# Multiple Criteria Decision Analysis: Procedures for Consensus Seeking in Natural Resources Management

T J Stewart , A Joubert, L Scott
T Louw

WRC Report No.512/1/97

# Multiple Criteria Decision Analysis: Procedures for Consensus Seeking in Natural Resources Management

T J Stewart, A Joubert, L Scott and T Low

Department of Statistical Sciences
University of Cape Town
Rondebosch
7700
South Africa

Report to the Water Research Commission on the project "The development of procedures for decision support in water resources management"

Head of Department : T J Stewart
Project Leader : T J Stewart

WRC Report No : 512/1/97
ISBN No : 1 86845 318 9

# Contents

# Glossary of Terms

**AHP – Analytic Hierarchy Process:** A method of MCDA or MCDM (see below) developed by Saaty (1980) and based on pairwise comparisons of criteria and of alternatives.

**BCA:** *Benefit-Cost Analysis* – see Chapter 3 for discussion

**CBA:** *Cost-Benefit Analysis* – an alternative term for BCA

**Contingent Valuation Method:** A method of valuing conservation areas, for example, by the amount people are willing to pay to conserve the area, or by the compensation they would be willing to accept for its loss. See page 47.

**Criterion:** A particular point of view or interest according to which policy or decision alternatives may be compared.

**CVM:** Contingent Valuation Method.

**Decision conference:** A workshop involving different interest groups, at which alternative courses of action are defined and/or compared using formal methods of decision analysis, as described in Appendix A

**EIA:** Environmental Impact Assessment

**ELECTRE:** A method of decision analysis based on "outranking" (see below); the approach incorporates concepts of voting (the weighted number of criteria favouring a particular course of action) and vetoes (in which a course of action can only "outrank" another if there is no criterion which very strongly prefers the latter).

**IA:** Impact assessment

**LDC:** Less(er) developed country

**MAUT – Multi-Attribute Utility Theory:** A method for MCDA or MCDM based on the construction of value functions consistent with a number of axioms of rational decision making; this is a fundamental basis for the assessment phase of scenario based policy planning.

**MCDA – Multi-Criteria Decision Aid (or Analysis):** The field of management science concerned with providing decision support in contexts in which substantial conflict between goals or criteria exist.

**MCDM – Multi-Criteria Decision Making:** The process of decision making in contexts in which substantial conflict between goals or criteria exist.

**Outranking:** An approach to MCDA, largely arising in Europe, and particularly France, in which the unique approach is to identify strengths of evidence for and against the assertion that one alternative is more preferred than another. See also ELECTRE, which is the most common algorithmic implementation of outranking.

**Pareto optimality / Pareto criterion:** A principle whereby one course of action can be declared inefficient, dominated or undesirable if there exists another feasible alternative at least as good as it from all points of view (interests, criteria), and strictly better on at least one of these. A converse view is that one course of action can only be said definitely and unambiguously to be better than another if it dominates the latter in the above sense.

**Policy scenarios:** Descriptions of possible policies, defined at a sufficient level of detail to allow all parties and interests to distinguish and to express preferences between the set of policies under consideration. See Chapter 2.

**SBPP:** Scenario-Based Policy Planning, as described in Chapter 2 and Appendix A

**SMART – Simple Multi-Attribute Rating Technique:** An implementation of MAUT which is the core of the thermometer scale and swing weighting assessments (see below) of scenario based policy planning.

**SODA:** Strategic Options Development Analysis – see Section 6.2.

**SSM:** Soft systems methodology – see Section 6.3

**Swing weighting:** A method of determining importance weights for criteria, as described in Appendix A, page 123.

**Thermometer Scale:** A method of directly associating relative scores with a set of policy scenarios or courses of action, as described in Appendix A.2.3.

**WTP:** Willingness to pay – see page 41

# List of Figures

# List of Tables

# Executive Summary

## ES1 Introduction

This report documents the results of the research project entitled "Development of procedures for decision support in water resources management", which itself was a follow-up to an earlier project described in Water Research Commission Report No 296/1/93 ("Scenario based multicriteria policy planning for water management in South Africa", Stewart *et al*, 1993). The primary findings of this earlier study can be summarised as follows:

1. In the context of strategic decision making for the development of water resources, particularly when there are substantial conflicts between different interests and groups, it is necessary to adopt an iterative *scenario-based* approach. Each scenario is characterised by a set of key policy choice combinations, together with those consequences which can be established in the light of current knowledge (by expert evaluation or by the use of available hydrological, economic and other models). The level of detail included in these *policy scenarios* should be just sufficient at each iteration of the process to allow each interest or group to compare and to distinguish between them.

   At early stages, the scenarios will represent relatively broad policy directions (typically including a *status quo*, or "do nothing" scenario), which can be refined and split into sub-options (with consequently more detail required) as the better compromises or consensus emerge.

2. Evaluation and comparison of policy scenarios by different interests or groups can be done effectively by ordering the scenarios along a "thermometer scale", in which the "gaps" between successive scenarios in the preference ordering represent strengths of preferences. This provides a mechanism whereby interests and goals of differing levels of measurability or quantification can be compared and communicated between different groups. Even within a supposedly single interest (e.g. conservation) there may exist conflicting goals or criteria of evaluation. By use of weighted averages of the thermometer scales for each criterion, coupled to sensitivity analysis, an aggregate thermometer scale for this particular interest can emerge, and can, in turn, be compared with those coming from other interests.

Details of the scenario-based policy planning approach are given in Chapter 2 and Appendix A of this report. The previous study provided evidence that this approach is both easily understood and broadly acceptable to a wide range of interests and groups. The thermometer scales provide a transparent means of comparing "apples and oranges" (interests of widely differing natures), in such a way that the manner in which the different concerns have been taken into consideration is systematic, well-documented and broadly understood. This is a step beyond "cost-benefit" analysis, in which the translation of some consequences into monetary terms is often neither well-understood nor broadly acceptable. The procedure as a whole is consistent with concepts of integrated environmental management.

The present report deals with developments and refinements, testing and implementation of scenario-based policy planning as a decision aid for water resources management. The issues covered are:

1. Comparisons between the multi-criteria decision analysis framework of scenario-based policy planning and other approaches such as cost-benefit analysis (Chapter 3);

2. Experiences from three case studies (Chapter 4);

3. A variety of developments concerning the methodology and its implementation (Chapter 5), such as:

   (a) Adaptation of the procedures to cope with less precise scoring procedures than the thermometer scales;

   (b) Adaptation of the concepts for the purpose of constructing importance indices (for prioritising conservation areas for example), rather than the evaluation of policy scenarios directly; and

   (c) Theoretical testing of the robustness of the procedures to the underlying assumptions, by means of simulation models;

4. The use of "soft" methods of operations research to facilitate problem structuring (Chapter 6);

5. Conclusions and Recommendations (Chapters 7).

The remaining sections of this chapter (executive summary) elaborate on each of the above points.

# ES2 Multi-Criteria Decision Analysis (MCDA) and Economic Evaluations

An important component of the research reported herein has been an evaluation of the theoretical and practical differences and similarities between benefit-cost analyses (BCA) and multi-criteria decision analyses (MCDA), particularly as they may be applied to projects or policies which have environmental and income distributional impacts. BCA is relatively well-known, whereas MCDA and the distinctions between the underlying principles of the two approaches are less well understood by many.

Ideally, these two decision-aiding tools should be used in a complementary way with due consideration given to their respective weaknesses.

BCA has its roots in neo-classical economics and utility theory, while MCDA is essentially also utilitarian in origin. Important considerations when contrasting the two approaches are the role of the Hicks/Kaldor compensation principle (see next paragraph), the use of income distributional weights and utility maximisation within BCA.

