



Modelling uncertainty in stochastic multicriteria acceptability analysis[☆]

Ian N. Durbach^{a,b,*}, Jon M. Calder^a

^a Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, South Africa

^b African Institute for Mathematical Sciences, 6-8 Melrose Road, Muizenberg 7945, South Africa

ARTICLE INFO

Article history:

Received 27 February 2015

Accepted 29 October 2015

Keywords:

Decision making/process
Decision support systems
Multicriteria
Risk
Sensitivity analysis

ABSTRACT

This paper considers problem contexts in which decision makers are unable or unwilling to assess trade-off information precisely. A simulation experiment is used to assess (a) how closely a rank order of alternatives based on partial information and stochastic multicriteria acceptability analysis (SMAA) can approximate results obtained using full-information multi-attribute utility theory (MAUT) with multiplicative utility, and (b) which characteristics of the decision problem influence the accuracy of this approximation. We find that fairly good accuracy can be achieved with limited preference information, and is highest if either quantiles and probability distributions are used to represent uncertainty.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

When facilitating decisions it sometimes happens that some inputs to the preference model either cannot be assessed at all or can only be assessed within relatively large bounds of uncertainty (e.g. [1,12,13]). This can happen for a number of reasons: a lack of time, a politically sensitive problem context, or a lack of decision maker (DM) involvement, for example. Whatever the reason, in these cases the DM is unable or unwilling to express him or herself with the degree of precision required by conventional decision aids. We call decision problems which must be addressed under such conditions “low-involvement” decisions. The question is what, if any, decision support can be provided in such situations.

Stochastic multicriteria acceptability analysis (SMAA [25,19]) is a family of decision models that can be used with arbitrarily precise preference information. It addresses low-involvement decision-making by providing information about the types of preferences (if any) that would lead to the selection of each alternative.

In this paper we use a simulation experiment to evaluate the ability of SMAA to approximate results obtained using multi-attribute utility theory (MAUT) where preferences are represented by a multiplicative utility function. In particular, we ask how closely results computed from a key output from SMAA (the acceptability index) can approximate those obtained using MAUT.

In doing so we hope to provide a broad indication of the losses that are possible if facilitators choose to use a low-involvement decision aid such as SMAA rather than compelling DMs to be more precise in their assessment of certain types of preference information – for example, using more detailed problem structuring. We also wish to test the robustness of the SMAA approach to various aspects of the decision process: the size of the decision problem, the way attribute evaluations are distributed, the underlying preference functions, the accuracy of assessed information, the amount of preference information gathered, and the way in which the acceptability index is constructed.

In addition to the conventional SMAA model, which uses probability distributions to represent uncertainty in the attribute evaluations, we also introduce and evaluate a number of ‘simplified’ models which make use of summarised measures of uncertainty instead of a full probability distribution [6]. By assessing the accuracy of both conventional and simplified uncertainty models under a range of different conditions we hope to provide motivation for the use of simplified models in appropriate circumstances. A similar approach has been used in Durbach and Stewart [9] to assess the effect of using simplified uncertainty formats in general decision-making, and we employ a similar simulation structure in the current paper.

Our view is that in nearly all cases it is preferable to resolve preferential uncertainty through discussion and problem structuring rather than by employing more ‘lenient’ decision aids, because of the additional insight and opportunities for learning. We focus on those circumstances in which the DM is unable or unwilling to participate fully in this process. In using a simulation experiment, we acknowledge that we can only evaluate the extent to which

[☆]This manuscript was processed by Associate Editor Doumpos.

* Corresponding author at: Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, South Africa.

E-mail address: ian.durbach@uct.ac.za (I.N. Durbach).

using SMAA rather than MAUT might impact results. We cannot evaluate critical issues like whether the reduced time spent on problem structuring in SMAA is “worth” the reduction in decision quality, or the degree to which the problem structuring process, through the insight it generates for the DM, is useful as an end in itself. Simulation results are unable to provide general conclusions on the viability of different methods, but provide inputs to such discussion by identifying the potential trade-offs in accuracy that are implied when using a simplified model. Ultimately accuracy must be weighed against other factors to determine which decision model may be most appropriate for a problem.

The remainder of the paper is structured as follows. Section 2 provides a review of the relevant background literature and notation. Section 3 describes the structure of our simulation experiment. Section 4 presents the results in a direct fashion, delaying a more detailed discussion of these until Section 5, which also provides theoretical justification for some key findings using results from applied probability theory. Section 6 discusses implications of the simulation results for the use of SMAA as a prescriptive decision aid, and concludes.

2. Notation and background

Consider a decision problem consisting of I alternatives $\{a_1, a_2, \dots, a_I\}$ evaluated on J attributes $\{c_1, c_2, \dots, c_J\}$. Let Z_{ij} be a random variable denoting the attribute evaluation of a_i on c_j , and U be a multi-attribute utility function mapping the attribute evaluations of alternative a_i (denoted \mathbf{Z}_i) to a real value using a weight vector \mathbf{w} . A joint density function $f_X(\mathbf{Z})$ governs the generation of the Z_{ij} in the space $X \subseteq \mathbb{R}^{I \times J}$, and a second joint density function $g(\mathbf{w})$ governs the generation of imprecise or unknown weights in the weight space W . Total lack of knowledge is usually represented by a uniform distribution in W . If restrictions have been placed on W we denote the feasible weight space by W' .

The original SMAA method [20] analysed the combinations of attribute weights that result in each of a set of alternatives being selected when using an additive utility function. Subsequently, a number of SMAA variants have been developed. These differ in terms of the preference model used and thus the type of preference information that is imprecisely known, but are all based upon Monte Carlo simulation from distributions which govern unknown preference parameters (and attribute evaluations). For example, SMAA variants are available for value function [20,17], outranking [10], reference point [21,5], prospect theory [18], Choquet integral [2], and AHP [7] methods. Comprehensive reviews are given by Tervonen and Figueira [25] and Lahdelma and Salminen [19].

Given a particular weight vector \mathbf{w} , the global utility of each alternative can be computed and a rank ordering of alternatives obtained. SMAA-2 [17] is based on simulating a large number of random weight vectors from $g(\mathbf{w})$ and observing the proportion and distinguishing features of weight vectors which result in each alternative obtaining a particular rank r (usually the “best” rank, $r=1$), using an additive value function model. Let the set of weight vectors that result in alternative a_i obtaining rank r be denoted by W_i^r . SMAA is based on an analysis of these sets of weights using a number of descriptive measures, the most important of which are:

Acceptability indices: The rank- r acceptability index b_i^r measures the proportion of all simulation runs, i.e. weight vectors, that make alternative a_i obtain rank r . A cumulative form of the acceptability index called the R -best ranks acceptability index is defined as $B_i^R = \sum_{r=1}^R b_i^r$ and measures the proportion of all weight vectors for which alternative a_i appears anywhere in the best R ranks. In

the discussion in Section 5 we make use of ordered acceptability indices, where we denote the alternative with the k -th largest rank- r acceptability index by $a_{(k)}^r$, and its acceptability index by $b_{(k)}^r$. **Central weight vectors:** The central weight vector \mathbf{w}_i^r is defined as the expected center of gravity of the favourable weight space W_i^r . It gives a concise description of the “typical” preferences supporting the selection of a particular alternative a_i , and in practice is computed from the empirical (element-wise) averages of all weight vectors supporting the selection of a_i as the best alternative.

The exact number of Monte Carlo iterations that are required to achieve a given accuracy is discussed in Tervonen and Lahdelma [26]. To estimate the acceptability index within δ of the true value with 95% confidence, one requires $1.96^2/4\delta^2$ iterations – so that 10 000 iterations will usually be sufficient to achieve error bounds of 1%.

