

The analytic hierarchy process with stochastic judgements

Ian Durbach^{*†}, Risto Lahdelma[‡], Pekka Salminen[§]

Abstract

The analytic hierarchy process (AHP) is a widely-used method for multicriteria decision support based on the hierarchical decomposition of objectives, evaluation of preferences through pairwise comparisons, and a subsequent aggregation into global evaluations. The current paper integrates the AHP with stochastic multicriteria acceptability analysis (SMAA), an inverse-preference method, to allow the pairwise comparisons to be uncertain. A simulation experiment is used to assess how the consistency of judgements and the ability of the SMAA-AHP model to discern the best alternative deteriorates as uncertainty increases. Across a range of simulated problems results indicate that, according to conventional benchmarks, judgements are likely to remain consistent unless uncertainty is severe, but that the presence of uncertainty in almost any degree is sufficient to make the choice of best alternative unclear.

Keywords: decision analysis; multicriteria; analytic hierarchy process; uncertainty; simulation

1 Introduction

The analytic hierarchy process (Saaty, 1990) is a widely-used method for multicriteria decision support based on a hierarchical decomposition of a decision problem into multiple criteria, the assessment of preferences using pairwise comparisons, and an aggregation of these pairwise preferences into an overall evaluation of the alternatives. While a number of practical and theoretical aspects of the AHP have proved controversial (see for example the discussion in Belton and Stewart (2002)), it has found widespread application and acceptance in practice (e.g. Vaidya and Kumar, 2006), to the extent that it may well be among the most-frequently applied of currently available methods for decision support.

At the heart of the method is a nine-point semantic scale used by decision makers to express their preferences for one alternative over another on a particular criterion, and for

^{*}Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, South Africa. Email: ian.durbach@uct.ac.za

[†]African Institute for Mathematical Sciences, 6-8 Melrose Road, Muizenberg 7945, South Africa

[‡]Department of Energy Technology, School of Engineering, Aalto University, Otakaari 4, FIN-02150 Espoo, Finland. E-mail: risto.lahdelma@aalto.fi

[§]School of Business and Economics, University of Jyväskylä, P.O. Box 35, FIN-40014 University of Jyväskylä, Finland. Email: pekka.o.salminen@jyu.fi

how much one criterion is valued over another. It is clear that sometimes these assessments will be subject to uncertainty – meaning that the decision maker (DM) does not possess the necessary information to describe or deterministically predict the inputs required by the AHP (see Durbach and Stewart (2012) for a review of uncertainty in multicriteria decision support). Although the standard AHP method does not directly treat uncertainty or imprecision in its inputs, a number of extensions have been proposed to address this issue, using for example fuzzy set theory (Van Laarhoven and Pedrycz, 1983; Buckley, 1985; Boender et al., 1989), interval arithmetics (Salo and Hämäläinen, 1995), and various stochastic techniques (Saaty and Vargas, 1987; Hauser and Tadikamalla, 1996).

This paper adds to that body of work by introducing a simulation-based method for representing imprecise or uncertain pairwise comparison information from one or more DMs through stochastic distributions, and a computational method to treat this information in the analysis. The method is a variant of stochastic multicriteria acceptability analysis (SMAA; see Lahdelma et al. (1998); Lahdelma and Salminen (2001); Tervonen et al. (2008)), an inverse-preference methodology applied here to the case of the AHP. The resulting SMAA-AHP can be used with arbitrary independent or dependent distributions for the comparisons, and is based on Monte Carlo simulation from probability distributions appropriately defined over any uncertain pairwise comparisons and a subsequent collection of statistics summarising the performance of each alternative. SMAA-AHP is related to other simulation-based methods, most notably Hauser and Tadikamalla (1996), but presents additional information to the DM, defines uncertainty regions differently, and uses a different distribution for the uncertain judgements. SMAA-AHP also allows more flexible representation of weight constraints and can also be used with missing preference information.

The remainder of the paper is organized as follows. Section 2 reviews uncertainty modelling in the AHP. Section 3 describes the SMAA-AHP method. Section 4 demonstrates the method using a small example. Section 5 discusses the advantages and potential problems with the method, guided by the results of a simulation study. A final section concludes the paper.

2 Uncertainty modelling in the AHP

In the following, we consider a decision problem consisting of I alternatives, each evaluated on K criteria. Let z_{ik} be the evaluation of alternative i in terms of criterion k , according

to some suitable performance measure. In the standard AHP the DM performs pairwise comparisons at each node of the objectives hierarchy, expressing their preferences for one alternative over another on a particular criterion, or for how much one criterion is valued over another. The pairwise preference a_{ijk} for alternative i over alternative j on criterion k represents the ratio between evaluations z_{ik}/z_{jk} , expressed on a discrete scale from 1 to 9 (where 1 means equal preference and 9 denotes absolute preference). Where convenient, we drop the criterion subscript and refer simply to the pairwise evaluation a_{ij} . The same approach is used to compare the importance of criteria, in which case we refer to a pairwise preference a_{ij} for criterion i over criterion j representing the ratio between trade-off weights w_i/w_j . In cases where pairwise comparisons can be assessed precisely, a number of ways have been proposed to aggregate these into global measures of performance (Belton and Stewart, 2002). Most commonly, the eigenvector corresponding to the largest eigenvalue of the $(I \times I$ or $K \times K)$ pairwise comparison matrix $\mathbf{A} = [a_{ij}]$ is extracted (the so-called priority vector), and a global evaluation formed by a simple weighted sum.

