Scenario planning in the analytic hierarchy process

Ian Durbach[*][†]

February 26, 2019

Abstract

Multi-criteria decision analysis and scenario planning are complementary tools for supporting large-scale, strategic decision making, but there has been limited interaction between the fields. This paper describes how scenario planning can be integrated with the analytic hierarcy process (AHP), one of the most popular approaches in decision analysis and one that is arguably more accessible for new users. Scenario planners looking for an avenue into MCDA may find the AHP a useful introduction, and AHP practitioners may find in scenario planning a tool with which to address problems with large, structural or "deep" uncertainties. A common understanding of scenarios as plausible futures, rather than states of nature, is emphasised, as is how scenarios can be viewed as a kind of "meta-attribute" over which possible courses of action can be compared. A simulation experiment assesses the potential effects of ignoring scenario-specific information, as well as of different ways of constructing scenarios.

Keywords: scenario planning; decision analysis; multicriteria; analytic hierarchy process.

Word count: 9333 (including abstract and references)

1 Introduction

Multi-criteria decision analysis (MCDA) is concerned with the evaluation of a set of possible courses of action or alternatives, for the purpose of making the different types of decisions such as choosing a preferred alternative, or a subset of alternatives, or of creating new alternatives (e.g. Belton and Stewart, 2002). At the heart of MCDA is the view that better decisions are made when the overall evaluation of alternatives is decomposed into evaluations on a number of usually conflicting criteria relevant to the problem, and by explicitly recognising and quantifying trade-offs and value judgments rather than leaving these as unarticulated assumptions. Many multi-criteria methods exist; they differ primarily according to how they evaluate performances on each attribute, and aggregate evaluations across attributes to arrive at an overall or global evaluation. MCDA can be

[*]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

thought of as embodying the philosophy of evidence-based practices, but applied to the decision-making process itself.

MCDA and scenario planning (e.g. Van der Heijden, 1996) have much in common. Both are concerned with large-scale, strategic decisions of the kind that often face uncertainties that are complex and interrelated, and for which precise mathematical measures such as probabilities may be difficult to assess and operationalize in a meaningful way. Both recognize the importance of problem framing, and that this is a difficult process easily rendered useless if done carelessly. Both are usually conducted in interactive workshops under the guidance of a trained facilitator who has no direct interest in the decision.

Given the overlap in interests, there has perhaps been less interaction between the fields than might be expected. Probability-based models constitute the majority of approaches dealing with uncertainty in MCDA, and although attempts were made fairly early on to co-opt scenario planning ideas into decision analysis (Goodwin and Wright, 1998), and the general integration of scenario planning and MCDA has been described in a number of publications (e.g. Stewart, 2005; Durbach and Stewart, 2012b; Stewart et al., 2013) impactful applications remain thin on the ground (although see Marttunen et al. (2017) and Montibeller et al. (2006), for example). Unfortunately this work has not found its way back into the mainstream scenario planning literature, and the review by Amer et al. (2013) includes a section on quantitative scenario planning methods but does not mention any MCDA-based methods.

There are many possible reasons for the lack of greater integration of scenario planning and MCDA, but two that come to mind are that MCDA may place an emphasis on quantification that does not sit well with scenario planners, and that one of the most flexible and popular MCDA approaches, the analytic hierarchy process or AHP (Saaty, 1980) has to a certain extent been left out of developments that brought scenario planning ideas into MCDA (although see Levary and Wan (1999)).

In this paper I outline how scenario planning can be integrated with the AHP, and by doing so hope to enlarge the size of the intersection of MCDA and scenario planning; to persuade some scenario planners to try MCDA and some MCDA practitioners to incorporate scenario planning methods. I do this in two main ways: by providing a comprehensive description of how, practically, one can implement a scenario-based approach to AHP; and by reporting the results of a simulation experiment that, within the limited scope of a simulation experiment, show that reducing a large set of complex uncertainties to a small number of plausible futures may be a fruitful approach for multi-criteria decision support employing the AHP.

The remainder of the paper is organized as follows. Following some preliminaries on notation, Sections 3 and 4 reviews uncertainty modelling in the AHP and scenario-based approaches in MCDA respectively. Section 5 is the first "main" section of the paper, describing the scenario-based AHP. Sections 6 and 7 describe the design and results of the simulation study, and a final section concludes.

2 Notation

We consider a decision problem consisting of I alternatives $i \in \{1, \ldots, I\}$, each evaluated on J criteria $j \in \{1, \ldots, J\}$ and (possibly) also on K scenarios $k \in \{1, \ldots, K\}$. Let z_{ij} be the evaluation of alternative i in terms of criterion j, according to some suitable performance measure. In the standard AHP the decision maker (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. We often need to refer to comparison between two alternatives, or two attributes, or two scenarios; we denote an arbitrary pair of alternatives by i and i^*, attributes by j and j^*, scenarios by k and k^*. Thus, the pairwise preference $a_{ii^*}^{(j)}$ expresses the DM's preference for alternative i over alternative i^* on criterion j, and this represents the ratio between evaluations z_{ij}/z_{i^*j}, expressed on a discrete scale from 1 to 9 (where 1 means equal preference and 9 denotes absolute preference). These pairwise preferences are collected into the $I \times I$ pairwise comparison matrix (PCM) $\mathbf{A}_0^{(j)} = [a_{ii^*}^{(j)}]$ (the zero subscript is to differentiate this PCM from later scenario-based PCMs). Similarly, the pairwise preference b_{jj^*} expresses the DM's importance assessment for attribute j over j^*, representing the ratio between attribute weights w_j/w_{j^*}, and collected in the $J \times J$ PCM $\mathbf{B}_0 = [b_{jj^*}]$.

Our concern is with decision making situations in which the pairwise evaluations $a_{ii^*}^{(j)}$ and b_{jj^*} (and consequently computed values for z_{ij} and w_j respectively) are uncertain and modelled using scenarios. We let z_{ijk} denote the performance of alternative i on attribute j in scenario k, and w_{jk} is the joint importance of attribute j and scenario k i.e. $\sum_{j,k} w_{jk} = 1$. This weight can be decomposed as $w_{jk} = w_{j|k}\omega_k$, where $w_{j|k}$ is the importance weight for attribute j conditional on scenario k i.e. $\sum_j w_{j|k} = 1$, $\forall k$, and ω_k denotes the importance weight for scenario k. When attribute importance weights do not vary across scenarios this allows for the decomposition $w_{jk} = w_j \omega_k$, which simplifies assessment.

Much of the remainder of the paper is taken up with how different PCMs can be used to estimate z_{ijk} and w_{jk}. Notation becomes more complicated, essentially because the analyst faces a choice between bundling scenarios with alternatives or with attributes. All possibilities are listed here; not all combinations will be feasible, but discussion of this is delayed until Section 5.

Pairwise preferences can be expressed between *combinations* of alternatives and scenarios e.g. how much does the DM prefer alternative i in scenario k to alternative i^* in scenario k^*, and these preferences are denoted by $\mathbf{A}^{(j)} = [a^{(j)}_{(i,k)(i^*,k^*)}]$. Note that the PCM $\mathbf{A}^{(j)}$ has dimension $IK \times IK$. Alternatively, pairwise preferences can be expressed between alternatives *conditional on* a particular scenario, $\mathbf{A}^{(j,k)} = [a^{(j,k)}_{ii^*}]$, in which case each $\mathbf{A}^{(j,k)}$ has dimension $I \times I$.

