

# Towards robust ELECTRE III with simulation: theory and software of SMAA-III

Tommi Tervonen <sup>a,b,\*</sup>, José Rui Figueira <sup>a</sup>, Risto Lahdelma <sup>b</sup>,  
Pekka Salminen <sup>c</sup>

<sup>a</sup>*CEG-IST, Centre for Management Studies, Instituto Superior Técnico, Technical  
University of Lisbon, 2780-990 Porto Salvo, Portugal*

<sup>b</sup>*Department of Information Technology, University of Turku, FIN-20520 Turku, Finland*

<sup>c</sup>*School of Business and Economics, University of Jyväskylä, P.O. Box 35, FIN-40014  
Jyväskylä, Finland*

---

## Abstract

ELECTRE III is a well-established multiple criteria decision making method with a solid track of real-world applications. It requires precise values to be specified for the parameters and criteria measurements, which in some cases might not be available. In this paper we present a method, SMAA-III, that allows ELECTRE III to be applied with imprecise parameter values. By allowing imprecise values, the method also allows an easily applicable robustness analysis. In SMAA-III, simulation is used and descriptive measures are computed to characterize stability of the results. We present a software implementing the method and show the usage by re-analyzing an existing case study.

*Key words:* Decision support systems, robustness analysis, ELECTRE III, Stochastic Multicriteria Acceptability Analysis (SMAA)

---

## 1 Introduction

ELECTRE III is a well-established multiple criteria decision making (MCDM) method for ranking a discrete set of alternatives. It belongs to the ELECTRE family of methods that are based on constructing and exploiting an outranking relation

---

\* Corresponding author.

*Email addresses:* `tommi.tervonen@it.utu.fi` (Tommi Tervonen),  
`figueira@ist.utl.pt` (José Rui Figueira), `risto.lahdelma@cs.utu.fi` (Risto Lahdelma), `psalmine@econ.jyu.fi` (Pekka Salminen).

(see [4]). ELECTRE III has a long history of successful real-world applications in different areas. The inputs for ELECTRE III consist of criteria evaluations on a set of alternatives and preference information expressed as weights and thresholds.

ELECTRE III is a pseudo-criterion based model, and as such it uses a threshold to model indifference between pairs of alternatives. Although this threshold might be an easy concept for a common Decision Maker (DM) to understand, simulation studies have shown, that it causes the model be quite unstable with respect to changes in the indifference threshold value [8]. Because of this instability, robustness analysis should always be done by considering different values for the threshold.

Real-world decision-making problems in general include various types of uncertainties inherent in problem structuring and analysis [1]. Eliciting the DMs preferences in terms of relative criteria importance coefficients or weights is usually difficult. Such weights should always be considered imprecise, because humans usually do not think about preferences as exact numerical values, but as more vague concepts [14]. In some cases, weight information may be entirely missing, which corresponds to extremely imprecise weights.

This work presents a tool for dealing with imperfect knowledge within the ELECTRE III method. It can be used either when information is poor and/or when a robustness analysis needs to be done. The way robustness analysis is conducted comprises intensity of exploration in the parameters space. This is achieved by applying simulation in such a way that the parameter space is explored with a high concentration of discrete values. In addition to this, the exploration is coherent with the model. This means that, for example, when exploring the weight space, the meaning of weight is taken into account. In ELECTRE III weights represent the amount of “votes” criteria have.

Capability to derive robust conclusions when applying MCDM methods is nowadays of uttermost importance. The main sources of imperfect knowledge that are present in complex and multifaceted decision-making situations require a careful observation of the results, and make them dependent of an exploration of the neighbourhood of the parameters used mainly to represent preferences or technical aspects of the problem. If an alternative occupies almost always the first position when changing simultaneously all the parameters in a certain neighbourhood, it means that it can be a good choice for future implementation; these are the kind of robust conclusions we are interested in.

The method presented in this paper is based on Stochastic Multicriteria Acceptability Analysis (SMAA) [7], that is a family of decision support methods for aiding DMs in discrete decision-making problems. For a survey on SMAA methods, see [15]. The proposed method, SMAA-III, explores weight, criteria measurement, and threshold spaces, in order to describe which values result in certain ranks for the

alternatives. It allows ELECTRE III to be used with different kinds of imprecise or partially missing information. This brings numerous advantages. Firstly, SMAA-III allows performing an initial analysis without preference information in order to eliminate “inferior” alternatives. Secondly, it allows DMs to express their preferences imprecisely, which can lower the DMs’ cognitive effort compared to specifying precise weights. Thirdly, imprecise criteria measurements can be represented with arbitrary joint probability distributions, allowing to model imprecision in a coherent way not possible with ELECTRE III. Fourthly, it allows representing the preferences of a group of DMs. Fifthly, the method can be used for analyzing the robustness of the results by representing the imprecision of the elicited weights as constraints or as suitable probability distributions.

In SMAA-III, robustness is analyzed with respect to the weights, criteria measurements, and thresholds. Traditionally, robustness with ELECTRE methods is analyzed by considering discrete points in the weight space (see e.g. [12]). But in the case of ELECTRE III this is not enough: weights between these points that might give contradictory results are missed. There are simulation techniques for robustness analysis also outside the SMAA methodology (see [2]), but to our best knowledge, they have never before been applied to ELECTRE III.

User-friendly software is of crucial importance for a MCDM method to be of practical importance. We present the software for SMAA-III, and demonstrate its use by re-analyzing a real-life case study. The software presented is available for all major operating systems.

This paper is organized as follows: ELECTRE III is briefly introduced in Section 2. SMAA-III is presented in Section 3. The software and a re-analysis of a case-study are presented in Section 4. The paper ends with conclusions and avenues for future research in Section 5.

## 2 ELECTRE III

ELECTRE III is designed for solving a discrete ranking problem. It consists of  $m$  alternatives  $a_1, \dots, a_i, \dots, a_m$ , that are evaluated in terms of  $n$  criteria  $g_1, \dots, g_j, \dots, g_n$ . We denote by  $J$  the set of criterion indices.  $g_j(a_i)$  is the evaluation of criterion  $g_j$  for alternative  $a_i$ . Without loss of generality, we assume that all criteria are to be maximized.

Similarly to the other ELECTRE family methods, ELECTRE III is based on two phases. In the first phase, an outranking relation between pairs of alternatives is formed. The second phase consists of exploiting this relation, producing a final partial pre-order and a median pre-order.  $S$  denotes the *outranking* relation, that is,  $aSb$  denotes that “alternative  $a$  is at least as good as alternative  $b$ ”.

