Environmental Modelling and Software 150 (2022) 105361

Contents lists available at ScienceDirect

Environmental Modelling and Software

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

ValueDecisions, a web app to support decisions with conflicting objectives, multiple stakeholders, and uncertainty

Fridolin Haag[a,b,*], Alice H. Aubert[a], Judit Lienert[a]

[a] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600, Dübendorf, Switzerland
[b] Leibniz Centre for Tropical Marine Research (ZMT), Fahrenheitstraße 6, 28359, Bremen, Germany

ARTICLE INFO

Keywords:
Multi criteria decision analysis
MCDA software
Multi-attribute value theory
Environmental decision analysis
Open source
Population survey

ABSTRACT

Complex environmental and public policy decisions profit from structured procedures such as multi-criteria decision analysis (MCDA). To support such decisions, the new open source application ValueDecisions provides advanced analysis and visualization with no programming expected from users. Based on multi-attribute value theory (MAVT), it offers analysis for decisions with conflicting and interacting objectives, multiple stakeholders, and uncertain consequences of options. Programmed in R, the shiny web framework makes it accessible via a graphical user interface in the browser. We exemplify using ValueDecisions for a wastewater infrastructure planning case in the Paris region. We surveyed preferences of 655 citizens and conducted sensitivity analysis of preference parameters. The best management options were robust across a range of preference profiles and assumptions. To evaluate the app, we developed a novel usability test based on the ISO standard for software quality and surveyed students using ValueDecisions for case studies.

1. Introduction

Environmental decisions and other public policy problems have characteristics that make them messy, complex, and difficult to tackle in a rational way. They need to be addressed in a social context often concerning many stakeholders (or decision-makers) with diverse views (French and Geldermann, 2005). These may include citizens, future generations, and other interest groups with opposed values, resulting in conflicts of interest. Because of the plurality of stakeholder views and the extent of decision consequences in space and time, multiple objectives need to be considered in these decisions (Gregory et al., 2012; Haag et al., 2019c). Since it is usually not possible to fully achieve all objectives, difficult trade-offs are required. Furthermore, the outcomes of decision options and their environmental impact is uncertain (Reichert et al., 2015).

In response to these challenges, structured, transparent procedures such as multi-criteria decision analysis (MCDA) have been developed. While the umbrella term MCDA encompasses various approaches (see Belton and Stewart, 2002; Greco et al., 2016; Cinelli et al., 2020), the common element is the development of a quantitative or qualitative model of the decision situation. These models support structuring information, making trade-offs, and choosing decision options to implement. MCDA approaches are increasingly considered and applied for environmental problems (Hajkowicz and Collins, 2007; Huang et al., 2011; Cegan et al., 2017; Esmail and Geneletti, 2018) and used in government agencies to address public issues (Kurth et al., 2017). Our work is based on multi-attribute value and utility theory (MAVT/MAUT; Keeney and Raiffa, 1993). In environmental management, decision support based on this conceptual background has also been popularized under the term “structured decision making” (Gregory et al., 2012).

Decision-making processes can be facilitated by software tools that allow interactive model building, exploration, and visualization of results. Accordingly, MCDA approaches have often been developed in tandem with respective decision support software (see Korhonen et al., 1992 for an early review). Especially group decision support has a tradition of technology-driven approaches (Morton et al., 2003). Consequently, a wide variety of MCDA software has been developed; online collections list more than 45 different products (EWG-MCDA, 2020; MCDM-Society, 2020). Systematic reviews and comparisons of software products are provided by French and Xu (2005) and Weistroffer and Li (2016) for general applications and in Mustajoki and Marttunen (2017) specifically with regard to environmental planning processes.

When deciding which software to use, one faces trade-offs, for instance, regarding feature-richness, adaptability, user-friendliness,

* Corresponding author. Leibniz Centre for Tropical Marine Research (ZMT), Fahrenheitstraße 6, 28359, Bremen, Germany.
E-mail address: fridolin.haag@leibniz-zmt.de (F. Haag).

https://doi.org/10.1016/j.envsoft.2022.105361
Received 7 July 2021; Received in revised form 10 November 2021; Accepted 14 February 2022
Available online 16 February 2022

speed, or costs. These objectives are often conflicting and software developments found different answers to these conflicts (Mustajoki and Marttunen, 2017). The fit of a software tool to the specific context in which it is used is crucial (Belton and Hodgkin, 1999; French and Xu, 2005). Based on applied research in participatory environmental and public policy decision support (e.g., Borsuk et al., 2008; Lienert et al., 2011; Zheng et al., 2016; Haag et al., 2019b; Harris-Lovett et al., 2019), we identified five common requirements for decision analysis software in theses contexts: (i) flexibility in representing stakeholder preferences across multiple, potentially conflicting, objectives; (ii) ability to consider - oftentimes large - uncertainty of predicted consequences and stakeholder preferences as well as support to understand robustness of conclusions; (iii) capability to compare results for multiple stakeholders; (iv) provision of analysis via a graphical user interface to users, such as consultants, who do not necessarily have programming skills; and (v) extendability and adaptability to new requirements. In addition to these context-specific requirements, there are fundamental requirements of software quality (see SI-5).

This study presents, tests, and evaluates a novel open source application, ValueDecisions, which we developed to target these requirements. As described above, there is a rich history of software that has been successfully used to support multi-criteria decisions (see Korhonen et al., 1992; French and Xu, 2005; Weistroffer and Li, 2016; Mustajoki and Marttunen, 2017). However, we found the flexibility to represent preferences and uncertain predictions often limited, and a unified approach to all five requirements at once missing. The main properties of ValueDecisions to tackle the requirements are: (i) support for complex, hierarchical MAVT models by combining non-linear lowest-level value functions with non-additive aggregation functions that can represent most preference structures; (ii) consideration of a range of probability distributions for predictions, interactive sensitivity analyses, and different visualizations of uncertainty; (iii) simultaneous comparison of multiple stakeholder preference profiles without forcing aggregation of individual preferences; (iv) a graphical user interface that can be accessed via a web browser with many options to save analysis results; and (v) open source development based on R (R Core Team, 2021), which allows for extensions by interested users. This combination of features is, to the best of our knowledge, unique. ValueDecisions attempts to strike a balance between simplicity and analytical possibilities that are relevant in the environmental field while being extensible, free of charge, and easily accessible.

For participatory decision support, other aspects than the sheer feature-richness of software are relevant. One aspect we want to highlight is the issue of access. Our intention is to make multi-criteria analysis accessible to interested users or stakeholders, such as municipalities, public authorities, consultants, and students. Even if users do not conduct the main analysis themselves, it would be helpful if they could engage with decision models created by researchers, for instance, to increase trust and acceptance in the analysis (Hämäläinen, 2015; Voinov et al., 2016).

Access can be limited by cost of commercial software, incompatibilities of hardware, software, or operating systems, and user rights in organizations. We require new ways to provide analysis and models to stakeholders. Web apps, as the app we introduce here, are one comparatively new possibility to mitigate these access impediments. The open source development means that the implementation is transparent and the app can be used free of charge. Access can also be limited by the required technical expertise to use software. As a community, we benefit from a variety of tools that fulfill needs of different target users (see section 3.2). Complex models and analysis are often only possible with frameworks that require programming, such as “utility” for R (Reichert et al., 2013) or “decisi-o-rama” for Python (Chacon-Hurtado and Scholten, 2020). This can make analytic capabilities that are useful in environmental decision support, for example, regarding uncertainty, inaccessible to interested users. With ValueDecisions we wanted to build a bridge by making advanced analysis available via a graphical user interface without the need to program. This benefits users without programming skills, but also facilitates interactive model exploration in workshop settings.

We propose that software development, similar to decision-making, can profit from value-focused thinking (Keeney, 1992). While all software developers have an interest to praise their own tool, the objectives of what should be achieved with the software should be kept at the forefront. Despite its importance, empirical evaluation of decision support systems remains rare (Walling and Vaneeckhaute, 2020). Therefore, this study includes a structured evaluation of ValueDecisions. We used the Software Product Quality Requirements and Evaluation (SQuaRE) series of standards, ISO/IEC 25000, as reference in our development and basis to evaluate ValueDecisions against defined objectives (ISO/IEC JTC 1/SC 7, 2011). This ISO standard defines two evaluation models. To gain insights into the usability of the app, we operationalized the *quality in use* model by a usability test in which representative users performed representative tasks. Additionally, we self-evaluated the application based on the *product quality* model of this framework (SI-5). As evaluation from an applied perspective, we analyzed an unpublished real-world environmental decision problem with ValueDecisions.

The remainder of this paper is organized as follows: In section 2, we introduce MCDA models, specifically multi-attribute value theory that forms the theoretical basis for ValueDecisions. Section 3 introduces the features and implementation of ValueDecisions. In section 4, we exemplify its use in a case study for wastewater infrastructure planning in the Paris region, based on a population survey with 655 respondents. Section 5 describes the development and results from the usability evaluation. We end with a general discussion and conclusions in section 6.

2. Decision support based on multi-attribute value theory (MAVT)

2.1. The decision analysis process

A decision-making process can be structured as a sequence of phases, which are intertwined and iterative in practice. We can differentiate between phases of problem structuring, data and information collection, decision analysis, negotiation and conflict resolution, implementation, and monitoring. Within this process, the aim of (multi-criteria) decision analysis is to develop and evaluate a set or sets of options (also called alternatives, strategies, actions, variants, or scenarios depending on the literature) on a set of objectives (also called criteria) and provide insights - e.g., about optimal decisions, involved trade-offs, and stakeholder perspectives - based on a decision model. The ValueDecisions app supports the phase of analyzing the decision options from different stakeholder perspectives. It builds upon an adequate structuring of the decision problem, predictions of (uncertain) outcomes of options, and the completed elicitation of stakeholder preferences.

The first steps of an analysis are fundamentally important, not least due to path dependencies (Lahtinen et al., 2017a). In the problem structuring phase, the boundaries of the decision problem are set and stakeholders identified (e.g., Lienert et al., 2015; Marttunen et al., 2017). Then, often in a participatory process, basic decision elements are defined: the objectives (what stakeholders find fundamentally important to achieve), the decision options (with which alternatives, strategies, actions the objectives can be achieved), and attributes (how to quantify or measure the performance of options) (e.g., Gregory et al., 2012).

Next, predictions of how each option performs, as measured by the attributes, are needed. Predicting potential consequences of options is a major undertaking of natural and engineering science and a large component of some decision analysis projects. Predictions can be derived from quantitative models, scientific literature, or expert judgment. As predictions are inherently uncertain in environmental problems, we propose to express predictions probabilistically (Reichert et al., 2015). The stakeholders' preferences are elicited over these potential outcomes of options. Stakeholder preferences are usually elicited from individuals in interviews or group workshops (Eisenführ et al., 2010) or increasingly online (Lienert et al., 2016; Aubert et al., 2020).

The defined structure and input data then allow us to build an evaluation model of the decision options (see section 2.2). Predictions and preferences are integrated and results calculated. As a result, each option receives a value or utility score allowing us to compare the performance of the options for each stakeholder perspective. Since any model-based analysis depends upon inputs, parameters, and model structure, further analysis of the model and conclusions, for instance, with sensitivity analyses, is crucial.

Based on the analysis, we can recommend one or several options to implement. Negotiation or consensus finding among stakeholders as well as implementation and monitoring of a decision demand additional activities (see Gregory et al., 2012; French and Argyris, 2018).

2.2. Decision models based on multi-attribute value theory

The theoretical underpinning of ValueDecisions is multi-attribute value theory (MAVT; Keeney and Raiffa, 1993), which is based on few rationality axioms. It has several advantages as a concept for environmental decision support (discussed in Reichert et al., 2015). MAVT is based on the concept of value functions. A value function maps from consequences, as measured by attributes, to the degree of relative achievement of objectives, usually in the interval [0,1]. The larger the value, the better this option meets the objectives, given the stakeholders' preferences. The options can then be ranked according to their values. In case of uncertain outcomes, a utility function provides this ranking of options. In a rational decision, the option with the highest value should be chosen. When a decision problem covers multiple objectives, a multi-attribute value function is built, which returns a single value based on the consequences of each option on several attributes.

A stepwise, hierarchical procedure helps constructing this multi-attribute value function: (1) determine an objectives hierarchy to structure the model, (2) identify shape and parameters of the lowest-level value functions (i.e., marginal value functions), (3) find the type and parameters for aggregating values upwards in the hierarchy to calculate the overall value, and, optionally, (4) convert values into utilities. Below, we briefly explain each concept; see Reichert et al. (2013, 2015) or Haag et al. (2019a) for an in-depth treatment. We propose to create a separate decision model for each stakeholder and compare the outcomes, though ideas to aggregate over stakeholders also exist (see section 3.4.5).

2.2.1. Objectives hierarchy

Objectives can be structured in form of a hierarchy (Keeney, 1992). This also determines the structure of the evaluation model. In many cases, it is advisable to include no more than 10–15 objectives (Marttunen et al., 2018) and reduce a higher number of objectives (Marttunen et al., 2019). An example objectives hierarchy is given in the case study (Fig. 3).

2.2.2. Lowest-level value functions

Each lowest-level objective is evaluated with respect to its attributes with help of a value function. This leads to a non-monetary common unit that allows comparing different types of attributes (e.g., costs in € with high phosphorus recovery in %). For simplicity, we only consider lowest-level value functions v_i for single attributes x_i, not over several attributes. Such a value function, $v_i(x_i)$, is defined over the range which the attribute can take in the decision case covering all options (e.g., costs from 90 € to 150 €). A value of 0 corresponds to the worst possible level given the range of the attribute (e.g., costs of 150 €) and a value of 1 to the best possible level (e.g., costs of 90 €). The value function thus measures the relative degree of achievement of an objective.

The shape of the lowest-level value functions depends on stakeholder preferences. Sometimes, attribute levels may be translated linearly to values (Fig. 1, lower panels), but often stakeholders have non-linear preferences (e.g., Scholten et al., 2015; Zheng et al., 2016; Langhans and Lienert, 2016). Several methods for eliciting the shapes of the value functions exist (Eisenführ et al., 2010). Commonly, we will obtain pairs of outcomes and corresponding values and can then interpolate between them or fit a parameterized function.

2.2.3. Aggregation model

To calculate the total or overall value $v(\boldsymbol{x}_a)$ of an option a across all consequences $\boldsymbol{x}_a = x_{1,a}, \ldots, x_{n,a}$, the performance on all objectives is aggregated.This means, the values of lower-level objectives $v_i(x_{i,a})$ are aggregated to a value at the next-higher level of the hierarchy, and so forth, until a single total value $v(\boldsymbol{x}_a)$ for each option a is reached (see Reichert et al., 2013, 2015; Haag et al., 2019a).