Similarly, when assessing the importance of attributes, the DM can either assess how much more important is the *combination* of attribute j and scenario k than attribute j^* and scenario k^*, denoted by the $JK \times JK$ PCM $\mathbf{B} = [b_{(j,k)(j^*,k^*)}]$, or they can compare the importance of attributes j and j^* *conditional on* scenario k, $\mathbf{B}^{(k)} = [b^{(k)}_{jj^*}]$, in which case each $\mathbf{B}^{(k)}$ has dimension $J \times J$.

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 PCM is extracted (the so-called priority vector), and these values are scaled to sum to 100 (or one). Whichever method is employed, the estimate of the attribute evaluation z_{ijk} is denoted by $\hat{z}_{ijk}$, the estimate of the attribute importance weight w_{jk} by $\hat{w}_{jk}$, and the estimate of the scenario importance weight ω_k by $\hat{\omega}_k$.

3 Uncertainty modelling in the AHP

The analytic hierarchy process (Saaty, 1980) is a widely-used method for multicriteria decision support founded upon using a nine-point semantic scale to express pairwise preferences for one alternative over another on a particular criterion, and for how much one criterion is valued over another. These can then be aggregated into an overall evaluation of the alternatives using an additive aggregation e.g. $\sum_j \hat{w}_j \hat{z}_{ij}$

Aspects of the AHP are controversial. The AHP uses additive aggregation to evaluate alternatives, which implies a very specific interpretation of criterion weights – they make differences in performance comparable across criteria, in the sense that if alternative a outperforms another alternative b by two units on criterion j, and b outperforms a by 5 units on criterion j^*, but overall they are equally preferred, then $w_j/w_{j^*} = 5/2$. Because the AHP scales evaluations on each attribute to sum to 100, the differences that weights compare are differences in *total* (or average) scores, and the addition or removal of alternatives changes the meaning of a one-unit change in the associated attribute. Weights are much harder to conceptualise in this way, and in practice they are often assessed without reference to attribute ranges, but this raises questions of consistency and can also lead to problems of rank reversal when alternatives are included or excluded. This and other criticisms are covered in more detail in, for example, Belton and Stewart (2002). Nevertheless, the AHP has found widespread application and acceptance in practice (e.g. Vaidya and Kumar, 2006), largely due to its ease of use relative to competing schools of MCDA.

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). The vast majority of AHP models that treat uncertainty in the attribute evaluations do so using some form of Monte Carlo simulation to randomly generate pairwise evaluations from the distributions specified by decision makers (Hauser and Tadikamalla, 1996; Levary and Wan, 1998, 1999; Basak, 1998; Banuelas and Antony, 2007), although some notable early research attempted a more analytical approach pairing distributional forms of pairwise judgements and priority vectors (Vargas, 1982; Saaty and Vargas, 1987; Basak, 1989, 1991). The simulation-based approaches all follow the same basic approach, first expressed by Hauser and Tadikamalla (1996). The decision maker expresses pairwise comparisons on the usual 1-9 scale, except that these comparisons are now random variables with associated probability distributions. Hauser and Tadikamalla generated random judgements uniformly on the interval $[a_{ii^*} - da_{ii*}, a_{ii^*} - da_{ii^*}]$, with d an uncertainty factor, before transforming any values less than one. PCMs are generated from these distributions, from which priority vectors are obtained, and alternatives ranked according to some measure of centrality of the priorities obtained (usually the mean).

Notably for the current paper, the AHP was one of the first MCDA methods to incorporate scenarios into their model. Levary and Wan (1998) and Levary and Wan (1999) asked decision makers to assess PCMs for each scenario, before aggregating over scenarios using probabilities. As within-scenario judgments were allowed to be stochastic, and the scenarios themselves were interpreted probabilistically (essentially, as states of nature), a simulation approach was used to randomly generate pairwise judgements within each scenario. In this paper I advocate a somewhat different approach, but one which still owes much to that of Levary and Wan.

Other simulation-based AHP models have essentially added minor extensions to accommodate specific challenges. For example, 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. Durbach (2014) integrate stochastic multicriteria acceptability analysis with the AHP, which presents additional information to the DM, and also advise drawing random values from a different distribution to Hauser and Tadikamalla (1996).

4 Scenario planning in multi-criteria decision analysis

Scenario-based MCDA approaches have been proposed in all the major schools of MCDA. Value function approaches have been proposed in Montibeller et al. (2006); Ram et al. (2010); van der Pas et al. (2010), outranking in Durbach (2014), AHP in Levary and Wan (1999), goal programming in Durbach and Stewart (2003). The general integration of scenario planning and MCDA has been thoroughly described in a number of places (e.g. Goodwin and Wright, 2009; Stewart, 2005; Stewart et al., 2013; Durbach and Stewart, 2012b; Stewart and Durbach, 2016), and Marttunen et al. (2017) reviews the use of scenario planning as a problem structuring tool, and summarizes features of a number of applications of scenario-based MCDA. Although the emphasis of the current paper is how multicriteria assessments can be combined across scenarios, scenario planning has also been used in a more qualitative way to support problem structuring for MCDA, for example to think about external uncertainties (Marttunen et al., 2017).

Reasons for the integration of scenarios in MCDA are easy to imagine – scenario planning is a mature field with a comprehensive literature and a well-established set of best practices (e.g. Schoemaker, 1991; Wright and Goodwin, 1999; Van der Heijden, 1996;

Schoemaker, 1995), particularly around practical issues like scenario construction, and it is often used to tackle the kinds of large-scale, strategic decisions that multicriteria decision models also aim to address, and to which probability-based approaches are arguably less suited. However, scenario planning has traditionally preferred to avoid formal quantitative modeling when using scenarios to arrive at a final decision about which course of action to take, and to use informed but informal judgment instead (e.g. Cairns et al., 2004). This stance is in obvious conflict with MCDA goals like transparency and explicitly recognising and quantifying value judgments, leaving an equally obvious opportunity for MCDA to plug what it perceives as a gap.

The general scenario-based approach identifies a small number of scenarios, and assesses attribute evaluations in each scenario. Assessments within each scenario are almost always deterministic. There is no theoretical reason why assessments could not also be allowed to be uncertain – the performance of alternative i on attribute j in scenario k can be represented, for example, by a probability density function. Practically, scenarios and scenario planning is used as a means of reducing complexity and managing strategic uncertainties that may be difficult to express probabilistically. Admitting both scenarios *and* allowing for further higher-order uncertainties within scenarios may be too onerous a task to be practically viable, and might negatively affect ease of use and transparency.