Our concern is with decision making situations in which the pairwise evaluations a_{ijk} (and consequently computed values for z_{ik} and w_j) are uncertain. Early research into the modelling of probabilities in the AHP was largely concerned with deriving relationships between the distributional form of the uncertain pairwise judgements and the distributions of the marginal evaluations contained in the priority vector (Vargas, 1982; Saaty and Vargas, 1987; Basak, 1989, 1991). Subsequent probabilistic AHP models (Hauser and Tadikamalla, 1996; Levary and Wan, 1998, 1999; Basak, 1998; Banuelas and Antony, 2007) have focused on using Monte Carlo simulation to randomly generate pairwise evaluations from the distributions specified by decision makers. These approaches all follow the same basic approach, first expressed by Hauser and Tadikamalla (1996). The decision maker expresses pairwise comparisons in the usual way i.e. using the same 1-9 scale as for deterministic AHP, except that these comparisons are allowed to be random variables with associated probability distributions. Hauser and Tadikamalla generated random judgements a_{ij}^* uniformly on the interval $[a_{ij} - da_{ij}, a_{ij} + da_{ij}]$, with d an uncertainty factor, before transforming any values less than one using $f(a_{ij}^*) = 1/(2 - a_{ij}^*)$. Further restrictions may be placed on the types of distributions if necessary. Next sets of random pairwise judgements are generated using Monte Carlo simulation. For each set of randomly generated evaluation matrices the priority vector is computed. Repeating this process many times gives a distribution of priorities for each alternative, which can be used to rank the

alternatives, in most cases using the mean of the distribution.

Most authors make small embellishments around this general process. Levary and Wan (1998) incorporate scenarios into their model (see also Levary and Wan (1999)). Decision makers thus assess different (possibly stochastic) judgemental matrices for each scenario. Their simulation approach first generates a random number to specify which scenario is being used, and then generates further random numbers specifying the pairwise judgements within each scenario. Basak (1998) uses a Bayesian approach to integrate expert judgements with the decision maker's prior probabilistic assessments. Pairwise judgements are simulated by drawing from the posterior distributions. Banuelas and Antony (2007) add a sensitivity analysis phase to investigate the influence of the probabilistic judgements on the consistency index. As mentioned above the primary distinction between existing simulation-based AHP methods and SMAA-AHP is the additional information that is presented to DMs, which can be useful in facilitating a greater understanding of the decision problem and progressing towards a final decision. We discuss this information in the presentation of the SMAA-AHP given in the following section.

3 The SMAA-AHP method

In SMAA-AHP, the DMs may express their comparisons on a discrete scale from 1 to 9 or use arbitrary positive values. The DMs can give their pairwise comparisons either as precise values, as in AHP, or as intervals to express imprecise or uncertain preferences. The DMs can give the lower and upper bounds of the intervals explicitly, or express them as $[a_{ij}/d_{ij}, a_{ij}d_{ij}]$ where a_{ij} is the geometric mean of the interval and $d_{ij} \geq 1$ is the so-called imprecision factor of their pairwise comparison. For example, the interval $[0.5, 8]$ corresponds to the pairwise comparison 2 with imprecision factor 4. The imprecision factor is a meaningful way to express uncertainty on a ratio scale, where all values should be positive.

When the DMs express their pairwise comparisons, it should be checked that these are sufficiently consistent. In the original AHP, where pairwise comparisons are expressed deterministically, a popular approach for evaluating consistency is to compare λ_1 , the leading eigenvalue of an assessed pairwise comparison matrix, with I , the leading eigenvalue obtained from an $I \times I$ matrix of perfectly consistent judgements (in the sense that $a_{ik} = a_{ij}a_{jk}$, $\forall i, j, k$). To provide a measure of the severity of this deviation, $(\lambda_1 - I)/(I - 1)$ is compared with the mean inconsistency value derived from many randomly generated

reciprocal matrices of the same size. An inconsistency ratio of 0.1 or less is generally stated to be acceptable (Saaty, 1990), meaning that the inconsistency of the observed pairwise comparisons should be no more than 10% of what would be observed, on average, from completely random judgements. For an interval-based analysis, a natural analogue would be to suggest that the geometric mean of the comparisons should have a inconsistency ratio below 10%. Note, however, that this benchmark, as well as the general use of the inconsistency ratio, has been strongly criticised (see in particular Bana e Costa and Vansnick (2008))

After each DM has given his/her pairwise comparisons, we combine them into intervals $[a_{ij}^{\min}, a_{ij}^{\max}]$ where a_{ij}^{\min} is the minimal value that any DM has expressed and a_{ij}^{\max} is the maximal value. We represent then the aggregated comparison values by stochastic variables with suitable probability distributions. Technically, it is possible to use arbitrary distributions. However, in the absence of information about the distribution, we apply the truncated and scaled $1/x$ distribution. The PDF (probability density function) of the scaled $1/x$ distribution is given by $f(x) = \alpha/x$ when $x \in [x^{\min}, x^{\max}]$, and zero elsewhere. The scaling coefficient $\alpha = 1/\ln(x^{\max} - x^{\min})$ is determined so that the integral over the PDF equals one.

The motivation for using the scaled $1/x$ distribution to represent pairwise comparisons in an interval is that this distribution allocates equal probability mass for all sub-intervals $[x/d, xd]$ corresponding to the same imprecision factor d . For example, given a pairwise comparison interval $[0.5, 8]$, the scaled $1/x$ distribution allocates equal probability mass of $1/4$ for each of the subintervals $[0.5, 1]$, $[1, 2]$, $[2, 4]$, and $[4, 8]$. If the interval is degenerate, i.e. $a_{ij}^{\min} = a_{ij}^{\max}$, we use Dirac's delta function (the unit impulse function) as the distribution.

After representing the aggregated pairwise comparisons by suitable distributions, we analyse the performance of each alternative through stochastic simulation by simultaneously drawing pairwise comparisons from their corresponding distributions and computing the score for each alternative as in AHP. Observe that even if the pairwise comparisons given by each DM are (sufficiently) consistent, comparisons drawn from the combined intervals can be inconsistent. It is easy to reject during the simulation comparisons whose inconsistency ratio exceeds some threshold. However, inconsistent comparisons can also be included in the computations, because the random inconsistencies do not introduce any systematic bias to the results.