The Hicks/Kaldor compensation principle was proposed as a more realistic criterion for policy evaluation than that of Pareto, which is essentially a test of efficiency only. The Pareto principle requires only that the welfare of at least one person is improved without diminishing the welfare of anyone else. The Hicks/Kaldor principle suggested that welfare could only be said to have improved if those who "gain" could compensate those who "lose" in some manner, and still retain a positive net benefit. This provides a stricter ordering of alternatives than is possible by the Pareto principle, but can lead to a possible pro-rich bias for three main reasons. Firstly, the principle does not require that compensation actually is paid. Secondly, most people agree that at the same levels of monetary compensation, the marginal utility of the poor is greater than that of the rich. As these differences in marginal utility for the same payment are not considered in BCA, projects which benefit the rich may be favoured as they are both more willing and able to compensate those who have lost utility. Thirdly, when the environment is a source of direct subsistence to the rural poor, the cash value of the livelihood is not the issue, but rather the lack of any viable alternative.

This pro-rich bias may further be exacerbated by the manner in which income-distributional weights are or are not employed in BCA. Without such weighting, BCA does not automatically lead to the maximisation of social welfare, but the allocation of weights in the absence of explicit measurement of relative marginal utilities remains arbitrary. In MCDA, the actual participation of the stakeholders (including "the poor") in the process leads to a direct evaluation of changing marginal preferences (i.e. "utilities"). While weighting of these utilities directly may be used in MCDA together with extensive sensitivity analysis, the emphasis is on retaining measures of goal satisfaction for each interested party separately, in the search for equitable compromise (see Chapter 2).

The discussion of practical issues in Chapter 3 considers the different approaches to the multiple dimensions present in all resource management problems, and the implications of these differences for the manner in which trade-offs are made and the data which can be used. In BCA these all relate to the use of monetary scales. MCDA on the other hand, in its use of interval scaling and weights, and focussing on relative trade-offs within each dimension, avoids many of the problems associated with monetary evaluation techniques, while still permitting assessment of potential trade-offs between criteria. At the same time, MCDA allows the use of both quantitative and qualitative criteria which means that the types of issues considered are not constrained by the need for monetary values. These practical differences have implications in terms of acceptability, which may be further emphasised by the way in which the public are included in each of the processes. BCA includes public input either as a stage of an environmental impact assessment process, at which point the public may raise issues of concern, or in contingent evaluation type questionnaires (in which the rationale is that individual preferences are expressed through choices which are reflected in the prices people pay). The aggregate of individual preferences is then taken to reflect societal preferences. In the MCDA approach, public inputs occur through the raising of any issues of concern ("criteria"), while stakeholder groups are involved in the actual evaluation and comparison of alternatives based on the issues that they have decided are most important.

# ES3 Experience from Case Studies

In order to obtain practical empirical validation of the procedures, attempts were made to become involved in actual policy problem areas as part of the research project ("action research"). For a variety of reasons, this did not always turn out to be possible, but we report on three studies in which we had significant involvement and from which useful lessons were learnt.

**Kruger National Park Rivers Research Programme:** This study was undertaken in conjunction with the decision support systems sub-programme of the Kruger National Park Rivers Research Programme (KNPRRP). The scenario-based policy planning constituted only part of a series of workshops held to develop a conceptual decision support framework for the KNPRRP.

For the evaluation phase, five scenarios were considered, which included the *status quo* and pre-development scenarios (even though the latter was unachievable). Use of such benchmarks proved valuable, and could be recommended more routinely for scenario-based policy planning. A difficulty in this study was that there was insufficient time in the workshops to properly establish operationally meaningful and independent criteria, perhaps because the evaluation phase was tacked on at the end of the workshops, rather than being an integral part of them. This clearly reinforced the need for time and effort to be taken in the problem structuring phases to identify the relevant criteria, if the benefits of scenario based policy planning are to be realized.

In spite of the problems, however, the decision analysis phase did generate at least one useful insight. A clear conclusion drawn from the ordering of the scenarios was that the best management plan on offer would cause the same degree of environmental upheaval relative to present day conditions as the destruction involved in going from pristine (pre-development) conditions to the present day level of development (*status quo*). The ability to make this sort of statement in effect provides environmentalists with a "currency" within which to convey their attitudes/opinions about proposed strategies.

**Conservation importance of estuaries:** This case study provided the motivation for developing methods for constructing importance indices (see later). At the stage at which we became involved, there had already been attempts at constructing scores, but there had been some concern concerning the validity of these. A workshop was arranged, following the recommendations summarized in Appendix A, but with separate "scenarios" developed for each criterion. To a large degree, consensus was reached within the workshop, which was then followed up by a postal exercise using conjoint scaling (more fully described in § 4.2). A high degree of consistency between the two methods of assessment was observed, but the conjoint scaling exercise did highlight a lack of independence between certain of the water quality criteria. Nevertheless, a broadly acceptable scoring system was ultimately derived.

**Forestry land-use (Maclear district):** This project was carried out (and at time of writing is continuing) in conjunction with the Department of Nature Conservation of the University of Stellenbosch, and was the most comprehensive of three case studies reported, in terms of the use made of scenario-based policy planning. In one of the workshops, use was also made of the electronic meeting room operated by UCT's Centre for Information Systems. This experience revealed some of the weaknesses of this high technology approach (especially as a result of vast differences in computer literacy of different role players), but also suggested its potential as a decision aiding tool which needs to be followed up in later research.

The problem initially presented itself as that of needing to develop a transparent, justifiable and efficient procedure for the allocation of forestry permits. The first workshop over two days focussed on this aim, attempting to use the decision analysis framework to develop a procedure which did take interests of the forestry industry, conservation, water and the local community into account. The process of formulating measures of achievement on the criteria led to a clear realization that such measures made little sense while decisions were made incrementally (e.g. permits for one farm at a time). It was agreed that further workshops be arranged, at which development scenarios for forestry in the region as a whole would be considered. This reinforced our earlier conclusions that scenario-based planning was necessary to ensure policies which satisfied all interests in the long term.

Two further workshops have to date been carried out, focussing on a number of development scenarios, differing primarily in terms of the size and placement of forestry in the Maclear district. At the first of these latter two workshops, a high degree of consensus was reached as to the appropriate criteria according to which the scenarios should be compared and evaluated. The approach appeared to provide a far better basis for expressing preferences in principle, but at both workshops it emerged that there were still considerable differences of opinion as to precisely what each scenario represented, and the consequences of these. Two of the differences of this nature which arose, and which illustrate the problems of reaching a final conclusion, were as described below.

- One view was that land allocated to forestry was automatically lost to conservation, whereas the counter view was that forestry could and would never use all the ground permitted for forestry, which would retain conservation corridors.
- Order of magnitude differences emerged as to precisely what employment might be created (directly and indirectly through a "multiplier effect")

in the region as a result of the establishment of pulp mills for example; these issues need still to be resolved through the use of broadly acceptable economic models.

These differences strengthened our conclusion that considerable time and effort needs to go into defining scenarios, if the full benefits of scenario based policy planning are to be realized, and sound and justifiable decisions reached. The project is (as at December 1996) still continuing, with a view to developing more comprehensively defined scenarios.

# ES4 Development of Methodology

Apart from the action research components, a considerable part of the research, as described in Chapters 5 and 6, went into developing and refining the underlying methodology of scenario based policy planning. Some of these developments derived very much from experience with the case studies, but a systematic sequence of simulation studies was also conducted in order to investigate more theoretical properties of the procedures, with a view to understanding their potential shortcomings and to fine tune them for practice. These developments are described in the following subsections.