Any number of MCDA approaches can be used to evaluate and compare the performances of alternatives in each scenario. Indeed, as the general approach assesses deterministic evalutions in each scenario, literally any MCDA model could be applied. The real question then, is whether and how to aggregate performance over different scenarios to arrive at a global evaluation of alternatives. Broadly, responses to this question can be categorised into three groups

Do not aggregate at all The original scenario planning literature placed a strong emphasis on developing strategies that were robust across scenarios, and intentionally avoided aggregation over scenarios (e.g. Van der Heijden, 1996). Some decision analysts have agreed with this assessment, and so have emphasized the interpretation of results within scenarios, preferring not to aggregate evaluations over scenarios to arrive at a final global evaluation (e.g. Goodwin and Wright, 2001; Ram et al., 2010). There is a good deal of value in this interpretation, and there is an argument to be made that adhering to the philosophy of scenario planning would increase the likelihood of using scenario-based MCDA in organisations already using scenario planning. It is also a popular approach in practice, with a recent survey finding that a majority of scenario-based MCDA applications explored how rankings changed between scenarios rather than aggregating over them (Marttunen et al., 2017). However, many decision problems demand that *some* selection between alternatives is made, and in these cases the question is which approach will generate the most goal-directed solutions, unaided intuitive aggregation across scenarios, or an aggregation facilitated by MCDA. Many if not most researchers, myself included, fall into the second camp.

Treat scenarios as states of nature and use probabilities In this approach a relative likelihood is assessed for each scenario, indicating which scenarios are more or less likely to occur. This is a traditional probabilistic interpretation of uncertainty in which each attribute evaluation is a random variable and the unfolding of the future is the 'experiment' whose outcome (a scenario) determines the realization of the random variable. In this interpretation a scenario is formally equivalent to a "state of nature".

The probabilisitic approach has probably been the most popular way to aggregate across scenarios (e.g. Levary and Wan, 1999; Vilkkumaa et al., 2018), but has also been criticized Stewart et al. (2013). On one hand, these criticisms can appear quite techical and esoteric – the set of scenarios does not constitute a complete probability space, and so the scenario "likelihoods" are not probabilities; the scenarios also do not represent the same dimensions in probability space, and thus are not likelihoods (in a statistical sense) either. On the other hand, it is clear enough that a scenario is not equivalent to a state of nature, and that it is the subjective narrative component of scenarios that is responsible for the difference. States of nature are mutually exclusive and exhaustive; moreover, they are constructed from the same underlying dimensions – for example one could always ask questions like "does event x occur in state of nature k?" Scenarios are incomplete descriptions, dimensions of these descriptions differ, and questions like the above cannot be answered. One scenario may sketch out a political development without indicating the economic consequences; another scenario may say very little about politics but focus on economic events instead. When asked to compare the likelihoods of these two scenarios the DM is essentially being asked to "fill in the details" required to place these scenarios on the same multidimensional probability space, and then to integrate over the unspecified dimensions in each scenario. Nobody would claim that this happens, even approximately – but it is not clear what *does* happen when DMs are asked to assess likelihoods, and this is a problem. A related problem is that scenario "likelihoods" depend on the precision with which the scenarios are described – the likelihood of a scenario specifying only that an election takes place should be greater than that of a scenario specifying that the election is also peaceful. Cognitive biases such as the "Linda effect" (Tversky and Kahneman, 1983) suggest that DMs may not answer these questions in the way expected of them.

Treat scenarios as dimensions over which performance can be compared This approach maintains that aggregation should use relative (swing) weights on performance in different scenarios. In this interpretation scenarios are treated in the same way as criteria. In a value function context weights can be assessed by asking questions like "would you prefer to go from the worst to best performance on scenario a or scenario b?" This approach has the benefit of a theoretically sound, clear interpretation, and DMs could give a scenario more weight if they feel that improving performance is more important to them in a particular scenario, a judgement which may take likelihood considerations into account.

A drawback of the weight-based aggregation is that suffers from the practical problem that performances (say best or worst performances in scenario a or b) are already aggregated evaluations over criteria, on a difficult-to-interpret scale ("value"), and so may well be difficult for DMs to compare. Another drawback is that the conceptualization of a scenario as a "dimension" (effectively, playing the same role as a criterion) is potentially not nuanced enough to capture how scenario planners conceptualize scenarios, or may not place enough emphasis on the scenarios – which may explain the preference of scenario planners for not aggregating at all. Even if differences in performance between scenario are mathematically well-defined (in a way that probabilities are not) scenario planners may still believe it inappropriate to do so, or that the task is too cognitively challenging to do consistently.

5 Scenario planning in the AHP

A scenario-based AHP model is easily established and implemented, following the general principles used in scenario-based MCDA. There are two broad approaches that could be followed. The key difference between these two approaches are the way in which scenarios are treated. The first considers combinations of alternatives and scenarios as IK distinct outcomes or "meta-alternatives" (Durbach and Stewart, 2012b) to be evaluated in terms of the J attributes. Scenarios are thus combined with alternatives, and the joint "meta-alternatives" evaluated over attributes. The second procedure, which is arguably a more natural interpretation of scenarios (Stewart et al., 2013; Durbach, 2014) is to consider combinations of attributes and scenarios as JK distinct "meta- attributes", and evaluate the I alternatives in terms of each of these meta-attributes. It should be noted that these approaches both use a standard implementation of the AHP, and thus are subject to the same concerns raised in Section 3 e.g. regarding rank reversal and interpretability of weights. There are ways in which these concerns can be alleviated, for example by normalizing the attibute scores differently, but as these have not been taken up by the wider AHP community they are not pursued here either.

5.1 The "meta-alternative" approach: combining scenarios with alternatives

This approach requires:

1. For each of J attributes, the assessment of an $IK \times IK$ PCM $\mathbf{A}^{(j)} = [a^{(j)}_{(i,k)(i^*,k^*)}]$ representing preferences between alternative-scenario pairs, and from which a leading eigenvector of length IK can be extracted, these representing the performance of alternative-scenario pair (i,k) on criterion j i.e. $\hat{z}_{ijk}$. There are two lines of questioning that can be used to build this PCM:

   Joint assessment Vary both the alternative i and scenario k, for example comparing (i,k) and (i^*,k^*), where $i \neq i^*$ and $k \neq k^*$. A typical assessment question might be:

   > "Thinking about criterion j, by how much do you prefer alternative i in scenario k to alternative i^* in scenario k^*, on a 1-9 scale?"

   This approach requires a minimum of $IK - 1$ assessment questions, many of which would be cognitively difficult (those in which both alternatives and scenarios differ). Note that in this line of questioning scenario importance is built into the responses by construction, because the DM must compare performances across different scenarios.

Conditional assessment First, fix each scenario k one at a time, and compare the performance of pairs of alternatives e.g. compare (i,k) and (i^*,k), using a similar question to that above. For each scenario k, this gives an $I \times I$ PCM $\mathbf{A}^{(j,k)} = [a_{ii^*}^{(j,k)}]$. The elements of eigenvectors extracted from these PCMs provide *relative* information on which alternative perform better than others in each of the scenarios. What they lack is information on performance levels *between* scenarios – for example, performance in some scenarios may be generally lower than in others, and this would not be apparent from the $\mathbf{A}^{(j,k)}$'s. This information can be assessed by fixing an alternative i and comparing the performance of pairs of scenarios e.g. comparing (i,k) and (i,k^*), again using a similar question to that above –

> "Thinking about criterion j, by how much do you prefer alternative i in scenario k to alternative i in scenario k^*, on a 1-9 scale?"

