Munich Personal RePEc Archive

A natural language generation approach to support understanding and traceability of multi-dimensional preferential sensitivity analysis in multi-criteria decision making

Wulf, David and Bertsch, Valentin

Karlsruhe Institute of Technology, Economic and Social Research Institute, Trinity College Dublin

27 October 2016

Online at https://mpra.ub.uni-muenchen.de/75025/
MPRA Paper No. 75025, posted 15 Nov 2016 15:10 UTC

A natural language generation approach to support understanding and traceability of multi-dimensional preferential sensitivity analysis in multi-criteria decision making

- **David Wulf**
  Chair of Energy Economics, Karlsruhe Institute of Technology (KIT), Hertzstr. 16, 76187 Karlsruhe, Germany

- **Valentin Bertsch** (corresponding author)
  Economic and Social Research Institute, Whitaker Square, Sir John Rogerson's Quay, Dublin 2
  valentin.bertsch@esri.ie
  Phone: +353 1 863 2088
  Department of Economics, Trinity College Dublin

Key words

- Decision support systems; Multiple criteria analysis; Preferential uncertainty modelling; Natural language generation; Multi-dimensional preferential sensitivity analysis

A natural language generation approach to support understanding and traceability of multi-dimensional preferential sensitivity analysis in multi-criteria decision making

Abstract

Multi-Criteria Decision Analysis (MCDA) enables decision makers (DM) and decision analysts (DA) to analyse and understand decision situations in a structured and formalised way. With the increasing complexity of decision support systems (DSSs), it becomes challenging for both expert and novice users to understand and interpret the model results. Natural language generation (NLG) techniques are used in various DSSs to cope with this challenge as they reduce the cognitive effort to achieve understanding of decision situations. However, NLG techniques in MCDA have so far mainly been developed for deterministic decision situations or one-dimensional sensitivity analyses. In this paper, a concept for the generation of textual explanations for a multi-dimensional preferential sensitivity analysis in MCDA is developed. The key contribution is a NLG approach that provides detailed explanations of the implications of preferential uncertainties in Multi-Attribute Value Theory (MAVT). It generates a report that assesses the influences of simultaneous or separate variations of inter-criteria and intra-criteria preferential parameters determined within the decision analysis. We explore the added value of the natural language report in an online survey. Our results show that the NLG approach is particularly beneficial for difficult interpretational tasks.

1 Introduction

With the aim of enabling transparent and systematic support in complex decision situations, Multi-Criteria Decision Analysis (MCDA) represents a formalised framework for the analysis of different decision alternatives (Stewart, 1992; Geldermann *et al.*, 2009). While such decision support approaches are aimed at providing guidance to decision makers (DMs), their increasing mathematical complexity often hinders a straightforward understanding and traceability on the part of the DMs. Consequently, a lot of cognitive effort is required in order to analyse, interpret and derive adequate implications from the obtained model results which is particularly challenging for novice users (Spiegelhalter and Knill-Jones, 1984; Henrion and Druzdzel, 1991; Gregor and Benbasat, 1999). DMs consider such models as a 'black box', so they mistrust or even reject them (Brans and Mareschal, 1994; Bell *et al.*, 2003), which leads to a gap between available information on the one hand and processible information on the other hand.

To compensate for this, further explanations of decision analysis results promote understanding of the decision situation and thus help to increase trust and acceptability of the system (Greer *et al.*, 1994; Greef and Neerincx, 1995; Dhaliwal and Benbasat, 1996; Gregor and Benbasat, 1999; Parikh *et al.*, 2001; Geldermann, 2010). The use of natural language generation (NLG) techniques to generate such explanations automatically based on the model results has been proposed, for instance, by Papamichail and French (2003), Geldermann *et al.* (2009) or Clark *et al.* (2010). However, existing DSSs with explanatory functions focus mainly on the communication of their largely deterministic results. For instance, Papamichail and French (2000, 2003) developed an approach allowing for the generation of two kinds of reports. Their *comparative report* analyses and compares the performance of two alternatives in relation to each other. Their *sensitivity report* assesses the influences of varying the relative importance of a specified criterion on the alternatives' overall performance. However, their sensitivity report is limited to generate explanations for a one-dimensional sensitivity analysis only, i.e. when one weight parameter is varied at a time.

Aimed at analysing the impact of simultaneous variations of multiple preference parameters on the alternatives' overall performance, many approaches for multi-dimensional sensitivity analysis have been proposed in the MCDA literature (see e.g., French and Rios-Insua, 1991; Butler *et al.*, 1997; Lahdelma *et al.*, 1998; Bertsch *et al.*, 2007; Scholten *et al.* 2015; Bertsch and Fichtner, 2016). Many of these, including Bertsch and Fichtner (2016), are based on Multi-Attribute Value/Utility Theory (MAVT/MAUT, cf. Keeney and Raiffa, 1976). However, albeit their higher complexity in comparison to one-dimensional sensitivity analyses, these approaches do not include any advanced explanation systems. This shortcoming leads to an increase of the above mentioned gap between available and processible information.

Our contribution in this paper is therefore the presentation of a NLG approach providing explanations for multi-dimensional sensitivity analyses, i.e. explaining the results in the presence of multiple preferential uncertainties aimed at increasing user understanding in such decision situations. We developed explanatory text in an iterative process with experts and implemented the NLG approach in Matlab. We added the resulting explanation system as an extension module to the existing DSS SIMADA ('Simulation Based Multi-Attribute Decision Analysis', see Bertsch and Fichtner, 2016). In order to validate our concept with novice users we conducted an online survey and can show that the generated explanations are beneficial, particularly for rather difficult interpretational tasks.

We demonstrate what new results/explanations can be obtained by applying the developed NLG approach for a case study in the context of the energy sector transformation in Germany (see Bertsch and Fichtner, 2016). The energy sector typically involves long-term investment decisions in the presence of high uncertainties resulting from regulatory, techno-economic, ecological and social interdependencies as well as the preferential influences of different interest groups. This is the reason why a wide range of MCDA methods is used in this area which apply different methodologies in order to model the decision situation and the involved uncertainties (cf. Browne et al., 2010; Heo et al., 2010; Kaya and Kahraman, 2011; Streimikiene et al., 2012; Ribeiro et al., 2013; Ren et al., 2013; Lühn et al., 2014). For literature reviews of the application of MCDA methodologies in energy decision situations please refer to Greening and Bernow, 2004; Pohekar and Ramachandran, 2004; Diakoulaki et al., 2005; Zhou et al., 2006; Loken, 2007; Kowalski et al., 2009; Wang et al., 2009; Abu-Taha, 2011 and Scott et al., 2012.

This paper is structured as follows: In section 2, we review and summarise relevant literature related to (i) preferential uncertainty modelling in MCDA, (ii) benefits of explanation systems in general and (iii) existing explanatory features in MCDA tools. In section 3, we describe the conceptual structure and the main steps of the development process of our NLG approach before we present its evaluation in section 4. In section 5, we demonstrate the explanatory power of our NLG approach on the basis of a case study. Section 6 provides a discussion and limitations of the methodology. Section 7 concludes the paper.

2 Related work

In this section, we summarise and present relevant, existing literature related to our own work. While section 2.1 presents approaches for modelling preferential uncertainties in MCDA (with a focus on multi-dimensional sensitivity analysis), section 2.2 provides a short overview of user needs and benefits from explanation systems in DSSs. Section 2.3 presents existing implementations of explanation systems in MCDA tools. The selection of literature is adjusted to the focus of our paper, i.e. the presentation of a NLG approach explaining results of multi-dimensional sensitivity analyses. The selection of previous work can therefore not be comprehensive and will, to some extent, always be subjective.

2.1 Modelling of multiple preferential uncertainties

Uncertainties in decision situations originate from a variety of sources (Zimmermann, 2000), which arise with the application of different MCDA methodologies. This includes for example the preference elicitation which is influenced by behavioural effects as well as the limitations of modelling a decision situation in general (French, 1995; Gilovich et al., 2002; Hämäläinen and Alaja, 2008; Morton and Fasolo, 2009; Scholten et al., 2015). Numerous uncertainty classifications exist in the literature (Morgan et al., 1990; Belton and Stewart, 2002; Stewart, 2005; Bertsch, 2008; Geldermann, 2010) and there is also a variety of modelling approaches in MCDA theory to deal with them (cf. Durbach and Stewart, 2012; Broekhuizen et al., 2015).

Especially in group decision-making situations it can be challenging to attain a consensus regarding the individual preferences. The application of precise preference information in the DSS might therefore not be feasible motivating the application of parameter ranges for one or multiple preference parameters (Ríos Insua and French, 1991; Butler *et al.*, 1997; Matsatsinis and Samaras, 2001; Jiménez *et al.,* 2005; Mustajoki *et al.,* 2005; Mateos *et al.*, 2006; Mavrotas and Trifillis, 2006; Jessop, 2011; Jessop, 2014; Scholten *et al.*, 2015). Many of these approaches use Monte Carlo Simulation techniques, where a probability distribution needs to be defined for each preference parameter. In case there is none or strongly limited information on the parameter available, uniform distributions are often applied (see e.g., Bertsch and Fichtner, 2016).

For situations where very little or no preferential information is available or the DMs are unwilling to provide such information, Lahdelma *et al.* (1998) introduce Stochastic Multiobjective Acceptability Analysis (SMAA). SMAA has been used in many MCDA applications (Tervonen and Figueira, 2008). The method is designed as an inverse method aimed at exploring the weights for which an alternative achieves a certain rank, which is expressed in terms of a 'rank acceptability index' for each alternative (Lahdelma and Salminen, 2001). Tervonen (2014) developed an open source software which provides a user interface for the application of different SMAA approaches. While it definitively makes the analysis of a decision situation more convenient for a DA, the software does not include NLG techniques for explaining the used terminology or the implications for the decision situation that can be derived from the model results.

Bertsch *et al.* (2007) propose a simulation based approach for multi-dimensional sensitivity analysis in MAVT/MAUT similar to the one proposed by Butler *et al.* (1997). Both of these are designed as direct approaches as opposed to the inverse SMAA. The approach by Bertsch et al. (2007) has been implemented in the DSS SIMADA (Bertsch and Fichtner, 2016). SIMADA is mainly aimed at supporting two groups of users: decision makers (DM) and decision analysts (DA). Implemented in Matlab, SIMADA features a graphical user interface that provides various visualisations. These support a thorough analysis of the obtained model results and their sensitivity towards changes of various (uncertain) modelling parameters considered in the decision analysis, including inter-criteria preference parameters (i.e. weights) and intra-criteria preference parameters (i.e. value function shapes). For the latter, an exponential form is assumed (Kirkwood, 1997). With different preferential parameters applied in the Monte Carlo simulation runs, the aggregated model results in SIMADA are represented by value ranges of the alternatives' overall performance scores (OPS). In the MAVT module of SIMADA, the OPSs are calculated with an additive value function (Belton and Stewart, 2002; Basson and Petrie, 2007; French *et al.*, 2009). SIMADA also calculates the alternatives' expected overall performance scores (EOPS), which can be seen as an aggregate performance indicator (Durbach and Stewart, 2009). A selection of visualisations provided by SIMADA is presented in section 5.

Since the analysis of the influences of various uncertain preferential parameters on the model results can become rather complex, the focus of this paper is on the NLG approach to support the user in understanding the multi-dimensional sensitivities of the decision situation (further details of its conceptual structure and implementation are provided in sections 3 and 4).

2.2 User needs and benefits of explanations

To interpret and derive implications from a sensitivity analysis purely based on data and visualisations remains challenging for DMs and even for DAs (Hodgkin *et al.*, 2005). Simply providing the obtained model results of a decision situation under uncertainty does neither promote understanding nor supports judgmental performance of the users in an effective way (Hammond *et al.*, 1975; Brehmer, 1980; Hoffman *et al.*, 1981). This is due to the fact that the human brain is limited in processing large amounts of data (Silver, 1991a; Zimmermann, 2000; Linkov *et al.*, 2004; Kiker *et al.*, 2005) which may lead to systematic biases in the assessment of a decision situation (Sage, 1981; Hogarth, 1987; Parikh *et al.*, 2001; Bell *et al.*, 2003). For example, people pay inconsistent attention to the criteria (Gardiner and Edwards, 1975), neglect alternatives that do not reach a certain threshold performance (Tversky, 1972) or unintentionally focus on aspects that draw their initial attention (Kahneman and Knetsch, 1992). Furthermore, particularly for novice users, it is challenging to understand the underlying methodology of a DSS and its reasoning for a certain result (Spiegelhalter and Knill-Jones, 1984; Henrion and Druzdzel, 1991). This unfamiliarity leads to mistrust, especially in situations where they experience an expectation failure or anomaly comparing the output of the system to their own logic or belief (Ye, 1995; Gregor and Benbasat, 1999).

The described interpretational challenges of a DSS's model results can be partially resolved by explanations providing informative guidance with unbiased and relevant information (Silver, 1991a; Silver, 1991b). However, with respect to MCDA theory, this information should not include a specific suggestion in favour of or against a certain alternative. The aim is furthermore to describe the model's reasoning logic (Weiner, 1980; Buchanan and Shortliffe, 1984) and the used terminology (Gregor and Benbasat, 1999) as well as to support visual model results with explanations on how to interpret them which can be complemented by value tables or additional statistical analyses (Silver, 1991a).

Due to the increased transparency of the DSS, users are more likely to accept and trust its application on decision situations and the obtained results from their simulations (Swartout and Moore, 1993; Buchanan and Shortliffe, 1984; Greer *et al.*, 1994; Dhaliwal and Benbasat, 1996; Parikh *et al.*, 2001). Explanations that provide the user with relevant information and a meaningful interpretation of them reduce the cognitive effort of deriving these independently what results in higher acceptance and decision efficiency by the user (Hammond *et al.*, 1975; Brehmer, 1980; Hoffman *et al.*, 1981; Dhaliwal and Benbasat, 1996; Mao and Benbasat, 2000). In this way, the user can efficiently explore a decision situation and gain a more detailed understanding, which leads to more accurate judgement and increases the effectiveness of decision-making (Hayes and Reddy, 1983; Gregor and Benbasat, 1999; Parikh *et al.*, 2001; Tintarev and Masthoff, 2007). While novices use the explanations primarily to understand the obtained results, experts verify the underlying assumptions and resolve anomalies of the involved stakeholders. Therefore, explanatory DSSs have proven beneficial for fostering the understanding of both novices and expert users (Buchanan and Shortliffe, 1984; Greef and Neerincx, 1995; Ye, 1995; Gregor and Benbasat, 1999; Mao and Benbasat, 2000). However, it shall also be noted that the understanding of explanations is not unambiguous (Kahneman *et al.*, 1982) and can also cause behavioural influences on the DMs' judgement (Silver, 1991a).

2.3 Literature overview on MCDA DSSs with explanatory functions

Explanatory functions were originally developed for expert systems and in the field of artificial intelligence in order to improve human-computer interaction and communication (Amgoud and Prade, 2006; Geldermann, 2010; Ouerdane *et al.*, 2010). One possibility to provide the explanations is by NLG techniques (Holtzman, 1988; Silver, 1991a; Reiter and Dale, 1997). In the literature, we found three different approaches of how explanations are generated in MCDA DSSs.

1. The systems of the first category use MAUT in order to tailor their generated explanations to the user (cf. Greer *et al.*, 1994; Walker *et al.*, 2004; Carenini and Moore, 2006).
2. The DSSs introduced in Papamichail and French (2000), Papamichail and French (2003), Bélanger and Martel (2005), Labreuche *et al.* (2011), Labreuche *et al.* (2012), Greco *et al.* (2013), Sánchez-Hernández (2013) and Kadziński *et al.* (2014) provide user-independent explanations of the model results. Both user-dependent and user-independent NLG approaches use a template-based approach (cf. Reiter and Dale,

1997). The DSSs either adapt predefined text components to the obtained model results or choose entirely different templates for different user groups.

3. The concept presented in Bailey *et al.* (2011) constitutes a third category because it explains the results for a MCDA DSS by applying fuzzy logic (cf. Zadeh, 1965) to determine the corresponding explanations.