Uncertain attribute evaluations are conventionally treated in SMAA using probability distributions, with each simulation run drawing values at random from these distributions. Adapting SMAA to use other uncertainty formats, however, is generally straightforward, as described in Durbach and Davis [6]. Each uncertain attribute is simply replaced by a number of lower-level attributes which capture the uncertainty in the evaluations on that attribute, using one of many possible simplified uncertainty formats. In this paper, we test the effect of using three different formats: expected values; variances; and quantiles. This transforms the decision problem into one having the same appearance as a deterministic decision problem, which can be treated by any of the existing SMAA models with some minor modifications. We sometimes refer to these collectively as “SMAA models” although it should be clear that these are approximations of external uncertainty built on top of the same SMAA approach. The models are described in more detail in Section 3. Table 1 provides a summary of notation used in the paper.

3. Description of the simulation study

The general structure of the simulation¹ is summarised in Fig. 1 and implements the following basic steps:

1. Form a hypothetical problem context, generating the relevant attribute evaluations.
2. Apply a multiplicative MAUT model to derive idealised or ‘true’ utilities and thus find the ‘true’ rank ordering of alternatives.
3. Calculate summarised measures of uncertainty based on the generated data and incorporating observational errors.
4. Run different models using SMAA and the inputs from step 3, and then compare the model results against the ‘true’ utilities and rank order obtained from step 2.

3.1. Generating attribute evaluations

We consider a decision problem involving I alternatives evaluated over J attributes. External uncertainty about attribute performance is captured by simulating attribute evaluations for alternative a_i on attribute c_j from a gamma distribution with mean μ_{ij} , standard deviation σ_{ij} , and skewness ξ_{ij} . We denote this distribution $f_{ij}^r(Z_{ij})$. Each mean μ_{ij} is drawn randomly from a uniform distribution between 0.2 and 0.8 for all alternatives on all attributes. Means are then standardised across all the attributes to lie on the unit hypersphere i.e.

¹ All codes used to run the simulations described in this section are openly available from <http://dx.doi.org/10.5281/zenodo.30523>.

Table 1
Summary of all parameters used.

Problem context	
a_i	Alternative i
c_j	Attribute j
$c_{(k)}$	Attribute with k -th largest importance weight
Z_{ij}	Performance of a_i on c_j
$f_X(\mathbf{Z})$	Joint probability distribution over all Z_{ij}
$f_{ij}(Z_{ij})$ or f_{ij}	Marginal probability distribution for Z_{ij}
Attribute weights	
W	Total weight space
W'	Currently feasible weight space
W_i^1	Favourable weight space for a_i
w_j	Weight for c_j
$w_{(1)}$	k -th largest attribute importance weight
SMAA	
b_i^r	Rank- r acceptability index
$b_{(k)}^r$	k -th largest rank- r acceptability index
$a_{(k)}^r$	Alternative with k -th largest rank- r acceptability index
$B_{(k)}^1$	Random variable which when realised gives $b_{(k)}^1$
B_i^R	Cumulative R -best ranks acceptability index
Simulation: problem context	
I	Number of alternatives
J	Number of attributes
K	Number of simulated realisations of Z_{ij}
Z_{ijk}	Realisation k of Z_{ij}
μ_{ij}	Mean of attribute evaluations for a_i on c_j
σ_{ij}	Standard deviation of attribute evaluations for a_i on c_j
ξ_{ij}	Skewness of attribute evaluations for a_i on c_j
Simulation: DM preferences	
v_j	Marginal value function for c_j
τ_j	Reference level for v_j
λ_j	Value of v_j at reference level
V_i	Global value of a_i
U_i	Global utility of a_i
ζ	Risk preferences defined over value
ω	'True' attribute weights
Simulation: uncertainty modelling	
ρ	Proportion of importance ranks provided by DM
ϵ	Size of assessment errors
Z_{ij}	Sampled value from z_{ij} , used by 'TDist' model
\hat{Z}_{ij}	Sampled value from f_{ij} used by 'EDist' model
$E[Z_{ij}], \hat{E}[\hat{Z}_{ij}]$	Expected value of Z_{ij} before and after application of ϵ
$\text{VAR}[Z_{ij}], \hat{\text{VAR}}[\hat{Z}_{ij}]$	Variance of Z_{ij} before and after application of ϵ
$Q_r[Z_{ij}], \hat{Q}_r[\hat{Z}_{ij}]$	5%, 50% and 95% quantiles of Z_{ij} for $r = 1, 2, 3$ before and after application of ϵ
ϕ	Weight of $\text{VAR}[\hat{Z}_{ij}]$ relative to $E[\hat{Z}_{ij}]$

$\sum_{j=1}^J \mu_{ij}^2 = 1, \forall i$. This creates a set of non-dominated alternatives, at least in this 'average' sense. The standard deviations and skewnesses of the attribute evaluations are varied as parameters of the simulation (see Section 3.6 for details of chosen parameters values). A set of $K=15\,000$ realisations is generated to represent the uncertain attribute evaluation of each alternative on each attribute. These evaluations, denoted z_{ijk} , are then standardised within each attribute (across all alternatives and realisations), to have a minimum of 0 and a maximum of 1.

3.2. Generating preference structures

We employ a multiplicative MAUT model to simulate the 'true' underlying preference structure of a DM, with U_i denoting the overall utility of alternative a_i . For simplicity we assume that our simulated DM has a utility function that exhibits constant risk preferences over the scalar value attribute V_i . Specifically, this requires that the DM's preferences over the attribute space are compatible with an additive value function, that at least one attribute is utility independent of its complement, and that $J \geq 3$

[15, Theorem 6.11]. These conditions allow us to rewrite the global utility of alternative a_i as a function of its global value V_i , and more precisely as a function of its value as derived from an additive value model:

$$U_i = \begin{cases} e^{-\zeta V_i} = \exp\left[-\zeta \sum_{j=1}^J w_j v_j(Z_{ij})\right] & \zeta < 0 \\ V_i = \sum_{j=1}^J w_j v_j(Z_{ij}) & \zeta = 0 \\ -e^{-\zeta V_i} = -\exp\left[-\zeta \sum_{j=1}^J w_j v_j(Z_{ij})\right] & \zeta > 0 \end{cases} \quad (1)$$

where w_j denotes the relative importance weight of attribute c_j , v_j is the marginal value function for c_j , and ζ determines whether the DM is constant risk averse ($\zeta > 0$), risk neutral ($\zeta = 0$) or constant risk prone ($\zeta < 0$) with respect to value. The precise form of the marginal value function v_j is discussed in Section 3.2.1. To calculate the 'true' utility of an alternative a_i , expectations are taken over the full set of K realisations. That is, we (a) compute, for each k , the value V_{ik} and hence utility U_{ik} of alternative a_i using attribute evaluations z_{i1k}, \dots, z_{ijik} ; (b) average these utilities over k . This gives

$$U_i^{\text{MAUT}} = \begin{cases} E[e^{-\zeta V_i}] = (1/K) \sum_{k=1}^K \exp\left[-\zeta \sum_{j=1}^J w_j v_j(z_{ijk})\right] & \zeta < 0 \\ E[V_i] = (1/K) \sum_{k=1}^K \sum_{j=1}^J w_j v_j(z_{ijk}) & \zeta = 0 \\ -E[e^{-\zeta V_i}] = (1/K) \sum_{k=1}^K -\exp\left[-\zeta \sum_{j=1}^J w_j v_j(z_{ijk})\right] & \zeta > 0 \end{cases} \quad (2)$$

This 'true' measure of utility is based on perfect and complete information and, for the purposes of this simulation study, provides a set of idealised results to which results from the different SMAA models can be compared.

