Risk Analysis, Vol. 29, No. 1, 2009 DOI: 10.1111/j.1539-6924.2008.01130.x

Uses and Misuses of Multicriteria Decision Analysis (MCDA) in Environmental Decision Making

Katie Steele,[1,]* Yohay Carmel,[2] Jean Cross,[3] and Chris Wilcox[4]

We focus on a class of multicriteria methods that are commonly used in environmental decision making—those that employ the weighted linear average algorithm (and this includes the popular analytic hierarchy process (AHP)). While we do not doubt the potential benefits of using formal decision methods of this type, we draw attention to the consequences of not using them well. In particular, we highlight a property of these methods that should not be overlooked when they are applied in environmental and wider decision-making contexts: the final decision or ranking of options is dependent on the choice of performance scoring scales for the criteria when the criteria weights are held constant. We compare this "sensitivity" to a well-known criticism of the AHP, and we go on to describe the more general lesson when it comes to using weighted linear average methods—a lesson concerning the relationship between criteria weights and performance scoring scales.

KEY WORDS: Analytical hierarchy process (AHP); environmental decisions; multicriteria decision analysis (MCDA); weighted average

1. INTRODUCTION

Multicriteria decision analysis (MCDA) has been recognized as an important tool in environmental decision making for formalizing and addressing the problem of competing decision objectives (Janssen, 1992; Lahdelma *et al.*, 2000; Linkov *et al.*, 2006; Regan *et al.*, 2007; Yatsalo *et al.*, 2007). In general, the overall goal is to determine a preference ordering among a number of available options $\{O_1 \ldots O_n\}$. In a reserve design problem, for instance, the options are the different combinations of land parcels to be nominated as reserves (see, e.g., Rothley, 1999; Moffett & Sarkar, 2006). The decisionmaker's preferences over options depend on how well they perform according to a number of objectives or "criteria" $\{C_1 \ldots C_M\}$ that have been identified by relevant stakeholders to be the (only) issues on which a decision between options should be made. The criteria that are relevant to reserve design, for example, might include ecological objectives such as protecting biodiversity or maintaining intact landscapes, as well as other social prerogatives like recreation opportunities and cost to the community. Initially, the options are assessed according to each criterion separately. In other words, for each criterion C_j the decisionmaker must provide a "score" for each option $\{O_1 \ldots O_N\}$, whether in cardinal or ordinal terms. Multicriteria methods are then employed to combine the criteria scores obtained for each option into an overall preference ranking or choice of

[1] Postdoctoral Research Fellow with the Australian Centre of Excellence for Risk Analysis (ACERA) and the Department of Philosophy, University of Sydney, Sydney, Australia.
[2] Division of Environmental, Water and Agricultural Engineering, Technion—Israel Institute of Technology, Haifa, Israel.
[3] School of Safety Science, University of New South Wales, Sydney, NSW, Australia.
[4] CSIRO Marine and Atmospheric Research, Hobart, Tasmania, Australia.
* Address correspondence to Katie Steele, Department of Philosophy, University of Sydney, Sydney, Australia; tel: +61 2 9036 5189; fax: +61 2 9351 3918; katiesiobhansteele@yahoo.com.

0272-4332/08/0100-0026$22.00/1

option. There are many suggested methods for performing this aggregation, each with its own informational requirements and mathematical properties (for a detailed survey of major multicriteria methods, see Figueira *et al.*, 2005; for a briefer survey, see Moffett & Sarkar, 2006).

Here, we focus on a special class of MCDA methods that depend on cardinal rankings of options for each criterion, and also cardinal weightings for the criteria. We chose these methods because they have been employed widely in environmental and other decision-making contexts. Our chief concern is to highlight the impact that scoring scales and other assumptions in the process have on decision outcomes, and to suggest some resolutions for the identified problems. Importantly, the idea is not to criticize any particular multicriteria methods out of hand, but rather to indicate how they must be approached if the final decisions are to be meaningful.

