ORNL
MASTER COPY

JAN 19 1995

ORNL/SUB/94/03371

ENERGY DIVISION

IDEA
INTERNATIONAL DEVELOPMENT AND
ENERGY ASSOCIATES, INC.

B. F. Hobbs and P. Meier

INTEGRATED RESOURCE PLANNING
AND THE ENVIRONMENT: A GUIDE TO THE USE
OF MULTI-CRITERIA DECISION METHODS.

Submitted July 1994

RECEIVED
NOV 16 1995
OSTI

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Prepared by a subcontractor
of the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
managed by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-84OR21400

HH
PROCESSED FROM BEST AVAILABLE COPY

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

db
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

International Development and
Energy Associates, Inc.

934-9788

APR 7 1988

FILE

**INTEGRATED RESOURCE PLANNING
AND THE ENVIRONMENT: A GUIDE TO THE USE
OF MULTI-CRITERIA DECISION METHODS.**

B.F. Hobbs and P. Meier

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

**INTEGRATED RESOURCE PLANNING
AND THE ENVIRONMENT: A GUIDE TO THE USE
OF MULTI-CRITERIA DECISION METHODS.**

B.F. Hobbs and P. Meier

July 1994

Submitted to

Oak Ridge National Laboratory
P. O. Box 2002
Oak Ridge, Tenn. 37831-65011

by

IDEA, Inc.
Suite 710
1111 14th Street NW
Washington, DC 20005

ACKNOWLEDGMENTS

We thank Seattle City Light and Centerior Energy for making this study possible, and the participants in the workshops for their irreplaceable contributions. Particular thanks go to Allen F. Wilson and Wing Chen at Seattle City Light, and to Carlton Buford, Art Caune, and Rob Martinko at Centerior Energy for assistance with the modeling. Without the help of Allen F. Wilson of Seattle City Light, and Lee Czerniawski of Centerior Energy, this project would not have been possible: they acted as project liaisons, helped organize the workshops, and provided the essential guidance and support that kept the project focused on real utility planning concerns. We also thank Eric Hirst and Stan Hadley of Oak Ridge National Laboratory for their valuable reviews and support. However, despite the assistance received from others, the opinions expressed herein, and the responsibility for the contents of the report, are solely those of the writers, and do not necessarily represent the views of Seattle City Light, Centerior Energy, Oak Ridge National Laboratory or the US Department of Energy.

ABBREVIATIONS

AHP	analytic hierarchy process
AVF	additive value function
ANN	annual optimization model (used by Seattle City Light)
BPA	Bonneville Power Administration
DSM	demand side management
DPU	Department of Public Utilities (of Massachusetts).
ELOE	expected loss of energy
EIS	environmental impact statement
GP	goal programming
GHG	greenhouse gas
IRP	integrated resource planning
LOLP	loss of load probability
LP	linear programming
MCDM	multi-criteria decision making
NEPA	National Environmental Policy Act
NRC	Nuclear Regulatory Commission
SCL	Seattle City Light

ABSTRACT

This report is intended as a guide to the use of multi-criteria decision-making methods (MCDM) for incorporating environmental factors in electric utility integrated resource planning (IRP). Application of MCDM is emerging as an alternative and complementary method to explicit economic valuation for weighting environmental effects.

We provide a step-by-step guide to the elements that are common to all MCDM applications. The report discusses (1) how environmental attributes should be selected and defined; (2) how options should be selected (and how risk and uncertainty should be accounted for); (3) how environmental impacts should be quantified (with particular attention to the problems of location); (4) how screening should be conducted; (5) the construction and analysis of trade-off curves; (6) dominance analysis, which seeks to identify clearly superior options, and reject clearly inferior options; (7) scaling of impacts, in which we translate social, economic and environmental impacts into value functions; (8) the determination of weights, with particular emphasis on ensuring that the weights reflect the trade-offs that decision-makers are actually willing to make; (9) the amalgamation of attributes into overall plan rankings; and (10) the resolution of differences among methods, and between individuals.

There are many MCDM methods available for accomplishing these steps. They can differ in their appropriateness, ease of use, validity, and results. This report also includes an extensive review of past applications, in which we use the step-by-step guide to examine how these applications satisfied the criteria of appropriateness, ease of use, and validity. Case material is drawn from a wide field of utility applications, ranging from project-level environmental impact statements to capacity bidding programs, and from the results of two case studies conducted as part of this research. In these case studies we examined the incorporation of environmental factors in the IRP process of Seattle City Light, and Centerior Energy (of Ohio). Each case study is available as a more detailed separate report (B. Hobbs and P. Meier, *Multi-criteria methods for resource planning: A case study of Seattle City Light*, IDEA, Inc., Washington, DC, May 1994; and P. Meier and B. Hobbs, *Multi-criteria methods for resource planning: A case study of Centerior Energy*, IDEA, Inc., Washington, DC, May 1994).

EXECUTIVE SUMMARY

Many states have now implemented, or are considering, the inclusion of environmental externalities in resource acquisition decisions by electric utilities. Some states are also considering whether to ask utilities to factor externalities into operating decisions, such as dispatch. Some states have already settled on a particular approach on how environmental factors are to be considered in integrated resource planning (IRP) -- Massachusetts, for example, requires explicit monetization of air emissions using values determined by its utility regulatory agency.

Other States are still exploring how this should be done. For example, the Ohio Public Utility Commission, which has jurisdiction over one of the case study utilities examined in this study¹, is at an early stage of this process. They have directed Centerior Energy include consideration of externalities in their next resource plan submission to the Commission, without yet specifying in detail how this should be done. Seattle City Light, the other case study entity, is, as a municipal utility, not subject to the jurisdiction of the Washington Public Service Commission; however, it stands under the directive of the City Council, which has required consideration of environmental issues in its resource plans for some years.

Approaches for considering environmental factors in IRP.

Any proposed method of incorporating externalities needs to meet three criteria: validity, appropriateness, and practicality. A valid method is one that is scientifically rigorous and arrives at results that accurately reflect the values held by decision-makers, while a practical method is one that can be implemented with reasonable time and effort. But how this balance is to be achieved is in fact critical. On the one hand it is not too difficult to propose a practical method that uses values of damage costs determined by perceived consensus and compromise in a regulatory proceeding. Yet arbitrary values are no substitute for rigorous analysis and documentation of a system's environmental impacts. On the other hand, a rigorous method that is perceived as too complicated or academic will simply not be implemented.

Four major approaches to the incorporation of environmental factors in IRP can be identified in current practice: (1) qualitative approaches; (2) rule-based approaches; (3) monetization, and (4) multicriteria decision methods. Each strikes a rather different balance between the criteria of validity, practicality, and appropriateness.

Qualitative approaches: Qualitative approaches take a number of forms, ranging from a simple listing of externalities associated with each option, to one in which some subjective judgment is applied to the severity of the impact in question (such as "high", "medium", "insignificant", etc.). Such approaches have long been a standard part of project level environmental impact statements: their advocates claim their inherent simplicity and flexibility is a major advantage

¹ In this study we worked with two utilities: Seattle City Light and Centerior Energy of Ohio -- with a primary focus on whether different methods yielded different results, and how utility planners judged appropriateness and ease of use.

given the danger that basic issues can quickly get bogged down in complex details. But whether the obvious practicality of the approach compensates for the troubling lack of precision, quantification, and validity is subject to question.

Rule-based approaches: The "rule based" approach is nothing more than a euphemism for arbitrary adjustment: here the emphasis is again almost entirely on practicality. Several States have adopted such methods under such labels as "percentage adders". To be sure, such seemingly arbitrary adjustments in fact reflect the very real frustrations of obtaining credible and defensible estimates of actual environmental damages; and may be justified largely on the intuitive ground that demand side management (DSM) programs avoid environmental costs associated with supply side resources, and therefore require some special consideration. But there is little discussion of the *actual* effects of such quantitative consideration measures in the literature, perhaps because the procedures appear so far removed from rigorous cost/benefit analysis that nobody has bothered to examine them systematically.

Monetization: The third approach, namely monetization of environmental externalities and their subsequent incorporation into traditional cost-benefit evaluation, is the obvious preference of environmental economists, with its seeming emphasis on validity. Yet the approach is beset with practical difficulties, not least of which is the limited ability of economists to make realistic estimates of actual damage costs. Because of these difficulties, many have advocated the use of control costs as a proxy for damages, yet this too can readily be shown to violate the axioms of benefit/cost analysis. Perhaps a more important problem with the monetization approach is that only those environmental impacts for which a monetization calculation proves convenient are included. In practice this means that air emissions are so quantified, while the impact of other emissions, and different types of impacts, tend to be ignored.

A further issue is that not everyone accepts the premises of benefit cost analysis, namely (1) that only people's preferences matter (i.e. that other species or natural areas do not have rights); and (2) that the distribution of benefits and costs does not matter. Moreover, when different options have very different impacts (dollars, human health impacts, loss of biodiversity), many individuals prefer to consider each in their natural units, rather than obscure the trade-offs by putting them into common units (particularly when there may be sharp disagreements about exactly how such a conversion is undertaken in practice).

Multi-criteria decision methods: Finally there are the multicriteria methods. Because of the problems inherent in the explicit valuation of externalities, multicriteria methods have been used for some time -- indeed, they have long been part of environmental impact assessments and power plant siting studies. More recently, they have come into widespread use in bidding systems used by utilities to acquire resources from independent power producers (IPPs). Typically in such systems alternative options are scored (or ranked) on some scale for a number of environmental and cost attributes, the attributes are given weights, and an overall ranking derived by adding the weighted scores. Multi-criteria methods have at least five purposes:

1. To communicate information about tradeoffs among attributes
2. To help decision makers make more consistent and rational evaluations of these tradeoffs. When alternatives have many attributes (ten or more are common in utility planning; e.g., cost, land use, endangered species, flexibility, reliability, water quality, emissions of CO₂, NOx, particulates, SO₂ ...), psychological research has shown that decision makers are inconsistent in their subjective evaluations of alternatives. Generally, the mind focuses on two or three attributes and ignores the others, or it flits

inconsistently among the attributes. Multi-criteria methods ask questions of decision makers to determine their priorities, and use the answers to systematically weed out undesirable alternatives and to indicate what alternatives might be most preferred.

3. To help decision makers make more consistent and rational evaluations of risk and uncertainty. Psychological studies show that decision makers have a difficult time evaluating risk in a consistent manner. Some multicriteria risk methods attempt to measure decision makers' attitudes towards risk (e.g., risk seeking, risk averse) and to consistently apply them to the problem. Other methods for initial screening of plans attempt to preserve a range of choice so that if the detailed engineering phase finds one alternative to be infeasible, other feasible plans will be still be available.
4. To help understand the values and perspectives of different interests. Multicriteria methods provide a framework for measuring the priorities of different groups, and for showing what the implications of each group's values would be for the decision.
5. To document how decisions are made. The use of multicriteria methods in federal decision making was spurred by the National Environmental Policy Act and the subsequent court rulings (Calvert Cliffs) which required agencies to document the rationale for decisions.

A final point to emphasize at the outset is that very often people do *not* know what they want: and it is therefore naive to view MCDM as simply a problem of eliciting preferences firmly established in someone's head -- as if these were eternal, constant and well defined. In reality, during the course of a planning exercise, people's attitudes will evolve in response to new information, interactions with others, and viewing the problem from different perspectives.

Indeed, when a decision is difficult because it involves a unique problem along with strongly held and conflicting values, most people will be unsure of their priorities. There is ample evidence that in such situations, the value judgments they articulate will depend strongly on supposedly irrelevant details of methodology, such as the exact phrasing of questions. In this case, a critical role of MCDM is to help people understand the problem better.

THE APPLICATION OF MCDM METHODS

In Chapter 2 we present a discussion of the sequence of steps necessary to apply an MCDM method to a decision problem. These steps are largely independent of the particular application involved, and apply not just to the incorporation of environmental factors in IRP but to a whole host of other potential utility applications from power plant siting to bidding programs. We begin with an overview of the process, and then discuss each of the steps in turn.

The application of MCDM methods involves the following steps:

1. Selection and definition of attributes, say A_i , $i=1, \dots, N$, selected to reflect important planning objectives and/or environmental concerns. System cost, reliability, impact on rates, air quality impacts, or impact on fisheries are examples: in this step we select which of these will be used in an application, and precisely how they should be defined. There are many issues to be considered here, including the need to avoid proliferation of attributes, and to avoid double counting.
2. Definition of the alternatives to be analyzed. Very often this also involves definition of alternative futures that capture factors over which utility planner have little or no control (such as natural gas prices, or the price of SO_2 allowances)

- 3 Quantification of the levels A_{ij} of the i attributes estimated for each of the j alternatives. This generally requires the application of some model to predict the impacts. Uncertainty and risk in attribute levels is quantified at this time.
- 4 Preliminary screening of alternatives. However, it is important that the options that remain for further analysis reflect a sufficiently diverse set of attribute values so that trade-offs can be examined in a meaningful manner. If, say, all ten plans that survive a preliminary screening exhibit very small differences in environmental impacts, it is not likely that environmental groups would accept the end-result.
- 5 Construction and analysis of trade-off curves
- 6 Dominance analysis, in which an alternative is screened out if it is dominated by another option. An alternative is dominated if there exists another plan that is just as good in all attributes, and strictly better in at least one.
- 7 Scaling of attributes, in which the level of an attribute is translated into a measure of value, $V_i(A_i)$ (also known as an attribute value function).
- 8 Selection of weights for each attribute. There are a large number of different techniques to elicit weights, each of which has advantages and disadvantages. How questions are asked about individuals preferences proves to be very important.
- 9 Determination and application of an amalgamation rule. Such rules combine the weights and value functions into a single overall value or ranking of the available options, or which reduces the number of options for further consideration to a smaller number of candidate plans. For example, the widely used "weighting-summation" rule ranks options by the score:

$$\text{maximize } w_i V_i(A_i)$$

where w_i is the weight, and $V_i(A_i)$ is the value function of attribute A_i .

- 10 Resolution of differences between methods, and between and among individuals.

While the steps appear conceptually straight forward, in general most applications give relatively little attention to a number of theoretical requirements for the results to be valid. Equally important, even though we present these elements in the form of a sequential list, in any actual application a certain amount of iteration between these steps will be necessary. The inter-relationship between these steps is indicated on Figure 1. Each of these steps is discussed in detail in Chapter 2.

Figure 2.1: Steps in a MCDM problem

Trade-off curves, in which the values of two attributes are displayed graphically, are a particularly useful tool for an understanding of the relationships among attributes: on Figure 2 we show a trade-off curve between CO₂ emissions and system cost for Seattle City Light: this displays what price must be paid to lower emissions, and what technology options are available to achieve CO₂ reductions: each point on the graph represents a particular resource mix for example, "cvhi", which has the highest costs and the lowest CO₂ emissions, includes 181 Mw of new wind capacity, 50Mw of geothermal, and 200Mw of fuel cells, as well as a high level of DSM. At the other end is the basecase, with the highest CO₂ emissions, but the lowest costs.

Figure 2: Trade-off curve

Whatever the method used, however, resolving differences among individuals is one of the central issues in MCDM methods. Ultimately, of course, it is the regulatory commission who decides what trade-offs are to be made, and what modifications must be made to a utility's proposal for the future resource mix. Nevertheless, MCDM methods make explicit these differences, because weights are expressions of individual's values. This is illustrated on Figure 3, in which we show the average weights given to different environmental and non-environmental attributes in the two case studies we conducted as part of this study. In each case we have sorted the individuals from left to right by the total weight given to environmental factors: in Seattle this ranges from a minimum of 10% to a maximum of 92%; whereas at Centerior, the range is from 18% to a maximum of 40%. The greater variation in the Seattle case study is a reflection of the much greater diversity of individuals, which included representatives from a business group, and an NGO, while in the Centerior case study, all individuals were members of the utility (albeit including the utility's environmental department). Thus the latter is clearly a more homogenous group in terms of their values.

Figure 3: Average weights by individual

REVIEW OF MCDM APPLICATIONS

There are literally hundreds of MCDM applications in a very wide range of fields. Even in the electric utility planning area there have been numerous applications -- ranging from screening of power plant sites to weighting-summation methods used to evaluate bids for providing additional supply resources from Independent Power Producers (IPPs). In this report, in Chapter 3, we review a representative sample of MCDM applications in electric utility planning or fields very closely related to it, by examining each application in terms of the ten-step process developed in Chapter 2. The following applications are reviewed:

Environmental impact assessment

 Puget Sound reliability Plan EIS

 FERC system for evaluating pipeline proposals

Resource bidding

 Niagara Mohawk (New York)

Transmission and distribution system planning

Supply capacity planning

 Netherlands

 Taiwan

IRP

 Wisconsin

System dispatch

 Ohio

National energy planning

 Israel

Applications in other countries

 Japan

 Power plant siting in Sri Lanka

 Power plant siting in Indonesia

SELECTING AN MCDM METHOD

How then does one proceed with selecting an appropriate method for a particular application? Given the focus in this report on how environmental factors can be brought into IRP, what practical guidance can be offered to a utility planning department charged with actually implementing the IRP process? We propose three criteria for selecting a technique, examined in detail in Chapter 4:

1. Is a method appropriate? That is, does it provide the information that is needed, does it mesh with the decision making style of the organization, and is it compatible with the data that are available?
2. How easy is the method to use? How much time is required to apply it, what training is needed to understand it, and what computer facilities, if any, are needed? (This is the practicality criterion noted in the introduction)
3. How valid is the method? Do the results of the method (e.g., rankings of alternatives) accurately reflect the values held by the decision makers? Are the method's assumptions and the evaluations it yields consistent with the values of the decision makers? Evaluations by an invalid method might bear little relationship to the tradeoffs decision

makers and the public are willing to make, and the chosen alternative might be preferred by no one. Finally, do the decision makers trust the method and its results?

The last criterion is important if different methods result in different decisions; if so, one needs to worry about which method more accurately captures the priorities of decision makers. If different methods lead to the same results then a method can be chosen on the basis of appropriateness and ease of use alone.

Appropriateness

What characteristics should an "appropriate" method have? Most follow from the real reasons for applying an analytical method to the decision process in the first place:

Display of trade-offs: For public-sector decisions, the primary purpose of multicriteria methods should be to present information on tradeoffs. A method should show, in a vivid and easily digestible manner, how different alternatives perform on different attributes and, by doing so, make it easier for fully informed decisions to be made. An important by-product of displaying tradeoffs is the elimination of alternatives that are "dominated" (that is, perform no better on all attributes than another alternative, while performing strictly worse on at least one attribute).

Risk and uncertainty: An appropriate method should recognize that information about risks and uncertainties regarding the impacts of electric utility projects is crucial to choices among alternatives. A method should display information on how alternatives affect the risks associated with each of the attributes.

Understanding: An appropriate method should be easily understood, since people with a wide variety of backgrounds will be applying it and interpreting its results.

Simplification: An appropriate method should simplify the decision for those who have the final responsibility for it, but without oversimplifying it. It is desirable to simplify the problem by avoiding presenting too many alternatives or too many criteria at once. Psychologists cite the famous "7 plus or minus 2" rule for how many facts that a decision maker can keep in mind at once. If there is too much information, much of it will be ignored or inconsistently applied. But oversimplification can occur if too few options representing too narrow a range of tradeoffs are presented. An appropriate method will screen out alternatives using, as examples, the following procedures:

- drop dominated alternatives;
- drop alternatives whose performance in one or more attributes is so poor that they are likely to be rejected on that basis alone ("exclusionary screening");
- keep alternatives which are robust; that is, which do well under a wide range of assumptions concerning weights, risk attitudes, and probability distributions of inputs; and
- keep alternatives that emphasize different attributes so that the decision makers are presented with a real choice. (Otherwise, the real decision is being made by analysts, not those who should be responsible.)

Ease of Use

Ease of Use

Ease of use is an empirical question for which it is hard to set firm guidelines, and which is highly dependent upon the particular set of users. If the users are limited to just utility planners one may come to one set of conclusions; if, as we advocate, the user group includes representatives from interest groups and the public, one may come to an altogether different set of conclusions. Indeed, the two case studies that follow support this particular contention: in Seattle, the user group was much broader (and included members of local interest groups) than in the case of Centorior, where participation was limited to utility planners. The findings about ease of use of methods did indeed differ somewhat.

The time available proves to be an important issue. The effort that can be put into explaining a method (and of course the effectiveness of that training effort) will certainly affect how participants feel about ease of use. A number of studies have contrasted the time and effort required of decision-makers to use a variety of weighting and value-scaling techniques. Even the most difficult methods, such as indifference trade-off weighting or risky utility function scaling, have rarely taken more than three hours for individuals. However much less is reported in the literature about groups that include larger numbers of non-technically trained people. The Seattle case study, however, tends to suggest that the members from the citizen and environmental groups who participated were just as able to understand the methods as utility planners.

Validity

Of the three criteria, validity is the most difficult to determine in practice, because there are so many dimensions to the concept. The literature distinguishes between (1) predictive validity; (2) estimative validity; (3) construct validity; (4) convergent validity; (5) inter-expert validity; (6) inter-temporal validity, and (7) normative validity. Each of these addresses a different dimension, and is reviewed in some detail in Chapter 4.

1. BACKGROUND

Many states have now implemented, or are considering, the inclusion of environmental externalities in resource acquisition decisions by electric utilities. Some states are also considering whether to ask utilities to factor externalities into operating decisions, such as dispatch. Indeed, some states have already settled on a particular approach on how environmental factors are to be considered in integrated resource planning (IRP) -- Massachusetts, for example, requires explicit monetization of air emissions using values determined by its utility regulatory agency.¹

Others are still exploring how this should be done. For example, the Ohio Public Utility Commission, which has jurisdiction over one of the case study utilities examined in this study, is at an early stage of this process. They have directed Centerior Energy² to include consideration of externalities in their next resource plan submission to the Commission, without yet specifying in detail how this should be done. Seattle City Light, the other case study entity, is, as a municipal utility³, not subject to the jurisdiction of the Washington Public Service Commission; however, it stands under the directive of the City Council, which has required consideration of environmental issues in its resource plans for some years.

What motivates this trend towards explicit consideration of environmental externalities in integrated resource planning? Certainly over the past three decades numerous Federal and State statutes have been enacted that are directly aimed at reducing the environmental impacts of power generation -- ranging from the Clean Air Act and its amendments to legislation on water and solid waste. But even when in compliance with such laws, there are residual emissions that merit consideration in utility decision-making. Moreover, as evidenced by the new market-based concepts introduced by the 1990 Clean Air Act Amendments, traditional command-and-control mechanisms to limit key pollutants have not been totally successful. Some important air pollutants are not now regulated, such as greenhouse gas (GHG) emissions, and some toxics. Finally, and perhaps most importantly, there are many other environmental concerns and community preferences not explicitly regulated as such, but which ought to be given some consideration -- witness the high value placed on whitewater recreation opportunities in the Pacific Northwest. Indeed, it is the very essence of IRP that the proper criterion for planning is not the narrow one of minimizing the utility's costs, but of minimizing a much broader definition of social cost.⁴ Moreover, despite evident progress in considering

¹ For recent reviews of state actions to include externalities in utility planning, see e.g. Haites and Hashem (1993) or Niemi *et al.*, (1993).

² Centerior Energy is the holding company for Cleveland Electric and Toledo Edison

³ Seattle City Light (SCL) is technically a department of the City Government.

⁴ The central concept of IRP is not just that demand-side resources need to be considered consistently with supply-side resources, but that a much broader range of criteria be applied to evaluation procedures. Least cost planning has been defined a process which "...explicitly includes

CONTENTS

1. Background	1
1.1 Approaches for considering environmental factors in IRP	2
1.2 Objectives	5
1.3 Outline of report	6
2. Application of MCDM methods	7
2.1 Attribute selection and definition	9
2.2 Definition of options	13
2.3 Quantifying impacts	17
2.4 Construction and analysis of trade-off curves	22
2.5 Screening	23
2.6 Dominance analysis	25
2.7 Scaling	31
2.8 Weighting methods	37
2.9 Amalgamation methods	44
2.9.1 Weighting summation	46
2.9.2 Goal programming	46
2.9.3 Multiplicative utility functions	47
2.10 Resolving differences	49
Annex I: Trade-off analysis for the three-dimensional case	55
Annex II: Generating trade-off curves	57
3. A Review of MCDM applications	61
3.1 Environmental impact assessment	61
3.2 Resource bidding systems	66
3.3 Transmission and distribution system planning	69
3.4 Supply capacity planning	75
3.5 IRP	81
3.6 System dispatch	84
3.7 National energy planning	86
3.8 Applications in other countries	89
4. Criteria for selecting MCDM methods	98
4.1 Appropriateness	98
4.2 Ease of use	100
4.3 Validity	100
Literature cited	103

environmental effects at the project level -- forced largely by the environmental impact statement (EIS) requirements introduced by the National Environmental Policy Act (NEPA) -- project level mitigation measures have a rather local impact, whereas consideration of environmental considerations at the resource planning stage -- where basic decisions about the resource mix are taken -- have a much more profound impact.⁵

1.1 Approaches for considering environmental factors in IRP.

Any proposed method of incorporating externalities needs to meet three criteria: validity, appropriateness, and practicality. A valid method is one that is scientifically rigorous and arrives at results that accurately reflect the values held by decision-makers, while a practical method is one that can be implemented with reasonable time and effort. But how this balance is to be achieved is in fact critical. On the one hand it is not too difficult to propose a practical method that uses values of damage costs determined by perceived consensus and compromise in a regulatory proceeding. Yet arbitrary values are no substitute for rigorous analysis and documentation of a system's environmental impacts. On the other hand, a rigorous method that is perceived as too complicated or academic will simply not be implemented.

Four major approaches to the incorporation of environmental factors in IRP can be identified in current practice:⁶ (1) qualitative approaches; (2) rule-based approaches; (3) monetization, and (4) multicriteria decision methods. Each strikes a rather different balance between the criteria of validity, practicality, and appropriateness.

Qualitative approaches:

Qualitative approaches take a number of forms, ranging from a simple listing of externalities associated with each option, to one in which some subjective judgment is applied to the severity of the impact in question (such as "high", "medium", "insignificant", etc.). Such approaches have long been a standard part of project level environmental impact statements⁷: their

conservation and load management programs as capacity and energy resources; it considers environmental and social factors as well as direct economic costs; it involves public participation; and it carefully analyzes the uncertainties and risks posed by different resource portfolios and by external factors" (Goldman *et al.*, 1989.)

⁵ In fact, the proposition that environmental factors must be considered in decision-making is not new, for the National Environmental Policy Act has required such consideration in the case of major federal actions since 1970. However, the requirement that environmental externalities need not merely be identified, but also be quantified and valued in economic terms, is more recent; The Pacific Northwest Electric Power Planning and Conservation Act of 1980 appears to contain the first statutory requirement that environmental costs and benefits need to be explicitly incorporated in benefit-cost analysis when evaluating new power resources. But by the late 1980s, many state regulatory commissions outside the Pacific Northwest had begun to take an active interest in assessing environmental externalities (Weil, 1991; Wood and Naill, 1992).

⁶ This corresponds to the categories in the EPRI guide (Temple, Barker & Sloane, *op. cit.*, 1991). The Pace University Study (Ottinger *et al.*, 1990) uses a somewhat different classification based on regulatory practice rather than methodology, namely: quantitative consideration, rate of return consideration, qualitative consideration, avoided cost consideration, environmental dispatch, and collaborative consideration.

⁷ We review one such recent example, the EIS for the Puget Sound Reliability Study, in Section 3.1.

advocates claim their inherent simplicity and flexibility is a major advantage given the danger that basic issues can quickly get bogged down in complex details⁸. But whether the obvious practicality of the approach compensates for the troubling lack of precision, quantification, and validity is subject to question.

Rule-based approaches:

The "rule based" approach is nothing more than a euphemism for arbitrary adjustment: here the emphasis is again almost entirely on practicality. Several States have adopted such methods under such labels as "percentage adders"⁹. To be sure, such seemingly arbitrary adjustments in fact reflect the very real frustrations of obtaining credible and defensible estimates of actual environmental damages; and may be justified largely on the intuitive ground that demand side management (DSM) programs avoid environmental costs associated with supply side resources, and therefore require some special consideration. But there is little discussion of the *actual* effects of such quantitative consideration measures in the literature, perhaps because the procedures appear so far removed from rigorous cost/benefit analysis that nobody has bothered to examine them systematically.

