

Accepted manuscript of article:

Haag, F., Reichert, P., Maurer, M., and Lienert, J. 2019. 'Integrating uncertainty of preferences and predictions in decision models: An application to regional wastewater planning', *Journal of Environmental Management*, 252: 109652. <https://doi.org/10.1016/j.jenvman.2019.109652>

Integrating uncertainty of preferences and predictions in decision models: an application to regional wastewater planning

Fridolin Haag^{a,b,*}, Peter Reichert^{a,b}, Max Maurer^{a,c}, Judit Lienert^a

^a Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland

^b ETH Zürich, Institute of Biogeochemistry and Pollutant Dynamics, Universitätstrasse 16, 8092 Zürich, Switzerland

^c ETH Zürich, Institute of Environmental Engineering, Stefano-Franscini-Platz 5, 8093 Zürich, Switzerland

*Corresponding author: fridolin.haag@eawag.ch; co-authors: peter.reichert@eawag.ch; max.maurer@eawag.ch; judit.lienert@eawag.ch

Abstract

Decision-making in environmental management requires eliciting preferences of stakeholders and predicting outcomes of alternatives. Usually, preferences and predictions are both uncertain. Uncertainty in predictions can be tackled by multi-attribute utility theory, but the uncertainty in preferences remains a challenge. We demonstrate an approach for including both uncertainties in a multi-criteria decision analysis (MCDA), using utility theory and the concept of expected utility. For a decision regarding a regional merger of wastewater infrastructure in Switzerland, we constructed preference models for four stakeholders. These models also allowed for non-additive interactions between objectives. We evaluated the performance of eleven decision alternatives for which we predicted potential outcomes. Even though uncertainties were high, we could draw conclusions based on the expected utility of alternatives. Building a pipeline to discharge treated wastewater to a larger river emerged as a potential consensus alternative to mitigate the problem of micropollutants in a small stream. We investigated the robustness of the findings with sensitivity analysis regarding the preference parameters and the included objectives. In their actual decision, the stakeholders partly preferred other alternatives than those proposed by the model. Their choices could be explained by reduced decision models in which only few objectives were included. This may indicate the use of simplified choice heuristics by the stakeholders. The presented approach is feasible for supporting other difficult environmental or engineering decisions in practice, for which we give a number of recommendations.

Keywords:

Multi-criteria decision analysis; Urban water management; Uncertainty; Stakeholder preferences; Multi-attribute utility theory; Adaptive utility

1. Introduction

The uncertainties in planning and decision-making often increase when considering larger temporal and spatial scales. The uncertainty concerns both the knowledge about consequences of implementing a decision alternative as well as the values of stakeholders and their valuation of these consequences. Such uncertainties are particularly prominent in long-term, strategic, or regional planning, such as the planning of wastewater infrastructure (e.g., Lienert et al. 2006; Dominguez et al. 2011; Zheng et al. 2016).

Wastewater infrastructure provides vital services to society. In European countries, it is largely characterized by capital-intense assets with long lifespans and interconnected system components (Markard 2011). These system characteristics pose challenges to strategic planning. Long life-spans and capital intensity entail strong path dependencies and limit adaptive approaches to management (Allen et al. 2011). Instead, planning has to account for the future uncertainties from the onset.

Strategic decision-making can be facilitated by various procedures and methods (e.g., Truffer et al. 2010; Gregory et al. 2012; Vacik et al. 2014; Scholten et al. 2017). While a structured decision-making process cannot guarantee an optimal outcome, it can facilitate a rational and transparent approach to the decision's complexity and uncertainty and facilitate the involvement of stakeholders. One valuable tool to support the process is a decision model (e.g., Reichert et al. 2007; Reichert et al. 2015; Scholten et al. 2015; Zheng et al. 2016).

Such a model consists of two connected parts: (1) an outcome prediction model for estimating the outcomes of the alternatives, and (2) a preference model for representing stakeholder values. The prediction model is not necessarily a mathematical model; for instance, it can also be the mental model of an expert. The preference model is defined over the outcome space of the predictions and returns a valuation of the potential outcomes. When applied to the predictions of decision alternatives, it can be used to rank the alternatives according to the preferences.

Specifying and identifying decision models is complicated by at least two aspects. (i) Predictions about the consequences of implementing an alternative are often difficult to obtain. A large portion of modeling in the natural and engineering sciences is focused on improving such predictions. How to conceptualize and tackle these uncertainties is not a focus in this study and has been explored elsewhere (e.g., Durbach and Stewart 2012; Kelly et al. 2013; Reichert et al. 2015; Uusitalo et al. 2015). (ii) Stakeholders might not have clear preferences, cannot express them (numerically), and the preference elicitation process can introduce uncertainty (e.g., Slovic 1995; Payne et al. 1999). To simplify elicitation, often assumptions are made about the form and structure of preference models, which might not hold in practice (Haag et al. 2019a).

Taken together, the uncertainties in predictions and preferences can limit the conclusions that can be drawn from a decision model. The uncertainty can overwhelm the actual differences

between alternatives (cf. Reichert and Borsuk 2005). The simplistic solution of ignoring uncertainties may occasionally work without impacting major conclusions, but is not a general solution (see Durbach and Stewart 2009; Scholten et al. 2017). Thus, we need (a) a conceptually sound approach to dealing with uncertainties in decision-making and (b) a way of integrating the uncertainties to allow making rational recommendations given the knowledge we have.

The established theory of expected utility (Keeney and Raiffa 1976; Eisenführ et al. 2010) prescribes a way for making rational decisions while considering the uncertainty in predictions. However, it assumes known preferences. Intuition and counter-examples show that preferences are also uncertain and change in time (e.g., Slovic 1995; Chajewska et al. 2000; Houlding and Coolen 2011). This led to the development of the *adaptive utility* concept (Cyert and DeGroot 1975), though this seemingly did not have a major impact on mainstream decision analysis. Still, the explicit consideration of preferential uncertainty is equally relevant for good decision-making as the consideration of prediction uncertainty.

Various approaches for dealing with uncertainty of preferences have been developed. Sensitivity analysis can be used to determine how strongly preferences may vary before conclusions would change (e.g., Lahdelma and Salminen 2001; Scholten et al. 2015), or robust inferences can be made directly given uncertain preferences (e.g. Greco et al. 2008). In the stochastic multicriteria acceptability analysis (SMAA) family of methods, different indices are suggested to find well performing alternatives despite uncertainty in weight parameters, based on distributions of rankings of the alternatives (Lahdelma et al. 1998; Hokkanen et al. 2000; Lahdelma and Salminen 2001; Tervonen and Figueira 2008). However, the conceptual underpinning why decisions should be based on one or another index is so far lacking.

An alternative and possibly more natural way of considering uncertainty in preferences is offered by the concept of *expected expected utility* (EEU; Chajewska et al. 2000; Boutilier 2003). It is essentially equivalent to the adaptive utility concept. EEU allows us to jointly average over the uncertainty in predictions and preferences, while sharing the same axiomatic basis as utility theory (see section 2.2). This provides a sound basis for rational decision-making given uncertainties in predictions and preferences. In addition to considering these uncertainties, analyzing the robustness of the results regarding underlying model structures and chosen parameter distributions is always important to assess the uniqueness or ambiguity of the choice of the optimal alternative (see Roy 2010).

The aim of this paper is to show how to practically tackle the consideration and integration of uncertainties in preferences and predictions. While multi-criteria decision analysis (MCDA) based on utility theory is well established in environmental planning (Cegan et al. 2017), we are not aware of any application of EEU in that area. A case study on regional planning of wastewater infrastructure in Switzerland serves as our reference and reflection point. We demonstrate how

the concepts can be implemented for such a complex, real-world case and investigate the robustness of the results for the actual decision. The main research questions are:

- (1) How can the EEU concept be applied in a complex decision? Is it feasible and beneficial for making choices under uncertainty?
- (2) What is the value of eliciting more detailed preference information in comparison to using simplified assumptions? Does the consideration of uncertainty in preferences influence the ranking of the alternatives?
- (3) Can the objectives hierarchy be simplified by omitting objectives, without losing much of the power to discriminate the alternatives?
- (4) How does the actual choices of the stakeholders in the case match to the results from the decision model?

2. Expected (expected) utility

2.1. Decision models based on multi-attribute utility theory

Rational decision-making requires predicting the potential outcomes of alternatives and knowledge about preferences for these outcomes. If we can specify these elements, the theory of expected utility (Eisenführ et al. 2010) offers a well-founded rationale for evaluating alternatives and for making (environmental) decisions (Reichert et al. 2007; Reichert et al. 2015). Expected utility theory assumes that uncertainties about outcomes of alternatives can be quantified as probabilities and stakeholders can specify their preferences.

A decision model can help to select one or a subset of alternatives out of a set of feasible alternatives $\mathcal{A} = \{a, b, \dots\}$. A potential decision outcome can be characterized by attributes $\mathbf{x} = (x_1, \dots, x_n)$. Attributes are properties to describe the state of a system. Our knowledge about the outcome of an alternative $a \in \mathcal{A}$ is described either (a) by the levels of the attributes when implementing the alternative, \mathbf{x}_a , when disregarding uncertainty, or (b) by a probability distribution $p_a(\mathbf{x})$ that quantifies the uncertain knowledge about the outcomes.

We care about the outcomes of alternatives because they impact what we value. Taking such a perspective of value-focused thinking (Keeney 1992), alternatives are evaluated in terms of a stakeholder's objectives. We can think of objectives as qualitatively specifying the values of a stakeholder and thus determining the structure and extent of the evaluation. In all but the simplest cases, multiple objectives are relevant. In such multi-criteria decisions, the objectives are often structured in the form of a hierarchy with higher-level objectives and sub-objectives (Keeney and Raiffa 1976). In multi-attribute utility theory, this hierarchy can be used to structure the quantification of a stakeholder's preference.

The preferences of a stakeholder can be represented by a value or utility function. Let X denote the set of all possible outcomes for all attributes considered relevant in the decision. A

multi-attribute value function $v: X \rightarrow [0,1]$ returns the valuation of potential outcomes such that larger values represent preferred states (Keeney and Raiffa 1976). When the outcomes are not uncertain, a ranking of the alternatives can be found after calculating the value $v(a) = v(x_a)$ for each alternative in \mathcal{A} . A decision is rational, if the alternative with the highest value is chosen.

For decisions under risk, i.e., if decision outcomes are uncertain, a multi-attribute utility function $u: X \rightarrow [0,1]$ returns the valuation of potential outcomes (Keeney and Raiffa 1976). A utility function not only considers preferences about outcomes but also preferences about risk. For example, a stakeholder can be risk averse and have a preference to avoid an uncertain situation (see Eisenführ et al. 2010 for discussion). If a stakeholder is risk neutral, the utility function has the same shape as the value function. In the case of continuous distributions, the expected utility of an alternative a is given by:

$$EU(a) = \int_{\mathbf{x}} u(\mathbf{x}) p_a(\mathbf{x}) d\mathbf{x} \quad (\text{Equation 1})$$

By calculating the expected utilities of the considered alternatives, a ranking can be found. A decision is rational, if the alternative with the highest expected utility is chosen.

Directly identifying the multi-attribute utility function $u(\mathbf{x})$ of a stakeholder is difficult; instead, it can be constructed in a stepwise and hierarchical manner (Reichert et al. 2015; Haag et al. 2019a). Once we elicited the model structure in form of an objectives hierarchy, constructing a utility function for the evaluation of the overall objective can be partitioned into three steps.

First, each lowest-level objective is evaluated with respect to its attributes $\{x_k, \dots, x_l\}$ with a value function $v_i(x_k, \dots, x_l, \theta_i)$ with parameters θ_i . Normally, due to the narrow definition of these objectives, the value function only depends on a single attribute or few attributes.

Secondly, each higher-level objective is evaluated by aggregating the evaluations of its sub-objectives. We also call this objective an *aggregation node*. It depends on the attributes only indirectly. The multi-attribute value function over lower-level objectives $\{o_p, \dots, o_q\}$ on a specific hierarchical level can be written as:

$$v_{p,q}(x_1, \dots, x_n, \theta) = F(v_p(x_1, \dots, x_n, \theta_p), \dots, v_q(x_1, \dots, x_n, \theta_q), \theta_{pq}) \quad (\text{Equation 2})$$

In practice, each value function v_i will only depend on a subset of the attributes $\{x_1, \dots, x_n\}$. The function F is an aggregation function (Grabisch et al. 2009). The aggregation function depends on the values of the sub-objectives. Form and parameters of the function depend on a stakeholder's preferences. For a hierarchy with several levels, evaluations are aggregated in a stepwise manner along the hierarchy, until an overall evaluation is reached. This means that multiple aggregation functions are nested according to the hierarchical structure (Haag et al. 2019a).