## ES4.1 Imprecise inputs

A key part of the scenario based policy planning procedure as initially formulated was the evaluation of scenarios on the thermometer scales. Our experience has been that respondents have little difficulty in expressing their views on the scenarios in this way (refer to the case studies), but there have nevertheless been questions as to how to proceed if people find themselves unable to do this with precision. If there is major conflict within a group as to how to position the scenarios along the scale, then this indicates the existence of important sub-criteria or sub-interests, which need to be identified and made explicit, and the evaluations performed separately for each of these. By "imprecise" inputs, therefore, we do not mean substantial disagreement on the basic rank ordering of the scenarios in terms of a specified interest or criterion, but rather some uncertainty as to the relative sizes of the "gaps" between them. Simulation studies (see §ES4.4 below) have demonstrated that results are quite robust to such imprecision, but it is nevertheless useful to have a means of treating it.

In the above sense, the most imprecise inputs occur when only a rank ordering of the alternative scenarios according to each criterion or interest can be given. It is demonstrated in Section 5.1 that it is possible in principle to characterise all possible overall rankings of the alternatives which are consistent with stated orderings according to each criterion and the existence of some form of value function structure. The details require quite sophisticated linear programming formulations, but in essence it is possible to calculate for each alternative the highest ranking it could ever have in any prioritization consistent with the individual rank orders and the underlying rationality assumptions. In this way, at least a shortlist of scenarios can be generated for more detailed evaluation, in the knowledge that nothing is lost through discarding the remaining options.

The linear programming formulation can also be adapted to cater for situations in which intermediate levels of information (for example, relative ordering of the magnitudes of the "gaps", or bounds on the ratios of successive "gaps"), but not a precise thermometer scale, is available. In this way, the procedure can be made generally applicable to a much wider range of settings. These options have not yet been built into the software described in Appendix B, but the intention is to do so in a later version. When this is done, the user will not need any knowledge of linear programming in order to apply the ideas.

## ES4.2 Construction of importance indices

A situation which seems to arise relatively often (such as in the case study on the conservation importance of estuaries) is that in which there is not an immediate decision to be made amongst competing scenarios, but rather a need to prioritise a number of items for future action or attention. In the estuarine case study, the aim was to associate some form of index indicating the importance of conserving (protecting from development) each estuary. The aim, of course, is to provide a clear picture to policy and decision makers as to what areas should be earmarked for protection and which for development.

In principle, the same approach can be adopted as for scenario-based policy planning, with some modifications. As the importance rating system will need to be applied to potentially very many items (estuaries in the case study), some of which may not even be identified at the time of the workshop, it would not be feasible to compare all of these against each other in making the evaluations. In this case, one needs to construct small numbers of scenarios *separately* for each criterion, which are then ordered along the relevant thermometer scale in the usual way. Once this has been done for each criterion, weights can still sensibly be assessed for each criterion, taking into account the ranges of scenarios considered for each. In evaluating any specific item at some future date, it requires only that for each criterion in turn, the scenario which best describes the item be identified, and in this way a full score for the item can be constructed.

As discussed in the relevant case study, we have found it feasible and useful to apply a method called conjoint scaling (used in market research studies) as a consistency check on the scoring systems and weights generated in this manner.

## ES4.3 Problem structuring aids

Implementation of scenario-based policy planning involves two key problem structuring phases. The first is the generation of policy alternatives (or policy "elements" as described in the previous report) in an innovative and constructive way, while the second occurs within each workshop, as different interest groups come to grips with the definitions of the criteria by which they wish to evaluate and to compare the policy scenarios. Both of these structuring phases require the building of a shared understanding amongst members of the groups concerned, as to their perceptions of the dynamics of the situation, the goals to be attained, and the options which are available. In addition to the already structured processes of scoring scenarios along the thermometer scales, the above problem structuring processes also need facilitation.

Of recent years, within the general field of operational research/management science, there have emerged a number of techniques which have been termed "soft" operational research or "soft systems" approaches to facilitate processes such as those described in the previous paragraph. The aim of all these approaches is to assist individuals or groups to build up their understanding of a system in a structured and logical framework (often with the aid of appropriate computer support aids), but without (initially at least) forcing this structure into a rigid mathematical framework, which was the classical OR approach.

Although there was not time within the current project to test out these problem structuring techniques in detail, an evaluation has been done of the relevant literature and its applicability to water resources planning in the scenario-based framework. This is described in more detail in Chapter 6 of the report.

## ES4.4 Theoretical testing by simulation models

One difficulty for research into decision support procedures is that the "correct" answer is unknowable, and thus one cannot directly assess the quality of solutions obtained. Furthermore, the same strategic decision will not be repeated many times, and thus statistical tests of differences between the quality of results obtained with and without the use of decision support are also not possible. The original report did record some measures of user satisfaction with the approach adopted, but even this is of limited value in attempting some form of scientific validation of the approach.

For this reason, we have adopted a simulation approach. Hypothetical decision contexts, defined in terms of alternatives differing on a range of attributes and ideal preference structures according to each interest or criterion, are randomly generated on the computer. The procedures are then applied to this context, building in non-idealities (such as imprecision in stating preferences or scores). By having knowledge of the "true" preferences which exist, it is possible to evaluate the extent to which the procedures facilitate the finding of the best compromises, and/or the extent to which the procedures may bias results inadvertently in one or other direction. Two sets of such studies were conducted during the course of the project, viz:

1. **A general study of the scenario-based policy planning procedure as a whole:** The decision context was broadly based on the Sabie-Sand rivers catchment case study described in the previous report, and was aimed at evaluating different methods of screening alternatives based on the responses of different interest groups. The context was that of an initial set of "background scenarios", from which a smaller number of "foreground scenarios" were chosen for evaluation by a number of simulated interest groups. On the basis of these evaluations, a certain number of scenarios were eliminated, and a new set selected for a second round of evaluation. After further evaluations by the interest groups, a final choice was made, and compared with the best compromises known to exist (since the simulated preference structures are known). Different simulations evaluated different methods of screening and different set

sizes. The basic conclusion was that the best methods of screening were either a simple weighted sum of thermometer scale scores (as proposed in the previous report) or a worst case analysis, i.e. attempting to maximize the minimum thermometer scale score taken across all interests. Both of these approaches came consistently very close to the best options available, and were robust to sizeable imprecisions in input data, with the "max-min" approach demonstrating slightly greater robustness.

Overall, therefore, this set of simulations demonstrated the validity of the proposed procedure in a different sense to what can be achieved by case studies. This work formed the basis of an MSc dissertation which was accepted with distinction (Heynes, 1995).

2. **A study of the validity of using additive functions**, i.e. weighted averages of thermometer scales: The use of additive functions can be justified mathematically on the basis of a small number of axioms of rational decision behaviour. In practice, however, the validity can be compromised by definitions of interests and criteria in a manner which violate a technical form of independence; by evaluation procedures which do not give adequate attention to different "gap" sizes between successive scenarios (for example by simply using rank orders, as discussed earlier in §ES4.1); and by a variety of imprecisions in evaluations which may arise.

   This simulation study investigated a simple once-off choice from amongst a set of randomly generated alternatives, for a fully-specified idealized preference structure. The choice was based on an additive function incorporating the various non-idealities described in the previous paragraph. As with the first set of studies, the methods showed considerable robustness to substantial imprecision in input. However, overly linearized responses and definitions of interests which were too highly interdependent were found to lead to potentially quite serious biases in the results. This reinforces our recommendations for quite careful implementation of scenario-based policy planning in workshop contexts (as described in Appendix A). This work has been published in an international journal (Stewart, 1996).