2. MCDA METHODS IN ENVIRONMENTAL DECISION-MAKING

MCDA methods vary with respect to the core decision rules that they implement in evaluating options in terms of the criteria (for a broad taxonomy of MCDA methods, see Keeney & Raiffa, 1993, ch. 3). This article focuses on cardinal multicriteria methods that permit "tradeoffs" between criteria. A prominent kind of cardinal method is the weighted linear average—it is usually specified in terms of normalized weightings for each criterion, as well as normalized scores for all options relative to each of the criteria. The final utility U for each option O_i is then calculated as follows:[5]

$$U(O_i) = \sum_{k=1}^{M} Z_k(O_i) \times w(C_k),$$

where $Z_k(O_i)$ is the normalized score of option O_i under criterion C_k and $w(C_k)$ is the normalized weighting for criterion C_k.

There are various methods and accompanying software packages based on the weighted linear average algorithm. These methods have the same underlying logic, but they differ in how they elicit and then calculate the weighting function w, and how they determine the scoring function Z_k for each criterion C_k.

[5]The dominant methods use a weighted *linear* average to determine the overall utility for options (i.e., they are additive), but there are potential alternative algorithms that also employ cardinal weights and utilities, such as the weighted *geometric* average.

Table I. AHP Matrix of Pair-Wise Criteria Comparisons

	C_1	C_2	...	C_M
C_1	C_1/C_1	C_1/C_2	...	C_1/C_M
C_2	C_2/C_1	C_2/C_2	...	C_2/C_M
...	...	...	...	...
C_M	C_M/C_1	C_M/C_1	...	C_M/C_M

2.1. Determining Criteria Weights

One very popular method, Saaty's (1980) analytic hierarchy process (AHP), determines the criteria weights indirectly based on scores of relative importance for each in pair-wise comparisons.[6] The comparison ratings are on a scale of one to nine, resulting in a ratio of importance for each pair with the maximum difference that one criterion is nine times more important than another. A matrix of pair-wise comparisons is determined in this way as per Table I (where C_i/C_j is just shorthand for the relative importance of C_i to C_j).

In the AHP, the final weightings for the criteria are the normalized values of the eigenvector that is associated with the maximum eigenvalue for this matrix. Saaty (1980) suggests that this procedure is the best way to minimize the impact of inconsistencies in the ratios.

There will be many different procedures for determining a set of positive weights that add to one. For instance, with reference to the above procedure, we might allow ratio comparisons of the importance of criteria to range between one and, say, 100, rather than one and nine.[7] Or else, we might disagree with the weights being determined by a normalized eigenvector. An alternative is to simply add the scores for each row in the matrix (i.e., add the scores for each criterion) and then normalize these sums in the same way that the eigenvector entries are normalized. Indeed, one might depart from the AHP altogether; the decisionmaker could, for example, directly assign numerical weights to criteria, as opposed to entertaining a series of pair-wise comparisons. Angelidis and Kamizoulis (2005) and Redpath *et al.* (2004) assign weights directly like this in their

[6]Many environmental decision-making studies make use of the AHP, including Anselin *et al.* (1989), Duke and Aull-Hyde (2002), Herath (2004), Li *et al.* (1999), Mardle *et al.* (2004), Mawapanga and Debertin (1996), Qureshi and Harrison (2001), Sadiq *et al.* (2003), Schmoldt *et al.* (1994), and Villa *et al.* (2002).

[7]Note, however, Saaty (1980) gives psychometric reasons for using a 1–9 scale, as opposed to a more fine-grained/course-grained scale, for assessing the relative importance of criteria.

Table II. Example Decision Problem with Two Criteria: Cost and Ecosystem Functioning

	Cost (Weight = 0.5)	Ecosystem Functioning (Weight = 0.5)	Overall Performance
Option 1	−10,000	0.2	$-10{,}000 \times 0.5 + 0.2 \times 0.5 = -4{,}999.9$
Option 2	−15,000	0.9	$-15{,}000 \times 0.5 + 0.9 \times 0.5 = -7{,}499.55$

respective environmental decision-making applications. In general, we simply want the criteria weights to be meaningful in the sense that they properly reflect the decisionmaker's attitudes/choice dispositions. This turns out to be more difficult than it looks. In later sections, we indicate why some straightforward and seemingly innocuous ways in which criteria weights are elicited yield decisions with an element of arbitrariness.