The most commonly used aggregation model for MAVT is the additive model (Belton and Stewart, 2002). Here, the weighted arithmetic mean is used to calculate the aggregated value $v(\boldsymbol{x})$:

$$v(x_1, \ldots, x_n, \boldsymbol{w}) = \sum_{i=1}^{n} w_i \cdot v_i(x_i) \qquad \text{(Eq. 1)}$$

where $v(x_1, \ldots, x_n)$ is the total value over the consequences $x_1, \ldots x_n$, $v_i(x_i)$ is the value for the consequence measured by attribute i, and w_i is a weight parameter for attribute i with $\sum_{i=1}^{m} w_i = 1$. Thus, in addition to the lowest-level value functions, this model requires to specify weight parameters (also called scaling factors). Weights can strongly affect the results (see section 4.3.2), and weight elicitation is especially prone to biases (Morton and Fasolo, 2009; Riabacke et al., 2012). Therefore, careful weight elicitation with tested methods, such as swing or trade-off, is advisable (Eisenführ et al., 2010). Elicited weights may still remain ambiguous, for example, because stakeholders may be inherently uncertain, because preferences change in different future scenarios, or because mistakes during elicitation occurred. A possible procedure to investigate the effects of this uncertainty is local sensitivity analysis on each weight (Eisenführ et al., 2010; section 4.3).

The additive model is not always an adequate representation of stakeholder preferences. An implication of the additive model is the possibility to compensate: If one objective performs poorly (e.g., high phosphorus recovery is not achieved), it can be compensated by another well-performing objective (e.g., high water savings) according to the given weights. Practical experience in many cases indicates that stakeholder preferences actually do not agree with some assumptions and implications of the additive model (e.g., Rowley et al., 2012; Langhans and Lienert, 2016; Zheng et al., 2016; Haag et al., 2019a; Reichert et al., 2019). Other non-additive aggregation models have less strict assumptions and allow for interaction between objectives (Grabisch et al., 2009). However, non-additive aggregation is so far rare in MCDA software for MAVT/MAUT and not even mentioned in the reviews by Mustajoki and Marttunen (2017) or Weistroffer and Li (2016). Exceptions are Logical Decisions (Logical Decisions, 2020), which supports multiplicative aggregation and the tool by Cinelli et al. (2021), which supports several mean functions.

An interesting alternative to the weighted arithmetic mean for aggregation in MAVT is the weighted power mean, also called weighted generalized mean (Haag et al., 2019a; 2019b). Using this aggregation function, we are able to model various interactions between objectives, such as partial non-compensation. To our knowledge, ValueDecisions is the first MCDA software with graphical user interface that implements this aggregation. The power mean is given by:

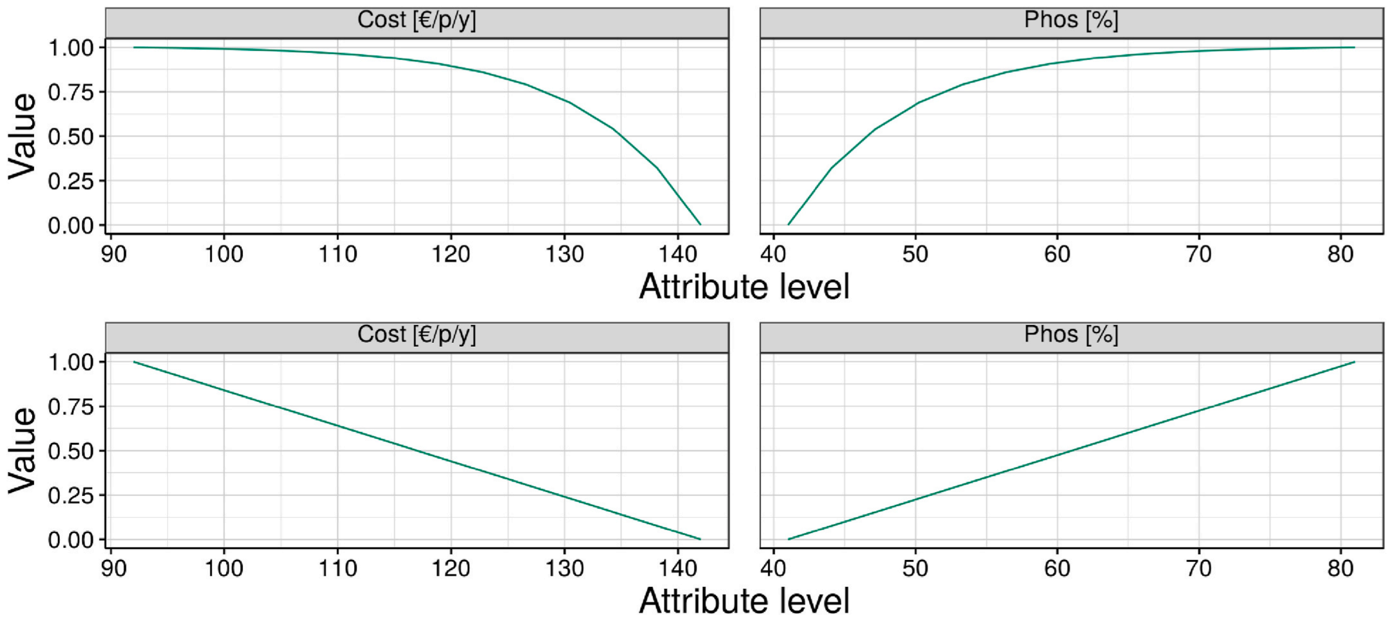

Fig. 1. Example of linear (lower panels), and exponential (with curvature c = 5, top panels) value functions (see SI-2 for equations) for the objectives low cost and high phosphorus recovery. Attribute levels (x-axis; e.g., costs in € per person and year) are mapped to a value (y-axis) in [0,1].

$$v(x_1,\ldots,x_n,\boldsymbol{w},\gamma)=\begin{cases}\left(\sum_{i=1}^{n} w_i\cdot v_i(x_i)^{\gamma}\right)^{\frac{1}{\gamma}} & \mathit{for}\ \gamma\neq 0\\ \prod_{i=1}^{n} v_i(x_i)^{w_i} & \mathit{for}\ \gamma=0\end{cases} \quad \text{(Eq. 2)}$$

The weights and value function parameters are as in Eq. (1). We informally refer to the additional parameter γ as the non-additivity parameter, because for $\gamma = 1$, the model reduces to the additive model and by varying γ we can express different non-additive behaviors. The weighted power mean covers a number of other aggregation functions of interest as special cases for certain values of γ (see SI-1). For an in-depth discussion of their properties, we refer to Langhans et al. (2014) and Grabisch et al. (2009).

2.2.4. *Uncertainty and risk: utility functions*

When outcomes of a decision are uncertain, a stakeholder's risk attitude determines how to evaluate this uncertainty. This can be captured with multi-attribute utility theory (MAUT; Keeney and Raiffa, 1993). Directly eliciting multi-attribute utility functions from stakeholders is demanding (Eisenführ et al., 2010), and to our knowledge rarely done in complex practice applications. However, it is possible to transform the value function at the highest level of the objectives hierarchy to a utility function (Dyer and Sarin, 1982; Reichert et al., 2015). Instead of calculating the value of options, the expected utility is calculated. This integrates over the uncertain predictions, taking into account the risk attitude and leads to a unique ranking based on the expected utility. In practice contexts, we found it can be advantageous to work with values and explicit visualization of their uncertainties instead of discussing expected utilities.

3. The ValueDecisions app

3.1. *Overview*

ValueDecisions supports multi-criteria decision analysis modeled with multi-attribute value theory. It provides a middle ground between sophistication and user friendliness, is flexible to extend and adapt, and is open source. Programmed in R (R Core Team, 2021), it is accessible to users online as a web application that does not require software installation. Additionally, it is available as a standard R package, which can be run locally. The users (section 3.2) need to be familiar with the principles of MCDA to sensibly use ValueDecisions, but programming capabilities beyond uploading two spreadsheets in the required format is not required. ValueDecisions is available at https://www.eawag.ch/en/department/ess/main-focus/decision-analysis-da/tools/.

ValueDecisions focuses on the MCDA modeling stage (i.e., the integration of predictions and preferences). This includes calculating the MCDA result (i.e., identifying the best performing options) and sensitivity analyses to further explore model inputs, assumptions, and results. ValueDecisions allows setting up and running a MCDA model in a graphical user interface for complex decision problems, including uncertainty of predictions. Many options can be evaluated in parallel for multiple stakeholders with possibly conflicting preferences. Results are presented in tables and visualized by various graphs. A focus is on diverse visualizations of results that can be explored interactively, for instance, in a workshop. Users can easily change preference parameters to analyze the sensitivity of the results in steps, including shapes of lowest-level value functions, non-additive aggregation models, and local sensitivity analyses of weights. They can download intermediate and final results as individual graphs and tables or compiled in a report.

3.2. *Target users and use cases*

We can differentiate between three potential user groups for MCDA software (Mustajoki and Marttunen, 2017): (1) "do-it- yourself" users (Belton and Hodgkin, 1999) who would like to apply MCDA, but have no specific training or experience, (2) analysts and facilitators who need software support for facilitation, analysis, and visualization, and (3) analysts and consultants who want to carry out sophisticated analyses.

ValueDecisions is mainly targeted for the latter two user groups. It can support consultants working in environmental decision cases, public planning authorities, municipalities, other practitioners, or students in the analysis phase of a MCDA process. These users may lack finances, time, or expertise, to work with sophisticated software or programming (R, Python, Matlab, etc.), but wish to go beyond the limited capabilities of spreadsheet software. Additionally, ValueDecisions is powerful enough that academic researchers can explore applied MCDA problems. The app can be used in workshops with interactive exploration of options or in backroom analyses. Since ValueDecisions requires a carefully conducted problem structuring and data collection phase to have taken place, it is not meant for "do-it-yourself" users. That said, if stakeholders are provided the input files, they can further explore the decision model, for example, to facilitate transparency and traceability (Voinov et al., 2016).

3.3. Elements of MCDA model and their implementation

3.3.1. Input data on predictions and preferences

Users need to upload a file with information on the predicted performance of each decision option for each attribute ("predictions") and a file with stakeholder preferences about these predictions ("preferences"). These are uploaded in two spreadsheets (excel or tab-separated values). In the files, the users enter all information for the analysis; templates for the files can be downloaded. Once uploaded, ValueDecisions first conducts several checks for file validation.

The predictions file (example in Table SI–1) specifies for each option the predicted performance of each attribute. Since the consequences of options for environmental decisions are usually uncertain, probability distributions for the predicted outcomes can be specified for each option and attribute (see section 3.3.2).

The second file (example in Table SI–2) specifies the structure of the objectives hierarchy of stakeholders used for evaluation, as well as information on stakeholders' preference parameters. Only the objectives hierarchy is essential to run ValueDecisions, including the names of upper-level (L2) and lower-level (L1) objectives, attribute names and units, and the best- and worst-possible outcome of attributes. Each stakeholder can have a different objectives hierarchy, however, the interpretation of results can become more complex if stakeholders do not share objectives. In addition to the objectives hierarchy, preference parameters for several different stakeholders can be defined, including the shape of lowest-level value functions, global weights assigned to attributes, and type of aggregation to be used in the model. If no preference data are specified or a group of parameters is missing, defaults are assumed. These are that lowest-level value functions are linear, attributes have equal weights, and the aggregation function is the weighted arithmetic mean (additive model). This can be useful for interactive use of ValueDecisions. For instance, in a workshop, the options can be explored together with the stakeholders and the preference parameters can be varied directly in the app interface.

3.3.2. Uncertainty of predictions

In many environmental and public policy decisions there is large uncertainty about the consequences of management options. In ValueDecisions, users can either calculate the results based on point predictions for each attribute and option, or provide parameters of probability distributions for these predictions. Implemented probability distributions are uniform, normal, lognormal, triangular, beta, generalized extreme value distributions, and discrete distributions. ValueDecisions will generate samples from the specified distributions and propagate the uncertainty to the results with Monte Carlo simulation. This leads to a distribution of the total value $v(\boldsymbol{x}_a)$ of each option and to a distribution of rankings. Currently, only independent probability distributions can be used, i.e., the predictions of one attribute cannot depend on the predictions of another. With other software (e.g., utility; Reichert et al., 2013) it is possible to work with predicted samples of non-independent distributions; however, this requires that users create these samples beforehand.

With some creativity, it is possible in ValueDecisions to combine MCDA with scenario analysis to account for large future uncertainty (e.g., Stewart et al., 2013; Scholten et al., 2015; Zheng et al., 2016). The performance of options given a scenario can be entered as separate options (e.g., option "A" would occur twice in the prediction file, as "A_scenario1" and as "A_scenario2") and the results can be analyzed across scenarios. Alternatively, different predictions and/or preferences for each of the future scenarios can be uploaded and the results compiled outside of ValueDecisions.

3.3.3. Lowest-level value functions

For each attribute on the lowest level of the hierarchy, outcome-value pairs that were elicited from stakeholders can be provided. Using least-squares, ValueDecisions then fits a single-attribute value function to these outcome-value pairs. The fit can be evaluated visually. For lowest-level value functions, only single-attribute value functions, not multi-attribute value functions, are possible. A wide variety of shapes can be specified, such as linear, exponential, or sigmoid functions (see SI-2). Alternatively, linear interpolation between provided values or discrete value functions can be used. Thus, users can represent many types of non-linear preferences with fewer restrictions than in many other software products. If no preference information is provided for an attribute, linear value functions are assumed as default. The effect of this assumption can be explored with sensitivity analysis (see section 4.2.2).

3.3.4. Aggregation for multi-attribute value functions

In ValueDecisions, the objectives hierarchies used for evaluation can have two levels. As default, ValueDecisions uses the additive model to calculate overall values. The weight parameters for objectives need to be elicited from the stakeholders. If no preference information for weights is provided, equal weights are assumed for all attributes.

Because ValueDecisions also implements the weighted power mean as aggregation function (Eq. (2)), many alternative ways for aggregation besides the additive model can be used (see SI-1). This flexible approach to aggregation together with non-linear lowest-level value functions allows representing complex preference structures. For each branch of the objectives hierarchy the γ parameter, which specifies the degree of non-additivity, can be entered in the preferences input file in addition to the weights. Alternatively, within ValueDecisions a slider allows adapting the value of the parameter γ across the entire hierarchy, allowing to explore the effects of using different weighted power mean functions.

3.4. Features and flow

3.4.1. Structure and controls

To users, ValueDecisions is essentially a website. The basic flow of the app is organized as consecutive pages. Users can switch between pages via a navigation menu on top (Fig. 2). On the Start and Information pages, information and checklists are provided. On the Upload page, users upload the two input files to the analysis (predictions and preferences). On the Analysis page, an interactive MCDA analysis is run, and on the Reporting page, a report containing graphics and tables of current results can be downloaded, for instance, as word file. The resulting plots and tables of all steps of the analysis can also be saved individually.

