustness analysis. A sample of 25 free MADM software was assessed using 29 criteria across eight domains. Finally, recommendations are provided based on an effort estimation required to extend the software functionality. Results show that the MCDA community aims to provide practitioners with reliable free MADM software. However, weaknesses are identified in the software assessed, particularly in the functionality for stakeholder's involvement, output variability analysis, and problem structuring. Collaborative initiatives involving SA practitioners, software developers and the MCDA community, can help to accordingly enhance MADM software for SA.

List of abbreviations

AHP	Analytical hierarchy process
ANP	Analytical network process
DEX	Decision expert
DPSIR	Drivers, Pressures, State, Impacts, Responses
DRSA	The dominance-based rough set approach
ELECTRE	ÉLimination Et Choix Traduisant la REalité ("Elimination and Choice Translating Reality").
ELECTRE ME	ELECTRE with Multiple Evaluators
ELECTRE Tri	ELECTRE for sorting of alternatives
FITradeoff	Flexible and Interactive Tradeoff
GIS	Geographic Information System
LCA	Life Cycle Assessment
LCSA	Life Cycle Sustainability Assessment
MAVT	Multi attribute value theory
MADM	Multi-Criteria Decision Analysis
OVA	Output variability analysis
PROMETHEE	Preference Ranking Organisation Method for Enrichment Evaluations
PSM	Problem structuring method
SA	Sustainability Assessment
SAW	Simple Additive Weighting
SMART	Simple Multi-Attribute Rating Technique
SMARTER	Simple Multi Attribute Rating Technique Exploiting Ranks
SMARTS	Simple Multi Attribute Rating Technique using Swings
SMCE	Social Multi-Criteria Evaluation
SWOT	Strengths, Weaknesses, Opportunities, and Threats
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
VC-DRSA	Variable Consistency Dominance-based Rough Set Approach

(continued on next page)

* Corresponding author.
E-mail address: laura.mesa-estrada@kit.de (L.S. Mesa Estrada).

https://doi.org/10.1016/j.esr.2025.102016
Received 28 May 2025; Received in revised form 16 December 2025; Accepted 16 December 2025
Available online 21 January 2026

(*continued*)

VIKOR	Vlse Kriterijumska Optimizacija I Kompromisno Resenj (Multicriteria Optimization and Compromise Solution)
VFT	Value-focused thinking
WASPAS	Weighted Aggregated Sum Product Assessment

1. Introduction

The transition towards a sustainable energy system is a complex process that entails a wide set of economic, environmental and societal aspects covering conversion, distribution, storage and use of energy. Accordingly, a wide set of stakeholders ranging from e.g. end users, grid operators and technology developers are involved leading to multi folded decision-making challenges inherent in the transformation of the energy system. It is thus imperative to consider the environmental, economic, and social dimensions in a holistic manner, with minimal trade-offs whilst incorporating the perspectives of relevant stakeholders [1]. However, the assessment of each dimension requires the use of a variety of different methods, e.g. environmental life cycle assessment, levelized cost of electricity, social life cycle assessment, with very different impact categories, criteria and implications. In addition, their results are always interdependent due to the underlying assumptions made (e.g. efficiency levels, materials used etc.). Due to the different types of data and scales involved, it is challenging to aggregate them into a single figure that can effectively inform decision-making processes. While this is a complex task, the Multi-Criteria Decision Analysis (MCDA) field provides structured methods to support decision-making by integrating technical knowledge, societal values, and related uncertainties. MCDA is a subdiscipline of Operations Research that supports decision-makers (DMs) in identifying a solution from a given set of finite or infinite alternatives. MCDA methods can be divided into two categories: Multi-Attribute Decision Making (MADM) and Multi-Objective Decision Making (MODM) [2]. MADM methods allow DMs to choose from a finite number of explicit alternatives and criteria. By contrast, MODM methods address problems where the alternatives are only implicitly known and the criteria are expressed as mathematical objective functions to be optimized. The selection of these methods depends on the characteristics of the decision problem. In some cases, they could be complementary, i.e. the objective (formal mathematical approach) and subjective (including human judgement and preferences) formulation of the decision model [3]. Operational decision problems are commonly approached with MODM methods, whereas strategic decisions appear to be more commonly handled with MADM methods [4]. Sustainability assessment (SA) using MCDA methods has been widely applied as demonstrated in the reviews by Thies, Kieckhäfer, Spengler et al. [5], Lindfors [6], Dias, Caldeira and Sala [7], and Wulf, Mesa Estrada, Haase et al. [8]. Recognizing that the transformation of an energy system is a complex sociotechnical process in which strategic decisions play an important role, this paper focuses on SA with MADM (SA-MADM).

Extensive literature exists describing MADM methods and important concepts to be considered when conducting SA. These include the work of Cinelli, Coles and Kirwan [9], Cinelli, Koffler, Askham et al. [10], Huysveld, Taelman, Hackenhaar et al. [11], Munda [12], Talukder and Hipel [13], in which relevant concepts include e.g. type of aggregation method, type of information available, interactions between criteria, preferences and applicability. Furthermore, the underlying concepts of weak and strong sustainability can be supported by compensatory and non-compensatory MADM methods [12]. SA of energy technologies, materials, systems, pathways and scenarios have been conducted using hand tailored approaches focusing on a limited set of criteria and stakeholders [14]. Nevertheless, its application often requires high effort and resources such as people, time and money [15]. At the same time, decision-making processes regarding energy system transformation become more urgent and rapidly evolving, driven by faster technology development, shorter product life cycles and changing sustainability requirements (e.g., those arising from geopolitical tensions). In light of the challenges encountered in decision-making for the sustainable transformation of energy systems, computational solutions have the potential to facilitate the operationalization and acceleration of SA-MADM [14].

The advantages of using MADM software have been addressed in various publications through comparative assessments of the software's features in specific contexts. Mustajoki and Marttunen [16] compare MADM software features to support decision making in environmental planning processes, focusing on support for i) dealing with the systemic nature of impacts, ii) integration of multiple stakeholders, iii) geographical distribution of impacts, iv) dealing with uncertainties, and v) types of users. Moreno-Calderón, Tong and Thokala [17] evaluate MADM software to support health care priority settings based on i) aggregation method, ii) visualization, iii) sensitivity analysis, iv) cluster analysis, and vi) availability (cost), and v) documentation. There are also other sources that list and characterize different MADM software without providing a specific context, but rather a "directory" for software users. Weistroffer and Li [2] categorize the software according to: i) characteristics of the decision problem, ii) MADM methods implemented by the software, iii) the type of decision problem (group decision making vs single decision maker), and iv) the platform(s) supported by the software. Mohamad and Selamat [18] compare software supporting rough set theory based on: i) model construction, ii) rough set type, iii) criteria weighting, and iv) results analysis. The ORMS Today survey [19] characterizes MADM software based on: i) decision analysis applications, ii) usability analysis and visualizations, and iii) licensing and training. Cinelli, Spada, Kim et al. [20] compare MADM software for scoring and ranking with a focus on output variability analysis (and visualization) with uncertainty and sensitivity analysis of input data and models. The International Society on MCDM [21] presents MADM software by license: free software, semi-commercial and commercial software. Huang and Burgherr [22] compare free MADM software for ranking and scoring based on their support for i) problem structuring, ii) model building, and iii) challenging thinking (i.e., results analysis).

Despite the extensive literature on SA-MADM and MADM software, there is no literature that assesses systematically generic MADM software to support SA, both theoretically and from a user perspective. This research addresses the identified gap by conducting a systematic analysis of the strengths and weaknesses of MADM software tools' functionality with regard to SA, establishing a connection between existing features and recommendations for software enhancement. Given the lack of literature, it was decided to conduct the analysis of MADM software functionality from a general perspective of SA so it is relevant for decision-making in the energy context but also other fields. The software functionality refers to the range of operations that a software can perform which meet the stated and implied needs of users under specific conditions [23]. The type of users considered in this study are researchers of all levels of expertise in MCDA methods.

This paper presents a systematic analysis that consists of three stages. First, development of a set of criteria based on a literature review of MADM software users' motivations and their relation to the capabilities of SA-MADM described in the literature. Second, identification of MADM software tools that are regularly maintained and freely available from existing inventories, literature and websites. Third, assessment of the selected MADM software against the set of criteria identified in the first stage and to provide recommendations for MADM software enhancement based on an estimate of the effort required.

This paper is organized as follows: Section 2 presents the methodology used to develop a set of criteria for assessing MADM software, to select a sample of MADM software, and to assess MADM software functionality for SA. In section 3, the results of the criteria development, the assessment of selected MADM software, and recommendations for software extensions are presented. Section 4 discusses the main findings of the study. Finally, section 5 concludes on major findings and research gaps in the context of MADM software for SA.

2. Methodology

The methodological approach consisted of three stages as presented in Fig. 1. The first one being the development of criteria to assess MADM software applicability for SA. The second stage consisted of selecting a sample of MADM software, and the third stage involved the assessment of MADM software and elaboration of recommendations for software extensions. In this study, software is defined as an assembly of programs, procedures, rules, documentation and data, pertaining to the operation of an information processing system, in accordance with the International Electrotechnical Commission [24] definition.

2.1. Development of criteria to assess MADM software

In this stage, the procedures for the development of assessment criteria are described. These include (i) a literature review on motivations for the use of MADM software for SA (section 2.1.1), (ii) an analysis of the overlap between grouped motivations and MADM methods capabilities for SA (section 2.1.2), and (iii) a refinement of criteria for the resulting final set of assessment criteria (section 2.1.2), involving selection, exclusion and complementation.

2.1.1. Literature review on motivations for use of MADM software for SA

The development of criteria to assess software for SA-MADM was driven by the needs and expectations of users, i.e. motivations for the selected MADM software. This section aimed to identify the type of software used to conduct SA-MADM, e.g. existing software or self-programmed software, and the user's motivations for selecting them. A literature review was carried out to identify publications in which SA was performed using MADM software. Fig. 2 presents the methodology used in the literature review i.e. identification, screening, eligibility and inclusion of research articles. Identification consisted of a literature search performed using the Scopus data base with search string combinations consisting of the terms: "sustainability assessment" or "sustainability evaluation" and “tool*” or “software”, restricted to articles published in journals from July 2013 to November 2025, and written in English. The screening of titles and abstracts was performed using the semi-automated tool Rayyan [25] to identify articles using MADM methods. A variety of terms referring to MADM were used, including MCDA, multi-criteria decision analysis, multi-criteria techniques, MCDM, multi-criteria analysis, MCDA techniques, MCA, and the names

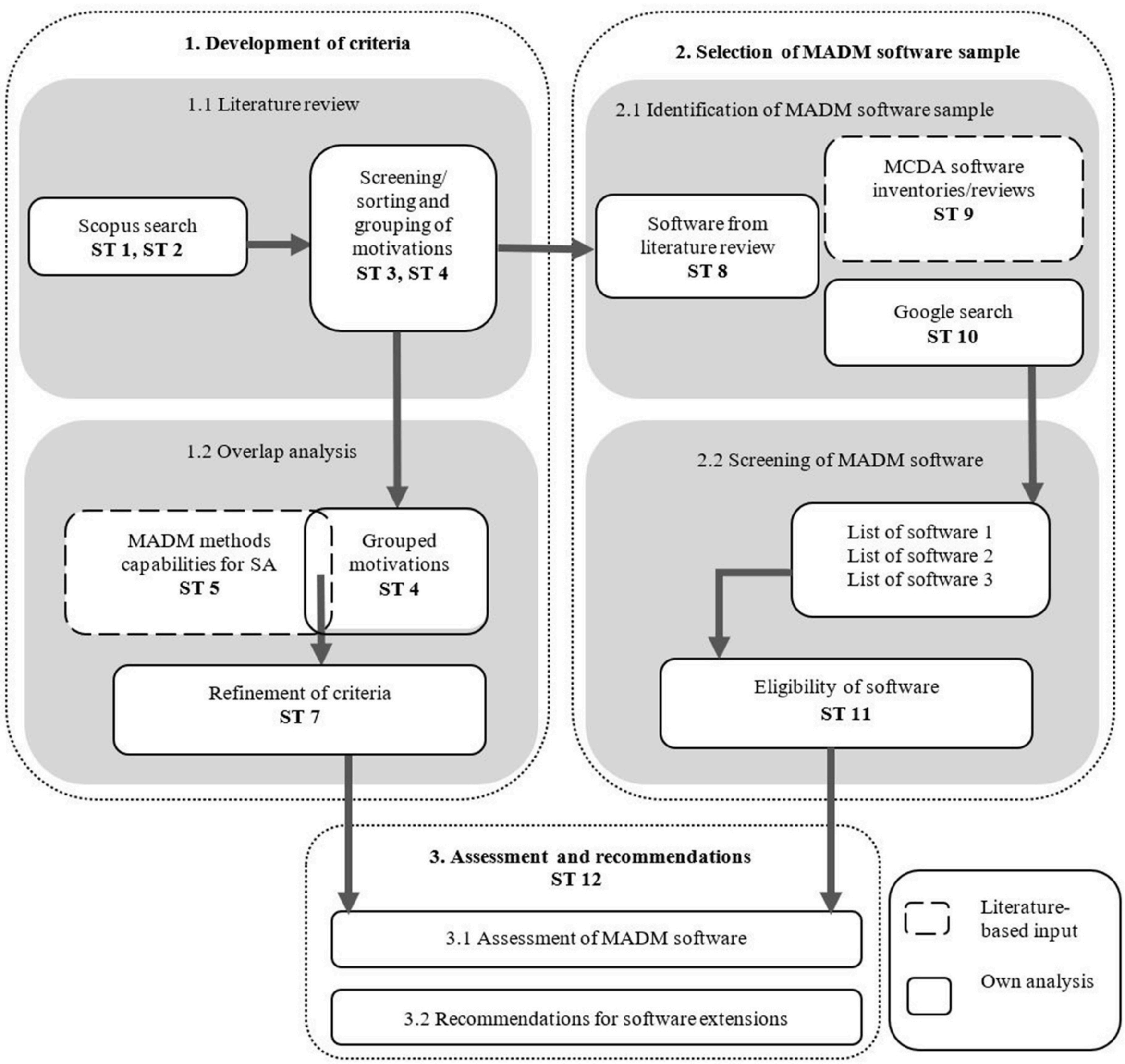

Fig. 1. Methodological approach for the development of criteria, selection of MADM software sample, and assessment of MADM software for SA. (ST = supplementary table).