Monetization:

The third approach, namely monetization of environmental externalities and their subsequent incorporation into traditional cost-benefit evaluation, is the obvious preference of environmental economists, with its seeming emphasis on validity. Yet the approach is beset with practical difficulties, not least of which is the limited ability of economists to make realistic estimates of actual damage costs. Because of these difficulties, many have advocated the use of control costs as a proxy for damages, yet this too can readily be shown to violate the axioms of benefit/cost analysis¹⁰. However, the encyclopedic review of electric utility environmental damages by Ottinger *et al.* (1989) tends to suggest that the uncertainties that surround actual damage estimates lead to valuations that are probably no more uncertain than those associated with the use of control costs.

⁸ And as we shall see later, the subjective judgments involved in these techniques are very often derived by the PSAT method -- literally "people sitting around a table" -- in which group discussion attempts to reach a consensus about the severity of a particular impact.

⁹ For example, Vermont gives a 15% adder to DSM; Montana gives a 2% adder for DSM. Wisconsin gives a 15% credit to resources that are not combustion based. See Ottinger *et al.* (1990) for a review. All of such rules presently in force discriminate in each case just between two categories; those resources that qualify, and those resources that do not. Thus such rules have limited ability to select among marginally cleaner or dirtier resources (Temple, Barker and Sloane, 1991, p.IV-2).

¹⁰ This approach is based on the proposition of a "revealed preference", that Congress would indeed impose standards up to the point where, implicitly, the marginal cost equals the marginal benefit. But see, e.g., Joskow (1992) for a critique of this theory. Massachusetts has adopted this approach, despite a great deal of controversy and contentious debate in the literature about its validity. In a recent regulatory proceeding, thousands of pages of testimony were generated that reflected the diversity of views on what are appropriate valuations for externalities (Commonwealth of Massachusetts, Department of Public Utilities, Docket 91-131. November 10, 1992). Nevertheless, whatever the shortcomings, the DPU concluded: "...the Department is reaffirming its strong commitment to the concept of considering environmental externalities in resource decision-making. At times, policy-makers have to accept a 'second best' solution as a means of moving in a desired direction."

Perhaps a more important problem with the monetization approach is that only those environmental impacts for which a monetization calculation proves convenient are included. In practice this means that air emissions are so quantified, while the impact of other emissions, and different types of impacts, tend to be ignored.

A further issue is that not everyone accepts the premises of benefit cost analysis, namely (1) that only people's preferences matter (i.e. that other species or natural areas do not have rights); and (2) that the distribution of benefits and costs does not matter. Moreover, when different options have very different impacts (dollars, human health impacts, loss of biodiversity), many individuals prefer to consider each in their natural units, rather than obscure the trade-offs by putting them into common units (particularly when there may be sharp disagreements about exactly how such a conversion is undertaken in practice).

Multi-criteria decision methods:

Finally there are the multicriteria methods¹¹. Because of the problems inherent in the explicit valuation of externalities, multicriteria methods have been used for some time -- indeed, they have been part of environmental impact assessments and power plant siting studies since the early 1970s (as a result of NEPA and the Calvert Cliffs decision) (Hobbs and Rowe, 1979)¹². More recently, they have come into widespread use in bidding systems used by utilities to acquire resources from independent power producers (IPPs)¹³. Typically in such systems alternative options are scored (or ranked) on some scale for a number of environmental and cost attributes, the attributes are given weights, and an overall ranking derived by adding the weighted scores. Multi-criteria methods have at least five purposes:

1. To communicate information about tradeoffs among attributes (Cohon, 1978). An example of a tradeoff between two attributes is: "in order to decrease the emissions of SO₂ by 10,000 tons, we would have to accept an increase in cost of \$2,000,000."
2. To help decision makers make more consistent and rational evaluations of these tradeoffs. When alternatives have many attributes (ten or more are common in utility planning; e.g., cost, land use, endangered species, flexibility, reliability, water quality, emissions of CO₂, NO_x, particulates, SO₂ ...), psychological research has shown that decision makers are inconsistent in their subjective evaluations of alternatives. Generally, the mind focuses on two or three attributes and ignores the others, or it flits inconsistently among the attributes (Shepard,

¹¹ The EPRI guide, *op. cit.*, calls this "rating and weighting" approaches: but in fact rating and weighting is only one of a much larger set of techniques that can broadly be defined as MCDM methods.

¹² MCDM methods are widely used outside the energy field to quantify environmental implications of decisions (Hobbs, 1985). For instance USEPA uses a rating and weighting approach to prioritize superfund sites (Caldwell and Ortiz, 1989). MCDM methods have also been used extensively for water resource planning by the U.S. Bureau of Reclamation (Brown, 1984) and U.S. Army Corps of Engineers (1975). Even though the "Principles and Guidelines" for federal water planning (U.S. Water Resources Council, 1983) no longer endorses the co-equal objective approach of the old "Principles and Standards" (U.S. Water Resources Council, 1973), multicriteria methods still can and do serve important roles in federal water resources planning (Stakhiv, 1986; Hobbs, Stakhiv, and Grayman, 1989).

¹³ See, e.g., Kahn, *et al.* (1989).

1964; Slovic and Lichtenstein, 1971). Multi-criteria methods ask questions of decision makers to determine their priorities, and use the answers to systematically weed out undesirable alternatives and to indicate what alternatives might be most preferred.

3. To help decision makers make more consistent and rational evaluations of risk and uncertainty. Psychological studies show that decision makers have a difficult time evaluating risk in a consistent manner (e.g., Krzysztofowicz and Duckstein 1980). Some multicriteria risk methods attempt to measure decision makers' attitudes towards risk (e.g., risk seeking, risk averse) and to consistently apply them to the problem. Other methods for initial screening of plans attempt to preserve a range of choice so that if the detailed engineering phase finds one alternative to be infeasible, other feasible plans will be still be available (Keeney, 1987).
4. To help understand the values and perspectives of different interests. Multicriteria methods provide a framework for measuring the priorities of different groups, and for showing what the implications of each group's values would be for the decision (Brown, 1984).
5. To document how decisions are made. The use of multicriteria methods in federal decision making was spurred by the National Environmental Policy Act and the subsequent court rulings (Calvert Cliffs) which required agencies to document the rationale for decisions.

There is no question that some MCDM methods are practical to apply, but whether they are also valid has come under increasing scrutiny. For example, Hirst and Goldman (1991), in discussing the scoring system used by the New England Electric System, conclude "...the method is easy to understand, but its lack of scientific basis is troubling. Therefore this approach may be useful primarily as an interim method."

A final point to emphasize at the outset is that very often people do *not* know what they want: and it is therefore naive to view MCDM as simply a problem of eliciting preferences firmly established in someone's head -- as if these were eternal, constant and well defined. In reality, during the course of a planning exercise, people's attitudes will evolve in response to new information, interactions with others, and viewing the problem from different perspectives.

Indeed, when a decision is difficult because it involves a unique problem along with strongly held and conflicting values, most people will be unsure of their priorities. There is ample evidence that in such situations, the value judgments they articulate will depend strongly on supposedly irrelevant details of methodology, such as the exact phrasing of questions (von Winterfelt and Edwards, 1986; Fischhoff *et al.*, 1979). In this case, a critical role of MCDM is to help people understand the problem better.

1.2 Objectives

This report addresses only the last of these four approaches. It is intended as a guide to the application of multi-criteria decision methods to the incorporation of environmental factors in electric utility integrated resource planning.

We provide a step-by-step guide to the elements that are common to all MCDM applications. We discuss (1) how environmental attributes should be selected and defined; (2) how options should be selected (and how risk and uncertainty should be accounted for); (3) how environmental impacts should be quantified (with particular attention to the problems of

location); (4) how screening should be conducted; (5) the construction and analysis of trade-off curves; (6) dominance analysis, which seeks to identify clearly superior options, and reject clearly inferior options; (7) scaling of impacts, in which we translate social, economic and environmental impacts into value functions; (8) the determination of weights, with particular emphasis on ensuring that the weights reflect the trade-offs that decision-makers are actually willing to make; (9) the amalgamation of attributes into overall plan rankings; and (10) the resolution of differences among methods, and between individuals. Throughout the discussion we stress validity, appropriateness, and practicality, recognizing that a good method necessarily draws a balance among these three criteria.

The primary question addressed by this report is therefore not *whether* it should be done, but *how* it should be done. Further, we limit ourselves to the *how* of MCDM approaches: we do not attempt the much broader task of addressing environmental externalities in general.¹⁴

1.3 Outline of the report

We begin, in Section 2, with a review of the elements of an MCDM application, and present a discussion of the salient features of the methods that are currently being used in IRP. The discussion is structured around the ten elements of MCDM.

Section 3 then reviews a selection of past applications, in which we use the step-by-step guide to examine how they satisfy the three main criteria of appropriateness, validity, and practicality. Case material is drawn from a wide field of utility applications, ranging from project level environmental impacts statements to capacity bidding programs, and from the results of two case studies conducted as part of this research project. We worked with two utilities: Seattle City Light and Centerior Energy of Ohio -- with a primary focus on whether different methods yielded different results, and how utility planners judged appropriateness and ease of use.

In Section 4 we discuss the criteria for selecting an MCDM technique, given the very large number of methods that are encountered in the literature and in practical applications. In addition to the already-mentioned criteria of practicality and validity, we also examine the extent to which methods are appropriate: that is, does the method provide the information that is actually needed by decision-makers?

¹⁴ There are several reviews that already fill this wider function, ranging from the elementary exposition in the EPRI Guide on Environmental Externalities (Temple, Barker, Sloane, 1991), to the treatise on monetization of environmental impacts prepared for the State of Washington Interagency Task Force on Environmental Costs (Dodds and Lessor, 1992) and the encyclopedic review in the Pace University study (Environmental Costs of Electricity, Ottinger *et al.*, 1990). Perhaps the best recent review of economic valuation techniques (albeit not limited to electric sector impacts), that emphasizes practical guidance and examples rather than theory, is the World Bank guide by Munasinghe (1993).

2. THE APPLICATION OF MCDM METHODS

In this Chapter we present a discussion of the sequence of steps necessary to apply an MCDM method to a decision problem. These steps are largely independent of the particular application involved, and apply not just to the incorporation of environmental factors in IRP but to a whole host of other potential utility applications from power plant siting to bidding programs. We begin with an overview of the process, and then discuss each of the steps in turn.

The application of MCDM methods involves the following steps:

1. Selection and definition of attributes, say A_i , $i=1, \dots, N$, selected to reflect important planning objectives and/or environmental concerns. System cost, reliability, impact on rates, air quality impacts, or impact on fisheries are examples: in this step we select which of these will be used in an application, and precisely how they should be defined.
2. Definition of the alternatives to be analyzed.
3. Quantification of the levels A_{ij} of the i attributes estimated for each of the j alternatives. This generally requires the application of some model to predict the impacts. Uncertainty and risk in attribute levels is quantified at this time.
4. Preliminary screening of alternatives
5. Construction and analysis of trade-off curves
6. Dominance analysis
7. Scaling of attributes, in which the level of an attribute is translated into a measure of value, $v_i(A_{ij})$ (also known as the attribute value function). This is sometimes combined with a normalization procedure, usually on a scale of zero to one (in which the lowest value of the attribute value function is assigned zero, the highest one).
8. Selection of weights w_i for each attribute.
9. Determination and application of an amalgamation rule. Such rules combine the weights and value functions into a single overall value or ranking of the available options, or which reduces the number of options for further consideration to a smaller number of candidate plans.
10. Resolution of differences between methods, and between and among individuals.

While the steps appear conceptually straight forward, in general most applications give relatively little attention to a number of theoretical requirements for the results to be valid. Equally important, even though we have presented these elements in the form of a sequential list, in any actual application a certain amount of iteration between these steps will be necessary -- as we shall see the subsequent discussions. The inter-relationship between these steps is indicated on Figure 2.1.

Figure 2.1: Steps in a MCDM problem

2.1: Attribute selection and definition (step 1)

Attributes, or measures of goodness, express the planning objectives to be considered in a given problem. Unfortunately, the need for care in the selection of attributes is not very evident from much of the literature of application of multi-attribute methods to power sector problems. Four main issues need to be addressed in the selection and definition of attributes:

Double counting

There are several types of double counting that need to be considered in a MCDM problem. The first relates to double counting in the attributes: for example, there ought not to be a separate attribute for the miles of associated transmission line if there are already other attributes that capture the relevant land use impacts. Provided one is always cognizant of precisely what impact a particular attribute is supposed to capture, this is not too difficult to avoid in practice.

Indeed, the attribute "land use" is one that is frequently encountered, often expressed as "acres", or "acres/Mw"¹⁵. Why this should be a separate attribute is generally unclear. The cost of land is (or should be) included in the system cost attribute, as should any costs associated with resettlement and compensation (often an issue in hydro projects). The opportunity costs of lost production are also easily calculated (say in the case of land in agricultural use). Therefore, it is only the *non-market* values of land that need to be captured. Yet if there are unique historical, ecological or archaeological resources that need to be explicitly considered, then they are surely not captured by a value for acres per Mw.

A more difficult double counting issue concerns the extent to which those pollutant emissions that are subject to tradable emissions allowances -- such as SO₂ -- ought also to be treated as an attribute (or, in monetization approaches, also assigned some externality value).¹⁶ The practical question is whether or not there should be an attribute for SO₂ emissions in an MCDM analysis, given (1) that the total quantity of SO₂ emissions in the US is fixed under the emissions allowance system established under the 1990 Clean Air Amendments,¹⁷ and (2) that the cost of purchasing the necessary allowances is already included in the cost attribute. At first glance the answer to that question might seem to be negative.

However, the argument that the allowances system has internalized the cost of SO₂ emissions is valid only as long as it does not matter where and when the emissions takes place. But the damages SO₂ emissions inflict do depend strongly on location. Emissions upwind of urban areas and sensitive ecosystems, such as New York's Adirondack lakes, are more harmful

¹⁵ For example, acres/Mw appears in the rating system of Niagara Mohawk in New York, reviewed below in Section 3.1

¹⁶ NO_x emissions in non-attainment areas are subject to "offset" provisions that can have much the same effect.

¹⁷ The SO₂ control program is implemented in two stages, Phase I and Phase II. Phase I will run from 1995 to 1999. During this time, allowances will only be issued to the 110 so-called Phase I plants which the title enumerates. These are generally larger coal-fired facilities whose emission rates exceed 2.5 lb/mmBtu. Allowances will be allocated to each at approximately that rate times the 1985-87 average heat input (in mmBtu) of the unit. The SO₂ emissions of other sulfur-emitting generating units, called Phase II units, will not require allowances at that time. In Phase II, starting in the year 2000, all generating units will come under the allowance system and, at most, 1.2 lb/mmBtu of allowances will be allocated to each. The total allowances issued nationwide during Phase II will be 8.95 million tons/year, approximately half of the emissions occurring in the 1980's.

than releases in areas with predominantly alkaline soils or few people downwind, such as Florida and Nevada.¹⁸

Elsewhere (Hobbs, 1992, 1993), it is shown that the social cost of an additional ton of SO₂ emissions at a particular location X is the sum of two quantities: (1) the price of an allowance, plus (2) the external cost resulting from the emissions at X, minus the external costs avoided because emissions at other locations will decrease. That is, the social cost accounts for how an emissions allowance would otherwise be used, and the resulting damage if it were not consumed at X. Since allowances are an internal cost anyway -- utilities must either buy allowances or forego selling them -- the appropriate adder is just the second of the two quantities. This adder might in fact be negative.¹⁹

In practice, therefore, whether or not there should be an attribute for SO₂ emissions would depend upon whether a case can be made that there is a differential locational effect -- *in addition to* the cost of allowances that should have been included in the system cost attribute. If such a case can be made, and SO₂ emissions are included as an attribute, it is likely that participants should be repeatedly reminded, particularly during the scaling and weighting elements of the MCDM process, that the attributes refers to only the *differential* impacts.

Value independence:

Attributes should be conceptually distinct, i.e. preference independent²⁰. For example, the attributes CO₂ emissions and SO₂ emissions would be preference independent if one were willing to spend \$15/ton to avoid a ton of CO₂ emissions regardless of whether SO₂ emissions were low or were high. Decision analysis differentiates between statistical independence and preferential independence; the former refers to the correlation structure of the alternatives, the latter to structure of the user's preferences -- a distinction that might be characterized as "facts" v. "values".

Additive value functions -- such as those used in weighting and summation methods -- assume certain type of *value* independence, but statistical independence is irrelevant to the validity of the additive form. For example, whether alternatives with high CO₂ emissions also tend to have high SO₂ emissions is a question of statistical, not preference, independence.

A good illustration of the importance of value independence is illustrated by the poor criteria definition adopted by the New York State hazardous waste facility siting board, which established "danger of fire and explosion" and "population near site" as separate criteria to be

¹⁸ For example, as noted by Joskow (1992), the net damage of an additional ton of SO₂ emitted in Massachusetts might be negative if that ton might otherwise be emitted upwind in Ohio, as much of the emissions at, for example, Massachusetts's Brayton Point plant, are blown out to sea.

¹⁹ The social cost of an incremental ton of SO₂ under the Title IV allowance system equals the increase in control costs that one or more utilities bear in order to free up the one allowance needed to cover that ton, plus the net increase in environmental damage stemming from the rearrangement of emissions. Under a well-functioning allowance system, the former cost equals the price of allowances. The theoretically appropriate adder is then the latter quantity--which is generally neither zero nor the difference between the gross damage caused by a ton emitted within its jurisdiction and the allowance price. Whether the net external cost is positive or negative depends on two variables: the relative magnitudes of marginal damages of emissions in different locations, and who would buy or sell allowances on the margin. If marginal damages in all locations are equal, then the net cost is zero.

²⁰ A set of attributes $\{X_1, X_2, \dots, X_n\}$ is preference independent of its complement (i.e. the other attributes $\{X_{n+1}, X_{n+2}, \dots, X_N\}$) if the trade-offs a decision maker is willing to make among the $\{X_1, X_2, \dots, X_n\}$ do not depend upon the levels of its complement. (Keeney and Raiffa, 1976).

added together. (New York Department of Environmental Conservation, 1980). But these are not preferentially independent -- a high likelihood of explosion would motivate a user to put a higher weight on the criterion "population near site". One possible approach to avoiding this kind of problem is to define a new attribute as the product of the two.

Specificity

The rationale for selecting a particular attribute needs careful consideration. For example one frequently finds a criterion "solid waste disposal" expressed as tons of waste generated. Precisely why the *quantity* of solid waste is relevant is unclear, except that it is easy to estimate. Since a large part of the total land area required for a coal-fired power plant may be for solid waste disposal, there may in any event be double counting if there is also a criterion for "land use" (expressed in acres). Moreover, the *quantity* of solid waste is perhaps primarily related to the *cost* of disposal (higher quantities correlate very well with higher disposal costs): yet again, such costs themselves should be included in the system cost attribute.

In fact the relevant environmental risk associated with ash and scrubber sludge disposal is the risk of toxic leachates contaminating nearby aquifers, which has very little relationship to the quantity of waste produced, but rather to characteristics such as soil permeability, depth to the groundwater table, and rainfall.

The problem, of course, is that actual environmental risk may not be very well determined at the long range planning stage: even if the general locations are known, site studies are unlikely to be available in the sort of detail necessary. Yet choosing the most convenient proxy, or even one with seeming precision, such as volume of solid waste generated, may in fact not be the most appropriate. In this particular case, the *number* of solid waste sites that have to be opened may be much more suitable.²¹

Proliferation of attributes

Many applications of MCDM methods suffer from the same malady that afflict environmental impact statements: in their desire to be comprehensive, they become tedious recitations of all possible impacts, with not much thought given to what is really important. For example, in the scoring system used in testimony in a Vermont proceeding -- which attempted to deal with the question of the relative environmental impacts of hydro imports from Canada, demand-side management measures, and decentralized renewable -- 36 different attributes were used (Mintzer *et al.*, 1990). These included separate attributes for each of the following material uses: steel, concrete, aluminum, silicon, glass, plastics and non-ferrous metals.²² Exactly what environmental concerns the consumption of these materials implies, that is not already reflected by other attributes, is unclear.

²¹ Elsewhere we argue that arbitrary "scoring" of impacts can be particularly misleading, particularly if a more rigorous quantification through modeling is possible (see discussion below under step 3). However, in this particular case, we see the dangers of using *inappropriate* -- and indeed quite spurious -- quantification: using the number of new solid waste sites to be opened as the attribute, if properly justified, is a lot less arbitrary than tons of waste.

²² This study also proposed an interesting approach to the determination of weights, based on the *reversibility* of the impact as the most important factor: high level radioactive waste from a nuclear power plant would be assigned a high weight because it remains radioactive for thousands of years; while thermal pollution of a river during low flow periods would be assigned a low weight. However, what is critical to validity is whether or not the weights actually reflect the trade-offs that decision makers would actually be prepared to make in practice.

A proliferation of attributes tends to make weighting more difficult, and may introduce a bias simply because one is reluctant to weight any particular attribute as near zero. Moreover, trade-offs become difficult to understand and display to decision-makers in comprehensible form if there are too many of them. The guiding principle ought to be that one starts from what are the most important impact issues (greenhouse gas emissions, health effects of fossil fuel pollution, the risk of groundwater contamination, the risk posed to aquatic ecosystems from thermal plumes, etc), and then select *one* attribute for each of these concerns²³.

In some sense, obviously, the definition and quantification of attributes goes hand in hand: MCDM methods require that attributes be quantified. However, there is the danger that what is difficult to quantify -- yet important -- may get ignored.

Two ways can be used to deal with this issue. First, in some cases, indirect valuation methods can be used without direct quantification: for example, some of the so-called hedonic pricing methods can be used to estimate the impact of a power facility. By examining real estate prices in the vicinity of a large generating facility, and comparing them to a comparable location without a generating plant in its vicinity, one may assess how individuals value the perceived environmental impacts -- without necessarily having to quantify the actual impact such as incremental air pollution (at a fossil plant) or incremental radiation or probability of catastrophic accident (at a nuclear plant).

A second way of dealing with impacts difficult to quantify is to use scenarios. For example, the health impacts associated with EHV transmission lines may be viewed as significant²⁴. Yet there is presently no basis for quantifying that effect. One way to include an assessment of the issue is to define a technology scenario that eliminates the need for additional EHV lines (by dispersed siting, or reduced reliability), and then asking whether one is willing to incur the incremental costs involved.

²³ All these points have long been established in the decision sciences literature, and were validated in a large number of applications in the 1970's to power plant siting particularly by Keeney and his colleagues (see e.g. Keeney and Raiffa (1976); Keeney and Nair (1977). Keeney *et al.* (1981) is one illustration of a well-founded application of decision theory to utility resource planning (a coal/nuclear choice).

²⁴ For a good review of the issues and uncertainties involved, see, e.g., Florig (1992).

2.2: Define options (Step 2).

Decision-makers like to be presented with some reasonable number of options that they can weigh against each other. What is reasonable in any set of circumstances may of course vary, but 20 may be too many, and two is likely to be too few. But whatever the number, the most important criterion is that the set selected show meaningful differences in resource mix choices and impacts. An analysis that looks, say, only at five different nuclear plant choices and a single level of DSM program implementation will satisfy nobody. At the same time, at the outset of an IRP process, one may not know which alternatives are the interesting ones to examine. In short, the definition of the options is likely to be part of an iterative process, rather than something determined irrevocably at the outset.

This problem is well illustrated by the Centerior case study (Meier and Hobbs, 1994). The possible combinations of different levels of DSM, unit career planning, dispatching method, and new generation and purchase options ran into several hundreds. This had to be reduced down to some manageable number. The question is how many, and by what criterion is the reduction process to take place?

Moreover, these plans had to be evaluated in light of many different uncertainties. A number of combinations of external circumstances beyond Centerior's control -- the purchase price of electricity from neighboring utilities, nuclear availability, DSM effectiveness, natural gas prices, to name a few -- all complicate the process.

This of course describes the classical problem of decision-making under uncertainty, for which decision trees make an excellent representation.²⁵ But with many chance nodes (representing the outcomes of the uncertain variables) and many decision nodes (which represent the alternative actions that can be taken), such trees often become very complex ("bushy") which may make them difficult for management and decision-makers to absorb.

Centerior Energy uses a somewhat different approach for its IRP, in which the various elements of uncertainty are combined into "futures". With management concern strongly focused on rate impacts, the futures defined for our case study represented the extremes that would combine into very favorable, or very unfavorable circumstances, for the evolution of rates. Thus, as shown on Table 2.1, Centerior defined "low stress", "reference", and "high stress" futures. These cases correspond to the likelihood that the utility would require rate relief, which in Centerior's case is the dominant concern of management (and of its relationship with the regulatory commission).

The process of defining such futures can stimulate thinking about what options might "work" in different futures, or be robust across these futures. This can lead to the addition of more plans to the initial list.

There is probably some danger of leaving option selection entirely to utility technical analysts: if the analysis is to be representative of the issues likely to come before the regulatory proceeding, it is better that options of interest to likely intervenor groups be included in the analysis, however unpalatable that may be, or however improbable that be viewed, by the utility resource planners. Thus we cannot stress too much the desirability of as broad, and as early a public participation in this process as possible -- a point that is hardly new.

²⁵ There is a growing literature on utility decision-making under uncertainty that is far too extensive to be reviewed here in any detail. Practical examples of such techniques include Hirst and Schweizer (1989); Merrill and Wood (1991); Hobbs and Maheshwari (1990); and Keeney and Sicherman (1986).

Table 2.1: Definition of futures: Centerior Energy Case Study

	low stress	reference	high stress
purchase power, \$/mwh	21	27	33
allowance price, \$/ton SO ₂	444	278	222
CO ₂ tax, \$/ton CO ₂	0	0	22
nuclear availability, %	80	80	60
DSM effectiveness, %	100	100	50
load growth, %/year	3	1.5	0
gas prices, cents/mmBtu	178	355	533

In any event, there also needs to be broad participation among different utility departments at this stage, with at a minimum representation from the environment, rate and DSM departments in addition to those directly involved in the IRP process and the preparation of the regular cycle of submissions to the regulatory commission.

As an example of the results of such an option definition process, in Figure 2.2 we show those branches of the total decision tree that represent the set of 15 options used in the Centerior case study (Reference, plus cases A through N, indicated on the far right hand side of this figure). The options were chosen to be representative of the diversity of resources and economic and environmental outcomes. The specific assumptions for these plan options are summarized on Table 2.2.

A different approach is represented by our Seattle case study. There, a mathematical programming-based IRP model was used to generate a range of resource mixes. This was accomplished by varying the relative weight of CO₂ and utility cost in the model's objective function, and by adding and deleting certain options. For instance, in one run a CO₂ penalty of \$20/ton was applied while new hydroelectric capacity was prohibited in response to fish concerns. The result of this process was 36 tentative resource plans which were then narrowed down to 12 using some of the methods described under step 6, below.

Figure 2.2: The Centerior case study decision tree

Table 2.2: Plan options in the Centerior Case Study

ref		DSM [Mw]	UCP [units]	New capacity [Type]	Percent reserve [%]	dispatching method
ref	reference	360	11	CT/CC	20	economic
a	reference with emissions dispatch	360	11	CT/CC	20	emissions
b	reference with 10% reserve	360	11	CT/CC	10	economic
c	ref. with wind as new capacity	360	11	200mW wind	20	economic
d	ref. with coal as new capacity	360	11	600mW coal	20	economic
e	ref. with Summit as new capacity	360	11	200mW Summit	20	economic
f	reduced UCP	360	7(-720)	CT/CC	20	economic
g	Reduced UCP with emissions dispatch	360	7(-720)	CT/CC	20	emissions
h	Reduced UCP with 500Mw purchase	360	7(-720)	500Mw purchase	20	economic
i	Reduced UCP with new nuclear	360	7(-720)	600Mw nuclear	20	economic
j	increased DSM	720	11	CT/CC	20	economic
k	increased DSM, reduced UCP+ emissions dispatch	720	7(-720)	CT/CC	20	emissions
l	increased DSM, reduced UCP+ wind and emissions dispatch	720	7(-720)	200Mw wind	20	emissions
m	load building	+360	11	CT/CC	20	economic
n	load building+reduced UCP	+360	7(-720)	CT/CC	20	economic

2.3: Quantify impacts (step 3)

How the impacts of the different alternatives are determined is extremely important to the ultimate validity of the method. In this report we stress the use of rigorous mathematical models to predict impacts on each of the attributes: indeed, impacts need to be quantified for each alternative under each future, and often for every year within the planning horizon. One should note, however, that in a number of applications -- particularly in power plant siting -- there has been a tendency to use "expert judgment" to score alternatives. In such procedures, groups of experts directly "score" different alternatives on some scale: this represents a combination of steps 3 and 4. As we shall see in the next chapter where we review applications of MCDM to siting, such short cuts lead to very questionable results.²⁶ In our view, whatever the application, there is no valid alternative to a rigorous calculation and quantification of impacts by application of an appropriate model. Even if all models are subject to certain assumptions and limitations, these can always be spelled out very clearly, so that decision makers clearly understand how results were derived.