The concept presented in Papamichail and French (2000) and Papamichail and French (2003) differs from these systems in the output format of the explanations which is either a comparative or a sensitivity report file generated by a NLG module. The explanations are represented by five different types of messages which are structured according to the attribute tree of the decision problem. As an example, reasoning explanations provide arguments in favour or against an alternative. The NLG module fills predefined text templates with linguistic information which can either be quantitative data from the model or qualitative semantic quantifiers which verbally express the quality of the analysed parameters. The latter are chosen by the NLG module based on statistical calculations.

In addition to the benefits from explanations mentioned above, this report-generating concept also increases the traceability of the decision process which can be followed easily from the facilitated documentation. It is validated with various users from different backgrounds (Papamichail and French, 2005) and successfully applied in a group decision situation in Geldermann *et al.* (2009). However, the sensitivity analysis in this approach is limited to the variation of only one weight parameter at a time.

Overall, the above descriptions show that many approaches exist for multi-dimensional sensitivity analysis (without explanation systems) and for NLG based explanation systems (limited to one-dimensional sensitivity analyses in MCDA). To our knowledge, however, NLG approaches for explaining results of multi-dimensional sensitivity analysis have not yet been proposed in the literature.

3 Explanatory concept for a preferential multi-dimensional sensitivity analysis

For a formal description of the approach to multi-dimensional sensitivity analysis implemented in SIMADA, please see Bertsch and Fichtner (2016). In order to facilitate the users' understanding of the multi-dimensional preferential sensitivity analysis of the model results, SIMADA is extended by a NLG approach to provide automatic explanations for these. Based on the well investigated and validated approach by Papamichail and French (2000) and Papamichail and French (2003), the concept in this paper also generates the explanations as a report in HTML format based on predefined text templates. Our main contribution is the extension of their approach to generate explanations for simultaneous variations of different weights and value function parameters which allows a more detailed assessment of the preferential uncertainties involved in the decision situation. Both user groups (DMs and DAs) of SIMADA are addressees of this concept. While DAs can use it to validate the model results with their own expectations, DMs benefit from the facilitated interpretation and enhanced traceability of the decision. This is particularly relevant in the light of an increasing demand from the media and the public for information and justification from authorities on how decisions are taken (Wybo, 2006). However, the explanations may still require support from an experienced DA due to the complexity of the report.

3.1 General approach to report generation

Similar to the approaches by Papamichail and French (2003) and Geldermann et al. (2009), the generation of textual explanations of multi-dimensional sensitivity analysis in our approach involves three stages: (i) content determination, (ii) discourse planning and (iii) sentence generation. Please note that, given our focus on multi-dimensional sensitivity analysis, our notion of content determination in this paper is slightly different from that by Papamichail and French (2003). While Papamichail and French (2003) use this term basically to describe the choice between generating a comparative report vs. generating a sensitivity report, we shall use the term for the step in which users determine which part(s) of the multi-dimensional sensitivity analysis results they wish to generate explanations for. The three stages are each described in further detail in the subsequent section 3.2.

3.2 Development process of the report structure, text messages and templates

Klein (1994) and Reiter and Dale (1997) introduce the theoretical approach of developing a NLG system. Successfully conducted in Papamichail and French (2003), a similar process is followed in this paper. As mentioned above, the generation of textual explanations follows a three-stage procedure. Several experts from the field of decision theory and MCDA were closely involved on the different stages, i.e. from developing a report structure to formulating explanatory text templates, in order to critically discuss and reflect suggestions and ensure coherence and understandability of the report as a whole. The experts originate from Germany, South Africa, Netherlands and Finland; their research fields vary from operations research and statistics to decision theory with a particular focus on MCDA. We now describe the three stages of report generation in more detail and how the experts were involved in each of these. We also provide information on further thoughts, activities carried out and additional resources involved for each stage.

3.2.1 Content determination

In a first step, all visualisations of multi-dimensional sensitivity analysis results provided by SIMADA were collected. On this basis, we initially identified explanation needs which were subsequently discussed and refined in four expert interviews. The refinement process also included suggestions for adding new visualisations and corresponding explanations, which were not yet provided by SIMADA but the experts perceived as important (e.g., information on stochastic dominance of alternatives). As a result of the interviews, the following nine message types emerged, which can be grouped into four broad categories.

A. Overarching overview
    1. Introduction and explanation what multi-dimensional sensitivity analysis is
    2. Executive summary
B. Sensitivities of overall performance scores
    3. Spread of results
    4. Cumulative performance
C. Ranking performance
    5. Cumulative performance by alternative
    6. Detailed ranking performance
    7. Stochastic dominance
D. Preference parameter exploration
    8. Weight space exploration
    9. Value function space exploration

The exact requirements for explanations in a decision situation will be context-specific to a large extent. In the content determination stage, users are therefore given the opportunity to choose which type of explanations they wish to generate. However, the overarching messages of category A will always be provided. The messages in the other three categories provide information on the spread and distribution of the attained OPSs of the alternatives (category B), their rank performances for all combinations of preference parameters varied in the multi-dimensional sensitivity analysis (category C) and the preferential sensitivities on the first ranking performance of a considered alternative, i.e. for which parameters or parameter combinations will a certain alternative achieve the highest OPS (category D). Additionally, a nomenclature defining the used scientific terminology is automatically added as an appendix.

3.2.2 Discourse planning

Once users have decided in the content determination stage, which explanations they would like to generate, the main target of discourse planning is to provide a structure (text plan) according to which the different text messages will be ordered to achieve a coherent report. The development process of the discourse planning stage therefore involved an initial proposal of a text plan for each type of explanation. Again, this proposal was discussed and refined in interviews with the same four experts as in the content determination stage. The following general structure emerged from the discussions: Each message type of the multi-dimensional preferential sensitivity report (with few exceptions, see Table 1) will include (i) the *visualisation* itself, (ii) an explanatory *introduction* explaining how to interpret it, (iii) a *value table* containing the analysed parameter values and (iv) information providing *insight*, i.e. a list of implications which can be derived from the respective analysis or visualisation. Figures 1-4 show the corresponding library of text plans for the multi-dimensional sensitivity analysis report.

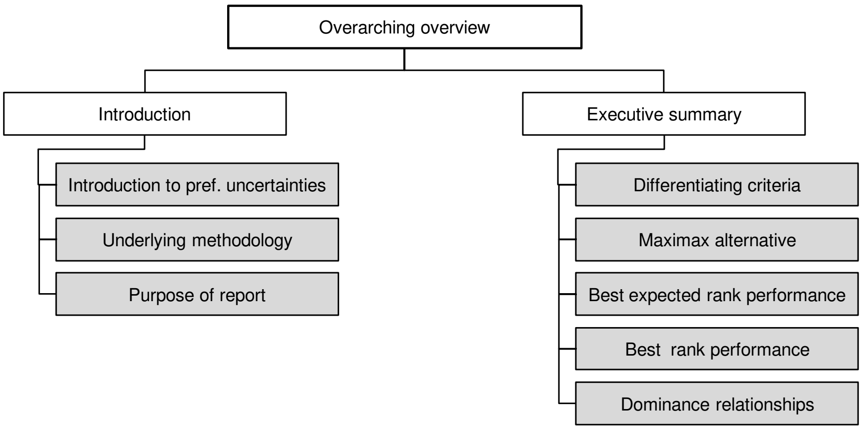


Figure 1: Text plan for the overarching overview messages

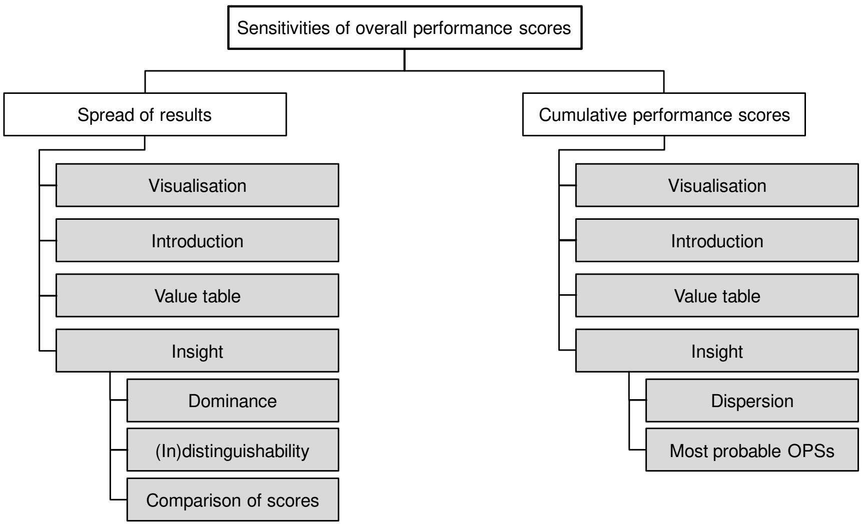


Figure 2: Text plan for the sensitivities of overall performance scores messages

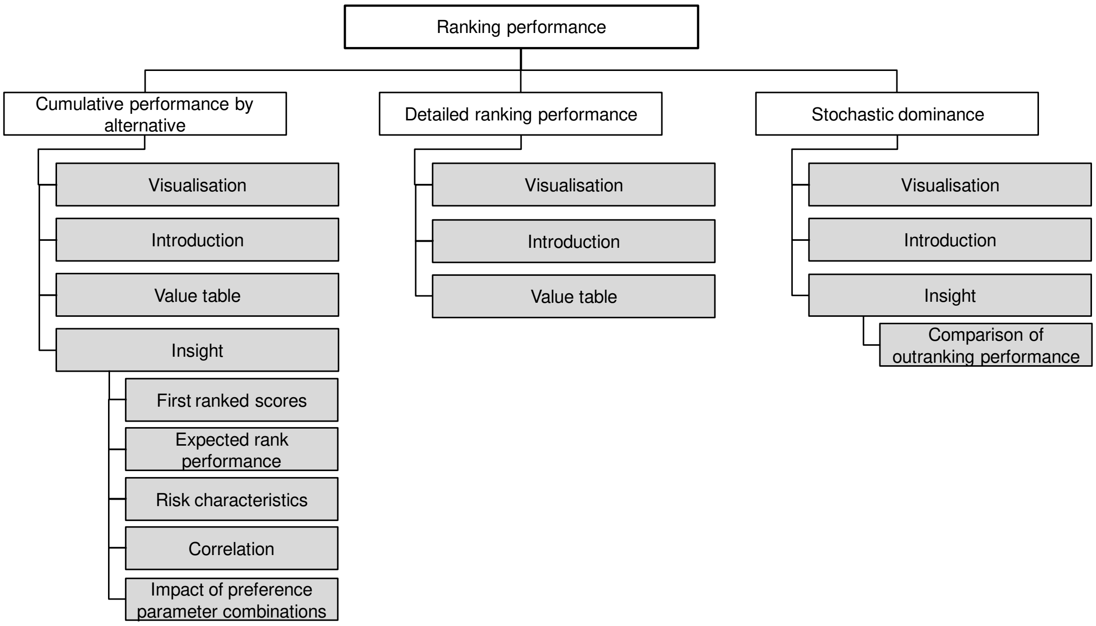


Figure 3: Text plan for the ranking performance messages

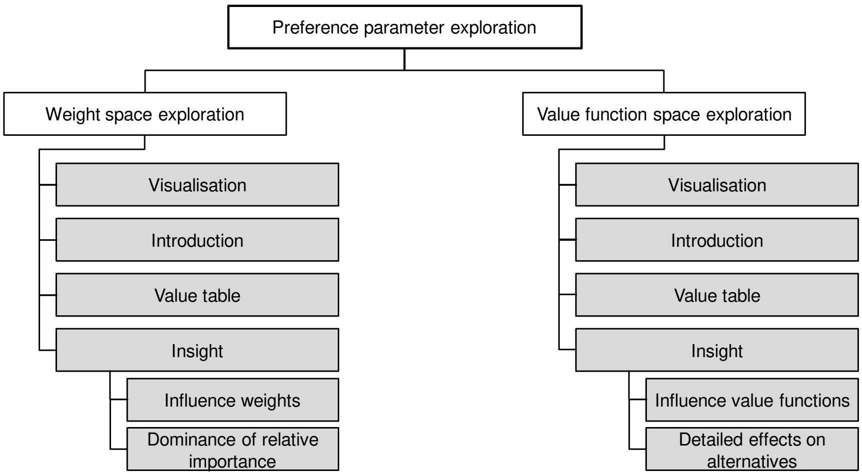


Figure 4: Text plan for the preference parameter exploration messages

Each of the grey boxes in Figures 1-4 represents at least one message to be conveyed to the user. The choice of wording for the explanatory messages and the structure of the descriptions are based on a number of literature sources (cf. Mareschal and Brans 1988; Brans and Mareschal 1994; Hodgkin et al. 2005; Treitz, 2006; Basson and Petrie 2007; Bertsch, 2008; Bertsch and Fichtner 2016) as well as self-developed concepts. Their formulation followed various design principles concerning relevance, conciseness and understanding (cf. Kass and Finin, 1988; Swartout and Moore, 1993; Nunes et al., 2012; Corrente et al., 2014). In terms of the input data required to generate each of the messages, they can be categorised into three broad groups (see below). For each group an example of a corresponding template is provided below (in *italics* in this section).

- Messages of the *first group* require *no input data*. These messages only require unchanged or ‘canned’ text (see Papamichail and French, 2003). For instance, report titles belong to this group, but also the messages related to the first message type in category A of the content determination (see section 3.2.1) are simply canned text messages. The introduction to the ‘spread of results’ visualisation is an example of this group of messages:

  The 'spread of results' graph shows the ranges of the overall performance scores for all alternatives. The vertical lines with tick marks at their ends represent the minimum and maximum results obtained from the simulation runs. The tick mark in their middle indicates the expected overall performance score (EOPS) for the respective alternative (e.g. the value represents the average overall performance score in case of symmetric distributions).

  Please note: This visualisation does not show the distribution of the overall performance scores for the alternatives.

- Messages of the *second group* require *directly available data*. Examples of directly available data include alternatives’ names or OPSs which are simply included in the sentences without any further transformation. An example of this group of messages is (where the words in <> brackets indicate slots in the templates which are to be filled by directly available data):

  The highest overall performance score is attained by <Alternative a_MaxiMax>. It maximises the upside potential of realizing the highest possible overall performance score.

- Messages of the *third group* require *computable data*. Messages of this kind are needed for almost all sentences related to the ‘insight’ boxes in Figures 1-4. The computable data is either included in the sentences in numerical form or in the form of a semantic quantifier, i.e. a verbal expression that may change depending on a numerical indicator taking different values or value ranges. Computable data in numerical form may either be included in individual sentences or in table form. An example of this group of messages including computed data in the form of a semantic quantifier is given by (where the words in <> brackets indicate slots in the templates which are to be filled by either directly available or computable data):

  The overall performance scores of <alternative a_cons> are <semantic quantifier> dispersed than those of <alternative(s) a_i>.

The nature of the text templates required to generate the various messages will be different for each of the three groups. While templates of the first group are simple in nature, the templates of the second and third group contain placeholders, which are replaced by strings or numerical values in the sentence generation stage. Table 1 below provides an overview of the structure of the template library.

Table 1: Structure of the library of templates

Determined content	Message type	Individual messages and sentences			
		Visualisation	Introduction	Value table	Insight
A. Overarching overview	1. Introduction	-	Template A1.1 (canned text)	-	-
	2. Executive summary	-	-	-	Templates A2.1-A2.5 (canned text, directly available and computable data)
B. Sensitivities of OPSs	3. Spread of results	Template B3.1 (Figure)	Template B3.2 (canned text)	Template B3.3 (canned text, directly available and computable data)	Templates B3.4-B3.6 (canned text, directly available and computable data)
	4. Cumulative performance scores	Template B4.1 (Figure)	Template B4.2 (canned text)	Template B4.3 (canned text, directly available and computable data)	Templates B4.4-B4.5 (canned text, directly available and computable data)
C. Rank performance	5. Cumulative performance sorted by alternative	Template C5.1 (Figure)	Template C5.2 (canned text)	Template C5.3 (canned text, directly available and computable data)	Templates C5.4-C5.8 (canned text, directly available and computable data)
	6. Detailed ranking performance	-	Template C6.1 (canned text)	Template C6.2 (canned text, directly available and computable data)	-
	7. Stochastic dominance	Template C7.1 (Figure)	Template C7.2 (canned text)	-	Template C7.3 (canned text, directly available and computable data)
D. Parameter exploration	8. Weight space exploration	Template D8.1 (Figure)	Template D8.2 (canned text)	Template D8.3 (canned text, directly available and computable data)	Templates D8.4-D8.6 (canned text, directly available and computable data)
	9. Value function space exploration	Template D9.1 (Figure)	Template D9.2 (canned text)	Template D9.3 (canned text, directly available and computable data)	Templates D9.4-D9.5 (canned text, directly available and computable data)

After refining and formulating all templates on the basis of the discussions in the four expert interviews, the resulting text templates were sent to four additional experts who provided written feedback which we incorporated in an additional iteration. Overall, all experts provided various ideas for reducing the complexity of the explanations (e.g., regarding sentence length and use of language), increasing user understanding by adding additional textual and visual information and suggesting new concepts (e.g., stochastic dominance).