# ES5 Conclusions and Recommendations

This report motivates and justifies the conclusion that the evaluation of water resource development alternatives should be conducted within the framework of clearly defined development scenarios, making use of the principles of multicriteria decision analysis (as described in Chapter 2 and Appendix A). In order both to "sell" this concept to interested and affected parties, and to refine the implementation of the procedures within the southern African context, the following key areas of continuing research are recommended:

- More "action research" in monitoring implementation of the procedures, and in taking the procedures to the communities;

- Development of the use of "soft" problem structuring methods, for both defining scenarios and facilitating identification of criteria, and the use of modern technologies such as electronic meeting rooms;
- Refinement of software to incorporate more of a "multi-media" approach.

A follow-up project addressing these issues has been approved by the Water Research Commission. This research, however, does need to go hand-in-hand with:

- the establishment of training courses which can be offered both to facilitators and to interested groups who are involved in debates over future water management policies; and
- the publication of a non-technical information document, describing how the procedures developed herein can assist the various stakeholders in the water community on South Africa.

# Chapter 1

# Updated Literature Survey

A literature review on the general topic of multiple criteria decision making (MCDM), or decision aid (MCDA), and their application to natural resources management, appeared as Appendix A of the previous WRC project report (Stewart et al., 1993). Much of the terminology used in the literature was defined and the main underlying principles of four main approaches to MCDA were described. The four approaches were classified as utility and value function approaches, goal programming and reference point techniques, outranking approaches, and game theoretical approaches. A critical comparison was made of the two most commonly used utility/value function approaches, namely multiattribute utility theory (MAUT), in forms such as "SMART", and the analytic hierarchy process (AHP). This served as part of the justification of the use of variations of SMART in practical application to the Sabie river catchment.

The previous review of the literature was divided into studies concerning: (1) Qualitative or quantitative goal assessment; (2) Operational and medium term planning of water resources; (3) Long term planning of water resources; (4) MCDM in other resource planning problems; and (5) Group decision support and models for conflict resolution.

A more technical review of MCDM and MCDA can also be found in Stewart (1992).

The primary aim of this chapter is to review additional relevant material which may have been published since the preparation of the previous report, but some older material is also included where this is considered necessary for completeness. The issues addressed here include: (1) The use of linguistic or semantic scales to elicit preference statements; (2) uncertainty or imprecision in information and in value judgements; (3) Conflict Resolution; (4) Comparative studies, including problem structuring, model complexity and user friendliness; and (5) Applications of decision-aid to water resource and catchment management.

## 1.1 Use of linguistic or semantic scales

There have been various initiatives over the last few years to formulate decision-aiding procedures which make use of linguistic or semantic statements of degree of preference which are then translated to a numerical scale. The forerunner of these approaches was the analytic hierarchy process (AHP), which has been subjected to criticism for various reasons (see for example, Stewart et al., 1993), but the use of linguistic or semantic statements of preference is appealing and various researchers have pursued the idea in different ways. Briefly, the original AHP used a nominal 9 point scale of preferences of alternatives which were represented linguistically to the user (in terms such as "absolute preference", "weak preference" or "indifference", with each of which were associated numerical values), as described in Saaty (1980).

One approach to addressing the criticisms of the initial AHP model has been developed at the Delft University of Technology. For example, Lootsma (1993) looked at scale sensitivity in AHP, and, in response to the various criticisms of the original AHP, proposed some adjustments which are included in their "REMBRANDT"[1] system. Criticisms of the original AHP included: the "fundamental" 9-point scale which is used to quantify human judgement, the use of eigenvectors to estimate the impact scores, and the use of an arithmetic aggregation rule. How these criticisms are addressed by the REMBRANDT approach is beyond the scope of the present review, but details may be found in Lootsma (1993) and Olson et al. (1995). The statements of preference, however, remain essentially the same as in AHP, but the numerical values and subsequent processing of these numbers differ. Much of the development of the REMBRANDT approach can be viewed as based on the theoretical work of Barzilai et al. (1987) and Barzilai and Golany (1994), in which it is demonstrated that only multiplicative (rather than additive) aggregation of scores obtained by pairwise ratio assessments is consistent with a number of desirable axioms.

Another European initiative (Portuguese and Belgian) has pursued a different approach in a procedure termed "MACBETH"[2], as described by Bana e Costa and Vansnick (1994). Pairs of alternatives are compared and respondents asked asked to allocate the difference in attractiveness between them to one of six categories. Using linear programming methods similar to those described in Appendices C.2 and C.3 of this report, consistent numerical scores are found. An important difference between this idea and that of AHP is that the judgements concern differences of attractiveness rather than ratios of priorities or importance, which allows for simple additive aggregation of scores.

Although in theory emerging from a different school of thought, certain methods based on fuzzy set theory are, in practical terms, rather similar to utility or value function approaches. Fuzzy information can be included in decision models in two different ways: by using linguistic variables, or by using fuzzy numbers (Munda et al., 1994). In their explanation of the fuzzy approach these authors introduced a function $u_A(x)$ which specifies the grade of membership of $x$ in the set $A$. In their practical example, criteria levels were found for each of three alternatives, on natural scales (e.g. Guilders) or qualitative scales (excellent, good, moderate). Then each of these was given a membership degree – on the interval [0,1] – which indicated membership of "each action to the interval of feasible and acceptable values defined on each criterion". In practice this is not dissimilar to rating alternatives on a [0,1] preference interval scale for each criterion. Munda et al. then continue by specifying

[1]An acronym for Ratio Estimation in Magnitudes or deciBels to Rate Alternatives which are Non-DominaTed

[2]An acronym for Measuring Attractiveness by a Categorically Based Evaluation TecHnique

the "degree of truth" of three linguistic comparisons of each pair of alternatives: $a_1$ is better than $a_2$, $a_1$ and $a_2$ are indifferent, $a_1$ is worse than $a_2$ etc. The process generates a partial ranking only. Smith (1994) also discusses in some detail the theory of fuzzy sets and numbers in the context of two examples: the siting of a hazardous waste disposal site and the building of a transport route, but do not illustrate the use of linguistic scales. Their explanation and approach is similar to Munda et al. (1994). A slightly different approach was followed by Xiang et al. (1992), who look at the application of fuzzy-sets to a land-use planning problem. A continuous real interval [0,1] is used as the preference measurement system, with descriptive labels for 'control points' on this interval i.e. 0 = absolutely unsuitable, 0.3 = unsuitable, 0.5 = moderate, 0.7 = suitable, 1.0 = perfect. A similar scale was used to derive the weights of the attributes: 0 = not important, 0.3 = less important, 0.5 = moderate, 0.7 =important, 1.0 = very important. A third scale measures the 'tolerance' of the respondent on each attribute. This seems to relate to what Munda et al. (1994) refer to as a measure of 'incertitude', but is not further explained. The meaning of the 'tolerance' scale of Smith (1994) is not clarified in the questionnaire used in the study, which must cast some doubt on the reliability of the results.

Apart from the usefulness of the linguistic statements there appear to be no substantial advantages in the 'fuzzy' approaches. They tend to be difficult to comprehend and seem to be individualistically used. Approaches such as MACBETH and REMBRANDT have used the advantages of linguistic judgements without the disadvantages of increased complexity, and remain consistent with the axiomatic foundations of decision analysis.

## 1.2 Uncertainty in information and value judgements

Decisions regarding projects which have environmental consequences may need to take into consideration various types of uncertainty such as long term environmental consequences, future political, social and economic conditions, and the dynamic relationships between these. Decision makers may wish to minimise the potential for loss of future opportunities or to 'avoid regret'. Uncertainty in preferences (which may be due to lack of precise information) is the main emphasis in fuzzy approaches, some of which were discussed in the previous section. MAUT, which addresses uncertainty through the analysis of expected values, is the only MCDM approach which formally attempts to include uncertainty in its axiomatic framework, although others may use expectation and/or variance as criteria (Stewart, 1995). Varis et al. (1994) look at how uncertainty and subjectivity in model structure, in objectives and in information can be structured and quantified. They divide decision-making into three components, viz. (1) the management objectives, goals, targets, criteria and constraints, (2) elements in the system that are under control, at least partially (decision variables), and (3) the information conditioning the above components (chance variables). Presentation or modelling of uncertainty in information may occur via non-numerical techniques (rule-based systems, production systems) or numerically via probability theory (e.g. Bayesian approaches, classical probabilistic approaches, fuzzy logic). Uncertainty in objectives (values, targets, preferences, constraints) is often dealt with through utility theory and analyses of attitudes to risk.