Models are in wide use by electric utilities for planning, and can be used to estimate total system costs and emissions. Traditionally their principal focus has been to determine the economic and financial consequences of alternative resource plans. Some are optimization models, that seek the least cost capacity expansion path subject to exogenously specified demand forecasts and reliability levels (such as EGEAS). Others are simulation models, in which cost, reliability, and rate impacts are calculated for a series of specific options specified by the user (such as MIDAS). Economic cost (specified as the present value of system costs), reliability (specified as loss of load probability or the present value of unserved energy), rate impact (specified as average cost per kwh), and capital requirements are attributes that will likely be viewed as part of every MCDM application to IRP, and for which these kinds of models are well suited.

Perhaps the most important criterion for selecting a techno-economic model is not theoretical rigor by some abstract academic standard, but the degree to which the analysts and decision-makers have confidence in model results. A model that has been in use for some time for resource planning is likely to be the best choice, for almost certainly the decision-makers will ask what confidence the analyst has in being able to predict the impacts of an option. The discussions are likely to be contentious enough without model validity becoming an additional issue.

Thus, in our two case studies, we used models that had been in use for many years: MIDAS in the case of Centerior, and the ANN model in the case of Seattle City Light. A number of possible shortcomings might have been cited in both models, but everyone seemed confident about their use, and their ability to calculate to within reasonable accuracy the trade-offs that were the principal focus of the exercise. In both cases the participants were confident that the models themselves would not be questioned by regulatory entity (the City Council in SCL's case) that would review the IRP.

²⁶ For instance, in the power plant siting example of Section 3.8, a comparison was made between judgmental scores for air quality impacts given by "experts" on a scale of zero to 30, and actual computation of air quality impacts using a dispersion model to estimate population exposure. The model calculations showed variations between the best and the worst site by a factor of 11, whereas the judgment scores showed a variation between the best and worst of less than 3.

Environmental impacts

Much more difficult is the quantification of the environmental *impacts*. Most of these techno-economic models have been augmented with the ability to calculate pollutant emissions, and to deal with environmental regulations that apply to the quantity of pollutants that may be discharged into the environment. Thus emissions of criteria pollutants, greenhouse gases, land use, and quantities of solid wastes are routinely compiled by such models.

However, the extent to which emissions address regulatory convenience rather than actual impacts is an important question. For some residuals, emissions may be much more appropriate than others as rational proxies for actual impacts. For example, we have already noted that tons of solid waste, easily calculated, is probably a rather poor indicator of the extent to which solid waste disposal actually poses an environmental risk. On the other hand, emissions of greenhouse gases is an appropriate proxy for the incremental contribution to global warming.

In the Centerior case study there was extensive discussion about solid waste, and how that concern would be captured in the process. Initially, almost as a matter of routine, estimates of tons of solid waste were generated by the MIDAS model. But after discussion with the participants, the shortcomings of such a quantification became clear. What was deemed to be a much better proxy for the actual environmental impacts, as well as utility's own transactions costs (the "hassle factor"), was the number of new solid waste disposal sites that had to be opened.²⁷

Environmental impacts and location:

That environmental impacts are a function of location is a generally recognized fact that has few exceptions²⁸. While emissions matter a great deal from the regulatory perspective, from the perspective of actual impacts it is not emissions that matter, but their fate in the environment. Two things follow from these observations: (1) that it matters a great deal where emissions occur, and (2) that any quantification and/or monetization of impacts are also likely to be dependent upon location. Most monetary valuations of air emissions as adopted by state regulatory commissions thus far, however, are independent of location: as noted earlier, it is entirely unclear why, if the regulatory objective is to minimize impacts, a ton of NO_x emitted from a combustion turbine with a relatively low stack in Boston should be subject to the same externality value as a ton emitted from a plant in the less populated areas of central Massachusetts. The values used for air emissions in some states are shown on Table 2.3.

Table 2.3: Monetary valuation of air emissions: \$/ton

	CO ₂	NO _x	SO ₂	particulates
New York	1.1	1,832	832	333
Massachusetts	22	720	1,700	4,400
Nevada	22	6,800	1,560	4,180
California (PG&E)	26	7,105	4,060	2,380

²⁷ Recall also the earlier discussion, under step 1, of specificity in attribute definition

²⁸ There are perhaps some pollutants -- such as plutonium -- that are so highly toxic that it is reasonable to assert that *any* emission at *any* location must be controlled. But such arguments would be difficult to invoke for most of the pollutants of concern to electric utility resource planning. Even in the case of hydro plants, where one might accept the premise that *any* hydro plant that eliminates whitewater recreation opportunities, or interferes with salmon migration, is unacceptable, there are certainly *some* locations where this would not be the case.

There are also situations where the exact opposite is true, most notably in the case of greenhouse gases: in this case it matters not one bit whether a ton of CO₂ is emitted in Manhattan, or rural upstate New York, or in Florida or Colorado. The incremental impact on global warming will be the same. Yet different regulatory commissions have chosen to value CO₂ emissions at very different levels (e.g. in Massachusetts at 2.7 \$/ton, in New York at 1.1 \$/ton).

To be sure, the distribution of the impacts of global warming will be very unevenly distributed: For example, Boston and Manhattan will be much more susceptible to the impacts of sea-level rise than Colorado. Yet even if the impacts of global warming were assessed by the Massachusetts PUC as more severe than elsewhere in the country, does it follow that imposing a higher externality value on greenhouse gas emissions in Massachusetts than elsewhere is rational? The quantification of environmental impacts requires the following sequence of steps (we use here the example of an air pollutant, but the framework is equally applicable, albeit with some modification, to other types of impacts):

1. *The timing and quantity of emissions.* This can generally be estimated fairly well with the kinds of models in general use by utilities. Models such as MIDAS or EGEAS can be run in a number of alternative dispatching modes to the traditional least cost dispatch, including dispatch to minimize specific emissions.
2. *The location of emissions.* At the IRP stage there may be high uncertainty about where particular new power plants may be located.
3. *The fate of pollutants in the environment.* For most of the important pollutants, enough is known about atmospheric transport to be able to make fairly good predictions of the changes in ambient concentrations resulting from a given quantum of emissions at a given location.²⁹ Such models are routinely used in Environmental Impact Statements. Modeling the fate of pollutants in surface waters can also be said to be relatively free of substantial uncertainty: and again many water quality models for rivers, lakes and coastal waters are generally accepted by the scientific community as valid. More difficult, and more subject to uncertainty, is modeling the fate of pollutants in the subsurface environment.
4. *Exposure of sensitive populations.* To calculate the aggregate dose requires knowledge of how large the affected population will be. Projecting population for local areas is notoriously difficult, and subject to large uncertainties. Equally important, an existing projection may be affected by the facility itself, for which the classic example is nuclear power plants: because of the uneven distribution of the property tax benefits, in some areas of the country nuclear power plants have attracted additional population far beyond the initial projections.
5. *Estimation of the dose-response function:* Epidemiological studies to estimate dose response functions are subject to great uncertainty because of the need to control for the large number of other factors that affect morbidity and mortality. Indeed, in regulatory proceedings, this is likely to be one of the most contentious issues: and it is quite unclear that Commissions and their staffs have the ability to resolve conflicting testimony presented by parties and their witnesses.

²⁹ However, this may be quite difficult for heavy metals and toxic hydrocarbons whose fate in the environment is complex.

6. *Valuation of the response:* Perhaps most controversial of all is the final step in this sequence, namely a monetary valuation of the impact of the affected populations to the dose in question. The most difficult of all is the economic valuation of human life, subject to such large uncertainties that use of any particular value is likely to be so controversial that attention is focused on validity of the valuation, rather than on the trade-offs that must be made. One of the advantages of MCDM methods, of course, is that this final step need not be taken.

The central difficulty in this sequence is that none of steps 3 to 6 can be undertaken unless location (step 2) is known. For this reason, while most project level environmental impact statements progress generally to step 3, to date at the IRP stage, one has been limited to the first step.³⁰

One approach to the problem of unknown location at the long range planning stage is the use of "surrogate sites". This was pioneered by the nuclear energy center studies conducted by the NRC in the mid 1970s (NRC, 1977). Charged by Congress with assessing the feasibility of clustering nuclear power plants into "energy centers", it was quickly recognized that in order to assess their environmental impacts, one would have to hypothesize particular sites. Unfortunately, in this specific instance of nuclear energy centers, residents near the selected sites had a hard time accepting the notion that one might study a site without actual plans to proceed to construction.

Advocates of monetary valuation argue that while it may well be true that location is indeed important, in the absence of practical ways to determine what differences in impact can be attributed to different locations, applying an average value to all locations is better than doing nothing at all; or, alternatively, applying the worst case value to all locations is better than making exceptions or imposing lower values at some other locations in the face of very imperfect information about the spatial variation in impacts.

³⁰ Critics of the monetization approach argue that going straight to step 6, without having done detailed studies at steps 4 and 5, is quite arbitrary.

Another way of expressing this is to note that such conclusions in fact tell us something about the risk preferences of utility regulators; and that in turn leads us to a decision analysis framework for examining these issues more scientifically. Rather than always assume the worst case, in the expectation that more detailed study, or resolving scientific uncertainties, takes too much time and money, the proper question is whether in fact developing additional information has value. Decision analysis, and its concept of the value of perfect information, provides just such a framework.³¹

Detailed environmental models -- particularly for air and water quality -- are of course routinely applied at the project level environmental impact stage to translate emissions into estimates of changes of ambient concentrations. In other words, Steps 1-5 in the above list can readily be determined at the project level. But at the long range planning stage, the necessary site specific data may simply not be available.

³¹ A good example of such an approach is the application of decision analysis to the identification of the optimal hedging strategy for the global climate change problem by Manne and Richels (1992). The paradigm they use is that of a portfolio of insurance options: what combination of insurance should be bought, if indeed any at all? "...what portion goes to R&D to resolve scientific uncertainties? What portion goes to the development of new supply and conservation technologies to reduce abatement costs? And what portion goes to immediate abatement of emissions? Particularly they focus on the value of information: how much accuracy is needed in climate modeling and impact assessment? Clearly, with *perfect information*, the best course of action can be charted immediately, and there is no need to hedge bets. Manne and Richels conclude that the need for precautionary near-term emissions reductions is inversely related to the sustained commitment to R&D to develop better climate information (which reduces the need to hedge against an uncertain and potentially hostile future).

2.4: Construction and analysis of trade-off curves (Step 4)

Any two attributes can readily be displayed in two-dimensional space: Figure 2.3, for example, shows rate impacts versus CO₂ emissions in the Centerior case study. If plan A is better than B in both cost and CO₂ emissions, then plan B is said to be dominated.

The individual points on this curve (a, b, c, etc) represent the plan options identified on Table 2.2 and Figure 2.2. The black line on this figure is the so-called trade-off curve, or non-inferior set: it represents the set of non-dominated solutions (see below, Step 6 for details). This is the set of options "closest" to the ideal (zero cost, zero emissions) at the origin.

Figure 2.3: Trade-off curve, rate impact v. CO₂ emissions (Centerior case study).

In Figure 2.3, plan c (which emphasizes wind power) has both higher emissions and higher electric rates than plan n (which eliminates some life extension of existing coal plants, and promotes load building).³² The trade-off curve connects those points which remain after all dominated options (such as k) are eliminated.

³² Of course there are more than two attributes of concern in the Centerior study, and plan c may have other advantages compared to n.

2.5: Screening (Step 5)

In exclusionary screening, one or more attributes are used to screen plans by dividing the range of each into two categories: "acceptable" and "unacceptable". Then, any alternative that is unacceptable in one or more attributes is dropped. Failure in one attribute is assumed to be as bad as failure in all, which is appropriate in the case of legal requirements (the most common), or where there exist unambiguous physical thresholds (less common, because of scientific uncertainty as to what the precise threshold values might be).

In the example below (Figure 2.4) the area to the right of the cutoff point for attribute A_1 , and above the cutoff point for attribute A_2 , are unacceptable: this leaves as the acceptable domain the area $XYZO$. The acceptable solutions will then lie to the northeast of the trade-off curve, in the pie-shaped slice WYV .

Figure 2.4: Exclusionary screening

For some attributes, there exist natural dividing points between "acceptable" and "unacceptable," often due to legal restrictions. Examples might include presence of endangered species (which might preclude development) and regulatory criteria imposed by the Nuclear Regulatory Commission (NRC) in the case of nuclear power plant siting (the only acceptable sites being those that strictly conform to the regulatory requirements): here failure in one attribute is indeed as bad as failure in all.

However, analysts are often tempted to treat attributes which lack such natural cutoff points (so called "discretionary attributes") as exclusionary. For instance, in a power plant siting study it might be decided a site must lie within 10 miles of a river. However, such a cutoff is not a natural one -- why is 9.5 miles acceptable but not 10.5 miles? The danger is that a plan might fail such a criterion by a small amount, yet be so good in other attributes that it would be found best in a later decision stage. More importantly, the violation of a particular constraint can often be mitigated through other added features of a selected plan without significantly affecting the benefit-cost ratio or its social desirability. In these cases,

exclusionary screening makes tradeoffs that decision makers would not accept if they could consider them explicitly.

Figure 2.5. Ratios of weights implied by cutoffs in exclusionary screening

This notion of implied tradeoffs is an important one. Consider Figure 2.5. The axes of the graph represent levels of two different discretionary attributes, both of which are to be minimized. From the point of view of exclusionary screening, plans C and B are "better" than the rest because only they are acceptable in both attributes. Tradeoffs are implied by these cutoffs, as A and D are each better in one attribute than C or B. By noting the slope of the line segments connecting acceptable and unacceptable alternatives, we can infer what weights are implied for the two attributes. First, the cutoff for attribute 1 implies that C is preferred to D, which means that the decision maker is (implicitly) willing to give up more than one unit of attribute 2 to gain a unit increase in attribute 1; that is, the weight of attribute 2 (w_2) is implied to be less than w_1 .³³ Second, the cutoff for attribute 2 implies that B is preferred to A. But if the decision maker's true weights were outside of those ranges, either A or D should actually be preferred to C and B.

Hence, contrary to the assertions of some researchers, attributes are not "weighted" equally in exclusionary screening. A change in the cutoffs can change the implied weights. Consequently, if exclusionary screening is to be used, the exclusionary cutoffs should be chosen carefully and conservatively. One should avoid using discretionary attributes to exclude plans, because the result may be to prematurely drop plans that might actually be preferred by some interests. If discretionary attributes must be used in exclusionary screening, then a sensitivity analysis should be undertaken to ascertain the effect of different cutoffs.

³³ As we discuss in Section 2.8, weights should represent the willingness of decision-makers to trade-off attributes.

2.6: Dominance analysis (Step 6)

In dominance analysis, an alternative is screened out if it is dominated by another option. An alternative is dominated if there exists another plan that is just as good in all attributes, and strictly better in at least one.

Dropping dominated alternatives is logical because a valid multicriteria method will never choose a dominated alternative over the option it is dominated by. However, the analyst should recognize that if there are other attributes that decision makers have in mind in addition to those in the model, it is possible that a dominated alternative may actually be preferred. Another consideration is that sensitivity analysis of parameters (such as interest rate, fuel prices, or hydrologic probabilities) may reveal that an apparently dominated alternative is advantageous under some conditions.

Another advantage of dominance analysis is that no value judgments are needed, other than whether each attribute is to be minimized or maximized. A disadvantage is that relatively few alternatives may be dominated. In that case, value judgments and tradeoffs will have to be made, perhaps with the aid of one of the other methods discussed in later sections.

The quadrant criterion:

A first concept is the so-called *quadrant criterion* (see Figure 2.6), which involves the comparison of selected options with respect to some base case, located at the origin. Assuming that less of each attribute is better, plans that lie in the quadrant III are *better* than the basecase on both criteria, while those in quadrant I are *worse* than the basecase on both criteria. In these cases there are no trade-offs to be made, because both interests (i.e. both economic efficiency and environment) are served in *adopting* options in quadrant III, and both are served by *avoiding* options in quadrant I. It is in quadrants II and IV where trade-offs must be made -- where an improvement in one objective necessarily implies a worsening in the other.³⁴

Figure 2.6: The quadrant criterion

³⁴ In the new lexicon of environmental advocacy, plans in the IIIrd quadrants are the "win -

Strict v. significant dominance

Because of the uncertainty that surrounds quantification, there is a need to distinguish between *strict* dominance, and *significant* dominance. To illustrate the difference between these two definitions, let the universe of initial options be denoted P_1, P_2 , etc. Suppose for the sake of clarity that there are only two attributes: cost, and an environmental attribute reflecting, for example, the population exposure to SO_2 . Figure 2.7 depicts the solution space for this problem, in which we plot the values of the two attributes for each plan.

The plan P_1 is said to *strictly dominate* plan P_2 if P_1 is better than (or equal to) -- i.e. dominates -- P_2 in terms of every attribute, and is strictly better in at least one attribute. Thus, as illustrated on Figure 2.6, P_1 is better than P_2 in both cost and SO_2 exposure. In fact, P_1 strictly dominates all of the plans northeast of the boundary AP_1B . By repeating such comparisons for all pairs of plans, and discarding all dominated plans, the remaining plans constitute the set of the so-called *non-inferior solutions* -- in our case the set of plans $\{P_6, P_4, P_1, P_{14}, P_{13}, P_{15}\}$. These points in turn define a *trade-off* or *non-inferior curve*, as indicated on Figure 2.6.

This procedure will in general provide a means for reducing a very large number of possible plans to some smaller number of plans -- a "short-list" or "candidate list" that is to be presented to decision-makers. In the example shown, using the concept of strict dominance, plan P_3 would not appear on the resulting short list. Yet one might argue that while P_3 is somewhat worse in both attributes than P_1 , it is not *significantly* worse, (as opposed to, say, P_2 and P_8 , which *are* significantly worse in both attributes). In particular, because of uncertainties associated with the calculations, one may be reluctant to discard a plan that is not significantly worse than another from a final short-list, (especially if such an option involves a generation or demand side measure not included in the other plan).

Figure 2.7: The two-attribute problem

win" options.

Figure 2.8: Significant dominance

This idea is captured by the concept of *significant dominance*³⁵. P_1 is said to significantly dominate P_2 if $A_i(P_1) + m_i \leq A_i(P_2)$ for at least one i , and if $A_i(P_1) - e_i \leq A_i(P_2)$ for all i , where m and e are significance parameters or tolerances. m_i is the smallest difference in values for attribute A_i such that one plan is considered to be much worse than the other, and e_i is the largest difference between values such that one plan is considered essentially equivalent to one another.³⁶

These concepts are illustrated on Figure 2.7. P_1 significantly dominates all of the plans in the shaded region. Note that P_2 , which is not *strictly* dominated by P_1 , is *significantly* dominated. That is, significant dominance also rejects solutions where the improvement in one attribute (in this case in the environmental objective) is bought at great cost in the other (in this case in cost). In other words, from the environmental perspective, for a plan to be significantly better than P_1 , it must have a value less than $P_1 - e_i$.

³⁵ This discussion uses the definitions of Crousillat and Merrill (1992).

³⁶ Given that preferences for attribute levels might be non-linear, it could be argued that m_i and e_i depend upon the level of $A_i(P_j)$, rather than being independent of level. For simplicity, we disregard that complication here.

Figure 2.9: Knee set definition

The *knee set* is the set of plans which are not significantly dominated by any other alternative. In the example of Figure 2.8, redrawn on Figure 2.9, the knee set is $\{P_1, P_4, P_9\}$. Note that this set differs from the non-inferior set, defined by the trade-off curve, which is the set $\{P_2, P_1, P_5\}$. From the practical standpoint of decision-making, the knee set is clearly the more useful, since the non-inferior set also includes solutions where slight improvements in one objective are bought only at great cost in the other -- such as P_5 and P_2 . Another way of expressing this is that the knee set consists of that set of plans for which decision-makers representing different objectives are most likely to agree upon. In the above example, for example, both P_5 and P_2 are unlikely candidates for agreement, even though both lie on the trade-off curve.

Obviously the knee set will be a function both of the values of the significance parameters used, as well as the shape of the trade-off curve. In the situation of Figure 2.9, the knee set is relatively compact, because the trade-off curve exhibits a sharp rate of change of curvature at P_1 . However, in the situation shown on Figure 2.10, there are no sharp changes in curvature of the trade-off curve: here the knee set is the same as the non-inferior set.

Figure 2.10: Knee set equivalent to the non-inferior set

All of the mathematical definitions of knee sets and non-inferior sets continue to be valid when there are more than two dimensions. In this case, one should properly refer to the trade-off *surface*, say $S = S(C, A_1, A_2, \dots, A_n)$, which is a n -dimensional non-inferior hyper-surface, where C is the cost attribute, and A_i ($i=1, \dots, n$) are n environmental attributes.

A rigorous definition of the trade-off between any *two* of these attributes would require that the values of all of the others are held constant. In practice that is rarely possible -- for example, finding a series of plans that all had the same water impact, but varied only in cost and air quality impact would be hard to find. What can be done is to simply plot the values of two attributes (e.g. C versus A_1), and to assume that the variations in the other attributes may be ignored. In other words, in the three-dimensional case, the third dimension is in effect projected onto the two-dimensional plane of the two attributes being displayed, ignoring variations in the third dimension.

It might be noted, however, that while looking at attributes only two at a time provides great insight, it may not in fact reveal all of the members of the non-inferior set. This is illustrated in Annex I.

On Figure 2.11 we show some of the actual trade-off curves in the Seattle case study. Figure 2.11A, which shows the trade-off between system cost and CO_2 emissions, has the classic shape as we have used in the above presentation: however in this particular case there is no clear knee-set.

Figure 2.11B shows a plot of CO_2 v. SO_2 emissions. Because CO_2 and SO_2 emissions are highly correlated, this is of a quite different shape: one may conclude that the two objectives of minimizing CO_2 emissions and minimizing SO_2 emissions are largely coincident, and no significant trade-offs must be made between them.

Figure 2.11: Trade-off curves in the Seattle case study

A. CO₂ emissions v. system cost (\$SCL)

B. CO₂ emissions v. SO₂ emissions

2.7: Scaling (Step 7)

The previous steps emphasized the generation and display of trade-off information, which represents the first of the two major purposes of MCDM methods. The second is quantifying value judgments that can lead to a choice (or ranking) from among the non-inferior set. This purpose is the subject of Steps 7-9 of our general MCDM procedure.

Most methods for amalgamation require that the decision maker *scale* attributes. Value scaling can be viewed as the creation of a value function $V_i(A_i)$ which translates a social, economic, or environmental attribute A_i into a measure of worth or desirability. $V_i(A_i)$ is usually defined so that the best outcome is assigned a value of one, the worst a value of zero.

For example, the widely used weighting summation method of amalgamation (discussed further below in Section 2.9.1), ranks options by the score:

$$\text{maximize } \sum V_i(A_i)$$

where w_i is the weight, and $V_i(A_i)$ is the value function of attribute A_i , $i=1, \dots, I$.

Both the value functions and weights can be obtained by several means, as described below: these methods for weighting and scaling apply also to other amalgamation rules (such as goal programming, discussed in section 3.10). A value function describes a person's preferences regarding different levels of an attribute under certainty. In contrast, utility functions also reflect attitudes toward risk.

In the example of Figure 2.12 we show alternative representations of a value function for damages resulting from the discharge of thermal effluents into an estuary. The simplest representation (attribute value function 1) is a simple linear scale that has the implication, for example, that a 1 degree rise in temperature from 26 to 27 °C produces the same damage as a 1 degree rise from 29 to 30 °C. Attribute value function 2, on the other hand, implies a non-linear relationship, and in particular a threshold at T^* . Temperature increases to below this value are

Figure 2.12: Attribute Scaling

assumed to cause little or no damage: while any temperature greater than T^* , all the way to the maximum observed, are perceived as equally bad!

Value functions:

Most amalgamation methods presume that attribute value functions are on an interval level of measurement with respect to value. In a valid interval scale, differences between numbers are meaningful but the zero point is arbitrary. For example, on the interval Fahrenheit temperature scale, it makes sense to say that 10° is halfway between 0° and 20° -- but it is incorrect to state that 20° is twice as warm as 10° . Likewise, in an appropriately scaled attribute value function, the difference between a value of 10 and one of 0 should be the same as the difference between 20 and 10. Only such interval scales can be weighted and otherwise manipulated algebraically³⁷.

A number of methods can be used to create interval-scaled value functions (Fishburn, 1967; Hobbs and Voelker, 1978). For example, the decision maker might be asked the following question:

What level of net benefits is halfway between \$0/yr (having a value of 0) and \$100,000,000/yr (having a value of 1) in value or desirability?"

Note that the answer does not have to be \$50,000,000; nonlinear value functions are possible and, indeed, likely. If linear, then the answer to that question would be a value of 0.5. Scaling techniques that fail to assure an interval level of measurement are widely applied, however, and are in violation of measurement theory. An example would be to simply ask the decision maker to sketch a value function for reliability, or solid waste impacts.

However, there has been no reported research that shows that interval scaled value functions are significantly different from value functions created using methods which do not guarantee interval scales (Hobbs, 1979). By "significantly", what is meant is that choice of scaling method makes a difference in the ranking of alternatives. Likewise, there has been no significant difference shown between utility functions (which include a decision maker's attitudes towards risk) and deterministic value functions, such as those derived from the methods just discussed (which do not include risk attitudes), although some studies have shown that the shapes of the functions have differed (Kryszytofowicz and Duckstein, 1980).

Therefore, it is suggested that scaling be done using any method that the decision makers feel comfortable with. It would be best to allow experts in the relevant fields to scale the attributes they are familiar with (Dyer and Sarin, 1979). For example, a biologist might choose the value function for wetland impacts, while an economist would scale net monetary benefits. It is recommended that the analyst not automatically assume that value functions are linear; there may be thresholds or other nonlinearities.

Indeed, in both the Centerior and SCL case studies some of the functions were strongly non-linear. In both cases the participants were presented with postulated linear value functions, which displayed the individual options and their values on the postulated linear scale (see Figure 2.13), and were then asked about whether linearity reflected their actual assessment of damages and values.

On Figure 2.14 we show some results of the group discussions on scaling in our case studies.³⁸ Figure 2.14a represents the scaling curve for short-term rate impacts in the Centerior

³⁷ The only important exception is utility functions, which, when correctly assessed, are not interval scaled in terms of value: see Dyer and Sarin, (1979). Utility functions are discussed below.

Figure 2.13: Scaling questionnaires in the Seattle case study

Figure 2.14: Results of the scaling discussions in the case studies

A. Centerior: short term rate impact

B. Seattle: Hydro impacts

study. The sharp threshold at zero reflected the low tolerance in the group for any short term rate increase: whether the rate increase was very small or more significant, in both cases a rate submission would have to be made, which was perceived as having a high transaction cost.

Figure 2.14b represents the scaling curve for the attribute *hydro Mw* in the Seattle study. The shape of this curve reflected the group consensus that the critical hydro impact was limitation of whitewater recreation opportunities; and that the extent of river affected was unlikely to be a function of the installed capacity (i.e. a 15Mw hydro plant would have the same undesirable effect in terms of miles of river affected as a 30Mw plant).

Utility functions:

This value scaling process described above is closely related to the concept of *utility functions*. Utility functions, however, are more general because they also capture the user's attitude towards taking risks. The purpose of utility functions is to find some function $u_i(A_i)$ for each attribute A_i , such that if $E(u_i(A_i))$ for plan A is greater than for plan B, then the decision maker prefers plan A, and vice versa. ($E(A)$ is the expected value or mean of A .) Such a function can reflect both 1) the innate value of different levels of the attribute and 2) the decision maker's attitudes towards risk (risk seeking, risk neutral, or risk averse). There is a well developed body of theory which describes under what circumstances utility functions accurately characterize decision maker preferences (Keeney and Raiffa, 1976). This approach is recommended by many risk analysts (e.g., Sage and White 1980) and has been used in water resources planning (Keeney and Wood, 1977; Krzysztofowicz, 1986)

If the utility function approach is adopted to deal with risk and uncertainty, the expected value of $u_i(A_i)$ would be calculated for each plan and subsequently traded off against other attributes. If a decision maker is risk neutral, then he or she will prefer the plan with the best value of $E(A_i)$, no matter how little or how much risk there is associated with each plan, for example, at the tails of the probability distribution of floods. For example, if A_i is an attribute for flood risks from a hydro plant, a single rare catastrophic flood might implicitly be weighted far less (by virtue its low probability) than more frequently occurring medium-sized floods. This means that the utility function is linear, and the simplest approach is to directly use the expected value of the attribute. Examples might include: expected generation cost, expected flood damages, expected loss of wildlife habitat.