3.2.3 Sentence generation

Once the users have chosen for which results they wish to generate explanations in the content determination stage and the overall structure of the text is determined by the corresponding text plan from the library of text plans in the discourse planning stage, the main task of the sentence generation stage is to fill the placeholder slots of the templates with numerical values or strings.

In a final stage, all generated messages are assembled in a HTML file. To ensure a coherent layout, a Cascading Style Sheet (CSS) produced in the discourse planning stage defines various style specifications for headlines, tables and paragraphs.

In the following subsections, details of the sentence generation stage are described for a selection of messages and templates. While the descriptions below generally cover all three groups of messages mentioned in section 3.2.2, we put special emphasis on messages requiring computable data. For each described message, we also provide the corresponding text template(s) (again in *italics*).

3.2.3.1 Overarching overview: Introduction

The overarching introduction into the report is mostly based on a ‘canned text’ template, where only very little adaptations are made. Even though scientific terminology is reduced to a minimum in the report, model-specific terms like ‘overall performance score’ or ‘simulation run’ cannot be completely avoided. In terms of layout, these terms are generally underlined in the templates indicating that they are linked to the nomenclature which is also represented as a predefined template in the NLG module to provide easily understandable definitions of such terms to users. They are also represented as tooltips in the report which appear when the cursor is moved over the respective link in the document.

This analysis examines the robustness of the simulation results of a decision situation with respect to the influences of preferential uncertainties. The decision situation is modelled by ***Multi-Attribute Value Theory (MAVT)*** *with an additive value function where the underlying preferential uncertainties are expressed by assigned parameter intervals. These* *include intra-criteria preferential uncertainties (regarding the criteria level ranges of the alternatives) as well as inter-criteria preferential uncertainties (regarding the relative importance of the criteria). The intra-criteria preferential uncertainties are modelled by the variation of the value function shapes (curvature variations) while the inter-criteria preferential uncertainties are represented by weight variations.*

*The aim of this analysis is to identify the most relevant preferential uncertainties in order to explore their respective impact on the results and to examine how the alternatives are distinguishable from each other in the light of these uncertainties. Therefore, this report presents the results of the analysis of **<amount> simulation runs** of the decision situation for which **<amount> different alternatives** are considered. In each of the simulation runs, the uncertain parameter samples are varied randomly with respect to the assigned interval boundaries.*

3.2.3.2 Sensitivities of overall performance scores: Explaining the 'spread of results' visualisation

The explanation of the spread of results includes the following information:

- Visualisation
- Introduction
- Value table
- Insight: Dominance, (In)distinguishability, Comparison of scores

Examples of the visualisation, introduction and value table are provided in section 5 as well as the appendix. Here, we focus on the sentence generation for the insight-related messages.

Dominance

As part of this explanation, strict dominance relationships between two alternatives are explored and reported. It occurs if all OPSs of an alternative $\mathrm{a_i}$ are strictly better than those of alternative $\mathrm{a_j}$ ($\mathrm{a_i} \neq \mathrm{a_j} \in \mathrm{A}$) in every simulation run. Therefore, the system verifies condition (1) for all alternatives if the alternative $\mathrm{a_i}$ strictly dominates another alternative $\mathrm{a_j}$ and fills the template below accordingly (it is not filled if the condition (3.4) is not met for an alternative):

- *<Alternative a_i> dominates alternative(s) <Alternative a_j> for all preference parameter combinations within the considered intervals.*

$$\min_{\mathrm{s_j}}\left\{\mathrm{OPS}_{\mathrm{a_i,s_j}}\right\} > \max_{\mathrm{s_j}}\left\{\mathrm{OPS}_{\mathrm{a_j,s_j}}\right\}, \forall\, \mathrm{a_i} \neq \mathrm{a_j} \in \mathrm{A} \tag{1}$$

(In)distinguishability

On the other hand, if the OPS ranges of the alternatives overlap, they are considered indistinguishable with regard to their performances. This is expressed in equation (2)(2).

- *The overall performance scores of <Alternative a_i> are indistinguishable from the overall performance scores of alternative(s) <Alternative a_j>.*

$$\min_{s_j}\left\{\mathrm{OPS}_{a_i,s_j}\right\} \leq \max_{s_j}\left\{\mathrm{OPS}_{a_j,s_j}\right\} \wedge \max_{s_j}\left\{\mathrm{OPS}_{a_i,s_j}\right\} \geq \min_{s_j}\left\{\mathrm{OPS}_{a_j,s_j}\right\},\ \forall\, a_i \neq a_j \in A \quad (2)$$

Comparison of scores

In decision theory under uncertainty, several concepts have been developed to identify the preferred alternative(s) in the presence of uncertainty (cf. Neumann, 1928). Making use of these, the NLG module identifies the MaxiMax, MiniMin and MaxiMin alternatives amongst all $a_i \in A$ respectively, where A is the set of all alternatives, $1 \leq i \leq m$ and $m \in \mathbb{N}$ is the total number of alternatives of the simulation. This is done by comparing the alternatives' OPSs obtained in the simulation runs $s_j \in S$, where S is the set of all simulation runs, $1 \leq j \leq n$ and $n \in \mathbb{N}$ is the total number of simulation runs, to the MaxiMax, MiniMin and MaxiMin OPSs of the simulation as defined in the conditions (3), (4) and (5). This information is then used to fill the text templates below.

$$a_{MaxiMax} = \left\{a_{i*} \in A \,\middle|\, OPS_{a_{i*},s_j} = \max_{a_i}\left\{\max_{s_j}\left\{\mathrm{OPS}_{a_i,s_j}\right\}\right\}\right\} \quad (3)$$

$$a_{MiniMin} = \left\{a_{i*} \in A \,\middle|\, OPS_{a_{i*},s_j} = \min_{a_i}\left\{\min_{s_j}\left\{\mathrm{OPS}_{a_i,s_j}\right\}\right\}\right\} \quad (4)$$

$$a_{MaxiMin} = \left\{a_{i*} \in A \,\middle|\, OPS_{a_{i*},s_j} = \max_{a_i}\left\{\min_{s_j}\left\{\mathrm{OPS}_{a_i,s_j}\right\}\right\}\right\} \quad (5)$$

- *The highest overall performance score is attained by <$Alternative$ $a_{MaxiMax}$>. It maximises the upside potential of realizing the highest possible overall performance score.*
- *The lowest overall performance score is attained by <$Alternative$ $a_{MiniMin}$> (<value>).*
- *Alternative <Alternative $a_{MaxiMin}$> attains the highest minimum of all alternatives (<value>). This alternative maximises the minimal overall performance scores of all alternatives.*

3.2.3.3 Sensitivities of overall performance scores: Explaining the 'cumulative performance' visualisation

The explanation of the cumulative performance includes the following information:

- Visualisation

- Introduction
- Value table
- Insight: Dispersion, Most probable OPSs

As in the previous section, we focus on the sentence generation for the insight-related messages here and refer to the appendix for further information related to the other messages.

Dispersion

The dispersion message uses semantic quantifiers for the verbal description of certain observations of the visualisations in the text templates. Motivated by Papamichail and French (2000), they are based on statistical calculations of certain model parameters. As an example, the predefined text template describing the dispersion of the OPSs of the considered alternative $\mathrm{a_{cons}}$ in pairwise comparison to the dispersion of the other alternatives' OPSs is shown below. These explanations complement the cumulative performance visualisation which shows distributional information for the alternatives' OPSs. The text templates are filled with a semantic quantifier to describe the degree of difference in dispersion between them. It is determined in the following way, where $\lambda = 1.96$ to calculate a 95 % confidence interval for the considered alternative's standard deviation $\sigma_{a_{cons}}$:

- *The overall performance scores of <Alternative a_{cons}> are <semantic quantifier> dispersed than those of <Alternative(s) a_i>.*

$$\text{if } \sigma_{a_{cons}} > \sigma_{\mathrm{a_i}} + \lambda * \sigma_\sigma \longrightarrow \text{'much more'}, \quad \forall\, \mathrm{a_i} \neq \mathrm{a_{cons}} \in \mathrm{A} \tag{6}$$

$$\text{if } \sigma_{\mathrm{a_i}} < \sigma_{a_{cons}} \leq \sigma_{\mathrm{a_i}} + \lambda * \sigma_\sigma \longrightarrow \text{'more'}, \quad \forall\, \mathrm{a_i} \neq \mathrm{a_{cons}} \in \mathrm{A} \tag{7}$$

$$\text{if } \sigma_{\mathrm{a_i}} = \sigma_{a_{cons}} \longrightarrow \text{'equally'}, \quad \forall\, \mathrm{a_i} \neq \mathrm{a_{cons}} \in \mathrm{A} \tag{8}$$

$$\text{if } \sigma_{\mathrm{a_i}} > \sigma_{a_{cons}} \geq \sigma_{\mathrm{a_i}} - \lambda * \sigma_\sigma \longrightarrow \text{'less'}, \quad \forall\, \mathrm{a_i} \neq \mathrm{a_{cons}} \in \mathrm{A} \tag{9}$$

$$\text{if } \sigma_{a_{cons}} < \sigma_{\mathrm{a_i}} - \lambda * \sigma_\sigma \longrightarrow \text{'much less'}, \quad \forall\, \mathrm{a_i} \neq \mathrm{a_{cons}} \in \mathrm{A} \tag{10}$$

Most probable OPSs

This section of the report analyses the inter-quantile range of the 5 % and 95 % quantile of the considered alternative's OPSs. The boundaries of the inter-quantile range are stated and compared and compared to the quantiles and/or EOPSs of the other alternatives. This can be seen as a weaker relationship than the strict dominance of alternatives and accounts for the fact that values outside of this inter-quantile range (5-95%) can potentially represent outliers.

- *The 90 % most probable overall performance scores of Alternative <Alternative name> in the executed simulation runs are between <value> and <value>.*

- *<All/The highest 95 %> overall performance scores of Alternative <Alternative name> dominates <the 95 % highest/all> overall performance scores of Alternative <Alternative name>.*
- *The expected overall performance score of Alternative <Alternative name> dominates <the 95 % highest/all> overall performance scores of Alternative <Alternative name>.*

3.2.3.4 Ranking performance: Explaining the 'cumulative performance sorted by alternative' visualisation

The explanation of the cumulative performance sorted by alternative includes the following information:

- Visualisation
- Introduction
- Value table
- Insight: First ranked scores, Expected rank performance, Risk characteristics, Correlation, Impact of preference parameter combinations

In contrast to the above sections, we focus on the sentence generation for the value table and the insight-related messages here and refer to the appendix for further information related to the other messages.

Value table

This value table provides information on the percentage of simulation runs in which they achieve a certain rank within a simulation for each alternative. The system calculates a ranking matrix as shown in equation (11) where an entry $\mathrm{d_{i,j}}$ is defined by the rank of alternative $\mathrm{a_i} \in \mathrm{A}$ in the simulation run $\mathrm{s_j} \in \mathrm{S}$. Afterwards, the number of equal entries in every line is divided by the number of simulation runs. This indicates the relative percentage of simulation runs that the alternatives achieved the respective rank. These numbers also represent the entries for the value table of the cumulative performance sorted by alternative visualisation.

$$\mathrm{ranking} = \begin{pmatrix} \mathrm{d_{1,1}} & \cdots & \mathrm{d_{1,j}} & \cdots & \mathrm{d_{1,n}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{d_{i,1}} & \cdots & \mathrm{d_{i,j}} & \cdots & \mathrm{d_{i,n}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathrm{d_{m,1}} & \cdots & \mathrm{d_{m,j}} & \cdots & \mathrm{d_{m,n}} \end{pmatrix} \tag{11}$$

$$\text{where } \mathrm{d_{i,j}} = \mathrm{rank}(a_i, s_j),\ \mathrm{a_i} \in \mathrm{A}, \mathrm{s_j} \in \mathrm{S}. \tag{12}$$

In addition to the table itself, a list of further aspects is provided based on the information in matrix (11):

- *The <highest/lowest> percentage of <No. 1/last> ranks is attained by Alternative <Alternative name>.*
- *Alternative <Alternative name> does never <become the preferred alternative/attain the last rank>.*
- *The overall performance scores that <Alternative* a_{cons}*> attains in the simulation runs where it ranks first range from <value> to <value>.*

Especially the last aspect provides detailed information on the maximal value that the considered alternative can attain in simulation runs where it becomes the preferred one.

First ranked scores

This paragraph analyses the results the considered alternative obtains in the simulation runs where it ranks first. It indicates the range of obtained overall performance scores and calculates the expected overall performance scores for these. Lastly, the first-ranking performances of the considered alternative are compared to the first-ranking performances of all alternatives. This lets the DM or DA interpret if the considered alternative attains high or low performance scores when it becomes the preferred alternative in the simulation.

- *The overall performance scores that Alternative <Alternative name> attains in the simulation runs where it ranks first range from <value> to <value>.*
- *The expected overall performance score of the simulation runs where Alternative <Alternative name> becomes the preferred alternative accumulates to <value>.*
- *This value is <value> % <higher/lower> compared to the expected overall performance score of all first ranking overall performance scores. This means, that in the simulation runs in which Alternative <Alternative name> ranks best, it also attains <high/low> overall performance scores.*

Expected rank performance

Based on equation (13) the system outputs the best, worst and in case the considered alternative is not one of these, also the expected rank of this alternative in text form.

$$rank_{a_i,expected} = \frac{\sum_{j=1}^{n} d_{i,j}}{n} \quad \forall\ \mathrm{a_i} \in A \tag{13}$$

- Alternative <Alternative name> does attain the <best/worst> expected rank of <expected rank> out of <number of alternatives> alternatives.
- Alternative <Alternative name> attains an expected rank of <value> out of <number of alternatives> alternatives.

Risk characteristics

In order to provide insights into the risk of choosing a low performing alternative, the report calculates the regret (i.e. opportunity loss) for every alternative by equation (14). This can be seen as the sum of the difference of OPS between one alternative and the first-ranking alternative of every simulation run (cf. Loomes and Sugden, 1982; Bell, 1982). The minimal regret value identifies the MiniMax regret alternative. A potential decision for this alternative minimises the risk of obtaining low performance results.

$$regret_{a_i} = \frac{\sum_{j=1}^{n} \max_{\mathrm{s_j}} \left\{ \mathrm{OPS}_{\mathrm{a_i,s_j}} \right\} - OPS_{a_i,s_j}}{n}, \mathrm{a_i} \in \mathrm{A}, \mathrm{s_j} \in \mathrm{S}. \tag{14}$$

- *Alternative <Alternative name> does achieve the highest expected overall performance score. This alternative minimises the downside risk of obtaining a low overall performance score with regard to the best alternative of every single simulation run. It can be expected that this alternative does attain an overall performance score that is <value> <higher/lower> than the best overall performance score of a simulation run.*

Correlation

In order to describe the linear correlation between the alternatives, Pearson's correlation coefficient is calculated with the alternatives' OPSs (for a discussion on correlation coefficients cf. Hauke and Kossowski, 2011; Bishara and Hittner, 2012; Pagano, 2013). Since for a high number of data samples already small effects can cause significant influences regarding the correlation of two alternatives, the effect of correlation is of particular interest (Ellis, 2013). This degree is expressed in the following text template by the semantic quantifiers 'very small', 'small', 'medium' and 'large' according to the effect size classification of correlation in Cohen (1988) and Ellis (2010). However, only if the correlation can be considered significant, the template is generated for the report. This is verified by calculating the p-value of the correlation with the t-statistic (Gosset, 1908) and a significance level of $\alpha = 5\ \%$.