During the simulation we collect statistics about the weights at different nodes of the hierarchy, the overall score of the alternatives, and their ranking. In particular, we collect the following statistics: the number of times alternative i obtains rank r (B_{ir}), the number of times alternative i scores better than alternative j (C_{ij}), the sum of scores for alternative i at node t in the hierarchy (S_{it}), the sum of the weights for criterion k during the iterations when alternative i obtained first rank (W_{ik}), and the number of times alternative i obtained first rank using its central weights (P_i , computation of which requires a second simulation round after the central weights have been computed). A sufficient number of simulation runs K to obtain 95% confidence limits of 0.01 around most outputs (most notably, acceptability) is approximately 10 000 (Tervonen and Lahdelma, 2007). Based on the statistics, we compute the following descriptive measures for evaluating the alternatives:

- *Average criterion score* for different alternatives at each criterion node. This generalizes the corresponding crisp AHP criterion scores to consider imprecise comparison values. The average criterion score is computed as $s_t^i = S_{it}/K$.
- *Average overall score* for different alternatives. This generalizes the crisp AHP overall score to consider imprecise comparison values. The average overall score is simply the average criterion score s_t^i for the root node t .
- The *rank acceptability index* b_i^r measures the variety of different preferences for which alternative i obtains rank r . The rank acceptability indices can be used for ranking the alternatives roughly, or for finding compromise alternatives in case no alternative obtains sufficient acceptability for the first rank. Potential compromise alternatives are those with high acceptability for the best ranks. Alternatives that obtain high acceptability for the worst ranks should be avoided. The rank acceptability index is computed as $b_i^r = B_{ir}/K$.
- The first rank *acceptability index* b_i^1 measures the variety of different preferences that make alternative i most preferred. In other words, the acceptability index measures how widely acceptable the alternative is. The acceptability index can be interpreted as the share of people voting for the alternative, assuming that the applied distribution for comparison values represents the voters' preferences. Zero acceptability means that the alternative is inefficient, i.e. no preferences make it best.

- The *central weight vector* \mathbf{w}_i^c describes what kinds of weights are favourable for alternative i , i.e. make it most preferred. The central weights 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 weights are undefined for inefficient alternatives. The central weights are computed as $w_{ik}^c = W_{ik}/B_{i1}$.
- The *pairwise winning index* c_{ij} is the probability for alternative i to score better than alternative j considering the uncertainty in the preference statements. The pairwise winning indices are computed as $c_{ij} = C_{ij}/K$.
- The *confidence factor* p_i^c is the probability for alternative i to be most preferred when the central weight vector for that alternative is applied. In other words, the confidence factor measures if the performance of the alternative has been assessed accurately enough, so that it can be selected under favourable preferences between criteria. The confidence factors are computed as $p_i^c = P_i/K$.

4 Illustrative example

To illustrate the SMAA-AHP method, we reconsider the AHP example originally used in Saaty (1990). First we consider the problem with precise comparison values, then with imprecise comparisons, and finally with missing comparisons.

Tables 1 and 2 show the pairwise comparisons assessed between the criteria and alternatives respectively. Only the upper triangle of each reciprocal comparison matrix is presented. The criterion comparison matrix in Table 1 is a little inconsistent with consistency index $CI = 0.14$ and inconsistency ratio $IR = 11\%$, as is the comparison matrix for criterion four ($CI = 0.10$, $IR = 20\%$). The remaining comparison matrices have $IR < 10\%$, the conventional benchmark (Saaty, 1990).

Crit.	2	3	4	5	6
1	5	7	5	3	1
2		3	1/5	1/6	1/6
3			1/4	1/5	1/5
4				1/5	1/6
5					1

Table 1: Pairwise comparisons between criteria.

Using the standard AHP approach, alternative A obtains the highest score (normalised to sum to 100) of 40 followed by B at 36 and C at 24. Suppose now that the inputs above

Alt.	B	C
A	(1/3,1,5,9,1/2,6)	(1/2,1,1,7,1,4)
B		(3,1,1/5,1/5,2,1/3)

Table 2: Pairwise comparisons between alternatives with respect to different criteria. For compactness comparisons are displayed in vector form $(a_{ij1}, a_{ij2}, \dots, a_{ijK})$.

Alt.	s_i^t	b^1	b^2	b^3	p^c	\mathbf{w}_i^c	c_{iA}	c_{iB}	c_{iC}
A	40	86	14	0	95	[33,5,3,9,23,28]	0	86	100
B	36	14	86	0.005	30	[35,5,3,9,23,25]	14	0	99.995
C	24	0	0.005	99.995	0	–	0	0.005	0

Table 3: Output from the SMAA-AHP model: rank acceptability indices, confidence factors, central weights, and pairwise winning indices (all expressed as %).

are not precise but are effectively expectations (or other forms of ‘best guesses’) around which substantial uncertainty exists. For simplicity we use the same uncertainty factor $d_{ij} = 1.5$ for all comparisons a_{ij} .

Output from the SMAA-AHP model is shown in Table 3. Alternative A remains the most likely candidate for the first rank, with a first rank acceptability of 86%. However alternative B , which obtains $b^1 = 14\%$, is also potentially optimal. In practice the DMs could either decide that this result is conclusive enough, or determine that more accurate evaluation of the alternatives may be necessary. Similar central weights are observed for the two potentially optimal alternatives – to be expected because these are restricted by the pairwise comparisons between criteria. The confidence factors show that under favourable preferences alternative A is very likely (95%) the most preferred one. In contrast alternative B receives only 30% confidence, meaning it is unlikely to be the best option even when its central weights are applied. Choosing B would require more precise comparisons to be collected, and for that new information to favour B .

The use of simulation allows for the collection of detailed statistics on consistency and its effect on results. Basic consistency information is summarized in Figure 1, which shows empirical distributions of inconsistency ratios for each of the seven pairwise comparison matrices. Clearly the greatest inconsistencies lie in the pairwise comparisons between criteria and within criterion 4, and to a lesser extent within criterion 1 and 6. For all of these, the conventional 10% upper bound on consistency is regularly violated (in 82%, 83%, 26%, and 26% of simulations, for the four cases respectively) by the matrices simulated as part of SMAA-AHP. This provides the user with information on the consistency of the preferences implied by the combination of initial judgements and imprecision factor d ,

either in the form of the full empirical distributions or summarized as the proportion of simulated judgments violating some consistency benchmark such as 10%.