The predominant aggregation function in applications is the weighted arithmetic mean. The resulting preference model is called the *additive model* (Keeney and Raiffa 1976). Assuming that each lowest-level value function depends only on one attribute, an additive (multi-attribute) value function has the form:

$$v(x_1, \dots, x_n, \Theta) = \sum_{i=1}^n w_i \cdot v_i(x_i, \theta_i) \quad (\text{Equation 3})$$

with parameters $\Theta = \{\theta, \mathbf{w}\}$ and weights $\mathbf{w} = (w_1, \dots, w_n)$ that sum to unity. However, this model is only a valid representation of a stakeholder's preferences if certain independence conditions are fulfilled (see Keeney and Raiffa 1976; Dyer and Sarin 1979; Eisenführ et al. 2010; Haag et al. 2019a for discussions). Many alternative aggregation functions with other premises exist (Grabisch et al. 2009).

Thirdly, a value function can be converted to a utility function, given the risk attitude of the stakeholder (Dyer and Sarin 1982). We only consider the case in which the value function is converted to a utility function at the highest hierarchical level, as proposed by Reichert et al. (2015). This requires knowledge about the stakeholder's risk attitude at this hierarchy level.

The objectives hierarchy determines the structure of the preference model and entails certain independence conditions (Haag et al. 2019a). If the aggregation function that is used is *decomposable*, the hierarchical structure of the preference model has no influence on the results, given proper weighting of objectives. Decomposability means that each element of a subset of arguments can be replaced by their partial aggregated value without changing the overall aggregated value (see Grabisch et al. 2009 p. 35). Then, the order in which values are aggregated does not matter. The arithmetic mean of the additive model is such a decomposable aggregation function. However, if (combinations of) other aggregation functions are used, the hierarchical structure can influence the results. This requires a careful structuring of the hierarchy according to the preferences of the stakeholder.

2.2. Expected expected utility

With utility functions we can calculate the expected utility for an alternative, conditional on the preferences and the uncertain outcomes. In case the preferences are uncertain, too – i.e., the utility function has an uncertain structure or parameters – this results in a distribution of expected utilities. Expected utility theory does not provide a rationale how to select or order alternatives based on such distributions of utilities.

To overcome this limitation, we propose to use the concept of expected expected utility (EEU), as has been formalized by Boutilier (2003). If a stakeholder is unsure about her or his utility function, she or he is unsure which expected utility she or he will receive when choosing an alternative. A rational decision-maker would now select the alternative with the highest expected value of these expected utilities, this means, the highest EEU.

The EEU of an alternative a , with the probability of outcomes \mathbf{x} given by $p_a(\mathbf{x})$, for a stakeholder m with a parameterized utility function $u(\mathbf{x}, \boldsymbol{\psi})$ and a distribution of these parameters $p_m(\boldsymbol{\psi})$, can be written as:

$$EEU(a, m) = \int_{\boldsymbol{\psi}} \left(\int_{\mathbf{x}} u(\mathbf{x}, \boldsymbol{\psi}) p_a(\mathbf{x}) d\mathbf{x} \right) p_m(\boldsymbol{\psi}) d\boldsymbol{\psi} = \int_{\boldsymbol{\psi}} \int_{\mathbf{x}} u(\mathbf{x}, \boldsymbol{\psi}) p_a(\mathbf{x}) p_m(\boldsymbol{\psi}) d\mathbf{x} d\boldsymbol{\psi}$$

(Equation 4)

The inner integral $-\int_{\mathbf{x}} u(\mathbf{x}, \boldsymbol{\psi}) p_a(\mathbf{x}) d\mathbf{x}$ – is simply the expected utility for specific parameter values $\boldsymbol{\psi}$ (Equation 1).

Boutilier (2003) has shown that under specific conditions EEU can be justified by appeal to the foundational axioms of decision theory. More precisely, all the considered utility functions have to be extremal equivalent. This means that all utility functions need to assign the same value to the best outcome (e.g. 1) and the same value to the worst outcome (e.g. 0). In our practical application this is given. However, this is an important condition, as in expected utility theory the utilities are only unique up to an affine transformation. This degree of freedom must be constrained if we are taking the expectation over a distribution of utilities.

3. Implementation

3.1. Application: a regional merger of wastewater treatment plants

3.1.1. Decision problem, stakeholders, and process

The decision case concerned a potential regional merger of wastewater treatment plants (WWTPs) in Switzerland. In a merger the number of WWTPs is reduced and the sewer lines are adapted correspondingly. In the presented case, the fragmented organizational structures would be merged in parallel with the WWTPs. At the moment, there are four WWTPs in the region. Three (NW, AZ, GM) discharge their treated wastewater into the larger river RA (Figure 1). One (BB) discharges into the smaller stream NZ.

As the first country worldwide, Switzerland demands certain WWTPs to take measures to remove micropollutants such as pharmaceuticals or pesticides (Eggen et al. 2014; Logar et al. 2014). Currently, none of the WWTPs in the region has a specific treatment for micropollutants. The WWTP BB is required to take measures until 2025, because the stream NZ is small and potentially negatively impacted. Micropollutants could be eliminated at WWTP BB if an additional treatment step was built. Alternatively, the wastewater could be rerouted to another WWTP and the WWTP BB could be decommissioned. This prompted a search process for a regional solution for the wastewater system with the year 2040 as the time horizon.

Currently, all WWTPs have mechanical and biological treatment, including nitrification and denitrification (Table SI 1). The WWTPs have some idle capacity concerning the pollution load, i.e., they are designed to treat a higher load than they currently receive. However, in case of mergers, extensions may be required. Currently, a lot of parasitic water (water that would not need treatment, for instance groundwater) enters the system (28%–69%). By reducing this amount, increases in the required hydraulic capacity due to mergers could be absorbed. At three WWTPs (BB, NW, GM), the sludge that accumulates in the treatment processes is digested for biogas

production. The gas is used in a combined heat and power plant at BB and GM, and at NW the biogas is utilized for a district heating system (Table SI 1).

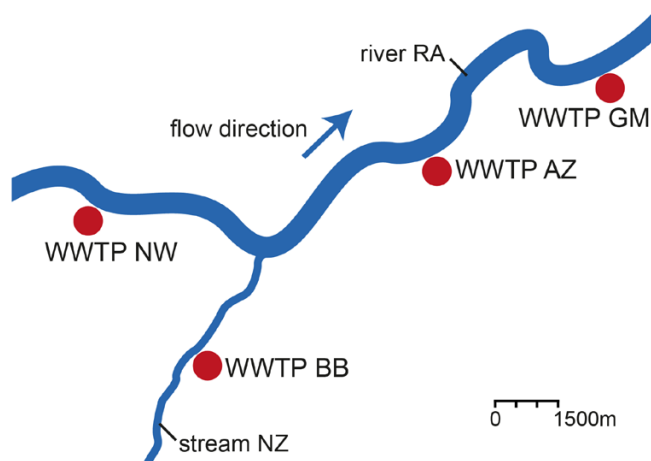


Figure 1: Conceptual map of the spatial configuration of the wastewater treatment plants (WWTP) in the case study. Further information on the WWTPs is given in Table SI 1.

The decision process was initiated by the cantonal authorities responsible for wastewater (stakeholder CT, Table 1). Four organizations that each operate a WWTP in the region were the other main stakeholders (Table 1). These organizations are responsible for the wastewater treatment of 9 to 18 associated municipalities each and represented their interests in the decision process. A decision committee was set up in which each of the five organizations had one or two delegates. An engineering consultancy facilitated the multi-organizational group decision process. Two further engineering offices were involved as consultants. The research team acted as adjunct consultants who conducted a MCDA and gave inputs.

Table 1: Organizations that were stakeholders in the decision process. The organizations' names are anonymized. The persons interviewed for this study were all delegates to the decision committee, with exception of the operator of organization GM.

Stakeholder / organization	Municipalities affiliated with organization	Inhabitants represented	Interviewees in this study
CT	n/a	n/a	The department head of the responsible cantonal authority
BB	13	15500	The president and the vice-president of the organization
GM	12	13500	The president of the organization and the chief plant operator
NW	9	8500	The chief plant operator
AZ	18	41000	Not interviewed

The decision process was organized in several phases. For the first phase, which took place in 2017, the aim was to structure the problem, generate alternatives, and select a subset of feasible alternatives for further analysis. In several meetings, the representatives of the organizations, the

cantonal authorities, the consulting engineers, and two of the authors met to discuss. The project aim for us as researchers was to support the selection of alternatives by building and analyzing a decision model. For this, we elicited preferences of representatives that were delegates in the decision committee (Table 1). From organization AZ, no representative was available for an interview at the time. Therefore, the possibility of extrapolation of the individual results to the group level is limited. We aimed at determining promising alternatives, given the high uncertainty in the preferences of the stakeholders and in the long-term consequences.

3.1.2. Decision objectives, attributes, and alternatives

Objectives were identified by the stakeholders in a workshop in April 2017 by simple brainstorming. The research team carried out a preliminary structuring using clustering and means-ends networks (Keeney 1992). The set of objectives was further developed by the lead engineering office. In a meeting in August 2017, the involved actors agreed to use this system of objectives for evaluation. For our MCDA we made two adaptations: we split the objective of water protection into one objective regarding micropollutants and one regarding the ecological state and removed the objective of political feasibility, as this is a constraint for the feasible set of alternatives rather than an objective. This led to a system of 14 objectives (Table 2).

Due to time constraints, the attributes (Table 2) were selected by the lead engineering office and the research team and not defined jointly with all stakeholders. Many fulfill the criteria for useful attributes as proposed by Keeney and Gregory (2005). However, for some objectives, such as the “Long-term solution” or “Flexibility to react”, we had to resort to proxy attributes.

Table 2: System of objectives for a good wastewater disposal for the region in 2040. The description of how the attributes were predicted is given in Table SI 3; a visualization of the objectives hierarchy in Figure SI 2. WWTP: wastewater treatment plant; Nm³: normal cubic meter.

Higher level objective	Objective	Short name	Description	Attribute
Economic viability	Low annual cost	cost	Annual cost for the wastewater treatment for the whole region (operational costs, depreciation, interest) in 2040.	Annual cost in 2040 [mio. CHF/year]
	Good ecological state of stream NZ	ecology_nz	Water protection is an important legal objective in Switzerland. Stream NZ is small and thus sensitive to the impacts of (treated) wastewater.	Ecological state of river NZ as given by the Swiss modular stream assessment* [points]
Environmental protection	High removal of micropollutants	micropollutants	Certain WWTPs are obligated to implement additional treatment to remove micropollutants. Beyond this legal requirement, this is also a general objective for the improvement of water quality.	Population for which micropollutants will be removed [number of people]
	Low impairment of protected areas	protected_areas	Minimize impacts to legally protected areas such as nature, landscape, or river banks. Extensions of WWTPs or other new buildings can impact such	Number of protected areas that could be impaired [areas impaired]

Higher level objective	Objective	Short name	Description	Attribute
Protection of resources	High amenity of the landscape	landscape	protected zones. The visual amenity of the landscape can be negatively impacted by large buildings of WWTPs, for example digester towers.	Number of buildings of WWTPs higher than 10m [buildings]
	Low electrical energy consumption	electricity	The wastewater treatment itself requires electrical energy. Additionally, energy is required for pumping wastewater and micropollutant removal.	Gross electrical energy required [MWh/year]
	High utilization of energy in sludge	sludge_utilization	Sewage sludge contains considerable amounts of energy, which should not be wasted. The sludge can be digested for producing biogas. Drained sludge can also be thermally utilized, either after a digestion or directly.	Production of biogas from sludge [thous. Nm ³ /year]
	Little land consumption	land_consumption	Little land should be used for buildings of wastewater infrastructure. Sometimes, additional land would need to be acquired, which can be difficult.	Land requirements for infrastructure [m ²]
Social compatibility	Preservation of jobs in the wastewater sector	jobs	The jobs in the wastewater sector of the region should be preserved.	Jobs in the wastewater sector in the region [full time equivalents]
	High degree of co-determination for municipalities	co_determination	Co-determination and direct democracy are important values in Switzerland. People should be able to influence decisions concerning wastewater management.	Number of municipalities in the largest organizational unit in the region [municipalities]
	Few nuisances to residents	residents	Wastewater treatment can have effects such as noise, odor, or traffic that affect nearby residents.	Number of residents in 1 km radius around WWTP that will possibly be negatively affected by a system change [residents]
Long-term perspective	Long-term solution	long_term	The alternative should be robust, i.e., it should be functional in 2040, even when changes in the surrounding conditions occur. This also should reduce the burden to coming generations.	Estimate of long-term suitability of the alternative on a four point scale [points]
	Flexibility to react	flexibility	The system should remain flexible to be changed when new changes in surrounding conditions occur.	Number of further alternatives that remain feasible once the alternative has been realized [paths]
	Synergies with district heating for organization NW	district_heating	The organization NW has a district heating system which also uses biogas. This biogas is produced directly at their WWTP. The sludge from wastewater is one important energy source.	Local production of biogas from sludge at WWTP NW [thous. Nm ³ /year]

In a workshop in April 2017, the stakeholders and consultants generated conceptual alternatives and the lead engineering office specified them further. In a next meeting, consensus on the alternatives to investigate was reached. Five strategies remained which were specified as ten concrete alternatives (Table 3). Some can be realized in a step-wise manner. For instance, it is possible to first divert the treated effluent of WWTP BB from stream NZ to the large river RA (alternative 1B) and later extend this rerouting to a larger WWTP and decommission WWTP BB (alternatives 3 and 4).