The heart of ValueDecisions is the analysis page (Fig. 2). On its left side, a control panel allows users to interact with the analysis. On the right side, users can switch between different types of analyses and displays that are organized in tabs. On the control panel, a Run/Update button re-runs the analysis after adjusting settings. It is possible to switch between analyses with or without uncertainties in the predictions and adjust the number of Monte Carlo samples.

Three filters can be applied to select:

1. stakeholders, for which results should be calculated;
2. decision options, for which results should be calculated;
3. objectives, for which results should be shown. This only applies to certain plots.

The following parameters of the preference models can be changed. This can be combined in arbitrary ways:

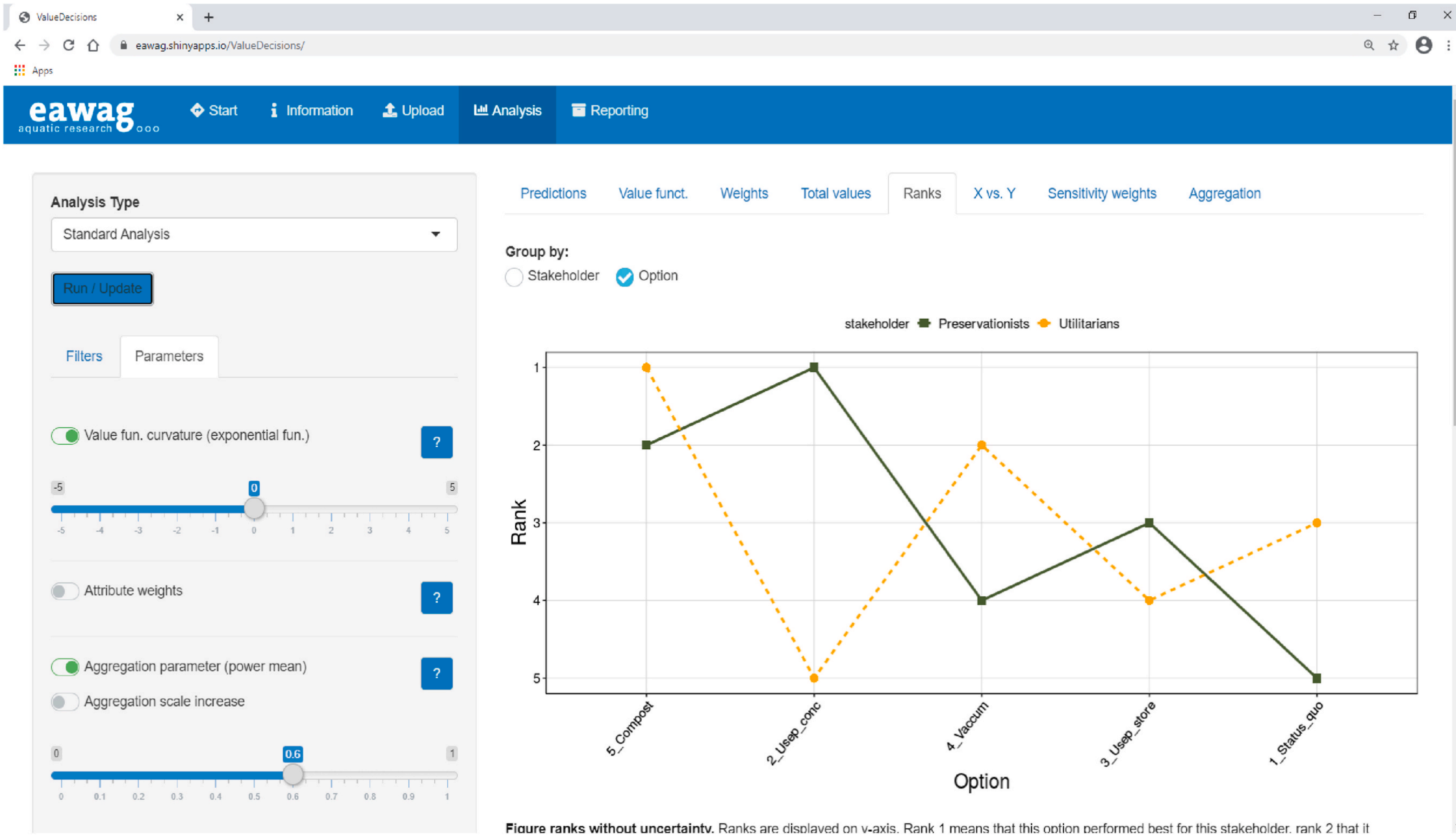

Fig. 2. Screenshot of the ValueDecisions app. For a description of the pages and analysis tabs (on the top), and the filters and parameters (left side panel) see text. In the depicted analysis, the ranks (y-axis) that each option achieved (x-axis) are shown, given two weight profiles (Preservationists, Utilitarians; colored lines in plot, see section 4.3.2), where rank 1 = best-performing, and rank 5 = worst-performing option. In this example, the aggregation model was changed (left side panel) as part of a possible sensitivity analysis.

1 The shape of the lowest-level value functions can be changed collectively. If this is active, *all* lowest-level value functions for *all* stakeholders are converted to exponential value functions with the curvature parameter *c* specified by a slider (see Eq. SI-3 for formula).
2 The weights assigned to objectives can be adjusted individually. Users can select *individual* objectives and give these a new weight using a slider. The newly assigned weight is applied for *every* stakeholder. The other weights, which were not modified, are renormalized, so that the previous relationships remain unchanged and the sum of weights equals one. Consequently, the newly assigned weight will be the same for all stakeholders, but the other weights remain stakeholder-specific.
3 The aggregation function can be changed collectively. If this is active, the aggregation model specified in the preferences input file (or without specification, the default additive aggregation) is switched to the weighted power mean (section 2.2.3). The value of the non-additivity parameter γ can be adjusted with a slider (see SI-1 for implications). The same value is assumed for all stakeholders and across all branches.

In addition to the control panel, there are controls in the tabs that show the analysis results. Here, users can switch between different presentations of the results, for instance, different types of graphs, or switching the x-axis variable.

3.4.2. Visualization of input data and intermediary steps

Exploratory visualization of the input data before calculating the full MCDA model allows checking for mistakes in data entering and already gives insights into differences between stakeholders and options. The structure of the evaluation is given by the hierarchy of objectives (example in Fig. 3). The first tab on the Analysis page shows the predicted performance of each option. When uncertainties are considered, the empirical distributions of the samples are shown, as obtained by Monte Carlo sampling (example in Fig. SI-2). On the next tab, the lowest-level value functions that were fitted to the elicited preference data or modified in the app are displayed (example in Fig. 1). A third tab shows the weights the stakeholders provided or that were specified in the app (example in Fig. 4).

3.4.3. MCDA results

The next three tabs of the Analysis page provide results of the decision model and additional analysis. There are several options for different types of graphics. In the Values tab, the overall and partial values of each option are displayed (example in Fig. 5). Visualization options differ if uncertainty of the predictions is included. In the Ranks tab, the overall rankings of the options are shown (example in Fig. 2) or the rank distributions in case of uncertainty (example in Fig. 9 and Fig. SI-7). In the X vs. Y tab, the resulting values for selected groups of objectives can be related to each other. The most common example is plotting costs against benefits (Fig. 10). The efficient frontier of cost-efficient options can also be shown.

3.4.4. Sensitivity analyses

ValueDecisions allows extensive step-wise sensitivity analyses and robustness analyses (examples in Schuwirth et al., 2012; Zheng et al., 2016; Haag et al., 2019b). One way to conduct this is via the controls for preference parameters (left side panel, Fig. 2). They can be varied interactively and in arbitrary combination by users to study the impacts on the results. It can be tested whether results are sensitive to (a) the assumption of linear lowest-level value functions or other functional shapes, (b) changes of individual or combinations of weights, and (c) the choice of the aggregation model. For didactic reasons, the effects of choosing a non-additive aggregation model can also be visualized for two attributes (as in Reichert et al., 2013).

Because results are often sensitive to weight parameters, an additional local sensitivity analysis for the weight of each objective is possible (example in Fig. 7). It shows the resulting overall values of the options when an objective's weight is varied from 0 (this objective is not at all relevant in this context) to 1 (only this objective is relevant, none of the other objectives matter). A more advanced approach, the weight stability interval (WSI), allows calculating intervals in which (partial) orders of options are stable (Mareschal, 1988). In ValueDecisions, such intervals can only be visually read off the figures (Fig. 7).

3.4.5. Comparing stakeholder groups

Preferences within and across groups can be considered in different ways in decision modeling (see, e.g., Eisenführ et al., 2010) and resulted in corresponding software implementations (Mustajoki and Marttunen, 2017). One option is to determine aggregated preference profiles. The simplest approach is to average preference parameters across stakeholders, for example, aggregating individual weights using a weighted mean or median (also see section 4.2.1). This can be sensible if we are interested in generalized preferences of large groups, such as a population sample. Another possibility is to aggregate the utility or value of decision options across stakeholders (Keeney, 2013). In environmental decisions, stakeholder groups commonly have conflicting or opposing views (Gregory et al., 2012). This calls for an approach that allows for comparisons by treating the preferences of individual stakeholders (or groups) separately. We need to see in how far conflicting preferences affect the results. This can then point towards leverage points for compromise options or be an input to further stakeholder deliberations. Learning about the stakeholders' opinion can often more important than calculating a final ranking of options (e.g., Mustajoki and Marttunen, 2017).

ValueDecisions follows this latter approach. An overall "consensus" is not enforced by the modeling approach, but instead several stakeholders can be analyzed individually and compared. Interactive variation of weights and other preference parameters (see section 3.4.4) can facilitate finding paths to consensus options. The explicit visualization of conflicting results can be a way forward to further negotiation and deliberation. This would be lost by averaging over the conflicting views.

3.5. Implementation as a reactive web app based on R

ValueDecisions is written in the programming language and software environment R (R Core Team, 2021). The MCDA modeling and calculations partly build on our earlier work (e.g., Haag et al., 2019b) as well as the R package "utility" (Reichert et al., 2013). Being open source and using R means that the app is relatively easy to change and extend by users familiar with R, in comparison to proprietary software using bespoke frameworks and data structures. The interactive graphical user interface in the form of a web application is achieved with the "shiny" framework (Chang et al., 2021). In addition, we use several extension packages to shiny for visualizing tables, different types of widgets, and others. Most figures rely on "ggplot2" (Wickham, 2016). The reporting as downloadable word file was achieved with "Rmarkdown" (Xie et al., 2018). The creation of an R package followed the "golem" framework (Fay et al., 2021).

More generally, the combination of specifying analysis in a programming language common in science (e.g., R, Python, or Julia) in the backend of an application and then building a web app to make the analysis available to users or stakeholders by a graphical user interface is a promising development model. It offers many possibilities for scientific work but also for interactive communication of scientific results to stakeholders.

Shiny applications are based on reactive programming which is different to an imperative "if this do that" programming style. In a shiny application, we define the endpoints (outputs) and their relation to sources (inputs) via conductors (intermediate steps). In other words, we define the relationship of elements, for instance, from input files via calculations to a graph with results. However, we do not specify the exact time or order that calculations should occur in. Instead,

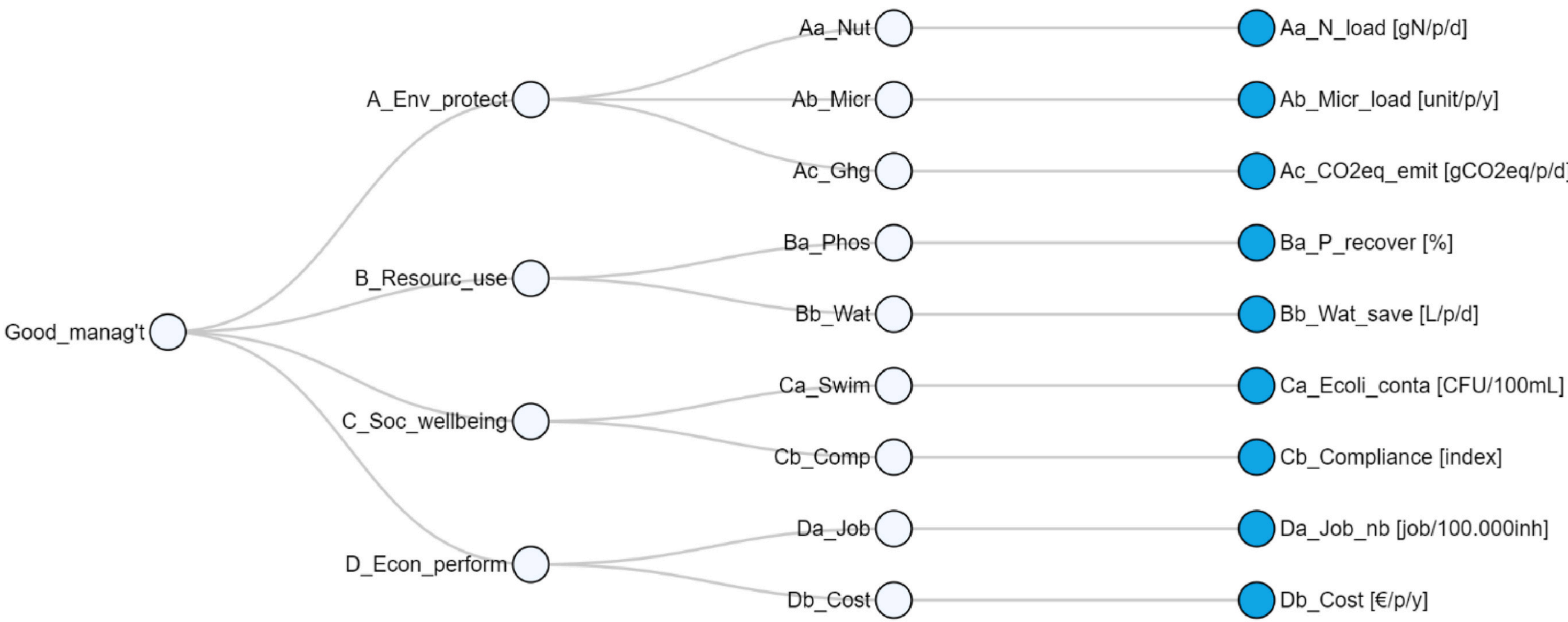

Fig. 3. Objectives hierarchy as displayed in ValueDecisions for deciding about Good management (Good_manag't) of wastewater in the Paris region. The lower-level objectives "low discharge of nitrogen (nutrients) to the rivers and air" (Aa_Nut), "low load of micropollutants in the rivers and soils" (Ab_Micr), and "low greenhouse gas emission" (Ac_Ghg) define the upper-level objective "high natural environment protection" (A_Env_protect). "High phosphorus recovery" (Ba_Phos) and "high water saving" (Bb_Wat) define "high sustainability in natural resource use" (B_Resourc_use). "High chance of compliance by end-users" (Ca_Comp) and "high possibility of swimming in rivers" (Cb_Swim) define "high societal well-being" (C_Soc_wellbeing). "High number of local jobs" (Da_Job) and "low cost" (Db_Cost) defined "high economic performance" (D_Econ_perform). Each objective is associated with an attribute and its unit in brackets (right). Table SI–3 further describes the meaning of the abbreviations.