3.2.1. Marginal value functions

Marginal value functions are assumed to be piecewise linear between the points $(0, 0)$, (τ_j, λ_j) , and $(1, 1)$. Convex, linear, and concave value functions can be generated by imposing $\tau_j > \lambda_j$, $\tau_j = \lambda_j$, and $\tau_j < \lambda_j$ respectively, allowing a wide range of preference to be simulated parsimoniously. Note that this is a simplified version of the approach used in Durbach and Stewart [9], where additional parameters were used to control the degree of curvature in the preference functions either side of τ_j . Marginal values are given by:

$$v_j(x) = \begin{cases} \frac{\lambda_j x}{\tau_j} & \text{for } 0 \leq x \leq \tau_j \\ \lambda_j + \frac{1-\lambda_j}{1-\tau_j}(x-\tau_j) & \text{for } \tau_j < x \leq 1 \end{cases} \quad (3)$$

3.2.2. Inter-attribute weights

'True' attribute weights ω are simulated to be uniformly distributed and sum to one, using the approach in Butler et al. [4]. This involves sorting $J-1$ randomly generated $Un[0, 1]$ variates in ascending order, appending 0 and 1 to the extremes of this sequence, and extracting the differences between adjacent values.

3.3. Simulating the application of SMAA

The 'true' attribute weights ω provide a rank ordering of attributes from most important to least important. Let $c_{(k)}$ denote

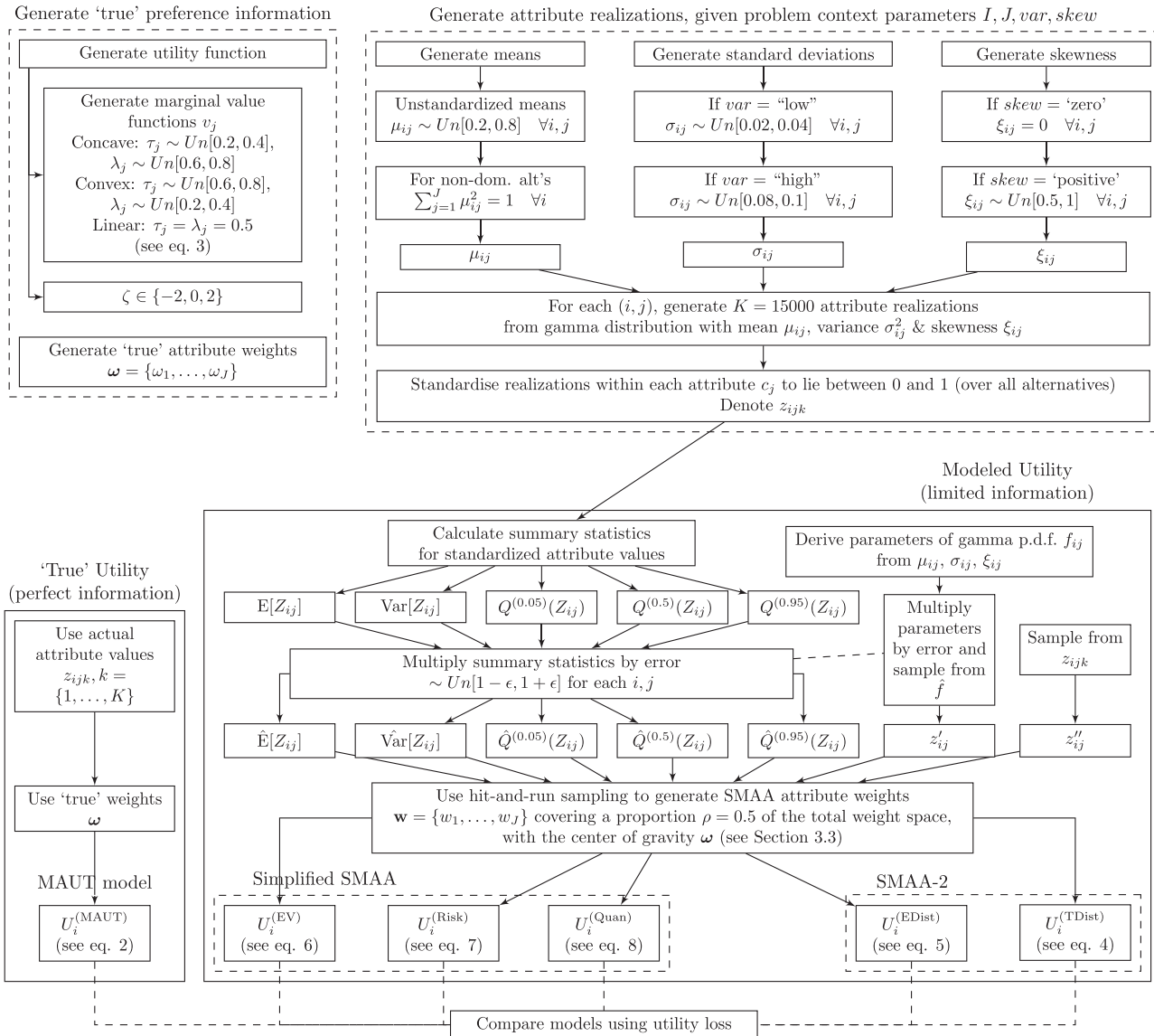


Fig. 1. Outline of a simulation run.

the k -th most important attribute. To emulate the inclusion of partial preference information in the SMAA models, we simulate a DM who provides the importance ranks of a proportion ρ of the full set of attributes (less one, since the rank of the final attribute is given by the others), in either decreasing (starting from $c_{(1)}$) or increasing (starting from $c_{(j)}$) order of importance. Total uncertainty regarding weights corresponds to $\rho = 0$ while for a complete rank ordering $\rho = 1$.

Efficient “hit-and-run” sampling procedures using Markov Chain Monte Carlo (MCMC) methods to simulate uniformly distributed weights on a simplex with arbitrary constraints have been described in Tervonen et al. [27] and van Valkenhoef et al. [28]. Importance ranks provide ordinal information that is easily implemented in this framework as linear constraints on the weight space e.g. $w_{(1)} > w_{(k)}, \forall k \neq 1$. The hit-and-run procedure is implemented in R using the ‘hitandrunk’ package [30]. The computational costs of the simulation are already fairly high, so that ρ is not included as one of the final simulation parameters but is rather the subject of an initial investigation before being set at a fixed value of 0.5 (see Section 4.1 for details).

Finally, our SMAA models assume that marginal value functions are known and assessed without error, but that the parameter ζ

governing risk preferences defined over value is unknown. Although in principle it is straightforward to simulate possible values of ζ , or parameters of marginal value functions, as part of each SMAA model, assuming known value functions but unknown weights is, perhaps regrettably, a common feature of SMAA, which we follow here to prevent the simulations from becoming overly complex. We thus assume that low-involvement models like SMAA may well assume $\zeta = 0$, approximating the multiplicative utility function with an additive value function and ranking alternatives according to their global values as returned by the respective SMAA models described below.

3.3.1. Model using a full probability distribution

We implement two SMAA models employing full probability distributions. The first random sample, at each SMAA iteration, values from the empirical distribution of the standardised attribute values, and thus directly samples from z_{ijk} . The resulting “true distribution” model is given by

$$U_i^{(TDist)} = \sum_{j=1}^J w_j v_j(z'_{ij}) \tag{4}$$

where z_{ij}^* is the sampled value from z_{ij} . Note that this model captures uncertainty in the attribute evaluations exactly, by construction, but is not easily amenable to modelling assessment errors. We therefore also use a second distributional model that randomly draws attribute evaluations from the $f_{ij}(Z_{ij})$ at each SMAA iteration. Assessment errors can be modelled in a straightforward way by imposing errors on the parameters of f_{ij} . The “estimated distribution” model is given by

$$U_i^{(EDist)} = \sum_{j=1}^J w_j v_j(z_{ij}^*) \quad (5)$$

where z_{ij}^* refers to the sampled attribute evaluations, now understood to be drawn from $f_{ij}(Z_{ij})$ rather than the empirical distributions.