2.2. Scoring the Options

There is a general agreement that, if final utilities for options are to be calculated as a weighted linear average, then for each criterion, options are best scored on a 0–1 scale, where zero is the worst-case outcome and one is the best-case outcome. For instance, consider a decision problem for which there are only two relevant criteria against which options are rated—cost and ecosystem functioning. It seems inappropriate to have cost measured on a very large scale and ecosystem functioning measured on a 0–1 scale, as in the example depicted in Table II. The costing scale is so large in the above example that this criterion "swamps" the ecosystem functioning criterion. If costing were, scaled back to a 0–1 scale (keeping the criteria weights at 0.5), then this would probably better reflect the relative importance of the two criteria.[8]

We divide methods of normalization into two categories. The first, *relative normalization,* adjusts the scores for the options such that they sum to one. The second, *absolute normalization*, scales the scores such that each falls between zero and one, but the scores for the different options need not sum to one. To illustrate the distinction, consider a case in which we are trying to determine the performance of a number of potential water reservoirs according to the criterion of whether they can hold sufficient water. Assume that the demand for water is x L/day. The first approach recommends that the performance of the reservoirs sum to one. Perhaps the capacity of each reservoir is divided by the total capacity of all the potential reservoirs to get its score. The second approach does not have this adding-to-one requirement. It might be the case that two reservoirs have a capacity that is greater than or equal to the demand of x L/day, and so they should both be attributed a score of one.

AHP users have typically employed relative, rather than absolute normalization (see Anselin *et al.*, 1989; Herath, 2004; Li *et al.*, 1999; Mardle *et al.*, 2004; Mawapanga & Debertin, 1996; Sadiq *et al.*, 2003). In fact, the traditional AHP recommends that scores for options relative to each criterion be determined in exactly the same way that criteria weights are determined (Saaty, 1980). Of course, there are other examples of relative normalization that do not involve matrix eigenvectors (such as in the reservoir example above). There are a number of multicriteria applications in environmental decision making that incorporate scales that conform to absolute normalization, that is, normalization with respect to a given minimum and maximum value. Examples can be found in Angelidis and Kamizoulis (2005), Ausseil *et al.* (2007), Janssen *et al.* (2005), Qureshi and Harrison (2001), Bojórquez-Tapia *et al.* (2004), and Redpath *et al*. (2004).

There are some good reasons for preferring absolute normalization to relative normalization. Performance scores do not play the same role as weights in the weighted linear average function, and there is no good reason for requiring that they sum to one for any criterion across the different options. Moreover, it is surely more efficient not to have to rescale the criteria whenever there is a change in the option set, as is the case for relative normalization.[9] We note

[8]It will become clear later in the article that normalizing the scoring scales for criteria is in fact neither necessary nor sufficient for properly representing the decisionmaker's attitudes regarding the relative importance of the criteria. Many find it easier, however, to work with normalized scoring scales.

[9]Note also that rescaling the criteria can lead to changes in the final ranking of options, if the criteria weights are not also changed appropriately. We discuss this issue in more depth in the subsequent sections of the article.

that some methods of absolute normalization depend on the nature of the option set because they assign minimum (zero) and maximum (one) values according to the worst and best performing options available. At least in this case, however, changes to the option set do not *necessarily* require changes in the performance-score scaling. It is only when changes to the option set result in new worst or best performance levels for any of the criteria that utility scales for those criteria must be recalibrated.

