Instituto de Engenharia de Sistemas e Computadores de Coimbra

Institute of Systems Engineering and Computers

INESC - Coimbra

Tervonen, T., Figueira, J., Lahdelma, R., Salminen, P.

Modelling MCDA Group Preferences for Public Human Resource Management

Evaluating the Quality of Education at The Department of Information Technology, The University of Turku (Finland)

No. 22 2004

ISSN: 1645-2631

Instituto de Engenharia de Sistemas e Computadores de Coimbra
INESC - Coimbra
Rua Antero de Quental, 199; 3000-033 Coimbra; Portugal
www.inescc.pt

MODELLING MCDA GROUP PREFERENCES FOR PUBLIC HUMAN RESOURCE MANAGEMENT

Evaluating the Quality of Education at The Department of Information Technology, The University of Turku (Finland)

Tommi Tervonen [*][†][‡], José Figueira [†][‡], Risto Lahdelma [*], Pekka Salminen [§]

20th December 2004

---

[*] Department of Information Technology, University of Turku, FIN-20520 Turku, Finland. E-mails: `tommi.tervonen@it.utu.fi`, `risto.lahdelma@cs.utu.fi`

[†] Faculdade de Economia da Universidade de Coimbra, Av. Dias de Silva 165, 3004-512 Coimbra, Portugal, Phone: (+351) 239 790 500, Fax: (+351) 239 790 514. E-mail: `figueira@fe.uc.pt`

[‡] INESC–Coimbra, R. Antero de Quental 199, 3000-033 Coimbra, Portugal

[§] University of Jyväskylä, School of Business and Economics, P.O. Box 35, FIN-40014 Jyväskylä, Finland. E-mail: `psalmine@econ.jyu.fi`

Contents

MODELLING MCDA GROUP PREFERENCES FOR PUBLIC HUMAN RESOURCE MANAGEMENT

Evaluating the Quality of Education at The Department of Information Technology, The University of Turku (Finland)

Abstract

Modern managerial decision making problems are discrete and multicriteria by nature, and involve several decision makers (DMs). One of the key questions in this type of problems is how the preferences of the DMs can be modelled. Usually the DMs are not sure of their preferences or will not tell them to the analyst, because they are not able to express their preferences directly. In these type of situations the decision support system should allow modelling of ignorance. Another key question is related to the treatment of conflicting preferences.

Stochastic Multicriteria Acceptability Analysis (SMAA) is a family of methods to aid DMs in discrete decision aiding problems. Dempster-Shafer theory of evidence (DST) allows modelling of ignorance by using belief functions. In this paper we show how the preferences of multiple DMs or other stakeholders can be modelled and combined using DST, and how this information can then be encoded as interval constraints for sets of weights in SMAA. We address both of the key questions: how to model ignorance and how to resolve conflicting preferences. Thus, the theoretical part of this work extends the SMAA methodology.

Our contribution has a wide range of practical applications. We present a real-world case study of human resource management in the public sector. It consists of evaluating the quality of education in the University of Turku, Finland. We analyze the performance of 10 courses in terms of 6 criteria, based on client satisfaction: we take into account the collective preferences elicited from a group of 58 students by applying the proposed method.

Keywords: *Stochastic Multicriteria Acceptability Analysis (SMAA); Multiple Criteria Decision Aiding (MCDA); Dempster-Shafer theory (DST); Preference Modelling; Group Decision Making; Human Resource Management*

1 Introduction

In modern managerial decisions, there exists a large amount of decision making situations, which are discrete (they comprise a finite set of alternatives) and multicriteria by nature. Situations of this type include choosing, ranking, and sorting problem statements or problematics (see Figueira et al., 2005, which provides also a wide selection of important methodologies to attack the different problem statements).

The *choosing problem* consists of selecting one alternative or a small set of alternatives as the "preferred ones", e.g. choosing a site for a landfill or harbour. In *ranking problems* each alternative needs to be attributed a position. *Sorting problems* differ from the ranking ones in that they include pre-defined categories (for example, "good", "medium", "bad"), in which alternatives need to be assigned to. Also the way the alternatives are compared is different for the other two problem statements (see Figueira et al., 2005, Ch. 4).

Managerial Multiple Criteria Decision Aiding (MCDA) problems include many practical questions which have been addressed widely in the literature. One of the most difficult questions is how the preferences of Decision Makers (DMs) or other stakeholders should be modelled? The difficulty inherent in sharing subjective views of the world has been acknowledged in science since dualism was introduced by Des Cartes (1641). The fact that MCDA methods are in general supported by mathematical models makes the problem even harder; DMs might find it too difficult to understand the method for preference elicitation.

Usually the preference information in MCDA problems is modelled by determining a weight for each criterion and other preference parameters (thresholds, catogory profiles, ...). Stochastic Multicriteria Acceptability Analysis (SMAA) is a family of decision support methods to aid DMs in discrete decision making problems with multiple DMs or other stakeholders. SMAA is a quite recent MCDA methodology, but it has had a tremendous impact for dealing with complex real-world decision making situations since 1998. SMAA methodology is based on inverse weight space analysis, which means *exploring the weight space in order to describe the weights that would make each alternative the most preferred one, of that would give a certain rank for a specific alternative*. Thus SMAA is suitable for dealing with ranking problem statements, but instead of a ranking, SMAA produces *rank profiles*, which are probabilities for alternatives to obtain different ranks.

Classical ranking methods have an input consisting of criteria measurements (evaluation matrix) and preference parameters. The output is a ranking of the alternatives according to their comprehensive utility values or scores (as in MAUT based methods), or according to some specific procedure designed to build the ranking (as in ELECTRE III). The classic methods have a drawback of requiring preference parameters. Sometimes the DMs can not provide any preference information. It might also be possible, that the DMs do not want to reveal their preferences, for example, in a political decision making of a city council. In this type of situations an inverse approach seems more suitable.

The inverse ranking approach mitigates the drawbacks of the classical approach. Input in inverse problems is the same as with classical ones, with parameters having imprecise values: criteria measurements might be uncertain and/or the DMs' preference structures might contain ignorance. In this paper ignorance means lack of information, while uncertainty means that the values are imprecise, but the underlying probability distribution is assumed to be known. Output of the inverse approach is a characterization of the problem, for example, a share of the parameter values (weights) which grant an alternative a certain rank. The inverse approach removes the need for the DMs to provide more preference information than they feel comfortable with, and thus it is very suitable for MCDA problems.

Instead of giving direct answers to the decision making problem, the SMAA methods analyze and characterize the problem, leaving the final decision for DMs to make. The main results of the analysis are *rank acceptability indices*, *central weight vectors* and *confidence factors* for different alternatives. The rank acceptability indices describe the variety of different preferences resulting in a certain rank for an alternative, the central weight vectors represent the typical preferences favouring each alternative, and the confidence factors measure whether the criteria measurements are sufficiently accurate for making an informed decision.

On the birth and development of SMAA methodology and other related methods. The weight space analysis, which forms the basis of SMAA methodology, was initially developed based on the work by Bana e Costa (1986, 1988). It is applicable to all MCDA methods where weights have to be provided.

In the original SMAA method by Lahdelma et al. (1998) the weight space analysis is performed based on an additive utility or value function and stochastic criteria measurements. The SMAA-2 method (Lahdelma and Salminen, 2001) generalized the analysis to a general utility or value function, to include various kinds of preference information and to consider holistically all ranks. The SMAA-3 method (Lahdelma and Salminen, 2002) applies ELECTRE III type pseudocriteria in the analysis. Another extension with ELECTRE III weight space analysis can be found in (Tervonen et al., 2004a). The SMAA-O method (Lahdelma et al., 2003) extends SMAA-2 for treating mixed ordinal and cardinal criteria in a comparable manner. The SMAA-A method models the preferences using reference points and achievement scalarizing functions (Lahdelma et al., 2004a). SMAA-D applies, instead of a value function, the efficiency score of Data Envelopment Analysis (DEA) (Lahdelma and Salminen, 2004).

Other methods applying the inverse approach have also been published. Charnetski (1973) and Charnetski and Soland (1978) introduced the comparative hypervolume criterion which is restricted to deterministic criteria measurements and an additive utility function. Rietveld (1980) and Rietveld and Ouwersloot (1992) presented similar methods for problems with ordinal criteria and ordinal preference information. A method similar to SMAA-A has been studied by Durbach (2004). SMAA differs from all other inverse methodologies in that it allows criteria measurements and preference information of arbitrary forms.

On the applicability of SMAA in managerial decisions. SMAA methods are applicable in many real-life managerial decision making situations for a number of reasons:

- The inverse weight space approach is suitable for many group decision making problems, where the DMs are unable or unwilling to provide preference information, or it is difficult to reach consensus over the preferences. In such cases SMAA can be used to provide descriptive information about the acceptability of different alternatives, and this can help the DMs to identify commonly suitable compromise solutions.