But decision makers are not generally risk neutral. In that case, utility functions are nonlinear, and the alternative with the best $E(A_i)$ might not be preferred because it might be riskier than other plans (for example, because it has a higher standard deviation). An example is the exponential utility function:

$$u_i(A_i) = a + be(cA_i),$$

in which a , b , and c are parameters, and x_i is the original attribute.

It must be noted, however, that nonlinearities in utility functions arise for two reasons (Dyer and Sarin, 1979; Chankong and Haimes, 1983):

³⁸ What was particularly noteworthy in both of the case studies was the relative ease with which consensus was attained. As each attribute was discussed, in turn, initially there was often wide disagreement about the shape of the curve, where thresholds might lie, etc. However, after several rounds of letting each participant express their views, with one or two exceptions, where one or more individuals insisted on a dissent, a consensus was reached.

1. The decision maker's preferences for levels of an attribute (as represented by a deterministic value function) are nonlinear. As an example, preferences for levels of dissolved oxygen in a stream are nonlinear because of threshold effects.
2. The decision maker is either relatively risk averse or risk seeking in general. An example involving bets with money might clarify the latter type of preference. Assume that a person's deterministic value function for money is linear. If the person is indifferent between \$50 and a 50-50 chance at \$100 (expected value = \$50), he/she is considered to be relatively risk neutral. If he/she would rather have the bet, he/she is relatively risk seeking. He/she would be risk averse if the certainty of \$50 is preferred. Formally, one is said to be relatively risk averse if the expected value of his/ her deterministic value function is greater than expected utility. He/she is relatively risk preferring if the opposite is true. It suffices to use deterministic value functions (which are easier to assess than utility functions) if the decision maker is relatively risk neutral or the level of each attribute for each plan is specified as a single number rather than as a probability distribution.

There are serious difficulties with the utility function method. Decision makers may not be able to respond to lottery questions or their answers may be inconsistent, leading to invalid functions. Krzysztofowicz and Duckstein (1980), for example, show that a person's utility function depends very much on purely arbitrary aspects of how the questions needed to calibrate that function are posed. And even if decision makers can choose meaningful functions, whose values and risk attitudes should be used? The utility functions of different people can differ drastically. Of course, this problem is not unique to utility functions. Different people's value functions can also differ as can their weights. In step 10, we address the question of how to resolve such conflicts.

2.8: Weighting Methods (Step 8)

Almost all ranking methods (such as additive value functions) require weights. After introducing several weighting methods below, what constitutes a "valid" weight is carefully defined, and the validity and ease of use of different techniques for picking weights are then contrasted. The methods are discussed as if they would be used in additive value functions; however, the discussion applies equally to other amalgamation methods which require weights. Recommendations concerning which weighting methods should be used conclude the section.

Equal weights:

The simplest approach to weighting is to use equal weights. However, to say that two attributes are given equal weight makes them equally important is incorrect. For example, say that $w_1 = w_2$. Then if the difference between $V_i(A_j)=0$ and $V_i(A_j)=1$ represents only a trivial change in A_1 but a significant change in A_2 , then, implicitly, a higher importance is assigned to A_1 . This is because a large improvement in A_2 is required to compensate for a small worsening in A_1 ³⁹. The relative importance of two attributes depends not just on their weights, but also their ranges, as reflected in the $V_i(A_j)$.

Observer derived weights:

Observer derived weighting estimates weights using multiple regression or linear programming to analyze unaided subjective rankings of plans in order to determine which weights best predict those evaluations. This method is sometimes called policy capturing, as it calculates the weights implicit in a person's unaided judgments (Hammond, 1986).

Direct weighting:

Direct weighting methods, in contrast, ask decision-makers to choose numerical values for weights directly. Examples of direct methods include point allocation (allocate 100 points among attributes in proportion to their "importance"), categorization (assign attributes to different categories of importance, each carrying a different weight), ranking (the least important attribute receives a weight of 1, the next lowest, a weight of 2, etc.), ratio questioning (which asks for ratios of the importance of attributes, two at a time), and rating (each attribute's importance is rated on a scale of, say, 0 to 10). Unfortunately, direct methods often fail to yield weights that correspond to tradeoffs people are willing to make: the reason is that subjective notions of attribute "importance" may have little to do with the rate at which a person is willing to trade off the $V_i(A_j)$ for two attributes.

Direct methods can assign weights to all criteria at once (Figure 2.15a). But if there are many criteria, such as the twelve in our case studies, then a hierarchical approach is often easier. This procedure groups criteria into general categories. In one version, the user first assigns a weight to each general category. Then, the user allocates each category's weight among its criteria. Another approach is to first assign weights within categories and then weight the categories themselves. Hierarchies can also be defined with several levels, as in Figure 2.15b.

³⁹

i.e., to leave the value of $w_i V_i(A_j)$ constant.

Figure 2.15: Hierarchical and non-hierarchical attribute structures

In figure 2.16 we show the "peakiness" of weights elicited by a variety of different methods in the Centerior case study: "Peakiness" is defined as the difference in weight between the highest weighted attribute, and the lowest weighted attribute: if all were given equal weight this measure would be zero. This measure was defined for each individual, and then averaged across individuals for each method.⁴⁰

Figure 2.16: Peakiness of weights in the Centerior case study

⁴⁰ Before making the calculations, all weights were normalized by rescaling them so that their sum was 1.0.

As one might expect (Stillwell *et al.*, 1987) the peakiness of non-hierarchical weighting ($p(w)=0.29$ and 0.37) is significantly higher than for non-hierarchical rating weights ($p(w)=0.19$ and 0.25). Swing weights are seen to give the "flattest" results; but AHP weights (see discussion below of this method) are also very peaked.⁴¹

The Analytic Hierarchy Process:

A popular version of ratio questioning is the Analytic Hierarchy Process (AHP; Saaty, 1980). It asks decision-makers to compare every possible pair of criteria using a scale from 1/9 to 9. For example:

Is criterion A (say air quality impact) 1/7, 1/5, 1/3, 1, 3, 5, 7, or 9 times more important as criterion C (say cost)

A matrix of these ratios is then set up as shown on Table 2.3 for a four -criteria example.

Table 2.3: Matrix of importance ratios.

	A	B	C	D
A	1	1/3	3	5
B	3	1	7	9
C	1/3	1/7	1	3
D	1/5	1/9	1/3	1

This shows, for example, that A is three times more important as C (and, thus, C is 1/3 as important as A). These ratios may not be completely consistent -- for example, A is 3 times as important as C, C is three times as important as D, but A is not 9 ($=3 \times 3$) times as important as D. Such inconsistencies are to be expected, since people's values are often inconsistent, or their quantitative expression may be inconsistent. AHP solves for the set of weights that are, in some sense, most consistent with these ratios by a mathematical procedure called eigenvector analysis. Let R be the above matrix of ratios. The W is the weight set that solves the equation

$$R W = \lambda W$$

while simultaneously yielding the largest possible value of λ (which is called the eigenvalue of the matrix R).⁴² In the above case, $W = \{0.24, 0.61, 0.10, 0.05\}$, and $\lambda = 4.09$. The AHP is used in the Centerior case study (Meier and Hobbs, 1994): Figure 2.17 shows the questionnaire that was used.

Swing weights:

A final variant of direct weighting is swing weighting (vonWinterfelt and Edwards, 1986). In this method, the user considers a hypothetical plan where attributes are all at their worst value. The user is then asked "which of the attributes would you prefer to "swing" from its worst value up to its best value?". Then the user is asked which attribute he or she would swing up second, and so on. After ranking the attributes in this manner, the user is asked to give the most important attribute a weight of 100, and then assign weights to the other attributes in proportion to the importance of their ranges, while being sure that the most highly ranked attributes receive the highest weight. Many decision analysts favor this method over other direct techniques

⁴¹ In the Seattle case study, trade-off weights were also very peaked.

⁴² W is unique up to a scaling constant. Usually, the weights are rescaled so they sum to 1.

Figure 2.17: The AHP weighting questionnaire in the Centerior case study.

The purpose of this questionnaire is to ask you to compare two criteria at a time. A set of weights that are consistent with the using the "Analytical Hierarchy Process." These weights will be used in a weighted utility model.

Criterion "importance" is defined as follows. An attribute X is more important than an attribute Y if alternative A is preferred to alternative B:

- A: X is at its best value, Y is at its worst value
B: X is at its worst value, Y is at its best value

The ratios of importance you state should represent your willingness to trade off the criteria. The following scale is suggested for ratio judgments:

The ratio of Criterion X over
Criterion Y should be:

IF:

- 1/9 Criterion Y is extremely more important than Criterion X
1/7 Criterion Y is strongly more important than Criterion X
1/5 Criterion Y is moderately more important than Criterion X
1/3 Criterion Y is slightly more important than Criterion X
1 The two criteria are equally important
3 Criterion X is slightly more important than Criterion Y
5 Criterion X is moderately more important than Criterion Y
7 Criterion X is strongly more important than Criterion Y
9 Criterion X is extremely more important than Criterion Y

1/8, 1/6, 1/4, 1/2, 2, 4, 6, 8 Intermediate values between two adjacent judgments

What is the ratio of importance of Criterion X to Criterion Y?

Criterion X (range)

Criterion Y (range)

Answer (1/9 - 9)

Economic Criteria

Revenue Reqs. (-41 - 339x10 ⁶ \$/y)	Capital Expenditures (-86 - 220x10 ⁶ \$/y)	1
Revenue Reqs. (-41 - 339x10 ⁶ \$/y)	Rates (1993-99) (-3.3 - 10.8 \$/MWh)	1/9
Revenue Reqs. (-41 - 339x10 ⁶ \$/y)	Rates (1993-2013) (-3.2 - 24.3 \$/MWh)	1/7
Capital Expend. (-86 - 220x10 ⁶ \$/y)	Rates (1993-99) (-3.3 - 10.8 \$/MWh)	1/4
Capital Expend. (-86 - 220x10 ⁶ \$/y)	Rates (1993-2013) (-3.2 - 24.3 \$/MWh)	1/7
Rates (1993-99) (-3.3 - 10.8 \$/MWh)	Rates (1993-2013) (-3.2 - 24.3 \$/MWh)	5

Air Criteria

SO ₂ (-28564 - 17,876 t/y)	CO ₂ (-4,437,000 - 3,047,000 t/y)	5
SO ₂ (-28564 - 17,876 t/y)	NO _x (-10,000 - 7,000 t/y)	5
CO ₂ (-4,437,000 - 3,047,000 t/y)	NO _x (-10,000 - 7,000 t/y)	1

because the user is asked to directly consider each attributes' ranges when ranking them.

Trade-off weights:

An alternative weighting approach which avoids this problem is the indifference tradeoff method. It asks decision makers to state how much of one attribute they would be willing to give up to obtain a given improvement in another attribute; then, the method calculates the weights implied by such indifference judgments, as in Eq.[1], below.

Properties of weights:

No matter which method is used to choose weights, the resulting values must be ratio scaled. If attribute A_1 is twice as "important" as A_2 , w_1 should be twice as large as w_2 . Furthermore, weights should represent the relative importance of unit changes in their attribute value functions. If a decision maker is indifferent between a change in $V_1(A_1)$ of 1 and a change in $V_2(A_2)$ of 1/3, then w_1 should equal 0.333 w_2 . Only this definition of importance should be used.

Another way to put this is that weights should represent the rate at which the user is willing to tradeoff one criterion for another. This can be seen imagining two alternatives A and B that differ in just two attributes, say h and i . If a decision maker finds A and B equally preferable, then from (1) we can derive:

$$w_h/w_i = [v_i(x_{iA}) - v_i(x_{iB})] / [v_h(x_{hB}) - v_h(x_{hA})] \quad [1]$$

That is, the ratio of the weights of two criteria should be inversely proportional to the rate at which the decision maker is willing to trade them off. For instance, if NO_x gets three times the weight of CO_2 , then an improvement of amount X in $v_{\text{NO}_x}(\text{NO}_x)$ should be valued just as much as an increase of $3X$ in $v_{\text{CO}_2}(\text{CO}_2)$.

Thus, an attribute's weight is inextricably linked to its value function. One cannot meaningfully choose a weight unless the value function (or at least its upper and lower bounds) has been defined; otherwise, tradeoffs may be made which do not reflect the true preferences of the decision maker. For example, one might assert, *a priori*, that "environment" is just as important as "economics." However, depending on the actual ranges of the environmental and economic attributes and how their value functions are defined, this could imply that each ton of NO_x emissions from a fossil fuel power plant is worth \$200 -- or \$2,000,000! Many applications have made the mistake of setting weights without defining or presenting attributes and their value functions (see Hobbs and Rowe, 1979, p. 122 for a number of citations). The resulting additive value functions are, at best, misleading representations of the decision maker's preferences, as there is no way to determine what tradeoffs he or she is actually willing to make.

Validity of weights

Some methods for choosing weights are specifically designed to produce valid weights -- that is, weights on a ratio scale representing the correct type of importance. Others are not. The former group of methods generally ask decision makers to trade off attributes (e.g., "how many dollars are you willing to sacrifice in system cost in order to decrease CO_2 emissions by 1 million tons"), and then calculate weights that are consistent with the tradeoffs they are willing to make. An example is the indifference tradeoff method. In contrast, methods that do not assure valid weights usually directly ask for judgmental assessments of an attribute's importance, where "importance" is ambiguously defined and may have nothing to do with the decision maker's willingness to trade it off against other attributes (Schoemaker, 1981; Goldstein and Beattie, 1990). Direct weighting methods, such as rating, are methods of this

type. Such methods make no effort to determine if weights are consistent with the tradeoffs decision makers are willing to make.

However, differences in the theoretical validity of weighting methods does not matter if, in practice, different weighting methods yield the same weights and rankings of alternatives. If that was true, the most easily applied method could be used. Unhappily, a number of experiments have shown that different methods can choose different weights and, ultimately, yield different decisions (Hobbs, 1986). In general, observer derived (multiple regression) weights differ strongly from those chosen by other methods, perhaps because decision makers tend to focus erratically on just two or three of the attributes when making the holistic evaluations of alternatives that are required by regression (Shepard, 1964). For this reason, we cannot recommend the use of regression methods.⁴³

In contrast, different direct methods, such as rating and point allocation, generally yield similar weights and evaluations of alternatives. But weights and evaluations resulting from indifference tradeoff and direct assessment have been shown in several experiments to be significantly different (Hobbs, 1986), a result that is in part confirmed in our two case studies as well. By "significantly", we mean that choice of weighting method can make as much or more a difference in the results as which person chooses the weights.

Conclusions:

Thus our recommendations on weighting methods are as follows. If the decision makers feel comfortable in answering tradeoff questions, then indifference tradeoff weights are to be preferred. Tradeoff questions force decision makers to think about what they value in very concrete terms; as a result, learning takes place and they feel that they know more about the problem.

However, trading off attributes is an unfamiliar task involving conflicting, strongly held values. As a result, decision makers may not know what tradeoffs they are willing to accept and the indifference tradeoff weights they choose may be unstable (Fischhoff *et al.*, 1979). It is therefore important to check for consistency by asking more than the minimum number of questions needed to determine weights. Consistency checks also aid the learning process by encouraging decision makers to reflect on their values and the problem.

On the other hand, our experiments with electric utility planners show that many (perhaps as much as a majority) of decision makers are distinctly uncomfortable with tradeoff questions. As a result of this discomfort, the answers may not be meaningful and the decision makers have little confidence in the results. In that case, weights should instead be chosen by a direct method, such as rating on a scale of 0 to 100. Hierarchical rating is certainly to be preferred over non-hierarchical weighting.

There are several precautions that should be taken if rating is used. First, those who are choosing the weights should be informed about the necessary properties of weights (ratio scales and the correct type of importance). Second, when picking weights, decision makers should keep in mind the ranges of the attributes. For example, if system cost has a range of \$0 to \$10,000,000 and CO₂ emissions has a range of 1 to 20 million tons, benefits should receive roughly twice the weight as would be the case if the range of system cost was only \$0 to \$5,000,000. (It is assumed that the value functions $V_i(A_i)$ are always on a 0 to 1 scale, with 0

⁴³ Except perhaps as a means of understanding which attributes are most salient in a person's unaided evaluation process. However, we believe that MCDM methods should be used to improve judgments, not imitate them, so weights derived by regression methods should not be used to rank alternatives and make recommendations.

being the value of the least desired outcome and 1 the value of the most one.) Third, some tradeoff questions should be asked as checks upon the weights. For instance:

"You stated that the weight for system cost be 75 and the weight of CO₂ emissions should be 50. Based on the ranges of the attributes, this implies that you would prefer an alternative with a cost of \$10,000,000 and 1 million tons of CO₂ emissions to one costing \$20,000,000 and 50,000 tons of CO₂ emissions. Is this true?"

2.9: Amalgamation methods (Step 9)

Amalgamation methods can be viewed as methods for approximating decision makers' "indifference curves," which show which alternatives are equally preferred. For example, Figure 2.18 shows a set of hypothetical indifference curves for one decision maker. The figure links the objective of system cost with that of air emissions. Plans P1 and P14 are equally preferred and therefore lie on the same indifference curve. So are P12, P9, and P8. Curves below and to the left represent more preferred combinations of the two attributes (assuming that less of each is better). From this perspective, the problem faced in amalgamation is to measure a person's values (as crystallized by the indifference curves) and then to find the most preferred option (the one that lies on the lowest curve) -- i.e. in the example of Figure 2.18, the curve P1-P14-P13.⁴⁴

Figure 2.18: Indifference curves

Thus most amalgamation methods, in essence, estimate a utility function which represents the decision maker's priorities. The isoquants of such a utility function are indifference curves, and the questions the method asks of decision makers fix the slopes and shapes of those curves. Figure 2.19 shows the functional forms of indifference curves assumed by two amalgamation methods: additive value functions (in which the weighted sum of scaled attributes is maximized) and goal programming (in which the distance to an "ideal" plan or set of goals is minimized; see below, section 2.9.2)

⁴⁴ However, this is subject to the caveat made in the introduction: very often people do *not* know what they want; and their views may evolve during a planning exercise as a result of more information, better understanding, and, most importantly, interaction with others who hold possibly different views.

Figure 2.19: Indifference curves implied by two amalgamation methods

2.9.1 Weighting summation

We have already made a number of references to the weighting summation method, which ranks options by the score:

$$\text{maximize } \sum w_i V_i(A_i)$$

where w_i is the weight, and $V_i(A_i)$ is the value function of attribute A_i . The resulting indifference curves are shown in Figure 2.19a. This is unquestionably the amalgamation method in widest use, and we have already pointed to the need to pay attention to how the weights and the value functions are derived in order for the results to be valid. Unhappily, as we shall see in our review of applications in the next Chapter, many are quite lacking in their understanding of the requirements: the problem seems to be that the concept and algebra of the method seem so simple, that issues of validity may seem esoteric.

Yet in fact they are not, as well documented in a number of studies. Over a decade ago the US Nuclear Regulatory Commission examined the application of such site selection models to nuclear power plant siting (Rowe *et al.*, 1979): ten of thirteen applications of weighting summation violated theoretical requirements (such as use of ordinal scales), eight of 13 violated the requirement for preference independence of attributes, and none acknowledged the existence of *any* theoretical requirements for validity.

2.9.2 Goal Programming

Goal programming (GP) is another amalgamation method that yields a ranking of plans. The user provides weights that reflect the relative importance of each attribute (using an appropriate weight assessment method discussed earlier), and the target level (goal) of each attribute the user would like to see achieved. These inputs make it possible to rank plans according to the weighted deviation from goal: the smaller the deviation, the more preferred the plan. The idea, therefore, is to choose the plan closest the goal by minimizing a distance measure of the form:

$$\text{Min } \sum (w_i |G_i - x_{ij}|^P)^{1/P}$$

where: $||$ = the absolute value operation

w_i = weight for attribute i (normalized so that the sum of the weights equals 1.0)

G_i = goal for attribute i

x_{ij} = impact (score) of plan j on attribute i

P = a positive constant signifying how much large deviations from goals should be penalized.

An alternative form has the weights w_i located outside the parentheses so that they are not raised to the power P ; in the case studies we used the formulation shown above.

If $P = 1$, the results represents the absolute value of deviations form ("city block distance"), which penalizes each size of deviations equally. When $P = 2$, the squared deviations form ("Euclidean" distance) results. Here, larger deviations are penalized more than smaller ones since each deviation is, in a sense, multiplied the attribute equal to itself. Figure 2.19b

displays the implied difference curves in that case. Finally, when $P = \text{infinity}$, the "minimax" distance results, in which a plan that minimizes the maximum (worst) weighted deviation from any goal is chosen. The choice of P clearly can affect results as different values of P have different implications concerning the tradeoffs among attributes. A guideline for choosing an appropriate value of P is discussed in Hobbs (1979).

The target level G_i of an attribute is often set at the best possible value achievable by any plan. For example, the ideal goals for cost, hydro Mw, and SO_2 emissions might be set at, respectively, \$0, 0 Mw, and 0 tons. Using these ideal goals, the only inputs required from the user are attribute weights and the choice of P , and goal programming can be viewed as a linear additive model with all value functions predetermined (except for the value of P). For example, when $P = 1$, goal programming using ideal goals reduces to a weighting summation approach with the value function for attribute i being $v_i(A_i) = A_i$, which implies constant tradeoffs among attributes for all values of A_i . The former is clearly subjected to any assumptions, pros, and cons that the latter has (see Hobbs (1979, pp. 105-115)).

Likewise, if $P = 2$, the method is equivalent to an additive model with $v_i(A_i) = (G_i - A_i)^2$ and weights $= w_i^2$, which represents a set of concentric ellipsoids centered at $(G_1, \dots, G_n)$. Here, the less of $v_i(A_i)$ one has, the less of it one is willing to give up for a unit of $v_j(A_j)$, and *vice versa*. This concave indifference curve favors plans with balanced impacts over ones with extreme high and low impacts more than the case where $P = 1$.

Finally, in the extreme case where $P = \text{infinity}$, the single attribute value function is the similar to when $P = 1$, i.e. $v_i(A_i) = (G_i - A_i)$, but the overall value function is $v = \max \{w_1(G_1 - A_1), \dots, w_n(G_n - A_n)\}$. Here, the user is not willing to worsen the attribute furthest from its goal even for large increases in other attributes. With equal weights, each goal will probably be achieved to the same degree, i.e. $w_i(G_i - A_i) = w_j(G_j - A_j)$ for all $i=j$.

In practice, the user is allowed to select target levels that are "acceptable" rather than "ideal". In this case, it is possible that the target levels will be chosen so that the final choice is indeed a dominated alternative (Cohon, 1978). This is particularly likely if the target levels are chosen in the absence of a complete knowledge of the set of choices and their impacts. The problem of dominated alternatives being chosen by goal programming is a well-known and serious drawback of the method.

2.9.3: Multiplicative utility functions

As noted previously, single attribute utility functions capture not only preferences regarding levels of the attributes, but also relative risk attitudes. Multiattribute utility functions do the same for several attributes simultaneously, usually by weighting and amalgamating single attribute utility functions. The expected value of the multiattribute function is then used to rank the alternatives.

The multiplicative form is the most widely applied multiattribute utility function (Keeney and Raiffa, 1976), and has found use in utility planning (Keeney and Nair, 1977; Keeney and Sicherman, 1983; Keeney and von Winterfeldt, 1987). A special case of the multiplicative form is the additive form. The remainder of this section describes in general how such a function is constructed. Two types of information are required from the decision maker:

1. single attribute utility functions $u_i(A_{ij})$ (defined in Section 2.7), and
2. a set of weights w_i

The utility functions $u_i(x_{ij})$ are scaled from 0 (worst) to 1 (best), and the weights each lie between zero and one. After they are chosen by one of the methods described above, the alternative that maximizes the expected value of one of the following multiattribute utility functions is picked:

Additive form: Maximize $\sum w_i u_i(A_{ij})$

Multiplicative form: Maximize $= \{[\prod (1 + Kw_i u_i(A_{ij}))] - 1\} / K$

Which form is used depends on the sum of the weights w_i . If the sum is 1.0 (indicating multiattribute risk neutrality), then the additive form applies; if not, then the multiplicative form is used (multiattribute risk aversion or seeking). The scaling parameter K is set to the value that causes the value of the multiplicative form to lie between 0 and 1.

The assumptions that both the additive and multiplicative utility functions make are:

1. Utility independence: the answers to gamble questions involving just one attribute do not depend on the levels of the other attributes (i.e., each attribute's utility function $u_i(A_{ij})$ is not a function of the level of the other attributes); and
2. Preference independence: this holds if the tradeoffs one is willing to make among any two attributes do not depend on the levels of the other attributes (that is, the weights are constant).

There are several ways to determine the $u_i(A_{ij})$ and w_i . SmartEdge, a popular computer package for decision analysis (Haviland-Lee, 1987), determines those functions and weights as follows. It asks the user to choose between a "sure thing" (an alternative whose attributes are known) and a "gamble" (an alternative whose attributes are random). In the case of the $w_i u_i(A_{ij})$ determinations, SmartEdge adjusts the value of the "sure thing" until the user is indifferent between the "sure thing" and "gamble". (This is called a "certainty equivalent" form of a gamble question.) In the case of the weights w_i , SmartEdge adjusts the probabilities in the "gamble" until the user is indifferent between the "sure thing" and "gamble". (This is a "probabilistic equivalence" form of a gamble question.) SmartEdge solves for w_i and $u_i(A_{ij})$ from the answers to these questions. In the case of the single attribute utility functions, the "gamble" is a plan which has probability 0.5 of obtaining the worst level of attribute i and probability 0.5 of getting the best level.

The "sure thing" has probability 1 of obtaining a level z between the best and worst levels. Because the best level has a utility of 1 and the worst has one of 0, the expected utility of z is 0.5. This type of question is then iterated to determine the utility of other points between the best and worst levels. It is also possible (and, in the opinion of the authors of this report, preferable) that some of the weights instead be determined using indifference tradeoff questions;

⁴⁵ For the weights, the "sure thing" is a plan with attribute i at its best level, and all other attributes at their worst levels. The "gamble" is an option which has a chance w_i of having all attributes take on their best levels, and chance $(1 - w_i)$ of having all attributes realize their worst levels.

⁴⁵ see Keeney and Raiffa (1976) or Keeney and von Winterfeldt (1987).

2.10: Resolving differences (Step 10).

We know of few if any applications of MCDMs which yielded completely consistent results. If more than one method was used, results may differ. Even given consistency of method, it is not very likely that all participants will arrive at the same ranking of options. Indeed, if they do achieve perfect consistency, the group in all likelihood has been poorly selected. If as we have argued earlier as being desirable, the group reflect very diverse views, including those of NGOs and likely intervenors in subsequent regulatory procedures, one should *expect* divergence of results. Finally, a participant applying the same method at the beginning and at the end of the process is very unlikely to arrive at the same result: as noted at the outset, one important purpose of the use of MCDM is to educate, to expose members of a diverse group to each other's views, to force people to really think about the issues and problems. The more successfully that objective is attained, the more likely is it that a participant may change some of his initial views.

Resolving differences between people.

This is by far the most important issue. Ultimately, of course, it is the regulatory commission who decides what trade-offs are to be made, and what modifications must be made to the utility's proposal for the future resource mix.⁴⁶ One should expect significant differences between individuals in weights, because the weights are expressions of their values.

On Figure 2.20, for example, we show the average weights given to the different attributes for selected participants in the Centerior and Seattle case studies. In each case we have sorted the individuals from left to right by the total weight given to environmental factors: in Seattle, this ranges from a minimum of 10% to a maximum of 92%; whereas at Centerior, the range is from a minimum of 18 % to a maximum of 40%.

The much greater variation in the Seattle case is simply a reflection of the much greater diversity of individuals, which included representatives from a business group, and an NGO, while no non-utility individuals were present at Centerior. Thus the latter was clearly a more homogenous group in terms of their values.

Clearly this is not a matter for resolution by analysts. No amount of massaging of results will change such fundamental differences. Instead, what can be accomplished is two things:

1. Identify agreement. Are there plans that everyone (or nearly so) agrees should be in, say, the top five? Are there plans that are clearly undesirable to everyone? Even when weights differ, some alternatives may appear as attractive on everyone's lists.
2. Identify disagreement. Are there alternatives that some people highly ranked, but other didn't. Because value judgments are explicit in MCDM, it then becomes possible to explore the exact reasons why disagreement occurs.

⁴⁶ The US Bureau of Reclamation has successfully used rating and weighting methods to solicit public input on water project design. Although interest groups chose greatly different weights in both of two studies, the top ranked alternatives were the same for all groups (Brown 1984, 1990).

Figure 2.20: Average weights by individual

Resolving differences among methods.