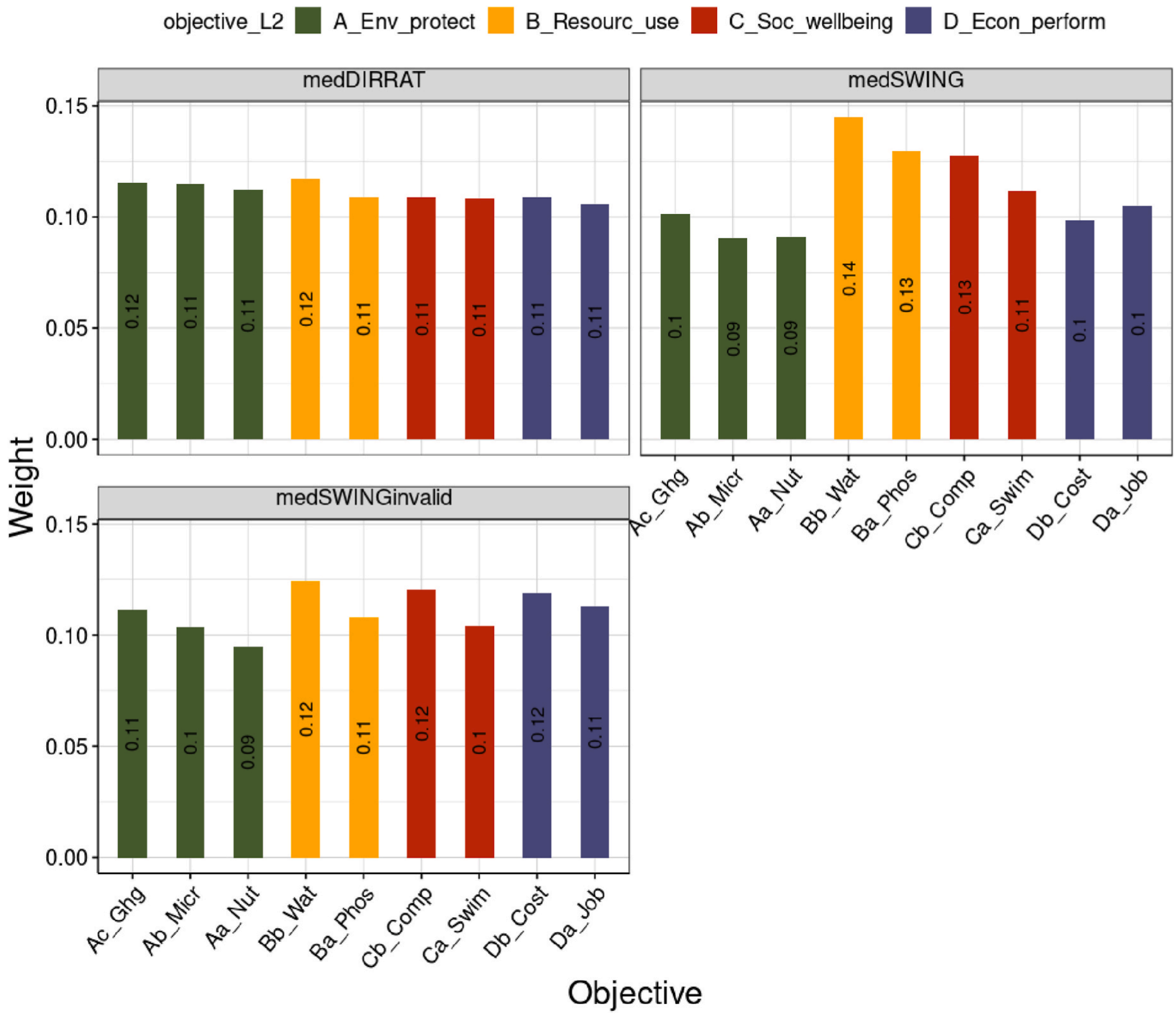

Fig. 4. Median weights (y-axis) for each lower-level objective (x-axis) for the three online survey groups using direct rating (medDIRRAT; N = 357), swing (medSWING; N = 36), or not complying with the swing instructions (medSWINGinvalid; N = 262). From left to right, the color or shade of the bar indicates the correponding top-level objective, namely green: environmental protection (A_ …); orange: resource use (B_ …); red: social well-being (C_ …); blue: economic performance (D_ …). Explanation of abbreviations are given Fig. 3 and Table SI–3.

calculations are run when they need to: ValueDecision will recalculate and update only those outputs that are affected by user changes. This is useful for interactive analysis as it lowers the required amount of recalculation. Because reactive programming focuses on the relation of elements, it is relatively easy to add or remove components without affecting the whole.

ValueDecisions can be accessed as a web app or downloaded as a standard R package to be executed locally. Building upon R makes the app platform independent, it can be run on Windows, Linux, or macOS systems. Deployment as a web app has advantages for users, since they can use ValueDecisions if they can run a modern web browser on their device, independent of operating systems, hardware, and other installed software. Because ValueDecisions is essentially a website, its design is responsive and it can also be viewed on devices such as tablets or smartphones.

4. Application: MCDA for wastewater infrastructure planning in the Paris region

4.1. Context and decision structure

The rivers of the greater Paris region face increasing pressure, partly caused by the wastewater management system (Esculier et al., 2015). Two factors contribute, typical for other large cities worldwide (Lossouarn et al., 2016): Firstly, continuous population growth leads to increases in wastewater production. Thus, the pollution load entering wastewater treatment plants increases. Secondly, regional river discharge is predicted to decrease due to climate change (Flipo et al., 2021). This will lead to reduced dilution of the treated wastewater in receiving rivers. Currently, wastewater management in the Paris region complies with the French transposition of the Water Framework Directive on nutrient concentrations (SIAAP, 2021). However, given the pressures of population growth and climate change, the main Paris wastewater management authority has to invest significantly to avoid serious future damages of water ecosystems. Today's point in time is also suitable to make changes because of new infrastructure projects in the Paris region (“le Grand Paris” project; DILA, 2021).

Innovative technical options have the potential to address the pressures at hopefully reasonable costs, but entail radical system changes (e.g., Larsen et al., 2016; Hoffmann et al., 2020). They imply moving from the centralized system, where wastewater is transported in sewer networks to large treatment plants, to semi- or fully decentralized systems, where wastewater is treated locally, for instance, in individual buildings. Waste streams, like urine, feces, and greywater, may be separated at the source, which has technical advantages and allows resource recycling (Larsen et al., 2009).

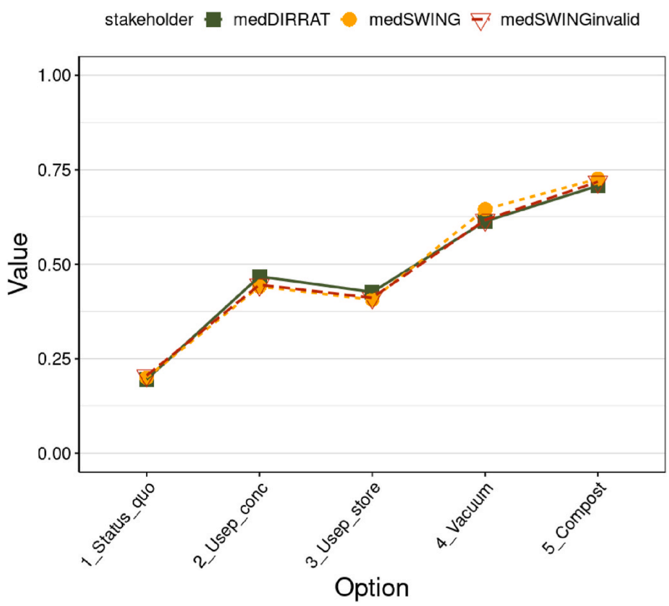

Fig. 5. Overall values for the different options obtained by the additive model in ValueDecisions. Values (y-axis) for each option (x-axis) for the three groups' weight profiles (symbols). Values range from 0 (worst possible achievement given the ranges of attributes) to 1 (best possible achievement given the ranges of attributes), given the predictions and the preferences of the three stakeholder groups.

Decisions concerning future wastewater management inevitably affect citizens. Before deciding, the main Paris wastewater authority needs to know how the public would perceive unconventional decentralized options. We conducted a study to evaluate options for urine and feces management for the Paris region. A workshop with 18 local stakeholders was held in June 2016 to select objectives for

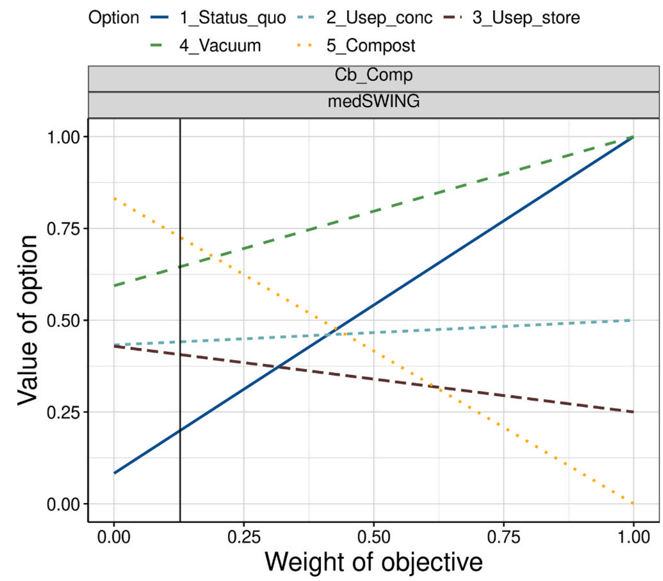

Fig. 7. Local sensitivity analysis of weights: value (y-axis) of the options (lines) varying as a function of the weight (x-axis) of the objective "High chance of compliance by end-users" (Cb_Comp). At weight 0, Cb_Comp is completely unimportant; at weight 1, Cb_Comp receives all the weight (and the eight other objectives 0 weight). The black vertical line indicates the weight currently given to this objective in the medSWING weight profile (0.13).

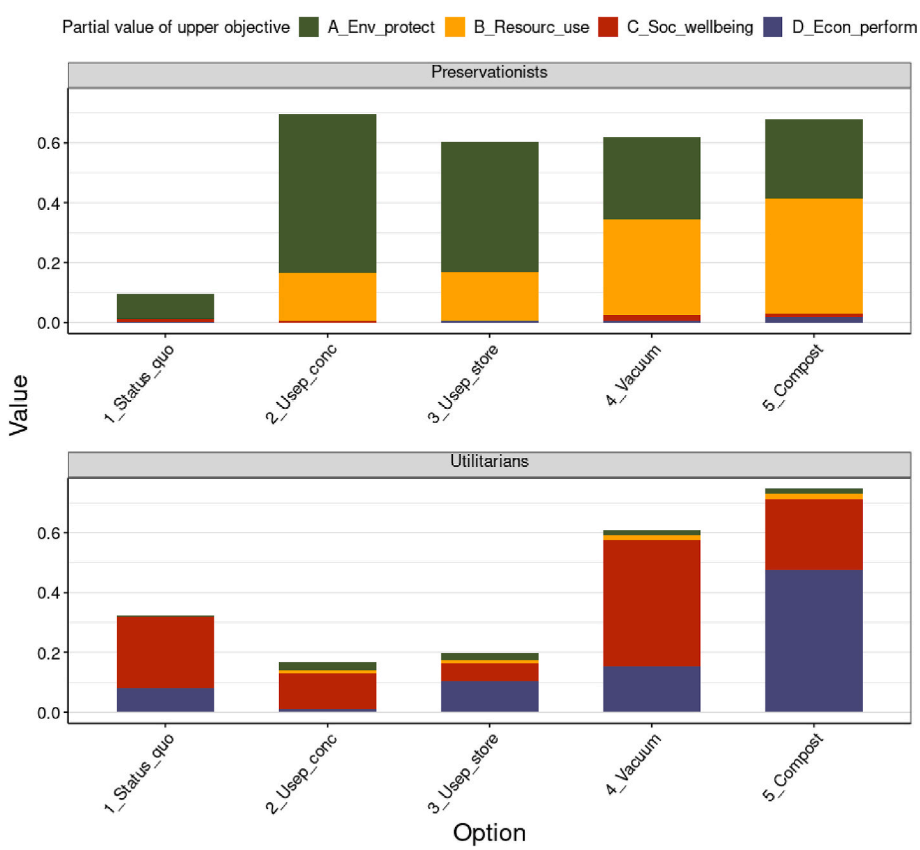

Fig. 6. Values (y-axis) for each option (x-axis) for the two groups with extreme weight profiles, preservationists (top panel) and utilitarians (lower panel). The colors or shades in the bars indicate the values achieved by the corresponding upper-level objective.

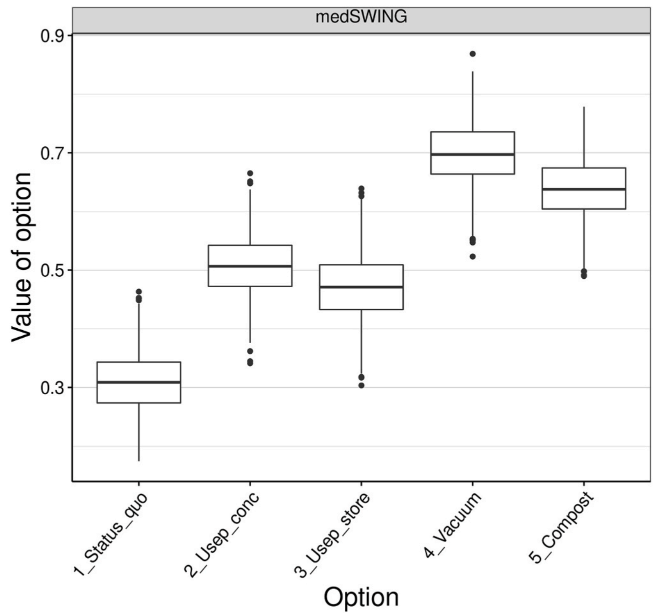

Fig. 8. Value (y-axis) of five options (x-axis) for the medSWING group, with uncertainty of the predictions with 2000 simulation runs. The boxplots show the 0.25 (lower), 0.5 (median), and 0.75 (upper) quartiles of the results concerning values. The whiskers extend to the minimum and maximum values within 1.5 times the interquartile range. Points outside the whiskers are outliers.