These assessments compare the performance of the same alternative across different scenarios, and thus fill in a small portion of the PCM $\mathbf{A}^{(j)} = [a_{(i,k)(i^*,k^*)}^{(j)}]$. Note that these assessments also capture the importance of performance in different scenarios. For example, an alternative i may return exactly the same value (say a profit of \$1000) in two scenarios but there is nothing forcing the DM to be indifferent between receiving this amount in either scenario – statements like "I prefer to receive \$1000 in scenario 1 than \$1000 in scenario 2" are perfectly permissable. Since relative performance levels between alternatives are already known within each scenario (from the previously assessed PCMs $\mathbf{A}^{(j,k)}$) this is sufficient to infer the performance of *any* alternative across different scenarios, and thus to fill in the remainder of the PCM $\mathbf{A}^{(j)}$, although other comparisons may be used as a consistency check. The conditional assessment approach requires the same number of responses as joint assessment – a minimum of $K(I-1)$ responses in the first step, and at least $K-1$ in the second step – but is (arguably) cognitively easier.

2. The assessment of a $J \times J$ PCM $\mathbf{B}_0 = [b_{jj^*}]$, whose entries compare the importance of attribute j and j^*. Note that because scenarios are bundled with alternatives in the meta-alternative approach, attribute importance must be assessed independently of scenarios, using the same assessment questions as for deterministic AHP. Elements

from the leading eigenvector of $\mathbf{B}_0$ give $\hat{w}_j$, estimates of the importance weight for attribute j.

These evaluations may then be aggregated over criteria in the usual way i.e. $V_{ik} = \sum_{j=1}^{J} \hat{w}_j \hat{z}_{ijk}$, giving the performance of alternative i in scenario k. As scenario importance (which can include considerations of how likely a scenario is to occur) has already been incorporated into the assessments, aggregation over scenarios is straightforward – one just sums up each alternative's performance over scenarios. Aggregation *can* take the form of a weighted average, with scenario "likelihoods" used as weights, though the caveats raised in the previous section about the difficulty of defining and assessing these apply.

5.2 The "meta-attribute" approach: combining scenarios with attributes

This procedure evaluates the I alternatives in terms of each of KJ scenarios-attribute combinations or "meta- attributes". It requires:

1. For each of KJ scenario-attribute combinations, the assessment of an $I \times I$ PCM $\mathbf{A}^{(j,k)} = [a_{ii^*}^{(j,k)}]$, from which a leading eigenvector of length I can be extracted, these representing the performance of alternative i on the scenario-attribute pair (k, j) i.e. $\hat{z}_{ijk}$. A typical assessment question might be:

   "Imagine you are in scenario k. Thinking of criterion j, by how much do you prefer alternative i to alternative i^*, on a 1-9 scale?"

   Questions of this type assess the usual AHP preferences for one alternative over another, but conditional on both attribute *and* scenario.

2. The assessment of a $KJ \times KJ$ PCM $\mathbf{B} = [b_{(j,k)(j^*,k^*)}]$, from which a leading eigenvector of length KJ can be extracted, these representing the importance of the scenario-attribute pair (k, j) i.e. $\hat{w}_{jk}$. In a similar fashion to the assessment of alternative-scenario combinations in the first procedure, pairwise comparisons of attribute-scenario combinations can be constructed in two ways:

   Joint assessment Vary both the attribute j and scenario k. A typical assessment question might be:

   "How important is achieving good performance on attribute j in scenario k compared to achieving good performance on attribute j^* in scenario k^*, on a 1-9 scale?"

This approach requires a minimum of $KJ-1$ assessment questions, those in which both attributes and scenarios differ being cognitively difficult. Note that, as for the joint assessment of the "meta-alternative" approach of the previous section, scenario importance is built into the responses by construction.

Conditional assessment First, fix each scenario k one at a time, and vary attributes e.g. compare (j,k) and (j^*,k), using a similar question to that above. For each scenario k, this gives an $J \times J$ PCM $\mathbf{B}^{(k)} = [b^{(k)}_{j,j^*}]$. Note that if attribute importance does not vary over scenarios then $\mathbf{B}^{(k)} = [b_{j,j^*}] = \mathbf{B}_0$. Next, assess the relative importance of scenarios under the assumption of a common attribute j. Again, since relative importance levels between attributes are already known within each scenario, comparing scenario importance for a specific attribute is sufficient to infer between-scenario importance for any attribute i.e. to complete the PCM $\mathbf{B} = [b_{(j,k)(j^*,k^*)}]$. Also, the conditional approach requires at least $KJ-1$ questions, the same as the minimum number of joint assessment questions, but these are probably cognitively easier for the DM to answer.

Both joint and conditional assessment methods thus return estimates of the joint importance weights $\hat{w}_{jk}$. In the former case these are directly assessed; in the latter they are assessed via the decomposition $\hat{w}_{jk} = \hat{w}_{j|k}\hat{\omega}_k$ which may, if attribute importance is constant over scenarios, simplify further to $\hat{w}_j\hat{\omega}_k$.

The aggegate performance of each alternative in scenario k can then be easily obtained as $V_{ik} = \sum_{j=1}^{J} \hat{w}_{jk}\hat{z}_{ijk}$, from which a rank order of alternatives in that scenario is obtained. The global performance of each alternative is given by

$$V_i = \sum_{k=1}^{K}\sum_{j=1}^{J} \hat{w}_{jk}\hat{z}_{ijk} = \sum_{k=1}^{K}\hat{\omega}_k \sum_{j=1}^{J} \hat{w}_{j|k}\hat{z}_{ijk}$$

from which a global rank order of alternatives is obtained.

6 Simulation study

In this section a simulation experiment is used to assess how different the outcome of a scenario-based AHP approach may be to an approach that takes into account all possible states of nature, in this case MAUT. The use of MAUT as a point of reference for scenario-based approaches must be accompanied with a note of caution. Scenario planning is not intended to be an "approximation" to MAUT, and constructing scenarios may well be more demanding than eliciting the probability distributions used as inputs to MAUT. In this simulation a very large number of states of nature are used, so that the use of MAUT would be practically impossible. Nevertheless it is still useful to ask how closely the results of the two approaches – scenario-based or not – might be. I refer to MAUT as an "idealised" model, but only in the sense that it uses all available states and that this is not practically feasible.

Moreover, using a simulation experiment to investigate a partially qualitative approach like scenario planning has obvious limitations. Scenario plan can benefit MCDA in several ways that are simply not amenable to simulation – encouraging thinking about unforeseen events, increasing stakeholder participation, generating new solutions, for example (Marttunen et al., 2017). Whether the specific implementation of scenario-based AHP advocated here realizes these benefits or not is left to future behavioural research to answer. Nevertheless, there are some aspects of scenario construction that can be simulated – for example the number of scenarios used, whether likelihoods are employed, whether preferences differ over scenarios – and that give useful insights into the potential performance of scenario-based approaches in MCDA.

The structure of the current simulation study closely follows that used and reported in previous studies (Durbach and Stewart, 2009, 2012a; Durbach, 2014; Durbach et al., 2014), and parameter values are largely informed by a simulation study of scenario-based outranking methods (Durbach, 2014), although some new elements, particularly between-state correlations, are added here. The structure of each simulation run is:

1. Select problem size parameters (number of alternatives, criteria, and states), and generate attribute evaluations and attribute importance weights, which are used to generate pairwise comparison matrices, these constituting the input data to the AHP.

2. Simulate the application of an AHP model that has access to all possible attribute evaluation and attribute importance weight information, referred to as an "idealised" model.

3. Simulate the application of various models that access only a subset of possible information, referred to as "non-idealised" models. These include a number of models that use scenario-based approaches.

4. Evaluate how closely the rank orders returned by non-idealised models approximate those returned by an idealised model.