3.3.2. Model using expected values

The model using only expected values ignores any uncertainty in the attributes and is defined as:

$$U_i^{(EV)} = \sum_{j=1}^J w_j v_j(E[Z_{ij}]) \quad (6)$$

Expected value models effectively reverse the order of expectation and utility operations. Of course if expectations can only be estimated from the full distributions then (6) is of little use, but it may be the case that the DM is able to assess an average fairly easily, without reference to the full distribution. In these cases (6) may be thought of as a “simplified” approach. The model is investigated in Kirkwood [16], Stewart [24], and Durbach and Stewart [8].

3.3.3. Model using expected values and variances

This model uses variances in addition to expected values and is referred to here as a “risk” model. It is defined as:

$$U_i^{(Risk)} = \sum_{j=1}^J w_j v_j(E[Z_{ij}]) - \phi \sum_{j=1}^J w_j VAR[Z_{ij}] \quad (7)$$

where ϕ is a multiplier indicating the contribution of the risk component relative to the expected value component. Here we set $\phi = 0.25$, which is smaller than the values used in [9] but follows results in that paper reporting poor performance with larger values of ϕ .

3.3.4. Model using quantiles

The quantile model makes use of the “extended Pearson–Tukey” approximation [14], a widely applicable three-point approximation for continuous probability distributions typical of those elicited via judgmental assessments:

$$U_i^{(Quan)} = \sum_{j=1}^J w_j \sum_{r=1}^3 x_r v_j(Q_r[Z_{ij}]) \quad (8)$$

with $x_1 = x_3 = 0.185$ and $x_2 = 0.63$ and where Q_1 , Q_2 , and Q_3 are the 5%, 50% and 95% quantiles respectively.

3.4. Generating errors in the assessment of uncertainty information

In order to allow for assessment errors, all model inputs relating to uncertainty are multiplied by randomly generated realisations from $Un[1 - \epsilon, 1 + \epsilon]$. The error factor ϵ is a parameter of the simulation – we use 0%, 10% and 20% errors. The error adjusted parameter values are then denoted using the ‘hat’ symbol i.e. assessed probability distributions $\hat{f}(Z_{ij})$, expected values $\hat{E}[Z_{ij}]$, variances $\hat{VAR}[Z_{ij}]$, and quantile values $\hat{Q}_r[Z_{ij}]$. Errors are not applied to the “true distribution” model sampling directly from z_{ijk} . In the case of the “estimated distribution” model errors are applied to both shape and scale parameters of the assumed

gamma distributions governing attribute evaluations before renormalising attribute evaluations to lie between zero and one.

3.5. Output measures

For each SMAA model, alternatives are ranked by their R -best ranks’ acceptability indices, where $R \in \{1, 3, 5\}$ is varied as a parameter of the simulation. To evaluate the quality of the alternatives selected by each SMAA model we use three outcomes: the probability that the alternative selected by MAUT appears in the top one, three, or five positions of the SMAA rank order. The first of these tests the ability of SMAA with respect to the choice problematique, the others with respect to the shortlisting of promising alternatives.

In previous simulations (e.g. [8,9]) we have primarily used ‘utility loss’ $UL = (U_{i^*} - U_{i^{set}})/(U_{i^*} - U_{i_w})$ to measure model accuracy, where U_{i^*} , U_{i_w} , and $U_{i^{set}}$ are respectively the utilities of the best, worst and chosen (by SMAA) alternatives according to the simulated ‘true’ rank order. In the current paper, we have elected to use rank-based measures as the analytical results reported in Section 5 are much easier to obtain with these than with utility losses, which depend both on what rank an alternative obtains and the distribution of utilities over alternatives. In Section 4.1 we report utility loss for preliminary simulations and show that it is strongly correlated to the probability that MAUT and SMAA select the same alternative.

3.6. Parameter values used

In simulation studies such as this it is usually necessary to include only a small number of possible values for each simulation parameter. This keeps the total computation time and model complexity within reasonable limits, and should not be interpreted as indicating that these are the only plausible values or even capture the full range of possible variation.

The effect of problem size is investigated using $I=9$ or 19 alternatives and $J=6$ or 12 attributes. The four combinations allow for an independent investigation of alternatives and attributes in decision environments located between ‘fairly small’ and ‘moderate’ in size. Similar values have been used in Barron and Barrett [3] and Salo and Hämäläinen [23].

Parameter values for the variance and skewness parameters (σ_{ij} and ξ_{ij} respectively) were chosen on the basis of trial-and-error experimentation. The variance parameter takes on either “low” ($\sigma_{ij} \sim Un[0.02, 0.04]$) or “high” ($\sigma_{ij} \sim Un[0.08, 0.1]$) values. For the symmetric case and prior to standardisation, a 6σ range will cover approximately 25% and 45% of the full range of possible attribute evaluations in the ‘low’ and ‘high’ variability conditions respectively. After standardisation these ranges are somewhat larger; pre-simulations indicated coverage of 40–50% and 65–70% respectively. This means that even in the ‘low’ variability condition there is still a considerable amount of uncertainty in the attribute evaluations. The skewness of the attribute evaluations is considered to be either zero (in which case the gamma distribution is equivalent to a Gaussian distribution), or a moderate positive value randomly drawn from a uniform distribution $\xi_{ij} \sim Un[0.5, 1] \forall i, j$. The assessment error parameter ϵ is key and so takes on one of three values: 0%, 10%, and 20%.

Parameter values for τ_j and λ_j are combined to give preferences that are convex ($\tau_j = 0.8, \lambda_j = 0.2$), concave ($\tau_j = 0.2, \lambda_j = 0.8$), or linear ($\tau_j = \lambda_j$). Note that τ_j and λ_j are not varied independently. A single parameter ζ controls preferences for risk in the utility function defined over scalar value, and thus maps global values into utilities using Eq. (1): we use $\zeta = 2$ (constant risk aversion with respect to value), $\zeta = 0$ (risk neutral), and $\zeta = -2$ (constant risk proneness).

SMAA models are based upon the best $R = \{1, 3, 5\}$ ranks. The effect of including partial preference information about the weights into the SMAA process is investigated as part of an initial set of simulations using $I = 19$ alternatives, $J = 12$ attributes, and $\rho \in \{0, 1/11, 3/11, 5/11, 8/11, 1\}$, with 500 runs for each value of ρ . We also simulated an additional ‘full-information’ case where the cardinal values of weights are known precisely. Other variables were held constant at high σ_{ij} and ξ_{ij} , $\epsilon = 0$, mostly convex utility (low τ_{ij} , high λ_{ij}), and $\zeta = 0$. On the basis of these results, reported in Section 4.1, an ‘intermediate’ value of $\rho = 0.5$ was selected for use in the full simulation study.

We used a full factorial design – with 16 parameter combinations for problem context, 3 for assessment error, and 6 for shape of utility function there are a total of 288 simulation parameter combinations. We perform 100 simulation runs for each combination of parameters, and each simulation run uses 10 000 random weight vectors to generate the SMAA results. Simulations were run using R version 3.1.3 [22] and used packages `hitandrun` version 0.5 and `smaa` version 0.2-3 [29].