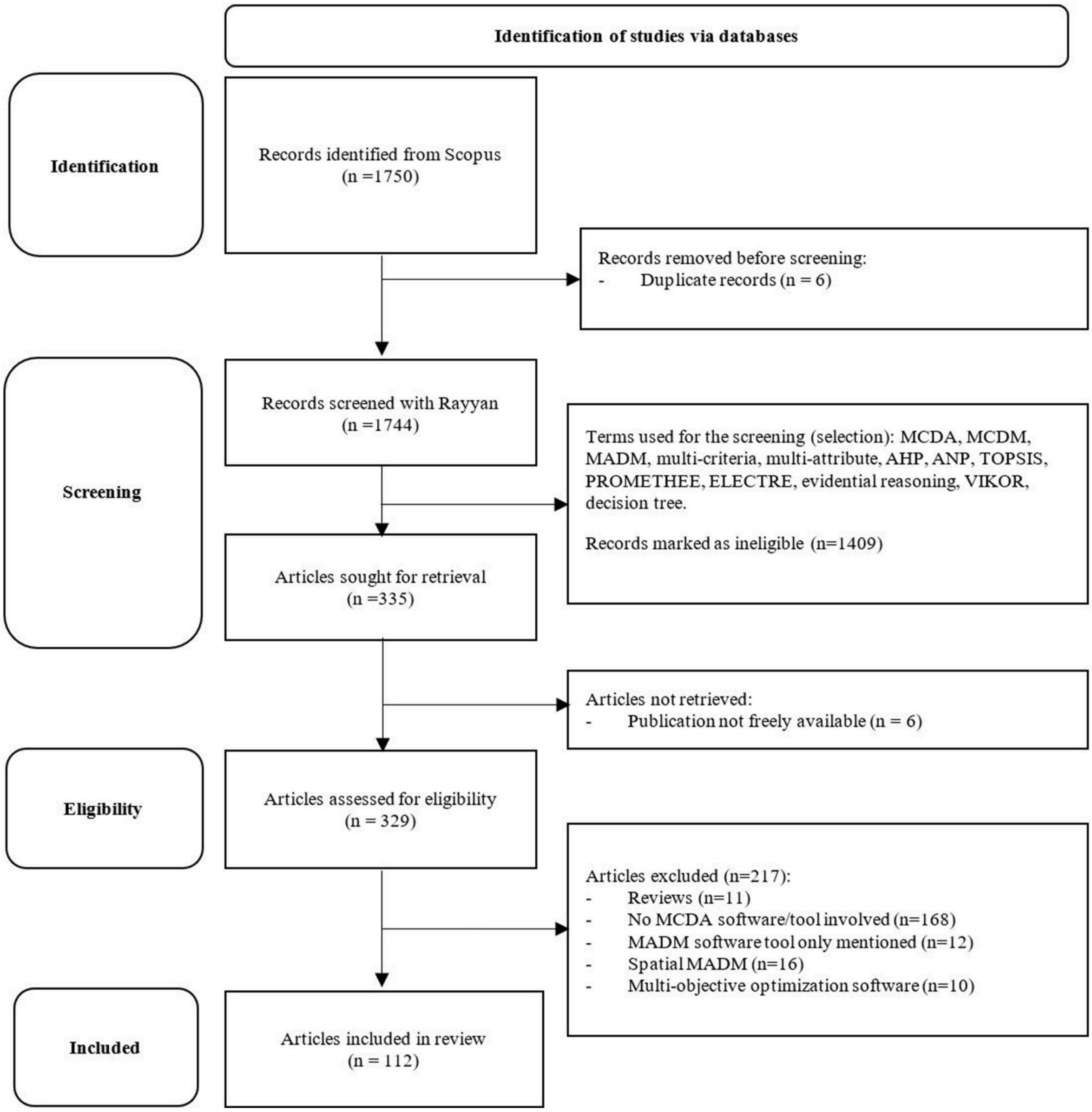

Fig. 2. Process of identification of studies performing SA with MADM software (Stage 1.1).

of specific methods, such as AHP, PROMETHEE, TOPSIS. This resulted in 335 articles that were searched for retrieval (see ST 1). As six publications were not freely available, finally, 329 articles were assessed for eligibility by excluding those where:

1. a literature review was conducted,
2. the terms “tools” and “software” where not explicitly mentioned in relation to MCDA methods
3. MADM software was mentioned but not used,
4. MADM methods were used within a geographic information system (GIS) software, and
5. a multi-objective optimization (MOO) software was used.

The selection process resulted in the identification of 112 articles in which MADM software was used to perform SA in diverse contexts (see Fig. 2 and ST 2).

The MADM software tools identified in the 112 articles were collated with information on the respective articles (e.g. year, authors, context, MADM method used) and divided into five software categories based on the general motivation for using a specific MADM software in the article, i.e. method-based and framework-based motivations:

I Method-based motivations
- Existing software - Generic
- Self-programmed software - Custom MADM method
- Self-programmed software - Existing MADM method

II Framework-based motivations
- Existing software - Application
- Self-programmed software - Application

The motivations of category I are centered on the method used and the need of a software to implement it. The motivations for category II are centered on the application of a sustainability framework within the software, which, in addition to a method, incorporates contextual criteria and, in some cases, sets of weights defined with experts in the specific context. Within these two categories, further distinctions emerge. The first distinction is between the use of existing software and self-programmed software, i.e. authors use programming languages or Excel to develop their own software. The second (relevant only for category I) pertains to the inclusion of existing or custom MADM methods for SA in the software. Having identified these five categories, a qualitative analysis of the specific motivations stated by the authors for using the respective software was conducted (see ST 3). To facilitate the overlapping process in stage 1.2, the identified specific motivations were sorted using different colors (see ST 3) and subsequently grouped (see ST 4).

2.1.2. Overlap of motivations for use of MADM software for SA and MADM methods capabilities for SA

In this section, it was analyzed how the user's motivations relate to the capabilities or theoretical concepts of MADM for SA described in the literature. For this, the groups of motivations (section 2.1.1) were overlapped with the MADM methods capabilities for SA to develop the criteria for assessing MADM software (ST 6). The capabilities were extracted from Lindfors [6], Dias, Caldeira and Sala [7], Cinelli, Coles and Kirwan [9], Huysveld, Taelman, Hackenhaar, Pihkola, Goedkoop, Isasa, Zanchi, Kujanpää, Harmens, Zamagni, Bianchi, Kamp, Bachmann, Alvarenga and Cordella [11], Munda [12], Talukder and Hipel [13], Cinelli, Spada, Kim, Zhang and Burgherr [20] (see ST 5). These were organized (if possible) according to the MCDA taxonomy of features for describing MCDA methods proposed by Cinelli, Kadziński, Miebs et al. [26]. This taxonomy is currently the most comprehensive framework of features to describe the MADM process and methods using a set of decision problem characteristics. These include features related to problem typology, preference model, preference elicitation, and types of decision recommendation. In instances of overlap, the criteria were named after the MCDA-MSS taxonomy of Cinelli, Kadziński, Miebs, Gonzalez and Słowiński [26] or the respective name given by the literature with which the group overlaps. In the specific case of output variability analysis (OVA), the proposed conceptualization of uncertainty and sensitivity analysis by Cinelli, Spada, Kim, Zhang and Burgherr [20] was used (see ST 5). It consists in the analysis of the effect of varying input data (performances and weights) and/or preference models (normalization methods and aggregation functions) in the results. Furthermore, additional criteria have been defined based on the contributions of Huysveld, Taelman, Hackenhaar, Pihkola, Goedkoop, Isasa, Zanchi, Kujanpää, Harmens, Zamagni, Bianchi, Kamp, Bachmann, Alvarenga and Cordella [11] and Talukder and Hipel [13] (see ST 5). In the absence of overlap, the criteria names were designated by the authors of this study. The domains were created following the same strategy indicated for the criteria, i.e. using existing literature when overlap occurred or created by the authors. Finally, the criteria refinement for the subsequent research step was performed, which led to exclusion and addition of criteria (see ST 7).

Finally, the resulting criteria were described and assigned a rationale to assess the level of fulfilment for each selected MADM software using a qualitative scale (weak, moderate, strong). The definitions of weak, moderate and strong were created for each criterion, in accordance with the existing literature on MADM methods capabilities for SA of Lindfors [6], Dias, Caldeira and Sala [7], Cinelli, Coles and Kirwan [9], Huysveld, Taelman, Hackenhaar, Pihkola, Goedkoop, Isasa, Zanchi, Kujanpää, Harmens, Zamagni, Bianchi, Kamp, Bachmann, Alvarenga and Cordella [11], Munda [12], Talukder and Hipel [13], Cinelli, Spada, Kim, Zhang and Burgherr [20], and the grouped motivations (see ST 6). The latter applies to criteria that are exclusively related to the motivations stated by the authors in the review, such as dynamic alternative/criteria management.

2.2. Selection of MADM software sample

The objective of this stage was to identify, from the extensive collection of MADM software that is available, those that are regularly maintained and freely available. The identification process was based on the literature review (see section 2.1.1), existing MADM software inventories/reviews, and a Google search (section 2.2.1). Afterwards, the MADM software was screened for eligibility (section 2.2.2).

2.2.1. Identification of MADM software

In this section, freely available MADM software was identified and characterized. Only non-commercial software is assessed in this study in consideration of the principles of open science, i.e. provide products from research activities, such as publications, data, and software free of charge and reusable to the user [27,28]. As stated above, the list of software 1 (see ST 8) was created based on the literature review of SA studies using MADM software. Second, the list of software 2 (see ST 9) was created using existing MADM software inventories available in Weistroffer and Li [2], Beekman [19], International Society on MCDM [21], and the review articles from Mustajoki and Marttunen [16], Moreno-Calderón, Tong and Thokala [17], Mohamad and Selamat [18], Cinelli, Spada, Kim, Zhang and Burgherr [20], Huang and Burgherr [22]. Third, the list of software 3 (ST 10) was created via Google search using the terms "MADM software" and "MADM tools". It should be noted that the purpose of this search strategy was to find a sample of MADM software and not to conduct an exhaustive search.

The following information was used to characterize the MADM software:

- Name of MADM software
- Context (only list of software 1 – software from literature review)
- Source (only for list of software 2 – software from inventories/review papers)
- Authors/developers: people and organizations
- Purpose of the software: generic (software which can be used for a variety of installations purely by the provision of application-specific data or algorithms, or both) or application (software that is specific to the solution of a problem submitted by a user)
- Platform: desktop, web-based, Excel, programming language, or client server
- MADM methods (for list of software 1 only the methods used in the reviewed article are mentioned, in the other lists all methods available in the software are mentioned)
- Availability: whether the software is currently available (online) or not
- Website
- Type of license: Open-source (software and source code freely available), freeware (software is freely available but not the source code), commercial (software developed for sale) and academic (software freely available only for academic purposes with or without restrictions).
- License
- Year of last version
- User interface or executable file

2.2.2. Screening of MADM software for eligibility

In this section, the software tools from list of software 1 (software from review on SA studies), list of software 2 (software from existing inventories/reviews) and list of software 3 (software from Google search) were screened using the following eligibility criteria and related rationales:

1. The software is available for download or can be accessed directly through a website. *Rationale:* Direct and easy access to the software as it accounts for user convenience.

2. The software is available for free, i.e. no payment is involved in accessing the software with full functionality, e.g. open source and freeware type of licenses. *Rationale:* Principles of open science and type of users considered in this study (researchers from different fields).
3. The latest version of the software was released between 2019 and 2025. *Rationale:* Higher probability of reaching software developers and/or an active user community which could support non-expert users.
4. The software has a user interface or executable file in the case of programming libraries. *Rationale:* Consideration of ease of use for potential users who may not be technical experts to use e.g. libraries in programming languages.

2.3. Assessment and recommendations

The aim of this stage was to identify how the selected MADM software tools fulfil the assessment criteria (which are based on user requirements and the capabilities of MADM for SA) and to provide recommendations to extend software functionality to take advantage of existing software functionality.

2.3.1. Assessment of MADM software

The selected MADM software was assessed against the developed set of criteria. The frequencies of weak, moderate and strong fulfilment per criterion were then calculated to identify the trends in the strengths and weaknesses of the software functionalities with respect to SA.

2.3.2. Recommendations of software extension for SA

Following the assessment of MADM software, an estimation of the software enhancement effort for SA was conducted. Software enhancement, as defined by Banker, Davis and Slaughter [29], refers to the modifications that extend, modify, and delete the functionality of an existing software for a specific use. The effort estimation is a time and cost estimation that is often dependent on software complexity, maintenance team experience, application size, application quality and the software functionality modified [29]. Given the different characteristics of the software in the sample, the estimation of effort was conducted according to the functionality of the software to be extended, in relation to the assessment criteria. A relative estimation of software enhancement effort was conducted based on the Story Points (SP) technique [30]. A SP is a metric of effort to execute a specific requirement (in this case assessment criteria) in terms of relative work, uncertainty and complexity [31]. The amount of work accounts for the resources and tasks involved in completing a requirement, for example, writing ten lines of code requires more work than writing one line. Complexity refers to the technical or cognitive difficulty in implementing a requirement. Uncertainty refers to unclear requirements that could challenge the implementation of the requirement, for example, a requirement that can be executed in different manners and requires not only technical expertise but a decision-making process. Therefore, lower values of SP, indicate solutions that can be implemented in shorter time using simple, standard and well-defined processes. Conversely, higher values of SP, indicate time-consuming solutions that involve not only technical challenges but collaboration among the developers to decide on the best way to proceed. In this study, instead of SP, a 4-point qualitative scale was used to compare the effort based on relative work, complexity and uncertainty related to the assessment criteria as presented in Table 1.

Table 1
4-point qualitative scale used to estimate the relative software enhancement effort for every criterion.

This study	Amount of work	Complexity	Uncertainty
Low	Little	None/little complexity	None
Low-moderate	Moderate	Little	None
Moderate-high	Moderate	Medium	Moderate
High	High	High	High

3. Results

3.1. Development of criteria to assess MADM software

This section presents an overview of the results from the screening and grouping of motivations (section 3.1.1) from the 112 selected articles. Following this, the results of the overlapping procedure (section 3.1.2) are presented, and then the refinement conducted to select the final assessment criteria.

3.1.1. Literature review on motivations for use of MADM software for SA

A total of 53 MADM software tools were identified from the 112 articles selected in the literature review (ST 3). Fig. 3 summarizes the information regarding the type of MADM software used in the reviewed literature categorized by the general motivations described in section 2.1.1. The majority of articles reviewed used existing software (60 %), with 37 % categorized as generic (e.g. Visual PROMETHEE, Dexi, and Super Decisions) and 23 % designated as application (e.g. SAFA Tol, SMART-farm tool and DEFINITE). Furthermore, 40 % used self-programmed software, 23 % with method-based motivation (existing and custom methods) and 17 % with application-based motivation. Despite the extensive range of software available for different methods, 19 % of the studies used self-programmed software with an existing method. Besides to the general motivations, Fig. 3 shows the platforms of the respective software, i.e. Desktop, Desktop/Web, Excel, Programming Language, Web. The 112 articles were manually screened to identify statements or keywords related to specific motivations for using the software (see section 2.1.1). About 60 % of the articles referred to method-based motivations, e.g. the method handles qualitative and quantitative criteria. The remaining 40 % used framework-based motivations e.g. the availability of criteria and indicators to support the operationalization of sustainability.

A total of 260 motivations were identified and sorted by colors according to the topics to which they referred, e.g. weights, alternatives, stakeholders, data (see ST 3). The outcome of this process was the definition of 41 groups, each of which was assigned a name that summarized the motivations within the group. For instance, the motivations "*has a wider application in natural resources applications*", "*widely used for sustainable agriculture assessments*", "*widely used in agriculture*", and "*widely used, one of the most comprehensive tools*" were sorted to Group 39 Widely used in a similar application/context. Fig. 4 presents the 41 groups (G) and the number of motivations that belong to each of them. G24 and G29 have the highest frequencies (n = 32): Users commonly mentioned the need of using a software to calculate and aggregate weights from preference information collected in questionnaires or surveys from several participants (G24) and selected a specific software based on the MADM method used to approach their decision problem (G29). Motivations in G24 and G29 were mostly found in articles referring to method-based motivations using existing generic software. For G31 (n = 22) users highlight the importance of a software including ready-to-use criteria and indicators sets that fit a specific context. This group was found in articles referring to framework-based motivations using existing and self-programmed software. Detailed information of the motivations that belong to each group is presented in ST 4.

3.1.2. Overlap of motivations for use of MADM software and MADM methods capabilities for SA

The 41 groups of motivations derived from the literature review were overlapped with the capabilities of MADM methods for SA. Overlapping means that groups of motivations were matched to criteria previously mentioned by authors in the context of MADM for SA. This process was guided by the aforementioned literature, with the aim of using consistent names for the criteria. In instances of overlap, the criteria names

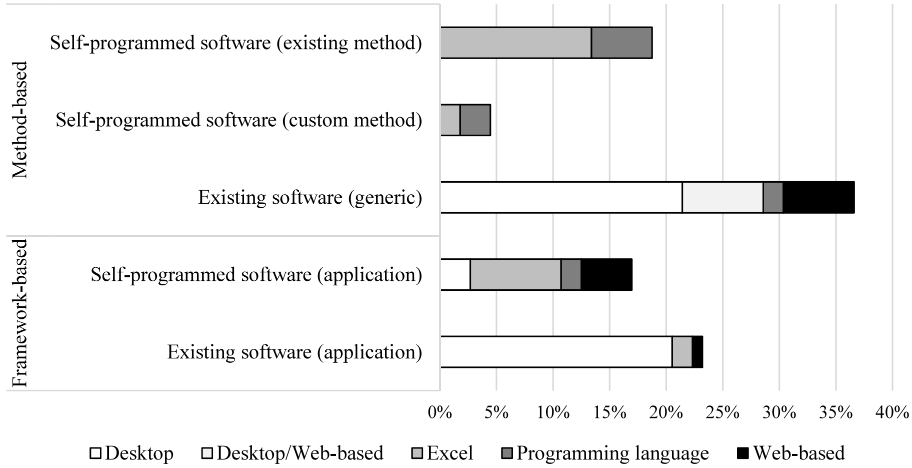

Fig. 3. Type of MADM software used in the reviewed articles (n = 112) categorized by the general motivations of the users (described in section 2.1.1.).

were referred to the literature, e.g. G11 Deal with compensation between input variables/criteria overlapped with *Compensation level between criteria* as suggested by the MADM taxonomy of Cinelli, Kadziński, Miebs, Gonzalez and Słowiński [26]. Conversely, in the absence of overlap, the criteria names were given by the authors of this study. For instance, the criterion *Customized weights* was assigned to G13 Includes pre-defined preferential information. The overlap process resulted in the identification of 29 criteria and nine domains (see ST 6). The refinement process resulted in the exclusion of the domain *Use case/context specific* and its related criteria. These criteria are out of the scope of this study (i.e. generic software) since they refer to the importance of providing ready-to-use decision models for sustainability assessment, i.e. framework with criteria, indicators, and weight sets. Conversely, six additional criteria were added to complement some of the remaining domains, e.g. criterion *Problem statement* to the domain *Problem typology*. This resulted in a total of 29 criteria distributed across eight domains (see ST 7).