However, Faucheux and Froger (1995) argue that the use of Bayesian and expected utility approaches are inadequate for the 'strong uncertainty' characteristic of projects with environmental consequences. They distinguish two types of uncertainty: risk (weak uncertainty) and strong uncertainty. They discuss different decision-making models, their rationality assumptions and their implications when there is high uncertainty, irreversibility and complexity. Two types of rationality are considered, viz. substantive rationality (the rationality of a decision is considered independently of the way in which it is made – rationality refers to the results of the choice) and procedural rationality (the rationality of a decision is in terms of the way in which it was made – rationality refers to the decision-making process itself). They consider that stochastic environmental decision-making approaches which rely on Bayesian theory are based on a substantive rationality hypothesis, require the listing of all consequences and their probabilities, and cannot include the strong uncertainty characteristic of environmental decisions. They regard the use of expected utility as similarly unsatisfactory because the range and the distribution of future environmental effects are not known and/or unknowable. They clain that we do not act 'as if' these are known or knowable, but do not give any justification for this statement. They also argue that expected utility analysis does not take account of the notion of regret (Loomes and Sugden, 1982), which may be inappropriate in situations of irreversibility of environmental impacts. (This is, however, debatable, as regret can formally be included as criterion in MAUT.) They therefore propose an approach which has procedural rationality, and which entails: (1) The creation of a tree of goals and sub-goals, where the goals may be unmeasurable but with sub-goals at some level which are measurable; (2) The satisficing principle is then used (as opposed to optimising) to choose between options taking into account social, economic and ecological considerations. This is in essence the same as the value tree used in MCDM approaches but they use energy as indicators in the lower level sub-goals.

The requirement of MAUT that knowledge of the full multivariate distribution of outcomes be known, is problematic. Stewart (1995) used Monte Carlo simulation to analyse the effect of reducing this requirement in discrete multi-criteria problems with uncertainty. Although decision-makers violate many of the assumptions of rational decision making, the complexities introduced into methods in order to account for this do not necessarily change the decisions which may result. The analyses showed that finding the utility of an alternative using the product of the expectation of the marginal utilities does not introduce significant errors compared to using the (theoretically correct) expectation of the product of the marginal utilities which has complex or impossible input requirements. The next stage showed that a further simplification — to use an additive rather than multiplicative model — has similarly little effect. However, incorrect elicitation of the marginal utilities (such as assuming a linear function when it is non-linear) may cause more serious errors. (Recent extensions of this work is included in Section 5.4.2 of the present report.) This means that only the expectations of marginal utilities are required for each alternative. As such expectations can be approximated by working in terms of 3- or 5-point discrete distributions, uncertainties may in effect be represented by 3 to 5 scenarios.

An approach which appears to combine the advantages of utility and fuzzy approaches in terms of expressing uncertainty in preferences is that of Salo and Hämäläinen (1995). This originates from the pair-wise comparisons of AHP, but by combining this with a MACBETH like-approach, the authors work in terms of interval statements of preference. This allows the DM to make approximate ratio statements as intervals of values on a ratio scale. The adaptation originated with Saaty and Vargas (1987) and was introduced by them because of the acknowledged difficulties of eliciting exact ratio estimates. These ideas have been incorporated in the software program INPRE which allows the user to adjust the intervals with a mouse while the dominance relationships are updated graphically on the same screen. An equivalent approach has been incorporated into the software COMPAIRS (Salo and Hämäläinen, 1992) which analyses value functions in a similar way by allowing interval statements of preference. This work may also be seen to be related to that of Cook and Kress (1991), some adaptations of which are discussed later in this report (Sections 5.1 and 5.2).

## 1.3 Consensus seeking and conflict resolution

Another characteristic of decisions regarding projects with environmental impacts is that there are likely to be fairly high levels of conflict, often simplistically perceived to be development versus conservation conflicts. Many environmental decision-making articles therefore discuss consensus seeking and conflict resolution and not all of these can be considered here. A few papers, which have a wide range of subjects are summarised below.

Van Huylenbroek (1995) combines the outranking approaches (as described in the previous report, Stewart et al., 1993) of ELECTRE and PROMETHEE, and the preference functions of the latter, together with the conflict analysis test of ORESTE in a model called the conflict analysis model (CAM). The CAM model, which can be applied to both ordinal and cardinal data, combines the notions of indifference, incomparability, weak and strong preference (from ELECTRE), different types of preference functions (from PROMETHEE), and the conflict test (from ORESTE). Alternatives are given evaluation scores $e_j(a)$ based on each criterion $j$. The decision maker has to identify a preference function from a choice of six which most closely reflects how these scores are reflected on a [0,1] interval. The difference in evaluation scores on each criterion are translated to preference scores $\hat{e}_j(a,b)$ by the preference function identified. Then overall dominance of alternative $a$ over $b$, measured by a preference indicator $P(a,b)$, is found from a weighted sum of the preference scores. The preference indicator $P(a,b)$ is used to identify degree of conflict in the third, conflict analysis stage of CAM. Threshold values are identified which allow the specification of strong and weak preference, indifference or incomparability, depending upon the magnitude of the absolute difference $|P(a,b) - P(\text{b,a})|$. All stages of this approach seem unnecessarily complicated, with little apparent benefit. Although referred to as a conflict analysis tool, this aspect is not expanded, and in fact the result is simply an indication of strength of preference, or dominance and so the extra effort does not seem warranted.

An approach which is perhaps more straightforward and useful is that of Lewis and

Butler (1993) who proposed a three stage interactive framework for multi-objective multiple decision-maker situations. The procedure combines the SIMOLP (Simplified Interactive Multiple Objective Linear Programming) and Tchebycheff MOLP (multiple objective linear programming) optimisation methods with a preference ranking tool and a consensus ranking heuristic based on minimisation of regret. The first stage makes use of either or both of the MOLP procedures. In the second stage, the non-dominated alternatives are ranked by the individuals. An aggregate of individual preferences is needed for group consensus and preference determination which implies interpersonal preference comparisons. However, ordinal preference rankings will not normally allow the interpersonal comparisons of preference necessary for aggregation of individual preferences. Two processes allow them to bypass this problem: the format of the decision process itself, and the use of Cook and Kress's (1985) preferences scale. Firstly, the format suggested is a combination of nominal and interacting groups, where the talk-analyse-talk iteration, may encourage those present to make internal interpersonal comparisons of utility, or at least to fix the end-points of their internal scales (much as in our SBPP procedures). Secondly, in using Cook and Kress's preference scale, the DM allocates each of $n$ alternatives into one of $q$ positions on a nominal preference scale (where $n < q$). The relative positions of the alternatives in this scale indicate intensity of preference and the ordinal rankings are converted to cardinal scale on a [0,1] interval. During the third stage, the minimum regret heuristic of Beck and Lin (1983) is used to calculate a consensus ranking. This heuristic weights "regret" (that a desired ordering is not achieved) and "agreement" (when it is achieved) differently. If the DMs agree to the solution which is generated, then the process stops; otherwise, the three stages are reiterated. Experience with 83 students who tested the process in 20 groups, indicated that the nominal interacting groups were very important to the process, and that the choices were fair reflections of the aggregation of individual preferences.