Whether different methods produce different results is a major issue. Most studies that have used more than one method, including our own studies of SCL and Centerior, find that results do vary by method. On Figure 2.21 we display this variability by displaying the average rank of different plans by method.

Figure 2.21: Average rank by different methods, Centerior case study.

Fortunately, however, it appears that the degree of variation among methods is not constant. On Figure 2.22 we display these results in a different way, plotting the standard deviation against the average rank. The variation for the best performing plans (a, ref, b), and for the worst plans (i, l and d) is much smaller than for those in the middle (only g, among the mid-ranked plans, has an unusually low variability).

Figure 2.22: Variability of method differences, Centerior case study

Whenever several methods are used, looking at the between-methods correlation coefficients provides useful guidance. On Table 2.4 we show the between-method correlation coefficients for the Centerior case study. One of the most striking feature of these results is the very low correlation between all goal programming methods except GP1,-. At the same time, there are also several methods that give similar results: AHP and the hierarchical rating methods are strongly correlated: this is a reflection that both are characterized by relatively "peaky" weights (recall the earlier discussion of section 2. 8).

Table 2.4: Between-method correlations, Centerior case study

	holistic	rating H-1	rating H-2	rating NH-1	rating NH-2	swing	swing_ high	swing_ low	AHP	GP1,-	GP1,+/-	GP5,-	GP5,+/-
holistic		0.58	0.78	0.73	0.58	0.58	0.57	0.58	0.58	0.76	0.62	0.65	0.4
rating H-1					0.82	0.79	0.79	0.81	0.8	0.73	0.36	0.52	0.26
rating H-2				0.88		0.83	0.83	0.93	0.96	0.89	0.64	0.69	0.32
rating NH-1						0.86	0.7	0.98	0.91	0.82	0.63	0.62	0.33
rating NH-2						0.93	0.84	0.94	0.72	0.63	0.32	0.43	0.28
swing							0.87		0.81	0.71	0.48	0.49	0.28
swing_ high									0.76	0.72	0.41	0.47	0.24
swing_ low									0.74	0.69	0.51	0.46	0.41
AHP										0.78	0.57	0.6	0.38
GP1,-											0.83	0.88	0.59
GP1,+/-												0.8	0.75
GP5,-													0.73
GP5,+/-													

Resolving differences for specific individuals.

This type of difference is perhaps the most easy to resolve by the simple method of follow-up interview. In some cases, significant discrepancies prove to be the result of some misunderstanding about how some particular question should be answered. In others, reflection and evolution of thought, as the *process* brings new information and perspectives to an individual's attention, is the reason for differences. We conducted follow-up interviews with all participants in both case studies, and they proved to be extremely useful.⁴⁷

It turns out that some individuals find it much easier than others to give quantitative expression to weights. For example, on Figure 2.23 we show the criterion weights expressed by two individuals in the Centerior case study for a variety of different methods. Individual X expressed weights that are highly consistent; whereas individual Y expressed weights that vary significantly across different methods. This does not mean that X's weights are "better"; it simply means that the choice of method matters more in some individuals than in others. Even in a consistent individual (such as X on Figure 2.23), there are some inconsistencies: for example, the weights given to job loss and emergency power in the AHP and non-hierarchical weighting methods are significantly higher than in others. This is the sort of point to be clarified in follow-up interviews.

We found that the process of examining and resolving such conflicts yields additional insight and more confidence in the results. Although more than one method and resolving inconsistencies takes more time, we strongly believe that it is worth the additional effort (approximately two hours more per person).

⁴⁷ The questionnaires used in these follow-up studies are contained as an Annex in Hobbs and Meier (1994).

Figure 2.23: Weights by individuals

Annex I

Trade-off analysis for the three-dimensional case

As an illustration of the caution that needs to be applied when looking only at the two-dimensional representations of the trade-off surface, consider a three-attribute example with attribute values for five plans as shown on Table 2.5.⁴⁸ The two-dimensional plots are shown on Figure 2.24.

Table 2.5: A three-dimensional example

	cost	air	water
P1	0.5	0.1	0.95
P2	1	0.5	0.1
P3	0.1	0.9	0.5
P4	0.6	0.6	0.6
P5	0.7	0.35	0.95

From Figure 2.24 one might infer that the non-inferior set is $\{P1, P2, P3\}$, since in two dimensions the members of the trade-off curve are $\{P1, P2\}$ for water versus air; $\{P3, P2\}$ for water versus cost, and $\{P1, P3\}$ for air versus cost. Both $P4$ and $P5$ appear to be strictly dominated.

But as inspection of Table 2.8 shows, in fact when one uses the mathematical definition of strict dominance in a simultaneous comparison of all three attributes, only $P5$ is seen to be strictly dominated (by $P1$). Option $P4$ is seen to be preferred over $P2$ in cost, preferred over $P3$ in air, and over $P1$ in water. Indeed, it may in fact be an attractive compromise solution to decision-makers representing different interests because it avoids very bad outcomes.

For example, from the environmental perspective one may accept a second best outcome for, say, water, if one can avoid the worst case outcome for air quality. Such a preference behavior on the part of decision-makers can be captured by an asymmetrical definition of the tolerance parameters in the knee-set definition: if $e \gg m$ (recall Figure 2.8), then it quickly follows by inspection from the above table that $P4$ is the only member of the knee set! For example, if $ei = 0.5$, and $mi = 0.1$, then because $P1$ is no longer *significantly better* than $P4$ for the air quality criterion, but is *significantly worse* than $P4$ for water, $P4$ significantly dominates $P1$, and so on.

None of this invalidates the two-dimensional plots of the trade-off surface: they still provides a useful view of those plans that are closest to the origin (i.e. best) for the two attributes displayed. However, it must simply be recognized that both knee-set and non-inferior sets must be defined on the basis of the simultaneous, multi-attribute inspection of attribute values: looking *only* at the two-dimensional displays may not uncover all of the solutions of actual interest to decision-makers.⁴⁹

⁴⁸ A similar example is given in United Nations Environment Programme (1988)

⁴⁹ It is of course also true that we have used a highly contrived example here to make our point!

Figure 2.24: A three-dimensional example

Annex II

Generating trade-off curves

Although a review of the literature reveals a great many different approaches to the problem of identifying the non-inferior set, three methods account for the bulk of practical applications. The first two, the so-called *constraint* method, and the *weighting* method, are derived from the theory and practice of multi-objective mathematical programming (Cohon 1978). These approaches are useful if a mathematical programming model is used for IRP, such as EGEAS, or SCL's ANN model. The third approach is simply to analyze defined plan alternatives using a simulation model, and determine the trade-off curve and knee set by inspection -- as described earlier in the Chapter. The latter approach is useful when a limited number of discrete plans are pre-defined -- as in the Centerior case study.⁵⁰

Let $COST$ and CO_2 be the two attributes of an analysis. In the constraint method, one solves a series of optimization problems of the type

$$\begin{aligned} & \text{Min } COST \\ & \text{subject to } CO_2 < CO_2(max)_i \end{aligned}$$

for a series of different values of $CO_2(max)_i$.

The principle is illustrated on Figure 2.25. In iteration 1, one minimizes cost subject to $CO_2 < CO_2(max)_1$, for which the minimum cost plan is clearly P_9 . In iteration 2, one tightens the CO_2 constraint to $CO_2(max)_2$ (defined as $CO_2(max)_1 - \Delta$), for which the least cost

Figure 2.25: The constraint method

⁵⁰ Other methods for generating non-dominated solutions are discussed in Zeleny (1981) and Stevens (1986).

solution is P_5 . In the third iteration, the least cost solution is P_{10} . In this way, by successive tightening of the CO_2 constraint, the non-inferior set $\{P_9, P_5, P_{10}\}$ is generated.

In practice, this method has its hazards: as one can see from Figure 2.25, the non-dominated points P_1 and P_3 have been missed, given the way in which the successive $CO_2(max)$ constraints have been defined. There are two ways of getting around this problem: the first is to make the Δ value very small, so that the probability of missing a point such as P_3 is minimized. For example, with Δ set at half the previous value, as indicated on Figure 2.26, P_1 is found to be non-dominated at the second iteration, and P_3 at the 4-th iteration.

Figure 2.26: Constraint re-definition

A second way of confirming the completeness of the non-inferior set is a two-phase procedure. In the first phase, one determines the non-inferior set, as described above. In the second phase, one reruns a series of optimizations using the CO_2 values of the initially defined non-inferior set step as the constraints. Thus, in phase two, the first optimization is subject to the constraint

$$CO_2 < CO_2(max)_1 - \delta$$

where δ is some very small increment. This optimization yields P_9 , already identified as being in the non-inferior set. Next one reruns the optimization subject to

$$CO_2 < CO_2(P_9) - \delta$$

which now yields P_1 , the previously unrevealed member of the non-inferior set. P_1 is added to the non-inferior set, and the process continues with the next optimization subject to

$$CO_2 < CO_2(P_1) - \delta$$

which will reveal P_5 as the next member of the non-inferior set. This procedure will ultimately reveal the correct non-inferior set $\{P_9, P_1, P_5, P_3, P_{10}\}$, as shown on Figure 2.27.

Figure 2.27: The two-step procedure

The weighting method solves a mathematical programming problem of the type

$$\text{Min}(1 - W(\text{CO}_2)) \text{ COST} + W(\text{CO}_2) \text{ CO}_2$$

over a range of values of $W(\text{CO}_2)$. When $W(\text{CO}_2)=0$, the result is the least cost solution (P9 on Figure 2.28). When $W(\text{CO}_2)=1$, the result yields that feasible solution that minimizes CO_2 emissions (which in practice might be an all hydro generation plan) — P1 of Figure 2.23. Intermediate values imply different slopes, and, as indicated on Figure 6.20, will yield intermediate points of the trade-off curve.

This method was used in the SCL case study, in which a large scale linear programming model, the Annual Optimization Model (ANN), was applied. Its objective function was defined as $\text{Min COST} + W(\text{CO}_2) \text{ CO}_2$, and solutions obtained for a range of values for $W(\text{CO}_2)$.

Figure 2.28: The weighting method

3. A REVIEW OF MCDM APPLICATIONS

There are literally hundreds of MCDM applications in a very wide range of fields¹. Even in the electric utility planning area there have been numerous applications -- ranging from screening of power plant sites to weighting-summation methods used to evaluate bids for providing additional supply resources from Independent Power Producers (IPPs). A complete review of this literature would be both tedious and necessarily superficial. Rather what seems indicated given the nature of the likely readership of this report is to review a representative sample of MCDM applications in electric utility planning or fields very closely related to it. In each area we review one or two applications in some detail, with briefer references to other examples: the primary emphasis in this review is to examine how these applications have adhered to the principles set forth in Chapter 2.

3.1 MCDM methods in Environmental Impact Assessment

MCDM methods have been used to derive indicators of environmental quality or impact since the early 1970's (Rau and Wooten, 1980; Hobbs, 1985). For instance, Dee *et al.* (1973) derived a weighting summation-based index of the environmental impact of water resources projects for use by the U.S. Bureau of Reclamation. Such indices can be used to examine the tradeoffs between environmental effects and economic and other criteria.

Qualitative consideration methods have long been a part of EISs, and range from checklists to more complex impact matrices. In fact, such methods have all the essential ingredients of a MCDM method: attributes, usually represented by the columns of the matrix; a set of discrete alternatives, usually represented by the rows of the matrix, and a value function $V_i(A_j)$. However, in such impact matrices, this value function often defined on a simple discrete scale, and also often displayed in shades of grey.

Puget Sound Area Reliability Plan EIS: a greyscale impact matrix

The strategy impact comparison matrix developed for the recent EIS for the Puget Sound Area Reliability Plan is a good example, and is illustrated on Figure 3.1. Four levels of impact are distinguished: high, moderate, low and minimal. These impact designations were developed by group consensus in a series of discussions by the team responsible for the EIS. The major problem with such an approach is perhaps not in the matrix itself -- for it may indeed be the case that a typical group of experienced environmental analysts can relatively easily reach a consensus about the environmental desirability of alternatives -- but in how the environmental rankings are integrated into the overall decision. In this particular study seven evaluation factors were used: the six shown on Table 3.1, and an environmental attribute. In this table, the strategy which ranks best in each attribute (i.e. each row) is shown in bold, and overall strategy

¹ See, e.g., the surveys by Chankong and Haimes (1983); Cohon (1978); Goicoechea *et al.* (1982); Hobbs (1979); United Nations (1988); and Zeleny (1982).

[illegible]**IDEA, Inc.**

Table 3.1: The Economic and technical evaluation summary

EVALUATION FACTORS	ALTERNATIVE STRATEGIES			
	Strategy 1	Strategy 2*	Strategy 3	Strategy 4
Net Present Value	\$67,000,000	<u>\$105,000,000</u>	-\$128,000,000	\$39,000,000
Sensitivity to Load Growth	\$196,000,000	\$126,000,000	<u>\$84,000,000</u>	\$102,000,000
Near Term Revenue Requirements	\$50,000,000	<u>\$25,000,000</u>	\$50,000,000	<u>\$20,000,000</u>
Long Term Revenue Requirements	\$75,000,000	<u>\$40,000,000</u>	\$110,000,000	\$105,000,000
Deliverability (1=Hi, 4=Low)	1.6	1.5	2.0	1.7
Reliability (1=Hi, 4=Low)	2.0	2.0	2.0	2.0
* Preferred Alternative				

Source: Bonneville Power Administration, Puget Sound Area Reliability Plan, Final Environmental Impact Statement, DOE/FEIS-0160, April 1992., p.4-21.

2 was selected as the preferred alternative.

It is interesting that the environmental rankings are not even included in this table. Yet from the environmental analysis, strategy 2 was only second best, and the preferred *environmental* alternative was strategy 3. There is no discussion at all as to whether the additional environmental impacts between strategies 2 and 3 were considered inconsequential.

Thus, trade-offs are entirely implicit, and nowhere in the EIS -- which runs to several hundred pages -- does the term trade-off even appear. Indeed, the "preferred" alternative was selected by a five-utility Plan Steering Committee, albeit with advice from a citizen review panel. In sum, it is extremely difficult to ascertain whether and how the environmental ranking influenced the actual decision.

A US Federal Energy Regulatory Commission approach: weighting summation.

The US Federal Energy Regulatory Commission has developed an environmental impact index to be applied to natural gas pipeline proposals (Stewart and Horowitz, 1991). The intention of the system is to fulfill the National Environmental Policy Act's mandate to "utilize a systematic, interdisciplinary approach which will ensure the integrated use of the natural and social sciences and the environmental design arts in planning and decision making."² The system includes 64 environmental attributes (Table 3.2). These attributes are combined into an environmental index using weighting summation (Section 2.9.1). The unique aspect of this system is how the weights for each of the attributes was derived.

The weight w_i for attribute i is calculated as follows (Stewart and Horowitz, 1991):

$$w_i = \sum_{k=1, \dots, 9} w_k d_k R_{ik}$$

² The same Act spurred utilities applying for nuclear plant operating licenses from the US Nuclear Regulatory Commission to use multicriteria approaches to compare alternative plant sites (Rowe *et al.*, 1979)

Table 3.2 : Environmental criteria in the FERC system

Water use and quality	Geology	
1. potable water supplies	24. active faults crossed by the pipeline	47. new pipeline right-of-way in orchards
2. municipal watersheds crossed by pipeline construction	25. areas with the potential for seismic soil liquefaction	48. temporary right-of-way in orchards and vineyards
3. toxic sediments at river or stream crossings	26. crossing of karst terrain	49. cranberry bogs crossed during pipeline construction
4. major water crossings	27. areas of potential landslide	50. construction of above-ground facilities in flood plains
5. minor stream crossings		51. inactive surface mines crossed by pipeline
6. crossings of lakes or reservoirs	Soils	52. new pipeline right-of-way near essential public facilities
	28 installing new pipeline in farmland	53. construction across hazardous waste sites
Vegetation and wildlife	29. installing pipeline loop in farmland	54. crossing of national and state trails
7. privately owned upland forest temporarily converted to open land for new or loop pipeline	30. soil with high erosion hazard crossed by pipeline	
8. privately owned upland forest permanently converted to open land	31. soil with moderate erosion hazard	55. compressor stations etc. within 1 mile of scenic area
9. forested or scrub-shrub wetlands converted to open wetland	32. prime farmland converted to non-farm use	56. new pipeline R-o-W within 1 mile of scenic areas
10. herbaceous wetlands cleared during construction	33. pipeline installed across annually cultivated fields with stony or clayey subsoil.	57. pipeline construction across federal or state wild and scenic rivers
11. significant fisheries	land use, recreation & aesthetics	58. sand dunes disrupted by construction
12. high-quality and cold-water fisheries	34. non-parallel right-of-way	59. commercial trawling or shellfish areas crossed
13. significant habitat crossings	35. new pipeline R-o-W within 50 ft of existing or planned residences	60. commercial anchorages and navigation channels crossed
14. national or state wildlife refuges crossed	36. compressor stations, etc. within 500 ft of residential areas	61. military reservations crossed by pipeline
15. national or state forest crossed	37. construction R-o-W within 50 ft of homes	air and noise quality
16. federally listed or proposed threatened or endangered species	38. natural or wild area crossed by pipeline	62. NOx emissions from compressor stations
17. critical habitat of federally-listed or proposed threatened or endangered species	39. natural or wild areas crossed by pipeline loop.	63. noise levels at noise sensitive areas
18. state-listed or proposed threatened, endangered or rare species.	40. active recreation areas crossed by pipeline.	64. abandonment and/or removal of pipeline facilities potentially contaminated with PCBs.
cultural resources	41. locally important, passive recreation area crossed by pipeline loop	
19. listed or eligible resources on the National Register of Historic Places (NRHP) in R-o-W.	42. locally important, passive recreation areas crossed by new pipeline	
20. listed National Historical Landmark (NHL) in R-o-W.	43. national parks crossed by pipeline	
21. cultural resources recommended as eligible for the NHL or NRHP.	44. state parks crossed by pipeline loop	
22. areas of high cultural resource sensitivity	45. state parks crossed by new pipeline	
23. areas of medium cultural resource sensitivity	46. disturbance to native American religious sites	

Source: Stewart and Horowitz (1991), Appendix, p.130-132.

where:

R_{ik} = the importance rating of attribute i in terms of importance dimension k . There are nine importance dimensions:

regional human environment (short, intermediate, and long term)

local human environment (short, intermediate, and long term)

natural environment (short, intermediate, and long term)

wd_k = weight for importance dimension k

Thus, the overall importance of an attribute is the weighted sum of nine dimensions of importance. The weights wd_k were initially assessed by the Analytic Hierarchy Process (AHP, Section 2.8). They were then adjusted so that the resulting w_i for a subset of twelve attributes were highly correlated with a set of attribute ranks directly chosen by two experts. Most of the resulting weight (62.5%) is assigned to natural environment, with the remainder split between regional and local human environment. In each of those categories, long term impact receives 57% of the weight, intermediate impact gets 29%, and short term effect receives the rest.

The R_{ik} are assessed for each of the 64 criteria using the following verbal scale: severe (all impacts likely to be considered unacceptable by FERC); considerable (some impacts unacceptable); moderate (impacts acceptable after mitigation and natural recovery); slight (impacts eliminated by mitigation and natural recovery); and none (no change in environmental quality). The verbal assessments for each criterion are then converted into an ordinal scale and multiplied by the wd_k 's to obtain the criterion's overall weight w_i . As an example, attribute $i=1$ (potable water) is judged to have "considerable" impact upon the short-term local human environment and a "moderate" impact upon the short-term natural environment. Its effect on the other seven dimensions is either "slight" or "none."

The main advantage of FERC's approach to weighting is that it provides a systematic way of assessing the importance of a very large number of criteria. Further, the authors made an effort to validate their system by comparing the weights with a ranking of attributes that experts chose directly. The major disadvantage of this system is that it appears that the ranges of the attributes were not explicitly considered when assessing weights. For instance, whether potable water impacts are important or not should depend on the amount of water involved. As another example, "long term", "intermediate term", and "short term" are vaguely defined, making the task of assessing the wd_k and R_{ik} ambiguous and difficult to replicate. Importance judgments made in the absence of such information are vacuous, and can be criticized for being arbitrary.

3.2 MCDM methods in resource bidding systems

What in effect are weighting summation MCDM systems have come into fairly wide use over the past few years for capacity bidding systems. Often such systems are referred to as "environmental rating" or "scaled scoring" systems. Mintzer *et al.* (1990) provide a typical rationale: "...scaled scoring systems are easy to use, facilitate comparisons of environmental impacts measured in different units, and are more comprehensive than economic analysis ... using weighted scaled scores, technologies (or a combination of technologies contained in an energy futures scenario) can be ranked from the least to the highest cost on the basis of all economic, environmental and social costs that can be quantified."

These are very impressive claims, but, unfortunately, many of the applications have paid little attention to methodological issues, with the consequence that the validity of the results is highly uncertain. New York is a typical case, and illustrates some of the issues.

A first general problem is consistency. In some bidding programs approved by the New York Public Utility Commission (NYPUC), the bid prices are adjusted to reflect externalities -- for example SO₂ emissions are valued at \$832/ton, and CO₂ at \$1.1/ton, valuations that are based on control costs. In others, the bidding programs require the use of so-called "environmental scoring forms". Table 3.2 presents such an example as used by the Niagara Mohawk utility in New York: there the "weight" assigned to SO₂ is 7, and to carbon dioxide 3³. The claim is made that the scoring forms are supposed to reflect "the environmental perceptions in each utility service area" (Putta, 1990); yet it seems fairly clear that "weights" can be manipulated in such a way as to produce almost any desired result.

The problem of validity in scaling is illustrated in Figure 3.2: the scaling function is arbitrary, and is unlikely to have interval properties. It is in fact quite common to encounter systems in which the physical attribute is converted into a step function corresponding to some number of points on a scale -- these are sometimes called "binned ranges".⁴ However, in the absence of some specific threshold effects that would justify some deviation from the presumption of linearity, much better would simply be to maintain the original physical attribute values, and simply normalize to a 0-1 scale.⁵ The important point here is that transformations ought not to be arbitrary.

³ Although at first glance the relative weights (SO₂ to CO₂) of 7 to 3 appear inconsistent with the NYPUC valuations, when one takes into consideration the scales used, the ratios are comparable. a reduction of 7 lb of SO₂ are seen to be worth $7 \times 5 = 35$ points, while a reduction of 1,500 lb CO₂ is seen to be worth $3 \times 5 = 15$ points. Thus 1 lb SO₂ is seen to be equivalent to 650lb CO₂ (which follows from the value w that satisfies the ratio $7/35 = w/1750/15$), which is not totally inconsistent with the \$1.1 to \$832/ton valuations of the NYPUC. The important point here is that the lb to lb trade-offs are not at all obvious when choosing weights!

⁴ Goldman *et al.* (1992) provide a detailed evaluation of binned v. linear scales (also using the Niagara Mohawk system as an example), and conclude that this may introduce serious distortions when used in bidding systems.

⁵ Uncertainty in bidder performance is explicitly quantified as probability distributions (which has apparently never been done), then the scaling might be non-linear due to risk attitudes (see section 2.7 on utility functions).

Environmental Attributes	Weight (W)	Points (P)						Score (W x P)
		0	1	2	3	4	5	
Air Emissions								
Sulfur oxides (lbs/MWh)	7	>6	4.0-6.0	2.5-3.9	1.5-2.4	0.5-1.4	<0.5	
Nitrogen oxides (lbs/MWh)	16	>6	4.0-6.0	2.5-3.9	1.5-2.4	0.1-1.4	<0.1	
Carbon dioxide (lbs/MWh)	3	>1500	1050-1500	650-1049	250-649	100-249	<100	
Particulates (lbs/MWh)	1	>0.3	0.2-0.3	0.1-0.199	0.05-0.099	0.01-0.049	<0.01	
Water Effects								+
Cooling water flow (Annual intake as % of lake volume)	1	80-100	60-79	40-59	20-39	5-19	<5	
Fish protection	1	None		Operational restrictions		Fish protection provided	No public water used	
NYS water quality classification of receiving water	1	A or better	B	C+	C	D	No water use or municipal water/wastewater utilized	
Land Effects								+
Acreage required (acres/MW)	1	0.3-0.5	0.2-0.29	0.1-0.19	0.05-0.09	0.01-0.05	<0.01	
Terrestrial	1	Unique ecological or historical value		Rural or low density suburban		Industrial area	No land used	
Visual aesthetics	1	Highly visible		Within existing developed area		Not visible from public roads		
Transmission	2	New OH >5 miles	New OH 1-5 miles	New UG >5 miles	New UG 1-5 miles	Use existing facilities	Energy conservation	
Noise (Leq - background L90)	2	5-10			0-4.9		<0	
Solid waste disposal (lbs/MWh)	2	>300	200-300	100-199	50-99	10-49	<10	
Solid waste as fuel (% of total Btu's)	1	0	1-30	31-50	51-80	81-90	91-100	
Fuel delivery	1	New RR spur	Truck & existing RR	New pipeline	Barge	Use existing pipeline	No fuel use	
Kilometers from sensitive receptor area (per 6 NYCRR 225.1.1 (b)(4))	1	<10	10-39	40-69	70-100	>100	Energy conservation	
Total Score (210 Max.)								=

Table 3.3: The environmental scoring form used by Niagara Mohawk
source: Putta (1990)

Figure 3.2: The attribute scaling function for CO₂ emissions in the Environmental Scoring System for Niagara Mohawk (New York).

Problems of attribute definition abound. For example, what exactly is the environmental consequence that is expressed by "land use" in acres (or acres/Mw)? As noted previously, differences in market value of the land should be reflected in the cost attribute, and the opportunity costs of lost production (say in the case of agricultural land) are also readily quantified in economic terms. But if the attribute reflects just non-market values, these are already captured in the attribute "terrestrial", in which sites of unique historical or ecological value are penalized. Finally, the "transmission" criterion may also double-count land area consumed. In short, it is unclear why acreage has any relevance.

There are similar problems with the attribute "fuel delivery". The scale used, ranging from new railroad spur to no fuel use, correspond very closely indeed to the costs of the fuel supply connection, which should be incorporated in the system cost attribute. If there are particular air emission problems associated with the construction of such new transportation links, then those should be included in the air emissions attribute. In any event, the requirement for preference independence is unlikely to be present: one might be motivated to put a higher weight on the "fuel delivery" criterion if barge transport takes place in relatively pristine rivers than if the water is instead very polluted.

3.3 MCDM Methods in Transmission and Distribution System Planning

T&D planning involves tradeoffs among system cost, reliability, and environmental considerations. Although MCDM methods have been applied less frequently in this area than to other aspects of utility planning, the following two examples illustrate how such methods can help the T&D planner. One focuses on tradeoffs between cost and reliability in system design, and the other looks at cost-environment tradeoffs in transmission line routing.

Analyzing Cost-Reliability Tradeoffs

Tradeoffs between cost and reliability are critical in T&D design, yet are rarely examined explicitly. Reasons for this include engineers' lack of training in probability and reliability and the absence of accepted methodologies for estimating the reliability of transmission or distribution network. Generally, standards based upon accepted engineering practice are used to determine when, where, and what type of circuits need to be added. AC load flow models are often used to simulate system performance under a set of system contingencies to see if standards are violated. For example, would voltages remain within an acceptable range if line A was to fail under the proposed system design?

However, we expect greater acceptance in the future of models for exploring cost-reliability tradeoffs. Ramirez-Rosado and Adams (1991) show one way of doing this for a distribution design problem. Below, we fit their procedure into the 10-step procedure of Figure 2.1.

Step 1: attribute selection

Two criteria are considered:

1. the sum over time and busses of weighted deviations of bus voltages from a target voltage (20 kV in this case), where the weights can depend on bus location and time of the deviation. This is similar in spirit to the goal programming amalgamation method of Section 2.9.2, and
2. the cost of line construction

Step 2: definition of options

The options considered are possible combinations of lines connecting seven load points with two substations, and the resulting power flows and voltages. The feasible configurations are defined by a set of constraints in a nonlinear mathematical program. A math programming approach has the advantage that it permits the planner to implicitly consider a very large number of alternatives, lowering the likelihood that very good solutions will be overlooked.

Step 3: quantifying impacts

Impacts of each configuration upon the two criteria are automatically calculated by the mathematical program, which expresses the two criteria as mathematical functions of the decision variables (line locations and bus voltages).

Step 4: screening

Alternatives are not explicitly screened. Instead, a few configurations were directly chosen by two amalgamation methods in Step 9.

Step 5: trade-off curves

Tradeoff curves are not explicitly developed, but could easily be obtained by weighting and summing the two criteria, minimizing that sum, plotting the resulting solution, and then permuting the weights. However, in Figure 3.3, we display some of the non-dominated solutions they generated using two

Figure 3.3: Voltage penalty v. line cost
(based on data in Ramirez-rosado and Adams (1991))

different amalgamation methods, described below. A more careful analysis would have generated more points to give the planner a fuller picture of the options that are available.