- *<Alternative a_{cons}> correlates <positively/negatively> with <Alternative a_i> to a <semantic quantifier> extent. This correlation is significant (p = <$p_{a_i,a_{cons}}$>).*

$$if\ \ 0 < r_{a_i,a_{cons}} \leq 0{,}1\ \rightarrow \text{'}very\ small\text{'}, \quad \forall\, \mathrm{a_i} \neq \mathrm{a_{cons}} \in \mathrm{A} \tag{15}$$

$$if\ \ 0{,}1 < r_{a_i,a_{cons}} \leq 0{,}3\ \rightarrow \text{'}small\text{'}, \quad \forall\, \mathrm{a_i} \neq \mathrm{a_{cons}} \in \mathrm{A} \tag{16}$$

$$if\ \ 0{,}3 < r_{a_i,a_{cons}} \leq 0{,}5 \rightarrow \text{'}medium\text{'}, \quad \forall\, \mathrm{a_i} \neq \mathrm{a_{cons}} \in \mathrm{A} \tag{17}$$

$$if\ \ r_{a_i,a_{cons}} > 0{,}5 \rightarrow \text{'}large\text{'}, \quad \forall\, \mathrm{a_i} \neq \mathrm{a_{cons}} \in \mathrm{A} \tag{18}$$

Impact of preference parameter combinations

In this section of the report, the influence of separate considerations of inter- and intra-criteria preferential uncertainties is compared. It is done by running the same simulation though considering different combinations of these uncertainty types. The first-ranking performance (percentage of first ranks obtained in the simulations) is used as an indicator for the considered alternative in the report to determine how sensitive it is with regard to the different uncertainty type combinations.

The overall performance score of Alternative <Alternative name> is most sensitive to <uncertainty type>.

- *With inter- as well as intra-criteria preferential uncertainties taken into account Alternative <Alternative name> attains rank 1 in <value> % of the cases.*
- *Considering only inter-criteria preferential uncertainties for this alternative, it ranks first in <value> % of the simulation runs.*
- *When only the intra-criteria preferential uncertainties are modelled, Alternative <Alternative name> achieves the first rank in <value> % of the simulation runs.*

3.2.3.5 Ranking performance: Explaining 'stochastic dominance' relations

The explanation of stochastic dominance includes the following information:

- Visualisation
- Introduction
- Insight: Comparison of outranking performance

As in most previous sections, we focus on the sentence generation for the insight-related messages here and refer to the appendix for further information related to the other messages.

Comparison of outranking performance

In contrast to the strict dominance relationship which considers the OPSs obtained by the compared alternatives as a whole set, stochastic dominance (see Levy, 1992; Graves and Ringuest, 2009; Eisenführ et al., 2010; Scholten et al., 2014) is defined in this paper as the percentage of simulation runs in which one alternative achieves an equal or higher OPS than a compared alternative (equation (19)). The following text output is generated for each pairwise comparison of the considered alternative $\mathrm{a_{cons}}$ of the report with all of the other alternatives $\mathrm{a_i} \neq a_{cons} \in A$.

- *<Considered alternative ID> dominates <alternative ID> in <$stochDom_{a_{cons},a_i}$> % of the simulation runs.*

$$stochDom_{a_{cons},a_i} = \frac{countIf\left(OPS_{a_{cons},s_j} \geq OPS_{a_i,s_j}\right)}{n} * 100, \forall\ \mathrm{a_i} \neq \mathrm{a_{cons}}\ \in \mathrm{A}, s_j \in S. \quad (19)$$

3.2.3.6 Preference parameter exploration: Explaining the weight space exploration visualisation

The explanation of the weight space exploration includes the following messages:

- Visualisation
- Introduction
- Value table
- Insight: Influence weights, Dominance of relative importance, (In)distinguishability of relative importance

As above, we focus on the sentence generation for the insight-related messages here and refer to the appendix for further information related to the other messages.

Influence weights

The weight space exploration visualisation compares the total weight space (containing the weights w_{tot,c_k}) for each individual criterion with the so-called limited weight space. This subset of the total weight space contains the weights w_{lim,c_k} which are applied in the simulation runs where the respective alternative obtains the first rank. We now compare the deviation of weight range (i.e. difference between the maximum and minimum weights) of the total and limited weight space for every criterion individually. If this value differs by at most 10 % from the maximum (minimum) of the deviations of all criteria range deviations, the respective criterion has a high (low) influence on the first-ranking performance of the analysed alternative. In this case, the condition (20a) (or (20b)) is fulfilled.

$$1-\frac{\max\left(w_{lim,c_k}\right)-\min\left(w_{lim,c_k}\right)}{\max\left(w_{tot,c_k}\right)-\min\left(w_{tot,c_k}\right)} \geq \max\left(\Delta w_{c_k}\right)-0{,}1*\left(\max\left(\Delta \mathrm{w}_{c_k}\right)-\min\left(\Delta \mathrm{w}_{c_k}\right)\right) \tag{20a}$$

$$1-\frac{\max\left(w_{lim,c_k}\right)-\min\left(w_{lim,c_k}\right)}{\max\left(w_{tot,c_k}\right)-\min\left(w_{tot,c_k}\right)} \leq \min\left(\Delta w_k\right)+0{,}1*\left(\max\left(\Delta \mathrm{w}_{c_k}\right)-\min\left(\Delta \mathrm{w}_{c_k}\right)\right) \tag{20b}$$

- *The following criterion is most sensitive for the ranking of Alternative <Alternative name> as preferred alternative: <list of criteria>.*
- *The relative importance of criteria <list of criteria> slightly affect the ranking of Alternative <Alternative name> as preferred alternative.*

Dominance of relative importance

Similar to the dominance relationship defined in equation (1) we also analyse potential dominance relationships between the relative importance of the criteria. However, we use each criterion's upper and lower boundaries of the total and limited weight space instead of the OPSs as input for equation (1). The following text is generated in case the dominance relationship occurs already with the application of the total weight space, i.e. in all simulation runs (first bullet point). The text of the second bullet point is generated in the case where there occurs a dominance relationship only when the simulation runs in which the alternative ranks first are considered (i.e. the limited weight space is applied).

- *Criterion <Criterion name> is more important than criterion <Criterion name> in all simulation runs.*
- *Only in the simulation runs where Alternative <Alternative name> ranks first, criterion <Criterion name> is more important than criterion <Criterion name>.*

4 Evaluation of the approach

After the implementation of the natural language generator on the basis of the concept presented in section 3, we have evaluated our approach with both expert users and novice users.

4.1 Expert users

Five experts, that were part of the feedback rounds referred to in section 3, were contacted again to provide additional comments on the concept after its implementation. The aim of this interview loop was twofold. First, we intended to gather assessments by experts of the added value provided by the implemented approach. Second, we sought feedback for final adjustments on the content of the explanations, especially focussing on the more detailed results that were available and explained in the report as a result of the implemented natural language generator. In relation to the first aim, the overall attitude expressed by the experts was very positive in general.

4.2 Novice user validation

In order to validate the usefulness of our NLG approach to explain a multi-dimensional preferential sensitivity analysis to novice users, an online survey was conducted. The survey was completed by a total of 268 participants. The comparison of the collected socio-demographic data of our sample with the average values of the German population particularly shows the following deviations. The share of participants aged between 18 and 30 years and with a high educational background is overrepresented as compared to the German population (according to the German Federal Statistical Office). Also the share of male participants (75%) is higher than the average of 49% of the German population.

The survey tested user understanding on the basis of the ‘spread of results’ visualisation for a hypothetical decision situation in a between subject design. After a brief introduction to MCDA and preferential uncertainty modelling in SIMADA, the participants had to answer questions which tested their understanding of the visualised information. While one sample group was provided with the explanations generated by the NLG approach presented in this paper in addition to the visualisation, the other half had to answer the questions without these explanations (i.e. only on the basis of the visualisation). The tasks differed in difficulty as they required different levels of interpretational capabilities. While some questions asked the participants to identify the value of an alternative's OPS (e.g. 'What is the maximum overall performance score of Alternative 2?'), another type of question requested them to compare the alternatives (e.g. 'What is the minimal overall performance score of all alternatives?'). With regard to the taxonomy of educational objectives by Bloom *et al.* (1984), these two question types verify if the participants understand the visualisation in terms of what it shows and if they can apply this knowledge when they are confronted with slightly more difficult comparison tasks. The two sample groups did not show a significant difference of understanding (proxied by the number of correct and wrong answers). Both sample groups answered over 80 % of the questions correctly. For these basic tasks, we can therefore conclude that the understanding of the participants did not depend on the provision of explanations.

A further question, however, also tested deeper analysis and interpretational capabilities of the participants. They were asked about distributional information on the alternatives' OPSs ('Are high overall performance scores more probable for Alternative 3 than low overall performance scores?'). In theory, this question was not more difficult to answer than the questions before for the sample group that was provided with explanations. They were provided with the explanation to answer this question correctly. The other user group, however, had to use their own interpretational capabilities to answer it. Almost twice as many participants of the former sample group (30 %) answered this question correctly as opposed to 17 % of the participants that were not provided with explanations. Pearson's Chi-squared test (Pearson, 1900) as well as Fisher's Exact test (Fisher, 1922) indicate a significant difference between the samples ($p=0.03$ and $p=0.02$ respectively). We thus conclude that the NLG approach is particularly useful for complex interpretational tasks (for the example of the spread of result visualisation) by reducing the user's cognitive effort to achieve understanding. However, the result also shows that a considerable number of participants who were provided with explanations did not read them carefully enough to retrieve the correct answer.

5 Demonstration of the natural language generation approach for a case study

We now apply our NLG approach to the MCDA case study presented in Bertsch and Fichtner (2016). The background of the case study is the multi-criteria evaluation of five decision alternatives in the context of the German energy transition. Each alternative includes a combination of different feed-in regimes for renewable electricity generation (i.e. curtailment options) and electricity transmission grid expansion possibilities (i.e. capacity reinforcement measures of existing lines or construction of new lines). The decision to be supported is therefore owned by the regulator and policy makers. The five considered alternatives are:

- A1: Economic RES injection management and economic grid expansion
- A2: Economic RES injection management and fixed, maximal grid expansion
- A3: Fixed, maximal RES injection and fixed, maximal grid expansion
- A4: RES injection fixed to 90% of max. and economic grid expansion
- A5: RES injection fixed to 90% of max. and fixed, maximal grid expansion

These five alternatives are evaluated with respect to the three traditional dimensions of energy policy (economic competitiveness, environmental sustainability, security of supply), which Bertsch and Fichtner (2016) augmented by public acceptance as a fourth key dimension. Figure 5 shows the corresponding attribute tree for their case study.

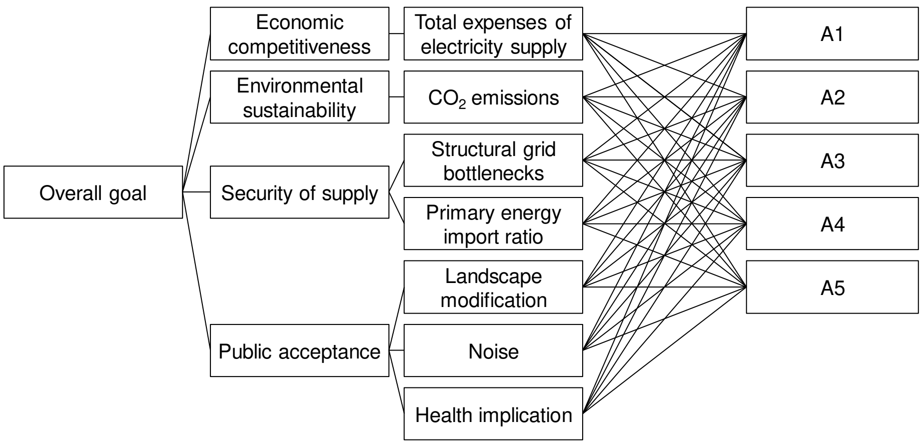


Figure 5: Attribute tree for the case study (Bertsch and Fichtner, 2016)

Table 2 shows the performance of the five alternatives with respect to the attributes of the attribute tree. For the top level criteria of the attribute tree, the following weight intervals are assumed for the multi-dimensional sensitivity analysis (Bertsch and Fichtner, 2016): Economic competitiveness (0.10-0.30), Environmental sustainability (0.15-0.55), Security of Supply (0.20-0.60), Public acceptance (0.05-0.25). While our NLG approach does provide textual explanations for varying the value functions' shapes, we focus on the text messages explaining the results of a multi-dimensional weight sensitivity analysis in this paper. The value functions are therefore assumed to be linear in the context of the case study. For all further details related to the case study in general as well as specific modelling assumptions, please see Bertsch and Fichtner (2016).

Table 2: Decision table for the case study (Bertsch and Fichtner, 2016)

Criterion	Scale	A 1	A 2	A 3	A 4	A 5
Total expenses of electricity supply	[Billion €]	182	200	224	204	220
CO_2 emissions	Million t CO_2/year	204	201	158	175	162
Structural grid bottlenecks	%	5	2	2	5	2
Primary energy import ratio	%	28	21	18	20	16
Landscape modification	Point scale	0	1.3	1.3	0.2	1.3
Noise	Point scale	0	0.2	0.2	0.05	0.2
Health implication	Point scale	0	0.4	0.4	0.05	0.4

While the original version of SIMADA used in Bertsch and Fichtner (2016) also supported multi-dimensional sensitivity analysis, the main difference is the following: The original version enabled the generation of mainly three visualisations. The version including the NLG approach presented in this paper enables, in addition to these visualisations, the automatic generation of a comprehensive natural language report explaining the impact of simultaneous variations of preference parameters on the MCDA results. As mentioned above, users have the opportunity to actively choose which type of explanations they wish to generate. However, the executive summary listing the key findings of the analysis is always provided in the beginning of each report. Such an executive summary is shown in the box below for the data and parameters of the case study. A full version, i.e. for users that chose to generate 'all' available explanations for simultaneous weight variations within the above mentioned intervals, of the multi-dimensional sensitivity report (including visualisations and textual explanations) is presented in Appendix A.

Executive summary

Overall, the following key aspects are characteristic for this decision situation and the executed simulation runs:

- The criterion CO_2 Emissions has the highest differentiating influence between all criteria of the decision problem.
- On the contrary, criterion Primary energy import ratio has the lowest differentiating influence between all criteria of the decision problem.
- The highest expected overall performance score is attained by Alternative 5 (0.316).
- Alternative 5 does attain the best expected rank of 1.26 out of the 5 alternatives considered in this decision situation.
- Alternative 5 attains most often the first rank in the executed simulation runs. In 76.2 % of the simulation runs it becomes the preferred alternative.
- There is no alternative with strictly higher overall performance scores than any of the other alternatives.

6 Discussion and limitations of the natural language generation approach

In the case study, Alternative 5 can be considered as a recommendable choice since it achieves the highest EOPS of all alternatives. Additionally, it attains the first or second rank in 98 % of the simulation runs while it stochastically dominates Alternative 1 in 79.5 % of the simulation runs and all other alternatives in at least 97.3 %. As this alternative also minimises the downside risk of achieving low OPSs, it is also the preferred alternative for a risk-averse DM. If the goal of the DMs is to maximise the upside potential, Alternative 3 might become more preferable as it achieves the maximum OPS(s) in the decision situation. However, this represents a rather risky choice because it is stochastically dominated by Alternative 5 in 97,3 % of the simulation runs.

Overall, we received positive feedback from the interviewed experts that were involved in the development process of the explanations. They appreciated the usefulness of the implications for a detailed preferential sensitivity analysis. Besides, we could also show its benefits for novice users for complex interpretational tasks of the spread of results visualisation. Critically reflecting our approach, we still lack more detailed knowledge about the explanatory preferences of novice users though. This includes on the one hand their benefits regarding other visualisations of the multi-dimensional sensitivity analysis. On the other hand, a more general assessment on how novice users perceive and accept the explanations could further improve this concept in future.