4. Results

4.1. Weight restrictions

Fig. 2 illustrates the relationship between the amount of weight information provided and various measures of model accuracy. In this and other plots, error bars indicate 95% confidence intervals around the mean values. If uncertainty about attribute evaluations is fully specified but no weight information is provided at all, the probability that a SMAA model selects the same alternative as MAUT is 0.15, and the probability of that alternative appearing in the top three and five ranks (out of 19) of the SMAA rank order is 0.33 and 0.50 respectively. If ordinal weight information is fully specified in decreasing order of importance these probabilities increase to 0.77, 0.97, and 0.99.

Marginal improvements in model accuracy diminish with the importance of the attribute for which ordinal information is specified. The greatest improvement in model accuracy comes from providing ordinal information about a few of the most important

attributes. In this case, information about attributes in the lower half of the importance rank order produces very little additional improvement in model accuracy. Marginal returns diminish, but more slowly, if information is provided on the least important attributes first.

Fig. 2 also shows that the way in which uncertainty about attribute evaluations is specified has a strong effect on model accuracy, particularly on the probability of the SMAA and MAUT models selecting the same alternative as best, and particularly when some ordinal information about attribute importance is provided.

Finally, Fig. 2 shows that our four outcome measures are strongly correlated, responding in much the same way to how the uncertainty about attribute evaluations is represented, and to the provision of ordinal weight information. Correlations range between 0.41 (between the probability of selecting the preferred MAUT alternative in the top 1 and top 5 positions of the SMAA rank order) and -0.58 (between utility loss and the probability that the SMAA model selects the same alternative as MAUT).

On the basis of this preliminary investigation, we fix ρ at an intermediate value of 0.5 and in the remainder of this section explore the effect that other parameters have in more detail. Fig. 3 reports mean model accuracy for different uncertainty models using R -best-ranks in the computation of the acceptability index, under various error conditions and values of R . Results have been averaged over the remaining simulation parameters.

We report three results here. Firstly, errors in the assessment of uncertainty can seriously compromise model accuracy. The greatest deterioration, as a result of assessment errors, occurs in the probability that the SMAA model selects the same alternative as MAUT (relative to the probability of shortlisting this alternative), and in SMAA models representing uncertainty with probability distributions. Some caution is needed in interpreting this latter result, since assessment errors are not directly comparable across uncertainty types. It is not clear, for example, whether it is cognitively ‘easier’ to make a 10% error when assessing parameters of probability distributions or when assessing expected values.

Secondly, models using simplified representations of uncertainty, particularly quantiles, can return very similar results to

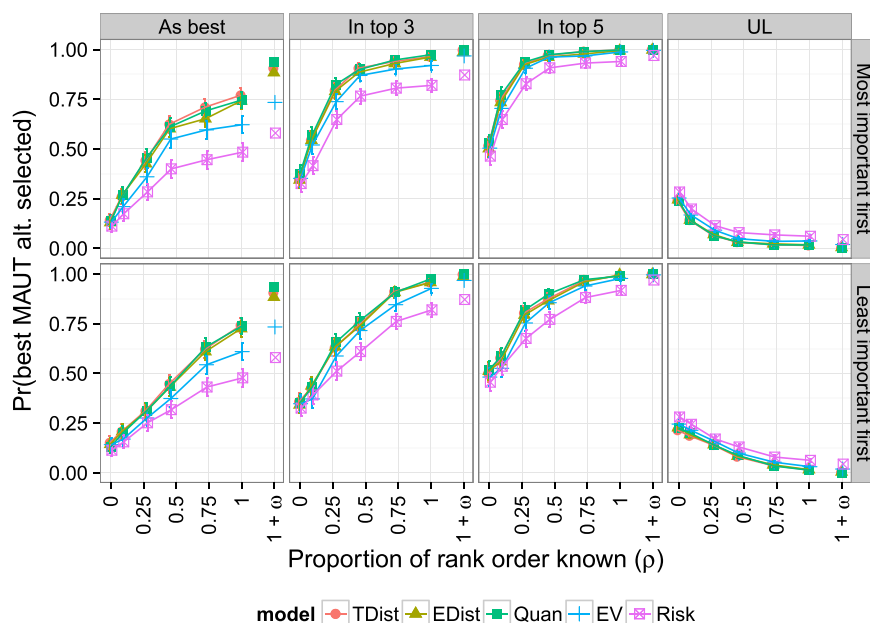


Fig. 2. Mean model accuracy as a function of the degree of imprecision in the weight space. Note ‘1+ ω ’ refers to the full-information case where the cardinal values of weights are known precisely.

models using the full probability distribution. Models using explicit risk measures, however, performed poorly on average. The same finding was reported in the context of utility/value function approaches with known weights in Durbach and Stewart [9]; the current result extends this to the SMAA context.

Finally, in the absence of assessment error model accuracy is almost always highest when only the first rank is used to construct the SMAA acceptability index. The sole exception is the risk-based model – the probability of the preferred MAUT alternative appearing in a shortlist is increased if the acceptability index is composed of the same number of ranks as the shortlist. However, when assessment errors are present, the same effect – the probability of the preferred MAUT alternative appearing in a shortlist is increased if the acceptability index is composed of the same number of ranks as the shortlist – is observed in other models too.

Fig. 4 shows two results related to the size of the decision problem considered. Model accuracy deteriorates as the number of alternatives increases, but improves as the number of attribute increases. This occurs for all of the simulated models, which to rough approximation experience the same magnitude of improvements and deteriorations.

Fig. 5 shows that in general skewness and variability in attribute evaluations lead to lower model accuracy. Models using simplified representations of uncertainty (those using quantiles, expected values, or variances) suffer far more from increases in variability or skewness of attribute evaluations than do SMAA models using full probability distributions. In fact, increasing attribute variability does not significantly change the accuracy of the SMAA models using full probability distributions. The net

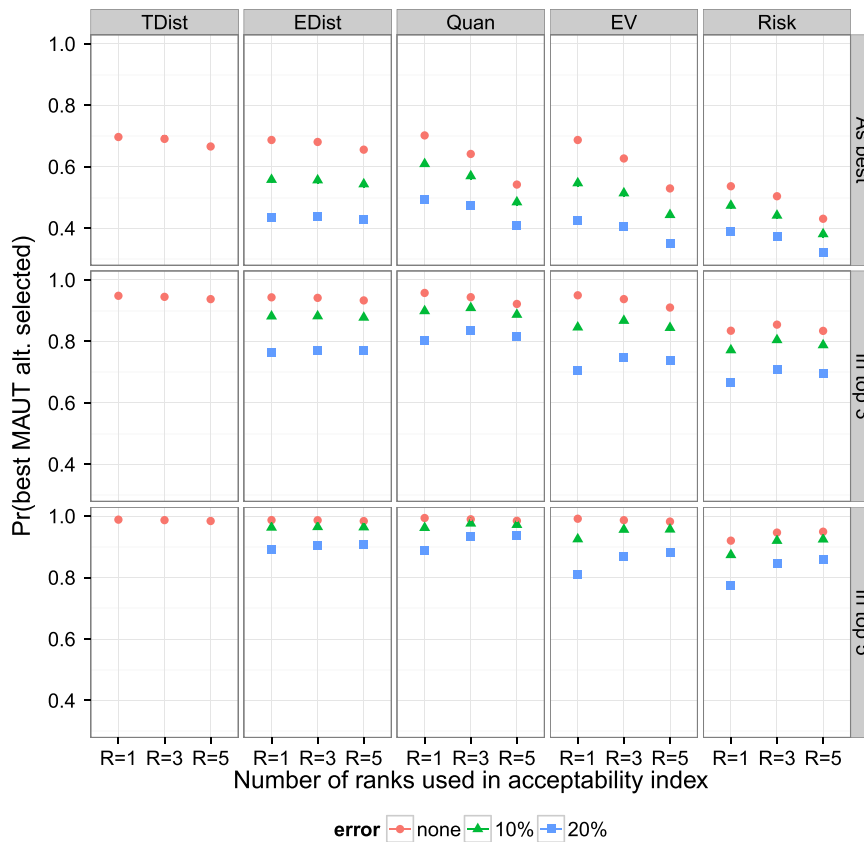


Fig. 3. Mean model accuracy as a function of assessment error levels and number of ranks used in the construction of acceptability indices.