3. VARIABLE SCORING SCALES: WHAT SENSITIVITY ANALYSIS NEGLECTS

Sensitivity analysis is a common practice conducted to test how changes in model parameters affect final outcomes (for a discussion of decision-analytic sensitivity analysis, see von Winterfeldt & Edwards, 1986, ch. 11 or Janssen, 1992, ch. 4). Many researchers test the sensitivity of a decision to the particular values of criteria weightings selected (whether via the AHP or by some other method). This might be a rather informal procedure whereby the weights assigned by different groups in the populace are compared, as in Duke and Aull-Hyde (2002). Or it might be more formal: Choo *et al.* (1999) demonstrate that the way in which questions are posed to elicit criteria weights affects these weights and thus the ordering of options. Redpath *et al.* (2004) graph the differences in final scores for options, given the varying criteria weights that are assigned by different sectors of the community. Similarly, Janssen *et al.* (2005) initially assume that each of their decision criteria should be weighted equally, and then consider how the final results change if the weights are altered to reflect the various political positions. Some focused more explicitly on the robustness of decisions based on particular sets of weights: Herath (2004) compared results for different interest groups in the community, and for each group, investigated how much change in the weights could be tolerated before the final ranking of options changed. Rothley (1999) also considers the sensitivity of a particular decision to perturbations in the proposed criteria weights.[10]

While the dependence of decisions on criteria weights has received a lot of attention, what might be overlooked by decisionmakers is the sensitivity of final rankings of options to the choice of scale for *indicators* of criteria performance *when weights are held constant*. Consider a specific hypothetical example where the decision is how much energy to produce via wind turbines (Table III). For simplicity, we assume there are only two relevant aspects of wind energy: it reduces CO_2 emissions to the atmosphere (replacing fossil fuels), while it may cause bird mortality (birds hit the turbines). Table III provides a summary of these criteria. We also assume that both indicators have a linear relationship with respect to the performance score for the criteria.[11]

Even if we have decided that the scale for the performance indicators for each criterion is linear, there remain many ways to set the scale, depending on how we want to fix the zero and one extreme points. In Table IV, we suggest two different rationales for fixing the zero and one levels (both examples of absolute normalization). For future reference, we refer to these as the "extreme scale" and the "target scale." They are both plausible and natural ways to set the scale (and are certainly not the only ones). Essentially, the extreme scale assigns a broader set of values to the 0–1 range, as compared to the target scale. Table IV also shows how "extreme" and "target" scaling might be settled for our particular decision problem.

The final ranking of options in this example decision problem can be sensitive to the choice of the zero and one endpoints for scoring the criteria (i.e., sensitive to the performance scoring scales for the criteria), *even when the weights for the criteria are held constant*. Assume (i) there are only two viable options, *O1* and *O2*, (ii) the options yield certain (as opposed to probabilistic) outcomes, (iii) the weights for the two criteria of reducing CO_2 emissions and bird mortality are equivalent, and (iv) the options have the characteristics described in Table V.

The final ranking of options *O1* and *O2* depends on the precise way in which the indicators for the criteria are scaled (Table V). When we use the "extreme scaling," which essentially maps a wider interval of bird kill and reduced CO_2 values to the 0–1 scales, *O1* scores better than *O2.* When we apply the narrower "target scaling," however, *O2* comes out better than *O1*. (The two criteria are assigned equivalent weightings of 0.5 in both cases.)

[10] There are many other applications of MCDA in the literature that incorporate sensitivity analysis, particularly with respect to criteria weights.

[11] In fact, throughout the article, we discuss variability in scaling in terms of how the endpoints (the minimum and maximum performance score values) are assigned. The shape of a criterion's scoring curve relative to the relevant indicator can also vary; it need not be linear. We set that complication aside, however.