Our survey results show that the explanations for the spread of results visualisation only reduce the complexity of cognitively demanding interpretational tasks. This needs to be further verified with a more representative sample of participants, as young people with high educational backgrounds were overrepresented in our evaluation study. Apart from this, we consider the spread of results visualisation as the cognitively easiest to understand within the preferential sensitivity report. We assume that due to the higher complexity of the other visualisations, such as the cumulative performance visualisation, we expect that our explanatory concept would show a considerably higher impact on their interpretation by novice users. We need to further validate this hypothesis, also in order to find out which explanations are really beneficial and which are less relevant for different groups of users. Further prioritisation of the explanations will also lead to a better interpretation by the users as we assume that they read shorter texts more carefully. In our survey we observed that the majority of the sample group that was provided with explanations did probably not use the information for the most difficult interpretational task. Focussing the explanations upon novice user needs can thus help reduce the effect that people do not notice or recall the explanations.

Since novice users were not included in the development phase of the explanations so far, we currently do not know how understandable and concise the generated explanations are for them. Another focus of research could thus be a more detailed assessment of this, for example in face-to-face settings aimed at revealing more detailed insights on the explanatory needs of different users.

7 Conclusions

Due to the complex and opaque character of DSSs in MCDA, explanatory functions that facilitate the analysis and interpretation of a decision situation within MCDA provide valuable benefits for DMs and DAs. One reason for this is the detailed modelling of preferential uncertainties of the decision situation and the comprehensive sensitivity analyses that these DSSs conduct. There are various DSSs existing which explain the obtained model results and partially also their sensitivities towards the variations of one weight parameter to the user. However, they do not provide explanations for the assessment of simultaneous variations of different preferential parameters.

The key contribution of this paper is a natural language generation approach to explain the influences of inter-criteria and/or intra-criteria preferential uncertainties on a decision situation. The aim of our NLG approach is to reduce the cognitive complexity to access, understand and interpret the influences of preferential uncertainties on a decision situation. The concept promotes deeper understanding for both expert and novice users as it provides relevant information on key implications that support their judgemental capabilities. The generated text provides complementary information for various sensitivity analysis visualisations. Both visualisations and text are provided in a multi-dimensional preferential sensitivity report which the developed tool can automatically generate for the user on the basis of the model results. This format furthermore increases the traceability and documentation of a decision-making process.

Beyond applying the presented NLG approach to an energy policy case study, we evaluated its benefits with both expert and novice users. The involved experts appreciated the level of detail of the multi-dimensional preferential sensitivity report. They considered it to be very useful in real decision-making situations as it conveniently explains implications related to the model results. The evaluation with novice users showed that our approach provides particularly beneficial support in cognitively demanding tasks that require a deep understanding and interpretational capabilities.

Future research activities in this area should include investigation and validation of the approach in a number of different decision situations. Moreover, a methodological extension including explanations for data uncertainties in MCDA is a relevant area for future research.

Acknowledgements

The authors would like to thank all involved external experts for their time and commitment in the interviews and for the written feedback which helped to improve the concept developed in this paper. This includes Lisa Scholten, Valentina Ferretti, Fridolin Haag, Ian Durbach, Johannes Siebert, Jyri Mustajoki, Mika Marttunen and Theodor Stewart.

References

Abu-Taha, R. (2011). Multi-criteria applications in renewable energy analysis: A literature review. paper presented at PICMET ‘11 Conference, July 31 - August 04, 2011, Portland. DOI: 10.1007/978-1-4471-5097-8_2.

Amgoud, L. and Prade, H. (2006). Explaining qualitative decision under uncertainty by argumentation. In Cohn, A. (Ed.), *Proceeding AAAI 2006 proceedings of the 21st national conference on Artificial intelligence, Boston, MA, July 16-20, 2006*, AAAI Press, Menlo Park, CA, pp. 219–224.

Bailey, D., Goonetilleke, A. and Deriche, M. (2011). A decision support system for site selection of large-scale infrastructure facilities using natural language. paper presented at PICMET ‘11 Conference, July 31 - August 04, 2011, Portland, available at: http://eprints.qut.edu.au/4223/ (accessed 12 January 2016).

Basson, L. and Petrie, J.G. (2007). An integrated approach for the consideration of uncertainty in decision making supported by Life Cycle Assessment. *Environmental Modelling & Software*, Vol. 22 No. 2, pp. 167–176. DOI: 10.1016/j.envsoft.2005.07.026.

Bélanger, M. and Martel, J.-M. (2005). An automated explanation approach for a decision support system based on MCDA. In Roth-Berghofer, T.R., Schulz, S. and Woody, A. (Eds.), *Explanation-aware computing: Papers from the AAAI Fall Symposium, Arlington, VA, November 04-06, 2005*, AAAI Press, Menlo Park, CA, pp. 21–34.

Bell, D.E. (1982). Regret in decision making under uncertainty. *Operations research*, Vol. 30 No. 5, pp. 961–981. DOI: 10.1287/opre.30.5.961.

Bell, M.L., Hobbs, B.F. and Ellis, H. (2003). The use of multi-criteria decision-making methods in the integrated assessment of climate change: Implications for IA practitioners. *Socio-Economic Planning Sciences*, Vol. 37 No. 4, pp. 289–316. DOI: 10.1016/S0038-0121(02)00047-2.

Belton, V. and Stewart, T.J. (2002), *Multiple Criteria Decision Analysis - An Integrated Approach*, Springer US, Boston.

Bertsch, V., Treitz, M., Geldermann, J. and Rentz, O. (2007). Sensitivity analyses in multi-attribute decision support for off-site nuclear emergency and recovery management. *International Journal of Energy Sector Management*, Vol. 1 No. 4, pp. 342-365.

Bertsch, V. (2008). Uncertainty Handling in Multi-Attribute Decision Support for Industrial Risk Management. University of Karlsruhe, Karlsruhe, 2008. DOI: 10.5445/KSP/1000007378.

Bertsch, V. and Fichtner, W. (2016). A participatory multi-criteria approach for power generation and transmission planning. *Annals of Operations Research*, Vol. 245 No. 1, pp. 177–207. DOI: 10.1007/s10479-015-1791-y.

Bishara, A.J. and Hittner, J.B. (2012). Testing the significance of a correlation with non-normal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological methods*, Vol. 17 No. 3, pp. 399–417. DOI: 10.1037/a0028087.

Bloom, B.S., Krathwohl, D.R. and Masia, B.B. (1984), *Taxonomy of educational objectives: The classification of educational goals Book 1: Cognitive Domain*, Longman, New York.

Brans, J.-P. and Mareschal, B. (1994). The PROMCALC & GAIA decision support system for multicriteria decision aid. *Decision Support Systems*, Vol. 12 No. 4-5, pp. 297–310. DOI: 10.1016/0167-9236(94)90048-5.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, Vol. 45 No. 1-3, pp. 223–241. DOI: 10.1016/0001-6918(80)90034-7.

Broekhuizen, H., Groothuis-Oudshoorn, C.G.M., van Til, J.A., Hummel, J.M. and Jzerman, M.J. (2015). A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. *PharmacoEconomics*, Vol. 33 No. 5, pp. 445–455. DOI: 10.1007/s40273-014-0251-x.

Browne, D., O'Regan, B. and Moles, R. (2010). Use of multi-criteria decision analysis to explore alternative domestic energy and electricity policy scenarios in an Irish city-region. *Energy*, Vol. 35 No. 2, pp. 518–528. DOI: 10.1016/j.energy.2009.10.020.

Buchanan, B.G. and Shortliffe, E.H. (1984), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project, The Addison-Wesley series in artificial intelligence*, Addison-Wesley, Reading, MA.

Butler, J., Jia, J. and Dyer, J. (1997). Simulation techniques for the sensitivity analysis of multi-criteria decision models. *European Journal of Operational Research*, Vol. 103 No. 3, pp. 531–546. DOI: 10.1016/S0377-2217(96)00307-4.

Carenini, G. and Moore, J.D. (2006). Generating and evaluating evaluative arguments. *Artificial Intelligence*, Vol. 170 No. 11, pp. 925–952. DOI: 10.1016/j.artint.2006.05.003.

Clark, A., Fox, C. and Lappin, S. (2010), *The handbook of computational linguistics and natural language processing, Blackwell handbooks in linguistics*, Wiley-Blackwell, Chichester, West Sussex, Malden, MA. DOI: 10.1002/9781444324044.

Cohen, J. (1988), *Statistical power analysis for the behavioral sciences,* 2nd ed., Lawrence Erlbaum Associates Inc., Hillsdale, N. J.

Corrente, S., French, S., Greco, S., Kadziński, M., Knowles, J.D., Mousseau, V., Siebert, J. and Słowiński, R. (2014). Drafting a Manifesto for DM-DSS Interaction (Working Group 'DM Sense'). In Greco, S., Knowles, J.D., Miettinen, K. and Zitzler, E. (Eds.), *Learning in Multiobjective Optimization: Report from Dagstuhl Seminar 2014*, *Dagstuhl Report*, pp. 72–81.

Dhaliwal, J.S. and Benbasat, I. (1996). The Use and Effects of Knowledge-Based System Explanations: Theoretical Foundations and a Framework for Empirical Evaluation. *Information Systems Research*, Vol. 7 No. 3, pp. 342–362. DOI: 10.1287/isre.7.3.342.

Diakoulaki, D., Antunes, C.H. and Martins, A.G. (2005). MCDA and energy planning. In Figueira, J.R., Greco, S. and Ehrgott, M. (Eds.), *Multiple criteria decision analysis: State of the art surveys*, *International Series in Operations Research & Management Science*, Springer, New York, pp. 859–890. DOI: 10.1007/0-387-23081-5_21.

Durbach, I.N. and Stewart, T.J. (2009). Using expected values to simplify decision making under uncertainty. *Omega*, Vol. 37 No. 2, pp. 312–330. DOI: 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*, Vol. 223 No. 1, pp. 1–14. DOI: 10.1016/j.ejor.2012.04.038.

Eisenführ, F., Weber, M. and Langer, T. (2010), *Rational Decision Making*, Springer, Berlin, London. DOI: 10.1007/978-3-642-02851-9.

Ellis, P.D. (2010). Effect sizes and the interpretation of research results in international business. *Journal of International Business Studies*, Vol. 41 No. 9, pp. 1581–1588. DOI: 10.1057/jibs.2010.39.

Ellis, P.D. (2013), *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results,* 6th ed., Cambridge University Press, Cambridge. DOI: 10.1017/CBO9780511761676.

Fisher, R.A. (1922). On the Interpretation of X^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society*, Vol. 85 No. 1, p. 87. DOI: 10.2307/2340521.

French, S. (1995). Uncertainty and Imprecision: Modelling and Analysis. *The Journal of the Operational Research Society*, Vol. 46 No. 1. DOI: 10.2307/2583837.

French, S., Maule, J. and Papamichail, K.N. (2009), *Decision behaviour, analysis and support*, Cambridge University Press, Cambridge, UK, New York.

Gardiner, P.C. and Edwards, W. (1975). Public values: Multiattribute-utility measurement for social decision making. In Kaplan, M.F. and Schwartz, S. (Eds.), *Human Judgment and Decision Process*, *Academic Press Series in cognition and perception*, Academic Press, New York, pp. 1–37. DOI: 10.1109/TSMC.1977.4309720.

Geldermann, J. (2010). Explanation Systems. In Ríos Insua, D. and French, S. (Eds.), *E-Democracy: A group decision and negotiation practice*, *Advances in Group Decision and Negotiation*, Vol. 5, Springer, Dordrecht, London, pp. 241–259.

Geldermann, J., Bertsch, V., Treitz, M., French, S., Papamichail, K.N. and Hämäläinen, R.P. (2009). Multi-criteria decision support and evaluation of strategies for nuclear remediation management. *Omega*, Vol. 37 No. 1, pp. 238–251. DOI: 10.1016/j.omega.2006.11.006.

Gilovich, T., Griffin, D.W. and Kahneman, D. (2002), *Heuristics and biases: The psychology of intuitive judgement*, Cambridge University Press, Cambridge, UK, New York.

Gosset, W.S. (1908). The probable error of a mean. *Biometrika*, Vol. 6 No. 1, pp. 1–25. DOI: 10.1093/biomet/6.1.1.

Graves, S.B. and Ringuest, J.L. (2009). Probabilistic dominance criteria for comparing uncertain alternatives: A tutorial. *Omega*, Vol. 37 No. 2, pp. 346–357. DOI: 10.1016/j.omega.2007.03.001.

Greco, S., Słowiński, R. and Zielniewicz, P. (2013). Putting Dominance-based Rough Set Approach and robust ordinal regression together. *Decision Support Systems*, Vol. 54 No. 2, pp. 891–903. DOI: 10.1016/j.dss.2012.09.013.

Greef, H.P.d. and Neerincx, M.A. (1995). Cognitive support: Designing aiding to supplement human knowledge. *International Journal of Human-Computer Studies*, Vol. 42 No. 5, pp. 531–571. DOI: 10.1006/ijhc.1995.1023.

Greening, L.A. and Bernow, S. (2004). Design of coordinated energy and environmental policies: Use of multi-criteria decision-making. *Energy Policy*, Vol. 32 No. 6, pp. 721–735. DOI: 10.1016/j.enpol.2003.08.017.

Greer, J.E., Falk, S., Greer, K.J. and Bentham, M.J. (1994). Explaining and justifying recommendations in an agriculture decision support system. *Computers and Electronics in Agriculture*, Vol. 11 No. 2-3, pp. 195–214. DOI: 10.1016/0168-1699(94)90008-6.

Gregor, S. and Benbasat, I. (1999). Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. *MIS Quarterly*, Vol. 23 No. 4, pp. 497–530. DOI: 10.2307/249487.

Hämäläinen, R.P. and Alaja, S. (2008). The threat of weighting biases in environmental decision analysis. *Ecological Economics*, Vol. 68 No. 1-2, pp. 556–569. DOI: 10.1016/j.ecolecon.2008.05.025.

Hammond, K.R., Stewart, T.R., Brehmer, B. and Steinmann, D.O. (1975). Social Judgment Theory. In Kaplan, M.F. and Schwartz, S. (Eds.), *Human Judgment and Decision Process*, *Academic Press Series in cognition and perception*, Academic Press, New York.

Hauke, J. and Kossowski, T. (2011). Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaestiones Geographicae*, Vol. 30 No. 2, pp. 87–93. DOI: 10.2478/v10117-011-0021-1.

Henrion, M. and Druzdzel, M.J. (1991). Qualitative Propagation and Scenario-based Explanation of Probabilistic Reasoning. In Bonissone, P.P., Henrion, M., Kanal, L.N. and Lenner, J.F. (Eds.), *Uncertainty and artificial intelligence*, 6th ed., pp. 17–32.

Heo, E., Kim, J. and Boo, K.-J. (2010). Analysis of the assessment factors for renewable energy dissemination program evaluation using fuzzy AHP. *Renewable and Sustainable Energy Reviews*, Vol. 14 No. 8, pp. 2214–2220. DOI: 10.1016/j.rser.2010.01.020.

Hodgkin, J., Belton, V. and Koulouri, A. (2005). Supporting the intelligent MCDA user: A case study in multi-person multi-criteria decision support. *European Journal of Operational Research*, Vol. 160 No. 1, pp. 172–189. DOI: 10.1016/j.ejor.2004.03.007.

Hoffman, P.J., Earle, T.C. and Slovic, P. (1981). Multidimensional functional learning (MFL) and some new conceptions of feedback. *Organizational Behavior and Human Performance*, Vol. 27 No. 1, pp. 75–102. DOI: 10.1016/0030-5073(81)90040-4.

Hogarth, R.M. (1987), *Judgement and choice: The psychology of decision,* 2nd ed., Wiley, Chichester, New York.

Holtzman, S. (1988), *Intelligent decision systems*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA.

Jessop, A. (2011). Using imprecise estimates for weights. *Journal of the Operational Research Society*, Vol. 62 No. 6, pp. 1048–1055.

Jessop, A. (2014). IMP: A decision aid for multiattribute evaluation using imprecise weight estimates. *Omega*, Vol. 49, pp. 18–29. DOI: 10.1016/j.omega.2014.05.001.