By performing a number of simulation runs using various combinations of parameters, aggregate statistics can be collected and the mean performance of scenario-based models assessed across a range of simulated conditions. Further details of the simulation structure are given below. All code used to run the simulation study is available at `https://github.com/iandurbach/ahp-scenarios`.

6.1 Generating pairwise comparison matrices

6.1.1 Generating attribute evaluations

The basis of the approach is to say that an alternative's performance (on an attribute) is determined by a two-step process. In the first step one of K possible "futures" is chosen, which determines in some sense the broad direction that the future takes. Each future k consists of M_k realizations or states – these are the actual values that Z_{ij} can take on. These values are drawn from a Gaussian probability distribution $f_{ij}^{(k)}$, with mean μ_{ijk} and variance σ^2_{ijk}. These parameters are used to control, for example, whether realizations are highly variable within each future (so that, even within a particular future, essentially any value is possible), or whether the primary variation in attribute values is between-future variation.

This two-step formulation has been used in a number of previous simulation studies investigating the performance of scenario-based approaches (e.g Durbach and Stewart, 2012a; Durbach, 2014). In this simulation I add the possibility that mean evaluations are correlated across futures – for example, that if an alternative performs well in one future (on average, relative to other alternatives), it tends to perform well in other futures too.

For each attribute j, I realizations are generated from a multivariate uniform distribution with a correlation matrix $\mathbf{\Gamma}^z = [\rho^z_{kk^*}]$, $k, k^* \in \{1, \dots, K\}$. The result, for each attribute j, is an $I \times K$ matrix containing the mean performance of alternative i in future k. These constitute the μ_{ijk}, the means of the probability distributions f^k_{ij}. For simplicity the current experiment uses the same correlation coefficient for all attributes and between all pairs of futures i.e. $\Gamma_{kk^*} = 1$ if $k = k^*$ and ρ^z otherwise. Alternatives are made Pareto optimal within each future by standardizing within each alternative a_i so that $\sum_j \mu_{ijk} = 1$. This standardization changes the correlations between mean evaluations across futures, but preliminary experiments showed that these changes were small (relative to the differences between levels of the simulation parameter ρ).

Within each future k, the simulation then generates unstandardized realizations by

1. Generating a standard deviation σ_{ijk} randomly on Uni$[0.01, \sigma^{(d)}]$, where $\sigma^{(d)}$ is a parameter of the simulation.

2. Setting the number of realizations M_k to be generated for future k. A total of $L = 400$ realizations for each Z_{ij} is used over all futures. These realizations are distributed over futures either "uniformly" (all futures contain 40 realizations) or "non-uniformly" (four futures contain 60 realizations each, four contain 20 realizations, and two contain 40 realizations). The distribution of realizations over futures is denoted by $\mathbf{M}$.

3. Generating M_k independent realizations from $N(\mu_{ijk}, \sigma^2_{ijk})$. The $1 \times M_k$ vector of realizations generated in future k is denoted $\mathbf{z}_{ij}^{(k)}$.

Once realizations have been generated for each future, these are concatenated into a single $1 \times L$ vector containing all the realizations for Z_{ij} i.e. $\mathbf{z}_{ij} = [\mathbf{z}_{ij}^{(1)}, \mathbf{z}_{ij}^{(2)}, \ldots, \mathbf{z}_{ij}^{(k)}, \ldots, \mathbf{z}_{ij}^{(K)}]$. Realizations are standardised so that, within each attribute, the largest realization is set to a random number generated uniformly between 0.75 and 0.9 (for each future), and the smallest realization weight is set to a random number generated uniformly between 0.1 and 0.15. This is done so that when pairwise comparison matrices are constructed the largest ratio of attribute importances lies somewhere between 6 and 9.

6.1.2 Generating attribute importance weights

The current study allows for the possibility that attribute weights may differ over futures – for example, that an attribute may be relatively important in one future but not in another. This is modelled by specifying a correlation structure between futures, in a similar way to what was done for attribute evaluations. Future-specific weights are generated from J realizations drawn from a multivariate uniform distribution with a correlation matrix $\mathbf{\Gamma}^w = [\rho^w_{kk^*}]$, $k, k^* \in \{1, \ldots, K\}$. Note that this means that weights differ between broad futures, but do not vary between different *states* belonging to the same future. The same correlation coefficient is used between all pairs of futures, so $\Gamma^w_{kk^*} = 1$ if $k = k^*$ and ρ^w otherwise. In a similar manner to attribute evaluations, weights are standardised so that, when pairwise comparison matrices are constructed, the largest ratio of attribute importances lies between 6 and 9.

6.1.3 From assessments to pairwise comparisons

Once attribute evaluations and attribute importance weights have been constructed, it is straightforward to compute pairwise comparison matrices from these by simply taking the appropriate ratios. In doing so DMs are assumed to be perfectly consistent in their assessments – something which is not realistic but the relaxation of which is not considered here. The output of this stage is $J \times \sum_k M_k$ attribute evaluation PCMs, one for each combination of attribute and state, and L weight importance PCMs, one for each future.

6.2 Simulating the application of "full uncertainty" AHP models

Given the inputs above, the following two models are simulated; both of these make use of all $J \times \sum_k M_k$ attribute evaluations i.e. all states are used. A short name given to each model and used in tables and figures below is given in bold below, leading into a description of the model.

Idealised uses PCMs in all 400 states, and assumes that the correct i.e. future-specific, weights are used to combine evaluations over states. Thus, for each state in a future k, a leading eigenvector can be extracted from the PCM for each attribute j, and this leading eigenvector can be combined with attribute importance weights appropriate for a state in future k. This gives

$$V_i = \sum_{k=1}^{K} \sum_{\ell=1}^{M_k} \sum_{j=1}^{J} \hat{w}_{jk} \hat{z}_{ijk\ell}$$

Average weights supposes that uncertainty in the attribute evaluations has been fully captured, and so that attribute evaluations from all 400 states are used, but that because a scenario process has not been followed the extra contextual information needed for the DM to realise that his or her weights differ between futures has not been gained. As a result, pairwise comparisons of attribute weights (a) do not differ between futures, and (b) are based on a comparison of attribute importance weights that have been averaged over futures, weighted by the number of states in each future i.e. based not on $\hat{w}_{jk}$ but on $\hat{w}_j = \sum_{k=1}^{K} M_k \hat{w}_{jk}$.

6.3 Simulating the application of scenario-based AHP models

Four scenario-based AHP models are simulated, which differ only in terms of the inputs that they use. These inputs are generated as follows:

1. Select which futures to use. This is done in two ways: a sample of size S is randomly drawn from the set of futures without replacement, either with each future having an equal probability of selection, or with futures selected in order of their likelihood i.e. how many states comprise the future, with ties broken at random.

2. For each future, determine which attribute evaluations to use. This is also done in two ways. A "mean" scenario model uses the mean μ_{ijk} in each of the S selected futures. A "random" scenario model randomly chooses one of the 40 generated states in a scenario, and uses the attribute evaluations in that state the represent the future.

Combining the two ways of selecting futures with the two ways of selecting evaluations within futures gives four possible scenario approaches.

ML-Mean uses the S most likely futures, and mean attribute evaluations to represent alternative performances in each future.

ML-Random uses the S most likely futures, and attribute evaluations from a randomly drawn state in each future to represent alternative performances in that future.