ELECTRE III applies pseudo-criteria in constructing the outranking relation. A pseudo-criterion is defined with two thresholds for modelling preferability: an indifference threshold  $q_j(g_j(\cdot))$  for defining the difference in criterion  $g_j$  that the DM deems insignificant, and a preference threshold  $p_j(g_j(\cdot))$  for the smallest difference that is considered absolutely preferred. Between these two is a zone of “hesitation” between indifference and strict preference. ELECTRE III also defines third threshold, a veto threshold  $v_j(g_j(\cdot))$ . It is the smallest (negative) difference that completely nullifies (raises “veto” against) the outranking relation. In addition to the thresholds, preferences are quantified through a weight vector  $w = (w_1, \dots, w_j, \dots, w_n)$ . Without loss of generality, we assume that  $\sum_{j \in J} w_j = 1$ .

Exploitation of the outranking relation produces a partial pre-order, in which every pair of alternatives is connected with either indifference ( $I$ ), incomparability ( $R$ ), or preference ( $\succ$ ) relation.

## 2.1 Constructing the outranking relation

The outranking relation between every pair of alternatives is constructed based on a comprehensive concordance index and partial discordance indices. The concordance index is computed by considering individually for each criterion  $g_j$  the support it provides for the assertion  $aS_jb$  “ $a$  outranks  $b$  with respect to criterion  $g_j$ ”. The partial concordance index is computed as follows, for all  $j \in J$ :

$$c_j(a, b) = \begin{cases} 1, & \text{if } g_j(b) - g_j(a) \leq q_j(g_j(a)), \\ \frac{g_j(a) + p_j(g_j(a)) - g_j(b)}{p_j(g_j(a)) - q_j(g_j(a))}, & \text{if } q_j(g_j(a)) < g_j(b) - g_j(a) \leq p_j(g_j(a)), \\ 0, & \text{if } g_j(b) - g_j(a) > p_j(g_j(a)). \end{cases} \quad (1)$$

After computing the partial concordance indices, the comprehensive concordance index is computed as follows,

$$c(a, b) = \sum_{j \in J} w_j c_j(a, b). \quad (2)$$

The discordance of criterion  $g_j$  describes the veto effect this criterion imposes against the assertion  $aSb$ . The partial discordance indices are computed separately

for each criterion  $j \in J$ :

$$d_j(a, b) = \begin{cases} 1, & \text{if } g_j(b) - g_j(a) \geq v_j(g_j(a)), \\ \frac{g_j(b) - g_j(a) - p_j(g_j(a))}{v_j(g_j(a)) - p_j(g_j(a))}, & \text{if } p_j(g_j(a)) \leq g_j(b) - g_j(a) < v_j(g_j(a)), \\ 0, & \text{if } g_j(b) - g_j(a) < p_j(g_j(a)). \end{cases} \quad (3)$$

By applying the pre-mentioned indices, the degree of credibility of the outranking assertion  $aSb$  is defined as

$$\rho(a, b) = \begin{cases} c(a, b) \prod_{j \in V} \frac{1 - d_j(a, b)}{1 - c(a, b)}, & \text{if } V \neq \emptyset, \\ c(a, b), & \text{otherwise,} \end{cases} \quad (4)$$

with

$$V = \{j \in J : d_j(a, b) > c(a, b)\}. \quad (5)$$

Notice that when  $d_j(a, b) = 1$  for any  $j \in J$ , this implies that  $\rho(a, b) = 0$ .

## 2.2 The exploitation procedure

The exploitation of the outranking relation consists of two phases. In the first phase, two complete pre-orders,  $Z_1$  (descending) and  $Z_2$  (ascending) are constructed with the so-called distillation procedures. In the second phase, a final partial pre-order or a complete median pre-order is computed based on these two pre-orders.

The distillation procedures work by iteratively cutting the fuzzy outranking relations with descending  $\lambda$ -cutting levels. With a given cutting level  $\lambda_*$ , alternative  $a$  outranks alternative  $b$  ( $aS^{\lambda_*}b$ ) if the following holds:

$$aS^{\lambda_*}b \iff \begin{cases} \rho(a, b) > \lambda_*, \text{ and} \\ \rho(a, b) > \rho(b, a) + s(\rho(a, b)), \end{cases} \quad (6)$$

where  $s(\cdot)$  is the distillation threshold, usually defined as [1]

$$s(x) = 0.3 - 0.15x. \quad (7)$$

The pre-orders are constructed in an iterative manner. In each step the alternatives with the highest or lowest qualification scores are distilled, depending on whether the distillation is descending or ascending. The qualification score is computed as a difference between the number of alternatives that the selected alternative out-

ranks and the number of alternatives that outrank it for a given cutting level. The procedure is presented in Algorithm 1.

---

**Algorithm 1** Distillations

---

- 1: Determine the maximum value of the credibility indices in the set under consideration. Assign this to  $\lambda$ .
  - 2: Determine  $\lambda_* = \max_{d(a,b) < \lambda - s(\lambda)} \{d(a,b)\}$ , where  $(a,b)$  belong to the set under consideration.
  - 3: If  $\lambda_* = 0$ , end this distillation.
  - 4: Determine for each alternative its *qualification* score, that is: the difference between the number of alternatives it outranks and the number of alternatives that outrank it. Outranking is determined according to  $\lambda_*$ .
  - 5: The set of alternatives having the largest (or smallest, if the distillation is ascending) qualification is the current distillate.
  - 6: If the number of alternatives in current distillate is larger than 1, repeat the process from step 2 inside the distillate.
  - 7: Form a new set under consideration by removing the distilled alternatives from the current one. If this set is not empty, repeat the process on the new set from step 1.
  - 8: The final pre-orders are ranked so that the alternatives in the first distillate are given rank 1, in the second rank 2, etc.
- 

In the original ELECTRE III, a median pre-order is computed based on the two complete pre-orders,  $Z_1$  and  $Z_2$ , and the final partial pre-order. The final partial pre-order is computed as the intersection of the two complete pre-orders in such a way that the following relations hold:

$$\begin{aligned}
 a \succ b &\iff (a \succ^{Z_1} b \wedge a \succ^{Z_2} b) \vee (a I^{Z_1} b \wedge a \succ^{Z_2} b) \vee (a \succ^{Z_1} b \wedge a I^{Z_2} b), \\
 a I b &\iff (a I^{Z_1} b \wedge a I^{Z_2} b), \\
 a R b &\iff (a \succ^{Z_1} b \wedge b \succ^{Z_2} a) \vee (b \succ^{Z_1} a \wedge a \succ^{Z_2} b).
 \end{aligned} \tag{8}$$

After this, the median pre-order can be computed by removing the incomparabilities with calculating the differences of ranks of an alternative in the two complete pre-orders.