Jiménez, A., Mateos, A. and Ríos-Insua, S. (2005). Monte Carlo Simulation Techniques in a Decision Support System for Group Decision Making. *Group Decision and Negotiation*, Vol. 14 No. 2, pp. 109–130. DOI: 10.1007/s10726-005-2406-9.

Kadziński, M., Corrente, S., Greco, S. and Słowiński, R. (2014). Preferential reducts and constructs in robust multiple criteria ranking and sorting. *OR Spectrum*, Vol. 36 No. 4, pp. 1021–1053. DOI: 10.1007/s00291-014-0361-z.

Kahneman, D. and Knetsch, J.L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management*, Vol. 22 No. 1, pp. 57–70. DOI: 10.1016/0095-0696(92)90019-S.

Kahneman, D., Slovic, P. and Tversky, A. (1982), *Judgment under uncertainty: Heuristics and biases*, Cambridge University Press, Cambridge, UK, New York.

Kass, R. and Finin, T. (1988). The Need for User Models in Generating Expert System Explanations. *International Journal of Expert Systems - Special Issue: Natural Language and Expert Systems*, Vol. 1 No. 4, pp. 345–375.

Kaya, T. and Kahraman, C. (2011). Multicriteria decision making in energy planning using a modified fuzzy TOPSIS methodology. *Expert Systems with Applications*, Vol. 38 No. 6, pp. 6577–6585. DOI: 10.1016/j.eswa.2010.11.081.

Keeney, R.L. and Raiffa, H. (1976), *Decisions with multiple objectives: Preferences and value tradeoffs*, Wiley, New York.

Kiker, G.A., Bridges, T.S., Varghese, A., Seager, T.P. and Linkov, I. (2005). Application of Multicriteria Decision Analysis in Environmental Decision Making. *Integrated Environmental Assessment and Management*, Vol. 1 No. 2, pp. 95–108. DOI: 10.1897/IEAM_2004a-015.1.

Kirkwood, C.W. (1997), *Strategic decision making: Multiobjective decision analysis with spreadsheets*, Duxbury Press, Belmont.

Kowalski, K., Stagl, S., Madlener, R. and Omann, I. (2009). Sustainable energy futures: Methodological challenges in combining scenarios and participatory multi-criteria analysis. *European Journal of Operational Research*, Vol. 197 No. 3, pp. 1063–1074. DOI: 10.1016/j.ejor.2007.12.049.

Labreuche, C., Maudet, N., Mousseau, V. and Ouerdane, W. (2012). Explanation of the robust additive preference model by even swap sequences. paper presented at 6th Multidisciplinary Workshop on Advances in Preference Handling, Montpellier.

Labreuche, C., Maudet, N. and Ouerdane, W. (2011). Minimal and Complete Explanations for Critical Multi-attribute Decisions. In Brafman, R.I., Roberts, F.S. and Tsoukiàs, A. (Eds.), *Algorithmic decision theory, Piscataway, NJ, October 26-28, 2011*, Springer, Berlin, Heidelberg, New York, pp. 121–134.

Lahdelma, R., Hokkanen, J. and Salminen, P. (1998). SMAA - Stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, Vol. 106 No. 1, pp. 137–143. DOI: 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*, Vol. 49 No. 3, pp. 444–454.

Levy, H. (1992). Stochastic Dominance and Expected Utility: Survey and Analysis. *Management Science*, Vol. 38 No. 4, pp. 555–593.

Linkov, I., Varghese, A. and Jamil, S. (2004). Multi-criteria decision analysis: A framework for structuring remedial decisions at contaminated sites. In Linkov, I. and Ramadan, A.B. (Eds.), *Comparative risk assessment and environmental decision making*, Kluwer Academic Publishers, Boston, MA, pp. 15–54. DOI: 10.1007/1-4020-2243-3_2.

Loken, E. (2007). Use of multicriteria decision analysis methods for energy planning problems. *Renewable and Sustainable Energy Reviews*, Vol. 11 No. 7, pp. 1584–1595. DOI: 10.1016/j.rser.2005.11.005.

Loomes, G. and Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. *Economic Journal*, Vol. 92 No. 4, pp. 805–824. DOI: 10.2307/2232669.

Lühn, T., Schlömer, G., Schmidtmann, G., Lehde, B., Schmiesing, J., Hofmann, L. and Geldermann, J. (2014). Multi-Criteria Analysis of Grid Expansion Concepts on the Low Voltage Level. *Zeitschrift für Energiewirtschaft*, Vol. 38 No. 3, pp. 183–200. DOI: 10.1007/s12398-014-0134-z.

Mao, J.-Y. and Benbasat, I. (2000). The Use of Explanations in Knowledge-Based Systems: Cognitive Perspectives and a Process-Tracing Analysis. *Journal of Management Information Systems*, Vol. 17 No. 2, pp. 153–179. DOI: 10.1080/07421222.2000.11045646.

Mareschal, B. and Brans, J.-P. (1988). Geometrical representations for MCDA. *European Journal of Operational Research*, Vol. 34 No. 1, pp. 69–77. DOI: 10.1016/S0377-2217(00)00038-2.

Mateos, A., Jiménez, A. and Ríos-Insua, S. (2006). Monte Carlo simulation techniques for group decision making with incomplete information. *European Journal of Operational Research*, Vol. 3 No. 174, pp. 1842–1864. DOI: 10.1016/j.ejor.2005.02.057.

Matsatsinis, N.F. and Samaras, A.P. (2001). MCDA and preference disaggregation in group decision support systems. *European Journal of Operational Research*, Vol. 130 No. 2, pp. 414–429.

Mavrotas, G. and Trifillis, P. (2006). Multicriteria decision analysis with minimum information: Combining DEA with MAVT. *Computers & Operations Research*, Vol. 33 No. 8, pp. 2083–2098. DOI: 10.1016/j.cor.2004.11.023.

Morgan, M.G., Henrion, M. and Small, M. (1990), *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*, Cambridge University Press, Cambridge UK, New York.

Morton, A. and Fasolo, B. (2009). Behavioural decision theory for multi-criteria decision analysis: A guided tour. *Journal of the Operational Research Society*, Vol. 60 No. 2, pp. 268–275. DOI: 10.1057/palgrave.jors.2602550.

Mustajoki, J., Hämäläinen, R.P. and Salo, A. (2005). Decision Support by Interval SMART/SWING - Incorporating Imprecision in the SMART and SWING Methods. *Decision Sciences*, Vol. 36 No. 2, pp. 317–339. DOI: 10.1111/j.1540-5414.2005.00075.x.

Nunes, I., Miles, S., Luck, M. and Lucena, C. (2012). Investigating Explanations to Justify Choice. In Masthoff, J., Mobasher, B., Desmarais, M.C. and Nkambou, R. (Eds.), *User modeling, adaptation, and personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20, 2012. Proceedings*, *Information Systems and Applications, incl. Internet/Web, and HCI*, Vol. 7379, Springer, Berlin, Heidelberg, pp. 212–224. DOI: 10.1007/978-3-642-31454-4_18.

Ouerdane, W., Maudet, N. and Tsoukiàs, A. (2010). Argumentation Theory and Decision Aiding. In Ehrgott, M., Figueira, J.R. and Greco, S. (Eds.), *Trends in Multiple Criteria Decision Analysis*, *International Series in Operations Research & Management Science*, Vol. 142, Springer US, pp. 177–208. DOI: 10.1007/978-1-4419-5904-1_7.

Pagano, R.R. (2013), *Understanding statistics in the behavioral sciences,* 10th ed., Wadsworth Publishing, Belmont, CA.

Papamichail, K.N. and French, S. (2000). Decision support in nuclear emergencies. *Journal of Hazardous Materials*, Vol. 71 No. 1-3, pp. 321–342. DOI: 10.1016/S0304-3894(99)00086-2.

Papamichail, K.N. and French, S. (2003). Explaining and justifying the advice of a decision support system - A natural language generation approach. *Expert Systems with Applications*, Vol. 24 No. 1, pp. 35–48. DOI: 10.1016/S0957-4174(02)00081-7.

Papamichail, K.N. and French, S. (2005). Design and evaluation of an intelligent decision support system for nuclear emergencies. *Decision Support Systems*, Vol. 41 No. 1, pp. 84–111. DOI: 10.1016/j.dss.2004.04.014.

Parikh, M., Fazlollahi, B. and Verma, S. (2001). The Effectiveness of Decisional Guidance: An Empirical Evaluation. *Decision Sciences*, Vol. 32 No. 2, pp. 303–332. DOI: 10.1111/j.1540-5915.2001.tb00962.x.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, Vol. 50 No. 302, pp. 157–175. DOI: 10.1080/14786440009463897.

Pohekar, S.D. and Ramachandran, M. (2004). Application of multi-criteria decision making to sustainable energy planning - A review. *Renewable and Sustainable Energy Reviews*, Vol. 8 No. 4, pp. 365–381. DOI: 10.1016/j.rser.2003.12.007.

Reiter, E. and Dale, R. (1997). Building applied natural language generation systems. *Natural Language Engineering*, Vol. 3 No. 1, pp. 57–87. DOI: 10.1017/S1351324997001502.

Ren, J., Fedele, A., Mason, M., Manzardo, A. and Scipioni, A. (2013). Fuzzy Multi-actor Multi-criteria Decision Making for sustainability assessment of biomass-based technologies for hydrogen production. *International Journal of Hydrogen Energy*, Vol. 38 No. 22, pp. 9111–9120. DOI: 10.1016/j.ijhydene.2013.05.074.

Ribeiro, F., Ferreira, P. and Araújo, M. (2013). Evaluating future scenarios for the power generation sector using a Multi-Criteria Decision Analysis (MCDA) tool: The Portuguese case. *Energy*, Vol. 52, pp. 126–136. DOI: 10.1016/j.energy.2012.12.036.

Ríos Insua, D. and French, S. (1991). A framework for sensitivity analysis in discrete multi-objective decision-making. *European Journal of Operational Research*, Vol. 54 No. 2, pp. 176–190. DOI: 10.1016/0377-2217(91)90296-8.

Sánchez-Hernández, G. (2013). A contribution to the ranking and description of classifications. Dissertation, Instituto de Organización y Control de Sistemas Industriales, Universitat politécnica de Catalunya, Barcelona, 2013.

Scholten, L., Scheidegger, A., Reichert, P., Mauer, M. and Lienert, J. (2014). Strategic rehabilitation planning of piped water networks using multi-criteria decision analysis. *Water research*, Vol. 49, pp. 124–143. DOI: 10.1016/j.watres.2013.11.017.

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*, Vol. 242 No. 1, pp. 243–260. DOI: 10.1016/j.ejor.2014.09.044.

Scott, J.A., Ho, W. and Dey, P.K. (2012). A review of multi-criteria decision-making methods for bioenergy systems. *Energy*, Vol. 42 No. 1, pp. 146–156. DOI: 10.1016/j.energy.2012.03.074.

Silver, M.S. (1991a). Decisional Guidance for Computer-Based Decision Support. *MIS Quarterly*, Vol. 15 No. 1, pp. 105–122. DOI: 10.2307/249441.

Silver, M.S. (1991b), *Systems that support decision makers: Description and analysis, John Wiley information systems series*, Wiley, Chichester, New York.

Spiegelhalter, D.J. and Knill-Jones, R.P. (1984). Statistical and Knowledge-Based Approaches to Clinical Decision-Support Systems, with an Application in Gastroenterology. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 147 No. 1, pp. 35–77. DOI: 10.2307/2981737.

Stewart, T.J. (1992). A critical survey on the status of multiple criteria decision making theory and practice. *Omega*, Vol. 20 No. 5-6, pp. 569–586. DOI: 10.1016/0305-0483(92)90003-P.

Stewart, T.J. (2005). Dealing with Uncertainties in MCDA. In Figueira, J.R., Greco, S. and Ehrgott, M. (Eds.), *Multiple criteria decision analysis: State of the art surveys*, *International Series in Operations Research & Management Science*, Vol. 78, Springer, New York, pp. 445–466. DOI: 10.1007/0-387-23081-5_11.

Streimikiene, D., Balezentis, T., Krisciukaitienė, I. and Balezentis, A. (2012). Prioritizing sustainable electricity production technologies: MCDM approach. *Renewable and Sustainable Energy Reviews*, Vol. 16 No. 5, pp. 3302–3311. DOI: 10.1016/j.rser.2012.02.067.

Swartout, W.R. and Moore, J.D. (1993). Explanation in Second Generation Expert Systems. In David, J.-M., Krivine, J.-P. and Simmons, R. (Eds.), *Second generation expert systems*, Springer Berlin Heidelberg, Berlin, pp. 543–585. DOI: 10.1007/978-3-642-77927-5_26.

Tervonen, T. (2014). JSMAA: Open source software for SMAA computations. *International Journal of Systems Science*, Vol. 45 No. 1, pp. 69–81. DOI: 10.1080/00207721.2012.659706.

Tervonen, T. and Figueira, J.R. (2008). A Survey on Stochastic Multicriteria Acceptability Analysis Methods. *Journal of multi-criteria decision analysis*, Vol. 15 No. 1-2, pp. 1–14. DOI: 10.1002/mcda.407.

Tintarev, N. and Masthoff, J. (2007). A Survey of Explanations in Recommender Systems. paper presented at IEEE 23rd International Conference on Data Engineering Workshop, April 15-20, 2007, Istanbul. DOI: 10.1109/ICDEW.2007.4401070.

Treitz, M. (2006), *Production process design using multi-criteria analysis*, Universitätsverlag Karlsruhe, Karlsruhe.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, Vol. 79 No. 4, pp. 281–299. DOI: 10.1037/h0032955.

Walker, M.A., Whittaker, S.J., Stent, A., Maloor, P., Moore, J.D., Johnston, M. and Vasireddy, G. (2004). Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science*, Vol. 28 No. 5, pp. 811–840. DOI: 10.1207/s15516709cog2805_8.

Wang, J.-J., Jing, Y.-Y., Zhang, C.-F. and Zhao, J.-H. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews*, Vol. 13 No. 9, pp. 2263–2278. DOI: 10.1016/j.rser.2009.06.021.

Weiner, J.L. (1980). BLAH, a system which explains its reasoning. *Artificial Intelligence*, Vol. 15 No. 1-2, pp. 19–48. DOI: 10.1016/0004-3702(80)90021-1.

Wybo, J.-L. (2006). Editorial. *International Journal of Emergency Management*, Vol. 3 No. 2/3, pp. 99-100.

Ye, R.L. (1995). The value of explanation in expert systems for auditing: An experimental investigation. *Expert Systems with Applications*, Vol. 9 No. 4, pp. 543–556. DOI: 10.1016/0957-4174(95)00023-2.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, Vol. 8 No. 3, pp. 338–353. DOI: 10.1016/S0019-9958(65)90241-X.

Zhou, P., Ang, B.W. and Poh, K.L. (2006). Decision analysis in energy and environmental modeling: An update. *Energy*, Vol. 31 No. 14, pp. 2604–2622. DOI: 10.1016/j.energy.2005.10.023.

Zimmermann, H.-J. (2000). An application-oriented view of modeling uncertainty. *European Journal of Operational Research*, Vol. 122 No. 2, pp. 190–198. DOI: 10.1016/S0377-2217(99)00228-3.

A. Appendix: Multi-dimensional preferential sensitivity report of the evaluation of energy strategies

SIMADA - Simulation-Based Multi-Attribute Decision Analysis
(c) Copyright IIP, Karlsruhe Institute of Technology (KIT)

Multi-dimensional sensitivity report

This report summarises the multi-dimensional sensitivity analysis regarding the criteria levels for the alternatives on them. on the model results regarding the preferential uncertainties provided by SIMADA.

Date: 26-Oct-2016 16:16:50
Number of alternatives: 5
Number of simulation runs: 1000
Considered alternative: Alternative 5
Uncertainties considered: Weight variations, Value function curvature variations, Variations of lower boundaries of value functions, Variations of upper boundaries of value functions

Table of contents

Scientific terms are underlined by a dotted line. Further explanations are shown if the mouse cursor is moved on the underlined terms.