Table III. Criteria for Wind Energy Decision Problem

Value/Criterion	← Relationship	Indicator	← Relationship	Measurable
Reducing CO_2 emissions	Linear (by assumption)	Total reduction in annual CO_2 emissions	?	D = annual CO_2 emissions for wind farm of size S
Bird mortality	Linear (by assumption)	Total # of killed birds	?	N = number of birds killed annually by a wind farm of size S

Table IV. Two Example Linear Scoring Scales (the "Extreme" and "Target" Scales) Applied to Each of the Wind Energy Criteria

Scaling	Criterion	Score = 0	Score = 1	Linear Relationship
Extreme		Intolerable level	Best that one cares about	
	Reducing CO_2 emissions	0% of CO_2 emissions	100% of CO_2 emissions	$y = 0.01 \times R$
	Bird mortality	5,000 birds annually	0 birds	$y = -0.0002 \times B + 1$
Target		Status quo level	Target level	
	Reducing CO_2 emissions	5% of CO_2 emissions	15% of CO_2 emissions	$y = 0.1 \times R - 0.5$
	Bird mortality	1,000 birds annually	0 birds	$y = -0.001 \times B + 1$

Table V. Effect of Scoring Scale Selection of the Utility of Wind Energy Options (When Criteria Weights Are Held Constant)

Option	Birds	Reduction CO_2	Overall Utility (Extreme Scaling)	Overall Utility (Target Scaling)
O1	400	5%	$0.5 \times (0.92 + .05) = 0.485$	$0.5 \times (0.6 + 0) = 0.3$
O2	1,000	15%	$0.5 \times (0.8 + 0.15) = 0.475$	$0.5 \times (0 + 1) = 0.5$

It is clear that the wider the scaling for an indicator, the less difference there is between the performance of options relative to that criterion and the less influence that criterion has on the final ranking of options. If, for example, the extreme bird kill had been taken to be 10,000 birds, *O1* would score $0.5 \times (0.96 + 0.05) = 0.505$ and *O2* would score $0.5 \times (0.9 + 0.15) = 0.525$. This time *O2* is favored (as per the "target scaling"); the significance of the difference in bird kill between the two options has diminished in the calculation of overall utility although the weights are unchanged. The problem is that there are infinitely many ways to set the scales for the performance indicators for the different criteria. Thus, not only is the final ranking of options sensitive to the choice of weightings, it is also sensitive to the choice of scales used to measure performance for the criteria.

4. INTERDEPENDENCE OF WEIGHTS AND UTILITY SCALES

One may propose to conduct sensitivity analysis both on criteria weights *and* on their choice of scales to measure performance for criteria. However, conducting more sensitivity analysis does not necessarily improve the quality of the decision solutions. The concern is rather that researchers appreciate and make explicit in their methodology the fact that criteria weights, taken on their own, are meaningless. It makes no sense to determine criteria weights independent of the scales used to score options against those criteria. This is the real lesson to be gained from recognizing that the final ranking of options can be sensitive to changes in the performance scoring scales for criteria *when weights are held constant*. The point is just that weights and performance scales must be aligned: if we change the one, then we must also change the other if we want to properly represent the decisionmaker's preferences.

Many environmental decisions are likely to involve criteria for which more than one set of indicator scales will seem very plausible and where values that could reasonably be chosen for zero and one end points lead to quite different ranges of scaling. If individual decisionmakers are not attentive to the relationship between weights and scales, it is possible that they will assign weights to criteria with some vague idea of performance scoring scales at the back of their mind, and yet couple these weights with entirely different scoring scales in their final decision model. In such cases the numbers used in the weighted average calculations will not be entirely meaningful and thus the final ranking of options will have an element of arbitrariness.[12] The situation is exacerbated where multiple stakeholders are involved. When asked to weight criteria without specific indicator scales defined, different stakeholders are likely to have different indicator scoring scales in mind when they are considering the relative importance of the criteria. This would make any aggregation of these individual weights into group weights rather meaningless.

Some multicriteria methods, in particular, the AHP, serve to further entrench the problem. The structure of the original AHP obscures to decision-makers the fact that criteria weightings and performance scoring scales should be aligned. We consider this to be the core problem with the AHP, but criticism of the method has tended to focus on what could be viewed as an upshot of the core problem—the ranking of options can change when the set of available options changes. Indeed, Dyer (1990, p. 252) demonstrates that the ranking of two options A and B according to the AHP can depend on whether some other option, D, is included in the option set. The rank reversal issue for the AHP arises because relative normalization is used (with constant weights); if performance scores for each criterion must add to one, the scores are necessarily sensitive to what options are in the option set. But this need not be an essential part of the AHP. In fact, Saaty (1987) suggests that options might alternatively be ranked according to an "absolute scale," or in our terms, via a scale that conforms to absolute normalization. For instance, the options might be judged as having "high," "medium," or "low" performance relative to each criterion, and then assigned corresponding numerical values. In this case, final rankings of options via the AHP will *not* be sensitive to the number and character of options in the option set.