### 2.3 Robustness analysis for weights

There are numerous weight elicitation techniques proposed for ELECTRE methods, the following being among the most recent and popular ones:

- (1) DIVAPIME by Mousseau [10] produces intervals for weights.
- (2) Hokkanen and Salminen [5] used two different weight elicitation procedures, and found that the normalized sets of weights had minor differences.

- (3) SRF by Figueira and Roy [3] allows weight elicitation in a user-friendly manner by using a technique based on a pack of “playing cards” to determine the relative importance criteria coefficients. It can produce interval weights and was also designed to support multiple DMs.
- (4) The approach proposed by Rogers and Bruen [11] uses pairwise comparisons to elicit the weights.

The first three techniques that produce intervals or two weight sets that may be used to define intervals, can directly be used in robustness analysis. When using the fourth weight elicitation technique, intervals (such as  $\pm 10\%$ ) could be defined around the original weights.

Traditionally the robustness analysis for ELECTRE methods has been an *ad hoc* investigation into the effect of changing values [1]. This type of investigation typically considers only discrete points (for example, extreme points) of the feasible weight space (e.g. weight intervals). The procedure of building the pre-orders is based on exploiting the fuzzy outranking relation, which is non-linear and discontinuous by nature. Therefore, instead of just a few discrete points, it is important to analyze the entire continuum of the weight space.

### 3 SMAA-III

In order to overcome the limitations of ELECTRE III, SMAA-III applies simulation and studies the effect of changing parameter values and criterion evaluations on the results. The imprecision is quantified through joint density functions in the corresponding spaces.

The weights are represented by a weight distribution with joint density function  $f_W(w)$  in the feasible weight space  $W$ . The weights are non-negative and normalized: the weight space is an  $n - 1$  dimensional simplex,

$$W = \left\{ w \in \mathbb{R}^n : w \geq 0 \text{ and } \sum_{j \in J} w_j = 1 \right\}. \quad (9)$$

Completely missing preference information is represented by a uniform (constant) weight distribution in  $W$ , that is,

$$f_W(w) = 1/\text{vol}(W). \quad (10)$$

If some kind of preference information is available, different weight distributions can be applied (see [7]). In practice, the preferences can usually be elicited as interval constraints for weights. In this case, a uniform distribution in the space bounded by the constraints is used. Figure 1 illustrates the restricted feasible weight space of

a 3-criterion problem with lower and upper bounds for  $w_1$ . In this paper the focus is on weight information provided as intervals, because:

- (1) if there are multiple DMs whose preferences need to be taken into account, the weight intervals in general can be determined to contain the preferences of all DMs (see [7]), and
- (2) weight intervals allow simple robustness analysis also when only deterministic weights are available, by specifying, for example, a  $\pm 10\%$  interval for each weight.

It should be observed that other forms of easily elicitable preference information can be used as well, such as ranking of the criteria. A ranking can be obtained by asking the DMs to identify their most important, second most important, etc. criterion. Figure 2 illustrates the feasible weight space for a three-criterion problem with the ranking  $w_1 \geq w_2 \geq w_3$ .

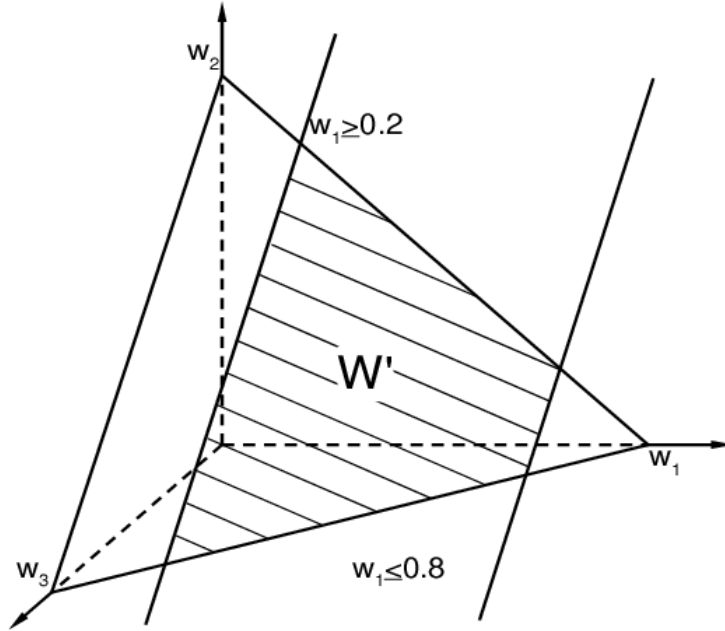


Figure 1. Feasible weight space of a 3-criterion problem with lower and upper bounds for  $w_1$ .

Imprecise thresholds are represented by stochastic functions  $\alpha_j(\cdot)$ ,  $\beta_j(\cdot)$ , and  $\gamma_j(\cdot)$ , corresponding to the deterministic thresholds  $p_j(\cdot)$ ,  $q_j(\cdot)$ , and  $v_j(\cdot)$ , respectively. To simplify the notation, we define a 3-tuple of thresholds  $\tau = (\alpha, \beta, \gamma)$ . It has a joint density function  $f_T$  in the space of possible values defining the functions. It should be noted that all feasible combinations of thresholds must satisfy  $q_j(a_i) < p_j(a_i) < v_j(a_i)$ .

Traditionally the thresholds in ELECTRE models have been used to model preferences of the DMs (e.g. differences deemed significant) as well as imprecision



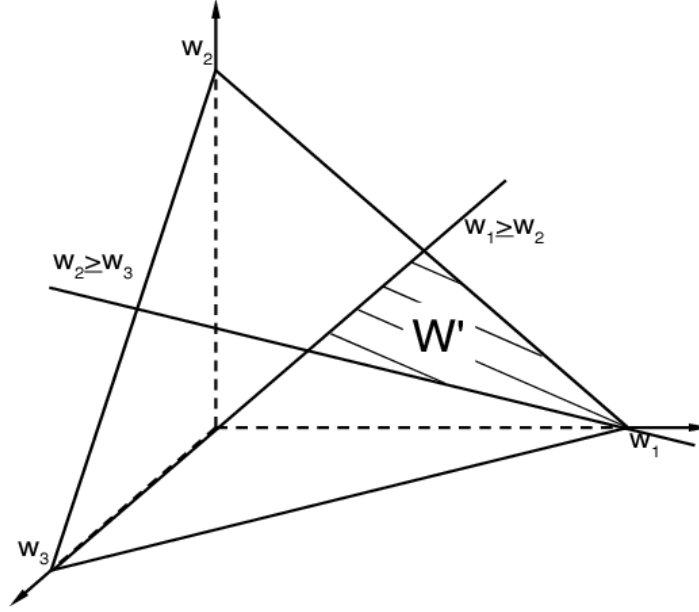


Figure 2. Feasible weight space of a 3-criterion problem with ranking of the criteria.

in the data. But it has been shown that the indifference threshold does not correspond to a linear imprecision interval [8]. Therefore, in SMAA-III thresholds are used only to model preferences (together with weights). Imprecision in the criteria measurements is modelled with stochastic variables.