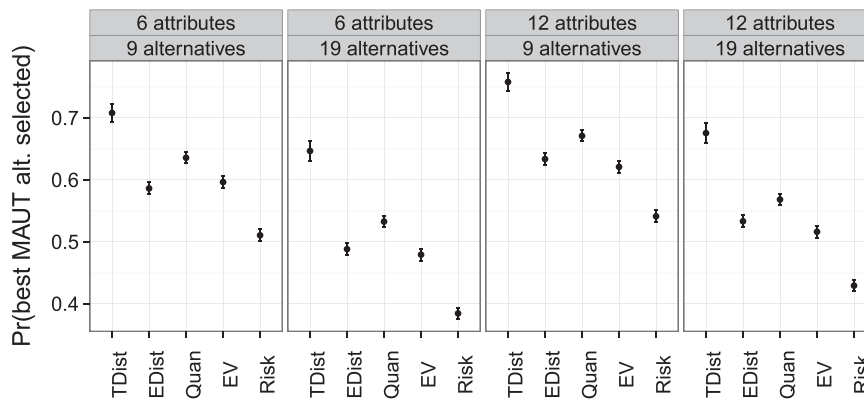


Fig. 4. Mean model accuracy as a function of the number of alternatives and attributes present in the decision problem.

effect is that quantile models become less attractive as distributions become more variable and skew.

Fig. 6 shows mean utility losses for different types of marginal value functions i.e. combinations of (τ, λ) , and risk preference parameter ζ . Model accuracy is highest for all models when marginal value functions are convex, deteriorating as these become linear and then concave. The effect of multi-attribute risk aversion ζ is substantially less than that of the shape of the marginal value functions, and contrasts are for the most part non-significant. The effect of risk preferences over value is only significant when marginal value functions are convex, shown in the right-hand plot in Fig. 6, in which case accuracy is highest when decision makers are multivariate risk neutral.

5. Discussion

Since SMAA always selects the alternative with the largest acceptability index, the probability that the SMAA model selects the same alternative as MAUT is equivalent to computing the probability that the decision maker's true preferences lie inside the largest favourable weight space. This is precisely what is computed by the largest rank-1 acceptability index. We therefore examine distributional properties of the random variable $B_{(k)}^1$ denoting the k -th largest rank-1 acceptability index, and in particular the expected value of the largest rank-1 acceptability index $E[B_{(1)}^1]$. Unfortunately deriving general analytical results for $B_{(k)}^1$ is difficult, so that we first consider an artificially simple case where

exact results can be obtained, and then extrapolate from these using heuristic arguments.

5.1. Model accuracy deteriorates as the number of alternatives increases

Suppose we have two attributes and I non-dominated alternatives, with known attribute evaluations and linear value functions. The weight space in this case consists of the line segment $[0, 1]$, i.e. the 1-simplex, from which w_1 and w_2 are obtained directly. The allocation of parts of the weight space to different alternatives, and hence the computation of acceptability indices, is analogous to the problem of breaking this line segment at random into I pieces, and randomly allocating each piece to an alternative. An alternative a_i is then selected with probability proportional to the length of the line segment allocated to it.

The distribution of the length of the k -th largest piece of the broken line segment is a well-studied problem in applied probability (where the line segment is often thought of as a "stick" of unit length) and is given by e.g. [11]

$$P(B_{(k)}^1 \leq x) = \sum_{m=0}^{k-1} \binom{I}{m} \sum_{n=0}^{I-m} (-1)^n \binom{I-m}{n} (1 - (m+n)x)_+^{I-1} \quad (9)$$

where $x_+ = \max(x, 0)$. Under the assumptions of the broken stick model,² this then is also the distribution of k -th largest rank-1 acceptability index, from which one obtains the expected k -th

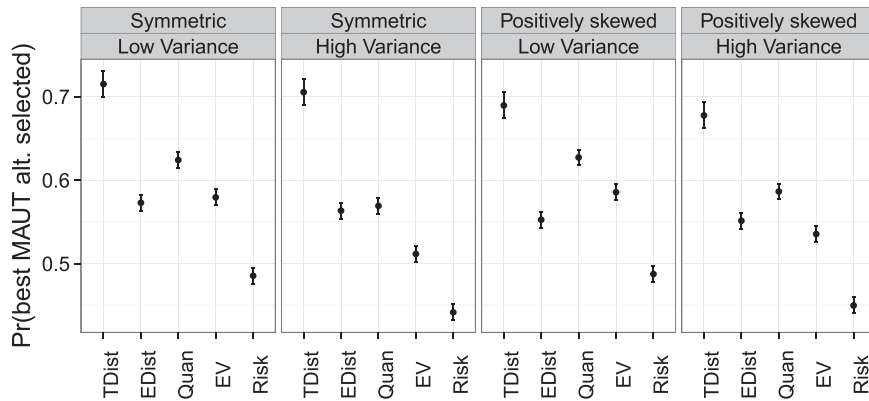


Fig. 5. Mean model accuracy as a function of the variance and skewness of distributions governing attribute evaluations.

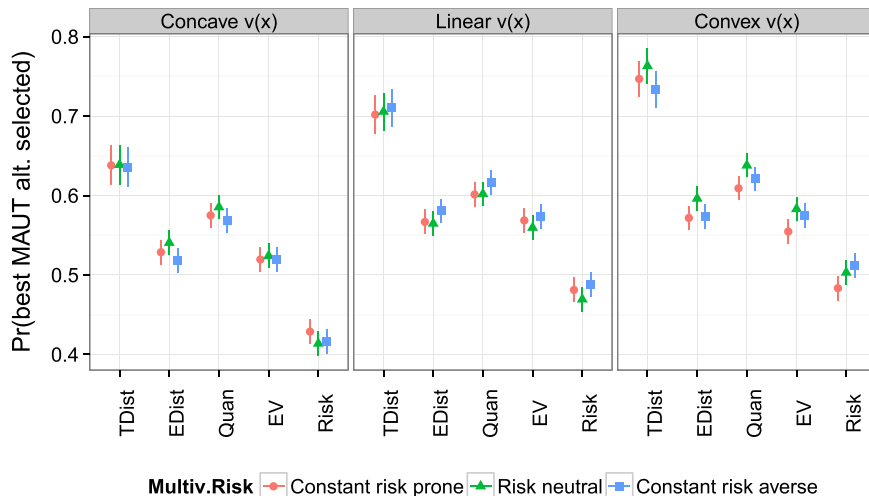


Fig. 6. Mean model accuracy as a function of DM preferences, as indicated by the shape of marginal value functions and risk preferences defined over global value.

largest rank-1 acceptability index as

$$E[B_{(k)}^1] = 1/I \sum_{m=0}^{I-k} 1/(I-m) \quad (10)$$

The largest rank-1 acceptability index, and hence the probability that the SMAA model selects the same alternative as MAUT, thus has expected value $E[B_{(1)}^1] = (1/I) \sum_{m=0}^{I-1} 1/(I-m) \sim \log(I)/I$, which decreases with the number of alternatives I .