- SMAA supports a very general and flexible way to model different kinds of uncertain, imprecise, or inaccurate information (imperfect knowledge) (for discussion on these definitions, see e.g. Roy, 1989) through stochastic distributions.
- The SMAA computations can be implemented very efficiently through numerical methods (Lahdelma et al., 2004b), making it possible to use the method in many different decision making contexts, including interactive decision processes.

As a consequence, SMAA methods have been successfully applied in a number of real-life managerial decision making situations mainly in Finland:

1. The first SMAA application was the Helsinki Harbour Environmental Impact Assessment (EIA) procedure (Hokkanen et al., 1999). In this application, 25 alternatives were evaluated on the basis of 11 criteria. The original SMAA method (Lahdelma et al., 1998) with a linear value function was applied to the problem. The uncertain criteria measurements and their dependencies were modelled through a joint probability distribution.
2. The EIA procedure for Kirkkonummi general plan (Hokkanen et al., 1998) considered the implementation order of the confirmed plan. In this application, 7 alternatives were evaluated in terms of 8 criteria. The SMAA-3 method was applied to the problem, and this involved using an ELECTRE III like double threshold model with indifference and preference thresholds.
3. The next SMAA application was the Toukolanranta technology competition for cleaning polluted soil in Helsinki (Hokkanen et al., 2000). The competition consisted of two phases. In the first phase three candidates were chosen based on submitted tenders. In the second phase, the finalists performed a test-cleaning of a small area and the winner received the contract for cleaning the entire area. The SMAA-2 method (Lahdelma and Salminen, 2001) was used in the first phase of this competition, where 9 alternatives were evaluated on the basis of 5 criteria. The SMAA-2 method was developed in conjunction with this application, because it was necessary to consider also other ranks than the first one.
4. A location problem of a waste treatment facility in South-Eastern Finland was studied by Lahdelma et al. (2002). In this application, 4 alternatives were evaluated on the basis of 17 criteria. The interesting property of this application was that only ordinal measurements for the criteria were available. The SMAA-O method (Lahdelma et al., 2003) for treating ordinal criteria was developed in conjunction with this application.
5. A locating problem was also addressed by Lahdelma and Salminen (2003), where the SMAA-O method was used to evaluate sites for processing and storing the by-products of a multi-fuel power plant in Pietarsaari. In this application, 4 alternative sites were evaluated

on the basis of 11 criteria. The novelty of this application was that ordinal scale interval constraints were specified. These constraints set bounds for cardinal values generated by a simulation of ordinal to cardinal mapping.

6. The next application was choosing a reparation method for a landfill (Lahdelma et al., 2001). In this application, 7 alternatives were evaluated on the basis of 13 criteria. SMAA-O was applied on mixed ordinal and cardinal data; 7 criteria were measured on ordinal scales and 6 on cardinal ones. This was the first real-life SMAA application where partial preference information was used. Preference information was collected by asking the DMs to rank the criteria according to their importance.

7. In (Lahdelma and Salminen, 2005) SMAA-2 was applied in a strategic decision making problem for an electricity retailer operating on the liberalized energy market. The characteristic feature of this problem was that it is inherently continuous. The problem was discretized by choosing different value combinations of continuous decision variables. The discrete alternatives were then quantified by using a stochastic simulation model that was also able to determine the dependencies between the different criteria measurements. This application was later re-analyzed by using the SMAA-2 method with a multivariate Gaussian distribution for criteria measurements (Lahdelma et al., 2005).

8. The SMAA-O method has also been applied for re-analysing a forest ecosystem management problem (Kangas et al., 2003b). In this application, 10 alternatives were evaluated on the basis of 5 criteria. Different possible ranks for the criteria were applied in order to understand their effects on the results.

9. Another re-analysis application in the forestry area (Kangas et al., 2003a) considered the influence of the correlation between the measurement errors across different criteria and across different alternatives. In this application, 20 alternatives were evaluated on the basis of 5 criteria.

Modelling preferences of multiple DMs. In many real-world complex managerial decision making problems the preferences of multiple decision makers (DMs) need to be taken into account. As an example of multiple DMs are company shareholders or a city council. If DMs have conflicting preferences, several possible approaches are available, including solving the conflicts or modelling the collective preference structure where the conflicts are still present. The interactive conflict resolution approach (negotiation analysis) has been extensively studied (see e.g. Raiffa et al., 2002). In this paper we introduce a method for modelling the conflicts in SMAA.

The SMAA methodology allows modelling of preference information in form of probability distributions, but in practice the preference information is given in simplified form, for example, as intervals for weights or ranking of criteria. When there are multiple DMs, the preference information must be aggregated before it can be used in SMAA computations. The aggregation method may have a great impact on the results of the SMAA analysis. It is crucial that the preference information from a large set of DMs can be elicited efficiently and aggregated in a consistent and theoretically sound way.

To our knowledge, prior to this work there has been no method for modelling preferences of multiple DMs in SMAA methodology, which would take into account ignorance in the DMs' preferences structures. Nor has there been any research on combining preferences in the context of preference modelling in SMAA. A theoretically sound method for preference combination is extremely important in a decision making context, where the stakeholders are different from the actual DMs. The actual DMs might be consultants or a managerial board, while the stakeholders could be company shareholders. In this context an iterative process for refining the preferences of stakeholders is not practical, especially if the number of stakeholders is large.

Dempster-Shafer theory of evidence (DST) is a generalization of Bayesian theory of probabilities that allows representation of ignorance in the probability assignments (Shafer, 1976). In DST, it is also possible to combine the probability assignments in a theoretically sound way. Multiple alternative methods have been introduced for combining the probability assignments (see e.g. Shafer, 1976; Yager, 1987). For these reasons, DST is very suitable for modelling and aggregating the preferences of multiple DMs. In this paper we show how the DST can be used to model and aggregate the preferences of multiple DMs, and how the aggregated preference information can be applied in SMAA models.

Case study. In the case study we have a ranking problem of 10 university courses which were taught at the Department of Information Technology of the University of Turku, Finland. The students who participated in the courses have been filling evaluation forms, and using these we defined Gaussian distributed criteria measurements for the courses. The preference information was elicited by asking 58 students to prioritize the 6 criteria used in evaluating the courses by filling a form. These preference statements were transformed and aggregated by applying the method introduced in this paper. With the criteria measurements and the collective preference information, we performed SMAA analysis to rank the quality of education between the courses.

The results of the case study were studied in collaboration with the persons being in charge of education at the department. The comments from the DMs provide client's insight to the method. Particularly, the DMs considered the method (1) to give valuable support for course planning and development, (2) to be robust and reliable because no exact weights need to be provided, and (3) to be useful because it can explicitly represent uncertain, imprecise and lacking information both in criteria measurements and in preferences.

Main contributions and outline of this paper. The contributions provided by this paper are both theoretical and practical. As a theoretical advance, we present a new method for modelling and aggregating the preferences of multiple DMs. We also present how the preferences can be applied for SMAA models. From the practical point of view, we present a case study where the new method is applied. The case study brings novelty also to the variety of SMAA applications: as far as we know, the SMAA methodology has never before been applied for human resource management.

This paper is organized as follows: We describe the outline of the new methodology in Section 2. The SMAA approach is introduced in Section 3. The Dempster-Shafer theory of evidence is briefly reviewed in Section 4. In Section 5, we present the framework for applying DST for preference modelling in SMAA. In Section 6 we present and analyze a small car ranking problem for illustrative purposes. Section 7 contains a case study of a real-world problem of course evaluation. We end this paper with conclusions and avenues for future research in Section 8.

2 Outline of the improved SMAA methodology

The methodology presented in this paper is for modelling preferences of multiple DMs in SMAA. The preferences of a single DM are modelled by using belief functions as defined by DST. The approach consists of several phases, which we will briefly describe next.

1. Transform the preferences of each DM into a belief function (*bpa*). This can be accomplished, for example, using the method presented by Tervonen et al. (2004b).
2. Combine the belief functions using the Yager's rule of combination. This results in a single belief function incorporating the preferences of all DMs.
3. Calculate characteristic belief and plausibility values from the combined belief function.
4. Transform the intervals provided by the charasteristic values [belief, plausibility] to constraints for sets of weights (the preferences are modelled in SMAA using interval constraints for weights).
5. Apply SMAA, with input consisting of criteria measurements and preference information (interval constraints for weights). The output are certain indicators (confidence factors, acceptability indices, and central weight vectors) characterizing the decision making problem.

The complete process is presented in Figure 1.