These stochastic variables are denoted with  $\xi_{ij}$  corresponding to the deterministic evaluations  $g_j(a_i)$ . They have a density function  $f_X(\xi)$  defined in the space  $X \subseteq \mathbb{R}^{m \times n}$ . In principle, arbitrary distributions can be used, but in practice a uniform distribution in a certain interval or a Gaussian distribution is used.

Incomparabilities between alternatives can be present in the final results of ELECTRE III. This is one of the main features of ELECTRE methods in comparison with the methods applying classical multi-attribute utility theory (see [6]). In the late seventies, it was considered a very important theoretical advance. But, in reality when dealing with practical situations, incomparabilities in the final result are inconvenient. This aspect was soon observed [13] and partial pre-orders were replaced by complete pre-orders or median pre-orders. This is quite logical, as why would DMs want to apply a tool that tells the alternatives are incomparable: this was already the starting position. We apply median pre-orders in computing rank acceptability indices. The only information lost in using the median pre-order as the primary measure of the ranking is the incomparability. As our method is also aimed to help analysts accustomed to ELECTRE III, we will later present another index to measure incomparability.

Since only the median pre-order is needed in SMAA-III, the intermediate step of constructing the final partial pre-order can be skipped; the median pre-order is con-

structed directly from the two complete pre-orders in the following way:

$$a \succ b \iff \begin{cases} (a \succ^{Z_1} b \wedge a \succ^{Z_2} b) \vee (a I^{Z_1} b \wedge a \succ^{Z_2} b) \vee (a \succ^{Z_1} b \wedge a I^{Z_2} b) \\ (a \succ^{Z_1} b \wedge b \succ^{Z_2} a) \wedge (|r^{Z_1}(a) - r^{Z_2}(a)| < |r^{Z_1}(b) - r^{Z_2}(b)|) , \\ (b \succ^{Z_1} a \wedge a \succ^{Z_2} b) \wedge (|r^{Z_1}(a) - r^{Z_2}(a)| < |r^{Z_1}(b) - r^{Z_2}(b)|) \end{cases}$$

$$a I b \iff \neg(b \succ a) \wedge \neg(a \succ b),$$

where  $r(\cdot)$  is the ranking of an alternative in the superscripted pre-order.

Monte Carlo simulation is used in SMAA-III to compute three types of descriptive measures: rank acceptability indices, pair-wise winning indices, and incomparability indices. In order to compute these indices, let us define a *ranking function* that evaluates the rank  $r$  of the alternative  $a_i$  with the corresponding parameter values:

$$rank(i, w, \xi, \tau). \quad (11)$$

The evaluation of this function corresponds to executing ELECTRE III and returning rank of the corresponding alternative in the resulting median pre-order. We will next introduce the three indices. Interpretation of their values is presented in Section 4 through various re-analyses.

### 3.1 Rank acceptability index

The rank acceptability index,  $b_i^r$ , measures the share of feasible weights that grant alternative  $a_i$  rank  $r$  in the median pre-order by taking into account simultaneously imprecisions in all parameters and criterion evaluations. It represents the share of all feasible parameter combinations that make the alternative acceptable for a particular rank, and it is most conveniently expressed percentage-wise.

The rank acceptability index  $b_i^r$  is computed numerically as a multidimensional integral over the spaces of feasible parameter values as

$$b_i^r = \int_{W: rank(i, w, \xi, \tau) = r} f_W(w) \int_X f_X(\xi) \int_T f_T(\tau) dT dw d\xi. \quad (12)$$

The most acceptable (“best”) alternatives are those with high acceptabilities for the best ranks. Evidently, the rank acceptability indices are within the range  $[0,1]$ , where 0 indicates that the alternative will never obtain a given rank and 1 indicates that it will obtain always the given rank with any feasible choice of parameters. Thus, the rank acceptability indices are a measure of robustness.

Using the rank acceptability indices as measures of robustness is quite straightforward. More caution should be put on interpreting the results when these indices are

computed without weight information to characterize the problem. If an alternative obtains a low score for the first rank acceptability, it does not necessarily mean that it is “inferior”. The DMs’ actual preferences may well lie within the corresponding (small) set of favourable first rank weights.

### 3.2 Pair-wise winning index

The pair-wise winning index [9],  $o_{ik}$ , describes the share of weights that place alternative  $a_i$  on a better rank than alternative  $a_k$ . An alternative  $a_i$  that has  $o_{ik} = 1$  for some  $k$  always obtains a better rank than alternative  $a_k$ , and can thus be said to *dominate* it.

The pair-wise winning index  $o_{ik}$  is computed numerically as a multidimensional integral over the space of weights that give alternative a lower rank than for another.

$$o_{ik} = \int_{w \in W: \text{rank}(i, w, \xi, \tau) < \text{rank}(k, w, \xi, \tau)} f_W(w) \int_X f_X(\xi) \int_T f_T(\tau) dT dw d\xi. \quad (13)$$

The pair-wise winning indices are especially useful when trying to distinguish between the ranking differences of two alternatives. Because the number of ranks in the median pre-order of different simulation runs varies, two alternatives might obtain similar rank acceptabilities although one is in fact inferior. In these cases looking at the pair-wise winning indices between this pair of alternatives can help to determine whether one of the alternatives is superior to the other or if they are equal in “goodness”.