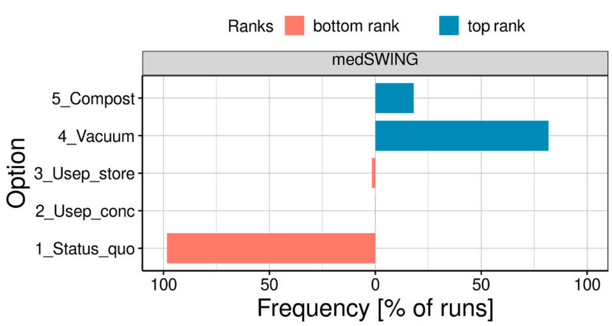

Fig. 9. Frequency (x-axis) for each option (y-axis) to be ranked first (blue bars to the right) and last (red bars to the left) for the 2000 simulation runs for the medSWING group preferences. Reading example: option 4_Vacuum and achieved the first (best) rank in more than 75% of the runs and never the last (worst) rank.

evaluating wastewater management options in new districts and to elaborate innovative options (details in SI-4.1). This process led to agreeing upon nine fundamental objectives and five wastewater management options. The five selected options are: status quo (or business as usual; 1_Status_quo), urine source separation with on-site concentration (2_Usep_conc), urine source separation with on-site storage (3_Usep_store), urine and feces collection in a separate vacuum network with decentralized treatment (4_Vacuum), and dry toilets with on-site underground composting chambers (5_Compost). The nine objectives were organized in a two-level objectives hierarchy (Fig. 3). The decision case was also the focus in an experimental paper addressing online preference elicitation for MCDA (Aubert et al., 2020). Here, we present novel results based on a representative population survey in the Paris region.

We deployed two online population surveys to collect citizens' preferences concerning the importance of objectives and trade-offs they are willing to make (which can be expressed by weights). The surveys, using swing weight elicitation and direct rating (Eisenführ et al., 2010), are presented in Aubert et al. (2020). The obtained samples were

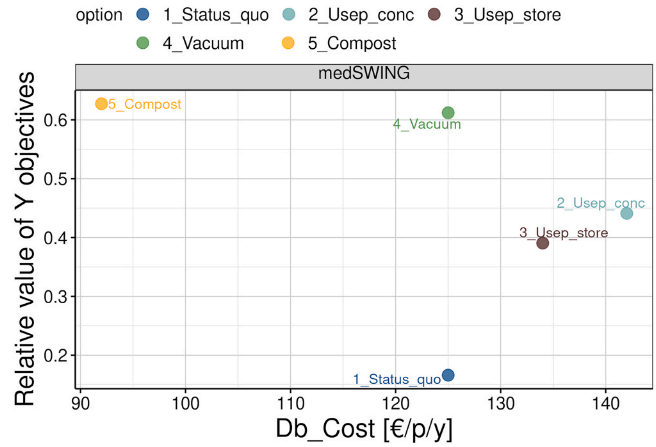

Fig. 10. Cost-benefit visualization. Relative value of the five options for the following eight objectives (Aa_Nut, Ab_Micr, Ac_Ghg, Ba_Phos, Bb_Wat, Ca_Swim, Cb_Comp, and Da_Job) (y-axis) as a function of cost in € per person and year (x-axis).

representative of the regional statistics in terms of gender, age, and occupation. The data from the population survey and the data on the predictions, i.e., how the options perform on each attribute were reformatted as required by ValueDecisions (see examples: predictions Table SI–1, preferences Table SI–2); these input files are available in the data package (see Data Availability).

We aimed to generate advice for the main Paris wastewater authority who needs to know how the public would perceive unconventional decentralized wastewater management options. The main question in this study is: Are the results of the MCDA robust concerning the best-performing wastewater options, despite possibly differing preferences of the sampled population and various uncertainties? Which wastewater options can we recommend to the Paris authorities? Using ValueDecisions, we analyzed the decision problem including a comprehensive sensitivity analysis. We investigated how robust the outcomes are to the weights, the non-elicited preference parameters, the uncertainty of the predictions, and visually explored costs of options versus their benefits.

4.2. *Analyses*

4.2.1. *Weight parameters and their influence*

First, we investigated whether the results of the MCDA differed between preferences of survey respondent groups of the Paris region population. One survey group did a direct rating of objectives (DIRRAT; N = 357), and the other went through a swing weight elicitation. Within the latter group, we discovered that 262 respondents did not understand the procedural instructions (discussed in Aubert et al., 2020). We treated this group separately (SWINGinvalid, N = 262) from the group that understood the instructions (SWING, N = 36). To obtain generalized weights, we aggregated the individuals' weight for each objective by taking the median weight for each of the three groups. To obtain a median weight profile for each group, we normalized the nine median weights to sum to one. To check the sensitivity of the MCDA results to the weight profiles, we ran the analysis in ValueDecisions handling the three groups as three stakeholders.

Secondly as a sensitivity analysis, we created two fictional, extreme weight profiles, as the weight profiles from the survey were similar between groups (Fig. 4). These profiles are based on the idea of environmental attitudes (Milfont and Duckitt, 2006) stating that one can have a preservationist or utilitarian environmental attitude. In the preservationist profile, we equally distributed the weights between the objectives of protecting the environment and sustainable resource use, ignoring the actual ranges of the attributes. In the utilitarian profile, we equally distributed the weights between social well-being and economic performance. We ran the sensitivity analysis in ValueDecisions handling these two extreme weight profiles as two stakeholders.

Thirdly, we used the median weights of the SWING group (medSWING) to carry out a local (one factor at a time) sensitivity analyses (e.g., Eisenführ et al., 2010) for two critical objectives. First, we doubled the weight of the low cost objective, which is often decisive for the authorities and can be systematically underestimated in MCDA (Marttunen et al., 2018). Additionally, using the local weight sensitivity analysis in ValueDecisions, we investigated the results for all possible weights for the objective high chance of compliance by end-users, which might be the most critical to achieve a paradigm shift in wastewater management.

4.2.2. Non-elicited preference parameters: lowest-level value functions, aggregation model

For time reasons, it was not possible to elicit lowest-level value functions and test whether the implications of using the additive model comply with the respondents' preferences in the survey. To test the robustness of the MCDA results, we carried out sensitivity analysis changing the two standard assumptions of using linear lowest-level value functions and the additive model.

We used the weight profile of the SWING group (N = 36) with median weights (medSWING) for this analysis. For the lowest-level value functions, we tested concave and convex shapes, using an exponential function and setting the curvature of all value functions to $c = -5$ or 5, respectively (see Eq. SI-3). We also repeated the analysis using the more general power mean aggregation to check the sensitivity of the results for following aggregation functions: additive (as comparison; the aggregation parameter γ being 1), approaching the weighted geometric mean ($\gamma = 0.04$), and exploring the space between additive and geometric mean (including $\gamma = 0.5$). Furthermore, we included the two extremes, maximum aggregation (γ approaching $+\infty$), and minimum aggregation (γ approaching $-\infty$).

4.2.3. Uncertainty of predictions

ValueDecisions enables exploring the effect of uncertain predictions by propagating the uncertainty to the results with Monte Carlo simulation. In our case, this is particularly relevant, given that high uncertainty was associated with some predictions (e.g., for the objective high number of local jobs, Fig. SI-2). We investigated how these uncertainties affect the outcome, to estimate whether it is worth, or necessary, to invest additional time to improve the predictions.

4.2.4. Cost-benefit visualization

ValueDecisions enables us to represent the aggregated value over any combination of objectives as a function of any other combination of objectives, such as low cost. A cost-benefit analysis is usually important for stakeholders. They need to know if paying more actually pays off, i.e., makes a difference in terms of achieving higher values on all other objectives. We used the "X vs. Y" analysis in ValueDecisions, selecting the objective low cost for visualization on the x-axis, and the aggregated value from the MCDA for the other eight objectives on the y-axis. Again, we used the SWING median weight profile.

4.3. Results and discussion of the analyses

4.3.1. Visualization of objectives, predictions, weights, and lowest-level value functions

ValueDecisions has various possibilities to visualize the input data as well as intermediate steps of the analysis. This is helpful to verify that data were entered correctly, but also to understand the results. The evaluation structure is given by the objectives hierarchy (Fig. 3). For the predictions, we checked that the data displayed in the graphs and the tables corresponded to our input files. For lowest-level value functions, we verified whether their shapes and slopes were as intended (example in Fig. 1). We also visualized the performance of all alternatives on the lowest-level objectives (Fig. SI-1). For the weights (Fig. 4) we checked correspondence of displayed weights with the original input file at the lower and upper level of the objectives hierarchy.

4.3.2. Effect of weight profiles

First, we compared the effect of the three groups' median weight profiles, SWING, SWINGinvalid, and DIRRAT, obtained through the survey. Weights were relatively uniform within each of the groups; in the most spread distribution (SWING), the weights varied from 0.09 to 0.14 (Fig. 4). The ordering of options and the obtained overall values was very similar across the three weight profiles (Fig. 5). Option 5 (dry toilet with composting chamber) had the highest overall values this was followed by option 4 (vacuum network with decentralized treatment). The worst performing option for all groups was the status quo. Despite somewhat different weight profiles, the MCDA results were thus similar for all three groups. Therefore, we focus on the SWING median weight profile for the following analyses.

As the median weights obtained by the survey were similar, we created two extreme weight profiles to explore possible effects of other weight distributions: preservationist (preserve environment) and utilitarian (high socio-economic performance; see section 4.2.1). With such contrasted weight profiles (Fig. SI-4), the values of options (Fig. 6) and their ranks of options differed greatly between the two groups. For preservationists, all options, except the status quo, performed similarly with values ranging from 0.62 to 0.70 (Fig. 6). For utilitarians, there was a similarly high achievement for option 5 (dry toilet with composting chamber; $v = 0.75$) and 4 (vacuum network with decentralized treatment; $v = 0.61$). This resulted from option 5 performing relatively well on all objectives, except on Social wellbeing (Fig. SI-1). However, for utilitarians the performance of option 3 (urine source separation with on-site storage; $v = 0.20$) and option 2 (urine source separation with on-site concentration; $v = 0.17$) was markedly poor compared to preservationists. The utilitarians' high weights on the objectives Social wellbeing and Economic performance, but low weights on Environmental protection and Resource use can explain this. Interestingly, the status quo option performed poorly for preservationists ($v = 0.10$), but considerably better for utilitarians ($v = 0.32$). However, despite extremely different weights, it was possible to find two consensus options for the groups, namely 4_Vacuum and 5_Compost.

Based on the assumption that the importance of low costs might have been underestimated by the population, but is decisive to authorities, we interactively doubled the weight of the cost objective from 0.1 to 0.2. For clarity of presentation, we focus on the medSWING original weight profile (Fig. 4) and the analysis without prediction uncertainty. The ranking of options remained the same despite this weight change and the values were of the same order of magnitude (Fig. SI-5): option 5 (0.76) > option 4 (0.61) > option 2 (0.39) > option 3 (0.38) > status quo (0.22). Thus, even if the importance of low costs might have been underestimated, the results were not sensitive to the doubling of the weight.

ValueDecisions also provides graphs for local sensitivity analyses of weights on each objective. We discuss results for the objective "High chance of compliance by end-users" (Cb_Comp). The weight assigned to this objective by survey respondents was relatively low. However, end-user compliance is actually a main concern of the Paris wastewater authority. As the results show, Cb_Comp is a relevant objective in the decision because the values and rankings of options can drastically change depending on the weight (Fig. 7). In the extreme case, option 5_Compost, best performing with the current weights, would achieve the lowest performance if the weight of Cb_Comp was increased to 0.62 or higher. Inversely, 1_Status_quo, worst performing with the current weights, would achieve the second best value if the weight was increased to more than 0.42. It would even perform best along with option 4_Vacuum if only objective Cb_Comp was considered (i.e., having weight 1). If the wastewater authority is not ready to bear a high risk regarding this objective, option 4_Vacuum might be recommendable. It is only outperformed by option 5_Compost when the weight for Cb_Comp is relatively low, but achieves higher values than option 5 as soon as the weight is higher than 0.19 (visualized by crossing lines for options in Fig. 7).

4.3.3. Lowest-level value functions: what if not linear?

ValueDecisions allows jointly changing all lowest-level value functions to an exponential shape and varying the curvature parameter c (see Eq. SI-3). We tested the effect of strong curvatures where $c = 5$ and $c = -5$ (example in Fig. 1), compared to using linear value functions as in the previous analyses. For $c = 5$, both the values and the ranking of options were partly impacted (Table 1). For $c = -5$ the overall values strongly changed, while the ranking remained consistent. Because option 5_Compost or option 4_Vacuum performed best in all three cases, our recommendation is not impacted. The robustness of the main conclusions also suggests that investing more time to elicit lowest-level value functions from stakeholders is unlikely to add value to the analysis. However, this is dependent on the specific case and the model assumptions and needs to be verified in every new application.

4.3.4. Aggregation model: what if not additive?

ValueDecisions allows varying the non-additivity parameter γ of the weighted power mean aggregation function to explore the effect of different aggregation models on the results (see Eq. (2)). This is relevant, since we could not ask stakeholders about their preferences concerning the assumptions and implications of the additive model, which often does not comply with people's preferences. We tested the effect of various aggregation functions (Table 2). A rank reversal between option 5_Compost being best, and option 4_Vacuum, occurred around $\gamma = 0.25$. With the minimum aggregation function, all options received a value of zero, because all options performed worst on at least one of the objectives (Fig. SI-1). Similarly, the high values for the maximum aggregation function are understandable when looking at Figure SI-1: all options performed best on at least one of the objectives (except option 3_Usep_store, which performed second-best and quite well on the objective Aa_Nut: low discharge of nitrogen to the river and air). This analysis again confirmed that the most robust recommendation would be option 5_Compost, closely followed by option 4_Vacuum.

4.3.5. Including uncertainty of predictions

Our predictions for some attributes were highly uncertain, for instance, for the number of local jobs (Da_Job), but rather certain for other attributes, for instance, water saving (Bb_Wat; Fig. SI-2). This prediction uncertainty can be propagated to the MCDA results and can be represented by barcharts with error bars (Fig. SI-3) or boxplots (Fig. 8). For clarity, we again focus on the medSWING weight profile (Fig. 4) and our standard assumptions (linear lowest-level value functions, additive model). After 2000 Monte Carlos simulation runs, option 4_Vacuum performed best (median value 0.70), closely followed by option 5_Compost (median value 0.64), while 1_Status_quo remained the worst performing option ($v = 0.31$).

Table 1
Results (values and ranks) for the five options for the medSWING group when modifying the shape of the lowest-level value functions (not elicited in our case, and linear ($c = 0$) as default) to concave and convex shapes. Prediction uncertainty is not considered.