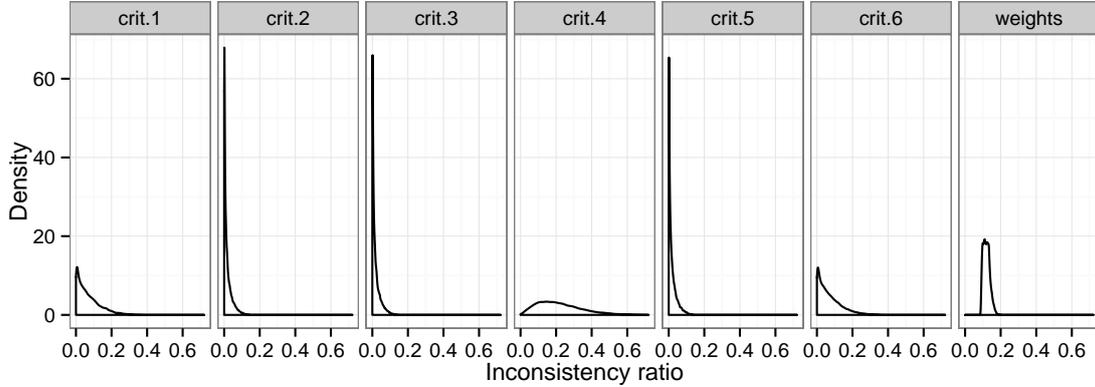


Figure 1: Distributions of inconsistencies in the simulated judgments.

Of particular interest is whether these inconsistent judgements are linked to preferences for particular alternatives. This information can be conveyed in a number of ways. Figure 2 shows the average first rank acceptability indices for alternatives as a function of the observed inconsistency in each comparison matrix. It is important to note that because consistency refers to the pairwise comparison matrix as a whole, two simulated matrices with the same inconsistency ratio may differ substantially in their pairwise judgments, and hence the ranking of alternatives and subsequent acceptabilities that are derived from them. Therefore, to ease interpretation, the averages shown in Figure 2 have been smoothed with a simple local smoother using a neighbourhood of the closest 10% of observations and a cubic inverse distance function to weight those observations, although our conclusions are relatively insensitive to these choices. Our aim in doing so is to provide a summarized view of how preferences change, on average, with increased inconsistency.

In Figure 2(a), which uses the same imprecision factor $d = 1.5$ for all judgements, it is clear that acceptabilities remain roughly the same regardless of the consistency of simulated judgements. However, this need not always be the case, particularly where uncertainty is not symmetric around the initial point judgement. Figure 2(b) provides an illustration of results obtained if only those between-criteria comparisons a_{12} , a_{15} , and a_{25} are uncertain and these are generated $U[0.5, 2]$; all within-criteria comparisons are still subject to an imprecision factor $d = 1.5$. In that case the overall acceptability of alternative B increases marginally to 18%, but the bottom panel of Figure 2(b) reveals that much of this increase in acceptability is derived from inconsistent simulated between-criteria

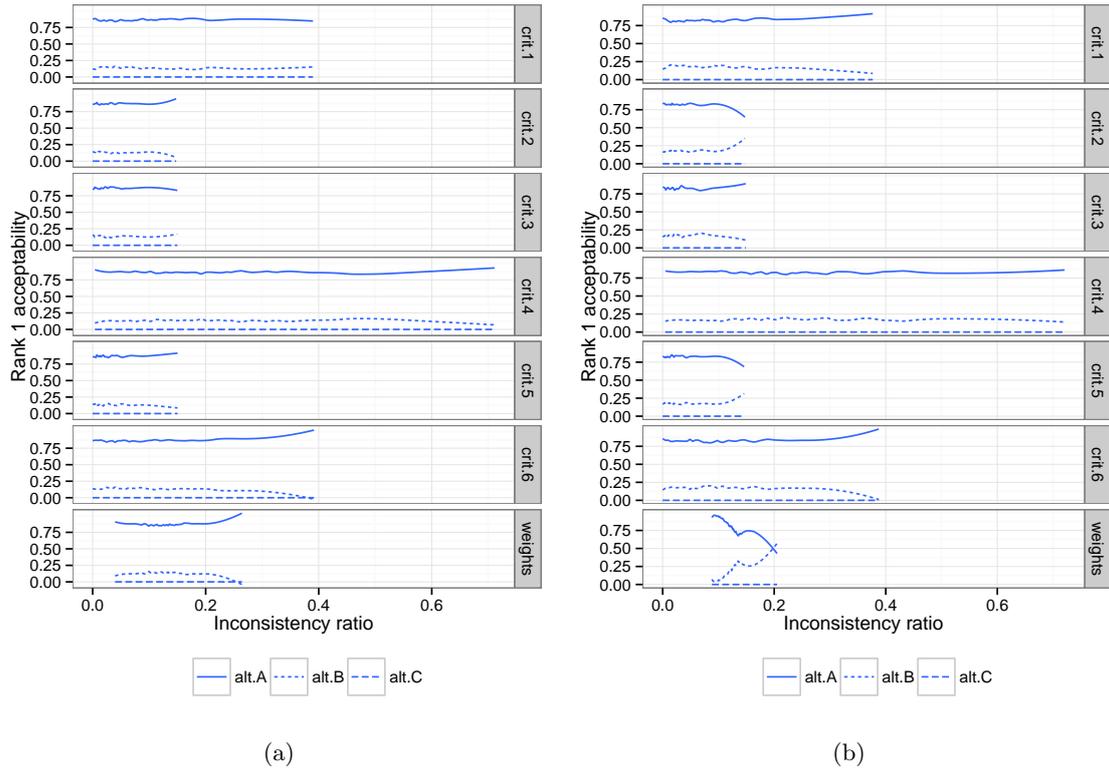


Figure 2: Effect of inconsistency on rank-1 acceptability indices.

judgements, and that under highly inconsistent between-criteria assessments alternative A attains higher acceptability, on average, than B .