### 3.3 Incomparability index

Because median pre-orders are used in computing the rank acceptability indices, it is not anymore possible to model incomparability. As some DMs might be accustomed to make decisions also based on incomparabilities, another index is introduced. Incomparability index  $p_{ik}$  measures the share of feasible parameter values that cause alternatives  $a_i$  and  $a_k$  to be incomparable. For this reason, we define the incomparability function:

$$R(i, k, \xi, \tau) = \begin{cases} 1, & \text{if alternatives } a_i \text{ and } a_k \text{ are judged incomparable,} \\ 0, & \text{if not.} \end{cases} \quad (14)$$

This function corresponds to a run of ELECTRE III with the given parameter values and checking if the alternatives are judged incomparable in the final partial

pre-order. In practice we do not compute the final partial pre-order, because this information can be extracted from the two partial pre-orders  $Z_1$  and  $Z_2$  as shown in (8). By using the incomparability function, the incomparability index is computed numerically as a multidimensional integral over the feasible parameter spaces as

$$\rho_{ik} = \int_W f_W(w) \int_X f_X(\xi) \int_T f_T(\tau) R(i, j, \xi, \tau) dT dw d\xi. \quad (15)$$

### 3.4 Computation

All of the indices mentioned above are computed with Monte Carlo simulation. The procedure is similar to that presented and analyzed by Tervonen and Lahdelma [16]. SMAA-III differs in the sense that it applies the ELECTRE III procedure to derive the descriptive values instead of a utility function.

In each simulation iteration, sample parameter values are generated from their corresponding distributions, and ELECTRE III is executed with these values. Then the corresponding hit-counters are updated as with the original SMAA. If standard distributions are used for defining the imprecise parameter values, then all sampling except weight generation are computationally very light. In the case of weight generation, if tight upper bounds are used, then we can have very high weight rejection ratios (up to 99,9%). Nevertheless, even with 99,9% weight rejection, the method is fast enough to use in an interactive decision making process with problems of reasonable size.

For obtaining sufficient accuracy for the indices, we suggest using at least 10000 simulation iterations. This gives error limits of less than 0.01 with 95% confidence [16].

## 4 Case study and software

ELECTRE III has been used to choose the best waste incineration strategy for the Eastern Switzerland region [12]. A total of 11 alternative strategies (alternatives) were evaluated in terms of 11 criteria. ELECTRE III was run separately for 6 interest groups, each of which had different preferences. The complete study will not be presented here. The interested reader should refer to [12, Section 6].

We re-analyze the study using the SMAA-III software. The software is programmed in C++ using portable user interface libraries, and is therefore available for various operating systems. Currently Linux, Macintosh OS X, and Windows XP are supported. The software allows easy input of all models data. The interface is designed in such a way that from each input phase, the user can move to any other one. For

examples of the interface, see tables for criterion type input in Figure 3 and criterion measurement input in Figure 4.

Name	Type	Direction	Indifference TH	Preference TH	Veto TH	Threshold type
C1.1	UNIF	DESC	1000	2000	Enable	ABS, DET
C1.2	UNIF	ASC	10 %	20 %	Enable	PROS, DET
C1.3	UNIF	DESC	10 %	20 %	Enable	PROS, DET
C2.1	UNIF	DESC	5	10	Enable	ABS, DET
C2.2	UNIF	DESC	10 %	20 %	Enable	PROS, DET
C3.1	UNIF	ASC	10 %	20 %	Enable	PROS, DET
C3.2	UNIF	ASC	10 %	20 %	Enable	PROS, DET
C3.3	UNIF	DESC	2	4	Enable	ABS, DET
C4.1	UNIF	DESC	0.2	0.4	Enable	ABS, DET
C4.2	Ordinal	DESC				
C4.3	Ordinal	DESC				

Figure 3. Input of criteria types in the software.

	C1.1	C1.2	C1.3	C2.1	C2.2	C3.1	C3.2	C3.3	C4.1	C4.2	C4.3
S1.1	125	866	9.81	218	1.41	542	483	23	1.5	1	1
S1.2	11980	900	11.45	189	1.45	452	303	12	1.5	6	6
S2.1	31054	883	9.86	172	1.82	341	311	0	0	3	3
S2.2	28219	840	10.38	171	1.95	339	318	0	0	3	3
S2.3	31579	903	10.74	165	1.7	312	281	0	0	5	5
S2.4	39364	922	13.87	167	1.65	287	269	0	0	8	7
S3.1	125	769	9.33	182	1.64	458	180	0	1.5	1	1
S3.2	8075	896	9.82	172	1.7	408	121	0	1.5	6	6
S4.1	3089	770	9.39	177	1.9	430	228	0	1	2	2
S4.2	6449	766	7.22	172	1.65	401	157	0	1	4	4
S4.3	12074	897	10.61	169	1.65	378	162	0	1	7	6

Figure 4. Input of criteria measurements in the software.

We re-analyze the “baseline run” of the case-study with weight information from the Switzerland's Federal Agency for the Environment. In this initial run, the veto thresholds were not defined. The weights were elicited by using two methods, the SRF method and the method by Rogers and Bruen [11]. According to [12], the differences in the weights obtained by using the two methods were minor. To see how small differences in the parameter values cause alteration of results, we re-analyze the problem with five different scenarios:

- (1) Original problem. Re-analysis by using the median pre-order.
- (2) Imprecise weights. Original problem with imprecise weight values.
- (3) Imprecise thresholds. Original problem with imprecise threshold values.
- (4) Imprecise criteria measurements. Original problem with imprecise measurements for the cardinal criteria.

- (5) Imprecision in all values. Original problem incorporating simultaneously all the above types of imprecision.

The goal of our re-analysis is to identify which parameters are the most sensitive, and what kind of robust conclusions can be derived from the results. We also show how the different index values should be interpreted. All these analyses are computed with 10000 Monte Carlo iterations.

#### 4.1 Original study with median pre-order

We ran the software with exact data from the original study, obtaining a median pre-order. Rank acceptability indices are shown in Figure 5. Notice that the indices for each alternative are 100% for a single rank, and 0% for the others, therefore defining a deterministic pre-order (S4.1 in the first rank, S3.1 in the second, etc).