Examination of the figure reveals that only two of the six plans are truly competitive: those labeled "Final STEM Solution" and "Goal programming, $P=1$." These two solutions would make up the knee set (see Section 2.6). Although the other solutions are undominated, they would certainly not be in the knee set because their advantages are very weak. For instance, the "Goal programming, $P = \text{infinity}$ " point is much worse in cost than the " $P=1$ " point but has only a negligibly better voltage penalty.

Step 6: dominance analysis

No dominance analysis is performed. If the mathematical program is correctly formulated, only non-dominated designs will be generated.

Step 7: scaling

The only scaling performed is as follows:

- Voltages above 19kV receive a small bonus (in the form of a negative penalty) in the first criterion
- Voltages below 19kV but above 18.6kV are penalized more heavily per unit of deviation
- Constraints in the mathematical program prohibit any bus voltage from falling below 18.6kV

Step 8: weighting

Weights are used by the two amalgamation methods applied. They appeared to be arbitrary, as no justification is given for them. In an actual application, planners would want to give careful consideration to the monetary value of additional reliability when choosing weights.

Step 9: amalgamation

Two amalgamation methods are applied:

- STEM, an iterative mathematical programming method that generates a solution, asks if the planner likes it, and then, if the planner is dissatisfied, generates another solution based on the criterion that the planner wishes to improve.
- Goal programming (Section 2.9.2), where the goals are set equal to the best possible values of the two criteria (-4 for voltage deviations and 787 for cost). Two values of the exponent P were considered: 1 and infinity. As we point out in Chapter 2, if $P=1$ and if goals are set equal to the best possible values, then this method will yield the same result as simply minimizing the weighted sum of the two criteria (i.e., weighting summation, Section 2.9.1).

In both cases, a nonlinear programming method must be used to find the feasible configuration which has the best value. The values of the two criteria for the solutions from these methods are shown in Figure 3.3. The two goal programming solutions are very similar, but STEM yields a much more expensive system with a better voltage profile.

Step 10: resolving differences

Not addressed in the published study.

The main advantage of the above MCDM approach is that its mathematical programming formulation allows a huge number of alternatives to be implicitly screened and evaluated. The disadvantage of the implementation is that Steps 8 and 9 seem arbitrary, with no justification given for the weights or amalgamation methods chosen. Without such a justification, it is difficult to be confident that the supposedly "best" solution yielded by a method really is the most preferred by the planner. However, since there are just two objectives here, it would be a simple matter to generate a range of options, as in Figure 3.3. Then the options would be presented to the responsible decision makers who could then directly decide what points are most acceptable. This is the approach we would recommend for problems with just two or three criteria.

Transmission Line Routing

Weedy (1989) reviews various methods for delineating corridors for new transmission facilities. One method which has found use since the 1970's is the network approach (e.g., Economides and Sharifi, 1978; Guidinger, 1987), in which a route is chosen to minimize a weighted sum of environmental, cost, and other impacts. Below we summarize how this method is typically applied in terms of the nine steps of the general MCDM procedure.

Step 1: attribute definition

The criteria may number in the dozens, and can be grouped as follows:

- damage to natural ecosystems
- conflict with existing and proposed land uses
- potential for sharing of right of ways
- impact on culturally significant features
- visual impact
- potential for exposure to electromagnetic fields (EMF), perhaps measured using a proxy such as population density within a certain distance, or population inversely weighted by distance
- cost
- transmission losses

Figure 3.4: Typical network structure showing alternative transmission links between nodes 2 and 39.
source: Weedy (1988) p.222

The most salient environmental issue today is EMF (Gulliver and Vito, 1993; DeCicco *et al.*, 1993), although aesthetics are also very important.

Step 2: definition of options

Options may be a few discrete routes predefined by planners. However, in network routing methods, they are defined as possible combinations of segments. As an example of the latter, Figure 3.4 shows a network for which there are many dozens of possible routes between nodes 2 and 39.

Step 3: quantifying impacts

Impacts are first quantified by tallying the environmental, financial, and other impacts of a line of the prescribed voltage for each segment. Then the total impact of entire route is implicitly defined as the sum of the impacts of the constituent segments.

Step 4: screening

Screening occurs in the selection of the segments and connections to be considered (Figure 2); for instance, no segments might be defined for some areas because of their high population densities.

Step 5: trade-off curves

Tradeoff curves are rarely generated by network methods, mainly because there are so many criteria. However, it might be possible to define a single environmental index that could be plotted against cost; the challenge would be to define an index that could be interpreted meaningfully in that context.

Step 6: dominance analysis

Dominance analysis is not explicitly performed. Instead, an amalgamation method is used to generate one or more routes (sequence of segments); these routes will be nondominated if weighting summation is used.

Step 7: scaling

Scaling is usually linear in network models. For instance, if segment A has twice as many meters of wetland to be traversed as segment B, then segment A would be penalized twice as much.

Step 8: weighting

The various criteria for each segment are assigned weights. All the considerations we outlined in Chapter 2 concerning appropriate choices of weights apply here: i.e., weights should be ratio scaled and represent the willingness of decision makers to trade off one criterion for another.

Step 9: amalgamation

A sequence of segments minimizing the weighted sum of criteria scores is chosen, usually by a linear programming or network solution method. This is a weighting summation amalgamation method, but with a difference: because the number of segments is not the same for all routes, the individual criteria scales must be ratio scaled, not merely interval scaled.

The reason for this can be illustrated by an example. Say there are just two routes between nodes 2 and 39. The first route traverses segments A and B, and the second route consists only of one segment, C. The criteria must have a non-arbitrary zero point (be ratio scaled), or else the ranking of the two routes could be changed by altering the arbitrary zero point. Consider, for example, the following scores for environmental impact for the three segments:

$$I_A = 2; I_B = 4; I_C = 7$$

The first route would then have the lowest total impact score ($2+4=6$, versus 7 for the second route). But if these impacts were only interval scaled, we could add an arbitrary constant to each, say 2. Then the new impact scores would be:

$$I_A = 4; I_B = 6; I_C = 9$$

and now the second route would have the lowest score (9, versus $4+6=10$ for the first route). This demonstrates that the criteria scales cannot have an arbitrary zero point.

Step 10: resolving differences

Not addressed in the published study.

The advantage of the network approach to routing is that it can consider many possible routes. Its disadvantages are:

- with many criteria, it is difficult to plot and visualize tradeoffs
- weighting summation is the most appropriate amalgamation method, but criteria must be measured on a ratio scale.

3.4 MCDM Methods in Supply Capacity Planning

We describe two basic approaches to multicriteria analysis of supply plans. One approach contrasts a few discrete plans and uses amalgamation methods to rank them. The other incorporates a MCDM method within a linear or dynamic programming model for capacity expansion.

Ranking Discrete Supply Plans

Traditional supply-side planning consists of defining alternative supply plans that meet future demands at a given reliability, and then choosing the plan that minimizes the present worth of cost for a given set of cost assumptions. Today, supply planning is more sophisticated in at least two ways. First, planners consider several criteria other than cost. Second, they explicitly recognize uncertainty in evaluating plans. To handle these complications, multicriteria methods, such as those described in Chapter 2, have been used to rank alternative supply plans.

Lootsma *et al.* (1990) use such an approach to rank electricity supply plans for the Netherlands. Rankings were made by each of 18 experts from universities, government ministries, and energy companies. This application is summarized below:

Step 1: attribute selection

The nine criteria considered are: average cost (cents/kWh); total cost; total SO₂, NO_x, and CO₂ emissions; domestic natural gas reserves; profits of the domestic natural gas industry; percentage of total energy supply by gas and oil; and installed nuclear capacity.

Step 2: definition of options

Three general strategies are compared: nuclear, coal, and gas.

Step 3: quantifying impacts

The performance of each strategy on each criterion is estimated for the years 2000 and 2010. Further, three scenarios are considered in each year:⁶

- Low oil prices, GNP, and electricity demand
- Medium, oil prices, GNP, and electricity demand
- High oil prices, GNP, and electricity demand

Step 4: screening

No screening procedures are described in the case study.

Step 5: trade-off curves

The authors explicitly reject using tradeoff curves or other devices to display the tradeoffs, asserting that decision makers would be confused by all that information and would not be able to make a reasoned choice. Instead, the authors emphasize Steps 7, 8, and 9 of the MCDM process. We believe, however, that tradeoff analysis is a useful complement to weighting and amalgamation, because it can help make decision makers more confident in the values they choose for weights, and in the resulting rankings.

Step 6: dominance analysis

No dominance analysis is performed.

⁶ These scenarios are logically similar to the three scenarios considered in the Centerior case study: low stress, reference, and high stress; see Section 2.2.

Step 7: scaling

An unusual scaling method is used in which decision makers are asked to perform pairwise comparisons of the strategies for each criterion. For example, the total SO₂ emissions of the nuclear strategy in the year 2000 under the low scenario would be 180,000 tons, while the coal strategy would yield 190,000 tons. The decision maker is asked which of the two values he prefers, and whether the preference was absent (score of 0), weak (2), strong (4), or very strong (6).

A scale or value function $v_i(A_{ij})$ for the attribute is then derived from those judgments as follows. First, the authors assume that perception is logarithmic; that is, statements of relative value of two alternatives are proportional to the logarithm of their difference. For instance, if a person states that A is twice as preferred as B, but four times as preferred as C, then the actual value ratios are e^2 for A:B and e^4 for A:C. The rationale for this is the finding by psychologists that this is how people compare physical magnitudes, such as brightness of light or loudness of sound (Stevens, 1966). This means that $v_i(A_{ij})$ is of the form:

$$v_i(A_{ij}) = \exp\{C x_{ij}\}$$

where $\exp\{x\}$ is 2.7128.... raised to the x power, and C is an arbitrary constant. Unfortunately, the choice of this arbitrary constant actually changes the ranks of alternatives, a result that understandably disturbs the decision makers!

The A_{ij} 's were not the original criteria scores (e.g., tons of SO₂), but rather are a scale that is statistically estimated from the pairwise responses. The A_{ij} are the values that result in the best (least squares) fit to the responses to all the pairwise comparisons.

Step 8: weighting

Weights are chosen by a procedure similar to Step 7, where each participant is asked to compare two criteria at a time, stating which is more preferred (more important) and to what extent.

Step 9: amalgamation

The resulting scales and ranks are combined by the following amalgamation rule ⁷

$$\text{Score}(\text{strategy } j) = \prod V_i(A_{ij})^{w_i}$$

where A_{ij} is the score on criterion i for strategy j , and w_i is that criterion's weight. This is called the "power law" method and it assumes logarithmic perceptions of relative importance of criteria. The score for each strategy is then averaged over the three scenarios by a complicated procedure that allows the participants to give the probability of each scenario as a range rather than a single point. For instance, someone could say that the probability of the low scenario was between 0.2 and 0.4. Decision makers may be more comfortable stating ranges than precise values.

Rankings are then produced for each of the eighteen experts. For a value of 0.5 for the arbitrary constant C , ten experts rank the nuclear strategy as best. Gas is ranked first by six experts, while coal is best only for one of the experts. The experts examined the scores, and their sensitivity to the choice of C . Several of the experts expressed their dislike of the dependence of the results on C .

For comparison purposes, weighting summation is also applied with the same scales and weights. Table 3.4 summarizes a set of strategy scores for each scenario and for two different assumed values of C in the exponential equation. For example, the nuclear strategy receives a weighting summation score of 34 under the medium scenario if $C = 0.5$. The sensitivity of the results to the arbitrary constant C is obvious: nuclear and coal are tied under the power law in the medium scenario if $C = 0.5$, but nuclear is far inferior if $C = 1$. The weighting summation ranks depend even more strongly upon C . The table also shows that the power law yields very different rank orders than weighting

⁷ The actual rule used by Lootsma *et al.* (1990) is somewhat more complex, but is equivalent to the one we present.

summation. Although the top option is usually the same, the relative ranking of coal and nuclear is often switched.⁸

Table 3.4 Netherlands Supply Planning: Power law vs. weighting summation scores for each strategy under different scenarios and different scaling constants C.

Oil Price/ GNP/ Demand Scenario	Amalgamation Method	Constant C = 0.5			Constant C = 1		
		Nuclear	Coal	Gas	Nuclear	Coal	Gas
Low	Power Law	29	34	37	7	43	50
	Weighting Summation	33	32	34	21	38	41
Medium	Power Law	31	31	39	9	34	57
	Weighting Summation	34	31	36	21	36	43
High	Power Law	30	33	37	7	39	54
	Weighting Summation	35	31	35	21	37	42

source: Lootsma *et al.* (1990).

Step 10: resolving differences

Not addressed in the published study.

A strength of this MCDM application was a careful laying out of the consequences of each strategy under several scenarios at different times in the future. Another strength was the attempt to measure the preferences of a diverse group of people. The major disadvantages of the methodology used in this case was the complexity and opaqueness of the scaling, weighting, and amalgamation procedures, and the dependence of the results on an arbitrary constant. A complex procedure such as the power law can be justified only if there is evidence that it is more valid than the simpler weighting summation method. However, there was apparently no attempt to confirm whether the participants' preferences conformed to the logarithmic model, or if the calculated weights were consistent with tradeoffs the people were willing to make. We believe that use of a model with unclear assumptions and complex computational procedures is less likely to be accepted by decision makers and the public.

Multiple Criteria Analysis of Supply Plans Using Linear and Dynamic Programming

As an alternative to considering a handful of predefined plans, many utilities use linear and dynamic programming models to sort through a large number of possible sequences of supply additions (Turvey and Anderson, 1977). Most applications of such models have emphasized the minimization of the present worth of cost. However, environmental and other non-cost objectives can be added to such models to allow the planner to generate alternatives representing a range of tradeoffs. For instance, Kavrakoglu and Kiziltan (1983) and Quaddus and Goh (1985) modified linear programming capacity expansion models in this manner (as we do in our Seattle case study (Hobbs and Meier, 1994)). The World Bank has also applied MCDM to electric sector planning in Costa Rica and Hungary (see Section 3.8). Below we summarize the dynamic programming-based method of Yang and Chen (1989) which they applied to the Taiwan Power Company system.

⁸ However, Lootsma *et al.* (1990) state that they have found in other projects that which amalgamation method is applied seems to make little difference when there are five or more alternatives.

Step 1: attribute selection

The selected attributes are thirteen in number, and include:

- Present worth of generation cost
- Environmental impact (SO₂, NO_x, and particulate emissions, in parts per million in the stack gas multiplied by the number of generating units times the number of years each unit is in service)
- Risk of plant disaster (measured as the cumulative capacity, in MW-years, of each fuel type: nuclear, hydro, coal, oil, and gas)
- Fuel import vulnerability (measured as the Mwh production by nuclear, coal, gas, and oil)

As we point out earlier in this report, total emissions are, at best, only a rough indication of environmental impact, as their effect depends on plant location. For instance, a plant located so that prevailing winds blow emissions out to sea might have minimal impact, even if emissions are large. The definitions and rationales of the last two groups of attributes is unclear. For instance, given a fixed nuclear and hydroelectric power capacity, why would additional oil capacity be considered to increase the "risk of plant disaster"?

Step 2: definition of options

The alternatives considered are different capacity expansion paths that could meet assumed demand growth through the year 2005. The alternative paths are implicitly defined by the dynamic programming network, which describes what additions can be made in which years.

Step 3: quantifying impacts

Impacts are automatically quantified by the dynamic program. Production costing runs for each year determine the amount of fuel cost, emissions, and "fuel import vulnerability." The capacity additions along the chosen path determine the capacity costs and "risk of plant disaster."

Step 4: screening

The only screening that is performed occurs when the dynamic programming network is set up. In defining the network, some combinations of generating plants will be excluded in order to limit the size of the problem; the excluded combinations will generally be those that the planner believes are unlikely to be economic (e.g., all baseload plants) or will fail to meet reliability criteria.

Step 5: trade-off curves

Trade-off curves could be generated by solving the dynamic program under several different weight sets. This is what Yang and Chen (1989) do; Figure 3.5 shows an example in which solutions from two different weight sets are compared. Present worth of cost, ppm of SO₂, and total Mwh generated from oil are shown. Perhaps because there are thirteen criteria, the authors do not present explicit tradeoff plots.⁹

Step 6: dominance analysis

No explicit dominance analysis is performed, but the dynamic program guarantees that a non-dominated solution will result from minimizing a weighted sum of the criteria.

Step 7: scaling

Linear scales are used for all attributes.

⁹ PROVIEW/PROSCREEN, another dynamic program-based expansion package, allows the user to define several different resource portfolios for the last year of the time horizon, and then uses tables to display the tradeoffs among their costs, rates, emissions, and other criteria (Energy Management Associates, 1993).

(a) Weights: Cost=.12, Environment=.16, Risk=.21, Import=.51
(Nuclear Engineers Group)

(b) Weights: Cost=.16, Environment=.13, Risk=.53, Import=.18
(Energy Economists Group)

Figure 3.5: Comparison of plans
source: Yang and Chen (1989)

Step 8: weighting

The Analytic Hierarchy Process (see Section 2.8) is applied by three different groups: nuclear engineers, power engineers, and energy economists. Each group has six people. First, the participants compare the four categories of attributes against each other using AHP's 1/9 to 9 scale. Then the attributes within the risk category are compared with each other. The four import attributes are also compared. Finally, the authors distribute each group's environmental weight among the three pollutants in proportion to estimates of the external costs they impose on the public. In addition, a poll was also taken of members of the public to determine a fourth set of weights.

The disadvantages of the criteria as defined in Step 1 become apparent when the AHP questions are phrased. For example, how can someone respond to the following question meaningfully?

Is gas 1/9, 1/7, 1/5, 1/3, 1, 3, 5, 7, or 9 times as important as coal in terms of the "plant disastrous risk" objective?

This question is too vague to be meaningful. Yet somehow people responded; for instance, the nuclear engineers, on average, found that gas is 80% as important as coal in terms of risk. This may simply mean that people will supply numbers when asked, even if they signify nothing.

Step 9: amalgamation

Amalgamation is done four times by weighting summation, once for each group's weights. The amalgamation is performed by defining the weighted sum of the criteria as a single objective function to be minimized by the dynamic program.

Step 10: resolving differences

As shown on Figure 3.5, the study addresses directly the fact that groups holding different values will arrive at rather different results.

The advantage of this approach to generation planning is that it allows some tradeoffs among various criteria to be considered. For instance, a comparison of Figures 3.5(a) and (b) show that the latter solution's higher weight on "plant disastrous risk" would result in much less reliance nuclear power. Under Yang and Chen (1989)'s assumptions, this decrease in risk would cost over five billion dollars.

However, the use of AHP's ambiguous questions in this context means that we suspect that the weights do not truly represent the values of the groups. If a recommendation for a plan is to be based on this type of methodology, each person should apply more than one weighting method as a check to ensure that the weights are valid. Preferably, a method that forces users to explicitly consider the units of the various criteria should be applied

3.5 MCDM Methods in IRP

Integrated resource planning considers not only the supply options included in the studies of the last section, but also DSM. As in supply studies, MCDM methods can be used in two general ways:

1. to rank a set of predefined plans
2. to create objective functions for linear programming or dynamic programming models. The models can then be used to generate tradeoff curves and to select an plan that represents an acceptable compromise among the objectives.

As an example of the mathematical programming approach, the dynamic programming package PROVIEW/PROSCREEN, which has been adopted by over 60 utilities, has the capability of generating tradeoff curves and optimizing a mix of objectives (Energy Management Associates, 1993). Cincinnati Gas & Electric used that model to compare several different strategies for complying with the SO₂ control program of the 1990 US Clean Air Act (Bloemer and Knue, 1993). A strategy consists of an commitment of resources (DSM, supply expansion, fuel choice, and emissions control) for the first few years of the planning period, followed by later resource commitments that can depend on what scenario of loads and prices for fuel, SO₂ allowance prices, and control equipment is realized. The plans are ranked in terms of the sum of their expected costs and emissions, which were equally weighted. The expectation is taken over the possible scenarios, each assumed to have the same probability.

A typical application of MCDM in IRP is the one used by a Wisconsin utility to rank four alternative resource plans (Hanson *et al.*, 1991).¹⁰ The nine steps of MCDM analysis are executed as follows in the Wisconsin study:

Step 1: attribute selection

Nine attributes are selected (Table 3.5). "AFUDC" indicates interest costs accumulated during plant construction; its ratio to net earnings is considered to be an indicator of "earnings quality." One quarrel we have with the list of attributes is that conservation and coal are given as criteria. They are more properly thought of as alternatives; it is more appropriate to define attributes that describe what it is about those options that planners like or dislike (e.g., emissions, public relations, etc.). There is a danger of double counting, in that conservation might be given a high weight because we want to avoid the pollution that results from fossil fuel use, but at the same time high weights are also directly given to SO₂ and solid waste.

Table 3.5: Summary of attribute values

alternative:	1	2	3	4
1 PV(customer cost), 1986 \$million	16.6	16.49	16.92	16.86
2 levelized rates, cents/1986 kwh	6.16	6.28	6.54	6.52
3 AFUDC/net income (average %)	15.7	10.5	6.3	9.2
4 SO ₂ emissions (average tons/year)	76.87	74.43	70.42	71.88
5 solid waste (1000 tons/year)	1,924	1,288	1,168	908
6 new employment (jobs created)	178	206	395	375
7 conservation (Mw)	0	150	445	445
8 baseload plant (in-service date)	1995	2006	2010	2008
9 coal consumed (million tons)	174	162	154	151

source: Hanson *et al.* (1991), p.41.

¹⁰ See also Hobbs and Meier (1994) for our own more detailed case study of Seattle City Light, and Meier and Hobbs (1994) for our own case study of Centerior Energy.

Step 2: definition of options

Four alternatives are considered:

1. Retire four existing 130 Mw coal units and implementing no DSM; as a result, a new coal unit will be required relatively soon
2. Retrofit those four units with atmospheric fluidized bed combustion (AFBC) and implement 150 Mw of DSM
3. Same as 2, except implement 445 Mw of DSM
4. Retrofit two units with AFBC and implement 445 Mw of DSM.

Step 3: quantifying impacts

Production costing and financial models are evidently used to quantify the attributes for each alternative (Table 3.5).

Step 4: screening

No screening procedures are discussed by the authors, although presumably some screening approach is used, for example, to determine the amounts and types of DSM in each plan.

Step 5: trade-off curves

With so few alternatives and so many criteria, it is probably easiest to present tradeoffs in the form of a table, rather than graphically. Table 3.5 shows that there are several interesting tradeoffs. Plan 1 has the cheapest rates, but Plan 2 has the lowest present worth of costs (due to its inclusion of relatively cheap DSM). The higher cost DSM plans (3 and 4) yield lower emissions and coal use, more employment, and less AFUDC. Clearly, the relative importance of the attributes will make a difference in the ranking of the plans.

Step 6: dominance analysis

Each of the four plans is non-dominated.

Step 7: scaling

Scaling is accomplished using the risky utility function approach of Keeney and Raiffa (1976) (see Section 2.7). Seven utility managers and outside experts were interviewed to obtain single attribute utility functions, along with the weights in the next step. The authors report that there are considerable differences in the shape of the utility functions for the attributes chosen by different people--consistent with what we find in the Seattle case study (Hobbs and Meier, 1994). As an extreme case, some people think that more conservation is undesirable (has lower utility) because it may be less reliable, while others believe that more conservation has higher utility because of the cost savings and experience that results. (Note that this indicates there is double counting between the conservation and cost attributes!)

Step 8: weighting

The authors say they used the methods of Keeney and Raiffa (1976) to weight the attributes, implying that the indifference tradeoff method was probably used. The weights differ greatly among the participants. For example, one person put most of his or her weight on levelized rates, while non-economic criteria (numbers IV-IX) were given most of the weight by a majority of the other participants.

Step 9: amalgamation

Either weighting summation (Section 2.9.1) or a multiplicative utility function (Section 2.9.3) is used to rank the alternatives. As a result, four of the seven experts rank plan 2 as best. Plan 4 is chosen by two experts, while the expert who placed most of the weight on rates picks plan 1. The heavy weighting of non-economic criteria was the reason for those results; if instead zero weight is assigned to those attributes, then plan 1 would be found best by six of the seven participants.

Step 10: resolving differences

Not addressed in the published study.

This study had the advantage of using a relatively transparent amalgamation method. Different value judgments, such as the expressed willingness to tradeoff attributes, can be easily translated into changes in scales, weights, and plan ranks. Such sensitivity analyses are much more difficult with, for example, the complex MCDM method used by Lootsma *et al.* (1990).

We have three concerns about the Hansen *et al.* (1991) study. First, as we mentioned, there is obvious double-counting in the attributes, resulting in a bias towards conservation. Second, a greater range of alternatives should have been explored, perhaps with the aid of a linear or dynamic programming capacity expansion model. For example, an amount of DSM between the extremes of 120 Mw and 445 Mw might be preferred by most of the experts.

Our final concern is that the ranks should have been presented to the experts to see if they agreed with them. Such feedback is useful for two reasons:

1. If the decision makers agree with the rankings, this provides evidence that the method is valid and trustworthy.
2. If the decision makers initially disagree with the ranks, then the reasons for the disagreement can be explored. There are two possible outcomes of this exploration. One is that the users adjust their scales and weights to better reflect their values. As we stated before, scales and weights are often inconsistent and unreliable because the decision makers' values themselves are not fully coherent (Fischhoff *et al.*, 1979). The advantage of a MCDM procedure with these types of checks is that it helps decision makers to think hard and make their values more consistent.¹¹ The other possible outcome is that, upon considering the reasons for the MCDM method's ranking, the decision maker gains some insight into the tradeoffs and decides that the ranking is reasonable after all.

¹¹ Of course, there is the danger that a user will arbitrarily adjust the parameters until the results agree with his or her prejudices about which option is best.

3.6 MCDM Applications in System Dispatch

There are several objectives that a power system dispatcher might consider. These include fuel cost, system security, minimization of use load controllers, and emissions. For instance, Maekawa *et al.* (1992) consider the problem of trading off fuel cost, transmission line overloads, and AFC generator margins (the difference between generator output and its upper and lower limits) in real-time dispatch. They use an amalgamation method based on fuzzy set theory to choose an optimal dispatch schedule. The scales in this method are termed "fuzzy set membership functions", which are then combined using an amalgamation method that the authors do not describe. We criticize this application for the same reason that we criticized the distribution planning method of Ramirez-Rosado and Adams (1991): with only three objectives, we believe that decision makers would appreciate being presented with a range of possible dispatch schedules in the form of tradeoff curves. A single solution resulting from a complex MCDM method would (deservedly) be distrusted.

Figures 3.6 and 3.7 present examples of such tradeoff curves. In Figure 3.6, non-dominated strategies for real-time operation of a power distribution system have been generated for two objectives (Muslu and Anderson, 1989):

1. minimization of fuel cost (in \$/day) (variable Z_1 on Figure 3.6)
2. minimization of curtailment of air conditioning loads that are participating in load control program (in MWh/day) (variable Z_2 on Figure 3.6).

Curtailling load lowers costs, but also results in customer complaints. Presented with the information in this curve, the operator can decide what level of curtailment represents a reasonable compromise between the two objectives.

Figure 3.7 focuses on cost-emission tradeoffs in dispatch (Heslin and Hobbs, 1989). Historically, emissions have not been considered in operating power systems. However, emissions dispatch is increasingly viewed as a cost-effective strategy for decreasing the impacts

Figure 3.6: Reliability-fuel cost trade-off curve

source: Muslu and Anderson (1989), Figure 6.

of power systems, and a number of models have been developed for exploring the tradeoffs between emissions and costs (Talaq *et al.*, 1994). "Emissions dispatch" is the operation of a power system in such a way that emissions are lowered, but costs are increased. Figure 3.7 shows tradeoffs between SO₂ emissions and cost for the state of Ohio, averaged over the year. On the lower right is the least-cost dispatch point, where fuel costs are minimized without regard to emissions. By shifting the plant merit order and, in some cases, changing fuel sources, emissions can be decreased, but at a cost.¹²

Figure 3.7: Ohio emissions-cost trade-off curves, with and without fuel switching

source: Heslin and Hobbs (1990), Figure 2.

¹² However, under the U.S. Clean Air Act, SO₂ emissions will be subject to an allowance system which effectively converts those emissions into just another cost. All a system operator needs to do is multiply each generator's emissions by the appropriate allowance price, add that cost to the generator's fuel expense, and then use traditional least-cost dispatching algorithms. Treating SO₂ as a separate attribute would constitute double-counting (see section 2.1 and Hobbs, 1992).