Random-Mean draws S randomly selected futures, without replacement, and mean attribute evaluations to represent alternative performances within each future.

Random-Random draws S randomly selected futures, without replacement, and uses attribute evaluations from a randomly drawn state in each future to represent alternative performances in that future.

The number of scenarios used S is a parameter of the simulation, taking on the values 1, 3, 5, and 10. These values must be interpreted in light of there being 10 possible futures – the simulation cannot, for instance, provide results on the general use of 1, 3, 5, or 10 scenarios. For this reason I refer to the *proportion* of futures selected, termed the 'coverage' provided by a scenario model.

6.4 Comparing model results

I consider the idealised model, which uses all states and future-specific attribute weights, as a kind of hypothetical ideal which a DM may strive for but is impossible to achieve, simply because of the amount of information demanded. The approximation of this idealised model by the rest of the models is assessed using the rank correlation between the non-idealised and idealised models' rank orders.

6.5 Parameter values used in the simulations

Table 1 summarizes the parameter values used in the simulation study. Two experiments were run. The first constitutes the main experiment, in which parameters relating to the implementation of scenario-based AHP are systematically varied (number of scenarios used, extent of between-scenario differences in DM preferences). To limit computation time, I held other parameters fixed. The second experiment varies these other parameters, which control problem context (number of alternatives and attributes, distributions of realizations over futures, within-future variation) and have been found by previous studies to exert some influence over model accuracy (e.g. Durbach and Stewart, 2012a; Durbach, 2014). Although a large degree of subjectivity in specifying simulation parameters is inevitable, the values here have been chosen on the basis of preliminary investigations and have been used to compare scenario models in the context of value function methods in Durbach and Stewart (2012a). A full factorial design is used, performing 100 simulation runs for each combination of parameters.

	Description	Experiment 1	Experiment 2
I	number of alternatives	9	9, 19
J	number of attributes	10	10, 20
$\mathbf{M}$	distribution of realizations over futures	Skew	Uniform, Skew
$\sigma^{(d)}$	controls within-future variation	0.05	0.01, 0.05, 0.10
ρ^z	correlation of attribute evaluations	0, 1/3, 2/3, 1	0
ρ^w	correlation of attribute weights	0, 1/3, 2/3, 1	0
S	number of scenarios used	1, 3, 5, 10	1, 3, 5, 10

Table 1: Parameter values used by the simulation study

7 Results

The accuracy with which scenario-based AHP models approximated the results of their full-uncertainty counterpart depended on two broad factors: one, the extent to which attribute evaluations and weights differed between futures; two, how scenarios were constructed. These factors are to a certain extent compensatory. When futures were very similar to one another, it mattered very little how many scenarios were used, or how these were constructed. However, when attribute evaluations and weights differed substantially between futures then scenario-based model accuracy was only high if a substantial proportion of these futures were sampled as scenarios and, to a lesser extent, if these scenarios were constructed by taking the mean performance in the most likely scenarios.

The largest positive effect of scenario construction was exercised by the number of scenarios that were used (scenario coverage), followed by whether futures were selected according to their likelihood or at random, and least of all by whether scenario-based evaluations used mean evaluations within the future or a randomly sampled state within the future. The performance of all scenario-based models was good when weights did not differ between futures at all ($\rho^w = 0.99$) and alternatives were also similar ($\rho^z \geq 0.66$), remaining above 0.75 even at low levels of scenario coverage. Beyond these levels of between-future similarity, the accuracy of scenario-based models dropped off substantially unless a high proportion (above 50%) of futures were covered by the scenarios. This deterioration in accuracy was slower when mean scenarios were used, and when scenarios were selected on the basis of likelihood. As a result, to attain the same level of accuracy, fewer scenarios could be used if these were constructed by the ML-Mean approach than if these were constructed by the Random-Random approach. For example, with both attribute weights and evaluations varying widely over futures ($\rho^z = \rho^w = 0.01$) the ML-Mean scenario model shows a steep improvement as scenario coverage increases from 10% of all futures to 50% of all futures, beyond which accuracy remains constant at around 0.8. In contrast, the accuracy Random-Random scenario model improves roughly linearly over the full range of scenario coverage values. Similar but weaker patterns hold when futures are more highly correlated.

When all states were taken into account but weights were assumed to be the same in all states, accuracy was of course highest when this assumption was true (top-right panel of Figure 1, "Average weights" with $\rho^z = 0.99$) and lowest when the assumption was patently incorrect (bottom-right panel of Figure 1, "Average weights" with $\rho^z = 0.01$). Equally obviously, results for the "Average weights" model did not depend on scenario coverage, as scenarios are not used to summarize uncertainty in the states. Accuracy decreased more as a result of weights differing between futures when attribute evaluations were *also* very different between futures.

The accuracy of scenario-based AHP models at intermediate levels of scenario coverage (30-50%) improved with a larger set of alternatives, but there were no differences when scenario coverage was very low or very high (Figure 2). Reasons for this are not clear, and