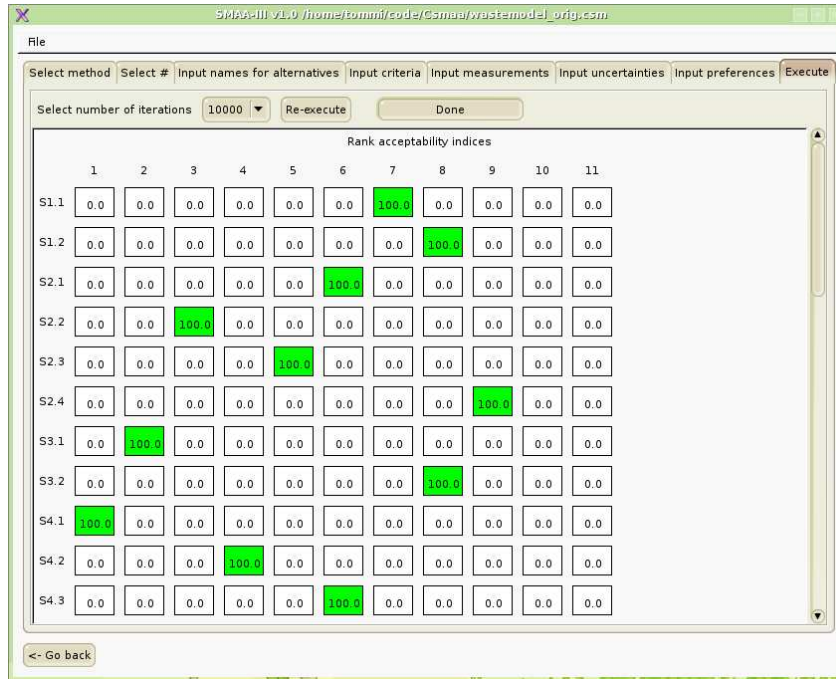


Figure 5. Rank acceptability indices of the pre-order of the original study.

#### 4.2 Imprecise weights

In the first re-analysis, we define the feasible weight space uniformly distributed and constrained to include original weights  $\pm 10\%$  in all dimensions. This weight information is introduced in the software as shown in Figure 6. The choice of intervals is quite arbitrary, and in a real-world application the intervals should be

approved by the DMs.

Figure 6. Input of ranges for weights in the software.

The effect of a variable number of ranks in different simulations makes it hard to interpret the results based on only the rank acceptability indices. Pair-wise winning indices are better for seeing how imprecise weights affect the ranking in comparison with exact weights. The pair-wise winning indices of this re-analysis are presented in Figure 7.

	S1.1	S1.2	S2.1	S2.2	S2.3	S2.4	S3.1	S3.2	S4.1	S4.2	S4.3
S1.1	0.0	100.0	0.0	0.0	0.0	100.0	0.0	77.5	0.0	0.0	44.4
S1.2	0.0	0.0	0.0	0.0	0.0	100.0	0.0	11.5	0.0	0.0	14.1
S2.1	100.0	100.0	0.0	0.0	30.3	100.0	0.0	91.5	0.0	1.2	63.2
S2.2	100.0	100.0	100.0	0.0	100.0	100.0	0.2	100.0	3.1	92.7	100.0
S2.3	100.0	100.0	69.5	0.0	0.0	100.0	0.0	100.0	0.0	1.3	100.0
S2.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S3.1	100.0	100.0	100.0	99.8	100.0	100.0	0.0	100.0	74.7	100.0	100.0
S3.2	22.5	43.7	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	8.2
S4.1	100.0	100.0	100.0	96.8	100.0	100.0	25.3	100.0	0.0	100.0	100.0
S4.2	100.0	100.0	97.6	0.1	98.7	100.0	0.0	100.0	0.0	0.0	100.0
S4.3	53.0	73.9	0.0	0.0	0.0	100.0	0.0	51.7	0.0	0.0	0.0

Figure 7. Pair-wise winning indices of the re-analysis with imprecise weights.

By looking at the pair-wise winning indices in Figure 7, we can see that the imprecise weights cause some alterations in the results. When the parameters are deterministic, there are only indices of 100% and 0%. Now with the imprecise values, for example, S2.1 is ranked better than S2.3 with 30% of the feasible weights values (had a worse rank in the original analysis). What is more important, is the change in ranking of alternatives that obtained good ranks in the original analysis. Although ELECTRE III is a ranking method, most of its applications, including this study, are in *selecting* an alternative to implement. The two best ranked alternatives in the original case study are S4.1 (1st rank) and S4.1 (2nd rank). But in this re-analysis incorporating imprecise weights, S3.1 obtains higher rank than S4.1 with 75% of the feasible weights. Therefore, by taking into consideration the purpose of the application (implementing the best alternative), we can say that weights are among the sensitive parameters of this model.

### 4.3 Imprecise thresholds

The second case re-analyzed is with imprecise thresholds. We define imprecise thresholds for cardinal criteria with  $\pm 10\%$  imprecision. We analyze the imprecision this time by looking at the rank acceptability indices, presented in Figure 8. This figure shows a common phenomenon with rank acceptability indices, emerging when the parameters are sensitive: the amount of ranks in different simulation runs changes. This is caused by some alternatives obtaining the same rank, therefore lowering the total number of ranks. This effect does not affect the first rank acceptabilities, but is cumulative in higher ranks.

For example, see the rank acceptabilities of alternative S2.4: 24.5% for rank 9, 63.1% for rank 10, and 12.4% for rank 11. Based on the rank acceptability indices, it would seem that the ranking of this alternative varies quite a lot. But by looking the pair-wise winning indices of the same re-analysis presented in Figure 9, more precise information is obtained. The column of alternative S2.4 tells the share of feasible parameter values for other alternatives to obtain a better rank than S2.4. All these (except for the alternatives own row, which is always 0%) are 100%, which means that alternative S2.4 is *always* ranked the last and never shares this rank with an another alternative.

Although this cumulative effect causes the rank acceptabilities to lose their intended meaning for the worse ranks, it does not hinder their help to decisions of selecting the best alternative. They also demonstrate in a comprehensive way the ranges of ranks for which alternatives can be assigned into, and whether these are overlapping with the corresponding ranges of another alternatives or not.



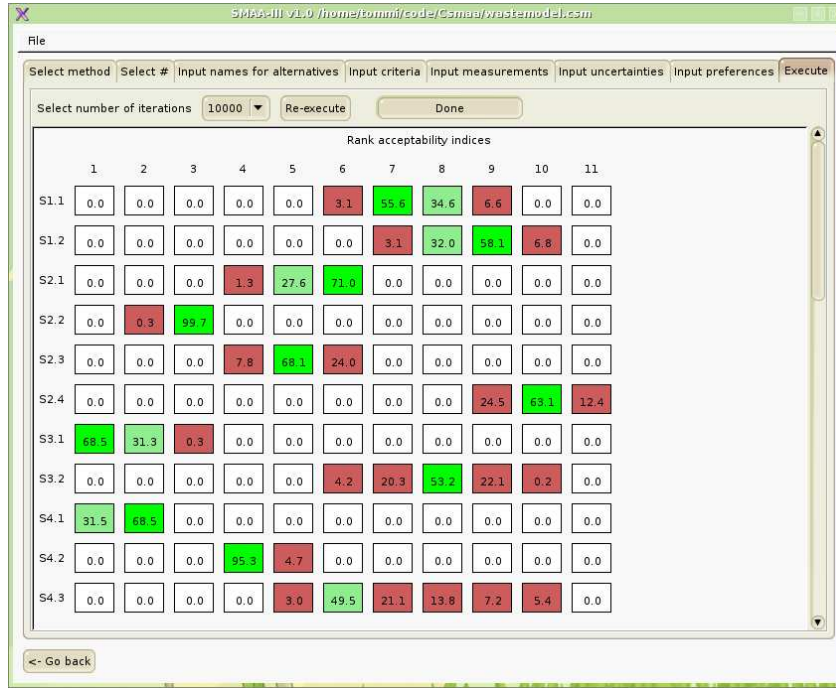


Figure 8. Rank acceptability indices of the re-analysis with imprecise thresholds.