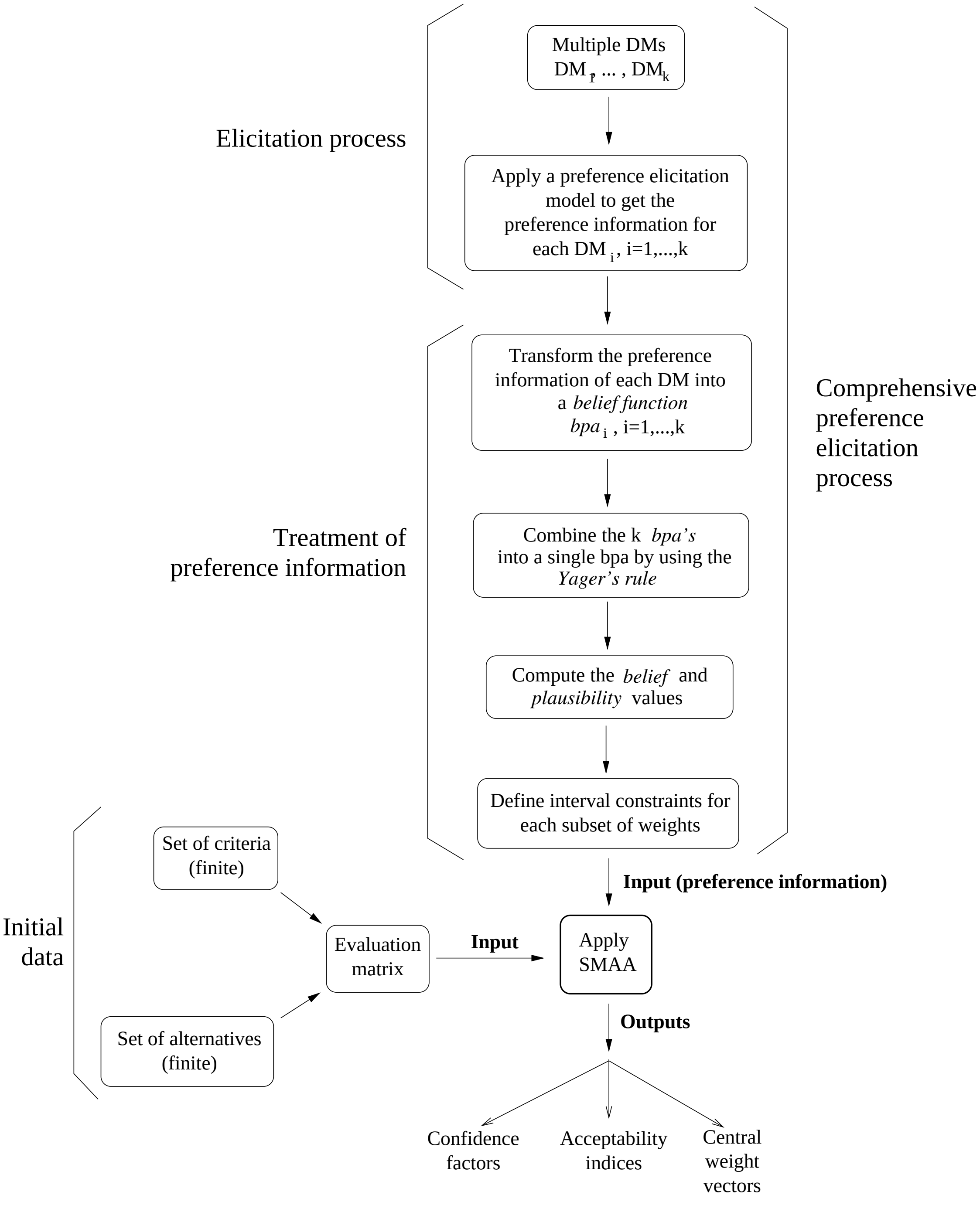

Figure 1: The procedure of using DST for preference modelling in SMAA.

3 The SMAA-2 approach

The SMAA-2 method (Lahdelma and Salminen, 2001) has been developed for discrete stochastic MCDA problems with multiple DMs. SMAA-2 applies inverse weight space analysis to describe for each alternative what kind of preferences make it the most preferred one, or place it on any particular position in the ranking.

3.1 Elementary notation

The following notation is used in this paper:

- $\{x_1, x_2, \ldots, x_m\}$ is the set of alternatives.
- $G = \{g_1, g_2, \ldots, g_n\}$ is the set or family of criteria.
- $w = (w_1, w_2, \ldots, w_n)$ is a vector of weights. The weights are represented by a weight distribution with joint density function $f_W(w)$ in the feasible weight space W.
- ξ_{ij} are stochastic variables representing imperfect knowledge of criteria measurements. They have a joint density function $f_X(\xi)$ in the space $X \subseteq R^{m \times n}$.
- $u(x_i, w)$ is the real-valued utility or value function representing the DMs preference structure. The value function maps the different alternatives to real values by using the weight vector w.

3.2 The weight space

The weight space can be defined according to needs, but typically, the weights are non-negative and normalized, that is; the weight space is an $n-1$ dimensional simplex in n dimensional space:

$$W = \left\{ w \in R^n : w \geq 0 \text{ and } \sum_{j=1}^{n} w_j = 1 \right\}. \tag{1}$$

Total lack of preference information is represented in 'Bayesian' spirit by a uniform weight distribution in W, that is, $f_W(w) = 1/\mathrm{vol}(W)$. It should be noticed here, that we use weights in the meaning of scale factors; the weights rescale the value or utility functions in such a way, that the full swing in the scaled function indicates the importance of the criterion (see Belton and Stewart, 2002, Section 5.4).

3.3 The value function

The value function is used to map the stochastic criteria and weight distributions into value distributions $u(\xi_i, w)$. Based on the value distributions, the rank of each alternative is defined as an integer from the best rank ($= 1$) to the worst rank ($= m$) by means of a ranking function

$$rank(i, \xi, w) = 1 + \sum_{k=1}^{m} \rho(u(\xi_k, w) > u(\xi_i, w)), \tag{2}$$

where $\rho(true) = 1$ and $\rho(false) = 0$. SMAA-2 is then based on analysing the stochastic sets of favourable rank weights

$$W_i^r(\xi) = \{w \in W : rank(i, \xi, w) = r\}. \tag{3}$$

Any weight $w \in W_i^r(\xi)$ results in such values for different alternatives, that alternative x_i obtains rank r.

3.4 Descriptive measures provided by SMAA-2

Three types of descriptive measures are obtained as outputs from a SMAA-2 analysis: (1) rank acceptability indices, (2) central weight vectors, and (3) confidence factors.

1. The *rank acceptability index (b_i^r)* measures the variety of different preferences that grant alternative x_i rank r. It is the share of all feasible weights 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 criteria distributions and the favourable rank weights as

   $$b_i^r = \int_{\xi \in X} f_X(\xi) \int_{w \in W_i^r(\xi)} f_W(w)\, dw\, d\xi. \tag{4}$$

   The most acceptable (best) alternatives are those with high acceptabilities for the best ranks. Evidently, the rank acceptability indices are in the range [0,1], where 0 indicates that the alternative will never obtain a given rank and 1 indicates that it will obtain the given rank always with any choice of weights.

   The first rank acceptability index b_i^1 is called the *acceptability index a_i*. The acceptability index is particularly interesting, because it is nonzero for stochastically efficient alternatives (alternatives that are efficient with some values for the stochastic criteria measurements) and zero for inefficient alternatives. The acceptability index not only identifies the efficient alternatives, but also measures the strength of the efficiency considering simultaneously the uncertainty in criteria measurements and the ignorance in DMs' preferences.

2. The *central weight vector (w_i^c)* is the expected centre of gravity (centroid) of the favourable first rank weights of an alternative. The central weight vector represents the preferences of a "typical" DM supporting this alternative. The central weights of different alternatives can be presented to the DMs in order to help them understand how different weights correspond to different choices with the assumed preference model. The central weight vector w_i^c is computed numerically as a multidimensional integral over the criteria distributions and the favourable first rank weights using

   $$w_i^c = \int_{\xi \in X} f_X(\xi) \int_{w \in W_i^1(\xi)} f_W(w) w\, dw\, d\xi / a_i. \tag{5}$$

3. The *confidence factor (p_i^c)* is the probability for an alternative to obtain the first rank when the central weight vector is chosen. The confidence factor is computed as a multidimensional integral over the criteria distributions using

$$p_i^c = \int_{\xi \in X : rank(i,\xi,w_i^c)=1} f_X(\xi)\, d\xi. \tag{6}$$

   Confidence factors can similarly be calculated for any given weight vectors. The confidence factors measure whether the criteria measurements are accurate enough to discern the efficient alternatives.

3.5 Additional theoretical issues

There are several different ways to handle partial preference information in SMAA methods (Lahdelma and Salminen, 2001). To our best knowledge, prior to this work there have been no methods, which allow DMs to express ignorance in their preference structures. The DST allows modelling of ignorance, and in this article we will present a novel method for modelling preference information in SMAA by applying DST. Using DST the information is aggregated, and modelled in SMAA as L interval constraints for sums of the subsets of weights. The constraints are given as

$$c_\ell^{min} \leq \sum_{j \in C_\ell} w_j \leq c_\ell^{max}, \forall \ell = 1, \ldots, L, \tag{7}$$

where C_ℓ is a set of criteria in constraint ℓ. The weight space analysis of SMAA is then performed in the restricted weight space

$$W' = \{w \in W | w \text{ satisfies (7)}\}. \tag{8}$$

This means that the uniform weight distribution $f_W(w)$ is redefined as

$$f_W(w) = \begin{cases} 1/\mathrm{vol}(W'), & \text{if } w \in W', \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Figure 2 illustrates the restricted weight space of a 3-criterion problem with a lower and an upper bound for w_1.