As mentioned earlier Lootsma (1989) examined the pair-wise comparison of concessions for conflict resolution using linguistic statements of degree of acceptance of a particular deal (very strong liking, strong liking, weak liking, indifference, weak aversion, strong aversion, very strong aversion). These are converted to numerical values on a geometric scale, although the subsequent comparisons by a mediator may be scale independent. The concessions are thus presented as perceived by the parties concerned rather than, for example their monetary value. The advantage of this approach over MCDM alone are that, for situations of real conflict, a 'rate of exchange' of concessions is established which could be useful to a mediator.

Although some of the above rather theoretical explorations had illustrative examples, it is clear that in practical applications less complex approaches are favoured, as illustrated in the practical applications to wildlife management conflicts discussed below.

Maguire and Boiney (1994) used expected utility within a conflict resolution framework in their hypothetical, but realistic example of northern white rhino conservation in Zaire (where at the time of writing only about 28 individuals remained in the wild). They consider alternatives such as CBA, 'initial decision analysis' and adaptive environmental management, but reject these in favour of MAUT. This framework allowed for the updating of utilities and probabilities (using a Bayesian approach) with new information which facilitated compromise and encouraged the creation of new alternatives. The separation of the utility and probability models avoided the confounding of judgements about facts and judgements about values. The two extreme positions held by Western zoos and Zairean wildlife officials respectively were: remove all remaining animals to captivity to ensure safety from poaching or, allow no more removals but concentrate resources on more vigilant anti-poaching activities. Expected utilities and probabilities of outcomes of the different approaches were elicited from both groups and compared. A distributive bargaining alternative (phased removal to captivity) was not favoured by either group, even with sensitivity analyses. However, a contingency based alternative whereby particular actions taken contingent on observation of a particular event, provided a compromise agreeable to both: removal of the wild population if it dropped below a critical threshold.

Ralls and Starfield (1995) discuss the use of two decision analysis approaches to choose between eight possible management strategies applied to the problem of mobbing deaths in the endangered Hawaiian monk seal about which different interest groups have conflicting objectives. Both decision approaches required estimates of various probabilities (such as the probability of a population of more than 400 after 20 years) derived from simulation models. The first approach (borrowed from preemptive goal programming) was to satisfy each goal sequentially (i.e. to satisfy the most important goal first, and then the next etc.). In this approach the goals do not have to be independent, and in fact their lack of independence could be advantageous as the goals of one group can be met part-way in the higher level goals and then readdressed later in lower level goals once the needs of others have been addressed. The other approach was a variation of SMART and the probabilities of various outcomes were used as surrogate criteria by which the alternative management strategies could be measured (or, the squares of the probabilities for a more risk averse view). The results of the SMART approach agreed well with those of the lexicographic approach and both approaches were found (1) to provide an explicit, well documented, reproducible decision process; (2) to be easy to use and to understand; (3) to promote and focus discussion on objectives and priorities; (4) to facilitate a structured examination of multiple objectives and trade-offs; (5) to be flexible, non-prescriptive and suitable for workshops; and (6) to be robust.

McDaniels (1995) used multiple objective decision analysis (specifically MAUT) to conduct an *ex post* analysis of a specific fisheries management decision involving conflicting long-term objectives for stocks of sockeye salmon. In the salmon fishery, delaying the start of the fish season by a day, based on incoming information relating to the peak of the run, can mean an increase in potential benefits of millions of dollars. The broad level objectives were: long term stock health, short and long-term economic benefits, social acceptability and opportunities for learning. Surrogates were used (stock size in the year 2010) for some attributes while for others dollar values and constructed scales were used. The attributes were found to be (for the range of values in the present case) utility and additively independent. Trade-offs between each pair of attributes were determined directly in order to determine scaling constants. Objectives other than those conventionally used in fisheries modelling were found to be useful for in-season management.

Brown et al. (1994) used GIS, production models, and what they termed "multiple accounts" (the efficient frontier between NPV and caribou habitat), in order to provide decision support to resolve forestry and wildlife conservation conflicts between commercial forestry operations and the conservation of the caribou herd in the Mount Revelstoke and Glacier National Parks, Canada. The area of conflict was divided into forestry cutblock areas and variables relating to harvesting (log value for different species, road building, maintenance and hauling costs) were determined. Costs differed depending on slope and altitude etc. and were modelled using standard forestry models to obtain Net Present Value for each cutblock. The cutblock areas were categorised into areas of high, medium and low habitat suitability for caribou, based on elevation, slope, slope position, dominant species, height class and crown closure class. Five different scenarios (type of harvesting regime and extent of protected area being some of the discriminating features of these scenarios) were simulated over the next 120 year harvesting cycle. The multiple accounts study determined that the establishment of a park and the preservation of old-growth scenarios were inefficient. The areas of highest potential conflict (high caribou value and high NPV of timber) could be identified by the GIS overlay system, and thus attention could be concentrated on these areas. Cocks and Ive, (1996), used similar 'conflict indicator maps' in another forest land allocation problem.

## 1.4 Problem structuring and model complexity

There is no way of prescribing a specific technique for a particular problem, and those interested in using decision analysis approaches may be faced with a bewildering array of approaches and software packages. Tecle (1992) in fact claims that there are more than 70 MCDM techniques and at least 49 criteria for evaluating them. While theoretical papers propose techniques which may be interesting, useful or offer greater discriminating power, as has already been seen, these are often of a complex nature which may be of little use in real world situations where expertise and backgrounds may be widely divergent, and techniques need to be easily understood and decisions accessibly recorded. In addition, an analyst's preference for a technique may be based on experience with that particular technique rather than on its inherent superiority, and this experience will naturally influence the people with whom an analyst may work. This is clearly seen in the studies which compared the effectiveness and ease of use of different approaches from particular points of view (Tecle, 1992 and Hobbs et al., 1992): those techniques which are favoured in one comparative study are ranked low in the next.

Hobbs et al. (1992) compared the use of four different multicriteria methods (goal programming, ELECTRE I, additive value functions, and multiplicative utility functions) and three techniques for choosing weights (direct rating, indifference trade-offs and the analytic hierarchy process) using a real water resource planning problem in the USA. Ten alternative plans were evaluated using ten criteria (both qualitative and quantitative data). Eight hypotheses were tested relating to appropriateness and ease of use, validity, and differences in results of the multicriteria and weighting approaches. They concluded that the simplest methods were preferred (particularly additive value functions but also goal programming), but that no method was endorsed by a majority of the 21 participants. They advised caution with the use of a direct rating approach to determining weights, as the relationship between the weights and the range of values of the criterion was seldom kept in mind, and so the weights may fail to represent the trade-offs that users are willing to make. Methods which were poorly understood (AHP, ELECTRE, utility functions which included risk attitudes, gambling questioning) were not favoured.

Fifteen different multi-criteria methods which were applied to a watershed resources management problem were compared by Tecle (1992). The performance of the different techniques were assessed in terms of 24 criteria using composite programming (itself an MCDM technique!). The criteria were divided into 4 groups which related to: (1) the characteristics of the decision maker and/or analyst involved; (2) the characteristics of the algorithm for solution; (3) the characteristics of the problem; and (4) the nature of the solution obtained. The evaluations stem from the author's own evaluations, those of graduate students and results of other comparative studies. Not surprisingly, considering that the preferences of the author are clear in his use of composite programming for this study, compromise programming is ranked highest overall and composite (a variation of compromise programming) second. MAUT ranked tenth, and goal programming and the surrogate worth trade-off method were 14th and 15th. The study highlighted that those approaches which are easiest to understand and flexible were preferable. Stansbury et al. (1991, see next section) noted that preference functions for MAUT may be difficult to assess for some problems, while goal values for goal programming may be difficult to find for others (e.g. for ecological or social criteria), and used composite programming for the following reasons: ease of use, the double weighting scheme provided flexibility, the aggregation of criteria through several levels improves weight analysis and analysis of the system and the graphical output improves understanding. Chang et al. (1994) also used compromise programming in their watershed decision aid (next section).