5.2. Model accuracy improves as the number of attributes increases

With $J=2$ attributes the pieces of the broken stick directly give the probability of selecting alternative a_i . Extending this to $J > 2$ attributes requires $J-1$ sticks, each of which is broken at random into I pieces, with pieces allocated at random to alternatives. Alternative a_i is then selected with probability proportional to the size of the hypervolume formed by the pieces allocated to it, computed as the product of the lengths of its $J-1$ pieces. This construction effectively assumes that the length of piece i of stick j represents the marginal likelihood of selecting alternative a_i after integrating over all dimensions of the weight space other than the one representing attribute c_j . Obtaining distributional properties of order statistics over these product terms is difficult, but a heuristic argument can be used to show that the expected value of the largest acceptability index increases in the number of attributes used.

Since each line segment is assumed independent and identically distributed, the expected product of the k -th longest piece of each line segment is given by $\prod_{j=1}^{J-1} 1/I \sum_{m=0}^{I-k} 1/(I-m) \sim (\log(I-k+1)/I)^{J-1}$. This expected product decreases as the number of attributes J increases, but decreases much faster for larger k i.e. for products of small pieces than for products of large pieces. Thus, when renormalising the hypervolumes, the net effect is that the expected value of the largest acceptability index increases with J . A simple numerical experiment was used to confirm that this remains true when products are formed from arbitrary pieces of broken line segments, as described above. With two alternatives and two (ten, thirty) attributes $E[B_{(1)}^1] = 0.75$ (0.83, 0.89).

The preceding argument is based upon a problem context in which weights are entirely unknown. In our simulations ordinal rank information was provided on half the attribute weights, and thus an alternate explanation for the increase in J is that we are providing absolutely more weight information. This explanation was ruled out by observing the same effect in additional simulations run with no weight information ($\rho = 0$).

5.3. Model accuracy and ordinal weight importance information

In the absence of any weight importance information the probability the decision maker's true preferences lie inside the largest favourable weight space is simply the proportion of the total feasible weight space occupied by the largest favourable weight space. Restricting the weight space increases this probability whenever it is done in such a way that the total feasible weight space W' shrinks by a greater proportion than the largest favourable weight space, provided that the feasible weight space still contains the decision maker's true preferences.³ In a

² The assumptions of the broken stick model – uniformly and independently distributed breakpoints – may be less likely to hold in SMAA where there may be interactions between the regions of optimality for different alternatives linked to, for example, correlations in their performance.

³ Once restrictions to the weight space are applied we must take into account the possibility that the restrictions incorrectly exclude the decision maker's true preferences. In this case the probability that the SMAA and MAUT models select the same alternative is $E[B_{(1)}^1]P(\omega \in W')$ where ω is the decision maker's true weights.

uniformly distributed weight space, each ordinal constraint of the form $w_j > w_k$ eliminates half of the remaining weight space from consideration. Thus the first constraint eliminates half the weight space, the second a quarter, and so on. Knowing the importance ranking of one attribute introduces $J-1$ ordinal constraints and thus eliminates $100(1-0.5^{J-1})\%$ of the feasible weight space. Knowing the importance ranking of a second attribute introduces $J-2$ ordinal constraints and thus eliminates a fraction $1-0.5^{J-2}$ of the remaining feasible weight space. For example, for $J=5$ knowing the importance rank of one attribute eliminates 15/16 of the weight space, while knowing the rank of a second attribute eliminates 7/8 of the remaining 1/16, that is just 7/128 of the original weight space. Thus it is unsurprising that marginal improvements in accuracy diminish as more attribute importance rankings become known, as less and less of the weight space is removed at each step. In fact, given the argument above it is interesting that marginal improvements diminish as slowly as they do. The reason for this must lie in each weight restriction eliminating only slightly less of the largest favourable weight space, as a proportion of its original volume, than is true of the total feasible weight space.

To the extent that weight restrictions increase $b_{(1)}^1$, they do this by reducing the rank-1 acceptability indices of other alternatives. Since weights express trade-offs, alternatives that do well on an unimportant attribute must, other things being equal, do much better on this attribute to have any chance of being selected. Information which sorts between relatively unimportant attributes therefore, on average, eliminates relatively fewer alternatives – or equivalently, reduces the acceptability indices of other alternatives less – than information sorting between more important attributes.

5.4. Model accuracy and value functions

Model accuracy is highest when marginal value functions are convex and multivariate risk attitude is neutral. As the SMAA models all make the simplifying assumption that the DM is multivariate risk neutral, the latter result is to be expected and can be interpreted as an artifact of the simulation structure. More notable is the relative insensitivity of results to multivariate risk aversion, although it may be reasonably argued that this is because we varied the risk aversion parameter $\zeta \in [-2, 2]$ too narrowly.

A convincing explanation for the effect of the shape of the marginal value functions is more elusive. We can only conjecture here that convex marginal value functions help differentiate between good alternatives. For convex functions $v_j(x) - v_j(x-d) > v_j(y) - v_j(y-d)$ iff $x > y$, so that equal size differences in performances on the attribute scale translate into larger value difference the better the performance levels are. Thus, the difference between the largest and second-largest acceptability indices, $b_{(1)}^1 - b_{(2)}^1$, is likely to be larger when marginal value functions are convex.

5.5. Shortlists are more robust to assessment errors than choices

Errors in the assessment of the uncertainty of attribute evaluations introduce errors in the estimation of acceptability indices. If these estimation errors are sufficiently large the alternative with the largest (true) rank-1 acceptability index $b_{(1)}^1$ may not appear first in the SMAA rank order. If this occurs it is clear that it is more likely that alternative $a_{(1)}$ appears, for example, in the second rank than in the worst rank, or more generally in rank k than in rank $m > k$. Thus, assessment errors affect the probability of including the preferred MAUT alternative in a shortlist less than the probability of selecting this alternative as best.

It may be possible to derive analytical results on the necessary errors from the expression for $P(B_{(k)}^1 \leq x)$ in (9) by assuming, for example, that a rank reversal occurs when $B_{(1)}^1 - B_{(2)}^1 \leq \delta$, where δ is a random error in the estimation of acceptability indices. Given the difficulty of deriving $P(B_{(k)}^1)$ for more complex decision problems, and of deriving real-world recommendations from the error term δ , we do not pursue this further here.

6. Conclusion

Our simulation evaluates the ability of acceptability indices derived from SMAA models to approximate MAUT, and in particular SMAA models that use simplified formats for representing uncertainty in the attribute evaluations. The previous section elaborated on possible reasons for some of the more important of the simulation results.

In this section we summarize the contribution of the paper and conclude with some tentative recommendations for practice, including when different SMAA models might be more or less useful. These recommendations should be interpreted with caution though, since they require a judgment about what level of performance (e.g. in terms of the probability of selecting the same alternative as MAUT) is acceptable, which is necessarily problem dependent. Furthermore, any recommendations regarding responses to changes in a problem input (e.g. more attributes) that are derived from simulation results are heavily dependent on what values are assumed for other variables, and others may reasonably disagree with ours. Thus, our recommendations only hold with these two caveats in mind.

Our main conclusions regarding the implementation of SMAA to support low-involvement decision making are as follows:

1. Acceptability indices obtained from SMAA models can be expected to return reasonable approximations to the full MAUT model in certain contexts and for certain problematques (see below).
2. The accuracy of SMAA models improves as more ordinal weight importance information is provided, but
 - (a) marginal improvements diminish rather quickly with the number of the attributes for which importance rankings are specified, and
 - (b) the rate at which marginal returns diminish is greatest when the most important attributes are specified first, and
 - (c) if no weight information is provided at all, the average accuracy of all SMAA models is poor.
3. It is not essential that the SMAA model uses probability distributions to represent attribute uncertainty. Good results can also be obtained using simplified uncertainty formats, particularly quantiles.
4. Errors in the assessment of the uncertainty of attribute evaluations affect the probability of including the preferred MAUT alternative in a shortlist less than the probability of selecting this alternative as best.
5. The accuracy of the SMAA models deteriorates as attribute evaluations become more skew and, for those models using simplified representations of uncertainty, more variable. Quantile models thus become less attractive in highly variable problem contexts.
6. The accuracy of all SMAA models deteriorates as the number of alternatives increases, but improves as the number of attributes increases.