The domains and criteria were then assigned a rationale for assessing the level to which MADM software fulfils them. For each criterion, three levels of fulfilment were defined, weak, moderate or strong. These were created based on the existing literature on the capabilities of MADM for SA [6,7,9,11–13,20] and the grouped motivations (see ST 6). The latter applies to criteria that are exclusively related to the motivations stated by the authors in the review, such as dynamic alternative/criteria management. For some criteria, only strong and weak fulfilment were defined. Table 2 shows the final set of domains, related criteria and rationale for assessing MADM software. The following paragraphs describe the domains (D) and criteria (C) resulting from the overlapping and refinement process.

3.1.2.1. D.1 Applicability and accessibility. The domain *applicability and accessibility* (D.1) focuses on the degree of complexity of access and use of the software, including the usability by those with different language backgrounds as well as customization requirements. The criteria included (and related groups) are *ease of use* (G10, G28, G41), *software customizability* (G9, G27), *interoperability* (G16, G4), and *availability and accessibility* (G26, G18, G19). In this domain, the criteria *personal information requirements* and *language inclusivity* were added as a proxy to specify the criterion *availability and accessibility*. The authors incorporated these two criteria to consider potential barriers to accessibility.

The definition of the criterion *ease of use* (C.1.1) in this study is “the ability of a user to successfully perform a task with the software without the need of training and/or the user manual”. The assessment of the software in this study is based on the premise that, in theory, the software should not be too challenging for researchers from different fields, provided that a support system is available. Therefore, it is considered whether the user interface includes a contextualized help system, i.e. specific information within the user interface referring to the task the user is engaged in. A contextual help system eases the user experience and flattens the learning curve associated with the software and the MADM method/process.

Software customizability (C.1.2) is relevant mainly for users who have the technical expertise to develop software and want to introduce new MADM methods, frameworks or simply modify existing methods. Open-source software provides a ready-to-use tool and the flexibility to copy, distribute, and/or modify the software, depending on the type of license, e.g. MIT license. Extending existing software may be a better way to disseminate MADM methods than developing new software.

Language inclusivity (C.1.3) becomes relevant when conducting MADM in contexts where involved stakeholders may not be proficient in the language used in the software and are expected to interact with it (e.g. Ref. [32]). A lack of language inclusivity could lead to misinterpretation of concepts or instructions when e.g. stakeholders or DMs are asked to provide preferences using an English-language interface. The criterion *language inclusivity* is illustrated by the application of Ssebunya, Schader, Baumgart et al. [33]. The publication reports that when using the SMART-farm tool (which is available in English only), users required additional preparation to avoid loss of information during interviews in local languages.

Personal information requirements (C.1.4), along with *Language inclusivity* (C1.3), serve to assess the accessibility of the software. While developers may be interested in monitoring the utilization of their software, prospective users may perceive requests for personal information/data as a barrier to accessing the software, even in the case of an email [28].

Interoperability (C.1.5) with other systems, such as Excel, enhances the efficiency of the process and facilitates the use of different platforms to complete different steps of the MADM process. For instance, a software is used for preference elicitation, and a different software is used for preference modelling [34]. Additionally, the data used in SA-MADM is derived from diverse methodologies, including Life Cycle Assessment (LCA), for which dedicated software is used.

3.1.2.2. D.2 Problem typology. The domain *problem typology* (D.2) defines the type and structure of the decision-making problem [26], i.e. *criteria structure* (G20), and *evaluation of alternatives on criteria* (G1, G5, G22). The criterion *problem statement* was added to this domain, as it is a relevant consideration for the selection of MADM methods irrespective of the context [35].

Problems in SA are characterized by a variety of *problem statements* (C2.1), i.e. ranking [36], sorting [37,38], clustering [39] and choice. The provision of a multifunctional software that can be utilized to

Fig. 4. Summary of groups of motivations and their frequencies in the reviewed publications.

Table 2
Assessment criteria per domain and respective rationale for the assessment ("-" indicates that the respective evaluation level was not considered for the respective criterion).

Domain	Criteria	Description	Strong (S)	Moderate (M)	Weak (W)
D1. Applicability and accessibility	**C1.1 Ease of use**	Degree of MADM knowledge required to use the software.	The help system is integrated as ‘Contextual Help’: This type of help provides information relevant to the task or feature the user is currently using.	Help systems are integrated into the software's user interface (UI). They can be accessed through menus, buttons, or shortcuts within the software.	Help systems are not integrated into the software's UI.
	C1.2 Software customizability	Type of permissions given to the user of the software, i.e. view, modify, and distribute.	Open-source license: The source code is freely available to the public.	Options for extending the software are available e.g. creation of plug-ins.	Non-free software license: no options available for extending/customizing software capabilities.
	C1.3 Language inclusivity	Capability to support several languages in the UI.	Multilingual support (including English).	Only English	Only other language (no English included).
	C1.4 Personal information requirements	Type of personal data required to get access to the software.	Software is free and publicly available online without restrictions.	Software is free and publicly available online with registration for full functionality.	Software is not available online, special request and verification needed.
	C1.5 Interoperability	Ability of the software to exchange information with e.g. external libraries, frameworks, or data sources.	Import and export formats available in Excel.	Only import or only export formats in Excel.	Not possible with Excel.
D2. Problem typology	**C2.1 Problem statement**	Ability of the software to support several decision recommendations, among which ranking, sorting, clustering, choice	Can support >2 decision recommendations.	Can support 2 decision recommendations.	Can support only 1 decision recommendation.
	C2.2 Criteria structure	Type of criteria structure supported by the software, i.e. flat and/or hierarchical.	Flat and hierarchical.	-	Only flat.
	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	Deterministic and uncertain.	-	Deterministic 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.	The software offers features to stimulate creative thinking in problem structuring e.g. brainstorming, objective definition (value-focused thinking, VFT)	-	Not available.
	C3.2 Dynamic alternatives management	Capability to manage dynamic and static sets of alternatives in the UI.	Alternatives can be added, deleted and selected deselected in the UI.	Alternatives can be added/deleted or selected/deselected in the UI.	Alternatives can only be added or deleted via file import. No selection/deselection available.
	C3.3 Dynamic criteria management	Capability of the user interface to manage dynamic and static sets of criteria.	Criteria can be added/deleted and selected /deselected in the UI.	Criteria can be added/deleted or selected/deselected in the UI.	Criteria can only be added or deleted via file import.No selection/deselection available.
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	Can support 3 types of aggregation.	Can support 2 types of aggregation.	Can support 1 type of aggregation.
	C4.2 Comparison of alternatives	Capability of the software to handle relative and absolute comparisons of alternatives performances.	Can support both relative and absolute comparisons: results are independent from new alternatives or deletion of existing ones.	-	Can support only relative comparisons: results are dependent on addition or deletion of alternatives.
	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.	At least one of the aggregation methods handles qualitative and quantitative or relative assessment.	-	Aggregation methods can only support qualitative or quantitative.
	C4.4 Compensation level between criteria	Capability to model situations of full, partial and null compensation between criteria.	Partial/null compensation can be modelled: No good performance on a criterion can compensate the poor performance on another criterion.	-	Only full compensation can be modelled: Good performance on a criterion can fully compensate the poor performance on another criterion
	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.	Software accounts for preference, indifference and other thresholds like veto.	Software accounts for preference and indifference thresholds	The software does not support methods with thresholds

(*continued on next page*)

Table 2 (*continued*)

Domain	Criteria	Description	Strong (S)	Moderate (M)	Weak (W)
	C4.6 Weights of criteria	Capability to support methods that include precise and imprecise weights of criteria.	Precise and imprecise weights	Precise weights	No weights can be used
	C4.7 Interactions between criteria	Capability to model interactions between criteria.	Positive/negative interactions.	-	Not possible
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	The software supports collaboration between stakeholders for problem structuring e.g. brainstorming (synchronous or asynchronous)	The software supports processing stakeholders' preferences for problem structuring e.g. brainstorming (asynchronous)	Does not support group settings.
	C5.2 Weights elicitation	Capability to support collecting and processing preference information elicited from different stakeholders from e.g. online surveys	The software facilitates data collection, enabling users to gather preference information synchronous or asynchronous.	The software facilitates processing preference information, enabling users to gather preference information asynchronous.	Does not support group preferences.
D6. Output variability analysis	**C6.1 Sensitivity analysis: input data**	Capability to carry out sensitivity analysis on weights and/or performance values	Performance values and weights	Performance values or weights	Not possible.
	C6.2 Sensitivity analysis: model	Capability to carry out sensitivity analysis on preference model, depending on the type of aggregation, e.g. normalization and aggregation function, discriminating thresholds.	Scoring: Normalization, aggregation function. Binary relations: preference function, discriminating thresholds (p,q). Decision rules: parameters of rules (consistency rules, thresholds)	-Normalization (if applicable) OR aggregation	Not possible.
	C6.3 Uncertainty: input data	Capability to carry out uncertainty analysis on weights and/or performance evaluations (single preference model)	Performance values and weights (e.g. uniform, normal, triangular probabilistic distribution).	Performance values or weights.	Not possible.
	C6.4 Uncertainty: model	Capability to carry out uncertainty analysis on multiple preference models	Different preferences of the DMs can be considered by accounting for different strategies to handle the data and to aggregate them (e.g. normalization methods, aggregation functions).	-	Not possible.
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	Documentation is available and regularly updated: manual version fits software version (same year) or documentation is max. 1 year older than the latest software version available.	Documentation is available but not updated, i.e. the available documentation is more than 1 year older than the latest software version available.	No documentation 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)	All applicable sections included.	Some sections missing.	No documentation available.
D8. Utility	**C8.1 Learning dimension**	Capability to acknowledge and accept new information revealed during the evolution of the procedure	Simultaneous comparison of the results with new information is possible (e.g. new alternatives)	-	No re-evaluation is possible and new software runs need to be performed and independently compared with previous ones
	C8.2 Interpretation of results	Capability to support interpretation of results with e. g. generate a report or summary with the main findings.	Features supporting interpretation of results, explaining the meaning of the figures.	Features showing the main results of the software. With this the user knows which are the main results to look at.	No software support provided
	C8.3 Graphical representation	Capability to represent results using figures and/or tables	Figures and tables	Only figures or tables	None, final script (list of alternatives)

manage different types of decision-making challenges may increase the likelihood of user software attachment [40]. In addition to that, it reduces the effort required by users to switch to a different software, i.e. the learning curve is reduced [22].

The *criteria structure* (C2.2) in SA is commonly guided by the concept of sustainability used e.g. triple-bottom line [41], SDGs [42]. Such concepts typically consider multiple dimensions and pillars, which are often represented through hierarchical structures. Moreover, the use of hierarchical structures can enhance the comprehension of decision models and the transparency of results interpretation (e.g. Refs. [7, 43–45]).

The *evaluation of alternatives on the criteria* (C2.3) in SA frequently involves dealing with uncertain, imprecise and even missing data [46]. The sources of uncertainty involved in MADM for SA are often associated with the input data, including imprecision of human judgement and poor data quality [6]. It is therefore important to use software that is capable of handling this type of information and that enables the user to use the data in its current form, for instance, using probability distributions.

3.1.2.3. D.3 Problem structure. The *problem structure* domain (D.3) includes criteria for assessing the options available to the user in defining a decision problem, i.e. constructing a model (G30), as well as to assessing the potential dynamism associated with its key components (G35), i.e. alternatives and criteria management. In this domain, the criterion *construction of models* was renamed as *problem structuring method (PSM)* to make the criterion more general, i.e. including not only decision trees but other techniques such as the strengths, weaknesses, opportunities, and threats (SWOT) analysis, and brainstorming.

Problem structuring methods (PSMs) (C3.1) are frequently used to support the identification of criteria and the development of alternatives in MADM [47]. Examples of PSMs in MADM for SA include value-focused thinking (VFT) [48], the Drivers, Pressures, State, Impacts, Responses (DPSIR) method [49], and the SWOT analysis [34].

In SA, problem structures can be static or dynamic, i.e. the set of alternatives and/or criteria can change depending on the decision problem. For example, a decision problem concerning land remediation strategies based on existing regulations involves static criteria [50]. In contrast, the assessment of chemical processes, which are known to be constantly evolving, require dynamic structures of alternatives and criteria [7]. Furthermore, there are methodologies like social-LCA for which indicators are still being developed [51]. In order to account for these important aspects, the criteria *dynamic alternatives management* and *dynamic criteria management* (C3.2 and C3.3) were used in the present assessment. These refer to the ability to add, delete, select or deselect criteria and alternatives using the user interface.

3.1.2.4. D.4 Preference model. The *preference model* domain (D.4) describes the characteristics of features that users would prefer to include within the model [26]. The criteria in this domain are: *type of aggregation of multiple criteria* evaluations (G29), *comparison of alternatives* (G3), *type of information available* (G21), *level of compensation between criteria* (G11), *weights of criteria* (G7, G37) and *interaction between criteria* (G14). A criterion was added, *per-criterion pairwise comparison thresholds*, given its importance in the literature of MADM capabilities for SA.

All *types of aggregation of multiple criteria evaluations* (C4.1), i.e. scoring functions, binary relations, or decision rules, are relevant for SA [6]. Scoring functions are useful in projects where having a final quantitative index is desired to assess the performance of the alternatives (e.g. Ref. [52]). Pairwise comparisons are useful in addressing problems where the objective is to sort alternatives into preference-ordered classes (e.g. Ref. [53]). Similarly, decision rules are effective when there is a need to link conditions of a particular decision problem to an outcome that may already be known based on measurements or expert judgement (e.g. Ref. [54]). As previously mentioned for the criterion *problem statement* (C2.1), software that handles multiple types of aggregation may increase the likelihood of user software attachment.

In terms of *comparison of alternatives* (C4.2), in some cases, practitioners aim to conduct an absolute sustainability assessment rather than a relative one [7]. This means that instead of comparing alternatives to each other, sustainability thresholds are used to reach a decision (e.g. certification) on the suitability or performance of a product, institution, or policy [55]. AHP and PROMETHEE methods work in relative terms, i.e. the result depends on the set of alternatives assessed. For ELECTRE methods, although most are relative, certain methods allow absolute assessments. For example, the sorting methods that work with class/-boundary profiles, do allow absolute assessments, as they do not depend on the alternatives in the set. Some practitioners use the figure of a "reference alternative" as a sort of threshold to conduct an absolute assessment [56]. Absolute SA can be conducted using methods such as TOPSIS, VIKOR, WSM and MAVT, using a normalization that refers to external references that ideally cover the full range of impacts or benefits [7].

Decision problems in SA are characterized by the inclusion of different types of information, with two overarching categories, qualitative and quantitative data. For example, quantitative data coming from LCA impact categories and qualitative information from social-LCA. The use of methods that can handle performance values in their original form (raw performances) support transparency [12]. For the assessment of the criterion *type of information* (C4.3), the scales used in the methods available in the software were used, i.e. relative, qualitative, quantitative.

The criterion *level of compensation between criteria* (C4.4) indicates whether different types of capital are substitutable in SA, i.e. manufactured, human, and natural capital [12,57]. Full compensation implies unlimited substitution between different types of capital (weak sustainability) [44]. For example, the economic benefits of a project (e.g. job creation) compensate for the environmental costs (e.g. loss of green space). Null or partial compensation indicate that substitution is limited or that non-substitution is allowed (strong sustainability) [58]. For example, chemicals/materials must meet both safety and environmental standards when a strong sustainability framing is used.

Per-criterion pairwise comparison thresholds (C4.5) can support SA in two ways. First, they allow to deal with the imperfect nature of knowledge (criteria selection and data) with discriminating thresholds (preference and indifference) [59]. For example, Wulf, Zapp, Schreiber et al. [60] present an approach to define preference and indifference thresholds based on uncertainty of Life Cycle Impact Assessment methods. Second, they enable DMs or stakeholders to set a condition under which an alternative is unacceptable regardless of its performance on other criteria, using veto thresholds. For example, when comparing energy technologies, setting a veto on greenhouse gas emissions would mean that an alternative would not be better than others regardless of its efficient technical performance and low cost.