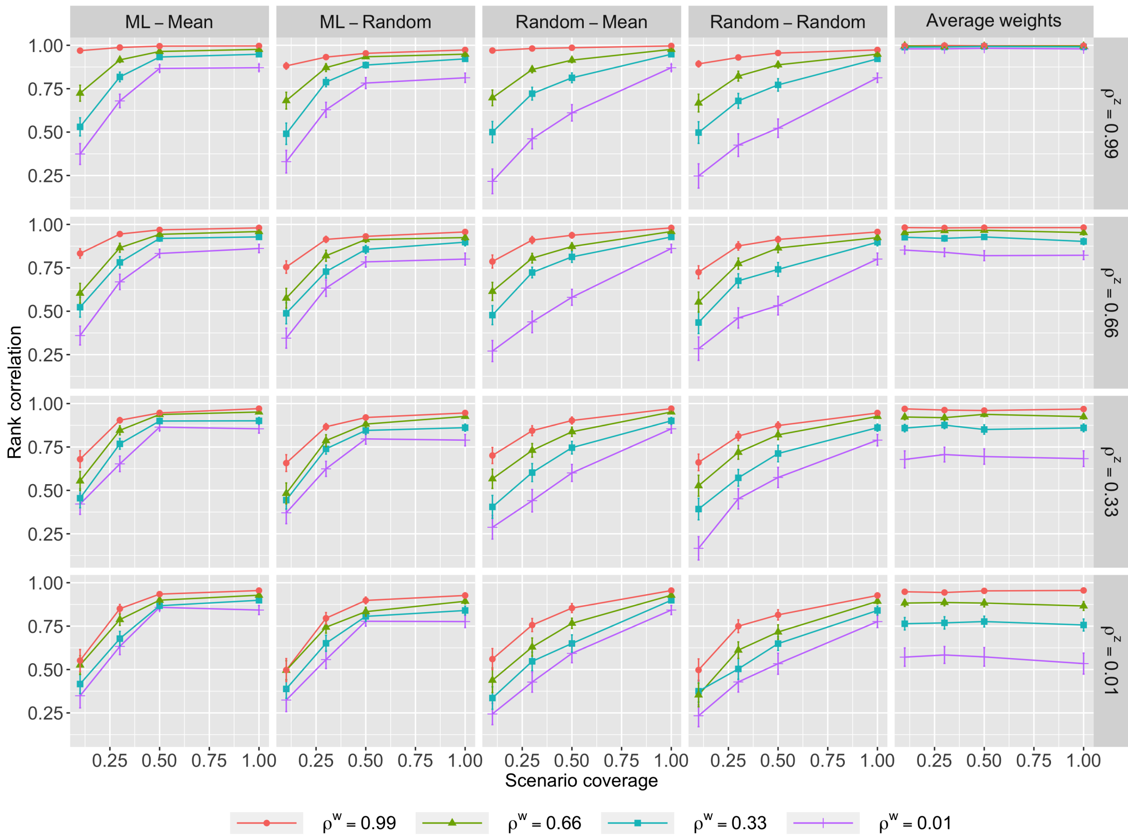

Figure 1: Accuracy of scenario-based AHP models, as measured by the correlation between each model's rank order and a rank order obtained from a full-uncertainty AHP model. For comparison, results from an "Average' weights" model that uses the full range of states but incorrectly assumes that criterion weights do not vary between scenarios is also shown (final column). Accuracy varies depending on the model used, the extent to which weights are correlated over futures (ρ^w), the extent to which attribute evaluations are correlated over futures (ρ^z), and the proportion of futures that are captured as scenarios (scenario coverage).

the finding is somewhat different from what was found for outranking methods (Durbach, 2014), where increasing the number of alternatives was associated with a moderate-sized improvement in performance across the full range of scenario coverage. Model performance was relatively insensitive to the number of criteria present in the decision problem, as was found in Durbach (2014).

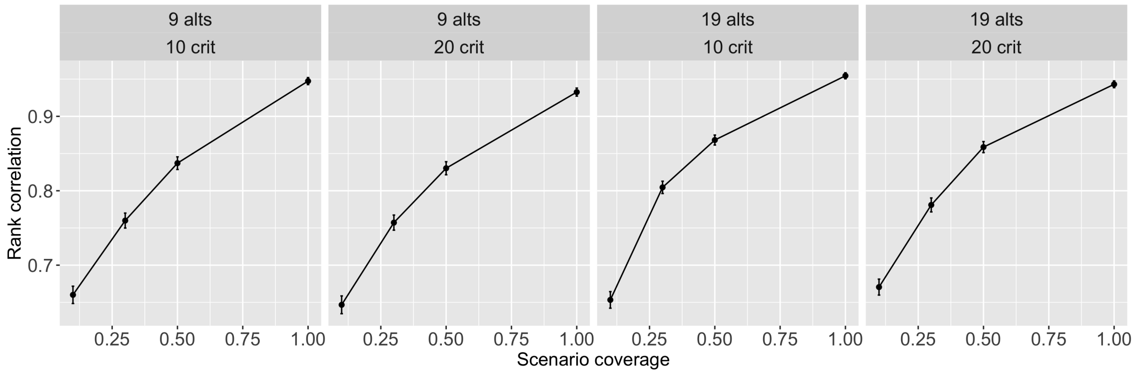

Figure 2: Mean accuracy of scenario-based AHP models as a function of the number of alternatives and criteria present in the decision problem. The accuracy of a model is measured by the correlation between its rank order and a rank order obtained from a full-uncertainty AHP model.

The accuracy of some scenario models – those that construct scenarios by randomly sampling a state from each selected future – deteriorated as attribute evaluations within each future became more variable (Figure 3). These deteriorations were much larger at high levels of scenario coverage. Models that construct scenarios by estimating mean performance within each selected future did not experience these deteriorations, presumably because means within a future are relatively unaffected by increasing variability. Model accuracy also deteriorated when futures were not all equally likely, this time equally so for all scenario-based models (Figure 3). This is readibly explained by the fact that the simulated scenario-based models do not weight futures by their likelihood. The findings in this study agree in broad terms with those reported for outranking-based scenario methods Durbach (2014). Scenario-based outranking model accuracy also deteriorated with attribute variability and if futures were not equally likely. However, that study found larger deteriorations at lower levels of scenario coverage, and very little effect when scenario coverage was high – the exact opposite of what is reported here for scenario-based AHP. The simulated conditions in the two studies are not precisely the same, and some outranking concepts do not map easily onto the AHP when constructing simulation studies, so that there is no reason to expect exactly the same findings from both studies. Nevertheless, the reason for the different way in which scenario coverage affects the relationship between attribute variability and model accuracy in the AHP and outranking methods is unclear and worth further investigation.

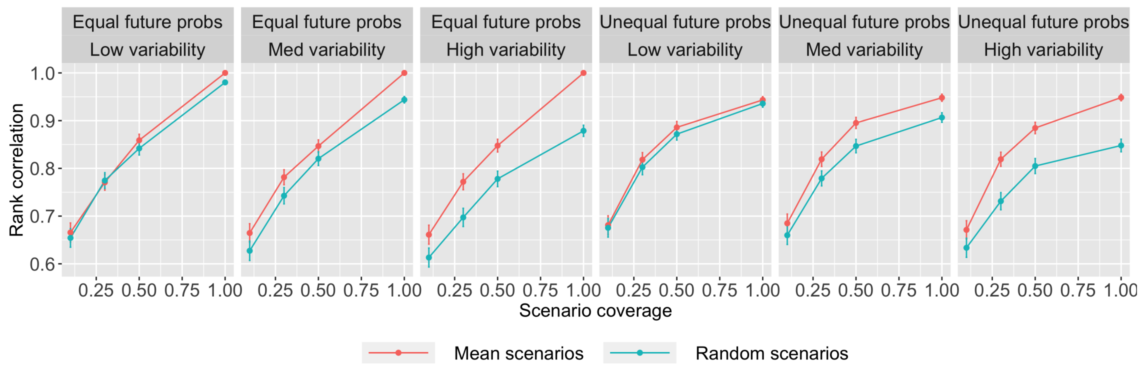

Figure 3: Mean accuracy of scenario-based AHP models as a function of how variable attribute evaluations are within futures and whether futures are equally likely to occur or not. The accuracy of a model is measured by the correlation between its rank order and a rank order obtained from a full-uncertainty AHP model.

8 Conclusions

Despite strong differences in opinion as to how much judgements and preferences should be quantified and mathematically aggregated, there is enough common ground between scenario planning and MCDA for a mutually beneficial exchange of ideas and techniques. For some decision problems at least, uncertainty may be large enough to preclude the use of probability-based methods but limited enough in scope that *some* quantification of performance is possible. Moreover circumstances may dictate that decisions should be transparent and justifiable. In such a case a scenario-based MCDA approach seems appropriate.

Scenario planning is more likely to find use in MCDA, and *vice versa*, through practitioners of its most popular variants. To this end this paper describes the integration of scenario planning techniques with the analytic hierarchy process (AHP). The AHP is one of the most popular MCDA approaches and one that is, arguably, more accessible to new users and more flexible in its approach to axiomatic principles and quantification. Scenario planners looking for an avenue into MCDA may find the AHP a useful introduction. Equally, there are probably more practitioners who favour AHP than any other MCDA method (although this is of course difficult to establish), so that if scenario planning ideas are take root in MCDA then acceptance by those working in AHP would be a step in the right direction. Ultimately, the best way to encourage the integration of MCDA with scenario planning is by applying these together for real-world decisions and reporting the successes and failures associated with these interventions.