Where acceptabilities are sensitive to judgmental consistency, two broad responses are possible. The first is simply to reject any simulated matrices whose inconsistency is above a certain threshold (e.g. the usual benchmark of 10%). A second option, arguably more closely aligned with the goal of decision support, is to interrogate the inconsistent judgements that are favouring a particular alternative, in an attempt to understand the sources of the inconsistency and ultimately to resolve these, either through more precise evaluations or rejecting certain judgements. Simulated comparison matrices favouring each alternative may be collected and stored as part of the SMAA process, so that those above a certain consistency threshold can be easily identified and summarized. To illustrate, in Table 4 we summarize 250 between-criteria comparison matrices which obtain a inconsistency ratio above 0.17 and hence, from Figure 2(b), can be seen to increasingly favour alternative B . Here, $a_{12} = 2.93$ and $a_{52} = 8.68$ so that it would appear that a_{51} should be greater than 1 and indeed perfect consistency would imply $a_{15} = 2.93/8.68 = 0.34$. The geometric mean of the inconsistent simulated comparisons indicates, however, a strong

	c_1	c_2	c_3	c_4	c_5	c_6
c_1	1	2.93	7	5	5.41	1
c_2	0.34	1	3	0.2	0.12	0.17
c_3	0.14	0.33	1	0.25	0.20	0.2
c_4	0.20	5	4	1	0.20	0.17
c_5	0.18	8.68	5	5	1	1
c_6	1	6	5	6	1	1

Table 4: Mean between-criteria pairwise comparisons, among those matrices obtaining a inconsistency ratio above 0.17.

Alt.	s_i^t	b^1	b^2	b^3	p^c	\mathbf{w}_i^c	c_{iA}	c_{iB}	c_{iC}
<i>A</i>	44	82	15	3	100	[14,17,18,19,15,19]	0	84	96
<i>B</i>	28	16	31	52	99.7	[33,17,10,7,26,8]	16	0	47
<i>C</i>	28	1	54	45	24	[16,23,40,4,13,5]	4	53	0

Table 5: Output from the SMAA-AHP model with missing information (all figures expressed as %).

preference for criteria 1 over 5 ($a_{15} = 5.41$). At this stage the decision maker would need to decide whether the inconsistent simulated judgments are merely artifacts of the simulation that can be summarily rejected, and/or whether further, more refined assessments are required.

Finally, we demonstrate the use of SMAA-AHP in the case where comparison information between criteria is completely missing. We represent this missing information using non-negative normalized weights that follow a joint uniform distribution. Table 5 shows the output from the revised SMAA-AHP model.

Average overall scores indicate that *A* increases its superiority, while *B* and *C* appear almost equally good. Rank acceptability indices show how the increased uncertainty in the comparisons is reflected as increased uncertainty in the ranking. The confidence factors show that both *A* and *B* can be chosen under favourable preferences, but *C* is likely not the best alternative under any preferences. The central weights identify what kind of trade-off ratios between the criteria make each alternative most preferred. We can see that different alternatives are favoured by dramatically different weights. For example, alternative *C* would require about 40% of the weight to be placed on criterion 3 alone. The pairwise winning indices reveal the surprising fact that while *B* is superior to *C* when competing about the first rank, *C* is preferred to *B* on average. An analysis of consistency information is not provided here, but would be conducted as for the partial-information case described above.

5 Simulation study

The purpose of the simulation study is to evaluate the influence of imprecision on model results. In particular we wish to evaluate the rate at which the consistency of the resulting judgements and the ability of the SMAA-AHP model to discern the best-performing alternative deteriorate as imprecision, as captured by the imprecision factor d , increases in magnitude.

5.1 Simulation design

The basic structure of each simulation run is as follows:

1. Select problem size parameters (number of alternatives I and criteria K) and generate $K + 1$ pairwise comparison matrices.
2. Select an imprecision factor d and select which of the pairwise comparison matrices are to be imprecise. Modify the appropriate pairwise judgements to reflect imprecision.
3. Simulate the application of SMAA-AHP to the resulting imprecise decision problem.
4. Evaluate model results.

By performing a number of simulation runs at each combination of parameters, aggregate statistics can be collected and the mean performance of the SMAA-AHP model assessed and compared statistically across a range of simulated conditions. Further details on the simulation structure are given below.

5.1.1 Generating pairwise comparison matrices

For each criterion k , pairwise comparisons a_{ij} are generated by first simulating the evaluation of alternative i on criterion k , denoted z_{ik} , for all i and k . The evaluations are drawn from $U[0.1, 0.9]$ and normalized so that each alternative's scores sums to one over all criteria. This ensures that all initial ratio judgements will lie between 1/9 and 9 as per the AHP scale, and that all alternatives are non-dominated. Pairwise comparisons are computed from the resulting normalized evaluations $a_{ij} = z_i/z_j$, where for convenience we have dropped the subscript k . Pairwise comparisons of criterion importance are done in the same way, except that no normalization is required. All initial judgements are therefore perfectly consistent.

5.1.2 Introducing imprecision into pairwise judgements

We introduce imprecision into the simulated pairwise comparisons via the imprecision factor d described in Section 3. This factor is left as a parameter of the simulation, taking on a range of values between 1 and 5. We also differentiate between imprecision involving evaluations of alternative performance and imprecision involving evaluations of criterion importance. This gives three conditions: (a) both pairwise comparisons of criterion importance and pairwise comparisons of performances on each criterion are imprecise; (b) pairwise comparisons of criterion importance are imprecise but pairwise comparisons of performances on each criterion are not; (c) pairwise comparisons of performances on each criterion are imprecise but pairwise comparisons of criterion importance are not. For simplicity, the same values of d are used for both types of imprecision.