3.7 MCDM methods in national energy planning

MCDM methods can help governments evaluate energy sector plans and policies. For instance, an expert committee appointed by the Government of Finland evaluated a range of policies using a formal multicriteria procedure (Hamalainen and Karjalainen, 1992). Criteria considered included a range of health and environmental risks, in addition to economic and political factors. The Analytic Hierarchy Process, a version of ratio questioning, was used to select weights (Section 2.8). Earlier, the parliament of Finland applied a similar procedure (Hamalainen, 1988).

As another example, Jones *et al.* (1990) asked 25 individuals in the United Kingdom to choose weights for 15 different criteria. Examples of the criteria include supply security, cost, contribution to the greenhouse effect, diversity, acid rain, and capital requirements. Weights are chosen by the rating method on a 1-100 scale. These weights are then used to evaluate five distinct energy strategies through the year 2010:

1. free market in energy
2. nuclear power emphasized
3. a mix of coal and nuclear emphasized, plus some DSM
4. an emphasis on DSM, with some additional fossil fuel supplies
5. a primarily DSM and renewable energy plan

Keeney *et al.* (1987) present another application of MCDM methods to national energy policy.

The application by Karni *et al.* (1992) to Israeli electricity pricing policy is particularly interesting because they use more than one MCDM method. We summarize the steps they went through below.

Step 1: attribute selection

Fifteen criteria are considered, grouped into six categories: reliability, cost, pricing efficiency, inflationary effects, social equity, and external goals (public acceptance, effects on national budget, and infrastructure).

Step 2: definition of options

Five pricing policies are compared, each with a different theme:

1. average cost-based pricing with no reference to marginal cost
2. price based on short run marginal cost
3. price based on long run marginal cost
4. conservationist pricing, where prices are set higher than cost for some customers
5. current policy, where prices are subsidized for some customers

Step 3: quantifying impacts

The value of each of the fifteen attributes for each policy is determined by panels of experts in the relevant fields. These are all determined on a subjective 1-5 scale, where 5 represents the best possible outcome and 1 stands for the worst.

Steps 4,5,6: Not performed

Step 7: scaling

The subjective scales created in Step 3 are directly used in the amalgamation.

Step 8: weighting

Nineteen policy makers, representatives of industry, and other stakeholders each choose two sets of weights for the six overall categories of criteria. One set results from direct allocation of 100 points in proportion to the each category's importance. The other set is chosen by the Analytic Hierarchy Process. The two weighting methods yield very similar category weights; the major difference is that many participants directly assigned zero weight to the last three categories of criteria. However, the eigenvector calculation procedure used by AHP always gave those categories non-zero (yet relatively small) weights.

Each person's fifteen individual attribute weights are then calculated by allocating his or her category weight among the attributes in the category using ratios chosen by panels of experts in relevant policy areas. One possible rationale for such a procedure is that policy makers should set priorities for general goals, while experts are better able to prioritize within categories. Another rationale is that policy makers have neither the time nor patience to weight a large number of criteria.

Step 9: amalgamation

Weighting summation is used to rank the five policies. In addition, each stakeholder directly and holistically ranks the five policies in terms of overall attractiveness. Some consistent patterns of result emerged from this analysis. Sixteen of the nineteen experts rank the current policy last by both the weighting summation and direct ranking methods. This is true even though the current policy is best in four the fifteen attributes. Furthermore, seventeen of the experts rank short term marginal cost pricing as best by weighting summation.

However, there are also inconsistencies. Only six of the experts preferred short term marginal cost pricing when asked directly. More would rather choose the long term approach. The correlation between the holistic and weighting summation rankings are actually negative for two of the stakeholders, and are 0.5 or lower for seven of the stakeholders. Unfortunately, the authors apparently did not confront the stakeholders with the inconsistencies between the two ranking approaches. However, the authors do suggest the following possible causes for the disagreements between the weighting summation and holistic ranks:

- the stakeholders disagreed with experts' assessments of the performance of each policy on the criteria (Step 3)
- stakeholders use a definition for "importance" when choosing weights that disagrees with their willingness to trade off the attributes when directly ranking options. (Footnote: In Section 2.8, we argue that this is likely for weighting methods that ask directly for weights because they leave the definition of attribute "importance" ambiguous.)
- when directly ranking options, the stakeholders consider latent attributes, such as implementability, that were excluded from the list (Step 1)

Step 10: resolving differences

Not addressed in the published study.

This case study has an important advantage over others we have reviewed in this Chapter; namely, the use of more than one method. Such comparisons promote learning and confidence in the results. For instance, because the point allocation and AHP methods give similar results, the authors can have more confidence in the validity of the weights.¹³ However, the differences between the direct holistic rankings of alternatives and the weighting summation rankings

¹³ Yet both methods are direct weighting methods; therefore their results may have converged because in each case the user was using the same--but *incorrect*--definition of attribute importance. A better check on the validity of the weights would have been a set of tradeoff judgments, as in the indifference tradeoff method. See Section 2.8.

indicated that there might be problems with the MCDM assessment, and that there is a need for further discussion. Unfortunately, as in the other cases in this Chapter, it seems that the stakeholders were not asked to resolve the disagreements.

3.8 MCDM methods in utility applications in other countries.

In this concluding section of our review of MCDM methods we examine the application of MCDM in other countries. With electricity growth rates in many developing countries in the range of 7-10% per year, and with an increasing awareness of environmental issues¹⁴, how environmental externalities ought to be incorporated into power sector planning has received growing attention. In particular, entities such as the World Bank have been grappling with the issue of how environmental factors should be incorporated already at the planning stage, because the shortcomings of waiting until the project implementation phase before environmental issues are addressed in a project EIS are all too apparent. Moreover, in Asia in particular there is a growing awareness of the greenhouse gas emission problem, given that India and China are expected to add over 50,000Mw of coal fired capacity just over the next decade. We thus review four representative areas of application of MCDM methods: (1) trade-off curve studies for analysis greenhouse gas emission reduction options; (2) MCDM methods in resource planning; (3) power plant siting; and (4) incorporating environmental factors in long range power sector planning.¹⁵

Trade-off curve studies and global warming

Over the past few years many studies have appeared on the costs of reducing greenhouse gas (GHG) emissions (GEF, 1993; Nordhaus, 1991, Manne & Richels, 1991). With electric utilities accounting for a significant fraction of such emissions in the industrialized countries, the search for cost effective ways for utilities to achieve GHG emission reductions has received particular attention. Trade-off curves are much in evidence in such analyses, and a good recent example is the study by Amagai and Leung (1991), who present an analysis of the trade-off between cost and CO₂ emissions in the Japanese power sector. It illustrates a number of important points about the use of such curves, and merits a review. It might be noted that the study only examined generation expansion and dispatch options, and did not examine mitigation alternatives, demand side management, or tax options. The results are displayed on the trade-off curve depicted on Figure 3.8.

Although the display of trade-off curves can be very useful to decision-makers, there are the usual dangers of misinterpretation caused by choice of scales. For example, there may be a great temptation, on intuitive grounds, to assume that the knee set of Figure 3.8 occurs near the point B. However, when one re-draws this figure on a relative scale -- Figure 3.9 -- such that the least cost solution (at A) is given the value of unity, the importance of *mathematical* rather than intuitive definition becomes clear.

¹⁴ As evidenced, for example, by the controversy over the World Bank's involvement in the Narmada hydro projects in India, that would displace 170,000 people in order to provide irrigation to 2 million hectares.

¹⁵ Integrated resource planning is not a term that is in general use in developing countries, although the need to go beyond the traditional emphasis on supply side expansion to include efficiency improvement and demand side management is receiving increasing attention (see e.g. Nadel (1991) for a study of DSM options in India, or USAID (1991) for a study of DSM options in Costa Rica).

Figure 3.8: Trade-off curve for the Japanese Power Sector

Source: Amagai and Leung (1991).

As is obvious from Figure 3.9, if "significantly worse" in cost means anything greater than, say 3% of the least cost solution, then all of the points on the segment BC are equivalent in cost, but do differ significantly in CO₂ emissions: that is, while the least cost and least CO₂ solutions differ by 3.5% in cost, they differ by almost 20% in CO₂ emissions.

Figure 3.9: Trade-off curve for the Japanese Power Sector on a relative scale.

A second caution relates to the interpretation of the slope of the trade-off curve itself. For example, on Figure 3.9, the slope of the trade-off curve changes dramatically from about \$10/ton on the right, to 360\$/ton on the left. The implication for policy-making would appear obvious: while one can reduce the first 12% of CO₂ emissions for 10\$/ton, further reductions incur very sharp cost penalties of 360\$/ton.

If CO₂ emissions were the only other attribute of interest, then such an interpretation of the slopes is indeed straight forward. However, if there are n attributes of interest, then the slope of a two-dimensional trade-off curve implies that all other $n-2$ attributes have been valued at zero. Thus in a three-dimensional case, say in which the third dimension represents population exposure to SO₂, the statement that the slope of the cost-CO₂ emissions curve is x \$/ton of CO₂ removed implies that the benefits of any concomitant SO₂ removal (or costs of increased externalities associated with more hydro in the system) are all valued at zero.

Thus, returning to the numerical example of Figure 3.8, since the solutions on the steep part of the curve involve increasing use of hydro and natural gas, CO₂ is not the only pollutant whose emissions are reduced in this portion of the curve. Consequently, the expenditure of the first \$360 may not only reduce CO₂ emissions by one ton, but also reduce SO₂ and NO_x emissions. It may also increase the externalities associated with additional hydro plants.

Capacity expansion planning

Largely at the insistence of the World Bank and other international financial institutions, developing countries have had to demonstrate that facilities proposed for financing are in the "least cost" plan. In practice this has meant the application of sophisticated capacity expansion optimization models -- such as the WASP model which is very widely used in developing countries -- whose objective function is the minimization of the present value of system expansion costs over some planning horizon and for some given level of system reliability.

Perhaps the most important objection to the use of models such as WASP and EGEAS is the difficulty with which they deal with the huge uncertainties currently faced by the sector, ranging from uncertainties in the prices of imported fuels to the impacts of construction cost overruns and droughts. In other words, the so called "least cost" plan may in fact be optimal only for a relatively narrow band of external circumstances assumed, and may deviate from optimality for most other combinations of external factors.

This is illustrated on Figure 3.10, which shows the results of successive annual "least cost plans" derived by use of the WASP optimization model. There are many reasons for the abrupt changes from year to year, including refinement of costs, changes in demand projections, and abrupt changes in fuel price projections. But whatever the explanation, it is clear that the environmental impacts associated with these changes are equally abrupt: in 1990 the expectation was that by 2005 one would need 300 Mw of new coal capacity; by 1991 this had changed to 900 Mw.

Given these kinds of difficulties the World Bank has begun to explore the use of MCDM in electricity supply planning, notably the studies by Crousillat (1989) and Crousillat and Merrill (1992) for Hungary and Costa Rica. These involved trade-off analysis to identify decision sets or "short lists" (using dominance analysis techniques described in Chapter 2), and subsequent risk analysis to measure robustness, quantify exposure, and develop hedges against particular uncertainties. On Figure 3.11 is shown a trade-off curve for total cost v. loss of load probability for the Hungarian power system: a plan was found at the knee of the five-attribute trade-off surface under all futures examined. One of the five attributes was SO₂ emissions. The main uncertainty was the extent to which imports were available from the former Soviet Union, and a number of hedging strategies were then developed.

Figure 3.10: Results of the annual capacity expansion planning studies

Figure 3.11: Trade-off curve for power supply in Hungary

Power plant siting

With US consulting firms very active in the fast-growing countries of Asia it comes as no surprise that the techniques used in the US for power plant siting in the 1960s and 1970s are now being applied to studies in these new markets. For example on Table 3.6 we show the weighting summation of impact score used in a siting study for a gas-fired thermal plant on the Island of Kalimantan, Indonesia. The ratings for each criterion (here the rows of the matrix) were on a discrete step-function scale (similar to that used earlier in New York), with the ratings -3, -2, -1, 0, +1, 2 and 3.

Table 3.6: The rating matrix for the Tanjung Batu gas fired thermal plant in Indonesia

Weighted Environmental Impact Scores For Gas Fueled Combined - Cycle Combustion Turbine										
Environmental Factors	Importance Weight	Gunung Lark		Tanjung Batu		Jongkang		Loa Duri		
		Impact rating	Im- pact Sco- res	Impact rating	Im- pact Sco- res	Impact rating	Im- pact Sco- res	Impact rating	Im- pact Sco- res	
I. Air Quality	10									
- Surrounding the project site	4	0	0	0	0	0	0	0	0	0
- Industrial Areas (power consumer)	6	3	18	3	18	3	18	3	18	18
II. Water Quality	10									
- Mahakam river	6	-1	-6	-2	-12	-2	-12	-3	-18	-18
- Surface and ground water	4	0	0	0	0	0	0	0	0	0
III. Water Quality	10									
- Mahakam river	6	-1	-6	-1	-6	-1	-6	-1	-6	-6
- Surface and ground water	4	0	0	0	0	0	0	0	0	0
IV. Biological Resources	18									
- Endangered species	6	0	0	0	0	0	0	0	0	0
- Wild Animals	6	0	0	0	0	0	0	0	0	0
- Economic Biota	6	-1	-6	-1	-6	-1	-6	-1	-6	-6
V. Land Use	24									
- Productive Cultivated Land	6	-1	-6	-2	-12	-3	-18	-3	-18	-18
- Protection Forest	6	-1	-6	0	0	0	0	0	0	0
- Human Settlements	6	0	0	0	0	0	0	0	0	0
- Facilities and Infrastructure	6	0	0	0	0	0	0	0	0	0
VI. Socio Economic	18									
- Number of Population	6	0	0	-1	-6	-1	-6	-2	-12	-12
- Job opportunities	6	2	12	2	12	2	12	2	12	12
- Local Economic Activities	6	1	6	1	6	2	12	2	12	12
VII. Social Cultures	10									
- Social Institutions	6	0	0	0	0	-1	-6	-2	-12	-12
- Historical Heritage	4	0	0	0	0	0	0	0	0	0
T o t a l	100		6		-6		-24		-54	

Many of the concerns about validity of past MCDM applications noted in Chapter 2 are also illustrated by the siting studies done in Sri Lanka in the mid 1980s in an attempt to find a suitable location for a first coal-fired power plant (with an ultimate capacity of 900Mw). A preliminary screen identified the Trincomalee Bay area as the best location because of the suburb natural deepwater harbor that would permit coal imports at relatively low cost¹⁶. Within this general area 7 sites were identified as possible candidates, and weighting summation used to rank them: the results are shown on Table 3.7. The most favorable site was assigned rank number 7, the least favorable rank number 1.¹⁷ The ranks were determined by "...specialists knowledgeable about each of these siting factors", and consensus weights were estimated by "several siting specialists". The final scores were calculated as follows: "...The rank numbers were multiplied by the factor weights for each factor and the products summed to produce a weighted rank total for each site." (Our emphasis). Unhappily this procedure violates most of the principles necessary for validity¹⁸.

Table 3.7: Site Rankings in the Trincomalee Study

site	coal unloading	transmission	site	generation facilities	supporting infrastructure	land use environment	overall ranking
"consensus weights>	25%	10%	15%	20%	10%	20%	
1	2	1	2	1	4	7	2.7
2a-1	7	3	3	5	6	4.5	5
2a-2	6	3	4	7	6	6	5.6
2a-3	5	3	6	6	6	4.5	5.2
3a-1	4	5.5	5	3	2.5	2	3.8
3a-2	3	5.5	7	4	2.5	2	3.6
4	1	7	1	2	1	2	2

Source: Trincomalee Thermal Power Project, Site Evaluation and Selection Study Phase II Final Report, Ceylon Electricity Board, September 1985, p.10-1.

This application illustrates another problem in the application of such methods. Five of the six attributes used here are in fact based on costs: indeed, we are told "...capital costs and expected costs of operation were generally used as a basis for ranking the sites". If in fact the purpose of the exercise is to examine the trade-off between "land use and environmental factors on the one hand", and factors related to system costs on the other, then everything related to cost should be subsumed into a single cost attribute.

The Trincomalee EIS proved highly controversial, and in the end the necessary permit from the Coast Conservation department was denied on grounds of unacceptable impacts of

¹⁶ Sri Lanka has no fossil resources of its own, and with few remaining hydro sites, must look to thermal baseload generation by the end of the 1990s to meet strongly growing electricity demand (expected to be around 7% per year).

¹⁷ Black and Veatch, *Trincomalee Thermal Power Project, Site Evaluation and Selection Study Phase II Final Report*, September 1985. The site selection procedure is described in Section 10 of this Report.

¹⁸ Even more unhappily, the procedure was justified as follows: "...Weighted ranking systems are commonly used in Engineering circles in the United States to evaluate factors other than costs: these ranking methods are particularly applicable to siting studies, since at least one of the major siting factors cannot be evaluated with costs. We have prepared many siting studies involving different geographical states in the United States which involved use of this method."

thermal discharges. In a subsequent more broader siting study, weighting summation was again used to rank sites. On Table 3.8 we show, as an example,, the "air quality scores". On the 30-point scale used for air quality, 30 is the best, 0 is the worst. The basis for the scoring is described only in very general terms as "number of sensitive receptors and the exposure risk of significant receptors". Even as *relative* indications these scores are very ambiguous. Finally, it is unclear whether the scores are comparable across the various generation types. One can only speculate, therefore, why a coal plant at Sapugaskanda would have an air quality score of 11, while a diesel plant is scored at 13.

Table 3.8: Air Quality Scores siting study

site ID	site	combined cycle	oil	coal	diesel
NC-1	Mundal East	21	20	20	20
NC-3	Bolawata	21	20	20	20
NC-5	Ekala	21	20	20	20
C-1	Mahawetta				7
C-3	Kelanitissa	6			5
C-6	Peliyagoda	11	6	6	11
C-7	Sugathasua	6			5
C-8	Yakbedda	11			11
C-10	Sapugaskanda	13	11	11	13
C-11	Siyambalape	13	11	11	13
C-12	Biyagama	13	11	11	13
G-3	Bope				13
G-6	Armitage Hill				13
G-7	Haliwate				13
G-8	Wakawella	13	9	10	13
G-9	Bataduwa	13	9		13
M-1	Kudawella East	30	28	29	27
M-16	Weligama	30	28	29	27

The most serious problem with such valuations is not that expert judgment may necessarily be incorrect, but that the relationship between the score and possible mitigation actions is not demonstrable. In other words, there is no way of subsequently performing a sensitivity analysis that might explore the use of alternative fuels, alternative pollution control strategies, or of alternative technologies. Indeed, as we shall note below, actual calculations of the impacts associated with emissions at these sites proved to show differences an order of magnitude different to the relative scores shown in this table.

Incorporating environmental externalities in power sector planning

With the World Bank under heavy criticism for its involvement in power projects that have major environmental impacts, and with project level EIS's under attack for being *ex post facto* justifications for decisions already made¹⁹, the Environment Department of the Bank has been examining the use of

¹⁹ The EIS prepared for the proposed coal plant at Trincomalee in Sri Lanka provoked almost as much controversy over the *process* as it did about actual impacts, with comments such as "...it is therefore evident that the entire feasibility study process, terminating with the EIA, has followed a policy known as "decide-announce-defend... the EIS is an *ex-post facto* document seeking to environmentally justify a project which was formulated without any environmental considerations whatsoever .. this EIA is a post planning exercise, whereas it ought to have been an ongoing exercise

MCDM methods to better incorporate environmental concerns at the planning stage. Many of the major environmental impacts of power projects in tropical countries are particularly difficult to monetize, perhaps most notably so in the case of the loss of bio-diversity associated with inundation of forests at major hydro projects. A case study of Sri Lanka (Meier and Munasinghe, 1993) is the first application of MCDM methods to environmental work at the World Bank.

The details of the methodology and the results of this study need not concern us here²⁰. Rather what is important and relevant to US IRP practice is the way in which the attributes were defined, and the importance placed upon location. Some care was taken to be sure that attributes reflected *actual impacts*, rather than measures of regulatory convenience. Moreover, attributes were sparingly selected, on grounds that a proliferation of attributes would needlessly clog the analysis, and cloud the presentation of essential conclusions to decision-makers. Each attribute was associated with an important concern of national importance; in addition to the six environmental attributes shown on Table 3.9 the analysis also included present value of system cost, leveled tariff, and reliability (expressed as unserved energy).

Table 3.9: Summary of environmental attributes

attribute	units	impact
emissions of carbon dioxide	[1000 tons]	global warming.
population exposure to air pollutants	person-ug/m ³ /yr	human health impacts
biodiversity index	[] ⁽¹⁾	diminution of biodiversity; impact on habitat of endemic species.
increase in surface temperature > 1°C	[ha]	ecosystem impacts from thermal plumes.
employment		discounted incremental employment
emissions of acid rain precursors (SO ₂ and NO _x)	[1000t/y]	potential for acid rain damages

Source: Meier and Munasinghe (1994)

from the inception of the project".

²⁰ The study made extensive use of the dominance and trade-off curve analysis concepts discussed in Chapter 2.

Rather than using SO₂, NO_x and/or particulate emissions each as separate attributes, the study derived a health impact index that reflected the population exposure to these pollutants. This was defined as

$$X = \sum C_j P_j$$

where X is the cumulative population exposure to the incremental ambient concentration attributable to the power plant.

P_j is the population in the j -th grid square

C_j is the incremental average annual concentration in the j -th square, attributable to the emissions from the source in question.

The C_j were calculated using a simple Gaussian plume model run for some given number of wind direction and wind speed combinations. The total concentration in square j is then simply the sum of the individual plume contributions:

$$C_j = \sum c_{jk}$$

where C_{jk} is the incremental contribution in square j attributable to the k -th plume, which is in turn given by:

$$C_{jk} = \frac{Q}{\pi u_k \sigma_y \sigma_z} \exp \left[-\frac{y^2}{2\sigma_y^2} - \frac{H^2}{2\sigma_z^2} \right]$$

where u is the wind speed (in meters/sec) of the k -th plume

Q is the source term (in g/sec)

H is the stack height (in meters)

The use of the health index had the important property that air quality impacts were location specific, and resulted in a number of important policy conclusions. For example, it was shown that it really did not matter very much whether base-load coal plants sited in remote locations and with relatively tall stacks were equipped with FGD systems; far more important to reduction of population impacts was NO_x and SO₂ control at urban diesel plants with low stacks that burn relatively high sulfur residual fueloil.

Another interesting finding concerned the difference between calculated values of health impacts and the air quality scoring system used in earlier siting studies. There the air quality scores, on a scale of 1 to 30, for a coal plant located on the South Coast and a diesel plant in Colombo were 29 and 11, respectively, a factor of about 2.7; actual calculation of impact showed a difference between the two by a factor of at least 11.

4. CRITERIA FOR SELECTING MCDM TECHNIQUES

As we have seen in the previous Chapter, the range of methods and applications of MCDM techniques is very large indeed. How then does one proceed with selecting an appropriate method for a particular application? Given the focus in this report on how environmental factors can be brought into IRP, what practical guidance can be offered to a utility planning department charged with actually implementing the IRP process? We propose three criteria for selecting a technique:

1. Is a method appropriate? That is, does it provide the information that is needed, does it mesh with the decision making style of the organization, and is it compatible with the data that are available?
2. How easy is the method to use? How much time is required to apply it, what training is needed to understand it, and what computer facilities, if any, are needed? (This is the practicality criterion noted in the introduction)
3. How valid is the method? Do the results of the method (e.g., rankings of alternatives) accurately reflect the values held by the decision makers? Are the method's assumptions and the evaluations it yields consistent with the values of the decision makers? Evaluations by an invalid method might bear little relationship to the tradeoffs decision makers and the public are willing to make, and the chosen alternative might be preferred by no one. Finally, do the decision makers trust the method and its results?

The last criterion is important if different methods result in different decisions; if so, one needs to worry about which method more accurately captures the priorities of decision makers. If different methods lead to the same results, then a method can be chosen on the basis of appropriateness and ease of use alone. Because there is ample evidence that choice of method can significantly affect decisions (sometimes making more difference than who applies the method) (Hobbs, 1986), the issue of validity should not be ignored.

All of these tests were put to the practical test in the two case studies, reported in the subsequent Chapters. However, before we move to a presentation of those results, each of the three criteria merit some further discussion.

4.1 Appropriateness

What characteristics should an "appropriate" method have? Most follow from the real reasons for applying an analytical method to the decision process in the first place:

Display of trade-offs:

For public-sector decisions, the primary purpose of multicriteria methods should be to present information on tradeoffs (Cohon, 1978). A method should show, in a vivid and easily

digestible manner, how different alternatives perform on different attributes and, by doing so, make it easier for fully informed decisions to be made. An important by-product of displaying tradeoffs is the elimination of alternatives that are "dominated" (that is, perform no better on all attributes than another alternative, while performing strictly worse on at least one attribute). A multicriteria method should not be a "machine" which takes as an input the alternatives and some simple value judgments (such as weights) and then determines which alternative is "best". Such a tool can be misleading and even dangerous in a situation where there are many interested parties who are bound to disagree fundamentally on priorities. No single set of "weights" can be defined meaningfully in that case.

Risk and uncertainty:

An appropriate method should recognize that information about risks and uncertainties regarding the impacts of electric utility projects is crucial to choices among alternatives.¹ A method should display information on how alternatives affect the risks associated with each of the attributes. Risks should be explicitly dealt with in decision making for three reasons (Haimes and Stakhiv, 1986, 1989, 1990; Duckstein and Plate, 1987; Hobbs *et al.*, 1988):

First, economic, social, and environmental systems are nonlinear, in that the expected value (mean) of a system's output (e.g., health impact from pollutants associated with fossil fuel combustion) does not equal the output calculated using the expected value of the system's input (e.g., fuel consumption). The mean and distribution of the output can only be estimated by considering the entire range of inputs and their probabilities.

Second, decision makers and the public are not, in general, risk neutral towards system outputs. That is, they do not base decisions solely on expected value. For example, a 10% chance per year of 1 person dying due to inhalation of fine particulates will not be viewed the same as a 0.01% chance per year of 1000 people dying. Decision makers in that case will want to know the entire probability distribution of outputs, not just their mean value.

Third, a number of generation options are dependent upon water resources (hydro plants, cooling water for thermal plants) whose characteristics are essentially stochastic.

Understanding:

An appropriate method should be easily understood, since people with a wide variety of backgrounds will be applying it and interpreting its results.

Simplification:

An appropriate method should simplify the decision for those who have the final responsibility for it, but without oversimplifying it. It is desirable to simplify the problem by avoiding presenting too many alternatives or too many criteria at once. Psychologists cite the famous "7 plus or minus 2" rule for how many facts that a decision maker can keep in mind at once (Morse, 1979). If there is too much information, much of it will be ignored or inconsistently applied. But oversimplification can occur if too few options representing too narrow a range of tradeoffs are presented. An appropriate method will screen out alternatives using, as examples, the following procedures:

- drop dominated alternatives;

¹ Here, risk is defined as occurring when randomness is present and can be quantified using probability distributions. Uncertainty, by contrast, is defined as occurring when the probability distributions are unknown. Different techniques are used to quantify risk and uncertainty and to include them in decisions.

- drop alternatives whose performance in one or more attributes is so poor that they are likely to be rejected on that basis alone ("exclusionary screening");
- keep alternatives which are robust; that is, which do well under a wide range of assumptions concerning weights, risk attitudes, and probability distributions of inputs; and
- keep alternatives that emphasize different attributes so that the decision makers are presented with a real choice. (Otherwise, the real decision is being made by analysts, not those who should be responsible.)

Another simplification procedure that is often appropriate is to combine the attributes associated with an objective into a single aggregate measure of performance on that objective, if it is possible to obtain a consensus on the relative weight that should be given each attribute. Such an aggregation process would result in a smaller set of criteria that the decision makers would have to consider when making a decision.

4.2 Ease of Use

Ease of use is an empirical question for which it is hard to set firm guidelines, and which is highly dependent upon the particular set of users. If the users are limited to just utility planners one may come to one set of conclusions; if, as we advocate, the user group includes representatives from interest groups and the public, one may come to an altogether different set of conclusions. Indeed, the two case studies that follow support this particular contention: in Seattle, the user group was much broader (and included members of local interest groups) than in the case of Centerior, where participation was limited to utility planners. The findings about ease of use of methods did indeed differ somewhat.

The time available proves to be an important issue. The effort that can be put into explaining a method (and of course the effectiveness of that training effort) will certainly affect how participants feel about ease of use. A number of studies (Adelman *et al.*, 1984; Eckenrode, 1965; Nutt, 1980; Sarin *et al.*, 1980) have contrasted the time and effort required of decision-makers to use a variety of weighting and value-scaling techniques. Even the most difficult methods, such as indifference trade-off weighting or risky utility function scaling, have rarely taken more than three hours for individuals. However much less is reported in the literature about groups that include larger numbers of non-technically trained people. The Seattle case study, however, tends to suggest that the members from the citizen and environmental groups who participated were just as able to understand the methods as utility planners.