4 The Dempster-Shafer theory (DST) of evidence

The DST is an extension of the classical Bayesian theory of probabilities, and it allows modelling of ignorance. But DST can also be used to represent other kind of information than probabilities. For example, in this paper we apply DST to model and to aggregate preference structures.

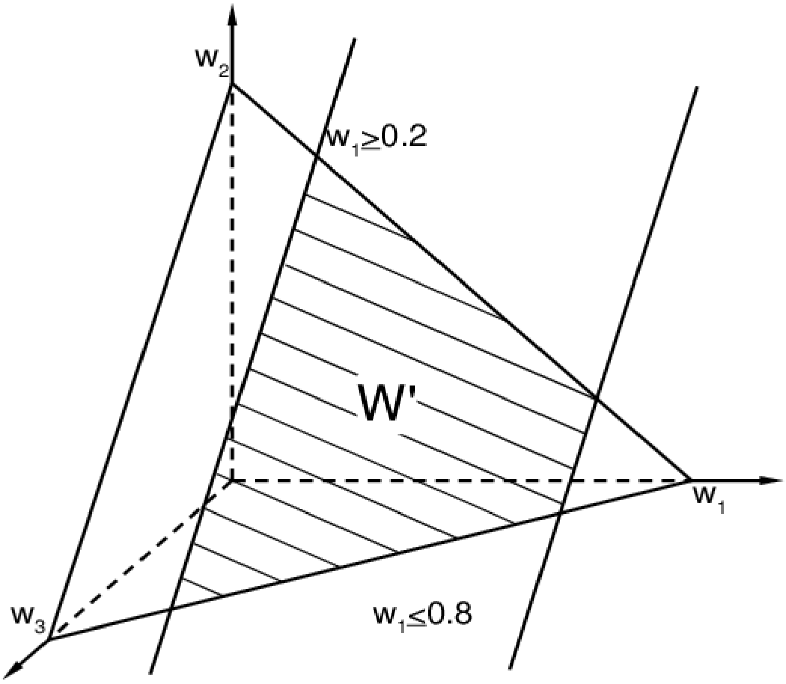

Figure 2: Feasible weight space of a 3-criterion problem with a lower and an upper bound for w_1, $0.2 \leq w_1 \leq 0.8$.

4.1 The basis of the theory

In the classical Bayesian theory, the probabilities are considered to be objective. This fact stems from the definition of probability: the relative frequency at which an event occurs with. This type of definition does not allow modelling of ignorance, which is important especially when modelling preferences of multiple DMs in the context of MCDA.

The DST extends the classical Bayesian theory of probabilities using belief functions. Instead of assigning probabilities to propositions, the belief functions assign probability masses to subsets of the propositions in the *frame of discernment* (denoted by Θ). The frame of discernment is the set of all propositions.

4.2 An illustrative example

Let us illustrate this by an example. Consider a situation where we have to decide whether a certain chemical substance is harmful to humans. Let:

- p_1 denote the proposition “substance is harmful”,
- p_2 denote the proposition “substance is not harmful”.

Most people do not possess sufficient knowledge required for an informed judgment between the proposition. According to Bayesian theory, in absence of knowledge a probability of 0.5 is assigned to both p_1 and p_2. If DST is applied, a probability of 0 is assigned to both p_1 and p_2, and 1 to the set Θ representing ignorance. Now suppose we have a third alternative,

- p_3 denoting the proposition "substance is slightly harmful for humans".

With the modified frame of discernment the Bayesian probabilities would be

$$P(p_1) = P(p_2) = P(p_3) = 0.33,$$

and the DST probabilities

$$P(p_1) = P(p_2) = P(p_3) = 0, P(\Theta) = 1.$$

The Bayesian probabilities thus strongly depend on the frame of discernment, and by looking at the example it is evident that DST is more consistent in assignment of the probability masses.

Based on the previous example, it should also be noted, that when the Bayesian approach is applied, knowledge and ignorance are indistinguishable. If DST is applied, ignorance can be modelled, and in addition certain metrics can be used for calculating the total amount of ignorance in the belief structure (see Shafer, 1976).

4.3 The basic probability assignment (bpa) functions

DST terminology defines a *basic probability assignment* (*bpa*) function as a belief function that assigns probabilities that sum to unity to the propositions and sets of propositions in the frame of discernment. It is defined as

$$\begin{aligned} m &: 2^{\Theta} \mapsto [0,1], \\ m(\varnothing) &= 0, \\ \sum_{A \subseteq \Theta} m(A) &= 1, \end{aligned} \tag{10}$$

where 2^{Θ} is the powerset of the frame of discernment containing all subsets of Θ (including $\varnothing$ and Θ). The *bpa*'s are also called *pieces* or *bodies of evidence*. If there are multiple *bpa*'s, they can be combined using a combination rule. The Dempster's rule of combination is (Shafer, 1976)

$$\begin{aligned} m(\varnothing) =& 0, \\ m(A) =& \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - K}, \text{ if } A \neq \varnothing, \end{aligned} \tag{11}$$

where

$$K = \sum_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j), \text{ and} \tag{12}$$

m_1 and m_2 are bpa's, m is the combined *bpa*, and $A_1, \ldots, A_k$ and $B_1, \ldots, B_l$ are their focal elements (subsets), respectively.

4.4 The weight of conflict and the Yager's rule

Let us consider K denoting the *weight of conflict*, which measures the conflict between two *bpa*'s. In the Dempster's rule, the weight of conflict is used for normalizing, meaning that it is distributed among all subsets of propositions. This approach has the downside of distributing the conflict, without any particular reason, to all sets of propositions. In some situations the alternative Yager's rule of combination may be more suitable (Yager, 1987):

$$\begin{aligned} m(\varnothing) &= 0, \\ m(A) &= \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j), \text{ if } A \neq \varnothing, \\ m(\Theta) &= \left(\sum_{A_i \cap B_j = \Theta} m_1(A_i) m_2(B_j) \right) + K. \end{aligned} \tag{13}$$

By using Yager's rule of combination, the weight of conflict is added to the set representing all propositions (the frame of discernment), so the conflict is added to the ignorance represented in the belief structure. We will use Yager's rule when applying DST for combining preferences in SMAA. There are also other ways of allocating conflict, for example, registering it to the empty set (Smets, 1990).

4.5 On the notion of belief and plausibility

After all available pieces of evidence have been combined, DST allows characteristic values to be calculated from the bpa's. The two most important are belief and plausibility. *Belief* in a set of propositions A measures the probability mass that is assigned to A or any subset of A. It measures the confidence we have in A, and is defined as

$$Bel(A) = \sum_{B \subseteq A} m(B), \text{ for all } A \subseteq \Theta. \tag{14}$$

Plausibility in a set of propositions A measures the probability mass that is assigned to sets that have common elements with A. It is the amount we fail to disbelieve A, and is defined as

$$Pls(A) = \sum_{B \cap A \neq \varnothing} m(B), \text{ for all } A \subseteq \Theta. \tag{15}$$

$[Bel(A), Pls(A)]$ is thus the interval for the "true" probability of A when ignorance is taken into account. Ignorance is present not only in basic probability assignments to Θ, but also in those to sets of multiple propositions. For example, assignment of probability 1.0 to a set of propositions $\{p_1, p_2\}$ means that either of the propositions is certainly true, but there exists no information about which one it is.

5 DST for preference modelling in SMAA

The DST can be used to model preferences of multiple DMs by considering the preferences of each DM as a body of evidence. This means that from the preferences of a DM, we have to form a *bpa* representing her/his preferences and ignorance. In this *bpa*, propositions will be the criteria $\{g_1, \dots, g_n\}$. After this, all *bpa*'s of different DMs are combined. From the combined *bpa* we calculate belief and plausibility values for all subsets of weights. The aggregated preferences are then represented as interval constraints [Bel, Pls] for the sums of subsets of weights.

Eliciting the preferences and transforming them into *bpa*'s can be accomplished in multiple ways. The method presented by Tervonen et al. (2004b) uses a technique (inspired by Beynon et al., 2000) to elicit the preferences and to transform them to *bpa*'s. This method is applied in the case study of Section 7.

When applying DST for modelling preferences in SMAA, it may be preferable to apply the Yager's rule of combination (13), instead of the Dempster's rule (11). The idea of Yager's rule is to assume nothing about the nature of conflict. Yager's rule adds the ignorance to Θ instead of distributing it between all elements. Thus, in presence of conflict, Yager's rule enlarges all the intervals [Bel,Pls], while Dempster's rule diminishes them. Because SMAA is designed to perform the analysis using all possible preferences, it is better to apply too wide intervals (and not to assume anything extra) than too narrow ones (and assume that the conflict can be distributed among all subsets). This makes Yager's rule more suitable in conjunction with SMAA.