Further alternatives had been discussed, but were not pursued (Table SI 2). Additional clarification or “back-of-the-envelope” calculation showed that they would likely be dominated by other alternatives or were infeasible within the required time frame. Of course, a premature exclusion of alternatives, narrow thinking, and a fixation on the status quo can be detrimental for good decisions and is a common bias in decision-making (Payne et al. 1999; Gregory et al. 2012; Montibeller and von Winterfeldt 2015). For this study, we introduced a do-nothing alternative (0) for comparison, even though it would not achieve the regulatory requirements.

Table 3: Alternatives under consideration in this study. See Figure 1 for a spatial overview and Table SI 1 for the status quo of the wastewater treatment plants (WWTPs).

Type	Alternative	Description
0	0	“Do nothing”. Keep the configuration of four WWTPs as it is. Only maintenance and rehabilitation work and, if needed, extension.
		Keep the general configuration of four WWTPs as is. Only the WWTP BB will take measures to reduce the discharge of micropollutants to stream NZ; all other WWTPs remain unchanged.
1	1A	A treatment step for micropollutant removal will be built for WWTP BB.
	1B	A gravity pipeline will be built to route the treated wastewater from WWTP BB to the larger river RA, so that no treated wastewater flows into the stream NZ. The pipeline would be built along the stream NZ.
2	2	Wastewater arriving at WWTP BB will be pumped to WWTP NW; these two systems and organizations will be merged. WWTP BB will be decommissioned. The other two WWTPs will remain independent.
		Wastewater arriving at WWTP BB will be pumped to WWTP AZ; these two systems and organizations will be merged. WWTP BB will be decommissioned. The other two WWTPs will remain independent.
3	3A	The sludge treatment at WWTP AZ will be kept as is.
	3B	A sludge digestion will be built at WWTP AZ.
		Wastewater arriving at WWTP BB and at WWTP NW will be pumped to WWTP AZ; these three systems and organizations will be merged. WWTP BB and NW will be decommissioned. WWTP GM remains independent.
4	4Aa	The sludge treatment at WWTP AZ will be kept as is.
	4Ab	The sludge treatment at WWTP AZ will be kept as is. In addition, a micropollutant removal step for WWTP AZ will be built.

Type	Alternative	Description
	4Ba	A sludge digestion will be built at WWTP AZ.
	4Bb	A sludge digestion will be built at WWTP AZ. In addition, a micropollutant removal step for WWTP AZ will be built.
5	5	Merger of all WWTPs to one system and organization. The site for this one WWTP remains open (possibly the WWTP AZ). The WWTP will have state-of-the-art micropollutant removal and sludge utilization. All other WWTPs will be decommissioned.

3.2. Generation and modeling of predictions and their uncertainties

The research team and the lead engineering office quantified the predictions and their uncertainty for each alternative and each attribute. The predictions were based either on simple models, expert estimates, or resulted directly from the definition of the alternatives (details are provided in Table SI 3). External uncertainty, e.g., concerning the future development of the region, was included where feasible. For instance, we calculated with uncertain population growth between 0 and 25 percent for predicting micropollutant removal, electricity consumption, and biogas production. The temporal framing was the situation in the year 2040. A regional perspective was taken: Impacts were calculated jointly for the region, not for individual organizations. Theoretically, a fair allocation of benefits and costs should later be possible if the situation was improved overall.

The uncertain predictions for the attributes were modeled with different probability distributions. Independence of the distributions for different attributes was assumed. The joint probability density of all attributes was thus calculated as the product of the probability densities of the individual attributes. Probability distributions with tails extending to infinity, for instance, all normal distributions, were truncated to reasonable ranges¹. The value functions were defined over these ranges.

The distributions were constructed by different methods: (1) some prediction models returned a distribution of outputs. For convenience, we used Monte-Carlo simulation to create an output sample and then fitted a parametric distribution – usually a normal distribution – to this sample. Histogram plots were used to evaluate the fit. (2) If a central point and a range was given (e.g. for costs), we assumed a normal distribution centered on this point with a standard deviation of $\frac{1}{4}$ of the range (i.e., about 95% of the values lie within the range). (3) If there were discrete possible outcomes (e.g. between 7 and 9 buildings), we modeled them with a uniform distribution.

¹ The minimum of the range was determined by subtracting twice the standard deviation from the mean values of alternatives and taking the minimum and vice versa for the maximum of the range. These numbers were then rounded.

3.3. Elicitation and modeling of preferences and their uncertainties

3.3.1. The preference model

Based on the empirical data gathered in the case study, we built a preference model as described in section 2.1. The objectives were structured hierarchically (Table 2, Figure SI 2). The model has the form:

$$u(x_1, \dots, x_{14}, \boldsymbol{\theta}, \mathbf{w}, \boldsymbol{\gamma}, r) = u(F_0 (v_1(x_1, \theta_{v1}), F_2(v_2(x_2, \theta_{v2}), v_3(x_3, \theta_{v3}), v_4(x_4, \theta_{v4}), v_5(x_5, \theta_{v5}), \mathbf{w}_{F2}, \gamma_{F2}), F_3(v_6(x_6, \theta_{v6}), v_7(x_7, \theta_{v7}), v_8(x_8, \theta_{v8}), \mathbf{w}_{F3}, \gamma_{F3}), F_4(v_9(x_9, \theta_{v9}), v_{10}(x_{10}, \theta_{v10}), v_{11}(x_{11}, \theta_{v11}), \mathbf{w}_{F4}, \gamma_{F4}), F_5(v_{12}(x_{12}, \theta_{v12}), v_{13}(x_{13}, \theta_{v13}), v_{14}(x_{14}, \theta_{v14}), \mathbf{w}_{F5}, \gamma_{F5}), \mathbf{w}_{F0}, \gamma_{F0}), r) \quad (\text{Equation 5})$$

with lowest-level value functions v_i , aggregation functions F_j , and a utility conversion function u .

In public infrastructure decisions usually multiple stakeholders are involved. Since utility theory focuses on individuals, each stakeholder could have their own model structure. In practice, it often makes sense to develop the model structure jointly, as this facilitates discussion. We also assumed a common model structure.

Based on the statements of the stakeholders (Table SI 5), the lowest level value functions v_i were modeled as (i) exponential functions (as in Scholten et al. 2015), (ii) sigmoid functions, or (iii) linear functions. The exponential function (i) has the form: $v(x_i, \theta) = \begin{cases} \frac{1 - \exp(-\theta \cdot \tilde{x}_i)}{1 - \exp(-\theta)} & \text{for } \theta \neq 0 \\ \tilde{x}_i & \text{for } \theta = 0 \end{cases}$, with $x_i \in [x_i^+, x_i^-]$, $\tilde{x}_i = (x_i - x_i^-)/(x_i^+ - x_i^-)$, and $\theta \in \mathbb{R}$. For the sigmoid function (ii), we used the cumulative density function of the beta distribution with parameters α, β : $v(x_i, \alpha, \beta) = \frac{B(x_i; \alpha, \beta)}{B(\alpha, \beta)}$ with $B(x; \alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt$ and $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$. The linear function (iii) is: $v(x_i) = (x_i - x_i^-)/(x_i^+ - x_i^-)$. All the lowest-level value functions were single-attribute value functions.

F_j is an aggregation function. We decided to not only use the weighted arithmetic mean of the additive model. Evidence is growing that in practice non-additive preferences are an important phenomenon (e.g., Rowley et al. 2012; Langhans and Lienert 2016; Haag et al. 2019a; Reichert et al. *subm.*). Instead, we chose to use the weighted power mean

$$F_j(\mathbf{v}, \mathbf{w}, \gamma_j) = (\sum_{i=1}^n w_i v_i^{\gamma_j})^{\frac{1}{\gamma_j}} \quad (\text{Equation 6})$$

with parameters $0 < \gamma_j \leq 1$ and \mathbf{w} a vector of weights, which sum to one. We call the parameter γ_j the *non-additivity parameter*, as it determines the degree of non-additivity. The function would take on further parameter values for γ_j ; however, by restricting γ_j to $]0,1]$ it is already able to

represent a wide range of relevant behaviors from the geometric mean ($\gamma_j \rightarrow 0$) to the arithmetic mean ($\gamma_j = 1$). In a previous study, the preferences of decision-makers always indicated values of γ_j in the range of 0.2 to 1 (Haag et al. 2019a). For a constant value of γ_j , i.e., $\gamma_j = c \forall j$, the weighted power mean is decomposable, given proper normalization of weights.

At the highest hierarchical level, the multi-attribute value function was converted to a utility function using an exponential function:

$$u(v(a)) = \begin{cases} \frac{1 - \exp(-r \cdot v(a))}{1 - \exp(-r)} & \text{for } r \neq 0 \\ v(a) & \text{for } r = 0 \end{cases} \quad (\text{Equation 7})$$

with $r \in \mathbb{R}$ (e.g. Keeney and Raiffa 1976). For $r = 0$ the stakeholder is risk-neutral, for $r < 0$ the stakeholder is risk-averse (concave utility function), for $r > 0$ the stakeholder is risk-seeking (convex utility function).

3.3.2. Elicitation of preference model parameters

The preferences of the stakeholders were elicited in face-to-face interviews. We used paper-based tools similar as described elsewhere (Scholten et al. 2015; Zheng et al. 2016), but implemented these in the form of cards that could be moved around by the stakeholders (Figure SI 3). We interviewed one representative of each of the organizations NW and CT, and two representatives of each of the organizations GM and BB (Table 1). The interviewees were asked to answer the questions in their role as representatives of the organizations. In case two interviewees were present, they discussed until they reached consensus.

Single attribute value functions: Value functions were elicited with the bisection method (mid-value splitting method) (Keeney and Raiffa 1976). Due to the inclusion of the do-nothing alternative and improvements in prediction modeling during the project, we had to extend or shift the ranges of four attributes after elicitation. A total of four elicited value functions needed to be transformed to these ranges using linear extrapolation (see Table SI 5 for details). We had no indication that a more sophisticated extrapolation would be better. For time reasons, we only elicited the value functions concerning objectives with the highest weights. Depending on the stakeholder, we collected preference information for 2–8 value functions.

Aggregation parameters: We elicited weights with the Swing method (Eisenführ et al. 2010). Because hierarchical weighting can lead to biases (Marttunen et al. 2018), we elicited weights in a non-hierarchical way. First, participants ranked and gave points to objectives in each branch separately; secondly, the highest-ranked objectives of each branch were compared and evaluated with Swing (Figure SI 3). From these $n - 1$ comparisons we calculated the weights. To account for the extension of ranges, an adjustment factor was included for three weights (see Table SI 5 for details). We did not quantify the non-additivity parameters γ_j in the interviews, but assumed a distribution that covers the expected range, including additive aggregation (Table 4).

Risk attitude: The assumptions about the distribution for the risk attitude at the highest level, which we used in the modeling (Table 4), was informed by the risk attitude for single attributes. We elicited this risk attitude for one or two attributes via the certainty equivalent (CE) method with a lottery question (Eisenführ et al. 2010). From this, we determined the risk attitude parameter r (Table SI 6). We propagated the uncertainty in the value functions to obtain the empirical distributions of risk attitudes.

3.3.3. Characterization of preference uncertainties

From the statements elicited from the interviewees (Table SI 5), we inferred the parameters of the preference model and their uncertainty distributions for each stakeholder (Table 4, “Uncertain information”). Because elicitation time was limited, we extrapolated uncertainty statements to cases where we did not elicit information. Even though the level of detail was higher than in many applications, the preference information was still rather rough, especially regarding the parameter distributions. Ideally, the parameters and the parametric uncertainty can both be inferred from statements of stakeholders (e.g., Haag et al. 2019a).

To explore how the level of detail in the preference information influences the results, we determined a second set of assumptions for each group of parameters (Table 4, “Simplifying assumptions”). For the lowest-level value functions, this assumption was linearity. For the weights, we assumed that (A) there was no uncertainty, and (B) only a ranking was known. For the aggregation functions, we assumed the weighted arithmetic mean for all nodes (the additive model). Lastly, for the risk attitude we assumed risk neutrality. The effect of these simplifications was tested (section 3.4.2).