Option	$c = 0$ (linear)		$c = 5$ (concave)		$c = -5$ (convex)	
	Rank	Value	Rank	Value	Rank	Value
1_Status_quo	5	0.20	5	0.29	5	0.14
2_Usep_conc	3	0.44	4	0.66	3	0.23
3_Usep_store	4	0.41	3	0.74	4	0.15
4_Vacuum	2	0.65	1	0.86	2	0.34
5_Compost	1	0.73	2	0.80	1	0.69

These findings also become clear with visualizations focusing on the ranks (Fig. 9, Fig. SI-7). Across all simulation runs, option 4_Vacuum and option 5_Compost achieved highest ranks often and never the bottom rank. Considering these results, it would be advisable to lower the uncertainty of the predictions for option 4_Vacuum and option 5_Compost, in order to establish with more certainty which of these would be the best-performing option.

4.3.6. Cost-benefit visualization

The visualization of costs (attribute Db_Cost) versus all other benefits (calculated as aggregated value of the remaining objectives using our standard MCDA model; Fig. 10) confirmed previous results: option 5_Compost and option 4_Vacuum performed almost equally well. However, option 5_Compost was markedly more cost-efficient than option 4_Vacuum. In fact, option 5_Compost, with estimated costs of 92 € per person and year, was the only cost-efficient option for the considered preferences, which is also why there is no efficient frontier in Fig. 10 (Fig. SI-6 depicts such an efficient frontier for the utilitarian and preservationist perspectives). Option 1_Status_quo clearly performed worst (lowest value, resp. benefit), and having estimated costs of 125 € per person and year is almost as expensive as option 4_Vacuum. The two most expensive options, 2_Usep_conc and 3_Usep_store, require a high cost investment for a rather small value gain. Since the predictions for cost had large uncertainties (Fig. SI-2), which is not visualized here, we recommend the decisions makers to obtain better predictions, at least for option 4_Vacuum and option 5_Compost, to increase trust in the robustness of these findings.

4.3.7. Implications for the case study

The results of the MCDA using ValueDecisions, given the model, the predictions, and the weight preferences elicited online from a sample of 655 people in the Paris region, clearly indicate two best-performing options: 4_Vacuum (urine and feces collection in a separate vacuum network with decentralized treatment) and 5_Compost (dry toilets with on-site underground composting chambers). These results are robust to changes in preferences, model assumptions, and to uncertainty in predictions. We are confident to recommend these two options for further evaluation to the main wastewater authority in the Paris region. Moreover, our analyses allow us to recommend further investigating the predictions of costs and of compliance by end-users for these two options. This will increase our confidence, which of the two options can be expected to perform best.

5. Evaluation of usability

5.1. Evaluation criteria and questionnaire items

We believe a value focused approach to software development is helpful. One important aspect is usability for the intended users, which is embodied in numerous user-centered design approaches (e.g., Brhel et al., 2015). We performed a usability test about the quality in use during the development of ValueDecisions that we formulated based on the ISO/EIC-25010 standard (ISO/IEC JTC 1/SC 7, 2011).

This standard suggests five criteria to evaluate usability. We focused on three of these – effectiveness, efficiency, and satisfaction – which all belong to the quality in use category. Each criterion is specified by one or several sub-criteria. We did not evaluate the criterion freedom of risk, as we foresee no economic risks (ValueDecisions is free and open source), no health and safety risks (e.g., no addiction risk), and no environmental risks (no special infrastructure or high electricity consumption is required). The criterion context coverage is, to our understanding, covered in our evaluation of the product quality characteristics (see section SI-5) and was not evaluated from the user perspective.

The ISO/EIC-25010 standard provides a structure for the evaluation, but no concrete format. Table 3 and in more detail Table SI–4 present

Table 2
Ranking (and values in parentheses) for the five options with different aggregation models for the medSWING group. The options are ordered according to the rank obtained with the default additive model (option 5_Compost achieving the best rank, 1_Status_quo the last rank).

Options	Additive model (γ = 1)	Approaching weighted geometric mean (γ = 0.04)	Between arithmetic and geometric mean (γ = 0.5)	Minimum (γ tends to -∞)	Maximum (γ tends to +∞)
1_Status_quo	5 (0.200)	5 (0)	5 (0.060)	0 (0)	3 (0.995)
2_Usep_conc	3 (0.441)	4 (0.001)	4 (0.314)	0 (0)	2 (0.996)
3_Usep_store	4 (0.406)	3 (0.020)	3 (0.335)	0 (0)	5 (0.960)
4_Vacuum	2 (0.646)	1 (0.062)	2 (0.569)	0 (0)	3 (0.995)
5_Compost	1 (0.726)	2 (0.024)	1 (0.593)	0 (0)	1 (0.999)

our operationalization of the standard. The evaluation is based on user feedback collected with a questionnaire and on observation. Based on literature, we developed between three to six items for each sub-criterion to measure their achievement (Table SI–4). The items are variants of a single question to measure a sub-criterion in a reliable and robust manner (Kline, 2000). We also asked users how often they required support (e.g., from the user guide, assistance from teaching assistant, assistance from other students). In two open text questions, users could note what they appreciated about ValueDecisions and what they recommend to improve. To measure the criterion effectiveness, we used the grades assigned to the student reports as a proxy indicator instead of questionnaire items.

Table 3
Overview of criteria, sub-criteria, and measures/items to operationalize the ISO/EIC-25010 standard and evaluate the quality in use of ValueDecisions. Details concerning the items, type of answers, and references see Table SI–4.

Criteria	Sub-criteria	Definitions from ISO/EIC-25010 standard	# of items
Effectiveness	–	"accuracy and completeness with which users achieve specified goals"	1 (measure)
Efficiency		"the resources expended in relation to the accuracy and completeness with which users achieve goals (e.g., time to complete the task)"	
	General	Cognitive resources involved (effort)	6
	Time	Time resources involved	3
Satisfaction		"degree to which user needs are satisfied when a product or system is used in a specified context of use"	
	Usefulness	"degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use". We distinguished between:	
		- Usefulness for analyzing **multi-stakeholder** decisions with MCDA, and	4
		- Usefulness for analyzing **uncertain** decisions with MCDA.	4
	Trust	"degree to which a user or other stakeholder has confidence that a product or system will behave as intended". We distinguished between two subsets of questions:	
		- nine questions (including four filter questions) about the **potential reasons for losing trust**, and	5
		- six questions about **trust as general attitude** towards ValueDecisions.	6
	Pleasure	"degree to which a user obtains pleasure from fulfilling their personal needs. Personal needs can include needs to acquire new knowledge and skills (…)". We distinguished between:	
		- user friendliness of interface, and	3
		- general pleasure of using the app.	3
	Comfort	"degree to which the user is satisfied with physical comfort"	7

5.2. Implementation and test users

The purpose of the usability test was to evaluate the app in a systematic way and to reveal how to improve and further develop it. For testing usability, representative users should perform representative tasks (Lazar et al., 2017). We conducted the usability test with master students with an environmental major who were learning MCDA based on MAVT and had a basic, but sound method knowledge. They are target users of the ValueDecisions app (see section 3.2): potentially, they will work as environmental consultants or in government agencies and may use the app in their future career. Sixteen students carried out four environmental case studies using ValueDecisions as part of a course at ETH Zurich in spring 2020 (Lienert, 2020). We repeated the usability test in the in the spring 2021 course; this time on a voluntary basis. Eight students from three case studies answered.

The students filled in the usability questionnaire as homework at the end of the course. They were fully informed that they were test users and signed an informed consent form. The survey was coded in LimeSurvey (Limesurvey GmbH, 2020). The questions appeared in random order except if they implied a logical development. All survey items are available in Table SI–4. The survey was pre-tested by two research assistants and the teaching assistant.

5.3. Results and discussion of evaluation and response

The evaluation of usability varied considerably across (sub-)criteria and among students (Table 4, Table SI–5). While our sample is too small to allow for any generalization, the usability testing nevertheless helped to better understand use of the ValueDecisions app and weaknesses. Based on this feedback, we made considerable efforts to improve the app with additions to the code and further explanations. Here, we highlight few important results and responses based on the test in 2020, a full description is provided in SI-6.

The efficiency of using the app, especially, the time efficiency, was rated relatively low. We suspect that this is largely due to difficulties related to preparing properly formatted and valid input files. The teaching assistant mainly supported students in finding formatting mistakes. Once they had set up correct input files, the students were capable of using ValueDecisions autonomously for carrying out their

Table 4
Usability results for each sub-criterion of the standard, averaged over the respective items of each sub-criterion. Neutral assessment is 3; above 3 to 5 is positive; below 3 negative. Results for each item are presented in Table SI–5.

	Sub-criterion	Min	Mean (SD)	Median	Max
Efficiency	General efficiency	2.33	3.28 (0.52)	3.5	4
	Time efficiency[a]	1	1.75 (1.08)	1.33	5
Satisfaction	Usefulness	2.88	4.09 (0.55)	4.19	5
	General trust	3	4.12 (0.67)	4.17	5
	Losing trust[a]	1.75	2.59 (0.38)	2.67	3.3
	Pleasure	2.5	3.91 (0.61)	3.92	4.83
	Comfort	2.57	3.77 (0.69)	3.93	4.86
Overall usability		2	3.13 (0.77)	3	4.25

[a] These two constructs required transformation of the single items scales to be handled as 5-point Likert scales (see section SI-6.2).

case studies. We also included questions to assess reasons for *loosing trust* in the app. Across all exercises, 14 of the 16 students had received error messages. Many students found that the messages did not indicate the source of the error clearly enough for them to know how to address it, resulting in low evaluations of these questionnaire items. Ten students faced moments when they did not know what to do next, and ten lost the running analyses and needed to restart the app. Hence, while the general trust in ValueDecisions was positive (mean 4.12), students lost trust (mean 2.59) due to technical difficulties.

As a response, we implemented additional validation of input files after upload and provided more details what users need to change when this validation fails. Furthermore, we added a checklist to support users in systematically controlling their files, information about common error messages and potential solutions, and extended the user guide. We also decided to better describe the output data and analyses within the app. Lastly, we provided an exemplary decision analysis "choosing the holiday destination for my extended family", which is understandable to many, with input files that can be downloaded as templates.

The other sub-criteria to satisfaction (usefulness, pleasure, and comfort) were on average rated positively (Table 4). Most students found ValueDecisions pleasant to use, and stated that it had a user-friendly interface. They also felt comfortable using the app, and found it easy to learn how to use it, clear and understandable, and flexible regarding the analysis possibilities. This increased our confidence that the approach taken for ValueDecisions is useable in practice and that people with basic understanding of MCDA can successfully use ValueDecisions and will have an effective and satisfying experience.

A repetition of the user feedback questionnaire in the 2021 MCDA course confirmed the evaluation (SI-6.4). While the time needed to prepare the input file was still detrimental to the overall evaluation, the other aspects were well-perceived. We conclude that users require a solid method understanding to understand why the information in the input files has to be specified in a certain way and to use the app well. As for any new software, a learning phase seems necessary prior to the efficient use of ValueDecisions.

In addition to the software evaluation with the user survey, we involved expert users from the Decision Analysis group at Eawag in developing ValueDecisions, and several external scientists are already using it in their projects. Their continuous use and feedback led to various other improvements, including bug fixes, support for additional visualizations, and download of interim simulation results. Furthermore, the product quality elements of the ISO/EIC-25010 standard (ISO/IEC JTC 1/SC 7, 2011) served as a checklist to guide the software development (see SI-5).

The usability test that we developed for this study based on the ISO/EIC-25010 standard (ISO/IEC JTC 1/SC 7, 2011) is general and can be adapted to other contexts and software. The full questionnaire is provided in Table SI–4. Our approach was pragmatic and a scientific validation of the scales (see Kline, 2000) would need further research. Nevertheless, as starting point, our scales showed satisfactory internal reliability (Table SI–5), which indicates that the items reliably and consistently measured the sub-criterion they were designed for. Generally, we found the value-focused approach to software development and the evaluation of this via usability testing valuable for guiding development efforts.

6. Discussion and conclusion

6.1. Supporting decisions with the ValueDecisions app

In an application concerning wastewater management in the Paris region (also see Aubert et al., 2020), we demonstrated how a public policy decision problem can be analyzed using ValueDecisions. We collected the preference data from the population before ValueDecisions was conceived, but could easily convert it to the appropriate format. By performing extensive sensitivity analyses, we found that key results were insensitive to preference parameters not collected in the online survey of 655 respondents. Such analysis supports determining for which aspects of the decision more information would be helpful and for which not. In our case, results are robust enough to conclude that eliciting more detailed preferences from stakeholders would likely not change the main results. This is especially relevant for online surveys with lay people, as in our case. To avoid cognitive overload and tiring of respondents (e.g., Riabacke et al., 2012), we did not elicit the shapes of single-attribute value functions or the best-fitting aggregation model. Earlier studies also concluded that not all preference parameters have to be known in detail (Schuwirth et al., 2012; Scholten et al., 2015), but this is always case-dependent. Uncertainty analyses allows us to recommend that it could be worthwhile to decrease the uncertainty of the input data for predictions if a clearer differentiation between the two best-performing options is desired. The focus on visualization given by ValueDecisions was key for generating insights and should ease communicating about the robustness of the results to the Paris decision-makers.

Two unconventional wastewater management options emerged as robust choices, namely dry toilets with on-site underground composting chambers and urine and feces collection in a separate vacuum network with decentralized treatment (section 4.3). This answers the main concern of the Paris wastewater authority: it seems reasonable to consider a radical system change for wastewater management in certain contexts (Larsen et al., 2016; Hoffmann et al., 2020), given the preferences of a representative sample of Paris region citizens. Our results thus validate the momentum in the Paris region toward unconventional systems: pilot projects are implementing some options we considered in this analysis.

6.2. Alternative software for environmental decision support

All tools have a context and use cases in which they are useful. Consequently, ValueDecisions was developed with specific use cases and users in mind (section 3.2). As far as we are aware, the combination of advanced features ValueDecisions offers, for instance, with regard to preference modeling, is unique; especially combined with a simple spreadsheet interface for the input data, a web interface, and automatic reporting. However, depending on the specific decision case, other software may be better suited (see Weistroffer and Li, 2016; Mustajoki and Marttunen, 2017). MCDA software development geared to different types of use and users is an active field of research with several recent additions (e.g., Chacon-Hurtado and Scholten, 2020; Cinelli et al., 2021; Preference AB, 2021) and we expect interesting developments in the future. While not being able to do justice to all existing software, we briefly point out some developments that address uses ValueDecisions is lacking.

Some MCDA software focuses more on guiding users through the process of decision-making. Mustajoki and Marttunen (2017) emphasize that for non-expert users, software needs to provide automatic guidance, for instance, to overcome commonly encountered biases. Entscheidungsnavi is one online tool that has extensive user support to overcome common biases and guides users through the entire decision-making process based on MAUT (von Nitzsch et al., 2020). Process guidance is also offered in software like V.I.S.A. Decisions (SIMUL8, 2021) or Logical Decisions (Logical Decisions, 2020) and others. Truly interactive elicitation of preferences is, for instance, possible with FITradeoff (de Almeida et al., 2016).