It must be noticed here, that the following links exist between the notations of DST and MCDA:

- $p_j = g_j$, meaning that the MCDA criteria are on the role of DST propositions, and
- $\Theta = G$, which means that the set of all criteria now replaces the set of all propositions.

6 An illustrative example

The following small example illustrates the use of DST for modelling preferences of multiple DMs in SMAA.

6.1 The initial data

The decision making problem consists of four cars $\{x_1, x_2, x_3, x_4\}$ from which a group of four equally important DMs $\{\mathrm{DM}_1, \mathrm{DM}_2, \mathrm{DM}_3, \mathrm{DM}_4\}$ have to choose one. The cars are evaluated based on four criteria:

- style (g_1),
- price (g_2),
- fuel consumption (g_3), and
- comfort (g_4).

The cardinal criteria measurements for the alternatives are presented in Table 1. The criteria are Gaussian distributed with standard deviation 10% of the mean value. Based on the criteria measurements, choosing the car depends completely on the preferences of the DMs: If price is preferred over all the remaining criteria, then alternative x_2 should be chosen, etc. Without any preference information an informed decision cannot be made.

Table 1: Criteria measurements for alternative cars (mean ± standard deviation).

Alternative	Style (g_1)	Price (g_2)	Fuel (g_3)	Comfort (g_4)
x_1	2 ± 0.2	1 ± 0.1	1 ± 0.1	1 ± 0.1
x_2	1 ± 0.1	2 ± 0.2	1 ± 0.1	1 ± 0.1
x_3	1 ± 0.1	1 ± 0.1	2 ± 0.2	1 ± 0.1
x_4	1 ± 0.1	1 ± 0.1	1 ± 0.1	2 ± 0.2

6.2 Determining and combining the bpa's

DST allows DMs to express their preferences by identifying individual criteria or sets of criteria that they consider important with respect to G. There are different ways how the preference statements can be elicited and transformed into *bpa*'s. In this example we have used the method presented by Tervonen et al. (2004b) for eliciting the preferences of DMs and for transforming them into *bpa*'s. We will not present the original preferences for brevity. After the transformation the following four *bpa*'s are obtained:

- $m_1(Style) = 0.63733$, $m_1(\{Price, Fuel\}) = 0.21244$, $m_1(G) = 0.15022$,
- $m_2(Fuel) = 0.83333$, $m_2(G) = 0.16667$,
- $m_3(\{Style, Price\}) = 0.8$, $m_3(G) = 0.2$,
- $m_4(\{Style, Comfort\}) = 0.85714$, $m_4(G) = 0.14286$.

For example, from the *bpa* m_1 it can be seen, that DM_1 gives a lot of weight to Style (0.63733), can not make a distinction between Price and Fuel, but thinks they are both quite important (weight of 0.21244), and has some ignorance present in her/his decisions ($m_1(G) = 0.15022$).

Combining the different DMs' *bpa*'s using the Yager's rule of combination (13) results in a *bpa* for the group of DMs. From this *bpa* we can calculate the belief and plausibility values using (14) and (15). The *bpa* of the group and the calculated belief and plausibility values are presented in Table 2.

Table 2: The combined bpa's and the belief and plausibility values for the example.

C_ℓ	$m(C_\ell)$	Bel(C_ℓ)	Pls(C_ℓ)
Style	0.015	0.015	0.272
Price	0.416	0.416	0.976
Fuel	0.009	0.009	0.202
Comfort	0.000	0.000	0.496
Style, Price	0.064	0.494	0.991
Style, Fuel	0.000	0.024	0.282
Style, Comfort	0.000	0.015	0.575
Price, Fuel	0.001	0.425	0.985
Price, Comfort	0.303	0.718	0.976
Fuel, Comfort	0.000	0.009	0.506
Style, Price, Fuel	0.000	0.504	1.000
Style, Price, Comfort	0.000	0.797	0.991
Style, Fuel, Comfort	0.000	0.024	0.584
Price, Fuel, Comfort	0.000	0.728	0.985
G	0.193	1.000	1.000

6.3 The SMAA model and analysis

Next we define constraints for weights (as in (7)) using the obtained belief and plausibility values. For each subset of weights C_ℓ, we define lower and upper bounding constraints for their sum as follows:

$$Bel(C_\ell) \leq \sum_{w_j \in C_\ell} w_j \leq Pls(C_\ell).$$

For example, the sum of weights for style, price and fuel criteria is constrainted within the range [0.504, 1],

$$0.504 \leq w_{style} + w_{price} + w_{fuel} \leq 1.$$

We ran SMAA analysis using the criteria values presented in Table 1 and the preference information. The scale for all criteria was normalized so that the interval [0.5,2.5] becomes [0,1]. The SMAA was executed using 100000 Monte Carlo iterations. The confidence factors and the rank acceptability indices are presented percentage-wise in Table 3, sorted by their (first rank) acceptability index. The rank acceptability indices are also presented graphically in Figure 3.

6.4 Comments on the results

From the results the effect of the preference information is evident. Alternative x_2, which has a larger value for the style criterion, has a large confidence factor and a high value for the acceptability index (b^1). As the preferences of the DMs are already present in the model, x_2 should be recommended as the car to buy. In contrast, alternatives x_1 and x_3, which have largest values for style and fuel criteria, respectively, obtain relatively small confidence factors and large acceptabilities for the worst rank (b^4). As a consequence, neither of those should be recommended if a compromise solution is sought. The combined preferences thus emphasize price more than comfort, and comfort more than style or fuel consumption.

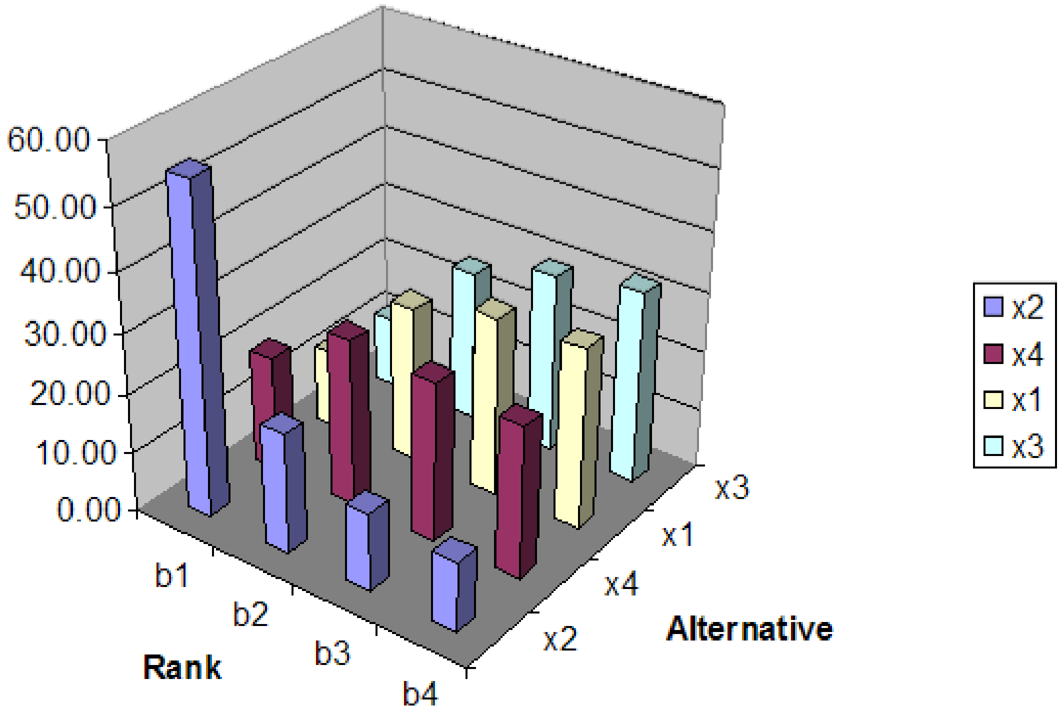

Figure 3: Rank acceptability indices for the alternatives (in %).

The central weight vectors are presented percentage-wise in Table 4. Because the weights are limited by interval constraints, the central weights differ only by small amounts. Nevertheless, it can be noticed that alternative x_2 is preferred by DMs that value style, and alternative x_4 by DMs that value comfort. This is consistent with our knowledge that alternative x_2 has high value for style and x_4 for comfort.

Table 3: Confidence factors and rank acceptabilities for the alternatives (in %). The highest acceptability for each rank appear in boldface, and the lowest italized.

Alt	p^c	b^1	b^2	b^3	b^4
x_2	57.97	**55.11**	*20.15*	*13.08*	*11.66*
x_4	22.68	19.70	**28.09**	26.76	25.46
x_1	13.90	13.25	26.27	30.03	30.46
x_3	12.52	*11.94*	25.49	**30.14**	**32.42**

Table 4: Central weight vectors for the alternatives (in %).