4.3 Validity

Of the three criteria, validity is the most difficult to determine in practice, because there are so many dimensions to the concept. The literature distinguishes between (1) predictive validity; (2) estimative validity; (3) construct validity; (4) convergent validity; (5) inter-expert validity; (6) inter-temporal validity, and (7) normative validity. Each of these addresses a different dimension, and is reviewed, *seriatim*, in the following paragraphs.

Predictive validity:

For behavioral scientists, the purpose of MCDM methods is to understand and simulate decisions: thus predictive validity -which evaluates a method on the basis of how well it replicates unaided, subjective evaluations -- is the primary criterion of validity. In our case

studies, the first task we put to the user groups was to make an initial ranking of the options, given all of the data about the options and estimates of the impacts -- which we term the "holistic rankings". Predictive validity was then established by comparing the results of the different methods with those of the initial holistic rankings.

Estimative validity:

The decision scientist's rejoinder to behavioral scientists is that predictive validity is not the appropriate standard, because unaided, holistic judgments may in fact be fundamentally biased and/or inconsistent. One possible way of looking at this problem is to ask whether a method produces a result that is consistent with that actually achieved by experts to evaluate alternatives in an actual situation.²

A practical approach to estimate validity is to have each participant repeatedly apply each method, alternating between them and contrasting their results, until their rankings converge and the user is satisfied (Lai and Hopkins, 1988). Then the most valid method is defined as the one whose initial rankings were closest to the final ones. This is, we believe, a particularly useful type of validity because a purpose of MCDM methods is to help people form their values; thus, their final rankings of plans are more trustworthy than their initial holistic judgments.

Construct validity: This type of validity is also referred to as empirical validity: a method is said to have construct validity when its evaluations change in appropriate ways when relevant problem characteristics are altered, but do not change when irrelevant characteristics are changed (Stewart and Ely, 1984). Therefore, we looked at this definition of validity in the case studies in some detail. Some observers claim that construct validity is a much more objective test than predictive validity, because the latter depends on unaided evaluations which are themselves untrustworthy and fail to provide an objective standard (Pitz et al., 1980).³

Convergent validity:

If two methods lead to similar parameters or evaluations, then confidence in their validity is increased. Important differences suggest that the methods cannot both be valid. But convergence does not guarantee validity, since both methods could have the same systematic bias. Convergence validity is examined in both of the case studies: and we do indeed find that some methods give quite different results than others.

Interexpert validity:

The assumptions of this criterion is that evaluations by fully knowledgeable and rational people should agree. Invalid methods are presumed to unduly exaggerate differences of opinion or value. For example, Newman (1975) discovered more convergence among evaluations by different judges' weighting summation rules than among their unaided judgments, and therefore concluded that the latter were less valid.

² A good example of such a study is Adelman *et al.* (1984), who compared the accuracy of five weighting techniques with two group discussion methods, adopting as a standard a model actually used by the U.S. Marines to evaluate combat readiness.

³ Perhaps one of the most obvious tests of construct validity is to change a decision-problem by adding a significantly dominated alternative: rankings ought not to change in a method that meets construct validity.

However, the inadequacy of the inter-expert convergence criterion is evident when one realizes that under this criterion, equal weighting would be the most valid weighting method. Moreover, since differences in values lie at the heart of many utility planning problems, forced uniformity would only bury disagreements that should be made explicit.

Intertemporal validity:

A method has inter-temporal validity if, when applied by the same person at different times, its evaluations do not change. Yet differences between a first and a second application may have as much to do with learning, or signal that the application of formal MCDM techniques has indeed achieved one of its aims, namely that it has forced the user to think harder about how he or she feels about the issues, and/or to gather more information, than it does about validity of the method itself.

Balzer *et al.* (1983) is one of the few studies to have looked at this question in a carefully controlled experiment: their results suggest the not unsurprising conclusion that direct weighting is more trustworthy if the users have been allowed to examine the alternatives and become very familiar with the problem.

Normative validity:

A method is said to have normative validity when it is based on the best available information, and is the result of careful thought (Stewart and Ely, 1984). The method must be understood, and judged reliable by the decision-maker -- who would then be committed to the method's results and willing and able to defend them. This is self-evidently a purely empirical question, because it can be tested only by asking the decision-makers for their opinions.

It may be difficult to treat this question as purely one of validity: the decision-maker's overall comfort with a particular method, and his willingness to defend decisions reached by that method, are likely to be as much questions of appropriateness and ease of use.

LITERATURE CITED

- Adelman, L., P. J. Sticha, and M. L. Donnell, 1984. The Role of Task Properties in Determining Relative Effectiveness of Multiattribute Weighting Techniques, Organ. Behav. Human Performance, 33, 243-262.
- Amagai, H., and P. Leung, 1989. Multiple Criteria Analysis for Japan's Electric Power Generation Mix. Energy Systems and Policy, 13, 219-236.
- _____, 1991. The Trade-off between Economic and Environmental Objectives in Japan's Power Sector, Energy Journal, 12,4,p.95-104.
- Andrews, C. J., 1992. The Marginality of Regulating Marginal Investments: Why We Need a Systemic Perspective on Environmental Externality Adders, Energy Policy, 20, 451-475.
- Andrews, C.J., 1992. Spurring Inventiveness by Analyzing Tradeoffs: A Public Look at New England's Electricity Alternatives, Environmental Impact Assessment Review, 12.
- Asian Development Bank, 1991. Economic Policies for Sustainable Development, Manila.
- Asian Development Bank, 1992. Environmental Considerations in Energy Planning, Manila.
- Balzer, W. K., J. Rohrbaugh, and K. R. Murphy, 1983. Reliability of Actual and Predicted Judgments Across Time. Organ. Behav. Hum. Perform., 32, 109-123.
- Baumol W. J. and W.E. Oates, 1971. The use of Standards and Pricing for Protection of the Environment, Swedish Journal of Economics, 7, 42-54.
- Belton, V., 1986. A Comparison of the Analytic Hierarchy Process and a Simple Multi-Attribute Value Function, European Journal of Operations Research, 26 (1), 7-21.
- Bernow, S. B. Biewald, and D. Marron, 1991. Full-cost Dispatch: Incorporating External Costs in Power System Operations, The Electricity Journal, 4 (2), 20-33.
- Bernow, S. *et al.*, 1992. Direct Environmental Impacts of Demand-Side Management, Proceedings, ACEEE 1992 Summer Study, American Council for an Energy Efficient Economy, Washington, DC.
- Biewald, B. *et al.*, 1991. Valuation of Environmental Externalities: Sulfur Dioxide and Greenhouse Gases, Tellus Institute, Boston, Massachusetts.
- Black and Veatch International, 1988a. Trincomalee Thermal Power Project: Social and Environmental Assessment. Report to Ceylon Electricity Board, Colombo, Sri Lanka.
- Black and Veatch International, 1988b. Thermal Generation Options Study, Report to the Ceylon Electricity Board, Colombo, Sri Lanka.
- Bloemer, J.G., and D.C. Knue, 1993. Integrated Resource Planning/CAAA Compliance Planning at the Cincinnati Gas & Electric Company, presented at the PROSCREEN/PROVIEW User's Forum, Atlanta, GA, September.
- Bloom, J. A., M. Caramanis, and L. Charny, 1984. Long-Range Generation Planning using Generalized Benders Decomposition: Implementation and Experience, Operations Research, 32, 314-342.
- Bohi, D. L. and D. Burtraw, 1992. Utility Investment Behavior and the Emission Trading Market, Resources and Energy 14, 129-153.
- Bohm, P. and C. S. Russell, 1985. Comparative Analysis of Alternative Policy Instruments, in Handbook of Natural Resources and Energy Economics (A. V. Kneese and J. L. Sweeney, Eds.), North-Holland, New York, NY.
- Bonneville Power Administration, 1986. Environmental Costs and Benefits Associated with Hydropower Development in the Pacific Northwest, Report DE-AC79-83BP11546, Portland, Oregon.
- Bonneville Power Authority, 1991. Environmental Costs and Benefits: Documentation and Supplementary Information, Portland, OR.
- Borcherding, K., T. Eppel, and D. von Winterfeldt, 1991. Comparison of

- Weighting Judgments in Multiattribute Utility Measurement, Management Science, 37, 1603-19.
- Braithwait, S. D. and D.W. Caves, 1994. Three Biases in Cost-Efficiency Tests of Utility Energy Efficiency Programs, The Energy Journal, in press.
- Brown, C.A., 1984. The Central Arizona Water Control Study: A Success Story for Multicriteria Planning and Public Involvement, Water Resources Bulletin, 20(3), 331-338.
- Brown, C.A., 1990. Managing Conflict Over a Dam Safety Problem, in Managing Water-Related Conflicts: The Engineer's Role, edited by W. Veissman, Jr. and E.T. Smerdon, American Society of Civil Engineers, New York, 223-237.
- Buhring, W. et al., 1991. ENPEP: An integrated approach for modeling national energy systems, Argonne National Laboratory, Environmental Assessment and Information Sciences Division.
- Burtraw, D., W. Harrington, A.M. Freeman III, and A. J. Krupnick, 1993. The Analytics of Social Costing in a Regulated Industry, Discussion Paper QE93-01, Resources for the Future, Washington, DC.
- Burtraw, D. et al., 1992. The Analytics of Social Costing in a Regulated Industry, unpublished manuscript, Resources for the Future, Washington, DC.
- Caldwell, S. and A. Ortiz, 1989. Overview of Proposed Revisions to the Superfund Hazard Ranking System, J. Air and Waste Management Association, 39 (6), 801-807.
- Casazza, J. A., 1993. Linking Power System Planners, the Public, and Energy Policy Makers, IEEE Power Engineering Review, 13 (1), 9-10.
- Caverhill, E. J., 1991. Prepared Testimony on Externalities on Behalf of Wisconsin Gas Company, Hearings of Public Service Commission of Wisconsin, Docket 05-EP-6, Nov. 22.
- Chamberlin, J. P. Herman, and M. Mauldin, 1993. The Value Test: Economic Efficiency and Demand-Side Management, Barakat & Chamberlin, Oakland, CA.
- Chankong, V. and Y. Haimes, 1983. Multi-objective Decision Making: Theory and Methods, North-Holland, Amsterdam.
- Chapel, S. 1989. CATALYST, Designing Flexible Business Strategies. EPRI Journal, Oct/Nov. 46-49.
- Chernick, P. and E. Caverhill, 1992. Testimony on Behalf of Boston Gas Company, Docket No. D.P.U. 91-131, Commonwealth of Massachusetts Department of Public Utilities, Boston, MA.
- Chernick, P. and E. Caverhill, 1983. The Valuation of Externalities from Energy Production, Delivery, and Use, Report to the Boston Gas Company, PLC, Inc., Boston, MA, Dec. 22.
- Chernick, P. and E. Caverhill, 1991. Methods of Valuing Environmental Externalities, The Electricity Journal 4(2), p46-53.
- Costello, K. W. and P.S. Galen, 1984. A Proposed Methodology for Evaluating Utility Conservation Programs, Dockets Nos. 83-0034 to 83-0043, Illinois Commerce Commission, Policy Analysis and Research Division, Springfield, IL.
- Crousillat, E., P. Dorfner, P. Alvarado, and H. M. Merrill, 1993. Conflicting Objectives and Risk in Power System Planning, IEEE Trans. Power Systems, 8.
- Crousillat, E. 1989. Incorporating Risk and Uncertainty in Power System Planning World Bank, Industry and Energy Department, Energy Series Paper #17, Washington, DC.
- Crousillat, E. and H. Merrill, 1992. The Trade-off/Risk Method: A Strategic Approach to Power Planning, World Bank, Industry and Energy Department, Washington, DC.
- DeCicco, J. M., S. S. Bernow, and J. Beyea, 1992. Environmental Concerns Regarding Electric Power Transmission in North America, Energy Policy, 20(1), 30-39.
- Dee, N., et al., 1973. An Environmental Evaluation System for Water Resource Planning, Water Resources Research, 9, 3, p.523.
- Delquié, P. 1994. Inconsistent Trade-offs Between Attributes: New Evidence in Preference Assessment Biases, Management Science, 39 (11), pp.1382-1395.
- Dobson, J., 1976. The Maryland Power Plant Siting Project: An Application of the ORNL-Land Use Screening Procedure, ORNL/NUREG/TM-79, Oak Ridge National Laboratory, Oak Ridge, TN.

- Hobbs, B. F., H.B. Rouse, and D.T. Hoog, 1993. Measuring the Economic Value of Demand-Side Resources in Integrated Resource Planning Models, IEEE Trans. on Power Systems, 8(3).
- Hobbs, B. F. and P. Meier, 1993. Multicriteria Methods for Resource Planning: An Experimental Comparison. IEEE Transactions on Power Systems, December 1993.
- Huber, G. P., 1974. Multiattribute Utility Models: A review of Field and Field-like Studies, Management Science, 20, 1393-1402.
- Hutchinson, M.A., 1991. Emission Caps, Externalities and Double Counting: Why Certain Externalities Should Not Be Included in New Resource Planning, in DSM and the Global Environment, Synergic Resources Corporation, Arlington, VA.
- International Atomic Energy Agency, 1991. Senior Expert Symposium on Electricity and the Environment: Key Issues Papers, Vienna, Austria.
- Isard, W. A., 1972. Ecologic-Economic Analysis for Regional Development, Free Press, New York, NY.
- Jia J. and G.W. Fischer, 1993. Evaluating Multiattribute Decision Quality: A Simulation Study, ORSA/TIMS Joint Natl. Meeting, Chicago, IL, May 1993.
- Jones, M., C. Hope, and R. Hughes. 1990. A Multi-Attribute Value Model for the Study of UK Energy Policy, Journal of the Operational Research Society, 41(10):919-929.
- Joskow, P.L., 1992. Weighing environmental externalities: Let's do it right!, The Electricity Journal 5(4), pp53-67.
- Kaarakka, P., M.S. Kanarek, and J.R. Lawrence, 1989. Assessment and Control of Indoor Air Pollution Resulting from Wood Burning Appliance Use, Environment International, 15, 635-642.
- Kahn, A.E., 1991. An economically rational approach to least-cost planning, Electricity Journal, 11-19.
- Kahn, E.P. *et al.*, 1989. Evaluation Methods in Competitive Bidding for Power. Lawrence Berkeley Laboratory, LBL-26924/UC-101, Berkeley, CA.
- Kahnemann, D., P. Slovic, and A. Tversky, 1982. Judgment under Uncertainty: Heuristics and Biases, Cambridge University Press, New York, NY.
- Karni, R., P. Feigin, and A. Breiner, 1992. Multicriterion Issues in Energy Policy-making, European Journal of Operational Research, 56:30-40.
- Kavrakoglu, I., and G. Kiziltan, 1983. Multi-Objective Strategies in Power Systems Planning, European Journal of Operational Research 12, 159-170.
- Keefer, D. L. and S. M. Pollack, 1980. Approximations and Sensitivity in Multi-objective Resource Allocation, Operations Research, 28(1), 114-128.
- Keeney, R.L., *et al.*, 1981. Decision Framework for Technology Choice - Volume 1: A Case Study of One Utility's Coal-Nuclear Choice, EPRI EA-2153, Electric Power Research Institute, Palo Alto, CA.
- Keeney, R. L. and H. Raiffa, 1976. Decisions with Multiple Objectives: Preferences and Value Trade-offs, New York: Wiley.
- Keeney, R.L., and K. Nair, 1977. Nuclear Siting Using Decision Analysis, Energy Policy, 5 (3), 223-231.
- Keeney R. L. and A. Sicherman, 1983. Illustrative Comparison of One Utility's Coal and Nuclear Choices, Operations Research, 31(1), 50-83.
- Keeney, R. L., O. Renn, and D. von Winterfeldt, 1987. Structuring West Germany's Energy Objectives, Energy Policy, 15, p.352-362.
- Keeney, R. L., and D. von Winterfeldt, 1987. Optimal Procedures to Evaluate Decisions with Multiple Objectives, EPRI EA-5433, Electric Power research Institute, Palo Alto, CA.
- Kirkwood, C. and R.K. Sarin, 1985. Ranking with Partial Information: A Method and an Application, Operations Research, 33, 38-48.
- Koenig, J. Q. *et al.*, 1988. Wood Smoke: Health Effects and Legislation, Northwest Environmental Journal, 4, 41-54.
- Koppe, R.H. 1992. Testimony on the Valuation of Environmental Externalities, Docket No. D.P.U. 91-131, Commonwealth of Massachusetts Department of Public Utilities, Boston, MA.
- Krupnick, A. J., W.E. Oates, and E. Van de Verg, 1983. On Marketable Air Pollution Permits: The Case for a System of Pollution

- Offsets, Journal of Environmental Economics and Management, 10, 233-247.
- Lai, S. K. and L.D. Hopkins, 1988. Experimental Assessment of Multiattribute Evaluation Techniques, ORSA/TIMS Joint Natl. Meeting, Denver, CO.
- Lootsma, F.A., *et al.*, 1990. Choice of a Long-Term Strategy for the National Electricity Supply via Scenario Analysis and Multi-Criteria Analysis, European Journal of Operational Research 48:189-203.
- Maekawa, T., *et al.* 1992. Fuzzy Coordination of Multiobjectives Reflecting the Operator's Intention for Dynamic Generation Rescheduling, Electrical Power and Energy Systems, 14(5): 314-320.
- Manne, A.S. and R. Richels, 1991. Global CO2 Emission Reductions: The Impacts of Rising Energy Cost, Energy Journal, 12, 1, 87.
- Markandya, A., 1990. Environmental Costs and Power System Planning. Utilities Policy, 1, p.13-27.
- Massachusetts Department of Public Utilities Decision, Docket 89-239, Aug. 31, 1990.
- Massam, B. H., and I. D. Askew, 1982. Methods for Comparing Policies using Multiple Criteria: An Urban Example, Omega, 10(2): 195-204.
- Meier, P., 1974. Evaluation of Power Facilities: A Reviewer's Handbook. Berkshire County Regional Planning Commission, U.S. Department of Housing and Urban Development, Washington DC.
- Meier, P. and L. Ruff, 1980. The Spatial Dimension of Regulatory Impact Assessment. Report to the National Commission on Air Quality, Washington, DC.
- Meier, P., M. Munasinghe, and T. Siyambalapitiya, 1993. Energy Sector Policy and the Environment: A Case Study of Sri Lanka. Environment Department, World Bank, Washington, DC.
- Meier, P. and M. Munasinghe, 1994. Incorporating Environmental Concerns into Power Sector Decision-making: A Case Study of Sri Lanka, Environment Department Paper 6, World Bank, Washington DC.
- Merrill, H. M. and A. J. Wood, 1991. Risk and Uncertainty in Power System Planning, Electrical Power and Energy Systems, 13, 2, pp.81-90.
- Metropolitan Environmental Improvement Program, 1992. Clean Air 2000: An Action Plan for the Colombo Metropolitan Area. Colombo, Sri Lanka.
- Mintzer, I. *et al.*, 1990. Environmental Externality Data For Energy Technologies, Center for Global Change, University of Maryland.
- Montana Power Co., 1991. Environmental Externality Considerations For Use in Integrated Least Cost Planning. Unpublished Manuscript.
- Montgomery, D., 1972. Markets in Licenses and Efficient Pollution Control Programs, Journal of Economic Theory, 5, 395-418.
- Munasinghe, M., 1993. Environmental Economics and Sustainable Development. World Bank Environment Paper 3, Washington, DC.
- Muslu, M., and M.D. Anderson. 1989. A Multiobjective Optimization Method for Electrical Power Distribution Systems with Load Control Alternatives, Electric Power Systems Research, 16(1):1-9.
- Nagelhout, M., 1993. Valuation of Environmental Externalities in Electric Resource Selection, Public Utilities Fortnightly, March 1, pp. 44-47.
- Nelson, M., 1993. Personal communication, Washington State Energy Extension Service, Olympia, WA.
- Nelson, M. and S. Kaufman, 1990. Evaluating Effects of Wood Smoke Control Legislation in Washington State on Electrical Consumers, Prepared by the Washington State Energy Office for the Bonneville Power Administration, Portland, OR.
- Newman, J. R. 1975. Assessing the Reliability and Validity of Multiple Attribute Utility Assessment Procedures, Social Science Research Institute, University of Southern California, Los Angeles.
- New York Department of Environmental Conservation, 1980. Hazardous Waste: The Siting Board, New York State Environment, 9 (Dec.18): 3-6.
- Niemi, E., E. Whitelaw, and B. Barkovich, 1993. Environmental Externalities and Electric Regulation, National Association of Regulatory Utility Commissions, Washington, DC.

- Nordhaus, W., 1991. The Cost of Slowing Climate Change: A Survey. Energy Journal, 12,1.
- Nutt, P.C., 1980. Comparing Methods for Weighting Decision Criteria. Omega, 8, 163-172.
- OKO Institute, 1989. Umweltanalyse von Energie Systemen: Gesamt Emissions Modell Integrierter Systeme (GEMIS), Ministry of Economy and Technology, Hessen, Darmstadt, Germany.
- Ottinger, R. L. *et al.*, 1990. Environmental Costs of Electricity, Pace University Center for Environmental Legal Studies, Oceana Publications, New York, NY.
- Otway, H., and W. Edwards, 1977. Application for Simple Multiattribute Rating techniques to the Evaluation of Nuclear Waste Disposal Sites: A Demonstration, RM-77-31, IIASA, Laxenberg, Austria.
- Palmer, K. and H. Dowlatabadi, 1991. Implementing Environmental Costing in the Electric Utility Industry, Discussion Paper QE91-13, Resources for the Future, Washington, DC.
- Pitz, C., J. Heerboth and N. Sachs, 1980. Assessing the Utility of Multi-attribute Utility Assessments, Org. Behav. Human Perform., 26,65-80.
- Putta, S., 1990, Weighing Externalities in New York State, Electricity Journal, July.
- _____, 1991. Methods for Valuing and Incorporating Environmental Costs in Electric Resource Planning and Acquisition, in O. Hohmeyer and R.L. Ottinger (eds.), External Environmental Costs of Electric Power, Springer-Verlag, NY.
- Quaddus, M.A., and T.N. Goh. 1985. Electric Power Generation Expansion: Planning with Multiple Objectives, Applied Energy, 19, 301-319.
- Rahman, S. and G. Shrestha, 1992. A Technique to Incorporate New Information in Evaluating Generation Alternatives, IEEE Trans. Pow. Systems, 7(2), 900-906.
- Ramirez-Rosado, I.J., and R.N. Adams. 1991. Multiobjective Planning of the Optimal Voltage Profile in Electric Power Distribution Systems, COMPEL, 10 (2), 115-128.
- Rau, J., and D. Wooten, 1980. Environmental Impact Analysis Handbook, McGraw-Hill, NY.
- Rowe, M. B. B.F. Hobbs, B.L. Pierce, and P.M. Meier, 1979. Evaluation of Nuclear Power Plant Siting Methodologies, NUREG/ CR-1689, Brookhaven National Laboratory, Upton, NY.
- Ruff, L. 1991. Internalizing Environmental Costs in Electric Utility Decisions, Putnam, Hayes, and Bartlett, Inc., London, England.
- Saaty, T. L., 1980. The Analytical Hierarchy Process: Planning, Priority Setting, Resource Allocation. McGraw-Hill, New York, NY.
- Sarin, R. K., J. S. Dyer and K. Nair, 1980. A Comparative Evaluation of Three Approaches for Preference Function Assessment. ORSA/TIMS Joint National Meeting.
- Schmalensee, R. L. 1992. Testimony of R.L. Schmalensee, Docket No. D.P.U. 91-131, Commonwealth of Massachusetts Department of Public Utilities, Boston, MA.
- Schoemaker, P. J. H., 1991. Behavioral Issues in Multiattribute Utility Modeling and Decision Analysis, in J.N. Morse (ed.), Organizations: Multiple Agents with Multiple Criteria, Springer-Verlag, NY.
- _____, 1991, When and how to use scenario planning: A heuristic approach with illustrations. Journal of Forecasting, 10, 549-564.
- Seattle Environmental Priorities Project Technical Advisory Committee, 1991. Environmental Risks in Seattle: A Comparative Assessment, City of Seattle, Office of Long-Range Planning, Seattle, WA.
- Seattle Post-Intelligencer, 1992. Seattle, WA, Oct. 6, p. B3.
- Seattle City Light, 1989. Long Run Forecasting Model, Seattle, WA.
- Seattle City Light, 1991. Strategic Corporate Plan 1990- 1991, Second Edition, City of Seattle.
- Stanton, T. 1990. Unpublished report, Michigan Public Service Commission, E.Lansing, MI.
- Staschus, K., 1991. A Multi-Attribute Evaluation Framework for Electric Resource Acquisition in California, Elect. Power & Energy Sys., 13(2), 73-80.
- Stevens, S. S., 1966. On the Operation Known as Judgment, American Scientist, 54, 4, p.385-401.

- Stewart, T. R. and D. W. Ely, 1984. Range Sensitivity: A Necessary Condition and a Test for the Validity of Weights. National Center for Atmospheric Research, Technical Report NCAR 3141-84/14. Boulder, CO.
- Stewart, W.R., Jr., and E.R. Horowitz. 1991. Environmental Factor Weighting at the Federal Energy Regulatory Commission, Socio-Economic Planning Sciences, 25(2):123-132.
- Stillwell, W. G., D. Von Winterfeldt, and R.S. John, 1987. Comparing Hierarchical and Nonhierarchical Weighting Methods for Eliciting Multiattribute Value Models, Management Science, 33(4), 442- 450.
- Synergic Resources Corporation, 1993. COMPASS, Bala Cynwyd, PA.
- Talaq, J. H., F. El-Hawary, and M. E. El-Hawary, 1994. A Summary of Environmental/Economic Dispatch Algorithms, IEEE Transactions on Power Systems, 9, In Press.
- Temple, Barker & Sloane, 1991. Environmental Externalities: An Overview of Theory and Practice. EPRI Report CU/EN-7294, Palo Alto, CA.
- Tietenberg, T. H., 1980. Transferable Discharge Permits and the Control of Stationary Source Air Pollution: A Survey and Synthesis, Land Economics 56, 391-416.
- Turvey, R. and D. Anderson, 1977. Electricity Economics: Essays and Case Studies, The Johns Hopkins University Press, Baltimore, MD.
- Vine E. L. and J.P. Harris, 1990. Evaluating Energy and Non-Energy Impacts of Energy Conservation Programs: A Supply Curve Framework of Analysis, Energy, Vol. 15(1), 11-21.
- von Nitzch, R., and M. Weber, 1993. The Effect of Attribute Ranges on Weights in Multiattribute Utility Measurements, Management Science, 39 (8), p.937-943.
- Von Winterfeldt, D., and W. Edwards, 1986. Decision Analysis and Behavioral Research, Cambridge University Press, New York.
- Weber, M. and K. Borcharding, 1993. Behavioural Influences on Weight Judgments in Multi-attribute Decision-making, European Journal of Operational research, 67, p.1-12.
- Weedy, B.M. 1989. Environmental Aspects of Route Selection for Overhead Lines in the U.S.A. Electric Power Systems Research, 16(3), 217-226.
- Weil, S. 1991. The New Environmental Accounting: A Status Report, Electricity Journal, 4(9), 46-54.
- Wile, J. H., 1991. Impacts of the 1990 Clean Air Act on Utility Planning, The Electricity Journal, 4(6), 46-51.
- Wilson, A. F., 1989. Value of Energy Study, Seattle City Light, Seattle, WA.
- Wood F. P. and R. Naill, 1992. Externalities in Utility Planning: What Would It Cost?, The Electricity Journal, 5(7), 35-43.
- Woodward-Clyde, Inc. 1981, Decision Framework for Technology Choice: Case Study of One Utility's Coal-Nuclear Choice, Electric Power Research Institute, EA-2153, Palo Alto, CA.
- World Bank, 1989, Striking a Balance: The Environmental Challenge of Development, Washington DC.
- Yang, H.T., and S.L. Chen. 1989. Incorporating a Multi-Criteria Decision Procedure into the Combined Dynamic Programming/Production Simulation Algorithm for Generation Expansion Planning, IEEE Transactions on Power Systems, 4(1), 165-175.
- Yuen, P.-F. and T. Larson, 1992. Receptor Modeling of Annual Average PM2.5 at a Residential Site in Seattle Washington, Proceedings, 1992 Annual Meeting, Pacific Northwest International Section of the Air & Waste Management Association, Bellevue, WA, Nov. 11-1