Supporting decisions with multiple stakeholders is an important property in public decisions. Explicit support for group decision-making is for instance provided by Logical Decisions (Logical Decisions, 2020) or Helision (Preference AB, 2021). Furthermore, spatial assessment can be a relevant feature of environmental problems. The combination of MCDA and geographic information systems (GIS) is an active research area with various software developments (e.g., Greene et al., 2011; Keenan and Jankowski, 2019). ValueDecisions does not allow for explicit spatial analysis, but attributes relevant to spatial features can be used, for example, “number of protected areas affected” or “space required”.

Other software allow for greater flexibility in modeling. This is provided by programming libraries and packages for R, such as “utility” (Reichert et al., 2013) or “MCDA” (Bigaret et al., 2017), and for Python, such as “decisi-o-rama” (Chacon-Hurtado and Scholten, 2020). The latter also supports portfolio decision analysis (Lahtinen et al., 2017b). With an appropriate workflow, software based on these libraries can also be coupled to complex prediction models. Examples of software with graphical user interface that supports different MCDA algorithms are DECERNS (Linkov et al., 2020) or DecSpace (Amador et al., 2018). Furthermore, the diviz initiative aims at providing a common interface to many algorithms (Meyer and Bigaret, 2012).

6.3. Further development of ValueDecisions

ValueDecisions can be extended in many ways. Thanks to its modular design, features can be added without changing the basic structure. Five examples of further development are:

1 Support for uncertain preference parameters. For instance, stakeholders may specify a range of weight parameters rather than point estimates (Scholten et al., 2015), or provide an estimate about how uncertain they were about their statements (Zheng et al., 2016). By allowing uncertain preferences, a more realistic picture of the uncertainty of results would be achieved. This would extend the current approach, which relies on performing sensitivity analyses regarding preferences. If uncertainty of predictions and preferences should be considered jointly, the concept of expected expected utility could be implemented (Haag et al., 2019b).
2 Support for multi-attribute utility theory (section 2.2.4). By including the risk attitude of stakeholders and calculating expected utilities instead of values, preferences about uncertain predictions could be directly considered (as in Scholten et al., 2015). Additionally, the risk-attitude could be interactively varied and robust options identified.
3 Support for SMAA methods and metrics for analysis (Stochastic Multi-criteria Acceptability Analysis; Lahdelma et al., 1998). These also can be useful for MCDA problems, where both predictions of options and preference information is uncertain (e.g., Zheng et al., 2016).
4 Interactively support users in data entry and/or data elicitation. Currently, users need to provide input data with two spreadsheets. This is error prone as issues with incompatible entries may arise, and users may be unsure how to provide this information properly. Integrating the data entry or even preference elicitation step into ValueDecisions would make for a more seamless experience. Such user guidance was also one recommendation of Mustajoki and Marttunen (2017).
5 The automatic reporting could be extended. One possibility is using the ideas of natural language generation to explain results to lay users (Wulf and Bertsch, 2017). We believe this to be an interesting feature for potential future users such as environmental consultants.

Scientific software development exhibits the long tail phenomenon, i.e., the large majority of software sees only very little uptake and reuse (e.g., Walling and Vaneeckhaute, 2020). Software sustainability is a major issue for academic software (e.g., Venters et al., 2014), and many tools in the long tail are eventually lost. ValueDecisions also faces this risk. Two aspects might mitigate this risk. Firstly, ValueDecisions is an open source development. The source code is available and can be reused or modified freely by anyone. Secondly, it is written in R, which is a programming language commonly used for analysis in academia and the app comes in the standardized form of an R package. This allows people without specific software development background to contribute to or change the app. Additionally, this makes ValueDecisions interoperable with other R packages for MCDA, such as “utility” (Reichert et al., 2013). It is possible to exchange the calculations and algorithms in the backend while keeping the visualization and frontend of the application.

7. Conclusions

Effective decision support is facilitated by software tools that help analyze, visualize, and understand the key aspects of a decision problem. This insight, well known in practice, has sparked numerous software developments since the 1980s (reviewed, e.g., by Korhonen et al., 1992; French and Xu, 2005; Weistroffer and Li, 2016; Mustajoki and Marttunen, 2017). Here, we introduced a novel open source development, the ValueDecisions app.

ValueDecisions was developed to support analysts, facilitators, and interested stakeholders during the modeling stage of an MCDA process. It is targeted towards environmental and public policy problems. Key properties of the app are: (i) the possibility to represent complex stakeholder preference structures over multiple objectives by building hierarchical MAVT models that combine non-linear lowest-level value functions with non-additive aggregation functions; (ii) the possibility to consider uncertainties by different probability distributions for predictions, interactive sensitivity analyses, and visualizing uncertainty in different ways; (iii) the comparison of results for multiple, potentially conflicting, stakeholder preference profiles; (iv) a graphical user interface accessible via a web browser that focuses on producing insightful visualizations; (v) open source development based on R, which allows for modifications and extensions of the algorithms and visualizations.

We tested ValueDecisions with data from an online survey of 655 citizens for an urban water management decision in the Paris region. ValueDecisions allowed us to clearly identify robust options for wastewater management by using different sensitivity analyses. To identify user needs, we developed a structured usability test based on the ISO/EIC-25010 standard (ISO/IEC JTC 1/SC 7, 2011). Students participating in a MCDA lecture (Lienert, 2020) used ValueDecisions to tackle environmental decision problems. Their responses in the usability test pointed us toward important improvements of the app. The usability survey we developed can now be used, tested, and improved in other applications.

Given the large role software plays in applied projects, we know surprisingly little about how it is used and what that use implies. Every software tool has certain affordances, and enables users to do certain things, but not others. The field of behavioral operations research (BOR; e.g., Franco and Hämäläinen, 2016; Hämäläinen, 2015) has picked up such research. We can only reiterate the call of Mustajoki and Marttunen (2017) to research “the interaction of people with MCDA software; for example, how people use software, what kind of support they need, and how the characteristics of the software affect people’s learning and interpretation of results”. One important step in this direction is user-centered evaluation (cf. Walling and Vaneeckhaute, 2020) as we have operationalized here.

Our basic motivation for developing ValueDecisions was making MCDA analysis for environmental decision problems accessible to a wider audience. Given the usability evaluations, we are hopeful that ValueDecisions can contribute to this. We chose to make the software available free of charge and accessible as a web application without any installation. We hope that the possibilities offered by the app will facilitate the application of decision analysis based on multi-attribute value/utility theory and eventually contribute to better structured, transparent, and well-informed decisions.

Software availability

- Name of software: ValueDecisions

- Developer and contact information: concept & design: Fridolin Haag, Judit Lienert; programming: Fridolin Haag, Kevin Schönholzer, Sara Schmid, ETH Scientific IT Services; contact: Judit.Lienert@eawag.ch
- Year first available: 2020
- Hardware required: No specific requirements
- Software required: Web browser; for the R package version: R base installation
- Availability: https://www.eawag.ch/en/department/ess/main-focus/decision-analysis-da/tools/
- Program language: R
- Cost: free, open source
- License: AGPL

Data availability

Data and software used in this paper are available on the Eawag Research Data Institutional Repository (https://opendata.eawag.ch/): https://doi.org/10.25678/00048S.

CRediT author statement

Fridolin Haag: Conceptualization, Methodology, Software, Writing (Original Draft, Review & Editing), Visualization, Supervision, Project administration; Alice Aubert: Conceptualization, Methodology, Investigation, Writing (Original Draft, Review & Editing), Supervision (usability test, case study); Judit Lienert: Conceptualization, Software (testing and improvement), Writing (Original Draft, Review & Editing), Funding acquisition, Supervision, Project administration.

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.

Acknowledgements

We are grateful to Kevin Schönholzer, Sara Schmid, and the ITS Scientific Software & Data Management group at ETH Zürich for their contributions to programming ValueDecisions. We thank our colleagues Qi Zeng and Sara Schmid for supporting data analysis, Harald von Waldow and Markus Kasper for valuable advice on software deployment, and Daniel Hofmann and Philipp Beutler for feedback. We thank the respondents of the Paris case study for answering our survey and Fabien Esculier for collaboration and interacting with the Paris region wastewater authorities. We are indebted to the students of the MCDA lecture in 2020 and 2021 for using ValueDecisions and participating in the survey. Alice Aubert was supported by an Eawag Postdoctoral Fellowship 2015 (grant number: 5221.00492.009.08 DF 15) and a Swiss National Science Foundation Ambizione grant (project 173973 – Edanaga). Eawag also funded this applied research as project (A.5. MCDA-Tool) under the 2019/2020 agreement between the Scientific IT Services of ETH Zürich and Eawag. We wish to sincerely thank the four reviewers and the editor for their constructive comments, which helped us to improve this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2022.105361.

References

Amador, J., Costa, A.S., Rodrigues, R., Figueira, J.R., Borbinha, J., 2018. Exploring MCDA methods with DecSpace. In: CAPSI 2018. Presented at the 18th Conference of the Portuguese Association for Information Systems.

Aubert, A.H., Esculier, F., Lienert, J., 2020. Recommendations for online elicitation of swing weights from citizens in environmental decision-making. Oper. Res. Perspect. 7, 100156. https://doi.org/10.1016/j.orp.2020.100156.

Belton, V., Hodgkin, J., 1999. Facilitators, decision makers, D.I.Y, users: is intelligent multicriteria decision support for all feasible or desirable? Eur. J. Oper. Res. 113, 247–260. https://doi.org/10.1016/S0377-2217(98)90214-4.

Belton, V., Stewart, T.J., 2002. Multiple Criteria Decision Analysis : an Integrated Approach. Kluwer Academic Publishers, Boston.

Bigaret, S., Hodgett, R.E., Meyer, P., Mironova, T., Olteanu, A.-L., 2017. Supporting the multi-criteria decision aiding process: R and the MCDA package. EURO J. Decis. Process. 5, 169–194. https://doi.org/10.1007/s40070-017-0064-1.

Borsuk, M.E., Maurer, M., Lienert, J., Larsen, T.A., 2008. Charting a path for innovative toilet technology using multicriteria decision analysis. Environ. Sci. Technol. 42, 1855–1862. https://doi.org/10.1021/es702184p.

Brhel, M., Meth, H., Maedche, A., Werder, K., 2015. Exploring principles of user-centered agile software development: a literature review. Inf. Softw. Technol. 61, 163–181. https://doi.org/10.1016/j.infsof.2015.01.004.

Cegan, J.C., Filion, A.M., Keisler, J.M., Linkov, I., 2017. Trends and applications of multi-criteria decision analysis in environmental sciences: literature review. Environ. Syst. Decis. 37, 123–133. https://doi.org/10.1007/s10669-017-9642-9.

Chacon-Hurtado, J.C., Scholten, L., 2020. Decisi-o-rama: an open-source Python library for multi-attribute value/utility decision analysis. Environ. Model. Software 104890. https://doi.org/10.1016/j.envsoft.2020.104890.

Chang, W., Cheng, J., Allaire, J.J., Carson, S., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2021. Shiny: Web Application Framework for R R Package Version 1.6.0. https://CRAN.r-project.org/package=shiny.

Cinelli, M., Kadziński, M., Gonzalez, M., Słowiński, R., 2020a. How to Support the Application of Multiple Criteria Decision Analysis? Let Us Start with a Comprehensive Taxonomy. Omega, p. 102261. https://doi.org/10.1016/j.omega.2020.102261.

Cinelli, M., Spada, M., Kim, W., et al., 2021. MCDA Index Tool: an interactive software to develop indices and rankings. Environ. Syst. Decis. 41, 82–109. https://doi.org/10.1007/s10669-020-09784-x.

de Almeida, A.T., de Almeida, J.A., Costa, A.P.C.S., de Almeida-Filho, A.T., 2016. A new method for elicitation of criteria weights in additive models: flexible and interactive tradeoff. Eur. J. Oper. Res. 250, 179–191. https://doi.org/10.1016/j.ejor.2015.08.058.

DILA, 2021. Grand Paris : un projet de développement pour une métropole de rang mondial. https://www.vie-publique.fr/eclairage/19461-grand-paris-projet-de-developpement-dune-metropole-de-rang-mondial. (Accessed 28 June 2021).

Dyer, J.S., Sarin, R.K., 1982. Relative risk-aversion. Manag. Sci. 28, 875–886. https://doi.org/10.1287/mnsc.28.8.875.

Eisenführ, F., Weber, M., Langer, T., 2010. Rational Decision Making. Springer, Berlin; London.

Esculier, F., Tabuchi, J.P., Créno, B., 2015. Nutrient and energy flows related to wastewater management in the Greater Paris: the potential of urine source separation under global change constraints. In: Presented at the UNCESCO International Conference on Water, Megacities and Global Change, Paris, France.

Esmail, B.A., Geneletti, D., 2018. Multi-criteria decision analysis for nature conservation: a review of 20 years of applications. Methods Ecol. Evol. 9, 42–53. https://doi.org/10.1111/2041-210X.12899.

EWG-MCDA, 2020. MCDA Software. http://www.cs.put.poznan.pl/ewgmcda/index.php/software. (Accessed 25 August 2020).

Fay, C., Rochette, S., Guyader, V., Girard, C., 2021. Engineering Production-Grade Shiny Apps. In: The R Series. Chapman and Hall/CRC.

Flipo, N., Gallois, N., Labarthe, B., Baratelli, F., Viennot, P., Schuite, J., Rivière, A., Bonnet, R., Boé, J., 2021. Pluri-annual water budget on the seine basin: past, current and future trends. In: Flipo, N., Labadie, P., Lestel, L. (Eds.), The Seine River Basin, the Handbook of Environmental Chemistry. Springer International Publishing, Cham, pp. 59–89. https://doi.org/10.1007/698_2019_392.

Franco, L.A., Hämäläinen, R.P., 2016. Behavioural operational research: returning to the roots of the OR profession. Eur. J. Oper. Res. 249, 791–795. https://doi.org/10.1016/j.ejor.2015.10.034.

French, S., Argyris, N., 2018. Decision analysis and political processes. Decis. Anal. 15, 208–222. https://doi.org/10.1287/deca.2018.0374.

French, S., Geldermann, J., 2005. The varied contexts of environmental decision problems and their implications for decision support. Environ. Sci. Policy 8, 378–391. https://doi.org/10.1016/j.envsci.2005.04.008.

French, S., Xu, D.-L., 2005. Comparison study of multi-attribute decision analytic software. J. Multi-Criteria Decis. Anal. 13, 65–80. https://doi.org/10.1002/mcda.372.

Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E., 2009. Aggregation Functions, Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge.

Multiple criteria decision analysis - state of the art surveys. In: Greco, S., Ehrgott, M., Figueira, J.R. (Eds.), 2016. International Series in Operations Research & Management Science, second ed. Springer, New York, NY. https://doi.org/10.1007/978-1-4939-3094-4_1.