Taken together, our results suggest some role for SMAA in low-involvement decision making. It seems that, with limited preference information and summarised representations of uncertain attribute evaluations, the SMAA acceptability index can be used to rank order

alternatives and select one or a small number which, on average, perform relatively well. However under the wrong kinds of conditions performance can be poor and thus avoiding these conditions, or responding adequately to them, is an important part of providing good decision support. To this end our recommendations for practice are:

1. Deciding between alternatives based on SMAA acceptability indices without some restriction of the weight space cannot be recommended. There is at most a 50% chance that the preferred MAUT alternative appears in the top five out of 19 alternatives, even when no assessment errors are made.
2. The SMAA acceptability index can be justifiably used for deciding between alternatives if either the weight space is restricted or indices are used to shortlist rather than choose between alternatives. Conservatively, the most important 25% of the attribute set should be identified and ordered, and a shortlist of 15% or more of the alternatives set extracted.
3. The acceptability index should always use the same number of ranks as the size of the constructed shortlist. These are more robust to assessment errors and offer only slightly worse performance in error-free environments.
4. Uncertainty about attribute evaluations should be assessed using either the full probability distributions, or quantiles. Expected values and variance measures are more sensitive to assessment errors and worse on average than other models, even if correctly assessed. These are thus not recommended.
5. Errors in the assessment of uncertain attribute evaluations cause substantial deteriorations in model quality. If assessment errors are expected e.g. because of limited time or experience, the size of the constructed shortlist should be extended (speculatively, on the basis of simulation results, perhaps doubled).
6. The approach outlined above is suitable for large attribute sets and in fact will return higher-quality shortlists as the number of attributes increases. Alternatively, the DM may be able to specify importance information for a slightly smaller proportion of the attribute set or use a smaller shortlist of alternatives.

In a simulation study such as this we are unable to assess a number of practical issues. For example: whether the reduced accuracy is "acceptable" for DMs; whether the trade-off between accuracy and the time saved on weight elicitation is felt to be worthwhile; or whether the limited weight information that must be assessed for the SMAA models is felt to still be too onerous a task. More broadly, while efficient sampling procedures are now available for a wide range of weight constraints in MCDA, the practical assessment of these constraints remains unresolved. For example, are DMs able to reliably restrict the weight space, in the sense of offering constraints that do not exclude their true preferences? What types of constraints are DMs more comfortable providing information on? What are the best procedures for aiding this process? Investigating these important practical issues offers a clear path for future research.

References

- [1] Ahn BS, Park H. Establishing dominance between strategies with interval judgments of state probabilities. *Omega* 2014;49:53–59.
- [2] Angilella S, Corrente S, Greco S. Stochastic multiobjective acceptability analysis for the choquet integral preference model and the scale construction problem. *European Journal of Operational Research* 2015;240(1):172–182.
- [3] Barron FH, Barrett BE. Decision quality using ranked attribute weights. *Management Science* 1996;42(11):1515–1523.
- [4] Butler J, Jia J, Dyer J. Simulation techniques for the sensitivity of multi-criteria decision models. *European Journal of Operational Research* 1997;103:531–546.

- [5] Durbach I. A simulation-based test of stochastic multicriteria acceptability analysis using achievement functions. *European Journal of Operational Research* 2006;170:923–934.
- [6] Durbach I, Davis S. Decision support for selecting a shortlist of electricity-saving options: a modified SMAA approach. *ORiON* 2013;28(2):99–116.
- [7] Durbach I, Lahdelma R, Salminen P. The analytic hierarchy process with stochastic judgements. *European Journal of Operational Research* 2014;238(2):552–559.
- [8] Durbach IN, Stewart TJ. Using expected values to simplify decision making under uncertainty. *Omega* 2009;37(2):312–330.
- [9] Durbach IN, Stewart TJ. A comparison of simplified value function approaches for treating uncertainty in multi-criteria decision analysis. *Omega* 2012;40:456–464.
- [10] Hokkanen J, Lahdelma R, Miettinen K, Salminen P. Determining the implementation order of a general plan by using a multicriteria method. *Journal of Multi-Criteria Decision Analysis* 1998;7(5):273–284.
- [11] Holst L. On the lengths of the pieces of a stick broken at random. *Journal of Applied Probability* 1980:623–634.
- [12] Jessop A. Imp: a decision aid for multiattribute evaluation using imprecise weight estimates. *Omega* 2014;49:18–29.
- [13] Kadziński M, Tervonen T, Figueira JR. Robust multi-criteria sorting with the outranking preference model and characteristic profiles. *Omega* 2015;55:126–140.
- [14] Keefer DL, Bodily SE. Three-point approximations for continuous random variables. *Management Science* 1983;29(May (5)):595–609.
- [15] Keeney R, Raiffa H. *Decisions with multiple objectives: preferences and value tradeoffs*. New York: John Wiley & Sons; 1976.
- [16] Kirkwood C. Estimating the impact of uncertainty on deterministic multi-attribute evaluation. *Management Science* 1992;38(6):819–826.
- [17] Lahdelma R, Salminen P. SMAA-2: stochastic multi-criteria acceptability analysis for group decision making. *Operations Research* 2001;49(3):444–454.
- [18] Lahdelma R, Salminen P. Prospect theory and stochastic multicriteria acceptability analysis (SMAA). *Omega* 2009;37(5):961–971.
- [19] Lahdelma R, Salminen P. Stochastic multicriteria acceptability analysis (SMAA). In: Greco S, Ehrgott M, Figueira JR, editors. *Trends in multiple criteria decision analysis*. New York: Springer; 2012 (chapter 10).
- [20] Lahdelma R, Hokkanen J, Salminen P. SMAA—Stochastic multiobjective acceptability analysis. *European Journal of Operational Research* 1998;106:137–143.
- [21] Lahdelma R, Miettinen K, Salminen P. Reference point approach for multiple decision makers. *European Journal of Operational Research* 2005;164(3):785–791.
- [22] R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL (<http://www.R-project.org/>); 2014.
- [23] Salo AA, Hämäläinen RP. Preference ratios in multiattribute evaluation (PRIME)—elicitation and decision procedures under incomplete information. *IEEE Transactions on Systems, Man, and Cybernetics, Part A* 2001;31(6):533–545.
- [24] Stewart TJ. Robustness of additive value function methods in MCDM. *Journal of Multi-criteria Decision Analysis* 1996;5:301–309.
- [25] Tervonen T, Figueira JR. A survey on stochastic multicriteria acceptability analysis methods. *Journal of Multi-Criteria Decision Analysis* 2008;15(12):1–14.
- [26] Tervonen T, Lahdelma R. Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research* 2007;178(3):500–513.
- [27] Tervonen T, van Valkenhoef G, Baştürk N, Postmus D. Hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research* 2013;224(3):552–559.
- [28] van Valkenhoef G, Tervonen T, Postmus D. Notes on hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research* 2014;239(3):865–867.
- [29] van Valkenhoef G. SMAA: stochastic multi-criteria acceptability analysis. R package version 0.2-3. URL (<http://CRAN.R-project.org/package=smaa>); 2014.
- [30] van Valkenhoef G, Tervonen T. hitandrun: hit and run and shake and bake for sampling uniformly from convex shapes. R package version 0.5. URL (<http://CRAN.R-project.org/package=hitandrun>); 2015.