Weights of criteria (C4.6) represent a powerful tool for incorporating societal values and interests in SA. These help to represent the preferences and priorities of the stakeholders involved in the decision-making process. Weights can be of two types: precise or imprecise. Depending on their meaning, precise weights can be trade-offs or importance coefficients [61]. In the context of sustainability, several authors highlight the importance of weights as importance coefficients which is directly related with null or partial compensation methods [6,12]. Imprecise weights are often used in situations where DMs/stakeholders preferences are incomplete, dynamic, or missing, for example when DMs refrain from revealing preference information to the public opinion [62]. Furthermore, the use of imprecise weights in group decision-making settings reduces the pressure on stakeholders to set a definitive weight value, instead focuses on reaching consensus on the order of priorities.

In the context of SA, the considerations that are involved in the construction of criteria usually interact with one another, reflecting the inherent dynamics of environmental and social systems [63]. *Interactions between criteria* (C4.7) are useful to model not only synergies but also redundancies between pairs of criteria. Despite exercising due care in the selection of criteria, there might be a pair of criteria for which interactions are difficult to avoid. This is exemplified by the case of LCA indicators [64,65], such as eutrophication and acidification potential, which are linked by NOx emissions. Another example is the case of understanding the interactions between the SDGs and how this can be vital for the design of appropriate policies that integrate various sectors, such as food production and land degradation [66]. Figueira, Greco and Roy [67] propose a categorization of interactions between criteria in ELECTRE methods, which includes mutual strengthening, mutual weakening, and antagonistic interactions.

3.1.2.5. D.5 Stakeholders involvement. The domain *stakeholders involvement* (D.5) has an important impact on the acceptance in different decision-making processes [13]. The dynamic nature of sustainability and the need for digital formats to reach a wider audience of participants require tools that facilitate the stakeholders' interactions with MADM models. The involvement of stakeholders in MADM can occur at various stages of the process, including i.e., problem structuring, criteria weighting, scoring, and results analysis. This domain includes criteria that assess the capability of the software to involve stakeholders into *problem structuring (groups)* (G15) and *weights elicitation* (G24, G8, G38).

In SA, as in any context where there may be conflicting opinions and preferences, *problem structuring involving groups* (C5.1) becomes more challenging than in the case of a single DM. This criterion assesses whether the software includes modules or features to support problem structuring in group settings, such as online group sessions.

Preference elicitation in group settings has two main considerations, i) how the preference data (i.e. weights) is collected in the software and ii) how they are aggregated. In this study, the criterion *weights elicitation* (C5.2) only considers the first aspect, assessing whether the software handles preferences of groups and the inclusion of a preference data collection module, i.e. manual data entry is required by the software user vs automatic input via e.g. online surveys.

3.1.2.6. D.6 Output variability analysis (OVA). The domain *output variability analysis* (OVA) (D.6) considers the software functionality to study the stability and robustness of the provided decision recommendation [20]. For this, the criteria *sensitivity analysis* (G25) and *uncertainty analysis* (G33) are used. They can be studied either at the level of input data (i.e. performance values, weights, and thresholds) or at the level of the preference model (i.e. normalization methods and aggregation functions).

SA-MADM frequently involves uncertainties related to the input data, including poor data quality and inaccuracy of human judgement [6]. The criterion *sensitivity analysis of input data* (C6.1) enables to derive conclusions regarding the influence of variations in the performance values and/or weights (one at a time) on the output of the model [20]. The examination of these variations support decision-makers to prioritize interventions, manage uncertainty, and enhance the adaptation of sustainability strategies.

The considerations for *sensitivity analysis of the preference model* (C6.2) depend on the type of aggregation, namely scoring function, binary relations, or decision rules. For example, for scoring functions, assessing OVA may involve variations of the normalization method and aggregation function [20]. For binary relations, it may involve changing the preference functions (e.g. PROMETHEE) and/or per-criterion pairwise comparison thresholds.

The criterion *uncertainty analysis of input data* (C6.3) assesses the capability to propagate the uncertainty of the inputs through the entire process, and evaluate the effect on the outcome. The importance of understanding, characterizing and propagating uncertainty is to provide decision-makers with insights into the likelihood that the uncertainty will lead to the selection of a different option [68].

The importance of the criterion *uncertainty analysis of preference model* (C6.4) relies on giving the analyst the possibility to include different preferences of the decision-maker on how to model the decision problem e.g. normalization and aggregation [20].

3.1.2.7. D.7 Transparency. The domain *transparency* (D.7) includes criteria that assess the comprehensiveness, availability and accessibility of information pertinent to the use of the software, including user manuals and dedicated publications. While this might not be relevant for some experienced users [22], it could be crucial for novice users, such as some practitioners of SA-MADM. Consequently, the criteria *traceability of documentation* and *transparency of documentation* were added to assess transparency (G32).

The criterion *traceability of documentation* (C7.1) entails an evaluation of the accessibility and timeliness of the written material that provides information about the development, use, and functionality of the software. Such documentation may take the form of user manuals, published articles or dedicated websites. A lack of accessible, timely documentation may discourage users from using the software or result in its inappropriate use.

The criterion *transparency of documentation* (C7.2) assesses the extent to which the available written material covers the relevant steps of the decision-making process, namely: i) problem structuring, ii) preference model, iii) output variability analysis, and iv) visual representation. Incomplete documentation may impede the use of the software.

3.1.2.8. D.8 Utility. The domain *utility* (D.8) is concerned with the manner in which the software supports the user's comprehension, comparison and interpretation of results. Three criteria belong to this domain: *learning dimension* (G40), *interpretation of results* (G17), and *graphical representation* (G2, G36).

Sustainability issues are continuously evolving, with new information on climate change, air and water quality, and societal stability emerging on a daily basis. MADM models are sensitive not only to these changing theories, but also to different assumptions made by the analyst and/or experts to tackle a specific decision problem. The criterion *learning dimension* (C8.1) assesses the ability of the software to incorporate new information in the model and compare different results simultaneously [9]. In theory, all software could do this by running the model, saving the results in e.g. excel format and comparing. However, this lack of flexibility results in a time-consuming process when the models are in constant change.

The *interpretation of results* (C8.2) is a very important step to support the decision-making process effectively. This requires a good understanding of the MADM methods and the input data. Therefore, such software feature is not intended to replace scientific analysis, but to provide a good starting point for understanding the results obtained from an MADM method. With such feature, novice MADM practitioners can rather learn how the software connects different pieces of information in the model and which results to focus on. For example, a summary of results can indicate the preferred alternative and how stable the ranking is. Similarly, interactions between criteria can be pointed out or the level of consensus in cases where different stakeholders are involved.

An effective visualization of results could make the difference while communicating MADM outcomes, especially in cases where stakeholders with different backgrounds are involved e.g. scientists and lay citizens. *Graphical representation* (C8.3) could lead to a better communication and understanding of the results (in comparison with tabular results).

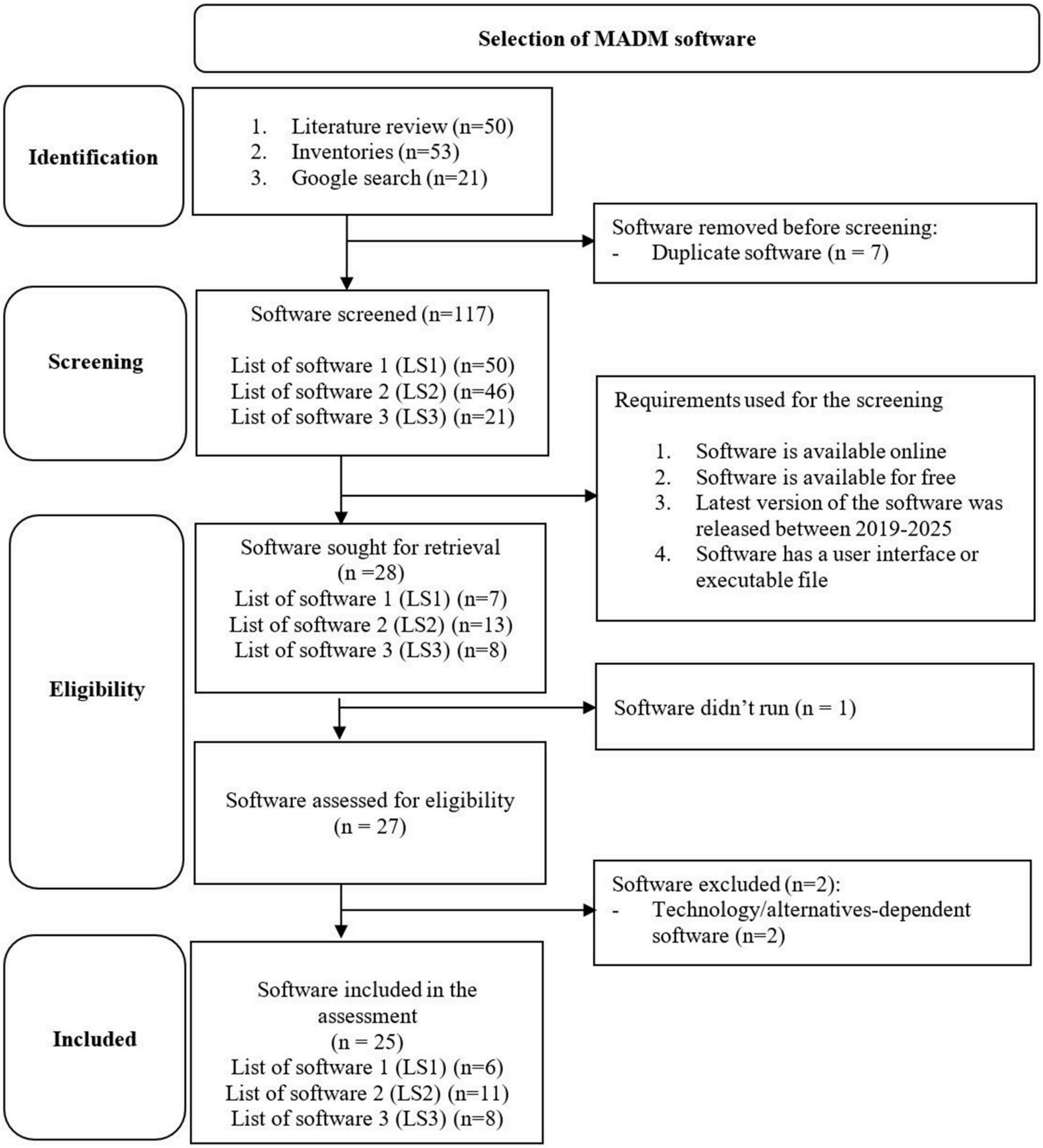

Fig. 5. Process of selecting a sample of MADM software for the assessment (Stage 2).

3.2. *Selection of MADM software sample*

In total, 117 MADM software tools were identified and characterized as described in section 2.2. Full characterization is available in ST 8–10. Fig. 5 summarizes the screening and eligibility process that resulted in a sample of 25 MADM software.

Table 3 presents the information of the MADM software selected for the assessment. This software sample matched the inclusion criteria: (i) free, (ii) with active versions released after 2019, and (iii) with a user interface.

3.3. *Assessment and recommendations*

3.3.1. *Assessment of selected software*

The 25 MADM software in Table 3 were assessed against the criteria (and respective rationales) presented in Table 2 (section 3.1.2). Table 4 presents the assessment results of each software (further information available in ST 12). The row *Software: cumulative assessment* presents the absolute frequencies of strong (S), moderate (M), and weak (W) fulfilment for each software. It can be observed that none of the software tools achieve a strong (S) fulfilment across the complete set of criteria. However, every software has distinct strengths from which users can benefit. Similarly, the column *Criteria: cumulative assessment* presents the relative frequencies (%) of the overall sample of 25 software for each criterion. This analysis elucidates the criteria that were found to be commonly addressed by the sample of MADM software. These include the acceptance of different *type of information* (C4.3), the execution of absolute and relative *comparisons of alternatives* (C4.2), and the *traceability and transparency of documentation* (C7.1 and C7.2). However, it also shows the criteria that have not been taken into consideration when developing MADM software and that are very important in the context of the transformation of energy systems. For instance, supporting

Table 3
Selected MADM software for assessment (more information available in ST 11).

	Software	Methods available	Platform	License	Website
1	**AHP-OS** [69]	AHP	Web-based	Open-source	https://bpmsg.com/ahp/ahp.php
2	**Apollo-Live** [70]	TOPSIS	Web-based	Freeware	https://apollo-live.epu.ntua.gr/
3	**Decision Radar**[a]	TOPSIS, SAW, ELECTRE	Web-based	Freeware	decision-radar.com
4	**DecSpace** [71]	Weighted sum, Categorization by Similarity-Dissimilarity (CAT-SD), Deck of Cards Method-Simos-Roy-Figueira (DCM-SRF), ELECTRE TRI-nC	Desktop	Freeware	http://decspace.sysresearch.org/content/homepage/about.html
5	**DEXiWin** [72]	DEX	Desktop	Open-source	https://dex.ijs.si/
6	**Decision Master** [73]	SMART, WASPAS, Taxonomy, REGIME, PROMETHEE, ELECTREE	Desktop	Open-source	https://github.com/BSTU/decisionmaster/tree/master
7	**Entscheidungs-navi** [74]	MAVT	Web-based	Open-source	https://entscheidungsnavi.de/
8	**EWMS** [75]	Entropy weight sum method	Desktop	Open-source	https://github.com/yinyixing/EWCProject
9	**FITradeoff**[a] [76]	FITradeoff	Web-based	Freeware	https://fitradeoff.org/
10	**HELDA**[a] [77]	AHP, SMART, SWING, DCM-SRF, TOPSIS, VIKOR, Weighted sum, PROMETHEE I and II, ELECTRE III	Desktop	Freeware	https://www.mcda-helmholtz.de/64.php
11	**J-Electre** [78]	ELECTRE I, ELECTRE I_S, ELECTRE I_V, ELECTRE II, ELECTRE III, ELECTRE IV, ELECTRE TRI AND ELECTRE TRI ME.	Desktop	Open-source	https://j-electre.sourceforge.io/
12	**Logical Decisions for Windows (LDW)**[a] [79]	AHP, SMART	Desktop	Freeware	http://www.logicaldecisionsshop.com/catalog/
13	**MakeDecision.it** [80]	ARAS, COCOSO, CODAS, COMET, COP RAS, EDAS, MABAC, MAIRCA, MARCOS, MOORA, OCRA, PROMETHEE II, SPOTIS, TOPSIS, VIKOR	Web-based	Open-source	https://make-decision.it/
14	**MAMCA** [81]	Hybrid weight elicitation method, SMART, AHP	Web-based	Freeware	https://www.mamca.eu/
15	**MCDA Calculator** [22]	MAVT, PROMETHEE II, SMART/SMARTS/SMARTER, TOPSIS, VIKOR	Web-based	Freeware	https://MADM-calculator.psi.ch/calculator
16	**MCDA Index Tool** [20]	Weighted aggregations, including multiple normalization functions	Web-based	Freeware	https://www.psi.ch/en/ta/mcdaindextool
17	**MCDA-ULaval** [82]	ELECTRE II, ELECTRE III, ELECTRE TRI	Desktop	Freeware	https://MADM.fsa.ULaval.ca/download/
18	**MCDMaker**[a]	AHP, ARAS, CILOS, COBRA, COCOSO, CODAS, COPRAS, CRITIC, DEMATEL, EAMR, EDAS, ELECTRE, ENTROPY, EXPROM, GRA, MABAC, MAIRCA, MARCOS, MAUT, MEREC, MOORA, OCRA, PROMETHEE, SPOTIS, TOPSIS, VIKOR, WSM, and more	Web-based	Freeware	https://mcdmaker-software.web.app/#home
19	**PROMETHEE-Cloud** [83]	PROMETHEE I, PROMETHEE II	Web-based	Freeware	https://promethee.pom.uni-due.de/
20	**RuLeStudio**[a] [84]	DRSA and VC-DRSA	Client-server	Open-source	https://sites.google.com/view/mszelag/software/RuLeStudio
21	**SilverDecisions** [85]	Decision trees	Web-based	Open-source	https://silverdecisions.pl/
22	**SOCRATES**[a] [86]	SMCE	Web-based	Freeware	https://web.jrc.ec.europa.eu/socrates/screen/home
23	**Super Decisions**[a] [87]	AHP, ANP	Desktop	Freeware	https://SuperDecisions.com/
24	**ValueDecisions** [88]	MAVT	Web-based	Open-source	https://eawag.shinyapps.io/ValueDecisions/
25	**Web-HIPRE 2.0**[a] [89]	MAVT, AHP, SMART, SWING, SMARTER	Web-based	Freeware	https://hipre.aalto.fi/

[a] Indicates software for which a dedicated publication is not available and, if available, an article related to the methodology used in the software is cited as reference.

problem structuring methods (C3.1) and *stakeholders' involvement* (C5.1 and C5.2). The following paragraphs present and discuss the assessment results of the software sample for every domain and criterion for SA.