Crucial to this integration is a common understanding of what a scenario is and how it should be handled for decision support. Although some authors interpret a scenario as synonymous with a "state of nature", this view is totally incompatible with the scenario planning view, and so in this paper I emphasise the view that scenarios are constructed as part of problem structuring, and that they have more in common with attributes than with states of nature. Specifically, scenarios are more naturally viewed as dimensions over which the decision maker can express preferences ("I prefer to reduce cost by $1000 in the first scenario than in the second") rather than as exogenous states over which quantities should be aggregated by mathematical expectation. Indeed, because the set of scenarios does not constitute a true probability space the concept of expectation is not defined. It is worth emphasising that although decision makers may be able to put some numbers to scenario likelihood if pressed, these numbers do not mean what we intend by either probability or likelihood.

Although other approaches are possible, the one advocated here is to view combinations of attributes and scenarios as "meta-attributes" over which alternatives can be evaluated using standard AHP questions. In some situations it may occur that a DM's preferences are scenario dependent, in the AHP context meaning that which attribute is viewed as most important may depend on which scenario occurs (preferences between alternatives will obviously differ too in non-trivial cases). Moreover it may only be possible to uncover these dependencies by going through the process of describing the possible futures in narrative detail, and that if this not done then a single set of preferences will be assumed (perhaps based on some unarticulated "average" scenario). Across a range of simulated decision problems a model that used every one of 400 states of nature (a practically impossible task for real decision problems) but failed to recognise strongly scenario-dependent preferences performed worse on average than a scenario-based approach that took only three of ten plausible scenarios in account, but in doing so was able to discern that weights differed between scenarios. This shows that there is a cost to ignoring the kind of information that only a scenario-based approach can provide, and that this cost can be substantial.

The same simulation results showed that the way that scenarios are constructed becomes more important as attribute evaluations and weights differ more between futures.

When futures were very similar to one another, the way scenarios are constructed and even how many scenarios are used did not matter much. When attribute evaluations and weights differed substantially between futures then models that sampled a substantial proportion of futures – what scenario planners term "bounding the future" (Schoemaker, 1991) – and if these scenarios were constructed by taking the mean performance in the most likely scenarios. The main effect of preferences and performance being similar across scenarios is thus that less scenarios are needed, and these are more robust to non-idealities in construction. Of course it would be difficult to know how similar preferences are across scenarios *a priori*, though our results indicated reasonable accuracy even with quite strongly different preferences between scenarios.

References

M. Amer, T. U. Daim, and A. Jetter. A review of scenario planning. *Futures*, 46:23–40, 2013.

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. Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach.* Kluwer Academic Publishers, Boston, 2002.

C. Boender, J. de Graan, and F. 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.

G. Cairns, G. Wright, R. Bradfield, K. van der Heijden, and G. Burt. Exploring e-government futures through the application of scenario planning. *Technological Forecasting & Social Change*, 71(3):217–238, 2004.

I. Durbach and T. Stewart. Integrating scenario planning and goal programming. *Journal of Multi-criteria Decision Analysis*, 12:261–271, 2003.

I. Durbach and T. Stewart. Using expected values to simplify decision making under uncertainty. *Omega*, 37(2):312–330, 2009.

I. Durbach and T. Stewart. A comparison of simplified value function approaches for treating uncertainty in multi-criteria decision analysis. *Omega*, 40:456–464, 2012a.

I. Durbach and T. Stewart. Modelling uncertainty in multi-criteria decision analysis. *European Journal of Operational Research*, 223(1):1–14, 2012b.

I. Durbach, R. Lahdelma, and P. Salminen. The analytic hierarchy process with stochastic judgements. *European Journal of Operational Research*, 238(2):552–559, 2014.

I. N. Durbach. Outranking under uncertainty using scenarios. *European Journal of Operational Research*, 232(1):98–108, 2014.

P. Goodwin and G. Wright. *Decision Analysis for Management Judgement.* John Wiley & Sons, Chichester, 2nd edition, 1998.

P. Goodwin and G. Wright. Enhancing Strategy Evaluation in Scenario Planning: a Role for Decision Analysis. *Journal of Management Studies*, 38(1):1–16, 2001.

P. Goodwin and G. Wright. *Decision Analysis for Management Judgement.* John Wiley & Sons, Chichester, 4th edition, 2009.

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. 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. 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.

M. Marttunen, J. Lienert, and V. Belton. Structuring problems for multi-criteria decision analysis in practice: A literature review of method combinations. *European Journal of Operational Research*, 263(1):1–17, 2017.

G. Montibeller, H. Gummer, and D. Tumidei. Combining scenario planning and multi-criteria decision analysis in practice. *Journal of Multi-criteria Decision Analysis*, 14: 5–20, 2006.

C. Ram, G. Montibeller, and A. Morton. Extending the use of scenario planning and MCDA for the evaluation of strategic options. *Journal of the Operational Research Society*, 62(5):817–829, 2010.

T. Saaty. *The Analytic Hierarchy Process*. McGraw-Hill, New York, 1980.

T. Saaty and L. Vargas. Uncertainty and rank order in the analytic hierarchy process. *European Journal of Operational Research*, 32(1):107–117, 1987.

A. Salo and R. Hämäläinen. Preference programming through approximate ratio comparisons. *European Journal of Operational Research*, 82(3):458–475, 1995.

P. Schoemaker. When and how to use scenario planning: a heuristic approach with illustration. *Journal of forecasting*, 10(6):549–564, 1991.

P. Schoemaker. Scenario Planning: A Tool for Strategic Thinking. *Sloan Management Review*, 36:25–25, 1995.

T. Stewart. Dealing with uncertainties in MCDA. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, chapter 11. Springer, 2005.

T. Stewart, S. French, and J. Rios. Integrating multicriteria decision analysis and scenario planning: review and extension. *Omega*, 41(4):679–688, 2013.

T. J. Stewart and I. Durbach. Dealing with uncertainties in MCDA. In S. Greco, M. Ehrgott, and J. R. Figueira, editors, *Multiple Criteria Decision Analysis*, pages 467–496. Springer, 2016.

A. Tversky and D. Kahneman. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90(4):293–315, 1983.

O. Vaidya and S. Kumar. Analytic hierarchy process: An overview of applications. *European Journal of Operational Research*, 169(1):1–29, 2006.

K. Van der Heijden. *Scenarios: The Art of Strategic Conversation.* John Wiley & Sons, New York, 1996.

J. van der Pas, W. Walker, V. Marchau, G. Van Wee, and D. Agusdinata. Exploratory MCDA for handling deep uncertainties: the case of intelligent speed adaptation implementation. *Journal of Multi-Criteria Decision Analysis*, 17(1/2):1–23, 2010.

P. Van Laarhoven and W. Pedrycz. A fuzzy extension of Saaty's priority theory. *Fuzzy Sets and Systems*, 11:199–227, 1983.

L. Vargas. Reciprocal matrices with random coefficients. *Mathematical Modelling*, 3(1): 69–81, 1982.

E. Vilkkumaa, J. Liesiö, A. Salo, and L. Ilmola-Sheppard. Scenario-based portfolio model for building robust and proactive strategies. *European Journal of Operational Research*, 266(1):205–220, 2018.

G. Wright and P. Goodwin. Future-focussed thinking: Combining scenario planning with decision analysis. *Journal of Multi-criteria Decision Analysis*, 8:311–321, 1999.