Table 4: Preference information used for modeling. The uncertain information (left) is based on elicitation in stakeholder interviews. The simplifying assumptions (right) are often assumed in practice applications and require less elicitation effort.

Preference element	Uncertain information	Simplifying assumptions
Lowest-level value functions	a) Shape convex or concave: a1) x for $v(x) = 0.5$ and shape known: Determine θ^* of exponential function based on the one point. Modeling as exponential function with θ normally distributed with $\mu = \theta^*$, $sd = 0.25$.	Assumption of linear value functions, no uncertainty.
	a2) Only shape known: Modeling as exponential function with θ uniformly distributed in $[0, 5]$ if concave or $[-5, 0]$ if convex.	
	b) Shape linear: Modeling as exponential function with θ normally distributed with $\mu = 0$, $sd = 0.25$	
	c) Shape sigmoid: Propagate uncertainty in $x_{0.1}$, $x_{0.5}$, and $x_{0.9}$ to find empirical distributions of function parameters α, β . Fitting of separate normal distributions to these empirical distributions.	

Preference element	Uncertain information	Simplifying assumptions
	d) No information on shape: Modeling as exponential function with θ uniformly distributed in $[-5,5]$. This was the case for 6–12 value functions, depending on the stakeholder.	
Weights	Non-hierarchical Swing weight elicitation. Forward propagation of uncertainty in assigned Swing points to calculate the empirical distribution of global weights. Uncertainty in assigned points: If a range was given by stakeholder, assumption of normal distribution with μ middle of range and $sd = \frac{1}{4}$ of range. If point estimate was given, assumption of normal distribution with $\mu =$ point estimate and $sd = 0.05$. Calculation of hierarchical local weights by summation and normalization. Global weights are shown in Figure SI 4.	A) Using precise weights and disregarding uncertainty in the statements.
		B) Ranking weights: The ranking of weights on each hierarchical level was based on the Swing weights. On each level with m objectives, sampling from a uniform distribution in the m -dimensional simplex considering the rank constraints. Weights were jointly sampled since they were dependent. The Hit-And-Run Markov Chain Monte Carlo sampling algorithm was used as implemented by Tervonen et al. (2013). In addition to respecting the ranking constraints, this ensured that each weight was in $[0,1]$ and the weights summed to one.
Aggregation parameter γ	On each hierarchical level and for each aggregation node j : Assumption of γ_j to be beta distributed on the interval $[0.1, 1]$ with $\alpha = 1.5$ and $\beta = 1$. This results in an expected value of $\gamma \approx 0.64$. With this value of γ the function is slightly non-additive, but not as extreme as the geometric mean ($\gamma \rightarrow 0$).	Assumption of additive model, i.e. $\gamma_j = 1$ for all nodes.
Overall risk attitude	Assumption that decision makers are generally not risk-seeking in infrastructure decisions. Assumption of the overall risk parameter to be uniformly distributed on the interval $[0,4]$. The elicited risk parameters were all <4 (Table SI 6).	Assumption of risk neutrality. Risk parameter $r = 0$.

3.4. Analysis

3.4.1. Modeling

The inputs and the parameters of our decision model were all uncertain, but specified by probability distributions. Sometimes these distributions were empirical distributions created by forward propagation of uncertainties, sometimes they were specified directly. To propagate the uncertainties through our model we used Monte Carlo simulations with 10000 samples. Each input and parameter was sampled independently, except for the weights in the ranking information case, which were sampled jointly (see Table 4). We randomly drew from the specified distributions and calculated the overall utility of each alternative with this sample. Thereafter, we calculated the EEU for each alternative and stakeholder (Equation 4). In practice, this meant taking the arithmetic mean of the utility over all Monte Carlo runs. By comparing the EEU of different alternatives we created a ranking.

For the comparison with ranking-based indices (see Figure SI 5), we drew subsamples with size 1000 of both the prediction and preference parameters. For each of the 1000 parametrizations of the preference model we calculated the EU given the prediction sample. Based on the EUs we calculated the ranking of alternatives and the rank distributions. Based on these distributions several indices could be calculated.

The modeling was conducted in R (R Core Team 2018). We used base R and “tidyverse” functionalities (<https://www.tidyverse.org/>) for most data handling and modeling, and the package “utility” (Reichert et al. 2013) for lowest-level value functions.

3.4.2. Sensitivity to assumptions about preferences

To test the sensitivity to the level of detail of preference information as described in Table 4, we repeated the analysis for six different layouts: (1) Uncertain information for all parameters, left column of Table 4; (2) assumption of linear value functions, uncertain information otherwise; (3) disregarding uncertainty in weights and using precise weights, uncertain information otherwise; (4) only ranking information on weights, uncertain information otherwise; (5) assumption of additive model, uncertain information otherwise; (6) assumption of risk neutrality, uncertain information otherwise; and (7) using all these simplifications (with precise weights), right column of Table 4. To determine the similarity in the rankings of alternatives for these parameter sets, we used Kendall’s rank correlation coefficient τ (Kendall 1938) as a similarity measure (e.g. Zheng et al. 2016).

3.4.3. Sensitivity to leaving out objectives

To determine the sensitivity to the different objectives, we used a form of leave-one-out cross-validation. This meant, we excluded one of the 14 objectives from the decision model, renormalized the weights to one, and evaluated the changes in the outcomes by simulation. This allowed determining the contribution of the objective to differentiating alternatives. The sensitivity of the result to including or excluding an objective depends on the objective’s weight, the dissimilarity of the alternatives performance with regard to that objective, and, in non-additive models, on the degree of interaction with other objectives. Thus, this is more conveniently explored by simulation than analytically. To determine the similarity in the rankings of the alternatives for different models, we again used Kendall’s τ (see above).

3.5. Comparison to actual choices

In a concluding workshop of the first project phase in November 2017, the participating representatives of the organizations made a direct choice about which alternatives to investigate in more detail in a second phase. We term this *holistic choice*. The stakeholders assigned green points to alternatives that they believed should be investigated further and red points to alternatives that they would exclude. By subsequent discussion, which included argumentation

for choosing or not choosing an alternative, a consensus was found. Not all interviewed representatives attended the meeting and of organization NW only a representative who we did not interview attended.

The stakeholders did not know the decision model results when they made their holistic choice. This gave us the chance to compare this choice and the model results. Depending on how many green and red points the stakeholders allocated, we selected the corresponding number of best and worst alternatives based on their ranking for each stakeholder (Table SI 7).

Secondly, we investigated whether the holistic choices could be approximated by decision models in which only few objectives were considered. In the full decision model, a broad range of concerns was included, specified as fourteen objectives (Table 2). From discussions in the group and in the interviews, we concluded that few of these objectives had high salience for the stakeholders. For each stakeholder, we built a *reduced decision model* that only included a subset of objectives. These reduced models are not a truthful representation of the actual thinking of the stakeholders, but an exploration of potential models. We identified possible important concerns of the stakeholders based on our knowledge and their statements in the interview and discussions.

For stakeholder CT, assumed key concerns were an economically efficient and long-term solution, the ecological state was seen as a hard constraint: the objectives cost and long_term were included (see Table 2 for an explanation of the abbreviations). For stakeholder BB, next to cost and the solution of the issues for stream NZ, using the energy from sludge and keeping options open were important concerns: objectives cost, ecology_nz, sludge_utilization, and flexibility were included. This was similar for stakeholder GM, but the sludge utilization was seen as less of an issue: objectives cost, ecology_nz, flexibility were included. For stakeholder NW, the local district heating system was a key constraint; additionally, the basic problem of the river NZ should be solved: objectives ecology_nz and district_heating were included. Again, for each stakeholder we classified the alternatives based on the resulting ranking with the reduced models (Table SI 10) and depending on the number of allocated points in the holistic choice.

Lastly, we compared the fourteen objectives of the full decision model with the results of an online survey on the decision objectives stated by nine participants involved in this decision (two per organization BB, NW, MG, AZ, and CT). The online survey was conducted in May and June 2018, about seven months after the concluding workshop; the details are given in Haag et al. (2019b). This survey was conducted in five different decisions on wastewater planning and aimed at testing different brainstorming methods to collect objectives online.

4. Results

4.1. Expected expected utility of alternatives

When we evaluate our decision model conditional on the best available information concerning the uncertainty distributions of the outcomes (Table SI 4) and the preference parameters (Table 4), we receive a wide distribution of overall utilities for each alternative (violin plots, Figure 2). While tendencies regarding better and worse performing alternatives are visible, the uncertainty distributions largely overlap. The uncertainty in the risk attitude parameter was one major factor contributing to the overall uncertainty in utilities, as can be seen when comparing the uncertainty distributions to a situation where this parameter was fixed (Figure SI 6).

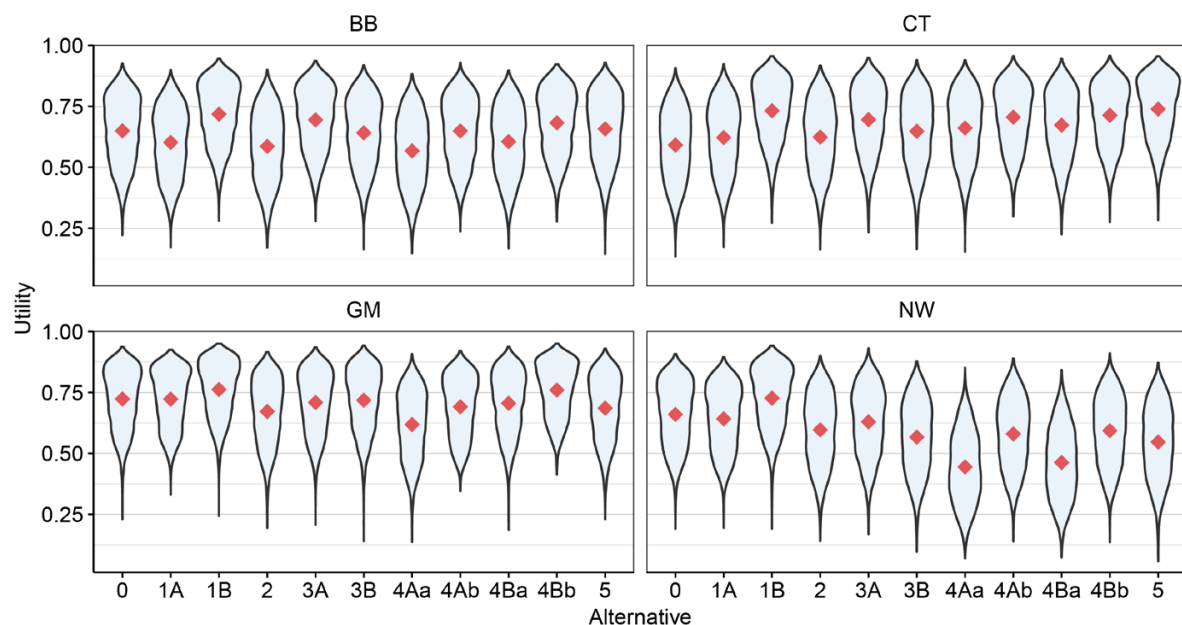


Figure 2: Results of the decision model for the four stakeholders: CT, BB, GM, and NW (Table 1). The violin plots show the mirrored probability density of overall utilities (y-axis) of the alternatives (x-axis; Table 3). Uncertainties in attribute outcomes, lowest-level value functions, weights, aggregation parameter γ , and risk attitude were propagated (Table 4). Diamonds indicate the expected expected utility of the alternatives.

The EEU concept allows us to collapse the large uncertainty in the overall utilities to a single number (diamonds in Figure 2, Table SI 7). If we determine the ranking of alternatives based on the EEU, this allows us to identify promising alternatives (Figure 3). One possible consensus alternative emerged, which is among the best for all four stakeholders, namely alternative 1B. This would keep the system configuration largely unchanged; only WWTP BB would discharge effluent into a larger river instead of the stream NZ.

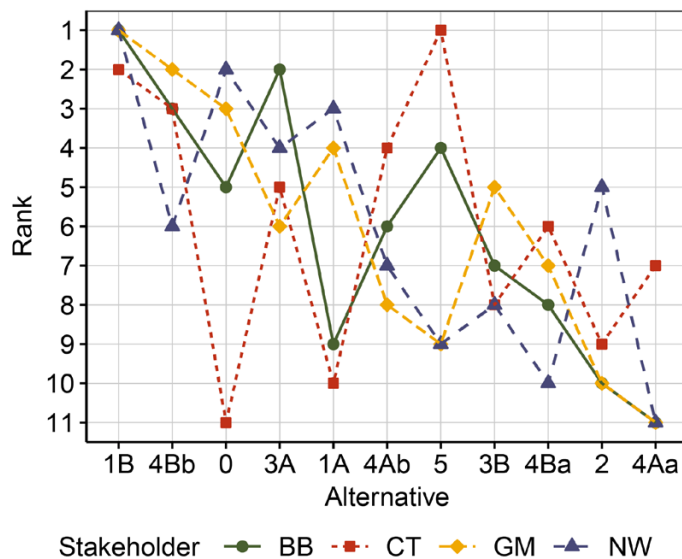


Figure 3: Ranks of alternatives (Table 3) based on their expected utility (Table SI 7), given the uncertain knowledge about predictions (Table SI 4) and preferences (Table 4, Table SI 5) for four stakeholders (Table 1). Alternatives on x-axis are ordered from left to right by median rank across stakeholders.