The possibility of rank reversals when the option set is changed and weights are kept constant is, however, not in fact the real problem with the traditional AHP. Or else, there is more than one way to solve this problem; rather than stipulating that an "absolute scale" should replace the eigenvector assignment of performance scores, we might instead have insisted that criteria weights cannot be kept constant when the option set is changed. The idea would be as follows: if you add new options, then you have necessarily changed the option scores relative to each criterion (because they must always add to one), and so there will need to be an appropriate adjustment in the relative weightings for the criteria.[13] Rank reversals amongst the original options would be a clear sign that the weightings for the criteria were not suitably recalculated. This solution to rank reversals in the AHP is the more revealing approach because it addresses the deeper issue with assigning weights and performance scores in a multicriteria decision model—the two must be aligned!

There are a number of suggestions in the literature regarding appropriate elicitation of weights. Edwards and Barron (1994, p. 315) note that "weights reflect the range of the attribute (criterion) being weighted, as well as its importance." They go on to suggest modifications of the weight-elicitation process in the SMART multicriteria model. In a similar spirit, Belton (1986) offers suggestions for improving weight elicitation in the AHP. Dyer (1990, p. 256) also makes some useful suggestions with regard to the alignment of criteria weights and scoring scales in his discussion of the AHP. When it comes to the elicitation of meaningful criteria weights, he says:

> the decision maker could be told the ranges over which the alternatives under consideration actually vary. Then he could be asked to answer the pair-wise comparisons regarding the importance of the criteria by considering the relative importance of a change from the least preferred to the most preferred values for criterion i compared to a similar change for criterion j.

Dyer is here appealing to absolute scoring scales for the criteria, but ones that have minimum and maximum endpoint values matching the least and best performing options. The general idea is to compare the importance of the same sized increment in performance scores for the various criteria, and regard

[12] It is possible that the decision model has errors in the coupling of weights and scoring scales, but that this does not affect the decision solutions. Indeed, it may be the case that a particular decision result is robust with respect to quite significant changes in the model parameters, and even to the multicriteria model itself (see Salminen *et al.*, 1998 and Yatsalo *et al.*, 2007 for studies comparing decision solutions across different multicriteria models). Not all decisions, however, will have solutions with this level of robustness. Moreover, it is simply good practice for a model user to properly interpret the parameters of the model, even if it makes no difference to the solutions in the case at hand.

[13] Dyer (1990, p. 254) notes that such a move amounts to a significant modification of the traditional AHP. The traditional version of the AHP involves the "principle of hierarchic composition." This principle stipulates that the weights on different levels of the hierarchy can be determined independently (i.e., weights for criteria and scores for options relative to each criterion can be determined independently). What we are suggesting here is in direct opposition to this principle.

this as their relative weightings.[14] This is a useful technique for eliciting criteria weights, whatever weighted average multicriteria method one chooses.

5. CONCLUDING REMARKS

In this article, we examined an error that is easy to make when using weighted average multicriteria decision models, and that is to treat the weights and performance scoring scales in the weighted average decision algorithm as completely separate measures. In such circumstances, it is possible to change the final ranking of options just by recalibrating the scoring scales for the criteria. This arbitrariness is not a feature or a fault of the formal model. It is rather a misuse of the weighted average decision method. To address the issue, decisionmakers must ensure that the numerical criteria weights reflect the relative importance of the criteria *given the way in which the performance-scoring scales for the criteria have been calibrated*.