5.1.3 Simulating the application of SMAA-AHP

Once the imprecision factor d has been chosen, a single iteration of the SMAA-AHP model involves replacing each pairwise comparison a_{ij} in the upper triangular part of the pairwise comparison matrix with a randomly drawn value a_{ij}^* from the scaled $1/x$ distribution defined between $[a_{ij}/d, a_{ij}d]$. In this case, $x^{\max}/x^{\min} = d^2$ and so $\alpha = 1/\ln(d^2)$. We do not generate random values in the lower triangular part of the pairwise comparison matrix so as to preserve the relation $a_{ij}^* = 1/a_{ij}^*$, and also allow the generated values to lie beyond the usual AHP 1-9 scale. Once all a_{ij}^* have been generated, scores are computed as in the usual AHP model. That is, for each pairwise comparison matrix we extract the eigenvector corresponding to the largest eigenvalue, and aggregate the resulting scores using an additive model. Alternatives are ranked in descending order of global score, and we note which alternative is ranked in each position (for later calculation of acceptability indices).

In implementing the SMAA-AHP model, acceptability results are based on 10 000 iterations of the procedure described above, following results in Tervonen and Lahdelma (2007). Note that the pairwise judgements generated by SMAA-AHP i.e. the a_{ij}^* , may well be inconsistent. We do not reject inconsistent judgements at this stage; rather we simply compute the average inconsistency and report how this varies with changes to the imprecision factor d as well as other parameters of the simulation.

5.1.4 Evaluating model results

Information is collected on the following outcome measures:

1. b^1 : Mean acceptability of the first-ranked alternative;
2. $b^1 - b^2$: Mean difference between the acceptability of the first-ranked and second-ranked alternatives, giving some indication of the degree to which a best alternative can be discerned;
3. $\Pr(b_i^1 > 0)$: Mean proportion of alternatives with non-zero acceptability indices, bearing in mind that all alternatives are by construction Pareto optimal.
4. IR : Mean inconsistency ratio evaluated across all matrices for which $d > 1$.
5. $\Pr(IR > 0.10)$: Mean proportion of imprecise matrices (i.e. those for which $d > 1$) with inconsistency ratios above the benchmark of 0.10.

5.1.5 Parameter values used in the simulation study

We carry out the following three sets of experiments:

1. The imprecision factor $d \in \{1.1, 1.2, 1.4, 1.6, 1.8, 2, 2.5, 3, 4, 5\}$ is varied, holding $I = 10$ and $K = 8$ fixed and with both types of imprecision present.
2. The type of imprecision present (criterion importance, alternative performance, or both) is varied, for each element in a subset of imprecision factors $d \in \{1, 1.5, 2, 2.5, 3, 4, 5\}$ and holding $I = 10$ and $K = 8$ fixed.
3. The problem size parameters governing the number of alternatives and criteria are varied, using four different problems sizes: $(I = 5, K = 4)$; $(I = 10, K = 16)$; $(I = 20, K = 8)$; and $(I = 20, K = 16)$ with both types of imprecision present and for each element in a subset of imprecision factors $d \in \{1, 1.5, 2, 2.5, 3, 4, 5\}$.

For each combination of simulation parameters 50 simulation runs are performed, giving standard errors of at most 0.05 for any group means discussed in the results.

5.2 Results

Figure 3 shows how the acceptability and consistency results obtained from SMAA-AHP change as a function of the imprecision parameter d , for the base case of $I = 10$ and

$K = 8$, with uncertainty in both alternative performance and criterion importance evaluations. With no imprecision ($d = 1$) the acceptability of the best-ranked alternative is by definition 1, so that all other alternatives receive zero acceptability. As imprecision increases the acceptability of the best-ranked alternative decreases sharply and more alternatives become potentially optimal. Our results indicate that at seemingly moderate degrees of imprecision (i.e. $d = 1.5$) the best-ranked alternative will receive on average around 70% acceptability, with the second-ranked alternative receiving around 20%. At this stage, upwards of 60% of the alternatives would be identified as Pareto optimal. Naturally these figures may vary greatly from case to case, but an acceptability of 20% would certainly prevent ruling out this alternative (and simply selecting the alternative with the highest acceptability index). Thus our results suggest that AHP results may be fairly sensitive to imprecision in their inputs.

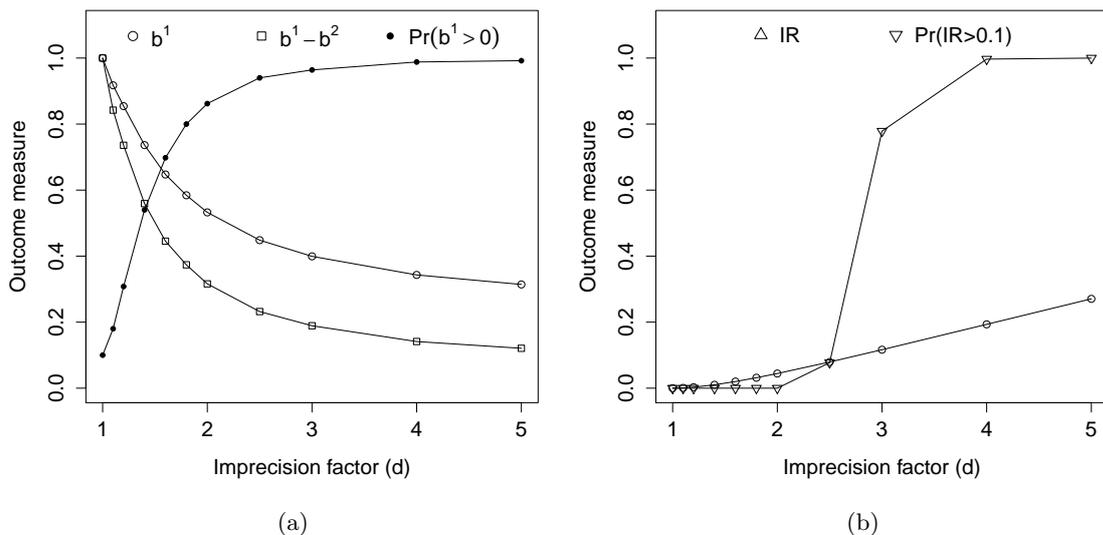


Figure 3: Main simulation results showing (a) mean acceptability outcomes and (b) consistency outcomes obtained from the SMAA-AHP model (for fixed number of alternatives $I = 10$ and criteria $K = 8$, and with both types of imprecision present).