Such a unique ranking could alternatively be based on other indices with a different conceptual background. We found little difference between a ranking based on EEU and rankings based on the holistic rank acceptability index of SMAA (Tervonen and Figueira 2008) or based on expected ranks (Figure SI 5). However, the acceptability index of SMAA that only considers the best rank can produce a quite different result (Figure SI 5).

4.2. Sensitivity to assumptions about preferences

In a stepwise sensitivity analysis, we tested the effect of other distributions of the preference parameters (simplifying assumptions; Table 4), compared to the uncertain information obtained in preference elicitation interviews. We calculated the ranks of the alternatives based on their EEU (Table SI 7) and determined the similarity of rankings with Kendall's rank correlation coefficient τ . In addition, we determined the sensitivity of the top three ranks (Figure SI 7).

The shape of the lowest-level value functions and the uncertainty distributions of their parameters (θ, a, b) had a low sensitivity in general. The sensitivity was stronger for stakeholders BB and GM, indicated by τ coefficients of 0.71 and 0.85, respectively (Figure 4). When using linear value functions for all lowest-level objectives, only for one stakeholder (BB) one change in the top three alternatives occurred (Figure SI 7). One explanation is that the differences between linear value functions and the elicited value functions were small.

Disregarding the uncertainty in the weights and using precise weights had no influence on the resulting ranking (Kendall's $\tau = 1$ for all stakeholders, Figure 4). Since the expected value of the weights was the same for the uncertain information and the precise weights analysis, the effect of the uncertainty might have averaged out at the level of EEU. In contrast, using

uncertainty distributions based on ranking information of the weights strongly affected the ranking. The rank similarity coefficient in comparison to the uncertain information case was in the range of 0.35–0.78 (Figure 4). For three stakeholders this led to one or two changes in the best three alternatives (Figure SI 7). A possible explanation is that the uncertainty distributions of weights were rather dissimilar between the two settings: For the ranking-based weights the resulting uncertainty in the weights was considerably larger.

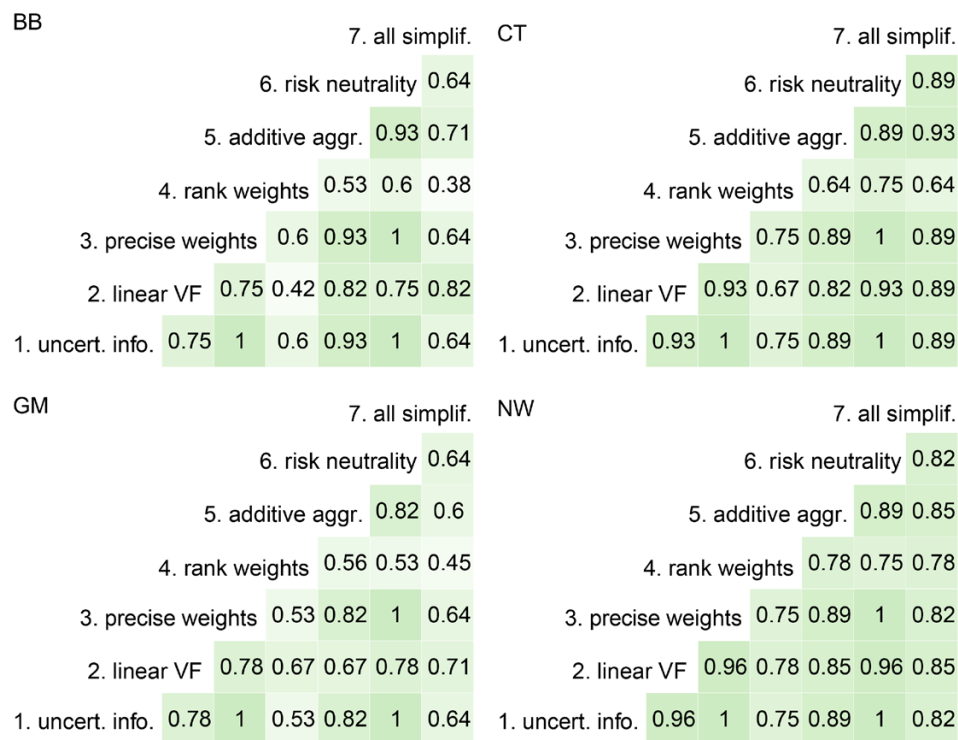


Figure 4: Similarity of alternatives' rankings (Kendall's τ) based on the alternatives' EEU (Table SI 7) for different preference sets (Table 4). Values closer to one and darker shading indicate higher similarity in ranking. We show the results for four stakeholders, CT, BB, GM, NW (Table 1). Uncert. info: uncertain preference information, corresponds to left column of Table 4; linear VF: all lowest-level value function assumed to be linear; precise weights: disregarding uncertainty in weights; rank weights: only ordinal information on weights is used; additive aggr: additive model, i.e. $\gamma = 1$ for all aggregation nodes; all simplif.: all simplifications are used at once, corresponding to right column of Table 4.

The ranking of alternatives was also sensitive to the aggregation model. The rank correlation coefficient was between 0.82–0.89 (Figure 4). However, only for one stakeholder (BB) one of the top three alternatives changed (Figure SI 7). The influence of the aggregation model depends on the assumptions about the values of the parameter γ , which determines the strength of non-additivity. In our case, the expected value of γ was 0.64 for all nodes and stakeholders, which is not extremely non-additive. However, in individual runs the model could be strongly non-additive.

The exact distribution of the risk attitude parameter (r) had negligible effect (Kendall's $\tau \geq 0.96$, Figure 4). The reason is that the evaluations of the alternatives, i.e., the overall values, had a similar spread for each alternative. The risk attitude does not help distinguishing

alternatives if their risks are similar. Thus, while strongly affecting the utilities and the EEU (Figure 2), the risk attitude had little influence on the ranking.

Taking all simplifying assumptions together – linear value functions, precise weights, additive model, risk neutrality – led to rankings with similarity coefficients between 0.64–0.89 (Figure 4) in comparison to the uncertain information case. For three of the stakeholders (BB, CT, GM), there were changes in the top three alternatives (Figure SI 7).

Because we used non-additive value functions, we suspected that the hierarchical arrangement of objectives could influence the results. Therefore, we tested different modifications to the hierarchical structure (Figure SI 8). We assumed the same distribution of the aggregation parameter γ_j for each aggregation node, but sampled independently. Thus, depending on the actual parameter values, the hierarchical structure can have considerable influence on the result of a single run (Figure SI 9). However, this effect averaged out at the level of EEU. Given the uncertain distributions of preferences (Table 4), modifications to the hierarchical structure had negligible effect on the EEU (Figure SI 9, Table SI 8) and thus on the resulting ranking of alternatives.

4.3. Sensitivity to leaving out objectives

We determined the sensitivity of the results to individual objectives by leaving them out one by one. We assumed the “uncertain information” throughout (Table 4). The results were sensitive to the exclusion of some objectives, but less for others. The sensitivity differed considerably between the stakeholders. The EEU and resulting ranks for all model runs is given in Table SI 9.

For four objectives, the rank similarity for having them in the analysis and leaving them out individually was high across all stakeholders (average $\tau \geq 0.90$, Table 5): cost, electricity, land consumption, and amenity of landscape (see Table 2 for a description of objectives). This suggests that these objectives contributed little to differentiate the ranking of alternatives and they were individually less relevant for the decision. These four objectives could individually be removed from the model without affecting the top three alternatives for three of the four stakeholders (Figure SI 10).

On the other hand, excluding flexibility, micropollutant removal, or sludge utilization had a strong influence on the rank correlation (average $\tau \leq 0.68$, Table 5). Thus, these objectives were key factors for determining the ranking of alternatives, given our current knowledge about predictions and preferences. If the micropollutant objective was excluded, the advantage of the larger mergers of WWTP that would include micropollutant removal (alternatives 4b and 5) diminished. For stakeholders for which some of these alternatives were among the top three (CT, BB, GM), none of the alternatives 4Ab, 4Bb, or 5 were among the top three anymore (Figure SI 9).

Table 5: Similarity of alternatives' rankings (Kendall's τ) based on their EEU (Table SI 9) for preference models in which one objective was left out in comparison to the full set of 14 objectives (Table 2). Values closer to one indicate higher similarity in ranking. Results are given for four stakeholders (Table 1) and ordered by increasing average rank similarity.

Objective left out	BB	CT	GM	NW	Average
landscape	0.89	0.93	0.96	1	0.95
electricity	0.96	0.96	0.85	0.96	0.94
cost	0.89	0.85	0.89	1	0.91
land_consumption	0.82	0.89	0.93	1	0.91
jobs	0.96	1	0.64	0.93	0.88
long_term	0.75	0.78	0.96	1	0.87
ecology_nz	0.85	0.64	0.93	0.96	0.85
residents	0.6	0.93	0.85	0.85	0.81
co_determination	0.93	1	0.45	0.82	0.8
district_heating	0.67	1	0.89	0.6	0.79
protected_areas	0.82	0.89	0.67	0.71	0.77
sludge_utilization	0.71	0.78	0.31	0.93	0.68
micropollutants	0.38	0.49	0.56	0.75	0.55
flexibility	0.27	0.78	0.35	0.56	0.49

Alternative 4Bb, the merger of four WWTPs, which was highly ranked for three stakeholders (CT, BB, NW), has two advantages. One is the micropollutant removal compared to alternatives 4Ba, and the second the higher sludge utilization compared to alternative 4Ab. Excluding the sludge utilization objective resulted in the alternative losing its advantage compared to alternative 4Ab and descending to a medium rank while alternative 4Aa advanced.

The exclusion of the flexibility objective (average $\tau = 0.55$, Table 5), led to a different pattern depending on the stakeholder. For stakeholders CT and BB, alternatives that keep options open, but do not change the general situation (0 and 1B) were now ranked lower (Figure SI 10). For stakeholders GM and NW the do-nothing alternative 0 was also ranked lower, but alternative 1B remained among the top three. The main reason why alternative 0 would be ranked well is because it allows postponing decision-making, not due to other characteristics.

This type of analysis is powerful to identify the key objectives needed to differentiate alternatives and explain the performance of alternatives in terms of objectives. This supports focusing discussions on the relevant trade-offs. It also indicates on which objectives an alternative would need to improve to receive a better ranking. For instance, if co-determination could be maintained in alternative 5, this alternative would become considerably more attractive for stakeholders GM and NW.

4.4. Comparison to actual choices

The holistic choice of the stakeholders which alternatives should or should not be further considered differed from the recommendations of the full decision model (Figure 5). Using reduced decision models (see section 3.5) to mimic the choice heuristic of the stakeholders

mitigated some of these discrepancies. In the reduced models, only few objectives were included, based on statements of the stakeholders in personal interviews and discussions.

In the meeting in which the choice was made, the stakeholders agreed to eliminate wastewater from the stream NZ. Therefore, alternative 1A was excluded; the do-nothing alternative 0 had been dismissed before. However, these alternatives were among those with the highest EEU for stakeholders NW and GM in the full decision model (Figure 2). They allow continuing the operations of WWTPs NW and GM without significant changes. Using the reduced decision models for evaluation, these two alternatives performed poorly for all stakeholders because the objective of stream ecology received relatively more weight.

Alternative 1B – the building of a pipe to divert the treated effluent of WWTP BB from stream NZ to the large river RA (Figure 1) – emerged as a top alternative for all interviewed stakeholders in the full decision model (Figure 3). Based on the modeling, this could be a promising consensus alternative. However, it was not considered a viable alternative in the meeting by stakeholders CT and AZ (Figure 5); they argued that this was not a long-term solution for the regional situation, but only an intermediate step. This holistic choice could be explained by the reduced model for stakeholder CT, which only considered cost and the long-term viability. In the project, it was decided that alternative 1B will be further considered as an intermediate step.

The full decision model and the group discussion agreed about excluding the direct merger between NW and BB (alternative 2). This alternative had been the initial suggestion by the authorities in 2014. For stakeholder NW this was still a conceivable option in the holistic choice. A reduced decision model for stakeholder NW, which only included the district heating system of NW and the ecological state of river NZ, matched this opinion.