Introduction

This analysis examines the robustness of the simulation results of a decision situation for Alternative 5 with respect to the influences of preferential uncertainties. The decision situation is modelled by Multi-Attribute Value Theory (MAVT) with an additive value function where the underlying preferential uncertainties are expressed by assigned parameter intervals. These include intra-criteria preferential uncertainties (regarding the criteria level ranges of the alternatives) as well as inter-criteria preferential uncertainties (regarding the relative importance of the criteria). The inter-criteria preferential uncertainties are modelled by the variation of the value function shapes (curvature as well as upper and lower boundary variations) while the inter-criteria preferential uncertainties are represented by weight variations.

The aim of this analysis is to identify the most relevant preferential uncertainties in order to explore their respective impact on the results and to examine how the alternatives are distinguishable from each other in the light of these uncertainties. Therefore, this report presents the results of the analysis of 1000 simulation runs of the decision situation for which 5 different alternatives are considered. In each of the simulation runs, the uncertain parameter samples are varied randomly with respect to the assigned interval boundaries.

This report assesses the influences of the following preferential uncertainties on the simulation result:

Inter-criteria preferential uncertainties

- Variations of the weights

Intra-criteria preferential uncertainties

- Variations of the value function curvatures
- Variations of the lower boundaries of the value functions
- Variations of the upper boundaries of the value functions

Top

Executive summary

Overall, the following key aspects are characteristic for this decision situation and the executed simulation runs:

- The criterion CO2 Emissions has the highest differentiating influence between all criteria of the decision problem.
- On the contrary, criterion Primary energy import ratio has the lowest differentiating influence between all criteria of the decision problem.
- The highest expected overall performance score is attained by Alternative 5 (0.316).
- Alternative 5 does attain the best expected rank of 1.26 out of the 5 alternatives considered in this decision situation.
- Alternative 5 attains most often the first rank in the executed simulation runs. In 76.2 % of the simulation runs it becomes the preferred alternative.

- There is no alternative with strictly higher overall performance scores than any of the other alternatives.

Top

Preferential sensitivities regarding the overall performance scores

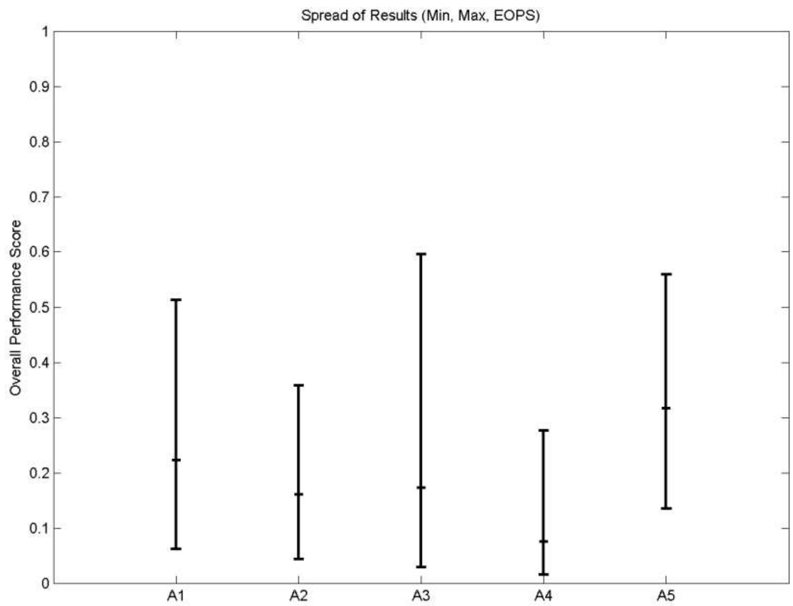
This paragraph compares the overall performance scores of the considered alternatives. The following observations are of general character, since the compared scores do not necessarily belong to the same simulation runs. The following aspects are examined:

a. Spread of overall performance scores
b. Cumulative performance

Top

Spread of overall performance scores

The 'spread of results' graph shows the ranges of the overall performance scores. The vertical lines with tick marks at their ends represent the minimum and maximum results obtained from the simulation runs. The tick mark in their middle indicates the expected overall performance score for the respective alternative (e.g. the value represents the average overall performance score in case of symmetric distributions).
Please note: This visualisation does not show the distribution of the overall performance scores for the alternatives.



Detailed results:

	Minimal overall performance scores	Expected overall performance score	Maximal overall performance scores
Alternative 1	0.061	0.223	0.513
Alternative 2	0.043	0.160	0.358
Alternative 3	0.029	0.172	0.596
Alternative 4	0.016	0.075	0.276
Alternative 5	0.135	0.316	0.559

Dominance of alternatives

In the following, all alternatives which have a dominance relation between each other are enumerated. This means that an Alternative A achieves strictly better overall performance scores than an Alternative B in all simulation runs.

- There is no alternative with strictly higher overall performance scores than any of the other alternatives in the executed simulation runs.

Indistinguishability of alternatives

The following alternatives cannot be distinguished from the visualisation only since their overall performance score value ranges do overlap:

- Alternative 1 is indistinguishable from all other alternatives.
- Alternative 2 is indistinguishable from all other alternatives.
- Alternative 3 is indistinguishable from all other alternatives.
- Alternative 4 is indistinguishable from all other alternatives.
- Alternative 5 is indistinguishable from all other alternatives.

Comparison of expected overall performance scores

- The expected overall performance score of Alternative 5 is higher than the maximum overall performance score of Alternative 4.

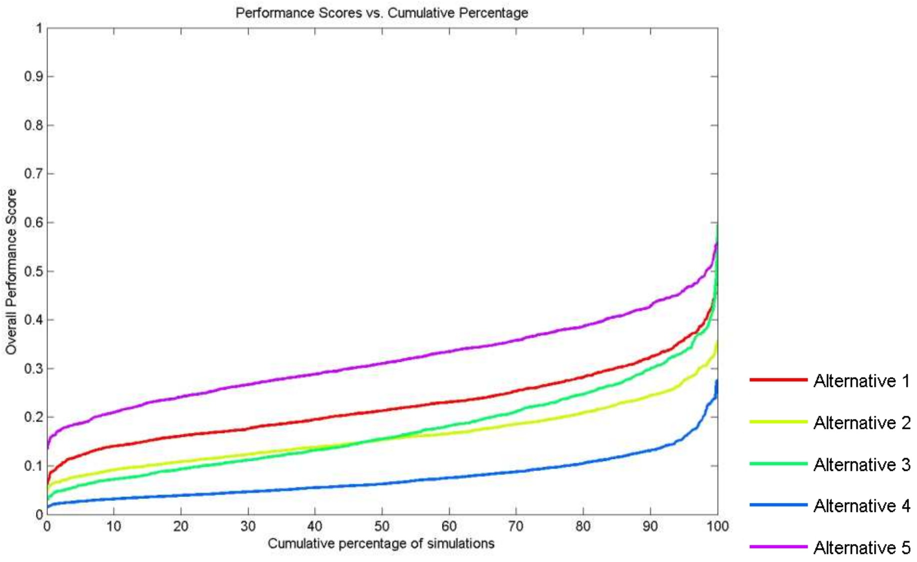
- The highest overall performance score is attained by Alternative 3 (0.596). It maximises the upside potential of realizing the highest possible overall performance score.
- The lowest overall performance score is attained by Alternative 4 (0.015).
- Alternative 5 attains the highest minimum of all alternatives (0.135). This alternative

maximises the minimal overall performance scores of all alternatives.

Top

Cumulative performance

The 'cumulative performance figure' shows the cumulative percentage (referring to 1000 simulation runs) and the respective overall performance score obtained by the alternatives. At any point on the graph, the cumulative percentage indicates the probability of an alternative to have an overall performance score equal or lower than at this point.
For every alternative the respective overall performance scores are sorted in ascending order individually so that the scores of different alternatives at one point do not necessarily belong to the same simulation run. This is why no implication on the ranking performance of the alternatives can be drawn.



Detailed results:

	Standard deviation	5 %-quantile	Inter-quantile range	95 %-quantile

Alternative 1	0.073	0.120	0.240	0.360
Alternative 2	0.059	0.079	0.198	0.277
Alternative 3	0.088	0.060	0.276	0.336
Alternative 4	0.044	0.027	0.133	0.160
Alternative 5	0.082	0.187	0.275	0.462

Dispersion of simulation results

The overall performance scores of Alternative 5 are neither more nor less dispersed with regard to the dispersion of the overall performance scores of all other alternatives. The standard deviation of 0.082 is a measure to express the dispersion of the overall performance scores of Alternative 5 around its expected overall performance score of 0.316.

Comparing the dispersion of the overall performance scores of Alternative 5 to the other alternatives individually, the following observations can be concluded:

- The overall performance scores of Alternative 4 are strongly less dispersed.
- The overall performance scores of Alternative 3 are more dispersed.
- The overall performance scores of Alternatives 1 and 2 are less dispersed.

Observations regarding the most probable overall performance scores

The 90 % most probable overall performance scores of Alternative 5 in the executed simulation runs are between 0.187 and 0.462.

- The highest 95 % overall performance scores of Alternative 4 dominate the 95 % highest overall performance scores of Alternative 5.
- The expected overall performance score of Alternative 5 is higher than 95 % of the overall performance scores achieved by Alternative 2.

Top

Preferential sensitivities regarding the ranking performances

This paragraph analyses the ranking performances of the considered alternatives. Therefore, the overall performance scores of the alternatives are compared to each other within the corresponding simulation runs. The following aspects are examined:

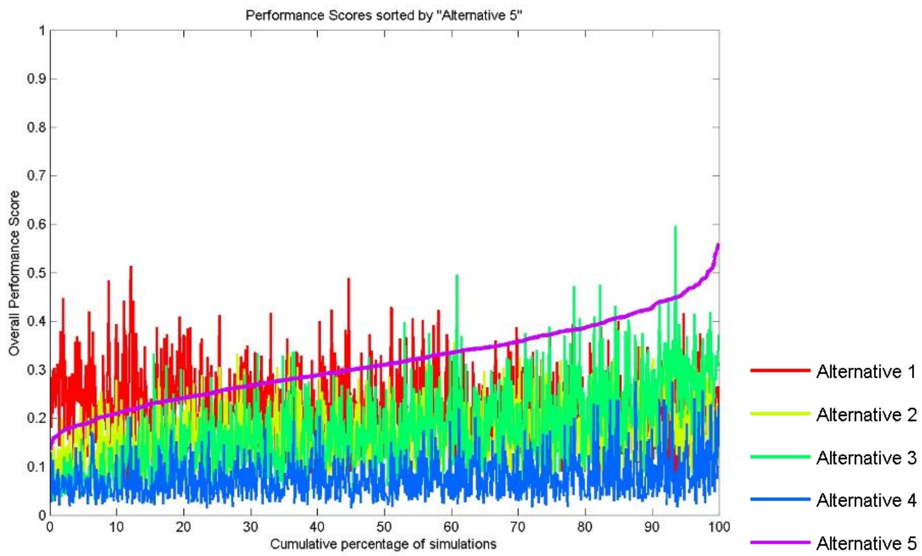
a. Cumulative performance sorted by Alternative 5
b. Detailed info on ranking performance of Alternative 5

c. Stochastic dominance

Top

Cumulative performance sorted by Alternative 5

The 'cumulative performance sorted by Alternative 5' visualisation shows the cumulative percentage (referring to 1000 simulation runs) and the respective overall performance score obtained by Alternative 5. At any point on the curve of Alternative 5 (represented by the continuous line), the cumulative percentage indicates the probability for this alternative to reach an overall performance score equal or lower than at this point. The overall performance scores of the other alternatives at a specific point belong to the same simulation run as the one of Alternative 5. This way, the value of the overall performance scores also determines the ranking of the alternatives in a particular simulation run.



Rank performance of the alternatives

The following table shows the percentage of the 1000 simulation runs that an alternative attained a certain rank:

	#1 rank	#2 rank	#3 rank	#4 rank	#5 rank
Alternative 1	20.3 %	36.4 %	28.3 %	14.3 %	0.7 %
Alternative 2	1 %	16.2 %	39.1 %	36.3 %	7.4 %
Alternative 3	2.5 %	25.6 %	28.4 %	36.6 %	6.9 %
Alternative 4	0 %	0 %	2.2 %	12.8 %	85 %
Alternative 5	76.2 %	21.8 %	2 %	0 %	0 %

- The highest percentage of No. 1 ranks (76.2 %) is attained by Alternative 5.
- The lowest percentage of No. 1 ranks (0 %) is attained by Alternative 4.
- The lowest percentage of last ranks (0 %) is attained by Alternative 5.
- The highest percentage of last ranks (85 %) is attained by Alternative 4.
- Alternative 4 does never become the preferred alternative.
- Alternative 5 does never attain the last rank.

First ranked overall performance scores of Alternative 5

In the simulation runs where Alternative 5 becomes the preferred alternative, the following implications on its obtained overall performance scores can be made:

- The overall performance scores that Alternative 5 attains in the simulation runs where it ranks first range from 0.169 to 0.559.
- The expected overall performance score of the simulation runs where Alternative 5 becomes the preferred alternative accumulates to 0.339.
- This value is 0.016 % higher compared to the expected overall performance score of all first ranking overall performance scores. This means, that in the 76.2 % of the simulation runs in which Alternative 5 ranks best, it also attains very high overall performance scores.

Expected rank performance of the alternatives

The following table shows the expected ranks attained by the alternatives in all 1000 simulation runs:

	Expected rank performance (out of 5 alternatives)
Alternative 1	2.387
Alternative 2	3.329
Alternative 3	3.198
Alternative 4	4.828
Alternative 5	1.258

- Alternative 5 does attain the best expected rank of 1.26 out of 5 alternatives.
- Alternative 4 does attain the worst expected rank of 4.83 out of 5 alternatives.

Risk characteristics of the alternatives

- Alternative 5 does achieve the highest expected overall performance score. This alternative minimises the downside risk of obtaining a low overall performance score with regard to the best alternative of every single simulation run.

It can be expected that Alternative 5 does attain an overall performance score that is 0.017 lower than the best overall performance score of a simulation run.

- On the other hand, Alternative 4 does achieve the lowest expected overall performance score. This alternative is characterised by the highest downside risk of obtaining a low overall performance score with regard to the best alternative of every single simulation run.
  It can be expected that Alternative 4 does attain an overall performance score that is 0.258 lower than the best overall performance score of a simulation run.

Correlation of the alternatives' overall performance

- Alternative 5 correlates negatively with Alternative 1 to a medium extent. This correlation is significant ($p = 0.000$).
- Alternative 5 correlates positively with Alternative 2 to a small extent. This correlation is significant ($p = 0.000$).
- Alternative 5 correlates positively with Alternative 3 to a large extent. This correlation is significant ($p = 0.000$).
- Alternative 5 correlates positively with Alternative 4 to a small extent. This correlation is significant ($p = 0.000$).

Influences of different types of preferential uncertainties

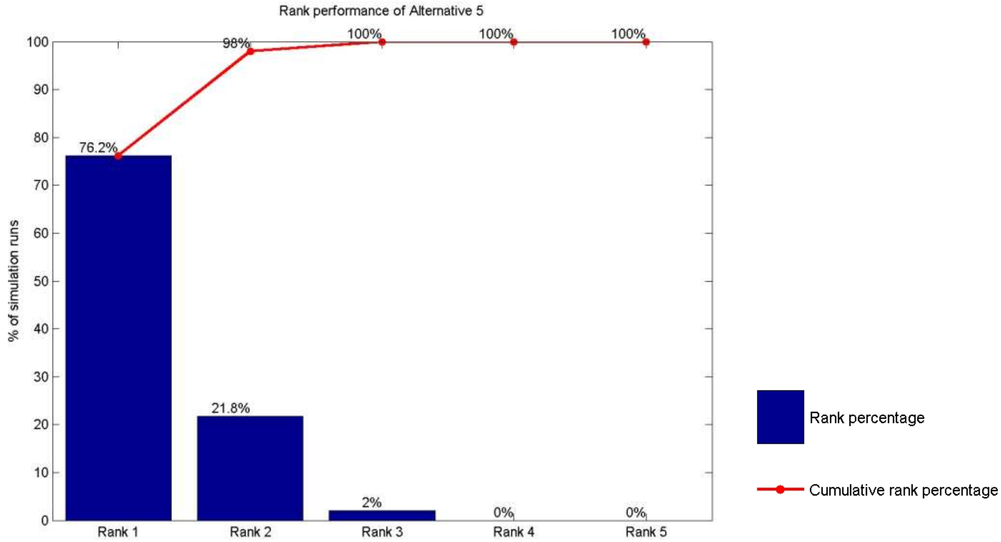
The overall performance score of Alternative 5 is most sensitive to a combination of inter- and intra-criteria preferential uncertainties. It is recommended to focus the initial discussion on this type of uncertainties.