On the other hand the consistency of judgements, as measured by inconsistency ratios, show a strong robustness to imprecision in the judgemental inputs. Imprecision must be extreme ($d \geq 2.8$) before the mean inconsistency ratio exceeds 0.10. The inconsistency ratio increases near-linearly in the imprecision factor d . Clearly, if a rejection policy was to be employed to remove inconsistent judgements from SMAA-AHP, then the proportion of rejected simulated matrices rises sharply around this level of imprecision. This would

appear to be a fairly unlikely scenario however, effectively allowing an initial judgement of 3 to lie anywhere between 1 and 9.

Figure 4 shows how the effect of imprecision type on acceptability and consistency, again holding the problem size fixed at $I = 10$ and $K = 8$. Average consistency results remain the same because these means are simply formed from those pairwise comparison matrices that are affected by imprecision i.e. $K + 1$ if both imprecision types are present, K if only alternative performances are imprecise, and 1 if only criterion importance is imprecise. As imprecision is increased the acceptability of the best-ranked alternative initially decreases much more slowly when only criterion importance is imprecise than when alternative performance is imprecise. Once the imprecision factor is large, however (i.e. $d \geq 3$), the rate of further decrease is very similar in both conditions. The net effect is that when only criterion importance is uncertain, the best-ranked alternative retains a high acceptability regardless of the level of imprecision, and fewer alternatives are identified as potentially optimal. Nevertheless with substantial acceptability indices for at least two alternatives it would still be necessary to resolve the uncertainty about criterion importance before a final choice could be justified.

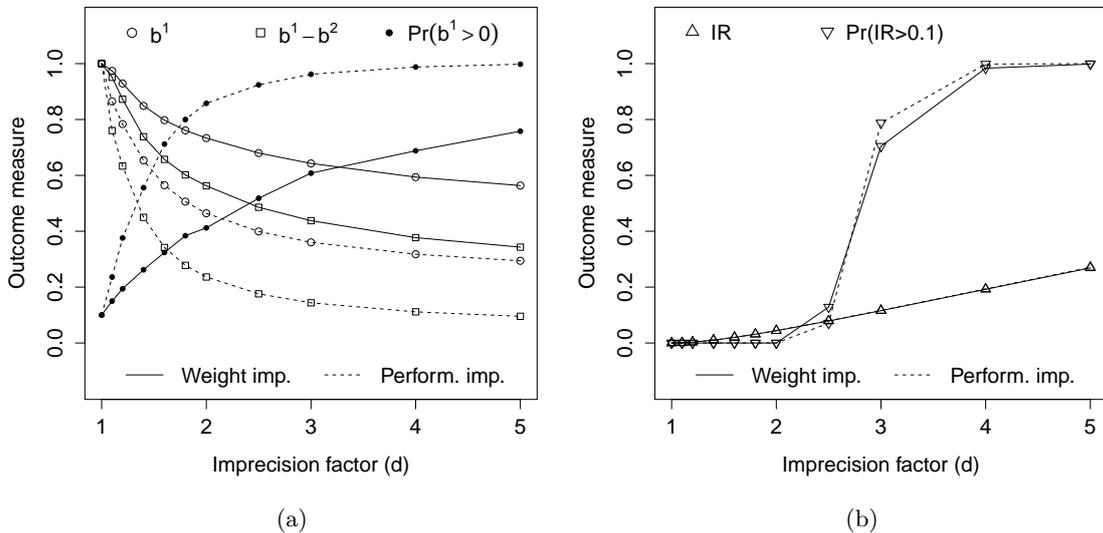


Figure 4: Effect of imprecision type on (a) mean acceptability outcomes and (b) consistency outcomes.

Finally we examine the effect of problem size on our acceptability and consistency results, now returning to the case where both types of imprecision are present. Figure 5 displays these results, which show that the relationship between mean consistency and

imprecision is remarkably consistent over the generated range of problem sizes. The proportion of inconsistent judgements experiences a sharp increase between $d = 2$ and $d = 3$, with the steepness of the transition depending on the number of pairwise comparisons made (note that the number of pairwise judgements increases linearly in the number of criteria but exponentially in the number of alternatives). Acceptability results are more complex. The relationship between the top-ranked acceptability indices and imprecision is also fairly consistent over problem sizes (see results for b^1 and $b^1 - b^2$), but the rate at which potentially optimal alternatives are identified (as d increases initially) is much greater when fewer pairwise comparisons are made (i.e. $I = 5, K = 4$) than when more are made ($I = 20, K = 16$). The reason for this is not immediately apparent.

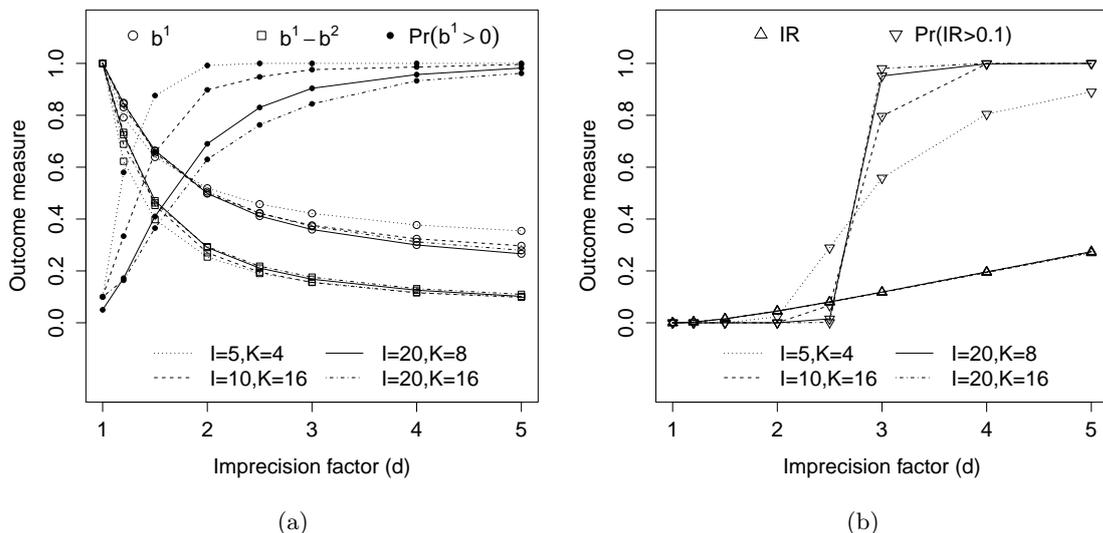


Figure 5: Effect of problem size on (a) mean acceptability outcomes and (b) consistency outcomes.