The comparisons made above are limited to studies where formal methods were applied. More often than not each problem encountered in a natural resource management situation is tackled afresh and a new system applied, as unfortunately the theoretical basis of MCDM techniques appears to have been slow to filter through to this field. Smith and Theberge (1987), look at 20 different studies which had evaluated natural areas in terms of biological importance or significance. The type of scales used in each study (nominal, ordinal, interval, or ratio), the type of analysis (simple additive weighting, utility theory etc.) and method and degree of aggregation of scores are assessed. A wide range of approaches was used, often *ad hoc*, and some incorrect (e.g. adding ordinal ranks, cf. Section 5.1). Most of the studies used simple additive weighting methods (7/20), and the so-called expected value method (6/20), or a disjunctive type model (either an overall ranking is based on the highest rank each area has for any criterion, or if an alternative meets a minimum standard of at least one criterion, regardless of its level for other criteria). The simple additive weighting approaches were often misapplied due to, for example, incorrect approaches to weighting or standardisation, while 6/20 of the studies interpreted ordinal ranks as interval and summed them for overall scores. The approach called "expected value method" in this study is basically a crude simple additive weighting approach, but where the criteria are ordinally ranked in terms of importance and in two of the studies the alternatives are also ordinally ranked for each criteria. The final score is a sum of the ranks for each criteria multiplied by the criteria score or rank (the latter two studies thus also performed non-permissible numerical operations on ordinal numbers). The disjunctive approach avoided the problem of hiding areas of particular importance for a specific criterion. A site may be selected because it has high diversity, irrespective of how low it may score for other criteria. While some felt that aggregation to one index of ecological value was desirable (or necessary), others felt this to rather confuse the issue than clarify (a view supported by Turpie, 1996). Supporters of the latter view prefer separate indices or ranks for separate criteria as an alternative with exceptional value for one criterion may be hidden in the aggregation. Some felt that unless the ecological relationships between criteria were reflected by the mathematical specification of criteria and aggregation the final score would be meaningless. Either way such a composite index does require a strong theoretical basis as yet not established for conservation evaluation. A more recent application (Faith and Walker, 1996), uses the concepts of utility and trade-off space in selecting protected areas, but limit themselves to two criteria: biodiversity and cost.

It is clear from these studies and those in Section 1.3 that the simpler the approach, and the less time-consuming the input required, the more often it will be applied in everyday situations. For these reasons the exhaustive pairwise comparisons required by AHP have been unfavourably compared to SMART (Marttunen and Hämäläinen, 1995, see next section), and approaches which use linguistic judgements (Section 1.1) or require only qualitative or ordinal inputs are promoted. Cook and Kress's (1985) approach to converting ordinal scales to preference scales, is an example of the latter, although their development of this (Cook and Kress, 1991) appeared to be more complex than many quantitative approaches. Larichev and Moskovich (1995), having concluded that quantitative evaluations are more difficult than qualitative or ordinal evaluations, discussed the use of ZAPROS, a qualitative method of evaluating multi-attribute alternatives. 'Joint ordinal scales' are constructed from ordinal pairwise comparisons, so that all possible values of all criteria are rank-ordered based on the DMs preferences. Vectors of combinations of criteria (all at their worst level, except for one at its best level) are used in a conjoint scaling type of exercise which requires simply stating preference, indifference or dis-preference. A rank ordering of the vectors is obtained, and the rank order of the relevant vector is substituted into the original vectors describing the alternatives. Substituting the so-called joint ordinal scaling into the vector describing the alternative gives ranks for each criterion. The dominance of one alternative over the others is then determined from these. This provides a ranking without too much effort on the part of the DM to differentiate between alternatives. However, with five criteria, for example, and four levels of each, the number of questions is already 60 (though some may be eliminated because of transitivity), so this may limit its usefulness.

## 1.5 Applications in water resource and catchment management

Water resource and catchment management applications (specifically flood control projects) were, in a sense, the origin of the MCDM and BCA approach to decision-making and this remains one of the main areas of application. Since the completion of the previous literature review, some studies have compared different MCDM approaches (Hobbs et al. 1992, Tecle 1992), particular approaches have been formally applied (e.g. Chang et al., 1994, Stansbury et al., 1992, Abu Taleb and Mareschal, 1995, and Marttunen and Hämäläinen, 1995), and less formal applications have been made (de Graaff and Kuyvenhoven, 1996). The comparative studies were summarised in the previous section.

Chang et al. (1994) used multi-objective linear programming (MOLP) and compromise programming in a watershed-based land-use decision making problem, to examine trade-offs among objectives for several planning scenarios. The objectives were grouped under the three broad headings of economic (regional and national) development, social welfare and environmental protection. The six objectives were to: minimise phosphorus and nitrogen pollution and discharge of biological oxygen demanding load, minimise the total sediment yield, maximise employment level and maximise income. Decision variables were the amounts of forest land, agricultural land, residential area, grassland, stock farming area and recreational area. The constraints were measured in terms of, for example, the minimum amount of forest land required for conservation by law. They used compromise programming (which determines the minimum distance from the ideal solution) to examine non-inferior solutions and trade-offs after rescaling the objectives to a [0,1] interval, and compared the results to those obtained by using a multi-objective simplex method. The solutions did not vary much with changes in the parameters of the compromise program nor with the use of the multi-objective simplex method

Stansbury et al. (1991) evaluated ten alternative water transfer options with a three module decision support system consisting of (1) surface- and ground-water models, (2) an impact analysis which made use of GIS, and (3) an MCDA module. The MCDA method used was composite programming, an extension of compromise programming. The 'base indicators' (criteria), were grouped into progressively fewer, more general groups (a value tree) and weights, 'balancing factors' and worst and best values for each criteria were determined. The balancing factors "indicate the importance of the maximal deviations of the indicators and limit the ability of one indicator to substitute for another", this means that "high balancing factors give more importance to large negative impacts on any indicator rather than allowing these impacts to be obscured by the trade-off process". The base indicators were quantitatively defined from models and GIS, transformed to reflect non-linearities, normalised to the [0,1] interval and then aggregated (composited) into first, second and final level indicators. The final level consisted of a composite indicator for ecology and one for socio-economics. Nine weight sets were defined for the different groups' perspectives (ecologists, water managers etc.) and the alternatives ranked on the basis of each of these. Apart from lack of clarity about how the weights and balancing factors were determined, this approach appeared easy to use problem focused and, importantly, easy for the reader to understand. In most respects it was similar to MAVT or MAUT, with the exception of the use of balancing factors.

AHP and SMART type approaches were used in the assessment of environmental impacts of two water development projects by Marttunen and Hämäläinen (1995). Computer aided interviews were held to clarify the values of the nine different stakeholder groups. The inclusion of these approaches within EIA is recommended as little attention is usually given to the use of EIA information subsequent to its collection and compilation, and the information needs to be put in a form which reflects its importance from different stakeholder perspectives (see also Chapter 4 and Gregory et al., 1992). Six different flood control policy alternatives were considered and the value hierarchy developed represented the problem definition from all stakeholder points of view. It had three higher level attributes: economics, social and the environment. The actual values of the lower level measurable attributes where given by the relevant experts and planners and based on the data collected in the EIA process. The interviewees had the opportunity to change the measured attributes previously assigned by the planners and EIA specialists. The importance of the criteria were then determined using the swing weighting approach (cf. Appendix A.2.3). The interviewees responded numerically, graphically or verbally. In the latter case AHP-like ratio scales were used. Some of the conclusions of this study were that the formulation of the value hierarchy was the most important step in ensuring a good framework for the process, particularly as the weights of the attributes were influenced by the number of attributes within a branch. Of the two approaches AHP was found to be cumbersome and time-consuming and the motivation of the participants was affected, possibly influencing the results. The SMART approach, on the other hand, was found to be easier to understand and apply, did not require as much time, and helped the participants to clarify issues and view them from a wider perspective. However, it appears that values were assumed to be linear functions of the underlying attributes, which may detrimentally affect results. The use of computers meant that the results would be immediately seen by the interviewees and sensitivity analyses could be held simultaneously.