The merger of three WWTPs (alternatives 4) was controversial. For stakeholder NW, this merger did not perform well, neither in the holistic choice, nor in the reduced decision model, nor in the full model – with exception of alternative 4Bb. It would entail giving up the local treatment plant of NW, with unknown consequences for the district heating system. For stakeholder BB, the limited usage of the energy contained in the sludge was the main argument to refuse the 4A alternatives in the holistic choice. The technically minimal alternative 4Aa was among the worst in the full model for three of four stakeholders (Figure 3). However, alternative 4Aa was among the best according to the holistic choice for stakeholders CT and AZ, given that it could be the alternative with lowest yearly cost. Accordingly, in the reduced model for stakeholder CT – which only included the objectives of cost and of long-term orientation – alternative 4Aa performed well.

None of the stakeholders chose one of the 4b alternatives that foresee micropollutant treatment in their holistic choice. They argued that they would not invest in it now, as long as micropollutant removal for such a WWTP is not compulsory. In contrast, based on the preference

statements in the interviews, the state-of-the-art alternative 4Bb with micropollutant removal was among the best in the full decision model.

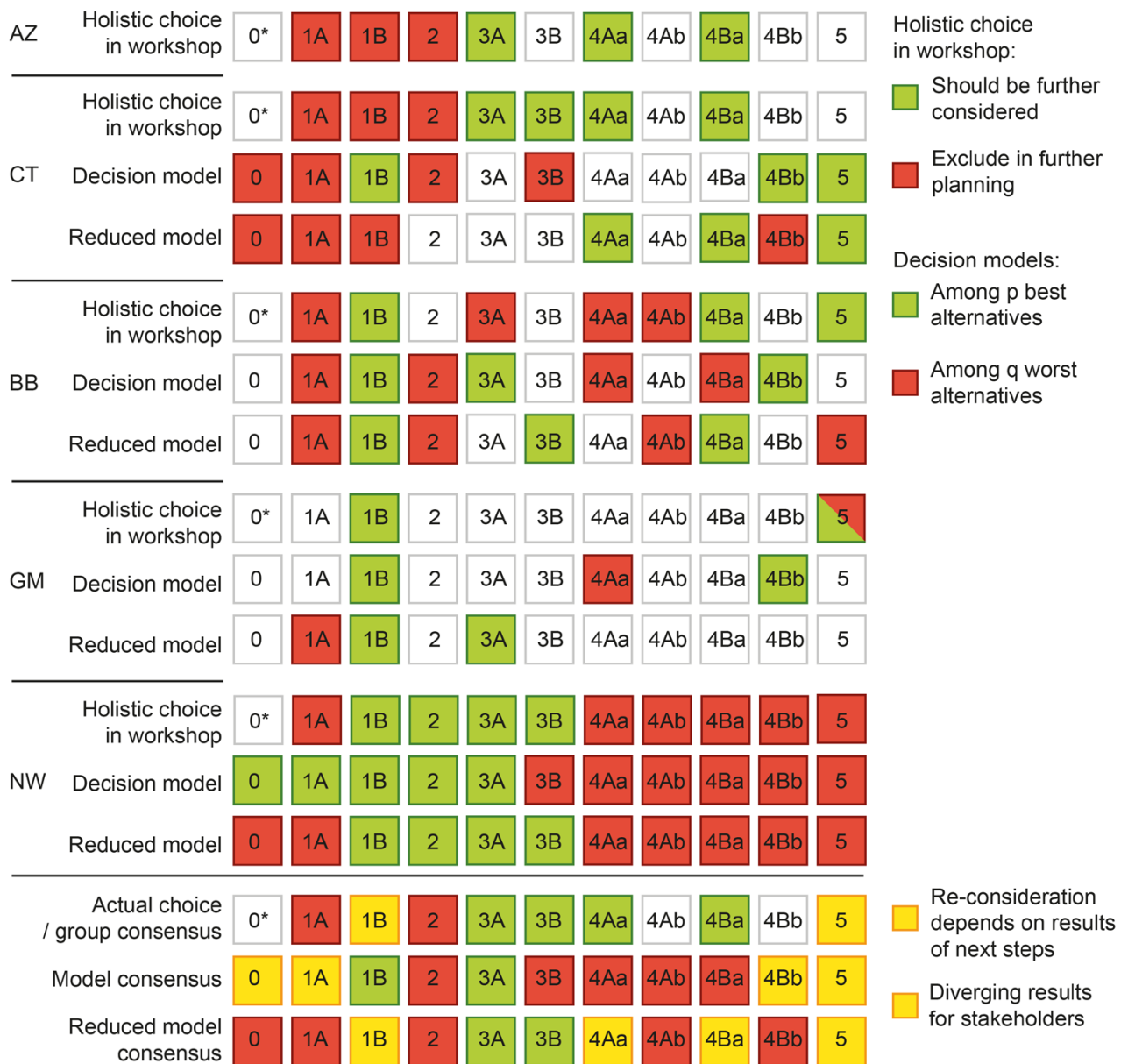


Figure 5: Comparison of the selection of alternatives (Table 3) (i) in a holistic choice in a group meeting, (ii) based on the full decision model, and (iii) based on reduced decision models. The holistic choice was made by stakeholders of the five organizations (CT, BB, NW, GM, and AZ, Table 1). Depending on how many “green” (to consider further) and “red” alternatives (to discard) each stakeholder identified in the meeting, we selected the corresponding number of p best and q worst alternatives based on their ranking results of the full and reduced decision models (Table SI 7 and Table SI 10). For stakeholder AZ only results for the holistic choice were available. In the reduced models, only few objectives were included (see section 3.5). For stakeholder CT: cost, long_term; BB: cost, ecology_nz, sludge_utilization, flexibility; GM: cost, ecology_nz, flexibility; NW: ecology_nz, district_heating. *The do-nothing alternative 0 was not considered in the group discussions; it was included in the modeling for comparison.

Based on the full decision model, the consensus solution across the interviewed stakeholders would be to focus on alternatives 1B, 3A, and 4Bb as they did not perform poorly for anyone. The conclusion from the holistic choice was to first focus on alternatives 3A and 4A and

then compare the better one with 3B or 4B, respectively. However, stakeholder AZ, who participated in the holistic choice, was not included in the modeling because they were not available for an interview in time. Using the reduced decision models, the consensus would be more similar to the holistic choice (Figure 5).

Discrepancies between the holistic choice and the full decision model could also arise because objectives were missing in the model. We compared the included objectives with results from a survey about objectives among the same stakeholders (Haag et al. 2019b). In the survey, thirty-three distinct objectives were stated by representatives of the five organizations. Of the 15 objectives that were perceived as most important for the decision in the survey, only half were part of the full decision model (see Figure SI 11). However, decision models that compare relative performance need to only include objectives which contribute to differentiating alternatives (Marttunen et al. 2019a). Some objectives that were perceived as important in the survey, such as the protection of ground water resources, were likely not relevant, as we would not expect substantial differences between the alternatives.

Two objectives related to the topic of justice might have been neglected in the MCDA. The objective of intergenerational equity was partly covered by the objectives of long-term solution and flexibility to react, but it was not discussed what this objective concretely meant to the stakeholders. The fair distribution of burdens and costs was perceived as one of the most important issues in the survey. It was disregarded in the MCDA, because a fair allocation of costs and benefits should be possible if an overall improved regional solution would be found. However, what such a fair allocation would mean was not discussed with the stakeholders. This might have influenced their holistic choice, prompting them to focus on alternatives with direct benefits for their own organization.

5. Discussion

5.1. Expected expected utility

To obtain a ranking of non-dominated alternatives in decisions with uncertainty, we need to employ a decision criterion. The EU criterion has a strong conceptual foundation (e.g., Keeney and Raiffa 1976) and is useful when predictions are uncertain, yet quantifiable by probabilities, and the preferences are known. However, alternative decision criteria have been proposed, for example, the maximin criterion.

The EEU concept is an extension of EU for uncertain preferences. In practice, preferences are generally uncertain. Accordingly, EEU can be useful in many MCDA applications. With EEU a unique ranking can be found in a way that is consistent with the axioms of utility theory, given specific assumptions (Boutillier 2003). We are not aware of other measures to aggregate uncertain utilities with this strong conceptual underpinning. Again, other decision criteria can be used, for

example those developed around the SMAA family of methods (Tervonen and Figueira 2008). On a more general level, different robustness criteria could be used to reason about the best alternatives for decisions under uncertainties (Roy 2010). Practically, the ranking of alternatives was relatively robust to basing it on the EEU or alternatively on the holistic rank acceptability index or on expected ranks (Figure SI 5). The acceptability index of SMAA often resulted in a different ranking; it does not consider the whole distribution of ranks but only the best rank and is thus more similar to a maximax criterion.

The unique ranking provided by EEU is conditional on the assumptions regarding model structures and prior probability distributions. It is important to investigate the robustness of the ranking regarding these assumptions to assess the uniqueness of the result and identify whether alternatives exist with similar performance than the one ranked best. In this study, we addressed this by sensitivity analysis. While a unique ranking can be relevant, inspecting the underlying distributions of expected utilities or of resulting rankings can also be instructive for decision-making. This information is lost when basing a decision on an aggregate measure such as EEU.

The EEU concept assumes risk neutrality with regard to uncertain expected utilities. A narrow distribution of expected utilities and a wide distribution of expected utilities can yield the same EEU. This might not always be the preference of stakeholders. However, it can be expanded for stakeholders who are not risk neutral in that regard. Houlding and Coolen (2011) have termed such a preference “trial averse” or “trial seeking” and have provided an interesting sketch of this concept. A practical application will require further developments, such as ways to elicit the degree of trial aversion from stakeholders.

In this study, we used EEU in analyzing the sensitivity to using different distributions for preference parameters and the sensitivity to leaving out objectives (sections 4.2 and 4.3). However, the potential is much greater. The EEU is a convenient variable to use in global, regional, and local sensitivity analysis of decision models. It can also be a target variable for a type of sensitivity analysis called *value of information analysis* (Borgonovo et al. 2016) that can be particularly insightful for decision models. Chajewska et al. (2000) provide an application.

Given a decision model with specified uncertainty distribution of predictions and preferences, the practical application of EEU is straightforward. It is efficiently calculated by jointly sampling from the space of uncertainties in predictions and preferences and Monte Carlo simulation. The difficulty rather lies in specifying the decision model and inferring its parameters as well as their uncertainties.

Estimating the uncertainty in the elicited statements is not necessarily more effort during the elicitation. Ideally, the uncertainties are estimated in the process of parameter inference (Haag et al. 2019a). However, often information is insufficient for such an approach. Instead, assumptions about the distributions can be made based on the preference statements themselves (Scholten et al. 2015). For instance, stakeholders may be asked to give interval statements (and a

best guess) instead of precise points. This information might actually be easier to provide than precise points. From such information an uncertainty distribution can be constructed.

5.2. Sensitivity of results

We tested the sensitivity of the obtained results – the ranking of alternatives based on EEU – to (1) different distributions of preference parameters, (2) the hierarchical structuring of objectives, and (3) leaving out objectives.

(1) The sensitivity of the results is influenced by the sensitivity of the model to the parameters, but also by the magnitude of changes in the distributional assumptions of the parameters. The results are, therefore, case-dependent. Changing all lowest-level value functions to linear value functions had minor impact. Similar observations have been made in other cases (Lahdelma and Salminen 2012; Zheng et al. 2016). In our case, few of the elicited value functions were strongly non-linear and for not-elicited value functions the expected value of the curvature parameter corresponded to linearity. This might explain the insensitivity. Disregarding uncertainty in the weights and using precise weights had no effect on the ranking. The effect due to uncertainty might have averaged out. In contrast, using only ranking information on weights instead of weights obtained by the Swing procedure had a large effect on the results. It seems that the ordinal information was not sufficient to specify the distributions narrowly enough. Using the commonly used weighted arithmetic mean for aggregation – instead of power means with an additional non-additivity parameter – had a moderate influence. So far, only few promising methods to practically estimate parameters of non-additive models exist (see Haag et al. 2019a). Finally, assuming an uncertain risk attitude, from risk aversion to risk neutrality, had almost no influence on the results compared to the assumption of risk neutrality. This can be expected when the risk of alternatives is similar. In these cases, there might not be added value in eliciting the risk attitude in detail (also see Scholten et al. 2015; Zheng et al. 2016).

(2) Generally, the exact hierarchical structure of the same objectives (e.g., flat or organized with several sub-levels, Figure SI 8), will affect the results when the overall aggregation function is not decomposable (see section 2.1). When the overall aggregation function is decomposable, the hierarchical structuring does not matter, given a proper normalization of weights. This is the case for preference models using the weighted power mean as aggregation function with the parameter γ constant across all aggregation nodes. Thus, this is also true for the special case of the additive model, where everywhere $\gamma=1$. In our application, we assumed the same uncertainty distribution of the γ parameter for each node; this may explain why effects due to the hierarchical structure averaged out at the level of EEU.