6 Conclusions

In this paper we have introduced the SMAA-AHP method for representing uncertain or imprecise information through stochastic distributions in the AHP and a simulation approach for analysing the resulting model. The method is suitable for group decision-making problems, where it is difficult to agree on precise pairwise comparisons. A particular strength of the method is that it allows flexible modelling of different kinds of imprecision, uncertainty, or missing preference information. This may be useful in many decision problems

where information is gradually refined during the process. The multiplicative uncertainty factor of SMAA-AHP is a natural way to express the uncertainty symmetrically around a midpoint, and has (in comparison to the uniform distribution, see for example Hauser and Tadikamalla (1996)) the advantage of allocating equal probability mass for equal proportional sub-intervals. However, the SMAA approach is not constrained by this choice, and any realistic distribution may be used (see Tervonen et al. (2013) for details).

One concern that users of the AHP may raise about the use of a simulation-based method such as SMAA is the potential for generating inconsistent judgements. Inconsistent judgements may of course be screened out using rejection sampling, but our simulation results indicate that they are unlikely to arise at all (given precise point judgements i.e. entirely due to the simulation process) unless imprecision is fairly severe. The simulation results also show that with even quite limited imprecision there tend to be several potentially optimal alternatives, with the difference between the largest and second-largest acceptability decreasing rapidly with imprecision. This means that when imprecision is present in almost any degree the choice of best alternative will not be clear. In some instances imprecision may be resolvable by further discussion with the DM; in others it may be an unavoidable element of the decision problem (for example, when it results from uncertainty about future events). In either case, the SMAA methodology provides a set of output measures that can be presented to DMs, so that they can better understand the kinds of preferences that support the selection of each of the alternatives. Even where progression to a final choice is not immediately possible this information can help DMs to gain greater insight into the choices facing them.

References

- C.A. Bana e Costa and J-C. Vansnick. A critical analysis of the eigenvalue method used to derive priorities in ahp. *European Journal of Operational Research*, 187(3):1422–1428, 2008.
- R. Banuelas and J. Antony. Application of stochastic analytic hierarchy process within a domestic appliance manufacturer. *Journal of the Operational Research Society*, 58(1): 29, 2007.
- I. Basak. Estimation of the multi-criteria worths of the alternatives in a hierarchical

- structure of comparisons. *Communications in Statistics-Theory and Methods*, 18(10): 3719–3738, 1989.
- I. Basak. Inference in pairwise comparison experiments based on ratio scales. *Journal of Mathematical Psychology*, 35(1):80–91, 1991.
- I. Basak. Probabilistic judgments specified partially in the analytic hierarchy process. *European Journal of Operational Research*, 108(1):153–164, 1998.
- V. Belton and T.J. Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers, Boston, 2002.
- C.G.E. Boender, J.G. de Graan, and F.A. Lootsma. Multicriteria decision analysis with fuzzy pairwise comparisons. *Fuzzy Sets and Systems*, 29(2):133–143, 1989.
- J. Buckley. Fuzzy hierarchical analysis. *Fuzzy Sets and Systems*, 17(3):233–247, 1985.
- I.N. Durbach and T.J. Stewart. Modelling uncertainty in multi-criteria decision analysis. *European Journal of Operational Research*, 223(1):1–14, 2012.
- D. Hauser and P. Tadikamalla. The analytic hierarchy process in an uncertain environment: a simulation approach. *European Journal of Operational Research*, 91(1):27–37, 1996.
- R Lahdelma and P Salminen. SMAA-2: stochastic multi-criteria acceptability analysis for group decision making. *Operations Research*, 49(3):444–454, 2001.
- R Lahdelma, J Hokkanen, and P Salminen. SMAA – Stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, 106:137–143, 1998.
- R.R. Levary and K. Wan. A simulation approach for handling uncertainty in the analytic hierarchy process. *European Journal of Operational Research*, 106(1):116–122, 1998.
- R.R. Levary and K. Wan. An analytic hierarchy process based simulation model for entry mode decision regarding foreign direct investment. *Omega*, 27(6):661–677, 1999.
- T.L. Saaty. *Multicriteria decision making: The Analytic Hierarchy Process*. RWS Publications, Pittsburgh, 1990.
- T.L. Saaty and L.G. Vargas. Uncertainty and rank order in the analytic hierarchy process. *European Journal of Operational Research*, 32(1):107–117, 1987.

- A.A. Salo and R.P. Hämäläinen. Preference programming through approximate ratio comparisons. *European Journal of Operational Research*, 82(3):458–475, 1995.
- T. Tervonen and R. Lahdelma. Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research*, 178(3):500–513, 2007.
- T. Tervonen, H. Hakonen, and R. Lahdelma. Elevator planning with stochastic multicriteria acceptability analysis. *Omega*, 36(3):352–362, 2008.
- T. Tervonen, G. van Valkenhoef, N. Baştürk, and D. Postmus. Hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. *European Journal of Operational Research*, 224(3):552–559, 2013.
- O.S. Vaidya and S. Kumar. Analytic hierarchy process: An overview of applications. *European Journal of Operational Research*, 169(1):1–29, 2006.
- PJM Van Laarhoven and W. Pedrycz. A fuzzy extension of Saaty’s priority theory. *Fuzzy Sets and Systems*, 11:199–227, 1983.
- L.G. Vargas. Reciprocal matrices with random coefficients. *Mathematical Modelling*, 3(1):69–81, 1982.