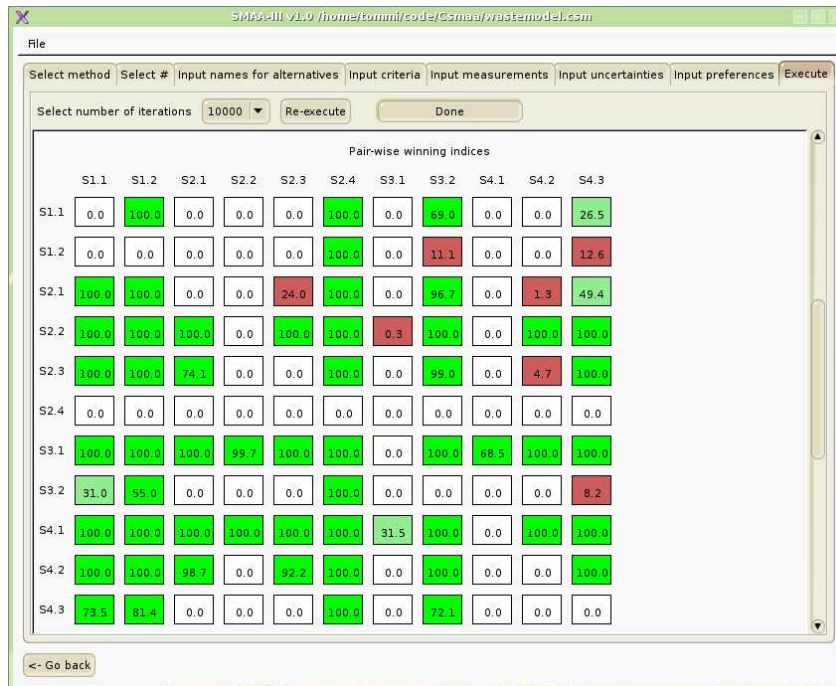


Figure 9. Pair-wise winning indices of the re-analysis with imprecise thresholds.

#### 4.4 Imprecise criteria measurements

In the third re-analysis we add imprecision to the criteria measurements. We define  $\pm 10\%$  imprecision interval for each cardinal criterion. The rank acceptability

indices of this re-analysis are presented in Figure 10. As was expected, the results show quite a large amount of dispersion in the indices. In a real application, the uncertainties could have been quantified more precisely. Anyhow, one crucial observation should be made: S3.1 obtains the 1st rank with 85% of the possible criteria measurements, while S4.1 achieves the same with only 15% of the measurements.

This observation confirms what has been reported by Lahdelma and Salminen [8]: thresholds cannot be used to model imprecision. In the original study the thresholds were used for modelling imprecision in the criteria measurements, and the analysts ended up in recommending S4.1 as the most “robust” option. But by considering the criteria measurements to have uniformly distributed values in an imprecision interval, S3.1 seems to be more robust candidate for the first rank.

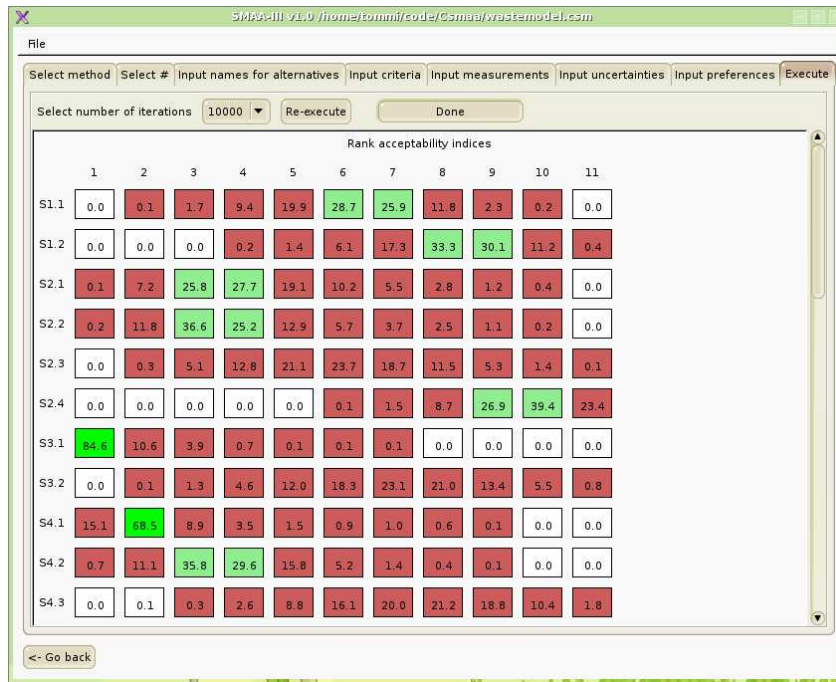


Figure 10. Rank acceptability indices of the re-analysis with imprecise criteria measurements.

#### 4.5 All values imprecise

The last re-analysis is with all types of imprecisions applied in the previous re-analyses. Rank acceptability indices and pair-wise winning indices of this one are similar to the previous re-analyses, showing (as expected) even more dispersion in the values. Therefore, we will not present these indices here.

The incomparability indices of this re-analysis are presented in Figure 11. What should be noticed from this figure, is the high amount of incomparability. In a real decision making situation, most (if not all) parameter values are defined with imprecise

cise values. This imprecision causes a quite large share of the pairs of alternatives to have incomparability indices of a reasonable magnitude. They can therefore not be left out of consideration is the DMs want to make the decision taking into account incomparability as well, and it brings an extra level of complexity to the decision making process. In our opinion, this is another reason why modelling incomparability is not useful in practical decision making.