3.3.1.1. D.1 Applicability and accessibility. C1.1 Ease of use: Only 32 % of the assessed software are strong in this criterion, incorporating a contextual help system within the user interface: Entscheidungsnavi, FITradeoff, HELDA, LDW, MAMCA, MCDA Index Tool, RuleStudio and SOCRATES. A further 8 % of the software has a moderate level of fulfilment, with help systems that can be accessed only via menus, buttons, or shortcuts. A significant proportion (60 %) of the software is weak in this respect as they do not include a contextualized help system. These either have a help button that refers to the online user manual, or do not include a help system at all.

C1.2 Software customizability: 48 % of the assessed software has strong fulfilment of this criterion by providing an open-source license. Conversely, 48 % exhibited weak fulfilment; users can not extend or customize the software's functionality. Only HELDA software (4 %) exhibited moderate fulfilment by allowing users to extend functionality with plug-ins.

C1.3 Language inclusivity: English is the dominant language of the user interfaces of the reviewed software. 72 % of the software exhibits moderate fulfilment, with an English-only user interface. 28 % of the software has strong fulfilment in this criterion including additional languages. These tend to be the native languages of the developers or the regions where they are based. This suggests that developers prefer to use software in their native language or the language of local entities in their respective contexts.

C1.4 Personal information requirements: The majority of the software (64 %) is strong in this criterion; it is available without any kind of registration or access requirements. Moderate fulfilment is assigned to the remaining 32 % of the software due to the requirement for user registration or account creation on the software website in order to obtain full access. MAMCA has weak fulfilment (4 %), it is provided to users upon request via email and developers verification of the application.

C1.5 Interoperability: The majority of the assessed software is interoperable with Excel. However, only 56 % of the software exhibits strong fulfilment, allowing both the import and export of data. Meanwhile, 28 % of the software exhibits moderate fulfilment, supporting either the import or export of data. The remaining 16 % of software is weak in this criterion, either not allowing the import/export of data or only allowing it in formats different to Excel (e.g. JSON, JPEG).

3.3.1.2. D.2 Problem typology. C2.1 Problem statement: The majority of the software tools (84 %) are weak in this criterion, as they only support one type of decision recommendation. Only 8 % of the software tools are

Table 4
Overview of the fulfilment of capabilities, according to the eight domains and respective criteria, of the assessed sample of MADM software for supporting sustainability assessment: Strong ●, moderate ◓, and weak ○.

	Criteria	AHP-OS	Apollo-Live	Decision Radar	DecSpace	DEXiWin	Decision Master	Entscheidungsnavi	EWMS	FITradeoff	HELDA	J-Electre	LDW	MakeDecision.it	MAMCA	MCDA Calculator	MCDA Index Tool	MCDA-ULaval	MCDMaker	PROMETHEE-Cloud	RuLeStudio	SilverDecisions	SOCRATES	Super Decisions	ValueDecisions	Web-HIPRE 2.0	Criteria: cumulative assessment % S ●	M ◓	W ○
1. Applicability and accessibility	1.1 Ease of use	○	○	○	○	○	○	●	○	●	●	○	●	○	●	○	●	○	○	○	●	○	●	◓	○	◓	32	8	60
	1.2 Software customizability	●	●	○	○	●	●	●	●	○	◓	●	●	●	○	○	○	○	○	○	●	●	○	○	●	○	48	4	48
	1.3 Language inclusivity	●	◓	◓	◓	◓	◓	●	◓	●	●	◓	◓	●	◓	◓	◓	●	◓	◓	◓	●	◓	◓	◓	◓	28	72	0
	1.4 Personal information requirements	◓	◓	●	●	●	●	●	●	◓	◓	●	◓	●	○	●	●	◓	●	◓	●	●	●	◓	●	●	64	32	4
	1.5 Interoperability	◓	●	○	◓	●	◓	●	●	●	●	○	●	●	●	●	●	●	◓	○	●	◓	◓	◓	●	○	56	28	16
2. Problem typology	2.1 Problem statement	○	○	○	◓	○	○	○	○	●	○	●	○	○	○	○	○	◓	○	○	○	○	○	○	○	○	8	8	84
	2.2 Criteria structure	●	○	○	○	●	○	●	●	○	●	○	●	○	●	○	○	○	○	○	○	●	●	●	●	●	48	0	52
	2.3 Evaluation of alternatives on the criteria	○	○	○	○	●	●	●	○	○	●	○	●	●	○	○	○	○	●	○	○	●	●	○	●	○	40	0	60
3. Problem structure	3.1 Problem structuring methods	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	8	0	92
	3.2 Dynamic alternatives management	○	◓	◓	◓	●	◓	●	◓	◓	●	◓	◓	◓	◓	◓	◓	●	◓	◓	●	◓	◓	◓	◓	◓	20	76	4
	3.3 Dynamic criteria management	○	◓	◓	◓	◓	◓	●	◓	◓	◓	◓	◓	◓	◓	◓	◓	●	◓	◓	●	◓	◓	◓	○	◓	12	80	8
4. Preference model	4.1 Type of aggregation of multiple criteria evaluations	○	○	◓	◓	○	◓	○	○	○	◓	○	○	◓	○	◓	○	○	◓	○	○	○	○	○	○	○	0	28	72
	4.2 Comparison of alternatives	○	○	●	○	●	○	●	●	●	●	●	●	●	●	●	●	●	●	○	●	●	○	●	●	●	76	0	24
	4.3 Type of information (scale)	●	●	●	●	●	●	●	○	●	●	●	●	●	●	●	○	●	●	●	●	●	●	●	●	●	92	0	8
	4.4 Compensation level	○	○	●	●	●	●	○	○	●	●	●	○	●	○	●	●	●	●	●	●	●	●	○	○	○	64	0	36

	Criteria	AHP-OS	Apollo-Live	Decision Radar	DecSpace	DEXiWin	Decision Master	Entscheidungsnavi	EWMS	FITradeoff	HELDA	J-Electre	LDW	MakeDecision.it	MAMCA	MCDA Calculator	MCDA Index Tool	MCDA-ULaval	MCDMaker	PROMETHEE-Cloud	RuLeStudio	SilverDecisions	SOCRATES	Super Decisions	ValueDecisions	Web-HIPRE 2.0	Criteria: cumulative assessment % S ●	M ◓	W ○
	between criteria																												
	4.5 Per-criterion pairwise comparison thresholds	○	○	○	●	○	◓	○	○	◓	●	●	○	◓	○	●	○	●	◓	◓	○	○	●	○	○	○	24	20	56
	4.6 Weights of criteria	◓	◓	◓	◓	◓	◓	●	○	●	●	◓	●	●	◓	◓	◓	◓	●	◓	○	◓	◓	◓	◓	◓	24	68	8
	4.7 Interactions between criteria	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	4	0	96
5. Stakeholders involvement	5.1 Problem structuring (groups)	○	○	○	○	○	○	●	○	○	●	○	○	○	●	○	○	○	○	○	○	○	◓	○	○	○	12	4	84
	5.2 Weights elicitation	●	●	○	○	○	○	●	○	○	●	○	○	○	●	○	○	○	○	○	○	○	◓	○	◓	○	20	8	72
6. Output variability analysis	6.1 Sensitivity analysis: input data	◓	○	○	○	◓	○	●	○	◓	●	○	◓	○	◓	○	○	◓	○	◓	○	◓	◓	◓	◓	◓	8	48	44
	6.2 Sensitivity analysis: model	◓	○	○	○	○	○	○	○	○	○	○	○	●	○	○	●	○	○	○	○	○	○	○	◓	○	8	8	84
	6.3 Uncertainty: input data	○	○	○	○	●	○	●	○	◓	●	○	●	○	○	○	○	◓	○	●	○	●	◓	○	◓	○	24	16	60
	6.4 Uncertainty: model	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	4	0	96
7. Transparency	7.1 Traceability of documentation	●	●	○	●	●	●	●	●	●	●	●	◓	●	●	●	●	●	○	●	●	●	◓	◓	◓	◓	72	20	8
	7.2 Transparency of documentation	●	◓	○	◓	●	◓	●	◓	●	●	●	●	◓	●	◓	●	●	○	●	●	●	●	●	●	●	68	24	8
8. Utility	8.1 Learningdimension	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	●	○	○	○	○	○	○	○	○	8	0	92
	8.2 Interpretation of results	◓	●	○	○	◓	○	●	○	○	●	○	○	○	◓	○	◓	○	◓	○	●	○	○	◓	●	○	20	24	56
	8.3 Graphical representation	◓	◓	◓	◓	●	◓	●	●	●	●	●	●	◓	●	◓	●	●	◓	●	◓	●	●	●	●	●	64	36	0
	S ●	7	6	4	5	13	6	21	7	11	19	10	12	12	10	7	10	13	6	6	12	12	9	5	10	6			
	M ◓	7	7	6	9	5	9	0	4	6	4	4	6	6	6	7	5	5	8	7	2	5	10	10	8	7			
	W ○	15	16	19	15	11	14	8	18	12	6	15	11	11	13	15	14	11	15	16	15	12	10	14	11	16			

strong, supporting more than two decision recommendations: J-Electre and FITradeoff. The remaining 8 % are moderate, as they include two decision recommendations: MCDA ULaval and DecSpace. In real-world applications, this can be problematic for dynamic decision-making challenges that require different decision recommendations for the same problem e.g., sorting energy technologies and then scoring those in the “good” category.

C2.2 Criteria structure: Only 48 % of the software tools are strong in this criterion, supporting both flat and hierarchical criteria structures. These include AHP-OS, DEXiWin, Entscheidungsnavi, LDW, EWMS, HELDA, MAMCA, SilverDecisions, SOCRATES, Super Decisions, ValueDecisions, and Web-HIPRE 2.0. From these, SOCRATES is the only software handling hierarchies in binary-relation-type aggregation methods. HELDA handles hierarchies only with scoring methods (see criterion C4.1). The remaining software tools (52 %) are weak supporting only flat criteria structures.

C2.3 Evaluation of alternatives on the criteria: Only 40 % of the software tools are strong in this criterion allowing both uncertain and deterministic evaluations of the alternatives on the criteria. For modelling uncertain data, some software tools use probabilistic functions: DEXiWin, DSS DecisionMaster, Entscheidungsnavi, HELDA, LDW, SilverDecisions, SOCRATES and ValueDecisions. Others use fuzzy sets: DEXiWin, MCDMaker and MakeDecisions.it. The remaining software tools (60 %) are weak, as they handle deterministic or uncertain evaluations only. From this, Apollo-Live is the only one handling only uncertain evaluations in the form of fuzzy sets.

3.3.1.3. D.3 Problem structuring. C3.1 Problem structuring methods: Only 8 % of the software tools are strong in this criterion, including a module to support PSMs. The software Entscheidungsnavi employs a value-focused thinking approach to guide the user through the process of decision problem definition. The software LDW includes a module for brainstorming. The majority of software tools (92 %) are weak, they lack a module to support PSMs. These findings align with the observations made by Mustajoki and Marttunen [16], who argue that the absence of PSMs in the software may be attributed, at least in part, to a lack of awareness regarding its importance in the MADM process.

C3.2 Dynamic alternatives management: Only five software tools (20 %) are strong in this criterion: DEXiWin, Entscheidungsnavi, HELDA, MCDA ULaval, and RuLeStudio. From these, MCDA ULaval and RuLeStudio allow full flexibility in the management of alternatives. This facilitates the execution of the model and the visualization of results exclusively for the selected alternatives. The remaining software under discussion allow selection and deselection of alternatives exclusively in the results section for specific analysis, e.g. ranking. 76 % of the software tools allow direct addition or deletion of alternatives in the user interface. In ValueDecisions it is only possible to import alternatives from Excel files. However, it allows selection and deselection of alternatives in the results section. Only AHP-OS is weak (4 %), offering limited options for managing alternatives.

C3.3 Dynamic criteria management: Only three software tools (12 %), Entscheidungsnavi, MCDA ULaval and RuLeStudio, perform strong while most of the software perform moderate (80 %). Most of the software tools have the same capabilities as for criterion C3.2 with a few exceptions. HELDA and DEXiWin have moderate fulfilment since criteria can be added and deleted within the user interface but there is no option for selection or deselection. Only AHP-OS and ValueDecisions perform weak (8 %) with limited options to handle criteria once a model is created.

3.3.1.4. D.4 Preference model. C4.1 Type of aggregation of multiple criteria evaluations: None of the assessed software tools provide support for all three types of aggregation. Only 28 % of the software tools exhibit moderate fulfilment by supporting binary relations and scoring functions. These are DecisionMaster, DecisionRadar, DecSpace, HELDA, MakeDecision.it, MCDMaker and the MCDA calculator. The remaining 72 % of the software tools are weak, supporting only one type of aggregation. Of these, the majority (44 %) corresponds to scoring functions (AHP OS, Apollo-Live, Entscheidungsnavi, EWMS, FITradeoff, ValueDecisions, LDW, Super Decisions, MAMCA, MCDA Index Tool and WebHIPRE 2.0), 16 % to binary relations (J-Electre, PROMETHEE-Cloud, SOCRATES, MCDA-ULaval), and 12 % to aggregation with decision rules (DexiWin, SilverDecisions, RuleStudio). Binary relations and decision rules-based methods are appropriate to build models with null or partial compensation between criteria. For scoring functions, the capacity for this purpose is dependent on e.g. the aggregation functions used as considered in the MCDA Index Tool and in ValueDecisions.

C4.2 Comparison of alternatives: 76 % of the software assessed is strong in this criterion, providing methods to support both relative and absolute comparisons. In contrast, 24 % of the software tools are weak, offering methods that support only relative comparisons.

C4.3 Type of information (scale): Most of the assessed software tools are strong in this criterion (92 %), being capable of handling qualitative and quantitative or relative evaluations of criteria. Some software tools offer more functions for qualitative scales. For example, SOCRATES includes predefined ordinal and linguistic scales, whereas J-ELECTRE uses only ordinal scales. Entscheidungsnavi and HELDA allow users to customize their own scales, including numerical, verbal and constructed scales. AHP-OS, FITradeoff, and Super Decisions handle only relative scales. MCDA Index Tool and EWMS perform weak since they handle quantitative criteria only (8 %).

C4.4 Compensation level between criteria: 64 % of the software tools are strong in this criterion by including aggregation methods that handle null/partial compensation between criteria. Of these, some include methods with full compensation as well: DecisionMaster, FITradeoff, HELDA, MCDMaker, MakeDecision.it, DEXiWin, MCDA Index Tool, and MCDA Calculator. In contrast, 36 % of the software tools are weak, supporting only full compensation methods: AHP-OS, Apollo-Live, Decision Radar, Decision Master, Entscheidungsnavi, EWMS, MAMCA, Web-HIPRE 2.0, and ValueDecisions.

C4.5 Per-criterion pairwise comparison thresholds: Only 24 % of the software tools are strong, giving users the flexibility to use three different thresholds in the preference model, i.e. preference, indifference and veto: DecSpace, HELDA, J-Electre, MCDA Calculator, MCDA ULaval and SOCRATES. 20 % exhibit moderate fulfilment, including two thresholds, mainly preference and indifference. An exception is the FITradeoff software which accepts indifference and veto thresholds. 56 % of the assessed software tools are weak in this criterion as their supported methods lack of discriminating thresholds.

C4.6 Weights of criteria: Only 24 % of the software tools are strong in this criterion supporting both precise and imprecise weights: Entscheidungsnavi, FITradeoff, HELDA, LDW, MakeDecision.it, and MCDMaker. 68 % exhibit moderate fulfilment, supporting precise weights only. Only two software tools are weak in this criterion (8 %), by supporting MADM methods that do not use weights. The first one is RuLeStudio, that uses decision rules. The second one is EWMS, which uses entropy weights that are a dispersion measure of the data and not the preference-based type of information expected in SA [75].