Alt	Style (g_1)	Price (g_2)	Fuel (g_3)	Comfort (g_4)
x_2	9.98	64.33	8.13	17.57
x_4	9.99	57.72	7.85	24.44
x_1	11.78	61.20	7.69	19.33
x_3	9.75	61.82	9.43	19.00

7 Case study: Evaluating the quality of education at the University of Turku

At the Department of Information Technology of the University of Turku, there has been a custom of collecting course feedback since 2001. After each course, students fill a form in which they rate different aspects of the course on a 5-point scale (1–5). Among the aspects rated are, for example, the lecturing skills of the teacher and the quality of the course material.

The objective of this study is to analyze a selection of 10 courses taught during the years 2003 and 2004. A grouping of the courses is needed, because we wish to provide recommendations for the staff of the department on how the quality of the education could be improved. The case study applies the method presented earlier in this paper for modelling collective preferences of the students. Also the criteria measurements are based on student evaluations. The courses are thus ranked according to *the collective opinions of the "clients"*.

7.1 Study data

We have chosen to evaluate the following courses:

- x_1 = Programming I (P1),
- x_2 = Programming II (P2),
- x_3 = Data Structures and Algorithms (DA),
- x_4 = Databases (DB),
- x_5 = Introduction to Computers (IC),
- x_6 = Inauguration of Information Systems (IIS),
- x_7 = Introduction to Computer Science I (ICS1),
- x_8 = Introduction to Computer Science II (ICS2),
- x_9 = Modelling of Information Systems II (MIS2), and
- x_{10} = Microprocessors (MP),

based on the following criteria:

- g_1 = Lecturers Level of Expertise (LLE),
- g_2 = Teaching Skills of Lecturer (LTS),
- g_3 = Usefulness of Lecture Material (LMU),
- g_4 = Demonstrators Level of Expertise (DLE),
- g_5 = Teaching Skills of Demostrator (DTS), and
- g_6 = Demonstrators Activity in Assuring Comprehension of Exercises (DAC).

The course criteria measurements are modelled as Gaussian distributed, with means and standard deviations calculated from the scores given by students in the evaluation forms. The criteria measurements are presented in Table 5 together with the number of evaluations the values are calculated from.

Table 5: Criteria measurements and the number of evaluations they are calculated from (in parenthesis) for the case study (mean ± standard deviation).

Course	LLE (g_1)	LTS (g_2)	LMU (g_3)
x_1 (P1),	4.42 ± 0.63 (106)	4.08 ± 0.77 (106)	3.38 ± 0.92 (106)
x_2 (P2)	4.25 ± 0.68 (96)	3.88 ± 0.78 (96)	3.68 ± 0.77 (96)
x_3 (DA)	4.37 ± 0.61 (76)	4.03 ± 0.75 (76)	4.11 ± 0.78 (76)
x_4 (DB)	4.59 ± 0.51 (17)	3.65 ± 0.93 (17)	3.71 ± 0.99 (17)
x_5 (IC)	4.03 ± 0.84 (36)	3.39 ± 0.99 (36)	3.74 ± 0.83 (34)
x_6 (IIS)	4.41 ± 0.59 (22)	3.45 ± 1.06 (22)	2.95 ± 1.02 (21)
x_7 (ICS1)	3.90 ± 0.78 (99)	3.04 ± 0.91 (102)	3.67 ± 0.90 (123)
x_8 (ICS2)	3.95 ± 0.60 (44)	2.82 ± 0.83 (44)	3.69 ± 0.83 (54)
x_9 (MIS2)	4.26 ± 0.54 (65)	3.89 ± 0.63 (63)	3.47 ± 0.87 (64)
x_{10} (MP)	3.81 ± 0.60 (21)	2.86 ± 0.99 (22)	3.36 ± 0.95 (22)
Course	DLE (g_4)	DTS (g_5)	DAC (g_6)
x_1 (P1)	3.92 ± 0.84 (106)	3.44 ± 1.05 (106)	3.49 ± 0.97 (106)
x_2 (P2)	4.14 ± 0.59 (96)	3.82 ± 0.71 (96)	3.72 ± 0.71 (96)
x_3 (DA)	4.32 ± 0.62 (76)	4.16 ± 0.63 (76)	4.07 ± 0.74 (76)
x_4 (DB)	4.41 ± 0.51 (17)	3.94 ± 0.66 (17)	3.88 ± 0.78 (17)
x_5 (IC)	4.33 ± 0.53 (36)	4.14 ± 0.69 (35)	3.63 ± 0.84 (35)
x_6 (IIS)	3.77 ± 0.87 (22)	3.27 ± 1.08 (22)	3.27 ± 1.12 (22)
x_7 (ICS1)	4.10 ± 0.65 (129)	3.78 ± 0.83 (130)	3.81 ± 0.84 (129)
x_8 (ICS2)	4.05 ± 0.53 (60)	3.66 ± 0.76 (58)	3.81 ± 0.78 (58)
x_9 (MIS2)	4.00 ± 0.55 (66)	3.86 ± 0.68 (66)	3.59 ± 0.70 (66)
x_{10} (MP)	3.91 ± 0.61 (22)	3.48 ± 0.79 (23)	3.00 ± 1.00 (23)

7.2 Preference information

For eliciting the preference information, a group of 58 students filled a questionnaire rating the criteria used in the study. The questionnaire listed all criteria, and students had to rate them on a 6-point scale (1–6). Half of the questionnaires had the first and last three criteria in inverse order (the 3 criteria concerning lecturer before the criteria concerning demostrator or *vice-versa*). Using the answers, we defined 58 bpa's using the method presented in (Tervonen et al., 2004b). We combined these bpa's using the Yager's rule, which provided as a result a single bpa. This bpa was then used to define lower- and upper bounds for sets of weights as in the illustrative example presented in Section 6. The components of the bpa where $m(\cdot) \geq 0.001$ with corresponding weight limits are presented in Table 6. The other components with smaller mass assignments are not presented for brevity.

Table 6: Components of the bpa of the case study with mass assignment ≥ 0.001

A	bpa	Bel(A) = w_{min}	Pls(A) = w_{max}
{LLE}	0.072	0.072	0.561
{LTS}	0.211	0.211	0.625
{LMU}	0.024	0.024	0.501
{DLE}	0.034	0.034	0.517
{DTS}	0.028	0.028	0.526
{DAC}	0.052	0.052	0.455
{LLE,DLE}	0.001	0.107	0.675
{LLE,DTS}	0.086	0.186	0.600
{LTS,DLE}	0.013	0.258	0.728
{LMU,DLE}	0.066	0.124	0.550
{LMU,DTS}	0.008	0.061	0.617
{DTS,DAC}	0.001	0.082	0.578
{LLE,LMU,DLE}	0.001	0.199	0.707
G	0.401	1.0	1.0

7.3 The SMAA model

The SMAA model was constructed with scaling of the value function by having 0-point at criterion value 1.0, and 1-point at criterion value 5.0. The scaling has a large impact on the results and the shape of the utility function should be accepted by the DMs. In this case study the scaling was done according to the natural scale of the evaluations: minimum score for evaluating a criterion of a course is 1, and the maximum 5.

We performed SMAA computations with 100 000 Monte Carlo iterations. The confidence factors and rank acceptability indices are presented percentage-wise in Table 7. The central weight vectors are presented percentage-wise in Table 8. The rank acceptability indices are illustrated graphically in Figure 4.

The strict weight intervals have a visible effect on the acceptability indices: the acceptability indices (b^1) are almost the same as confidence factors (p^c). If the weights are deterministic, the two will be equal.

The effect of very imprecise data has an influence on the results: the largest of the confidence factors (for course DA) is only 17.81. The reason for this can be found from the large standard deviations of the criteria measurements. Nevertheless, the results of the SMAA analysis are useful for making conclusions. The imprecision is an inherent attribute of the model, because we are modelling opinions of a large group of heterogenous DMs. When inspecting the results, we should bear in mind, that we have applied a model where ignorance and conflicting opinions are modelled in the last phase (SMAA run) as imprecision. As the preference information was elicited from 58 students, and the criteria measurements modelled based on hundreds of course evaluation forms, we should not expect results which are precise or easy to interpret.

Table 7: Rank acceptability indices and confidence factors of the case study (in %), sorted in descending order according to the confidence factors. The highest rank acceptability for each rank appear in boldface and the lowest italized.

Course	p^c	b^1	b^2	b^3	b^4	b^5	b^6	b^7	b^8	b^9	b^{10}
DA	17.8	**17.82**	**15.82**	**13.94**	**12.35**	10.65	9.17	*7.61*	*6.06*	*4.26*	*2.33*
DB	14.2	14.23	13.21	12.33	11.51	10.51	9.99	9.07	8.05	6.57	4.54
P1	12.7	12.72	11.12	10.60	10.12	10.00	9.71	9.41	9.30	8.95	8.07
P2	11.1	11.10	11.80	11.96	11.67	11.23	10.66	9.96	8.96	7.44	5.22
IC	10.9	10.81	10.24	10.02	9.81	10.08	9.91	10.14	10.01	9.81	9.17
IIS	9.1	9.18	7.92	7.66	7.86	8.20	*8.76*	9.48	10.71	12.63	17.62
MIS2	8.1	8.13	10.53	11.92	12.33	**12.45**	**11.94**	11.02	9.51	7.58	4.60
ICS1	7.6	7.33	8.10	8.43	8.95	9.28	10.02	10.59	11.53	12.40	13.37
ICS2	4.8	4.55	6.30	7.57	8.61	9.84	10.93	**12.20**	**13.26**	13.86	12.88
MP	4.1	*4.13*	*4.98*	*5.57*	*6.78*	*7.78*	8.91	10.52	12.63	**16.51**	**22.21**

Table 8: Central weight vectors of the case study (in %). The alternatives are in same order as in Table 7.