- With inter- as well as intra-criteria preferential uncertainties taken into account Alternative 5 attains in 76.2 % of the cases rank 1.
- Considering only inter-criteria preferential uncertainties for this alternative, it ranks first in 91.8 % of the simulation runs.
- When only the intra-criteria preferential are modelled, Alternative 5 achieves in 95.1 % of the simulation runs the first rank.
- In the deterministic simulation (without the consideration of any preferential uncertainties), Alternative 5 is ranked 1 out of 5 alternatives.

Top

Detailed analysis of ranking performance for Alternative 5

The following bar graph shows the percentage of simulation runs in which Alternative 5 attains a certain rank. The red line represents the cumulative percentage of these ranking performances, it indicates the percentage in which Alternative 5 has attained a respective rank or better.

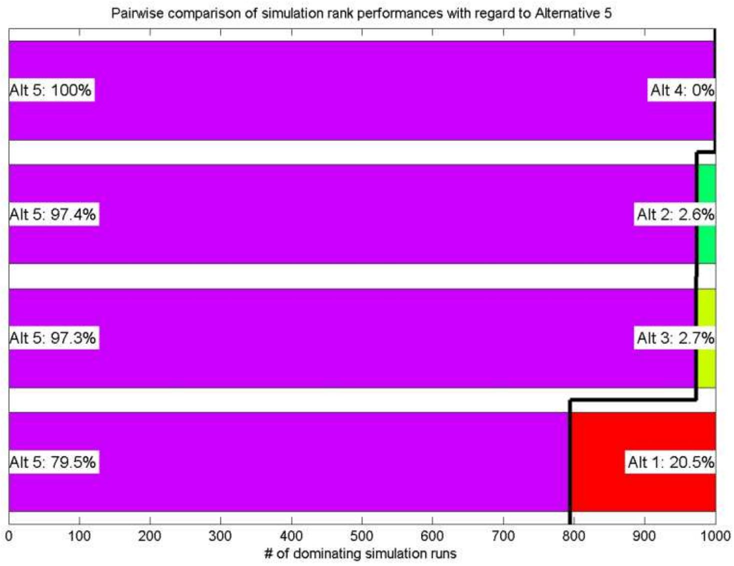


- Alternative 5 attains rank 1 in 76.2 % of the simulation runs.
- Alternative 5 ranks last in 0 % of the simulation runs. Alternative 5 never ranks worse than rank 3.
- In the majority of the simulation runs (76.2 %) Alternative 5 ranks 1.
- In 90 % of the simulation runs Alternative 5 attains rank 2 or better.
- Alternative 5 attains in 2 % of the simulation runs rank 3, overall this rank or better is achieved in 100 % of the cases.

Top

Stochastic dominance

The following graph shows a comparison of the overall performance scores of Alternative 5 with every other alternative individually. Each horizontal bar shows the number of simulation runs in which Alternative 5 outperforms the respective compared alternative. On the left, the number of simulation runs where Alternative 5 has a higher overall performance score than the compared alternative is shown. On the right the number of simulation runs in which the compared alternative dominates Alternative 5 can be seen. For each alternative comparison the relative percentage of outranking performances is indicated on the respective side of the horizontal bar.



- Alternative 5 dominates Alternative 4 in 100 % of the simulation runs. Also the expected overall performance score of Alternative 5 (0.316) is higher than the corresponding 0.0749 of Alternative 4.
- Alternative 5 dominates Alternative 2 in 97.4 % of the simulation runs. Also the expected overall performance score of Alternative 5 (0.316) is higher than the corresponding 0.16 of Alternative 2.
- Alternative 5 dominates Alternative 3 in 97.3 % of the simulation runs. Also the expected overall performance score of Alternative 5 (0.316) is higher than the corresponding 0.172 of Alternative 3.
- Alternative 5 dominates Alternative 1 in 79.5 % of the simulation runs. Also the expected overall performance score of Alternative 5 (0.316) is higher than the corresponding 0.223 of Alternative 1.

Top

Preferential sensitivities on first ranking performance of Alternative 5

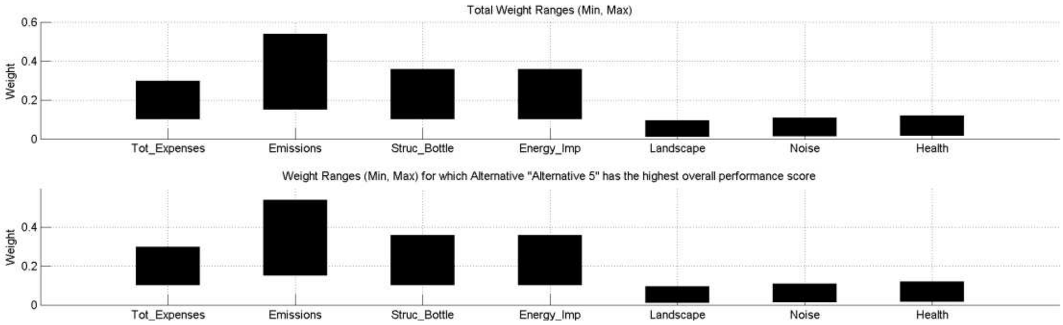
This paragraph analyses the preferential parameters that make Alternative 5 the preferred alternative based on the simulation results. The following aspects are examined:

a. Weight space exploration

Top

Weight space exploration

The 'weight range exploration visualisation' consists of two parts: The upper diagram shows the total weight space applied to the different simulation runs. In the lower diagram the black weight range without the red part represents all the weights for which Alternative 5 attains the first rank in the simulation runs.



Weight range for which Alternative 5 ranks first:

	Lower boundary	**Upper boundary**
Total Expenses of Electricity Supply	0.100	0.299
CO2 Emissions	0.150	0.540
Structural Bottlenecks	0.100	0.359
Primary energy import ratio	0.101	0.359
Landscape modification	0.009	0.095
Noise	0.012	0.110
Health implications	0.014	0.119

- As indicated by the red area the importance of the following criterion is most sensitive for the ranking of Alternative 5 as preferred alternative: Total Expenses of Electricity Supply.
- The relative importance of criteria CO2 Emissions and Structural Bottlenecks slightly affect the ranking of Alternative 5 as preferred alternative.

Dominance in relative importance

The following criteria dominate other criteria in importance. This means that a criterion A is considered strictly more important than a criterion B. All executed simulation runs and also exclusively those where Alternative 5 ranks first are considered.

- **Criterion Total Expenses of Electricity Supply** is more important than the criterion Landscape modification in all simulation runs.
- **Criterion CO2 Emissions** is more important than the criteria Landscape modification, Noise and Health implications in all simulation runs.

- **Criterion Structural Bottlenecks** is more important than the criterion Landscape modification in all simulation runs.
- **Criterion Primary energy import ratio** is more important than the criterion Landscape modification in all simulation runs.
- **Criterion Landscape modification** is less important than the criteria Total Expenses of Electricity Supply, CO2 Emissions, Structural Bottlenecks and Primary energy import ratio in all simulation runs.
- **Criterion Noise** is less important than the criterion CO2 Emissions in all simulation runs.
- **Criterion Health** implications is less important than the criterion CO2 Emissions in all simulation runs.

Indistinguishability of relative importance

The following criteria are indistinguishable with regard to their importance, i.e. their weight ranges overlap. All executed simulation runs and exclusively those where Alternative 5 ranks first are considered.

- The relative importance of **criterion Total Expenses of Electricity Supply** is indistinguishable from the relative importance of criteria CO2 Emissions, Structural Bottlenecks, Primary energy import ratio, Noise and Health implications when all simulation runs are considered.
- The relative importance of **criterion CO2 Emissions** and criteria Total Expenses of Electricity Supply, Structural Bottlenecks and Primary energy import ratio are indistinguishable when all simulation runs are considered.
- The relative importance of **criterion Structural Bottlenecks** is indistinguishable from the relative importance of criteria Total Expenses of Electricity Supply, CO2 Emissions, Primary energy import ratio, Noise and Health implications when all simulation runs are considered.
- The relative importance of **criterion Primary energy import ratio** is indistinguishable from the relative importance of criteria Total Expenses of Electricity Supply, CO2 Emissions, Structural Bottlenecks, Noise and Health implications when all simulation runs are considered.
- The relative importance of **criterion Landscape modification** is indistinguishable from the relative importance of criteria Noise and Health implications when all simulation runs are considered.
- The relative importance of **criterion Noise** is indistinguishable from the relative importance of criteria Total Expenses of Electricity Supply, Structural Bottlenecks, Primary energy import ratio, Landscape modification and Health implications when all simulation runs are considered.
- The relative importance of **criterion Health implications** is indistinguishable from the relative importance of criteria Total Expenses of Electricity Supply, Structural Bottlenecks, Primary energy import ratio, Landscape modification and Noise when all simulation runs are considered.
-

Top

Nomenclature

The following table defines scientific terms used in this report:

Term	**Definition**
5 %-quantile	5 % of all the values of a sample are smaller than this value.
95 %-quantile	95 % of all values of a sample are smaller than this value.
Additive value function	Possible aggregation function of Multi-Attribute Value Theory (MAVT): the individual performance scores on the criteria are multiplied with their respective weights and then summed up to an overall performance score.
Concave shape	The graph of a function is always above the intersection of any two points of the graph.
Convex shape	The graph of a function is always below the intersection of any two points of the graph.
Correlation	Describes the relationship between two samples of data. In this report it describes the linear relationship between the overall performance scores. A positive correlation means that with increasing overall performance scores of one alternative, also the other alternative's overall performance scores increase. A negative correlation indicates however that with increasing overall performance scores of one alternative, the other alternative's overall performance scores decrease.
Criterion level	This is the performance of an alternative on a criterion, e.g. Alternative A achieves 1.000 EUR on criterion price. This value is later transformed to a 0-1-scale by a value function.
Cumulative percentage	Indicates the percentage of simulation runs for which an alternative has an equal or lower overall performance score than at this point.
Cumulative performance	Indicates the overall performance scores of an alternative that are equal or lower than at this point.
Curvature of the value functions	Describes the curvature that distinguishes a function from a linear function. The higher the curvature the steeper (both positive and negative) the graph.

Data uncertainties	Term to describe uncertainties regarding the obtained criteria levels of the alternatives, especially when they are difficult to assess. For example, the acceptance of an alternative in the population cannot easily be expressed in one value and is thus influenced by uncertainties.
Decreasing preferences	Describes the preferences of an individual who prefers to have less of a good. For example, a low price is usually preferred to a high price.
Deterministic decision situation	Consideration of the alternatives without any uncertainty (data or preferential uncertainty) involved.
Differentiating influence	A criterion with high differentiating influence has high impact on the distinguishability of overall performance scores of the alternatives, i.e. this criterion is very important for the overall decision problem.
Distinguishability	Variables can be distinguished when their value ranges do not overlap.
Distribution of overall performance scores	The way how variables are distributed, e.g. normally distributed variables are distributed around their expected value.
Dominance	Alternative A dominates Alternative B when it is strictly better than B in every simulation run.
Expected overall performance score	In case of symmetric distributions, this value represents the average overall performance score of an alternative.
Expected rank performance	This value represents the average rank of an alternative for all simulation runs.
Increasing preferences	Describes the preferences of an individual who prefers to have more of a good. For example, a high income is usually preferred to a low income.
Indistinguishability	Variables cannot be distinguished when their value ranges overlap.
Inter-criteria preferential uncertainties	Uncertainties regarding the importance of different criteria, modelled by the weights of the criteria.
Inter-quantile range	Value difference of the 5 % and 95 %-quantile. In case of symmetric distributions, the inter-quantile range represents the 90 % most probable values of a sample around the expected

	value.
Intra-criteria preferential uncertainties	Uncertainties regarding the importance of differences of the performance scores of the alternatives with regard to every criterion individually, modelled by the value function shape.
Meaningful representation	The modelling of the decision situation holds certain conditions (e.g. regarding the weighting of the criteria or the adequate covering of the solution space; see meaningful analysis section of this report for more details).
Multi-Attribute Value Theory	Field of Multi-Criteria Decision Analysis that deals with a finite and discrete number of alternatives.
Multi-Criteria Decision Analysis	Field of Operations Research that formalises decision-making between different alternatives and multiple criteria.
Multi-dimensional sensitivity analysis	Assessment of the influence of simultaneous changes of the inter-criteria and intra-criteria preferential uncertainties on the model results.
Multivariate statistical methods	Statistical methods that analyse and observe simultaneously more than one outcome variable. For SIMADA these variables are the preferential parameters and performance scores of the alternatives.
Non-meaningful representation	The modelling of the decision situation does not hold certain conditions (e.g. regarding the weighting of the criteria or the adequate covering of the solution space; see meaningful analysis section of this report for more details).
Normalised vector	The vector is normalised to a length of 1. This is done for visualisation purposes of the principal component analysis visualisation.
Overall performance score	Aggregation of all performance scores on the considered criteria for an alternative. This represents the final performance result for an alternative.
Performance score	Score for an alternative on a criterion which is between 0 and 1. It is determined by the value function of the respective criterion.
Preferential parameters	These can include weight ranges, value function curvature ranges and the variations for upper and lower boundaries of the value functions.
Preferential uncertainties	In SIMADA these include intra-criteria and inter-criteria preferential uncertainties which can be analysed in a combined or

	separated way.
Principal component analysis	Projects multi-dimensional data in a 2-dimensional plane for better visualisation purposes by multivariate statistical methods. It aims to keep as much as possible from the original data, i.e. it minimises the loss of information between the original data set and the projection.
p-Value	Represents a mean for testing a statistical hypothesis. It describes the probability of obtaining a result equal to or "more extreme" than what was actually observed, assuming that the hypothesis under consideration is true. In this report, a low p-Value supports the hypothesis of the correlation This value represents the maximal probability of refusing the correlation hypothesis in the case where it is actually existing in the analysed data.
Ranking performance	This term describes the rank that an alternative obtains in a simulation run.
Scenario	Technique for dealing with uncertainty in decision situations. A scenario describes a possible set of future conditions (e.g. developments of prices). By comparing the results of different scenarios one is able to deduct cause-effect relations on the overall result of a decision situation.
Sensitivity analysis	Assessment of the most important influencing input factors on the model results.
SIMADA	Simulation-based Multi-Attribute Decision Analysis; name of the underlying decision model of this report.
Simulation run	A simulation run describes the modelling of the decision situation with a drawn sample of preferential parameters. In every simulation run that is executed different sets of preferential parameters are drawn. That's why the more simulation runs are executed the more the obtained results converge to the mathematical expected overall performance score for every alternative.
Solution space	The 2-dimensional plane of the PCA visualisation is the solution space. It needs to be entirely covered by the criteria axis.
Standard deviation	The standard deviation is a measure to express the dispersion of the overall performance scores of an alternative around its expected overall performance score.

Stochastic dominance	This term describes the probability of an alternative to dominate over another one, i.e. it is determined by the percentage of simulation runs in which an Alternative A attains a higher overall performance score than Alternative B.
Value function	Transforms the criteria level of the alternatives to a common 0-1-scale for comparison purposes. Also describes the differences between the criterion levels by its value function shapes.
Value function curvature variations	Modelling parameter that describes the uncertainty of differences for the performance scores of the alternatives.
Value function shape	Describes the curvature and slope of the value function.
Variance	Value to describe the dispersion of samples around their expected value.
Variations of lower boundary of the value functions	Parameters that describe the uncertainty of differences in performance scores for a criterion, especially with regard to the minimum of performance scores.
Variations of upper boundary of the value functions	Parameters that describe the uncertainty of differences in performance scores for a criterion, especially with regard to the maximum of performance scores.
Weight variations	Varied modelling parameters that describes the uncertainty of relative importance of the criteria.

Top