It might be argued that stakeholders cannot be expected to bear such complicated modeling details in mind when they are asked to nominate the relative importance of decision criteria. In many real examples of environmental decision making, indicators for criteria performance are not easily understood by stakeholders and the significance of performance levels is open to dispute. For instance, in Janssen *et al.*'s (2005) example concerning wetland management, there are a number of decision criteria with respective indicators, including climate change, evaluated in terms of net greenhouse gas storage capacity, and water quality, evaluated in terms of mass nitrogen and phosphorus export. The significance of a particular greenhouse gas storage capacity or a particular amount of nitrogen/phosphorus export might be difficult for nonexperts to appreciate. This kind of problem is not insurmountable, however; it simply highlights the importance of including, in the decision process, a discussion of the meaning of performance indicator scales so that the stakeholders can judge the comparative significance of an incremental change in the various criteria scores. We might also investigate ways to elicit stakeholder opinions that place more responsibility on the experienced decision modeler to interpret, rather than just "read off," the information provided by stakeholders, so as to appropriately translate their opinions into numerical form.

Finally, at some point, it must be recognized that decision making is an intricate business, and that those involved are wiser to tackle the difficulties, rather than simply ignore them. Whatever the process of elicitation in a weighted average multicriteria decision method, it is important that the relationship between criteria weights and performance scoring scales is properly appreciated, otherwise there is no good reason to think that the final decision model will accurately reflect stakeholder views.

[14]Note also that Keeney and Raiffa (1993, ch. 3) provide a comprehensive treatment of the problem of determining the appropriate "tradeoffs" between criteria, or in other words, the appropriate scoring functions and associated weightings for the criteria. The central idea is to examine the "indifference curves" when two criteria are mapped on an x-y axis, that is, examine how the scores for one criterion equate in terms of goodness to the scores for another criterion.

ACKNOWLEDGMENTS

This article developed out of the "Robust Multi-Criteria Decision Analysis Working Group" held in Hobart in August 2007, sponsored by the Australian Centre of Excellence for Risk Analysis (ACERA). We are particularly grateful to Helen Regan for organizing the workshop, and also for helpful feedback on an earlier draft of the article. Thanks also to two anonymous referees for critical and constructive comments.

REFERENCES

Angelidis, M. O., & Kamizoulis, G. (2005). A rapid decision-making method for the evaluation of pollution-sensitive coastal areas in the Mediterranean sea. *Environmental Management*, *35*(6), 811–820.

Anselin, A., Meire, P. M., & Anselin, L. (1989). Multicriteria techniques in ecological evaluation: An example using the analytic hierarchy process. *Biological Conservation*, *49*, 215–229.

Ausseil, A.-G. E., Dymond, J. R., & Shepherd, J. D. (2007). Rapid mapping and prioritisation of wetland sites in the Manawatu-Wanganui region, New Zealand. *Environmental Management*, *39*, 316–325.

Belton, V. (1986). A comparison of the analytic hierarchy process and a simple multi-attribute value function. *European Journal of Operational Research*, *26*, 7–21.

Bojórquez-Tapia, L. A., de la Cueva, H., Díaz, S., Melgarejo, D., Alcantar, G., Solares, M. J. *et al*. (2004). Environmental conflicts and nature reserves: Redesigning Sierra San Pedro Mártir National Park, Mexico. *Biological Conservation*, *117*, 111–126.

Choo, E. U., Sconer, B., & Wedley, W. C. (1999). Interpretation of criteria weights in multicriteria decision making. *Computers and Industrial Engineering*, *37*, 527–541.

Duke, J. M., & Aull-Hyde, R. (2002). Identifying public preferences for land preservation using the analytic hierarchy process. *Ecological Economics*, *42*, 131–145.

Dyer, J. S. (1990). Remarks on the analytic hierarchy process. *Management Science*, *36*(3), 249–258.

Edwards, W., & Barron, F. H. (1994). SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, *60*, 306–325.

Figueira, J., Greco, S., & Ehrqott, M. (2005). *Multiple Criteria Decision Analysis: State of the Art Surveys*. New York: Springer.

Herath, G. (2004). Incorporating community objectives in improved wetland management: The use of analytic hierarchy process. *Journal of Environmental Management*, *70*, 263–273.