(3) The inclusion or exclusion of objectives may change results severely, as our analysis showed. In our analysis, we identified objectives that could be excluded without changing the conclusions from the model. Thus, given our current knowledge about these objectives, it would

be possible to simplify the decision model. One candidate for exclusion was the objective of low annual cost, even though it was among those with the highest *perceived importance* for the stakeholders in the discussions. Of course, an objective's relevance for a decision can change, if the knowledge about impacts on the objective or the stakeholders' preferences towards it change.

5.3. Conclusions for the case study

In the case study region in Switzerland one WWTP was legally obligated to minimize its discharge of micropollutants to a small stream. While wastewater disposal in the region is generally functional, the situation might be optimized in the long-term with a regional approach. The initial suggestion by the authorities had been alternative 2: the merger of WWTP BB with WWTP NW. During the decision process and in our MCDA, it turned out that it was unlikely that this was the best alternative.

Alternative 1B emerged as potential consensus solution from the decision modeling (Figure 3). The treated effluent from WWTP BB would be discharged to a larger river instead of the stream NW, while keeping the regional configuration and organization of WWTPs as it currently is. This may seem simplistic, but would minimize impacts of the treated wastewater to the stream. The pipe that would be built could later be utilized as part of a rerouting to a larger treatment plant (alternatives 3, 4, and 5). In light of maintaining flexibility in wastewater systems (Spiller et al. 2015), this alternative could be promising as it reduces path dependencies and does not preclude later mergers. However, the actual flexibility would depend on the investments made at the WWTPs in the region, e.g., for rehabilitation, as long periods of depreciation limit opportunities for changes. Stakeholders CT and AZ considered the alternative 1B unsuitable as a long-term option for the year 2040 because it is not economically optimal given the applied cost calculation. The reduced decision model for stakeholder CT which only considered these two objectives came to the same conclusion (Figure 5).

For the treatment of wastewater, large mergers of WWTPs are attractive due to economies of scale (Abbott and Cohen 2009). Treatment performance for wastewater and sludge generally increases, while costs and resource needs per population equivalent might be lower, for instance, the specific energy consumption for treatment. However, if the entire wastewater system is considered, this becomes debatable. In the case study, the merger of four WWTPs (alternative 5) had higher yearly cost than mergers between three WWTPs. The results of our full decision model indicate that a merger of treatment plants is also not an a priori solution when a broader range of objectives is considered. In the decision model, the benefits of mergers could be outweighed by factors such as increased energy consumption for wastewater transport or the loss of flexibility. For the MCDA, some of these trade-offs were specified by the stakeholders. The merger 4Bb, which foresees increased biogas production and micropollutant removal, performed well, but

despite lower cost mergers without both advancements (4Aa, 4Ab, 4Ba) were usually ranked considerably lower (Figure 3).

5.4. Comparison between modeling and actual choice

There are several, empirically mingled, explanations for the discrepancies between the results of our MCDA and the holistic choices by the stakeholders in a meeting. The decision model might have been insufficient in its structure or the parameter distributions might not have represented the actual situation with respect to the preferences or the predictions. Likewise, when making the holistic choice, the stakeholders might have made different consideration than in the interviews where the problem was decomposed into small steps. For instance, they might have focused on few or different objectives, they might have made different trade-offs, or might have had different expectations about outcomes in mind.

Both the decision model as well as the heuristics employed by the stakeholders can suffer from the missing variable bias (Payne et al. 1999; Montibeller and von Winterfeldt 2015), i.e., an incomplete problem representation. As has been shown in section 4.3, excluding objectives can change the results. Conversely, the results might change if we had included further or other objectives, for example, those identified in a survey for the same case, especially concerning the topic of justice (section 4.4, Haag et al. 2019b). The holistic choices of the stakeholders could largely be mirrored when using reduced decision models in which most objectives were excluded. This could indicate that the stakeholders made their choice based on narrow considerations and disregarded many objectives which they actually considered relevant in the preference elicitation interviews.

The preference elicitation process itself can suffer from biases (Montibeller and von Winterfeldt 2015; Marttunen et al. 2018) and may also be a source of errors. The topic of micropollutants was the trigger for the decision problem. Due to this general importance, participants might have allocated too much weight to the respective objective when stating their preferences. In the holistic choice, alternatives that would treat micropollutants without legal obligation were dismissed. On the other hand, the cost objective might have received too little weight, as can regularly be observed in environmental decisions with many objectives (Marttunen et al. 2018).

5.5. Practical recommendations

Undertaking a structured decision-making process is certainly valuable in public planning, as the example of alternative 2 demonstrated, which was initially suggested in our case study, but turned out to be dispensable. The alignment of decision methods to the problem characteristics and requirements is decisive. The methods can range from mostly discursive (e.g., Dominguez et al. 2011; Franco 2013), to the use of quantitative prediction models and discussion among

stakeholders (e.g., Starkl et al. 2009; Gregory et al. 2012), to the specification of quantitative prediction and preference models as in this study (and, e.g., Zheng et al. 2016; Scholten et al. 2017).

We quantified the predictions of outcomes and the preferences of stakeholders and used the EEU concept to integrate their uncertainties. This required effort, but if the stakes in a decision are high enough such an approach can have advantages: It ensures the treatment of decision elements and the uncertainties surrounding them in the well-established framework of rational decision-making. The model building process demands to make all data, assumptions, and preferences explicit. This makes the decision process transparent. Conflicts can be identified and discussed at the appropriate level. Additionally, some insights only emerged with the modeling. Examples of this are: which objectives were actually decisive for the ranking of alternatives and which might be dismissed (section 4.3), or that the holistic choice might have neglected objectives that actually might be relevant for the stakeholders (section 4.4).

The importance of the structuring phase of a decision model is well established (e.g., Keeney 1992; Gregory et al. 2012). This phase includes the elicitation and structuring of objectives. This requires an opening up phase where a comprehensive overview of concerns is generated (Haag et al. 2019b), but also a consolidation of these concerns and a selection of objectives to include in the model (Marttunen et al. 2019a). In our application, a sensitivity analysis (section 4.3) indicated that we could perhaps have excluded some objectives at an earlier stage, which would have simplified the decision problem. The simplification process can be supported by methods such as an analysis of relevance (Marttunen et al. 2019a; Marttunen et al. 2019b). In practice, both the opening up and closing down phases can often be insufficient; therefore, we recommend applying structured procedures (see references above).

For preference elicitation, the choice and framing of the attributes is crucial (e.g., Keeney and Gregory 2005). This was affirmed in our interviews, where we received feedback from individual stakeholders that they had difficulty to specify their preferences for some attributes. For instance, for the energy objectives stakeholders were less interested in gross electricity demand, but cared more about the total net energy budget including sludge treatment, transport, etc., which we could not estimate with the given resources. As a result of these discussions, energy will be investigated in more depth in the next project phase of the case study.

The effort for elicitation depends on the number of parameters and the detail we want to achieve in our model. However, estimating the uncertainty in the elicited statements is not necessarily more effort during the elicitation itself, as discussed in section 5.1. The sensitivity of the results to different elements of the model is case-dependent (see section 5.2). Nevertheless, we can derive some insights for elicitation from our study.

We only elicited one point of the lowest-level value functions to receive an approximate shape for attributes with a high weight. It would be possible to elicit the shape in more detail. However, as our sensitivity analysis showed, and in line with earlier studies (see section 5.2), the

exact shape of the single-attribute value functions rarely changed relevant parts of the resulting ranking of the alternatives. Often, it might be less relevant to invest large efforts into this elicitation step. Similarly, we conclude that the added value in eliciting the risk attitude in detail might often not be large (also see Scholten et al. 2015; Zheng et al. 2016).

The opposite conclusion must be drawn for weight elicitation, which is often a major effort in real-world projects. We had hoped that ranking information on the weights could be sufficient, which would be fast and easy to elicit. However, the sensitivity analysis indicated that using only ordinal information on weights instead of precise or uncertain weights obtained by the Swing procedure had a large effect on the results. We are not aware of other research that has focused on the effects of different degrees of knowledge about weights and suggest this should be investigated further. The same applies to the aggregation model where further research is needed to corroborate in which cases the commonly used additive model is indeed a sufficient approximation of the stakeholders' preferences or if non-additive models are required instead.

To make best use of the decision model results, a more intensive discussion with the stakeholders would have been beneficial in the sense of facilitated modeling (Franco and Rouwette 2011). This might benefit from further research on the use of decision models in practice settings and on embedding decision analysis procedures into the actual political process of decision-making (French and Argyris 2018).

6. Conclusions

In this study, we demonstrated how stochastic decision models can be used in strategic regional planning of wastewater infrastructure. We built preference models based on multi-attribute utility theory. Instead of using the additive model, we used the power mean as aggregation function. This allowed us to represent the additive model as a special case, but also varying degrees of non-additivity. In practice, the alignment of model complexity to the purpose of modeling is decisive, but relatively complex preference models to better represent stakeholder preferences can be feasible. While our study gives some guidance for practice (section 5.5), more work is required to systematically investigate in which cases which simplifications to the modeling are appropriate.

Dealing with uncertainty is a challenge, but instead of ignoring it, we can embrace the inevitable. Once probability distributions have been specified for the inputs and parameters of a decision model, their forward propagation to the results is straightforward. As we introduced, the EEU concept can be used for rational decision-making when predictions and preferences are uncertain. Implementing EEU requires the quantification of the uncertainty in utilities (or utility parameters), but does not lead to a relevant increase in computational complexity. Using this concept allowed us to draw concrete and plausible conclusions on the alternatives. The top ranks of the alternatives were shown to be relatively robust against tested changes in distributional

assumptions of model parameters and risk attitude, except when only using ranking information about weights. The main conclusions did not differ when using ranking based indices instead of the EEU. In the future, it would be interesting to compare the EEU concept in more depth to alternative concepts for robust decision-making under uncertainty.

Most objectives included in the model were relevant for determining the ranking of alternatives, but some could also be omitted without great impacts to the ranking. Conversely, the ranking might change if we had included further or other objectives. Analyzing the influence of the selection of objectives and the selection of attributes in comparison to other sources of uncertainty in a further study could prove insightful for focusing efforts in model building and determining an appropriate level of model complexity.

The comparison of model results to choices in the actual decision process revealed some discrepancies. This can be the starting point to challenge the – perhaps simplistic – argumentation or heuristics by stakeholders as well as to challenge the model assumptions. In such an iterative fashion, decision models can bring rigor and clear thinking into difficult decision processes that are encountered in regional planning.

Acknowledgments

We are most grateful to all involved actors and stakeholders for their support and their participation in interviews, surveys, and meetings. We thank Philipp Beutler and Ulrike Feldmann for discussion, and two anonymous reviewers for their excellent suggestions. This work was supported with Eawag's Discretionary Funds, granted by the Eawag directorate.

Supplementary information

[Link to be updated.]

Bibliography

- Abbott, M., and Cohen, B. 2009. 'Productivity and efficiency in the water industry', *Utilities Policy*, 17(3): 233-244. <https://doi.org/10.1016/j.jup.2009.05.001>
- Allen, C.R., Fontaine, J.J., Pope, K.L., and Garmestani, A.S. 2011. 'Adaptive management for a turbulent future', *Journal of Environmental Management*, 92(5): 1339-1345. <https://doi.org/10.1016/j.jenvman.2010.11.019>
- Borgonovo, E., Hazen, G.B., and Plischke, E. 2016. 'A Common Rationale for Global Sensitivity Measures and Their Estimation', *Risk Analysis*, 36(10): 1871-1895. <https://doi.org/10.1111/risa.12555>
- Boutillier, C. 2003. "On the foundations of expected utility." In *Proceedings of the 18th international joint conference on Artificial intelligence*, 285-290. Acapulco, Mexico: Morgan Kaufmann Publishers Inc.
- Cegan, J.C., Filion, A.M., Keisler, J.M., and Linkov, I. 2017. 'Trends and applications of multi-criteria decision analysis in environmental sciences: literature review', *Environment Systems and Decisions*, 37(2): 123-133. <https://doi.org/10.1007/s10669-017-9642-9>
- Chajewska, U., Koller, D., and Parr, R. 2000. "Making Rational Decisions Using Adaptive Utility Elicitation." In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, 363-369. AAAI Press.
- Cyert, R.M., and DeGroot, M.H. 1975. 'Adaptive Utility.' in Richard H. Day and Theodore Groves (eds.), *Adaptive Economic Models* (Academic Press).