Course	LLE	LTS	LMU	DLE	DTS	DAC
DA	19.83	28.41	14.19	13.31	11.50	12.76
DB	20.14	28.36	14.01	13.44	11.38	12.68
P1	20.45	28.68	13.56	13.38	11.27	12.66
P2	20.09	28.49	13.89	13.42	11.40	12.71
IC	19.94	28.18	13.95	13.51	11.68	12.74
IIS	20.51	28.46	13.44	13.48	11.40	12.72
MIS2	19.96	28.55	14.13	13.17	11.59	12.60
ICS1	19.72	27.87	14.15	13.61	11.68	12.97
ICS2	19.59	27.89	14.53	13.27	11.62	13.10
MP	19.51	28.29	14.49	13.62	11.50	12.59

7.4 Results

From the results we can observe that courses can be splitted into three groups, which can be called “good”, “small improvements needed”, and “large improvements needed”. By looking at Table 7, we can distinguish certain courses that obtain high acceptabilities for the best ranks, and low acceptabilities for the worst ranks. DA, DB, P1, and P2 have relatively low acceptabilities for the three worst ranks ($b^8 - b^{10}$), and high acceptabilities for the three best ranks ($b^1 - b^3$). The confidence factors of these courses are also high compared to the other courses. Thus, the first conclusion drawn from the results is, that the majority of the students have an opinion that that these courses are taught well. These courses are separated to the group “good”.

The next course in Table 7 (IC) has high acceptabilities for the five best ranks, but it also ob-

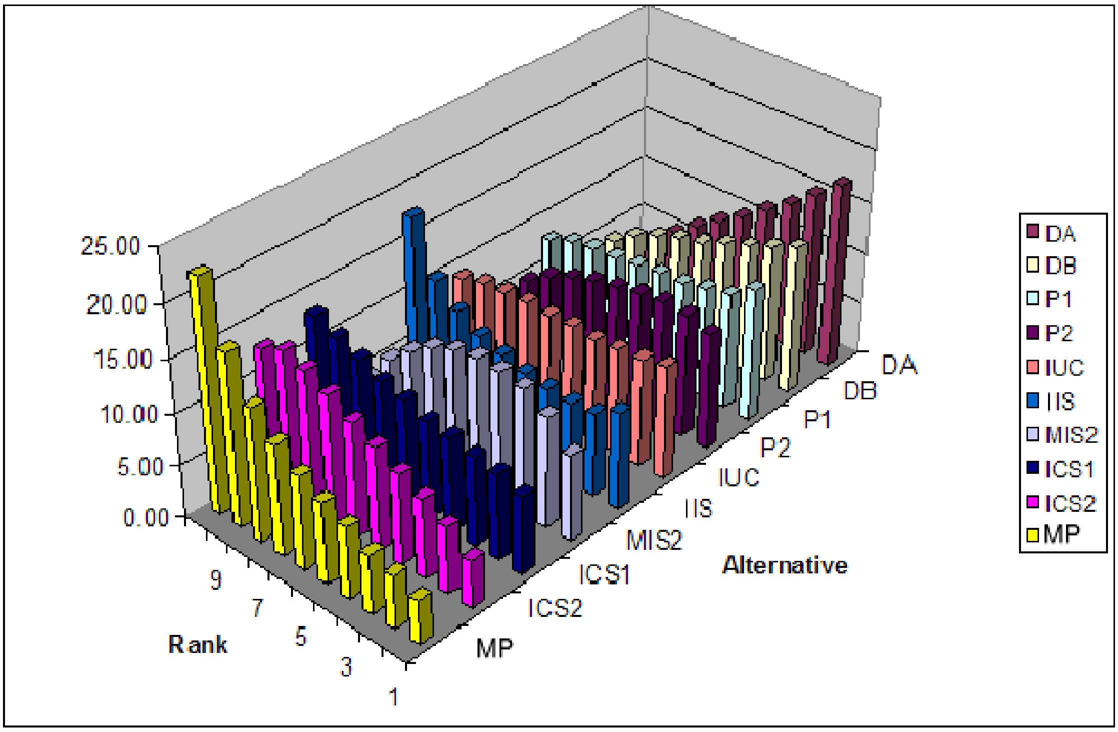

Figure 4: The rank acceptability indices of the case study.

tained relatively high acceptabilities for the three worst ranks. This course is separated to the group “small improvements needed”. To this group we put also MIS2, because it obtained relatively low acceptabilities for the best ranks. MIS2 is not separated to the worst group (“large improvements needed”), because it has quite low acceptabilities for the worst two ranks.

All of the courses left (IIS, ICS1, ICS2, MP) have reasonably high acceptabilities for the worst ranks, and thus we place them in the group “large improvements needed”. ICS2 and MP also have very low confidence factors (under 5%), so even if the acceptability for the best rank would be high, we should not place them into the best group. After the initial phase, we have 3 groups: The (4) courses which do not need improvement, the (2) courses which need small improvements, and the (4) courses which need a lot of improvements. The groups are presented in Table 9.

The preference information was present in the SMAA model as interval contraints for weights, and the effect of this is evident in the central weight vectors represented in Table 8. Usually (see, e.g. Lahdelma and Salminen, 2001) the central weight vectors have larger deviation. Now most of the central weights are within 1% for all criteria, because the feasible weight space has quite strict contraints. The magnitudes of the central weights for different criteria give insight to the collective preferences; the students seem to value lecturers level of expertise (LLE) and lecturer’s teaching skills (LTS) more than the expertise or the teaching skills of the demonstrator. It might surprise

Table 9: Initial grouping of the courses.

Group	**Courses**
Good	DA
	DB
	P1
	P2
Small improvements needed	IC
	MIS2
Large improvements needed	IIS
	ICS1
	ICS2
	MP

some, that the usefulness of the lecture material is deemed more important than the demostrator's level of expertise.

7.5 Managerial Implications

For making recommendations for the staff of the department we need to identify the faults in the evaluated courses. To do this for the courses in the group "small improvements needed", we need to find answer to the question: "why are courses IC and MIS2 graded worse than the ones in group "good" by majority of students?". The central weight vectors of the alternatives give answer to this question: if the central weight for an alternative is high, then the alternative is performing better with respect to that criteria than to the remaining ones. Because we had strict weight interval constraints present in the model, we have to compare the central weights to the ones of the better alternatives.

By looking at the central weights of IC, we can notice that it has a lower central weight for LTS than any other course. By looking at the criteria measurements at Table 5, we can see that is has lower mean (3.39) than any of the alternatives in the class "good". Thus as a recommendation, we think that the lecturer of IC should put more effort on her/his output during the lectures. For MIS2, there is no single central weight which is larger than the others. The criteria measurements of MIS2 are all thus lower than those of the courses in the group "good". The recommendation for the lecturer of the course MIS2 is thus based on the general preferences: it is most important to improve the level of expertise and the teaching skills.

From the central weights of courses ICS1 and ICS2, we can see that they are performing very bad with respect to criteria LTS. The recommendation for the lecturers of those courses is thus to put more effort on the lectures.

The low central weight and criteria measurement of LMU for IIS lead us to recommend that the lecturer should improve the lecture material distributed during that course. Also the means of the criteria measurements for the criteria involving demostrator are quite low, so as a second recommendation the demonstrator should improve her/his performance on all areas.

The last course, MP, has low criterion measurements for all criteria, and we recommend the lecturer and the demonstrator to think again how the course is taught. By looking at the criteria scores, we can see that LTS and DAC have very low mean values. Thus the first improvements should be that the lecturer improves her/his lecturing skills and the demonstrator spends more time ensuring that all students have understood the solutions of the exercises.

7.6 Comments from the DMs

A team of four professors responsible for planning and developing the teaching in the field of computer science participated in evaluating the method and the results. First the DMs observed that there are several possible sources for bias in both the original criteria data and in the preference information. The people who answer the questionnaires are not an unbiased sample. Drop-outs and those that do not participate in the lectures do not answer the questionnaires. One faction of the non-participants may be the competent students, who can complete the courses by self-study. The criteria evaluations may also reflect other things than they intend to measure. If the course topic is difficult, the student has lacking motivation, or is agitated due to bad performance or other reasons, the course may get worse evaluation. In contrast, a teacher with good social skills may get overly good evaluations regardless of learning results (but learning is of course promoted by a pleasant learning environment). All these factors result in great uncertainty in the input data, and should therefore be considered when making conclusions based on the results. Also, it is not obvious if the currently applied criteria are the right ones, or if they should be modified or complemented by additional criteria.