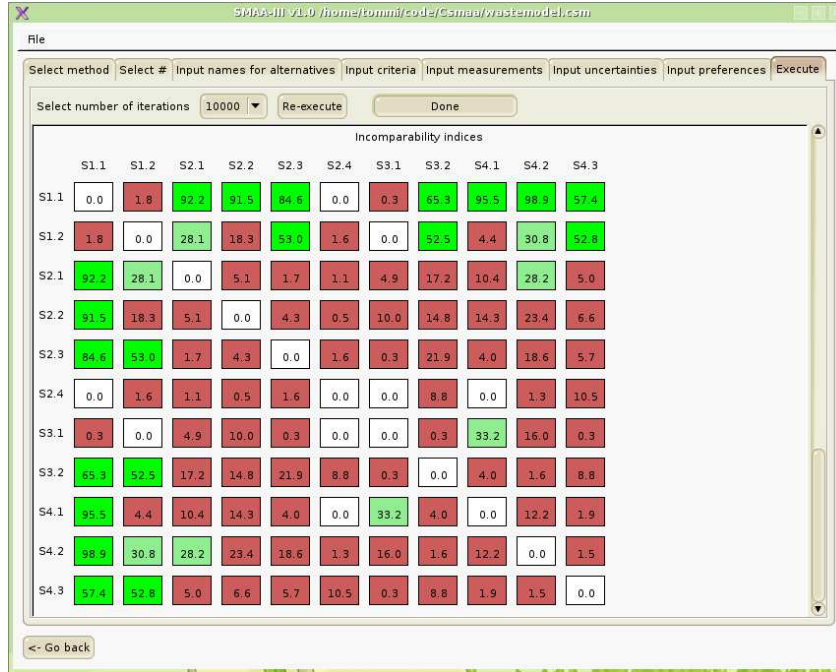


Figure 11. Incomparability indices of the re-analysis with all parameters imprecise.

Before summarizing the results of this section, we note that the authors of the case study also describe a robustness analysis of the results. But in their study, the robustness is analyzed through weights variations, by altering a single weight at a time. It is more a sensitivity analysis than a robustness one. As has been shown by our re-analyses, a different way should be considered to analyze robustness of ELECTRE III results; stability of all parameters should be analyzed in such a way, that the whole space of feasible parameter values is explored. Otherwise, nonlinearity of the ranking function, which ELECTRE III represents, can produce surprising results. The re-analyses of this particular case study showed that S3.1 would have been a more robust alternative than S4.1 for implementation.

## 5 Conclusions and avenues for future research

In this paper we introduced a new method, SMAA-III, that allows the parameters and criteria measurements of ELECTRE III to be imprecise, and to be defined with various types of constraints: no deterministic values are required. This has numerous advantages, especially in the context of MCDM with multiple DMs, because

the parameters can be determined as intervals that contain the preferences of all DMs. It also allows an easily applicable robustness analysis to be performed.

We presented a software<sup>1</sup> implementing the proposed method, and used it to re-analyze multiple times an existing real-world case study. These re-analyses were done to study the effect of imprecision in different parameters on the results. The analyses showed that in this case all the parameters of ELECTRE III were sensitive for reasonable changes. This confirms results of the simulation study by Lahdelma and Salminen [8]: pseudo-criterion based models are unstable with respect to changes in the threshold values and indifference threshold cannot be used to model imprecision in the data.

Future research should study usefulness of SMAA-III in real-life cases. It should be studied, if the indices of SMAA-III can be interpreted in a meaningful way by analysts less accustomed to SMAA methods. In addition, new techniques for visualizing the indices are needed.

## Acknowledgements

The work of Tommi Tervonen was supported by grants from Turun Yliopistosäätiö and the Finnish Cultural Foundation.

## References

- [1] V. Belton, T. J. Stewart, *Multiple Criteria Decision Analysis - An Integrated Approach*, Kluwer Academic Publishers, Dordrecht, 2002.
- [2] J. Butler, J. Dia, J. Dyer, Simulation techniques for the sensitivity analysis of multi-criteria decision models, *European Journal of Operational Research* 103 (3) (1997) 531–545.
- [3] J. Figueira, B. Roy, Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure, *European Journal of Operational Research* 139 (2002) 317–326.
- [4] J. Figueira, V. Mousseau, B. Roy, ELECTRE methods, in: J. Figueira, S. Greco, M. Ehrgott (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer Science+Business Media, Inc., New York, 2005, Ch. 4.
- [5] J. Hokkanen, P. Salminen, Choosing a solid waste management system using multicriteria decision analysis, *European Journal of Operational Research* 98 (1997) 19–36.

---

<sup>1</sup> The software presented in this paper is available from the corresponding author. It is distributed free for academic use.

- [6] R. Keeney, H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley & Sons, New York, 1976.
- [7] R. Lahdelma, P. Salminen, SMAA-2: Stochastic multicriteria acceptability analysis for group decision making, *Operations Research* 49 (3) (2001) 444–454.
- [8] R. Lahdelma, P. Salminen, Pseudo-criteria versus linear utility function in stochastic multi-criteria acceptability analysis, *European Journal of Operational Research* 14 (2002) 454–469.
- [9] P. Leskinen, J. Viitanen, A. Kangas, J. Kangas, Alternatives to incorporate uncertainty and risk attitude in multicriteria evaluation of forest plans, *Forest Science* 52 (3) (2006) 304–312.
- [10] V. Mousseau, Eliciting information concerning the relative importance of criteria, in: Y. Pardalos, C. Siskos, C. Zopounidis (Eds.), *Advances in Multicriteria analysis*, Kluwer Academic Publishers, Dordrecht, 1995, pp. 17–43.
- [11] M. Rogers, M. Bruen, A new system for weighting environmental criteria for use within ELECTRE III, *European Journal of Operational Research* 107 (1998) 552–563.
- [12] M. Rogers, M. Bruen, L.-Y. Maystre, *ELECTRE and Decision Support: Methods and Applications in Engineering and Infrastructure Investment*, Kluwer Academic Publishers, Dordrecht, Massachusetts, U.S.A., 2000.
- [13] B. Roy, M. Présent, D. Silhol, A programming method for determining which Paris metro stations should be renovated, *European Journal of Operational Research* 24 (1986) 318–334.
- [14] P. Smets, Varieties of ignorance and the need for well-founded theories, *Information Sciences* 57–58 (1991) 135–144.
- [15] T. Tervonen, J. R. Figueira, A survey on stochastic multicriteria acceptability analysis methods, *Journal of Multi-Criteria Decision Analysis* (to appear).
- [16] T. Tervonen, R. Lahdelma, Implementing stochastic multicriteria acceptability analysis, *European Journal of Operational Research* 178 (2) (2007) 500–513.