- Dominguez, D., Truffer, B., and Gujer, W. 2011. 'Tackling uncertainties in infrastructure sectors through strategic planning: the contribution of discursive approaches in the urban water sector', *Water Policy*, 13(3): 299-316. <https://doi.org/10.2166/wp.2010.109>
- Durbach, I.N., and Stewart, T.J. 2009. 'Using expected values to simplify decision making under uncertainty', *Omega*, 37(2): 312-330. <https://doi.org/10.1016/j.omega.2007.02.001>
- Durbach, I.N., and Stewart, T.J. 2012. 'Modeling uncertainty in multi-criteria decision analysis', *European Journal of Operational Research*, 223(1): 1-14. <https://doi.org/10.1016/j.ejor.2012.04.038>
- Dyer, J.S., and Sarin, R.K. 1979. 'Measurable Multiattribute Value Functions', *Operations Research*, 27(4): 810-822. <https://doi.org/10.1287/opre.27.4.810>
- Dyer, J.S., and Sarin, R.K. 1982. 'Relative Risk-Aversion', *Management Science*, 28(8): 875-886. <https://doi.org/10.1287/mnsc.28.8.875>
- Eggen, R.I.L., Hollender, J., Joss, A., Schäfer, M., and Stamm, C. 2014. 'Reducing the Discharge of Micropollutants in the Aquatic Environment: The Benefits of Upgrading Wastewater Treatment Plants', *Environmental Science & Technology*, 48(14): 7683-7689. 10.1021/es500907n
- Eisenführ, F., Weber, M., and Langer, T. 2010. *Rational decision making* (Springer: Berlin; London).
- Franco, L.A. 2013. 'Rethinking soft or interventions: Models as boundary objects', *European Journal of Operational Research*, 231(3): 720-733. <https://doi.org/10.1016/j.ejor.2013.06.033>
- Franco, L.A., and Rouwette, E.A.J.A. 2011. 'Decision development in facilitated modelling workshops', *European Journal of Operational Research*, 212(1): 164-178. <https://doi.org/10.1016/j.ejor.2011.01.039>
- French, S., and Argyris, N. 2018. 'Decision Analysis and Political Processes', 15(4): 208-222. <https://doi.org/10.1287/deca.2018.0374>
- Grabisch, M., Marichal, J.-L., Mesiar, R., and Pap, E. 2009. *Aggregation functions* (Cambridge University Press: Cambridge).
- Greco, S., Mousseau, V., and Slowinski, R. 2008. 'Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions', *European Journal of Operational Research*, 191(2): 416-435. <https://doi.org/10.1016/j.ejor.2007.08.013>
- Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T., and Ohlson, D. 2012. *Structured decision making: a practical guide to environmental management choices* (Wiley-Blackwell: Chichester, West Sussex; Hoboken, N.J.).
- Haag, F., Lienert, J., Schuwirth, N., and 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., Zürcher, S., and Lienert, J. 2019b. 'Enhancing the elicitation of diverse decision objectives for public planning', *European Journal of Operational Research*, 279(3): 912-928. <https://doi.org/10.1016/j.ejor.2019.06.002>
- Hokkanen, J., Lahdelma, R., and Salminen, P. 2000. 'Multicriteria decision support in a technology competition for cleaning polluted soil in Helsinki', *Journal of Environmental Management*, 60(4): 339-348. <https://doi.org/10.1006/jema.2000.0389>
- Houlding, B., and Coolen, F.P.A. 2011. 'Adaptive utility and trial aversion', *Journal of Statistical Planning and Inference*, 141(2): 734-747. <https://doi.org/10.1016/j.jspi.2010.07.023>
- Keeney, R.L. 1992. *Value-focused thinking: a path to creative decisionmaking* (Harvard University Press: Cambridge, Mass.).
- Keeney, R.L., and Gregory, R.S. 2005. 'Selecting attributes to measure the achievement of objectives', *Operations Research*, 53(1): 1-11. <https://doi.org/10.1287/opre.1040.0158>
- Keeney, R.L., and Raiffa, H. 1976. *Decision with multiple objectives* (Wiley: New York).
- Kelly, R.A., Jakeman, A.J., Barreteau, O., Borsuk, M.E., ElSawah, S., Hamilton, S.H., Henriksen, H.J., Kuikka, S., Maier, H.R., Rizzoli, A.E., van Delden, H., and Voinov, A.A. 2013. 'Selecting among five common modelling approaches for integrated environmental assessment and management', *Environmental Modelling & Software*, 47(0): 159-181. <http://dx.doi.org/10.1016/j.envsoft.2013.05.005>
- Kendall, M.G. 1938. 'A new measure of rank correlation', *Biometrika*, 30(1-2): 81-93. <https://doi.org/10.2307/2332226>
- Lahdelma, R., Hokkanen, J., and Salminen, P. 1998. 'SMAA - Stochastic multiobjective acceptability analysis', *European Journal of Operational Research*, 106(1): 137-143. [https://doi.org/10.1016/S0377-2217\(97\)00163-X](https://doi.org/10.1016/S0377-2217(97)00163-X)
- Lahdelma, R., and Salminen, P. 2001. 'SMAA-2: Stochastic multicriteria acceptability analysis for group decision making', *Operations Research*, 49(3): 444-454. <https://doi.org/10.1287/opre.49.3.444.11220>
- Lahdelma, R., and Salminen, P. 2012. 'The shape of the utility or value function in stochastic multicriteria acceptability analysis', *OR Spectrum*, 34(4): 785-802. <https://doi.org/10.1007/s00291-011-0244-5>
- Langhans, S.D., and Lienert, J. 2016. 'Four Common Simplifications of Multi-Criteria Decision Analysis do not hold for River Rehabilitation', *PLoS ONE*, 11(3): e0150695. <https://doi.org/10.1371/journal.pone.0150695>

- Lienert, J., Monstadt, J., and Truffer, B. 2006. 'Future scenarios for a sustainable water sector: A case study from Switzerland', *Environmental Science & Technology*, 40(2): 436-442.
<https://doi.org/10.1021/es0514139>
- Logar, I., Brouwer, R., Maurer, M., and Ort, C. 2014. 'Cost-benefit analysis of the Swiss National Policy on reducing micropollutants in treated wastewater', *Environmental science & technology*, 48(21): 12500-12508. <https://doi.org/10.1021/es502338j>
- Markard, J. 2011. 'Transformation of Infrastructures: Sector Characteristics and Implications for Fundamental Change', *Journal of Infrastructure Systems*, 17(3): 107-117.
[https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000056](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000056)
- Marttunen, M., Belton, V., and Lienert, J. 2018. 'Are objectives hierarchy related biases observed in practice? A meta-analysis of environmental and energy applications of Multi-Criteria Decision Analysis', *European Journal of Operational Research*, 265(1): 178-194.
<https://doi.org/10.1016/j.ejor.2017.02.038>
- Marttunen, M., Haag, F., Belton, V., Mustajoki, J., and Lienert, J. 2019a. 'Methods to inform the development of concise objectives hierarchies in multi-criteria decision analysis', *European Journal of Operational Research*, 277(2): 604-620. <https://doi.org/10.1016/j.ejor.2019.02.039>
- Marttunen, M., Weber, C., Åberg, U., and Lienert, J. 2019b. 'Identifying relevant objectives in environmental management decisions: An application to a national monitoring program for river restoration', *Ecological Indicators*, 101: 851-866. <https://doi.org/10.1016/j.ecolind.2018.11.042>
- Montibeller, G., and von Winterfeldt, D. 2015. 'Cognitive and motivational biases in decision and risk analysis', *Risk Analysis*, 35(7): 1230-1251. <https://doi.org/10.1111/risa.12360>
- Payne, J.W., Bettman, J.R., and Schkade, D.A. 1999. 'Measuring constructed preferences: Towards a building code', *Journal of Risk and Uncertainty*, 19(1-3): 243-270. <https://doi.org/10.1023/A:1007843931054>
- R Core Team. 2018. *R: A Language and Environment for Statistical Computing, Version 3.3.1* (R Foundation for Statistical Computing: Vienna, Austria).
- Reichert, P., and Borsuk, M. 2005. 'Does high forecast uncertainty preclude effective decision support?', *Environmental Modelling & Software*, 20(8): 991-1001. <https://doi.org/10.1016/j.envsoft.2004.10.005>
- Reichert, P., Borsuk, M., Hostmann, M., Schweizer, S., Sporri, C., Tockner, K., and Truffer, B. 2007. 'Concepts of decision support for river rehabilitation', *Environmental Modelling & Software*, 22(2): 188-201.
<https://doi.org/10.1016/j.envsoft.2005.07.017>
- Reichert, P., Langhans, S.D., Lienert, J., and Schuwirth, N. 2015. 'The conceptual foundation of environmental decision support', *Journal of Environmental Management*, 154(0): 316-332.
<https://doi.org/10.1016/j.jenvman.2015.01.053>
- Reichert, P., Niederberger, K., Rey, P., Helg, U., and Haertel-Borer, S. subm. 'The Need for Unconventional Value Aggregation Techniques - Experiences from Eliciting Stakeholder Preferences in Environmental Management'.
- Reichert, P., Schuwirth, N., and Langhans, S. 2013. 'Constructing, evaluating and visualizing value and utility functions for decision support', *Environmental Modelling & Software*, 46: 283-291.
<https://doi.org/10.1016/j.envsoft.2013.01.017>
- Rowley, H.V., Peters, G.M., Lundie, S., and Moore, S.J. 2012. 'Aggregating sustainability indicators: Beyond the weighted sum', *Journal of Environmental Management*, 111(0): 24-33.
<https://doi.org/10.1016/j.jenvman.2012.05.004>
- Roy, B. 2010. 'Robustness in operational research and decision aiding: A multi-faceted issue', *European Journal of Operational Research*, 200(3): 629-638. <https://doi.org/10.1016/j.ejor.2008.12.036>
- Scholten, L., Maurer, M., and Lienert, J. 2017. 'Comparing multi-criteria decision analysis and integrated assessment to support long-term water supply planning', *PLoS ONE*, 12(5): e0176663.
<https://doi.org/10.1371/journal.pone.0176663>
- Scholten, L., Schuwirth, N., Reichert, P., and Lienert, J. 2015. 'Tackling uncertainty in multi-criteria decision analysis - An application to water supply infrastructure planning', *European Journal of Operational Research*, 242(1): 243-260. <https://doi.org/10.1016/j.ejor.2014.09.044>
- Slovic, P. 1995. 'The Construction of Preference', *American Psychologist*, 50(5): 364-371.
<https://doi.org/10.1037//0003-066x.50.5.364>
- Spiller, M., Vreeburg, J.H.G., Leusbrock, I., and Zeeman, G. 2015. 'Flexible design in water and wastewater engineering – Definitions, literature and decision guide', *Journal of Environmental Management*, 149: 271-281. <https://doi.org/10.1016/j.jenvman.2014.09.031>
- Starkl, M., Brunner, N., Flogl, W., and Wimmer, J. 2009. 'Design of an institutional decision-making process: The case of urban water management', *Journal of Environmental Management*, 90(2): 1030-1042.
<https://doi.org/10.1016/j.jenvman.2008.03.012>
- Tervonen, T., and Figueira, J.R. 2008. 'A survey on stochastic multicriteria acceptability analysis methods', *Journal of Multi-Criteria Decision Analysis*, 15(1-2): 1-14. <https://doi.org/10.1002/mcda.407>

- Tervonen, T., van Valkenhoef, G., Baştürk, N., and Postmus, D. 2013. 'Hit-And-Run enables efficient weight generation for simulation-based multiple criteria decision analysis', *European Journal of Operational Research*, 224(3): 552-559. <https://doi.org/10.1016/j.ejor.2012.08.026>
- Truffer, B., Stormer, E., Maurer, M., and Ruef, A. 2010. 'Local Strategic Planning Processes and Sustainability Transitions in Infrastructure Sectors', *Environmental Policy and Governance*, 20(4): 258-269. <https://doi.org/10.1002/eet.550>
- Uusitalo, L., Lehtikoinen, A., Helle, I., and Myrberg, K. 2015. 'An overview of methods to evaluate uncertainty of deterministic models in decision support', *Environmental Modelling & Software*, 63: 24-31. <https://doi.org/10.1016/j.envsoft.2014.09.017>
- Vacik, H., Kurttila, M., Hujala, T., Khadka, C., Haara, A., Pykäläinen, J., Honkakoski, P., Wolfslehner, B., and Tikkanen, J. 2014. 'Evaluating collaborative planning methods supporting programme-based planning in natural resource management', *Journal of Environmental Management*, 144: 304-315. <https://doi.org/10.1016/j.jenvman.2014.05.029>
- Zheng, J., Egger, C., and Lienert, J. 2016. 'A scenario-based MCDA framework for wastewater infrastructure planning under uncertainty', *Journal of Environmental Management*, 183, Part 3(Pt 3): 895-908. <https://doi.org/10.1016/j.jenvman.2016.09.027>