Greene, R., Devillers, R., Luther, J.E., Eddy, B.G., 2011. GIS-based multiple-criteria decision analysis. Geogr. Compass 5, 412–432. https://doi.org/10.1111/j.1749-8198.2011.00431.x.

Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T., Ohlson, D., 2012. Structured Decision Making: a Practical Guide to Environmental Management Choices. Wiley-Blackwell, Chichester, West Sussex, UK.

Haag, F., Lienert, J., Schuwirth, N., Reichert, P., 2019a. Identifying non-additive multi-attribute value functions based on uncertain indifference statements. Omega 85, 49–67. https://doi.org/10.1016/j.omega.2018.05.011.

Haag, F., Reichert, P., Maurer, M., Lienert, J., 2019b. Integrating uncertainty of preferences and predictions in decision models: an application to regional wastewater planning. J. Environ. Manag. 252, 109652. https://doi.org/10.1016/j.jenvman.2019.109652.

Haag, F., Zürcher, S., Lienert, J., 2019c. Enhancing the elicitation of diverse decision objectives for public planning. Eur. J. Oper. Res. 279, 912–928. https://doi.org/10.1016/j.ejor.2019.06.002.

Hajkowicz, S., Collins, K., 2007. A review of multiple criteria analysis for water resource planning and management. Water Resour. Manag. 21, 1553–1566. https://doi.org/10.1007/s11269-006-9112-5.

Hämäläinen, R.P., 2015. Behavioural issues in environmental modelling – the missing perspective. Environ. Model. Software 73, 244–253. https://doi.org/10.1016/j.envsoft.2015.08.019.

Harris-Lovett, S., Lienert, J., Sedlak, D., 2019. A mixed-methods approach to strategic planning for multi-benefit regional water infrastructure. J. Environ. Manag. 233, 218–237. https://doi.org/10.1016/j.jenvman.2018.11.112.

Hoffmann, S., Feldmann, U., Bach, P.M., Binz, C., Farrelly, M., Frantzeskaki, N., Hiessl, H., Inauen, J., Larsen, T.A., Lienert, J., Londong, J., Lüthi, C., Maurer, M., Mitchell, C., Morgenroth, E., Nelson, K.L., Scholten, L., Truffer, B., Udert, K.M., 2020. A research agenda for the future of urban water management: exploring the potential of nongrid, small-grid, and hybrid solutions. Environ. Sci. Technol. 54, 5312–5322. https://doi.org/10.1021/acs.est.9b05222.

Huang, I.B., Keisler, J., Linkov, I., 2011. Multi-criteria decision analysis in environmental sciences: ten years of applications and trends. Sci. Total Environ. 409, 3578–3594. https://doi.org/10.1016/j.scitotenv.2011.06.022.

ISO/IEC JTC 1/SC 7, 2011. Systems and Software Engineering — Systems and Software Quality Requirements and Evaluation (SQuaRE) — System and Software Quality Models. No. ISO/IEC 25010).

Keenan, P.B., Jankowski, P., 2019. Spatial decision support systems: three decades on. Decis. Support Syst. 116, 64–76. https://doi.org/10.1016/j.dss.2018.10.010.

Keeney, R.L., 1992. Value-focused Thinking: a Path to Creative Decisionmaking. Harvard University Press, Cambridge, Mass.

Keeney, R.L., 2013. Foundations for group decision analysis. Decis. Anal. 10, 103–120. https://doi.org/10.1287/deca.2013.0265.

Keeney, R.L., Raiffa, H., 1993. Decisions with Multiple Objectives: Preferences and Value Trade-Offs, 2nd. ↱Cambridge University Press.

Kline, P., 2000. The Handbook of Psychological Testing, 2nd. Routledge, London; New York.

Korhonen, P., Moskowitz, H., Wallenius, J., 1992. Multiple criteria decision support - a review. Eur. J. Oper. Res. 63, 361–375. https://doi.org/10.1016/0377-2217(92)90155-3.

Kurth, M.H., Larkin, S., Keisler, J.M., Linkov, I., 2017. Trends and applications of multi-criteria decision analysis: use in government agencies. Environ. Syst. Decis 37, 134–143. https://doi.org/10.1007/s10669-017-9644-7.

Lahdelma, R., Hokkanen, J., Salminen, P., 1998. SMAA - Stochastic multiobjective acceptability analysis. Eur. J. Oper. Res. 106, 137–143. https://doi.org/10.1016/S0377-2217(97)00163-X.

Lahtinen, T.J., Guillaume, J.H.A., Hämäläinen, R.P., 2017a. Why pay attention to paths in the practice of environmental modelling? Environ. Model. Software 92, 74–81. https://doi.org/10.1016/j.envsoft.2017.02.019.

Lahtinen, T.J., Hämäläinen, R.P., Liesiö, J., 2017b. Portfolio decision analysis methods in environmental decision making. Environ. Model. Software 94, 73–86. https://doi.org/10.1016/j.envsoft.2017.04.001.

Langhans, S.D., Lienert, J., 2016. Four common simplifications of multi-criteria decision analysis do not hold for river rehabilitation. PLoS One 11, e0150695. https://doi.org/10.1371/journal.pone.0150695.

Langhans, S.D., Reichert, P., Schuwirth, N., 2014. The method matters: a guide for indicator aggregation in ecological assessments. Ecol. Indic. 45, 494–507. https://doi.org/10.1016/j.ecolind.2014.05.014.

Larsen, T.A., Alder, A.C., Eggen, R.I.L., Maurer, M., Lienert, J., 2009. Source separation: will we see a paradigm shift in wastewater handling? Environ. Sci. Technol. 43, 6121–6125. https://doi.org/10.1021/es803001r.

Larsen, T.A., Hoffmann, S., Lüthi, C., Truffer, B., Maurer, M., 2016. Emerging solutions to the water challenges of an urbanizing world. Science 352, 928–933. https://doi.org/10.1126/science.aad8641.

Lazar, J., Feng, J.H., Hochheiser, H., 2017. Chapter 10 - usability testing. In: Lazar, J., Feng, J.H., Hochheiser, H. (Eds.), Research Methods in Human Computer Interaction, second ed. Morgan Kaufmann, Boston, pp. 263–298. https://doi.org/10.1016/B978-0-12-805390-4.00010-8.

Lienert, J., 2020. Multi-Criteria Decision Analysis", Course 701-1522-00L. Department of Environmental Systems Sciences. ETH Zürich. http://www.vvz.ethz.ch/lerneinheitPre.do?semkez=2020S&lerneinheitId=135759&lang=en.

Lienert, J., Koller, M., Konrad, J., McArdell, C.S., Schuwirth, N., 2011. Multiple-criteria decision analysis reveals high stakeholder preference to remove pharmaceuticals from hospital wastewater. Environ. Sci. Technol. 45, 3848–3857. https://doi.org/10.1021/Es1031294.

Lienert, J., Scholten, L., Egger, C., Maurer, M., 2015. Structured decision-making for sustainable water infrastructure planning and four future scenarios. EURO J. Decis. Process. 3, 107–140. https://doi.org/10.1007/s40070-014-0030-0.

Lienert, J., Duygan, M., Zheng, J., 2016. Preference stability over time with multiple elicitation methods to support wastewater infrastructure decision-making. Eur. J. Oper. Res. 253, 746–760. https://doi.org/10.1016/j.ejor.2016.03.010.

Limesurvey GmbH, 2020. LimeSurvey: an Open Source Survey Tool. LimeSurvey GmbH, Hamburg, Germany.

Linkov, I., Moberg, E., Trump, B.D., Yatsalo, B., Keisler, J.M., 2020. Multi-Criteria Decision Analysis: Case Studies in Engineering and the Environment, second ed. CRC Press, Boca Raton. https://doi.org/10.1201/9780429326448.

Logical Decisions, 2020. Software for More Effective Decisions. http://www.logicaldecisions.com/. (Accessed 25 August 2020).

Lossouarn, C., Pangare, V., Fenghua, S., Lentini, E.J., Loftus, T., Dickinson, M.A., Ngoc, T.D.T., Perset, M., Strady, E., Phan, T.S.H., 2016. Water, Megacities and Global Change. UNESCO & ARCEAU-IdF.

Mareschal, B., 1988. Weight stability intervals in multicriteria decision aid. Eur. J. Oper. Res. 33, 54–64. https://doi.org/10.1016/0377-2217(88)90254-8.

Marttunen, M., Lienert, J., Belton, V., 2017. Structuring problems for Multi-Criteria Decision Analysis in practice: a literature review of method combinations. Eur. J. Oper. Res. 263, 1–17. https://doi.org/10.1016/j.ejor.2017.04.041.

Marttunen, M., Belton, V., Lienert, J., 2018. Are objectives hierarchy related biases observed in practice? A meta-analysis of environmental and energy applications of Multi-Criteria Decision Analysis. Eur. J. Oper. Res. 265, 178–194. https://doi.org/10.1016/j.ejor.2017.02.038.

Marttunen, M., Haag, F., Belton, V., Mustajoki, J., Lienert, J., 2019. Methods to inform the development of concise objectives hierarchies in multi-criteria decision analysis. Eur. J. Oper. Res. 277, 604–620. https://doi.org/10.1016/j.ejor.2019.02.039.

MCDM-Society, 2020. Software Related to MCDM [WWW Document]. URL. https://www.mcdmsociety.org/content/software-related-mcdm-0. (Accessed 25 August 2020).

Meyer, P., Bigaret, S., 2012. Diviz: a software for modeling, processing and sharing algorithmic workflows in MCDA. Intell. Decis. Technol. 6, 283–296. https://doi.org/10.3233/IDT-2012-0144.

Milfont, T.L., Duckitt, J., 2006. Preservation and utilization: understanding the structure of environmental attitudes. Medio Ambiente Comport Hum 7, 29–50.

Morton, A., Fasolo, B., 2009. Behavioural decision theory for multi-criteria decision analysis: a guided tour. J. Oper. Res. Soc. 60, 268–275. https://doi.org/10.1057/palgrave.jors.2602550.

Morton, A., Ackermann, F., Belton, V., 2003. Technology-driven and model-driven approaches to group decision support: focus, research philosophy, and key concepts. Eur. J. Inf. Syst. 12, 110–126. https://doi.org/10.1057/palgrave.ejis.3000455.

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

Preference AB, 2021. Helision. https://helision.com/. (Accessed 15 March 2021).

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/.

Reichert, P., Schuwirth, N., Langhans, S., 2013. Constructing, evaluating and visualizing value and utility functions for decision support. Environ. Model. Software 46, 283–291. https://doi.org/10.1016/j.envsoft.2013.01.017.

Reichert, P., Langhans, S.D., Lienert, J., Schuwirth, N., 2015. The conceptual foundation of environmental decision support. J. Environ. Manag. 154, 316–332. https://doi.org/10.1016/j.jenvman.2015.01.053.

Reichert, P., Niederberger, K., Rey, P., Helg, U., Haertel-Borer, S., 2019. The need for unconventional value aggregation techniques: experiences from eliciting stakeholder preferences in environmental management. EURO J. Decis. Process. 7, 197–219. https://doi.org/10.1007/s40070-019-00101-9.

Riabacke, M., Danielson, M., Ekenberg, L., 2012. State-of-the-Art prescriptive criteria weight elicitation. Adv. Decis. Sci. 24 https://doi.org/10.1155/2012/276584, 2012.

Rowley, H.V., Peters, G.M., Lundie, S., Moore, S.J., 2012. Aggregating sustainability indicators: beyond the weighted sum. J. Environ. Manag. 111, 24–33. https://doi.org/10.1016/j.jenvman.2012.05.004.

Scholten, L., Schuwirth, N., Reichert, P., Lienert, J., 2015. Tackling uncertainty in multi-criteria decision analysis - an application to water supply infrastructure planning. Eur. J. Oper. Res. 242, 243–260. https://doi.org/10.1016/j.ejor.2014.09.044.

Schuwirth, N., Reichert, P., Lienert, J., 2012. Methodological aspects of multi-criteria decision analysis for policy support: a case study on pharmaceutical removal from hospital wastewater. Eur. J. Oper. Res. 220, 472–483. https://doi.org/10.1016/j.ejor.2012.01.055.

SIAAP, 2021. Lois sur l'eau. https://www.siaap.fr/le-siaap/lois-sur-leau/. (Accessed 5 July 2021).

SIMUL8, 2021. V.I.S.A Decisions Multicriteria Decision Making Software. http://www.visadecisions.com/index.php. (Accessed 7 June 2021).

Stewart, T.J., French, S., Rios, J., 2013. Integrating multicriteria decision analysis and scenario planning—review and extension. Omega 41, 679–688. https://doi.org/10.1016/j.omega.2012.09.003.

Venters, C., Lau, L., Griffiths, M., Holmes, V., Ward, R., Jay, C., Dibsdale, C., Xu, J., 2014. The blind men and the elephant: towards an empirical evaluation framework for software sustainability. J. Open Res. Software 2. https://doi.org/10.5334/jors.ao.

Voinov, A., Kolagani, N., McCall, M.K., Glynn, P.D., Kragt, M.E., Ostermann, F.O., Pierce, S.A., Ramu, P., 2016. Modelling with stakeholders – next generation. Environ. Model. Software 77, 196–220. https://doi.org/10.1016/j.envsoft.2015.11.016.

von Nitzsch, R., Tönsfeuerborn, M., Siebert, J.U., 2020. Decision skill training with the entscheidungsnavi. In: de Almeida, A.T., Morais, D.C. (Eds.), Innovation for Systems Information and Decision, Lecture Notes in Business Information Processing. Springer International Publishing, Cham, pp. 15–30. https://doi.org/10.1007/978-3-030-64399-7_2.

Walling, E., Vaneeckhaute, C., 2020. Developing successful environmental decision support systems: challenges and best practices. J. Environ. Manag. 264, 110513. https://doi.org/10.1016/j.jenvman.2020.110513.

Weistroffer, H.R., Li, Y., 2016. Multiple criteria decision analysis software. In: Greco, S., Ehrgott, M., Figueira, J.R. (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, International Series in Operations Research & Management Science. Springer, New York, NY, pp. 1301–1341. https://doi.org/10.1007/978-1-4939-3094-4_29.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Use R! Springer-Verlag, New York. https://doi.org/10.1007/978-0-387-98141-3.

Wulf, D., Bertsch, V., 2017. A natural language generation approach to support understanding and traceability of multi-dimensional preferential sensitivity analysis in multi-criteria decision making. Expert Syst. Appl. 83, 131–144. https://doi.org/10.1016/j.eswa.2017.04.041.

Xie, Y., Allaire, J.J., Grolemund, G., 2018. R Markdown the Definitive Guide. The R Series. Chapman and Hall/CRC.

Zheng, J., Egger, C., Lienert, J., 2016. A scenario-based MCDA framework for wastewater infrastructure planning under uncertainty. J. Environ. Manag. 183 (Part 3), 895–908. https://doi.org/10.1016/j.jenvman.2016.09.027.