C4.7 Interactions between criteria: Only one software exhibits strong fulfilment. The MCDA ULaval software allows modelling interactions between criteria in the forms of mutual strengthening, mutual weakening and antagonistic effects. The remaining software tools (96 %) are weak in this criterion without options to model interactions between criteria.

3.3.1.5. D.5 Stakeholders involvement. C5.1 Problem structuring (groups): Entscheidungsnavi, HELDA and MAMCA (12 %) are strong in this criterion. HELDA and MAMCA offer online, survey-based approaches to elicit preferences of stakeholders in real-time, to make suggestions to add or delete criteria to the decision model. Similarly, Entscheidungsnavi offers a module for asynchronous participation based on the VFT approach. SOCRATES has a moderate performance (4 %), it involves a module called “Equity analysis” where suggested alternatives by the actors can be added to the evaluation matrix by the software user. All other software tools are weak (84 %) without any option to support group settings.

C5.2 Weights elicitation: Only 20 % of the software provide features to collect and process criteria weights from groups: AHP-OS, Apollo-Live, Entscheidungsnavi, HELDA and MAMCA. AHP-OS, Apollo-Live, HELDA and MAMCA allow the creation of online surveys. Of these, only HELDA can support binary relations methods. The remaining support the aggregation method of scoring functions. SOCRATES and ValueDecisions have a moderate fulfilment (8 %) of this criterion, by processing different preferences from stakeholders manually entered by the user. The remaining 72 % of assessed software tools are weak, without features to support weights elicitation of groups.

3.3.1.6. D.6 Output variability analysis (OVA). C6.1 Sensitivity analysis - input data: Only two software (8 %), Entscheidungsnavi and HELDA, are strong in this criterion by allowing sensitivity analysis of both weights and performance values. 48 % of the software tools exhibit moderate fulfilment, being only able to perform one of these, having most often sensitivity analysis of weights: AHP-OS, FITradeoff, MAMCA, MCDA ULaval, PROMETHEE-Cloud, SilverDecisions, SOCRATES, Super Decisions, Web-HIPRE, and ValueDecisions. The remaining 44 % of the software tools are weak, as this functionality is not available.

C6.2 Sensitivity analysis - model: The capabilities of the sample of software for this criterion are overall weak (84 %). Only four software tools (16 %) are strong in this criterion, allowing sensitivity analysis of the preference model for the same data. For scoring functions, the MCDA Index Tool and ValueDecisions allow sensitivity analysis of the aggregation function. Additionally, the MCDA Index Tool allows the user to vary the normalization method. For binary relations, both HELDA and MCDA ULaval allow sensitivity analysis of the per-criterion pairwise comparison thresholds. None of the software that handles decision rules allows sensitivity analysis of the preference model.

C6.3 Uncertainty - input data: Of the assessed software, 24 % are strong since they allow modelling uncertainty in both weights and performances using mostly Monte Carlo simulation: DEXiWin, Entscheidungsnavi, HELDA, LDW, PROMETHEE-Cloud, and SilverDecisions. Meanwhile 16 % have moderate fulfilment allowing to model only one of them. 60 % of the software are weak without capabilities to conduct uncertainty analysis.

*C*6.4 Uncertainty - model: MCDA Index Tool is the only software that has strong fulfilment in this criterion. The rest of the software are weak (96 %).

3.3.1.7. D.7 Transparency. C7.1 Traceability of documentation: A total of 72 % of the software are strong, providing both accessible and current documentation, indicating that the versions of the software and the documentation are the same. A total of 20 % of the software have moderate fulfilment because documentation is outdated, i.e. a release date for the documentation that is more than one year older than the latest version of the software. Only two software tools (8 %), DecisionRadar and MCDMaker, are weak in this criterion because supporting documentation was not found.

C7.2 Transparency of documentation: 68 % of the assessed software tools are strong in this criterion providing documentation that included complete information from relevant MADM sections (where applicable). In contrast, 24 % of the software assessed have moderate fulfilment because there was missing or incomplete information in the documentation. As stated above, for DecisionRadar and MCDMaker (8 %) supporting documentation was not found.

3.3.1.8. D.8 Utility. C8.1 Learning dimension: Only two software tools (8 %) are strong in this criterion. MCDA ULaval and MakeDecision.it allow comparing results obtained with different inputs simultaneously. The remaining software tools are weak (92 %) because it requires the analysis to be rerun and the results stored in order to make comparisons.

C8.2 Interpretation of results: Only few software tools (20 %) are strong including features to support interpretation of results, i.e. a reporting section summarizing, and in some cases explaining, input and output data. These include Apollo-Live, Entscheidungsnavi, HELDA, ValueDecisions, and RuleStudio. 24 % of the software have moderate fulfilment with features pointing out the main results of the software and 56 % are weak lacking features to support interpretation.

C8.3 Graphical representation: 64 % of the software are strong including both graphical and tabular visualization of results, whereas the remaining 36 % are moderate including either tabular or graphical only.

According to the results, the software ranking from best to worst based on the frequency of strong, moderate and weak capabilities in modelling SA is presented in Table 5.

Table 5
Resulting ranking of MADM software according to the assessment criteria in section 3.1.2.

Ranking position	MADM software
1.	Entscheidungsnavi
2.	HELDA
3.	MCDA-ULaval
4.	DEXiWin
5.	LDW
6.	SilverDecisions
7.	RuLeStudio
8.	ValueDecisions
9.	FITradeoff / MakeDecision.it
10.	MAMCA
11.	MCDA Index Tool
12.	J-Electre
13.	SOCRATES
14.	MCDA Calculator
15.	AHP-OS
16.	EWMS
17.	DecisionMaster
18.	MCDMaker
19.	PROMETHEE-Cloud / Apollo-Live / WebHIPRE 2.0
20.	Super Decisions
21.	DecSpace
22.	Decision Radar

3.3.2. Recommendations for software extension for SA

The estimation of effort required to extent software functionality for SA according to section 2.3.2 is presented in Fig. 6. Of the 29 assessment criteria, 13.8 % are categorized as low effort, 34.5 % as low-moderate, 41.4 % as moderate-high, and 10.3 % as high effort. Fig. 6 was divided into four quadrants or sections, designated as QI, QII, QIII and QIV. The left side of the figure, QI and QII, comprises criteria that are deemed "easy wins", that is, criteria that would require only low or low-moderate effort to achieve a strong fulfilment. The right side of the figure, QIII and QIV, contains criteria that are consider "hard wins", that is, criteria that would require moderate-high or high effort to achieve strong fulfilment. Similarly, the upper section, QI and QIII, contains criteria in which the majority of the software assessed exhibits a high level of weakness (W > 50 %), and thus it is reasonable to give them higher priority in the proposed recommendations. The lower section, QII and QIV, contains criteria in which the majority of the software assessed has a low level of weakness (W < 50 %). Consequently, these criteria have a lower priority in the recommendations, as there are already some software solutions available for that. The following paragraphs present further information on the criteria in each quadrant and strategies that could be implemented to extend the functionality of existing software. To confirm the considerations of amount of work, complexity and uncertainty related to the effort estimation, an interview was conducted with an MADM software developer.

In QI (upper left), criterion 8.1 *Learning dimension* notably exhibits the "weakest" fulfilment throughout the software assessed in this quadrant. Enhancing the software in this direction entails relatively low cognitive effort and work to enable multiple runs and simultaneous viewing and comparison of results by the user. Criteria 1.1, 6.2 and 8.2 are categorized in low-moderate effort because of the low cognitive effort and uncertainty, but higher amount of work. Criterion 1.1 *Ease of use* requires incorporating contextual help. Similarly, criterion 8.2 *Interpretation of results* requires a module or section to explain the main findings of the results. The fulfilment of criterion 6.2 *Sensitivity analysis: model* requires the integration of two or more normalization or aggregation functions and the ability to analyze the variation of results.

In QII (lower left), criteria 1.4, 3.2, and 3.3, are categorized with low effort because they do not require cognitive effort from the developers (low complexity and uncertainty). The fulfilment of criterion 1.4 *Personal information requirements* and criterion 1.2 *Software customizability* depends on the developers' preferences. In the first one (C 1.4), the

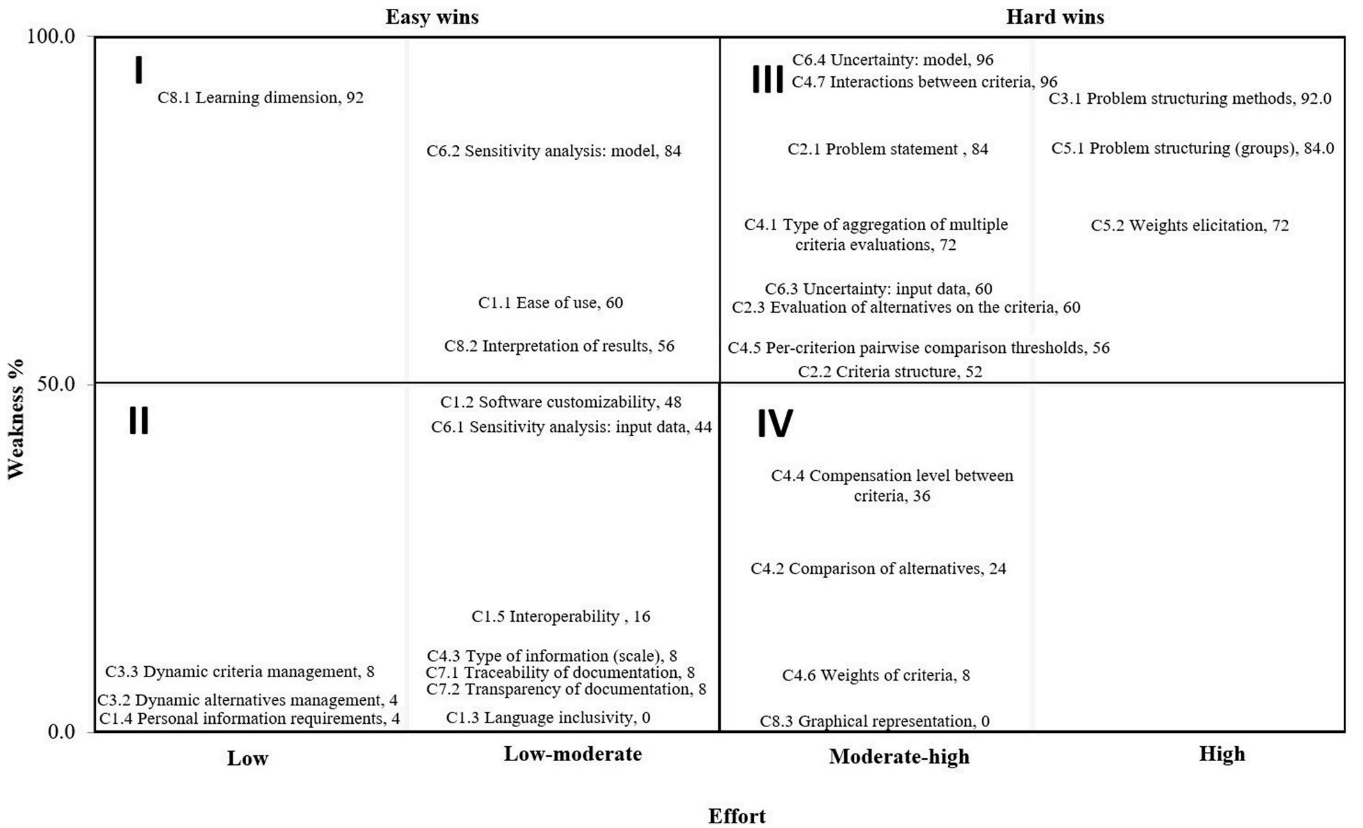

Fig. 6. Mapping of the effort to extend software functionality for sustainability assessment (x-axis) in relation to the level of weak fulfilment (Weakness %) of the group of software in each criterion (y-axis).

developer's decision concerns the request of credentials for software access, a process that is straightforward in terms of execution. In the second one (C 1.2), the developer determines whether to grant users the right to use and modify the software's code. However, once the decision to use an open-source license is made, a considerable investment of resources is required to prepare the necessary documentation and platform for the management and dissemination of the code. The fulfilment of criteria 3.2 *Dynamic alternatives management* and 3.3 *Dynamic criteria management* require minimal alterations to the code to include features for adding/selection and deleting/de-selection of criteria. Criteria 1.3, 1.5, 4.3, 6.1, 7.1, and 7.2 are categorized with low-moderate effort, and are associated with low complexity and uncertainty. However, their implementation would require a substantial amount of work. Criterion 1.3 *Language inclusivity* requires providing a user interface in different languages, this can be supported with the use of translation software e.g. artificial intelligence tools. The execution of criterion 1.5 *Interoperability* involves enhancing the software to import and export data using Excel files. Criterion 4.3 *Type of information (scale)* could be fulfilled with different strategies, e.g. a module for transforming data, adapting existing methods, or implementation of methods that accept different type of information. Criterion 6.1 *Sensitivity analysis: input data* requires code adaptation to allow for the calculations and graphical representation of the changes in the output of the model when varying one parameter (weights or performance values) at a time. Criterion 7.1 *Traceability of documentation* requires to regularly update the documentation (if necessary) and make it easily available, for example, via a software dedicated website. Criterion 7.2 *Transparency of documentation* requires upgrading the documentation to present information related to the following four topics i) problem formulation, ii) model creation (weights, aggregation functions and other parameters), iii) output variability analysis, and iv) visual representation.

In QIII (upper right), Criteria 3.1, 5.1 and 5.2 are categorized as high effort since their upgrade requires high complexity, high uncertainty and high amount of work. Criterion 3.1 *Problem structuring methods* requires not only programming effort but choosing an appropriate PSM e.g. VFT approach, brainstorming, SWOT analysis. Criteria 5.1 *Problem structuring (groups)* and 5.2 *Weights elicitation* require to integrate features for remote access or communication of stakeholders with the model. In addition, developers should bear in mind important considerations such as how to handle weights in group settings. Efforts to enhance the software in the direction of criteria 2.1 *Problem statement*, 4.1 *Type of aggregation*, 4.5 *Per criterion pairwise comparison thresholds*, and 4.7 *Interactions between criteria* involve moderate-high effort to either adapt existing method(s), or implement new methods in the software. The cognitive effort and uncertainty are moderate since existing literature on MCDA methods characterization can support the identification of appropriate methods to fulfil the criteria. The criteria *2.2 Criteria structure*, 2.3 *Evaluation of alternatives on the criteria,* 6.4 *Uncertainty: model,* and *6.3 Uncertainty: input data* are categorized with moderate-high effort because of the moderate level of complexity and uncertainty when deciding which type of uncertainty to implement in the software. For criterion 2.2 changes to the existing method or implementation of new methods is required. In the case of criterion 2.3 this involves adapting the code to accept uncertain data (e.g. performances using probability distributions) and performing calculations using this information. Criterion 6.3 requires i) adapting the algorithm to work with e.g. ranges of variation for input data and ii) performing calculations with multiple parameters varying simultaneously. For criterion 6.4, the effort may involve one or two steps, i) integrating two or more normalization or aggregation functions (if they are not available in the software), and ii) adapting the code to analyze the variation of output with different normalization or aggregation functions.

QIV (lower right) contains criteria in which most of the assessed software exhibited strong fulfilment ($W < 50\,\%$) despite of the moderate-high effort that would be required to improve the fulfilment. Criteria 4.2, 4.4, and 4.6 are categorized as moderate-high effort because of the moderate complexity and uncertainty required to either adapt existing method(s), or implement new methods in the software. Criterion 4.2 *Comparison of alternatives* requires changes to the existing method and implementation of new methods in order to support both, relative and absolute comparisons. Criterion 4.4 *Compensation level between criteria* requires integrating new methods (with low compensation) or adjustments to the algorithm to reduce the level of compensation e.g. use of thresholds. Criterion 4.6 *Weights of criteria* is categorized with moderate-high effort because changes to existing methods or the implementation of new methods is required as well as adaptions of the algorithm to the code. For criterion 4.6, the effort to add imprecise weights includes adapting the algorithm to receive this type of weights (e.g. interval weights), use them to perform calculations, and present the results in a relevant form. Criterion 8.3 *Graphical representation*, requires moderate-high effort for identifying optimal ways to present results for different methods, since it is not the case that all types of graphs are appropriate for every method (moderate uncertainty and amount of work).

4. Discussion

This section discusses the results in terms of four main topics: i) MADM software suitable for supporting sustainable transformation of energy systems; ii) the type of users for which the assessed software is designed; iii) relevant features for achieving long-term usability of the software; and iv) the limitations of the study.