Janssen, R. (1992). *Multiobjective Decision Support for Environmental Management*. Dordrecht: Kluwer Academic Publishers.

Janssen, R., Goosen, H., Verhoeven, M. L., Verhoeven, J. T. A., Omtzigt, A. Q. A., & Maltby, E. (2005). Decision support for integrated wetland management. *Environmental Modelling and Software*, *20*, 215–229.

Keeney, R., & Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge & New York: Cambridge University Press.

Lahdelma, R., Salminen, P., & Hokkanen, J. (2000). Using multicriteria methods in environmental planning and management. *Environmental Management*, *26*(6), 595–605.

Li, W., Wang, Z., & Tang, H. (1999). Designing the buffer zone of a nature reserve: A case study in Yancheng Biosphere Reserve, China. *Biological Conservation*, *90*, 159–165.

Linkov, I., Satterstrom, F. K., Kiker, G., Batchelor, C., Bridges, T., & Ferguson, E. (2006). From comparative risk assessment to multi criteria decision analysis and adaptive management: Recent developments and applications. *Environment International*, *32*, 1072–1093.

Mardle, S., Pascoe, S., & Herrero, I. (2004). Management objective importance in fisheries: An evaluation using the analytic hierarchy process. *Environmental Management*, *33*(1), 1–11.

Mawapanga, M. N., & Debertin, D. L. (1996). Choosing between alternative farming systems: An application of the analytic hierarchy process. *Review of Agricultural Economics*, *18*(3), 385–401.

Moffett, A., & Sarkar, S. (2006). Incorporating multiple criteria into the design of conservation area networks: A minireview with recommendations. *Diversity and Distributions*, *12*, 125–137.

Qureshi, M. E., & Harrison, S. R. (2001). A decision support process to compare riparian revegetation options in Scheu Creek catchment in North Queensland. *Journal of Environmental Management*, *62*, 101–112.

Redpath, S. M., Arroyo, B. E., Leckie, F. M., Bacon, P., Bayfield, N., Gutiérrez, R. J. *et al.* (2004). Using decision modeling with stakeholders to reduce human-wildlife conflict: A raptor-grouse case study. *Conservation Biology*, *18*(2), 350–359.

Regan, H. M., Davis, F. W., Andelman, S. J., Widyanata, A., & Frecse, M. (2007). Comprehensive criteria for biodiversity evaluation in conservation planning. *Biodiversity Conservation*, *16*, 2715–2728.

Rothley, K. D. (1999). Designing bioreserve networks to satisfy multiple, conflicting demands. *Ecological Applications*, *9*(3), 741–750.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.

Saaty, T. L. (1987). Rank generation, preservation, and reversal in the analytic hierarchy decision process. *Decision Science*, *18*, 157–177.

Sadiq, R., Husain, T., Veitch, B., & Bose, N. (2003). Evaluation of generic types of drilling fluid using a risk-based analytic hierarchy process. *Environmental Management*, *32*(6), 778–787.

Salminen, P., Hokkanen, J., & Lahdelma R. (1998). Comparing multicriteria methods in the context of environmental problems. *European Journal of Operational Research*, *104*, 485–496.

Schmoldt, D. L., Peterson, D. L., & Silsbee, D. G. (1994). Developing inventory and monitoring programs based on multiple objectives. *Environmental Management*, *18*(5), 707–727.

Villa, F., Tunesi, L., & Agardy, T. (2002). Zoning marine protected areas through spatial multiple-criteria analysis: The case of the Asinara Island National Marine Reserve of Italy. *Conservation Biology*, *16*(2), 515–526.

von Winterfeldt, D., & Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge, UK: Cambridge University Press.

Yatsalo, B., Kiker, G. A., Kim, J., Bridges, T. S., Seager, T. P., Gardner, K. *et al*. (2007). Application of multicriteria decision analysis tools to two contaminated sediment case studies. *Integrated Environmental Assessment and Management*, *3*(2), 223–233.