With the above factors in mind, it was considered strength of the method, that it can explicitly represent the uncertain, imprecise, and lacking information, both in criteria measurements and in preferences. In particular, it is impossible to objectively specify precise weights for the different criteria. Therefore, it is very important that the method can handle vague preference statements. In addition, the course evaluations and preference statements may typically contain some outliers (answers that differ radically from the mainstream). Also such opinions must be respected. A definite advantage of the proposed method is that it is able to consider explicitly all information, not just the average of different answers.

The classification of the alternatives into three categories may be too arbitrary in this case. For example, the difference between the acceptability index of the fourth and fifth alternative was only 0.14 percentage points, but still they were assigned different classes. The results indicate that maybe only the best and the two last alternatives can be discerned from the remaining ones. In any case, results like this should not be used to start immediate managerial actions, but as diagnostic, to help identifying potential problems and best practices, and in finding ways to improve the organization of courses.

It was concluded that the method can give valuable support for course planning and development. That no precise weights are needed gives robustness and reliability to the method. However, it is necessary to repeat the analysis to gain more experience. It is also important to ask the right questions; it is not obvious that the current criteria are the ideal ones. Additional surveys with free-form answers could be used to better identify potential problems and to improve the set of criteria. The analysis could also be made more in-depth by analyzing the correlation between the survey answers and student profiles. However, such analyses may be difficult to make, because the survey answers are anonymous. Anonymity is important, because then the students are more likely to express their true opinions.

8 Conclusions and avenues for future research

In this paper we have presented a novel method for handling preference information from multiple DMs in SMAA. The method is based on collecting the DMs' preferences as weights for subsets of criteria, aggregating them using the Yager's rule of combination and representing them as interval constraints for sums of sets of weights in SMAA. We demonstrated the method using a small car-selection problem, and presented a case study of university course evaluation in which the method was applied.

In the case study we evaluated the quality of education in the Department of Information Technology at the University of Turku, Finland. We used course feedback forms to establish stochastic criteria measurements for the courses, and used questionnaires to elicit preference information from a large group of students. The preference information was modelled and aggregated using the method introduced in this paper. Based on the results of the analysis, a ranking of the courses could be established, and for the courses with weak performance, improvements could be suggested. The results of the case study were presented to the personnel in charge of education at the department, and they considered the method to give valuable support for course planning and development.

This work should not be a “dead-end”. On the contrary, it should be taken further and exploited in other real-world applications of MCDA methods in managerial decisions. The future theoretical work should address new methods for collecting preferences from the DMs and for transforming them into bpa's. We hope that the results from the case study give rise in number to applications of SMAA to public decision making.

Acknowledgements

The work of Tommi Tervonen was supported by a grant from Turun Yliopistosäätiö. The work of José Figueira was partially supported by the grant SFRH/BDP/6800/2001 (Fundação para a Ciência e Tecnologia, Portugal) and MONET research project (POCTI/GES/37707/2001).

References

Bana e Costa, C. A., 1986. A multicriteria decision aid methodology to deal with conflicting situations on the weights. European Journal of Operational Research 26, 22–34.

Bana e Costa, C. A., 1988. A methodology for sensitivity analysis in three-criteria problems: A case study in municipal management. European Journal of Operational Research 33, 159–173.

Belton, V., Stewart, T. J., 2002. Multiple Criteria Decision Analysis - An Integrated Approach. Kluwer Academic Publishers, Dordrecht.

Beynon, M., Curry, B., Morgan, P., 2000. The Dempster-Shafer theory of evidence: An alternative approach to multicriteria decision modelling. Omega 28, 37–50.

Charnetski, J., 1973. The multiple attribute problem with partial information: The expected value and comparative hypervolume methods. Ph.D. thesis, University of Texas at Austin.

Charnetski, J., Soland, R., 1978. Multiple-attribute decision making with partial information: The comparative hypervolume criterion. Naval Research Logistics Quarterly 25, 279–288.

Des Cartes, R., 1641. Meditationes de prima philosophia, in qua Dei existentia et animae immortalitas demostratur. Parisiis, Apud Michaelem Soly.

Durbach, I., 2004. A simulation-based test of stochastic multicriteria acceptability analysis using achievement functions. European Journal of Operational Research (to appear).

Figueira, J., Greco, S., Ehrgott, M. (Eds.), 2005. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Science+Business Media, Inc., New York.

Hokkanen, J., Lahdelma, R., Miettinen, K., Salminen, P., 1998. Determining the implementation order of a general plan by using a multicriteria method. Journal of Multi-Criteria Decision Analysis 7 (5), 273–284.

Hokkanen, J., Lahdelma, R., Salminen, P., 1999. A multiple criteria decision model for analyzing and choosing among different development patterns for the Helsinki cargo harbor. Socio-Economic Planning Sciences 33, 1–23.

Hokkanen, J., Lahdelma, R., Salminen, P., 2000. Multicriteria decision support in a technology competition for cleaning polluted soil in Helsinki. Journal of Environmental Management 60, 339–348.

Kangas, A., Kangas, J., Lahdelma, R., Salminen, P., 2003a. Decision support model with dependent uncertainties in forest planning, submitted manuscript.

Kangas, J., Hokkanen, J., Kangas, A., Lahdelma, R., Salminen, P., 2003b. Applying stochastic multicriteria acceptability analysis to forest ecosystem management with both cardinal and ordinal criteria. Forest Science 49 (6), 928–937.

Lahdelma, R., Hokkanen, J., Salminen, P., 1998. SMAA – stochastic multiobjective acceptability analysis. European Journal of Operational Research 106, 137–143.

Lahdelma, R., Makkonen, S., Salminen, P., 2005. Multivariate Gaussian criteria in SMAA. European Journal of Operational Research (to appear).

Lahdelma, R., Miettinen, K., Salminen, P., 2003. Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA). European Journal of Operational Research 147, 117–127.

Lahdelma, R., Miettinen, K., Salminen, P., 2004a. Reference point approach for multiple decision makers. European Journal of Operational Research (to appear).

Lahdelma, R., Miettinen, K., Salminen, P., Tervonen, T., 2004b. Computational methods for stochastic multicriteria acceptability analysis. In: Neittaanmäki, P., Rossi, T., Korotov, S., Onate, E., Périaux, J., Knörzer, D. (Eds.), Proceedings of the 4th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS). Vol. 2. Jyväskylä, Finland, 24–28 July, 2004.

Lahdelma, R., Salminen, P., 2001. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. Operations Research 49 (3), 444–454.

Lahdelma, R., Salminen, P., 2002. Pseudo-criteria versus linear utility function in stochastic multicriteria acceptability analysis. European Journal of Operational Research 14, 454–469.

Lahdelma, R., Salminen, P., 2003. An ordinal multicriteria method in an environmental impact assessment procedure, manuscript.

Lahdelma, R., Salminen, P., 2004. Stochastic multicriteria acceptability analysis using the data envelopment model. European Journal of Operational Research (to appear).

Lahdelma, R., Salminen, P., 2005. Classifying alternatives in SMAA using cross confidence factors. European Journal of Operational Research (to appear).

Lahdelma, R., Salminen, P., Hokkanen, J., 2002. Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria. European Journal of Operational Research 142, 345–356.

Lahdelma, R., Salminen, P., Simonen, A., Hokkanen, J., 2001. Choosing a reparation method for a landfill using the SMAA-O multicriteria method. In: Multiple Criteria Decision Making in the New Millenium, Lecture Notes in Economics and Mathematical Systems. Vol. 507. Springer-Verlag, Berlin, pp. 380–389.

Raiffa, R., Richardson, J., Metcalfe, D., 2002. Negotiation Analysis: The science and art of collaborative decision making. The Belknap Press of Harvard University Press, Cambridge, Massachusetts, U.S.A.

Rietveld, P., 1980. Multiple Objective Decision Methods and Regional Planning. North-Holland, Amsterdam.

Rietveld, P., Ouwersloot, H., 1992. Ordinal data in multicriteria decision making, a stochastic dominance approach to siting nuclear power plants. European Journal of Operational Research 56, 249–262.

Roy, B., 1989. Main sources of inaccurate determination, uncertainty and imprecision in decision models. Mathl. Comput. Modelling 12 (10/11), 1245–1254.

Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press.

Smets, P., 1990. The combination of evidence in the transferable belief model. IEEE Trans. PAMI 12, 447–458.

Tervonen, T., Figueira, J., Lahdelma, R., Salminen, P., 2004a. An inverse approach for ELECTRE III. Research Report 20/2004 of The Institute of Systems Engineering and Computers (INESC-Coimbra), Coimbra, Portugal, http://www.inescc.pt.

Tervonen, T., Lahdelma, R., Salminen, P., 2004b. A method for elicitating and combining group preferences for Stochastic Multicriteria Acceptability Analysis. Tech. Rep. 638, TUCS - Turku Center for Computer Science, http://www.tucs.fi.

Yager, R. R., 1987. On the Dempster-Shafer framework and new combination rules. Information Sciences 41, 93–137.