4.1. MADM software for the sustainable transformation of energy systems

The transformation of energy systems towards sustainability involves decision-making processes concerning technologies and social issues, such as acceptance, user behavior, and governance [90]. There are several contexts within this transformation in which sustainability assessment could take place, e.g., energy planning, environmental management, or policy advice. These processes have one aspect in common: they involve the use of expert knowledge for democratic processes of opinion formation and decision-making [91]. This study identified several MADM software capable of handling expert-knowledge; however, only a few are ready-to-use software designed to support the integration of societal implications that energy technologies and systems have. Among the assessed software, there are several prominent tools with strong features that fulfil this requirement: AHP-OS, Apollo-Live, Entscheidungsnavi, HELDA, MAMCA, SOCRATES, and ValueDecisions. These software tools handle both, quantitative and qualitative information and facilitate the integration of stakeholders' preferences into decision models. In AHP-OS, Apollo-Live, Entscheidungsnavi and HELDA, the integration goes beyond merely collecting preferences to provide software that can support deliberation and consensus-reaching processes. Different software tools are suitable for different types of information or context. For example, AHP-OS, Entscheidungsnavi, HELDA, MAMCA, SOCRATES and ValueDecisions could support sustainability assessments for science-based policy advice. These software tools can integrate information coming from different methods, such as LCA, techno-economic assessment and/or optimization models, as well as societal preferences, which is important for that application [12]. It is not the case for Apollo-Live which is more useful in cases of high uncertainty when evaluating alternatives against criteria where only value-based judgements are feasible (e.g. when exploring energy policy objectives) [70]. It is important to address uncertainty in decision models for the energy transition, even when reliable performance data is available. The software mentioned in this section is capable of conducting different types of output variability analysis to assess the robustness of the decision recommendation.

4.2. Different MADM software for each user

This study shows that there is a variety of MADM software available for users with different levels of expertise in MADM methods. Many of the assessed software tools were designed to support MCDA calculation processes. These are recommended for experienced MCDA users looking for simple, efficient and effective tools to support their decision-making processes. Some examples include DecSpace, Decision Radar, MCDA Calculator, MCDMaker, and MakeDecision.it. These software tools provide users with a wide range of aggregation methods, giving them flexibility in the types of decision models they can develop. The available documentation for these software tools is usually simple and straightforward; in a few cases, however, no documentation is available (e.g., MCDMaker and Decision Radar). A second category involves a community of users familiar with a single method. The software tools in this category tend to have more capabilities than those in the first category and provide specific features for results analysis. Examples include DEXiWin for DEX, FITradeoff for MAVT, PROMETHEE-Cloud for PROMETHEE methods, MCDA ULaval for ELECTRE methods, Entscheidungsnavi for MAVT, and RuleStudio for DRSA. This type of software usually comes with good documentation for novice users, including dedicated websites with examples and publications demonstrating how to use the software. The FITradeoff website even provides training material. A third category involves software that was originally developed for a specific purpose, but which is also available and suitable for different contexts. Examples include Apollo-Live, which was developed for group decision making in energy and climate, SOCRATES which was developed for impact assessment problems, HELDA, which was developed to support SA in the energy sector, and ValueDecisions, which was developed for environmental and public policy decisions. The level of expertise in MCDA to use this type of software is expected to be lower, since the main target audience is not necessarily within the MCDA community but comes from other fields, e.g. LCA practitioners, policy analysts.

4.3. Long-term usability of MADM software

A common concern among software developers is that their software is not seen as a long-term solution by users. Features that can support this goal include licensing models, software maintenance capacity, and software user communities. This study revealed that a significant amount of free MCDA software is available, which users are taking advantage of for their studies. Much of the software assessed has an open-source license, which gives users more flexibility to extend or adapt a specific code to their needs and manage their data on their own servers, in the case of web-based applications. For extending or complementing software with additional MADM methods, developers could make use of existing libraries, e.g. pyDecision [92] or pyrepo-mcda [93] Python libraries. Relevant software that allows this includes ValueDecisions, SilverDecisions, MakeDecision.it, DEXiWin and Entscheidungsnavi. Software maintenance is a critical consideration, and for some developers, it may depend heavily on third-party funding. This implies that the software must be constantly developed to attract funding. Such development should keep up with research trends in order to solve specific issues. For instance, this study observed that MADM software has been extended or improved in recent years to integrate group preferences and participatory approaches. MADM software development paths could involve integrating artificial intelligence (AI) based methods and decision analysis. While this study did not identify any software capable of handling this yet, it is expected that ongoing research will result in this development.

Enhancing the interoperability of MADM software tools would be important for ensuring their long-term usability. From one side, it would be advantageous for MADM software developers to have some type of agreement on a standard format and definitions, to facilitate information exchange between software. The Decision Deck Project made an effort to establish a standard data format, which was named the XMCDA standard [94]. This initiative aimed to establish a universal modelling language to express MCDA concepts and generic decision aid processes. Unfortunately, there was no team that could maintain the infrastructure and make sure that all the new methods added did fit with the coding requirements. This is a challenging task and might even require a top-down strategy led by the MADM community to ensure acceptability and implementation. A disadvantage of such a strategy is that switching between software platforms is undesirable when thinking about the learning curve of the users for each tool [22]. Another consideration, is the interoperability with different types of software tools, such as LCA software. This would improve the use of these MADM software tools within communities conducting SA, e.g. EERA[1] (European Energy Research Alliance), SETAC[2] (Society of environmental toxicology and chemistry), and the ISSST[3] (International Symposium on Sustainable Systems and Technology).

4.4. Limitations of the study

While this study provides important insights into the capabilities of MADM software for SA, limitations must be considered when interpreting the results. One important limitation is that only software with versions released in the last 6 years are considered. The list of software could be expanded by considering software with the latest version released since 2015.

The qualitative analysis carried out to identify users' motivations from the respective publications is limited, as often, users do not state all the considerations for software selection in their publications. This research could be complemented by interviewing the authors of the publications from which the motivations were screened, to either confirm or complement the motivations already identified. In addition to that, the perspectives of MADM experts and non-MADM experts could be used to identify differences.

Another limitation relates to the simplification of some criteria. For example, the visualization of results (C 8.3) only considers tabular or graphical visualization, which does not necessarily indicate effective communication with stakeholders. Additional consideration should be given to the appropriateness of different visualization methods in conveying various types of information. A more detailed rationale for the assessment of specific criteria would enhance the precision and utility of the results of this study.

A further limitation of this study is that the software functionality was verified, but not tested in a use case by the authors of this study. Future work may include testing a selection of software in real applications (e.g., from previous publications) and comparing the functionality from a more practical perspective.

In addition, the assessment of effort for software enhancement presented in Fig. 6 is based on the functionality to be extended. Consequently, the effort required may vary for different software, depending on its structure, current functionality, and experience of the developing team. Further discussion with software developers is needed to confirm the considerations presented in this study, and to elucidate their relation to specific software. This would help to understand the main difficulties for extending each software, and to enable a comparative analysis.

5. Conclusions

The sustainable transformation of energy systems involves urgent and rapidly evolving decision-making problems. Sustainability assessment (SA) with multi-attribute decision making (MADM) methods is recognized as a valuable solution to this challenge. However, its execution can be regarded as complex and resource-consuming. In order to effectively tackle this complex issue, computational solutions are required to operationalize and accelerate decision-making processes. This paper conducts a systematic analysis to assess how to benefit from the functionality of existing MADM software to perform SA. A set of assessment criteria is developed, and a sample of MADM software is assessed to determine their strengths and weaknesses for supporting MADM methods in SA. Unlike previous MCDA software assessments, this study focuses on free MADM software and considers user needs and expectations derived from real-world SA use cases. In addition, it presents recommendations for extending the functionality of these software tools to align more closely with the needs of the SA-MADM community.

In this study, a sample of 25 free MADM software was assessed using 29 criteria and 8 domains. These criteria and domains were identified through: i) a literature review of 112 articles, to identify MADM software users' motivations for selecting MADM software, and ii) the capabilities of MADM methods to fulfil requirements for SA available in the literature. The domains assessed include applicability and accessibility, problem typology, problem structure, preference model, stakeholders' involvement, output variability analysis, transparency, and utility.

The assessment revealed the great efforts done by the MADM community to provide practitioners with robust, free MADM software for different types of decision problems. However, regarding SA, this study found that very few software tools are capable of modelling decision problems using the concept of strong sustainability, and eliciting preferences from a group of stakeholders. Besides that, most of the software, as of today, were primarily designed for users with a good level of expertise in MADM. This is supported, e.g., by the lack of contextual help, up-to-date manuals, and features to aid in result interpretation and learning. Our findings also confirm previous reviews that highlighted the scarcity of software capable of supporting problem structuring approaches. The software sample also tends to lack flexibility, which reduces the probability of potential users to attach, or even to use the software. For example, most of the software are specialized in one type of problem statement and one type of aggregation method. A significant gap identified in this study is the limited capability of the assessed software to support output variability analysis. Significant progress has been made in the last years to develop software that facilitates broader stakeholder participation in MADM processes. Such improvements not only reduce the time required for decision-making but also enhance the inclusivity and robustness of SA.

While none of the software evaluated in this study fully meets all identified requirements for SA, they provide a solid foundation for further development (extension) and adaptation. To address the current challenges posed by SA, this paper proposes a set of recommendations that outline strategies for enhancing the software tool's capabilities. These recommendations involve primarily software developers but also the whole MADM community. From this, some “easy wins” include improving the learning dimension, sensitivity analysis, result interpretation, and ease of use. Some “hard wins” include more complex solutions such as extending the functionality to integrate stakeholders in the respective MADM software, as well as advancing the preference models that can be managed by each software.

The successful implementation of these recommendations will require a collaborative effort to bridge the gap between SA practitioners, software developers, and the MCDA community,[4,5,6,7]. SA practitioners include, e.g., researchers from universities or research centers as well as from organizations like EERA, SETAC, ISSST, and ISIE[8] (International

[1] https://www.eera-set.eu/.
[2] https://www.setac.org/.
[3] https://issst.net/.
[4] https://www.mcdmsociety.org/.
[5] https://www.cs.put.poznan.pl/ewgmcda/.
[6] https://connect.informs.org/multiple-criteria-decision-making/home.
[7] https://connect.informs.org/das/home.
[8] https://is4ie.org/

Society for Industrial Ecology). This connection could help to foster greater visibility, adoption, and advancement of powerful software to aid SA.

The findings of this study imply that the provision of ready-to-use MADM software, encompassing the capabilities or concepts pertinent to conducting SA, enhances strategic decision-making processes in the context of energy system transformation.

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.esr.2025.102016.

Data availability

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

References

[1] K. Koasidis, A. Nikas, A. Karamaneas, et al., Climate and sustainability co-governance in Kenya: a multi-criteria analysis of stakeholders' perceptions and consensus, Energy Sustain. Dev. 68 (2022) 457–471, https://doi.org/10.1016/j.esd.2022.05.003.
[2] H. Weistroffer, Y. Li, Multiple criteria decision analysis software, in: M.E. Salvatore Greco, José Rui Figueira (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer, New York, NY, 2016, pp. 1301–1341.
[3] M.M. Wiecek, M. Ehrgott, A. Engau, Continuous multiobjective programming, in: S. Greco, et al. (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer New York, New York, NY, 2016, pp. 739–815.
[4] A. Azapagic, S. Perdan, An integrated sustainability decision-support framework Part II: problem analysis, Int. J. Sustain. Dev. World Ecol. 12 (2005) 112–131, https://doi.org/10.1080/13504500509469623.
[5] C. Thies, K. Kieckhäfer, T.S. Spengler, et al., Operations research for sustainability assessment of products: a review, Eur. J. Oper. Res. 274 (2019) 1–21, https://doi.org/10.1016/j.ejor.2018.04.039.
[6] A. Lindfors, Assessing sustainability with multi-criteria methods: a methodologically focused literature review, Environ. Sustain. Indic. 12 (2021) 100149, https://doi.org/10.1016/j.indic.2021.100149.
[7] L.C. Dias, C. Caldeira, S. Sala, Multiple criteria decision analysis to support the design of safe and sustainable chemicals and materials, Sci. Total Environ. 916 (2024) 169599, https://doi.org/10.1016/j.scitotenv.2023.169599.
[8] C. Wulf, L.S. Mesa Estrada, M. Haase, et al., MCDA for the sustainability assessment of energy technologies and systems: identifying challenges and opportunities, Energy Sustain. Soc. 15 (2025) 45, https://doi.org/10.1186/s13705-025-00546-8.
[9] M. Cinelli, S.R. Coles, K. Kirwan, Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment, Ecol. Indic. 46 (2014) 138–148, https://doi.org/10.1016/j.ecolind.2014.06.011.
[10] M. Cinelli, C. Koffler, C. Askham, et al., Criteria used to review weighting methods as part of the UN environment life cycle initiative's global guidance on environmental life cycle impact assessment indicators (GLAM) Project, Abstract book Conference SETAC Europe 32nd Annual Meeting. Society of Environmental Toxicology and Chemistry (2022) 559–560.
[11] S. Huysveld, S.E. Taelman, I. Hackenhaar, et al., ORIENTING - D1.5 Critical Evaluation of Sustainability Integration Approaches, ORIENTING project, 2022, https://doi.org/10.13140/RG.2.2.36610.96960.
[12] G. Munda, Multiple criteria decision analysis and sustainable development, in: S. Greco, et al. (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer New York, New York, NY, 2016, pp. 1235–1267.
[13] B. Talukder, K.W. Hipel, Review and selection of multi-criteria decision analysis (MCDA) technique for sustainability assessment, in: J. Ren (Ed.), Energy Systems Evaluation, vol. 1, Green Energy and Technology, Cham, 2021, https://doi.org/10.1007/978-3-030-67529-5_7.
[14] L.S. Mesa Estrada, M. Haase, C. Wulf, et al., MCDA for Sustainability Assessment – Insights to Helmholtz Association Activities, Zenodo, 2022, https://doi.org/10.5281/zenodo.7273635.
[15] M. Dean, Including multiple perspectives in participatory multi-criteria analysis: a framework for investigation, Evaluation 28 (2022) 505–539, https://doi.org/10.1177/13563890221123822.
[16] J. Mustajoki, M. Marttunen, Comparison of multi-criteria decision analytical software for supporting environmental planning processes, Environ. Model. Software 93 (2017) 78–91, https://doi.org/10.1016/j.envsoft.2017.02.026.
[17] A. Moreno-Calderón, T.S. Tong, P. Thokala, Multi-criteria decision analysis software in healthcare priority setting: a systematic review, Pharmacoeconomics 38 (2020) 269–283, https://doi.org/10.1007/s40273-019-00863-9.
[18] M. Mohamad, A. Selamat, An analysis of rough set-based application tools in the decision-making process, in: F. Saeed, et al. (Eds.), Recent Trends in Information and Communication Technology, Lecture Notes on Data Engineering and Communications Technologies, vol. 5, Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-59427-9_49.
[19] J. Beekman, Decision Analysis Software Survey. https://pubsonline.informs.org/do/10.1287/orms.2020.06.04/full/, 2020. (Accessed 15 August 2025).
[20] M. Cinelli, M. Spada, W. Kim, et al., MCDA Index Tool: an interactive software to develop indices and rankings, Environ. Syst. Decis. 41 (2021) 82–109, https://doi.org/10.1007/s10669-020-09784-x.
[21] International Society on MCDM, Software related to MCDM. https://www.mcdmsociety.org/2025/01/12/software/, 2025. (Accessed 10 December 2025).
[22] H. Huang, P. Burgherr, MCDA calculator: a streamlined decision support System for multi-criteria decision analysis, in: S.P. Duarte, et al. (Eds.), Lecture Notes in Business Information Processing, vol 506, Springer, Cham, 2024, https://doi.org/10.1007/978-3-031-59376-5_3. Decision Support Systems XIV. Human-Centric Group Decision, Negotiation and Decision Support Systems for Societal Transitions. ICDSST 2024.
[23] S. Fahmy, N. Ngah, W. Othman, et al., Evaluating the quality of software in e-Book using the ISO 9126 model, International Journal of Control and Automation 5 (2012) 115–122.
[24] International Electrotechnical Commission, Electropedia: the world's online electrotechnical vocabulary-international electrotechnical vocabulary (IEV). https://www.electropedia.org/iev, 2019.
[25] M. Ouzzani, H. Hammady, Z. Fedorowicz, et al., Rayyan—A web and mobile app for systematic reviews, Syst. Rev. 5 (2016), https://doi.org/10.1186/s13643-016-0384-4.
[26] M. Cinelli, M. Kadziński, G. Miebs, et al., Recommending multiple criteria decision analysis methods with a new taxonomy-based decision support system, Eur. J. Oper. Res. 302 (2022) 633–651, https://doi.org/10.1016/j.ejor.2022.01.011.
[27] European Commission: Directorate-General for Research Innovation, Horizon Europe, Open Science – Early Knowledge and Data Sharing, and Open Collaboration, Publications Office of the European Union, 2021. https://data.europa.eu/doi/10.2777/18252.
[28] D. Mendez, D. Graziotin, S. Wagner, et al., Open science in software engineering, in: M. Felderer, et al. (Eds.), Contemporary Empirical Methods in Software Engineering, Springer International Publishing, Cham, 2020, 477-501.
[29] R.D. Banker, G.B. Davis, S.A. Slaughter, Software development practices, software complexity, and software maintenance performance: a field Study, Manag. Sci. 44 (1998) 433–450, https://doi.org/10.1287/mnsc.44.4.433.
[30] J. Choudhari, U. Suman, Story points based effort estimation model for software maintenance, Procedia Technol. 4 (2012) 761–765, https://doi.org/10.1016/j.protcy.2012.05.124.
[31] Project Management Institute, Agile Practice Guide, Project Management Institute, 2017.
[32] H. Halland, L. Lamprinakis, I. Kvalvik, et al., Learning for sustainability in horticultural production in Arctic Norway. https://doi.org/10.3389/fsufs.2021.686104, 2021.
[33] B.R. Ssebunya, C. Schader, L. Baumgart, et al., Sustainability performance of certified and non-certified smallholder coffee farms in Uganda. https://doi.org/10.1016/j.ecolecon.2018.09.004, 2019.
[34] F. Abastante, C. Caprioli, M. Gaballo, The economic evaluation of projects as a structuring discipline of learning processes to support decision-making in sustainable urban transformations, Int. J. Sustain. Dev. Plann. 17 (2022) 1297–1307, https://doi.org/10.18280/ijsdp.170427.
[35] B. Roy, Méthodologie Multicritère D'Aide À la Décision, 1985.
[36] C. Wulf, P. Zapp, A. Schreiber, et al., Lessons learned from a life cycle sustainability assessment of rare Earth permanent magnets, J. Ind. Ecol. 21 (2017) 1578–1590, https://doi.org/10.1111/jiec.12575.
[37] M. Cinelli, S.R. Coles, M.N. Nadagouda, et al., Robustness analysis of a green chemistry-based model for the classification of silver nanoparticles synthesis processes, J. Clean. Prod. 162 (2017) 938–948, https://doi.org/10.1016/j.jclepro.2017.06.113.
[38] A.R. Domingues, P. Marques, R. Garcia, et al., Applying multi-criteria decision analysis to the life-cycle assessment of vehicles, J. Clean. Prod. 107 (2015) 749–759, https://doi.org/10.1016/j.jclepro.2015.05.086.
[39] C.-C. Huang, H.-W. Ma, A multidimensional environmental evaluation of packaging materials, Sci. Total Environ. 324 (2004) 161–172, https://doi.org/10.1016/j.scitotenv.2003.10.039.
[40] J. Lou, L. Deng, D. Wang, Understanding the deep structure use of mobile phones – an attachment perspective, Behav. Inf. Technol. 41 (2022) 3191–3209, https://doi.org/10.1080/0144929X.2021.1975822.
[41] A. Mishra, S.K. Das, K.R. Reddy, Life cycle triple bottom line sustainability assessment of coal mine overburden sand versus river sand, Resour. Policy 86 (2023) 104140, https://doi.org/10.1016/j.resourpol.2023.104140.
[42] E. Eberling, S. Langkau, Utilizing SDGs in sustainability assessments of innovations: deriving methodological recommendations from existing approaches, J. Clean. Prod. 437 (2024) 140383, https://doi.org/10.1016/j.jclepro.2023.140383.
[43] M. C, M. Sabaghi, P. Baptiste, R. Rostamzadeh, Sustainability Assessment Using fuzzy-inference Technique (SAFT): a Methodology Toward Green Products, vol. 56, 2016, https://doi.org/10.1016/j.eswa.2016.02.038.

[44] G.L. Søndergaard, P.J. Binning, M. Bondgaard, et al., Multi-criteria assessment tool for sustainability appraisal of remediation alternatives for a contaminated site. https://doi.org/10.1007/s11368-017-1805-2, 2018.
[45] M.G. Riera Pérez, E. Rey, A multi-criteria approach to compare urban renewal scenarios for an existing neighborhood. Case study in Lausanne (Switzerland), Build. Environ. 65 (2013) 58–70, https://doi.org/10.1016/j.buildenv.2013.03.017.
[46] S.M. Moni, R. Mahmud, K. High, et al., Life cycle assessment of emerging technologies: a review, J. Ind. Ecol. 24 (2020) 52–63, https://doi.org/10.1111/jiec.12965.
[47] M. Marttunen, J. Lienert, V. Belton, Structuring problems for Multi-Criteria Decision Analysis in practice: a literature review of method combinations, Eur. J. Oper. Res. 263 (2017) 1–17, https://doi.org/10.1016/j.ejor.2017.04.041.
[48] R. McKenna, V. Bertsch, K. Mainzer, et al., Combining local preferences with multi-criteria decision analysis and linear optimization to develop feasible energy concepts in small communities, Eur. J. Oper. Res. 268 (2018) 1092–1110, https://doi.org/10.1016/j.ejor.2018.01.036.
[49] M.A. Moktadir, J. Ren, Promoting sustainable management of hazardous waste-to-wealth practices: an innovative integrated DPSIR and decision-making framework, J. Environ. Manag. 344 (2023) 118470, https://doi.org/10.1016/j.jenvman.2023.118470.
[50] M. Cinelli, M.A. Gonzalez, R. Ford, et al., Supporting contaminated sites management with Multiple Criteria Decision Analysis: demonstration of a regulation-consistent approach, J. Clean. Prod. 316 (2021) 128347, https://doi.org/10.1016/j.jclepro.2021.128347.
[51] E. Sigcha, D. Sucozhañay, F. Cabrera, et al., Applying social life cycle assessment in the informal recycling sector: understanding challenges and limitations, Waste Manag. 181 (2024) 20–33, https://doi.org/10.1016/j.wasman.2024.03.029.
[52] A. Gani, A.T. James, M. Asjad, et al., Development of a manufacturing sustainability index for MSMEs using a structural approach, J. Clean. Prod. 353 (2022) 131687, https://doi.org/10.1016/j.jclepro.2022.131687.
[53] P. Ziyeh, M. Cinelli, A framework to navigate eco-labels in the textile and clothing industry, Sustainability 15 (2023) 14170, https://doi.org/10.3390/su151914170.
[54] M. Cinelli, S.R. Coles, M.N. Nadagouda, et al., A green chemistry-based classification model for the synthesis of silver nanoparticles, Green Chem. 17 (2015) 2825–2839, https://doi.org/10.1039/C4GC02088J.
[55] M. Ghaleb, S. Taghipour, Assessing the impact of maintenance practices on asset's sustainability, Reliab. Eng. Syst. Saf. 228 (2022), https://doi.org/10.1016/j.ress.2022.108810.
[56] L. Rosén, P.-E. Backb, T. Söderqvist, et al., SCORE: a Novel multi-criteria Decision Analysis Approach to Assessing the Sustainability of Contaminated Land Remediation, 2015, p. 511.
[57] P. Victor, S. Hanna, A. Kubursi, How strong is weak sustainability? in: S. Faucheux, et al. (Eds.), Sustainable Development: Concepts, Rationalities and Strategies Springer Netherlands, Dordrecht, 1998, pp. 195–210.
[58] J. Wątróbski, A. Karczmarczyk, A. Bączkiewicz, Using the TOSS method in semi-autonomous passenger car selection, Sustain. Energy Technol. Assessments 58 (2023), https://doi.org/10.1016/j.seta.2023.103367.
[59] B. Roy, J.R. Figueira, J. Almeida-Dias, Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: theoretical results and practical issues, Omega 43 (2014) 9–20, https://doi.org/10.1016/j.omega.2013.05.003.
[60] C. Wulf, P. Zapp, A. Schreiber, et al., Setting thresholds to define indifferences and preferences in PROMETHEE for life cycle sustainability assessment of European hydrogen production, Sustainability 13 (2021) 7009, https://doi.org/10.3390/su13137009.
[61] M. Cinelli, P. Burgherr, M. Kadziński, et al., Proper and improper uses of MCDA methods in energy systems analysis, Decis. Support Syst. 163 (2022) 113848, https://doi.org/10.1016/j.dss.2022.113848.
[62] R. Pelissari, M.C. Oliveira, S.B. Amor, et al., SMAA methods and their applications: a literature review and future research directions, Ann. Oper. Res. 293 (2020) 433–493, https://doi.org/10.1007/s10479-019-03151-z.
[63] M. Bottero, V. Ferretti, J.R. Figueira, et al., Dealing with a multiple criteria environmental problem with interaction effects between criteria through an extension of the Electre III method, Eur. J. Oper. Res. 245 (2015) 837–850, https://doi.org/10.1016/j.ejor.2015.04.005.
[64] H.V. Rowley, A. Geschke, M. Lenzen, A practical approach for estimating weights of interacting criteria from profile sets, Fuzzy Set Syst. 272 (2015) 70–88, https://doi.org/10.1016/j.fss.2015.01.011.
[65] H.W. Kua, On life cycle sustainability unified analysis, J. Ind. Ecol. 21 (2017) 1488–1506, https://doi.org/10.1111/jiec.12665.
[66] R. Bandari, E.A. Moallemi, R.E. Lester, et al., Prioritising Sustainable Development Goals, characterising interactions, and identifying solutions for local sustainability, Environ. Sci. Pol. 127 (2022) 325–336, https://doi.org/10.1016/j.envsci.2021.09.016.
[67] J.R. Figueira, S. Greco, B. Roy, ELECTRE methods with interaction between criteria: an extension of the concordance index, Eur. J. Oper. Res. 199 (2009) 478–495, https://doi.org/10.1016/j.ejor.2008.11.025.
[68] G. Dorini, Z. Kapelan, A. Azapagic, Managing uncertainty in multiple-criteria decision making related to sustainability assessment, Clean Technol. Environ. Policy 13 (2011) 133–139, https://doi.org/10.1007/s10098-010-0291-7.
[69] K.D. Goepel, Implementation of an online software tool for the Analytic Hierarchy Process (AHP-OS), International Journal of the Analytic Hierarchy Process (2018), https://doi.org/10.13033/ijahp.v10i3.590.
[70] K. Koasidis, A. Soursos, G. Xexakis, et al., 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 (2025), https://doi.org/10.12688/openreseurope.19614.3.
[71] A.S. Costa, R. Rodrigues, A. Xiang, et al., Supporting the use of decision aiding methods by non-specialists, in: B. Shishkov (Ed.), Business Modeling and Software Design. Business Modeling and Software Design, Springer International Publishing, 2019, https://doi.org/10.1007/978-3-030-24854-3_6.
[72] M. Bohanec, DEX (Decision EXpert): a qualitative hierarchical multi-criteria method, in: A.J. Kulkarni (Ed.), Multiple Criteria Decision Making: Techniques, Analysis and Applications, Springer Singapore, Singapore, 2022, pp. 39–78.
[73] A. Podvesovskii, A. Zakharova, D. Korostelyov, et al., DecisionMaster: a multi-criteria decision support system with ability to combine different decision rules, SoftwareX 16 (2021) 100814, https://doi.org/10.1016/j.softx.2021.100814.
[74] C. Hannes, R. von Nitzsch, A systematic approach to problem formulation in a decision support system, International Journal of Decision Support Systems 5 (2024) 14–29, https://doi.org/10.1504/IJDSS.2024.137233.
[75] Y. Yin, C. Song, Y. Jing, et al., EWMS: a software tool for interactively using entropy weight coefficient method for aggregating sustainability indicators, Environ. Model. Software 191 (2025) 106500, https://doi.org/10.1016/j.envsoft.2025.106500.
[76] A. Almeida, E. Frej, L. Roselli, et al., A summary on FITRADEOFF method with methodological and practical develpoments and future perspectives, Pesqui. Oper. 43 (2023), https://doi.org/10.1590/0101-7438.2023.043spe1.00268356.
[77] T. Müller, L. Mesa Estrada, M. Haase, HELDA - helmholtz MCDA Tool. User manual. https://www.mcda-helmholtz.de/downloads/User%20manual/HELDA_Manual_v1.pdf, 2025. (Accessed 10 December 2025).
[78] V. Pereira, L.D. de Oliveira Nepomuceno, J-Electre v3. 0, User guide. https://www.researchgate.net/profile/Valdecy-Pereira/publication/351087560_J-Electre-v30/links/608474f5907dcf667bc0a52f/J-Electre-v30.pdf, 2021.
[79] W. Edwards, Utility Theories: Measurements and Applications, 1 ed., Springer Dordrecht, 2013.
[80] J. Więckowski, W. Sałabun, MakeDecision: online system for the graphical design of decision-making models in crisp and fuzzy environments, SoftwareX 26 (2024) 101658, https://doi.org/10.1016/j.softx.2024.101658.
[81] H. Huang, B. Peter, C. Macharis, A collaborative group decision-support system: the survey based multi-actor multi-criteria analysis (MAMCA) software, J. Oper. Res. Soc. 76 (2025) 844–865, https://doi.org/10.1080/01605682.2024.2398114.
[82] N. Couture-Grenier, I. Abi-Zeid, MCDA ULaval: a multi-platform Multicriteria software. http://www.cs.put.poznan.pl/ewgmcda/newsletter/Spr15.pdf, 2015. (Accessed 12 December 2025).
[83] E. Pohl, J. Geldermann, PROMETHEE-Cloud: a web app to support multi-criteria decisions, EURO Journal on Decision Processes 12 (2024) 100053, https://doi.org/10.1016/j.ejdp.2024.100053.
[84] M. Szeląg, J. Błaszczyński, S. Roman, Rough set analysis of classification data with missing values, in: L. Polkowski, et al. (Eds.), Rough Sets, Springer, Cham, 2017, pp. 552–565.
[85] B. Kamiński, M. Jakubczyk, P. Szufel, A framework for sensitivity analysis of decision trees, Cent. Eur. J. Oper. Res. 26 (2018) 135–159, https://doi.org/10.1007/s10100-017-0479-6.
[86] G. Munda, Qualitative reasoning or quantitative aggregation rules for impact assessment of policy options? A multiple criteria framework, Qual. Quantity 56 (2022) 3259–3277, https://doi.org/10.1007/s11135-021-01267-8.
[87] E. Mu, M. Pereyra-Rojas, Practical Decision Making Using Super Decisions v3: an Introduction to the Analytic Hierarchy Process, 1 ed., Springer Cham, 2017.
[88] F. Haag, A.H. Aubert, J. Lienert, ValueDecisions, a web app to support decisions with conflicting objectives, multiple stakeholders, and uncertainty, Environ. Model. Software 150 (2022), https://doi.org/10.1016/j.envsoft.2022.105361.
[89] J. Mustajoki, R.P. Hämäläinen, Web-Hipre: Global decision support by value tree and AHP analysis, INFOR Inf. Syst. Oper. Res. 38 (2000) 208–220, https://doi.org/10.1080/03155986.2000.11732409.
[90] A. Grunwald, Technology assessment and Policy advice in the field of sustainable development, in: L.W. Zacher (Ed.), Technology, Society and Sustainability, Springer, Cham, 2017.
[91] A. Grunwald, Technology assessment at the German Bundestag: 'expertising' democracy for 'democratising' expertise, Sci. Publ. Pol. 30 (2003) 193–198, https://doi.org/10.3152/147154303781780515.
[92] V. Pereira, M.P. Basilio, C.H.T. Santos, Enhancing decision analysis with a large language model: pydecision a comprehensive library of MCDA methods in Python, J. Model. Manag. (2025), https://doi.org/10.1108/JM2-04-2024-0118.
[93] J. Wątróbski, A. Bączkiewicz, W. Sałabun, pyrepo-mcda — reference objects based MCDA software package, SoftwareX 19 (2022) 101107, https://doi.org/10.1016/j.softx.2022.101107.
[94] S. Bigaret, P. Meyer, XMCDA: an xml-based encoding standard for MCDA data, in: R. Bisdorff, et al. (Eds.), Evaluation and Decision Models with Multiple Criteria. International Handbooks on Information Systems, Springer, Berlin, Heidelberg, 2015.