Abu Taleb and Mareschal (1995) used the PROMETHEE V (outranking) approach to choose among various water resource development options. The 42 options were grouped into technical, managerial, pricing, and regulatory options and the 18 objectives or criteria were ranked and weighted. The criteria in PROMETHEE are classified into six types of 'generalised' criteria whereby a preference function $P_j(a,b)$ measures the decision-makers preference for alternative $a$ over alternative $b$ on the interval [0,1] with respect to criterion $j$ such that if $P_j(a,b) = 0$ the DM is indifferent, while if $P_j(a,b) = 1$ there is strict preference of $a$ over $b$. A multicriteria preference indicating the overall preference for $a$ over $b$ is then a weighted sum of the preference for $a$ over $b$ for each criterion. The preferences for $a$ over all other options are summed, the preferences of all options over $a$ are summed and the difference between these two sums give what is termed a 'net flow' from which the overall ranks can be derived. The complexity of this approach seems unwarranted, as there appears to be no gains in discriminatory power or understanding of the problem. (See also Section 7 of Stewart, 1992.)

BCA and MCDA were used to look at the effects of land-use conversions in an Indonesian watershed development which was an attempt to increase forest and agricultural productivity while reducing sedimentation and stream-flow fluctuations (de Graaff and Kuyvenhoven, 1996). The actual developments in this integrated watershed development project were compared to three hypothetical options (a conservation oriented option, an agricultural production oriented option, and an autonomous option where there is no government interference). The evaluation criteria were grouped under the headings efficiency (maximising net benefits of food, and of export produce or import substitutes, minimising investment and recurring costs), equity (share of income to people in watershed as compared to those downstream) and sustainability (minimise soil erosion and natural forest loss, maintain hydrological regime). Three stakeholder groups (central government, local agencies, upland farmers) gave rise to three weight sets on a scale of 1 to 8, from least to most important. The BCA reflected the efficiency of the four options, but could not take into account biodiversity features of forests and equity attributes whereas using the weight sets of the different stakeholder groups to the original criteria allowed all aspects of the problem to be considered.

# Chapter 2

# Review of the Principles of Scenario Based Policy Planning

## 2.1 Overview of Scenario-Based Policy Planning Procedures

Scenario-Based Policy Planning (*SBPP*) is fully described in the Water Research Commission report of Stewart et al. (1993). It is an approach which has been developed for incorporating diverse and conflicting interests and objectives into public sector policy evaluation, in a systematic and coherent manner. SBPP was developed in the first instance for water resources planning in South Africa, but is in principle applicable to a much wider range of problems. A characteristic feature of SBPP is that it provides a uniform framework for handling and comparing both the tangible and intangible goals of society, without reducing these to monetary or similar terms.

The motivation for the SBPP approach is based on a few fundamental principles, and in particular the following:

- The values of society, especially for intangible costs and benefits, have no absolute scale of measurement. Any particular policy alternative or plan of action will be assessed from the point of view of a specific goal or societal interest by comparison with, or reference to, alternative policies or plans, the status quo, or various hypothetical ideals (such as, for example, a pre-development "pristine" condition of the environment). The SBPP approach seeks to make these reference points explicit, and to provide a simple relative scale along which policy alternatives and other reference points can be arranged, with particular attention paid to the "gaps" between them (as described in Appendix A.2).
- In any complex planning problem, it is never feasible to define policy alternatives down to the finest detail: there is insufficient time and resources to do so, and in any case a point is reached after which additional detail obscures rather than clarifies the issues. The objective of SBPP is to work in terms of relatively small numbers of alternative *policy scenarios*, defined to a sufficient level of detail, but still concise enough, to allow direct judgemental comparisons. Initially, these policy scenarios may be defined at a very coarse level of detail, but will cover a

very wide range of options. As the process proceeds, the range of options will narrow, while the detail needed to distinguish between the remaining options will increase.

The basic development of, and justification for, Scenario-Based Policy Planning, is fully described in the previous WRC report (Stewart et al., 1993) and in Stewart and Scott (1995). The key features of the approach as described in the above references are the following.

1. **Selection of a relatively small number of potential policy scenarios, specified and described to a level of detail sufficient to allow the various stakeholder and interest groups to distinguish between them in terms of their preferences**: The implication is that these scenarios are evaluated prior to the value judgement steps (see below) in terms of relatively objectively determinable "attributes", as obtained from broadly accepted models, field studies, etc. This forms part of the "description". Initially, the policy scenarios would cover a wide range of alternatives and would differ substantially, thus requiring relatively little detail to distinguish them. The process is, however, meant to be iterative, in the sense that as the process proceeds, some scenarios will be eliminated, while the remainder are refined and subdivided into more detailed alternative scenarios. As this continues, so that greater detail will become necessary.

   The construction of policy scenarios requires considerable thought and effort. Firstly, the range of options considered needs, initially at least, to be sufficiently broad so that no interest perceives itself marginalized or excluded at the outset. On the other hand, limitations to human cognitive ability restrict the number of scenarios which can be considered at a time to about 7. In the initial report (Stewart et al., 1993) we suggested a two-stage process for generating scenarios, viz. a systematic generation of combinations of key policy elements (the "*background set*" of scenarios), followed a second level of screening to obtain a small number of "*foreground scenarios*". Requisite detail on all scenarios in the background set can be assembled by running the relevant models (hydrological, economic, etc.), by expert evaluation and by public scoping. This detail will then available on line as scenarios are entered into or deleted from the (foreground) set under evaluation (see below). The initial foreground set will consist of quite widely differing scenarios, needing relatively little detail to distinguish them. As scenarios are eliminated, there will be a tighter focus around the remaining scenarios, and the requisite level of detail will increase.

2. **Within-interest group decision conferences or workshops**: Each interest group is encouraged first to assess its own preferences between the alternative policy scenarios generated in the first step. This is done within workshops, using a range of decision analysis aids, such as commercially available packages (VISA or HIVIEW) or the SDAW software (available to the water community in South Africa at a nominal charge, and described in Appendix B). The key activities in this phase are the identification of the group's own criteria and goals, the assessment and ranking of the policy scenarios in terms of each criterion in turn

on "thermometer scales" or "mid-value splitting" (as described and illustrated in Appendix A.2), and the aggregation of these scales to obtain an overall evaluation of the scenarios from the point of view of this interest. The results will typically be represented on a further thermometer scale, and subjected to extensive sensitivity analysis.

An important by-product of this process is the potential for identifying critical research needs. If the group finds itself unable to rank order the scenarios under consideration, or if there is substantial disagreement within the group as to the rank orderings, then this points to one of two things: either the group includes divergent interests which have yet to be made explicit, or the group is not agreed on the impacts or implications of one or more scenarios. The first case should be resolved by extending the hierarchy of criteria being used. The second case points to the critical research needs, in the sense of identifying those issues in which additional data or understanding is critical to the decision process itself.

3. **Between-interest group comparisons and the search for consensus:** The assessments of the scenarios, which emerge from the within-interest group decision conferences in the form of thermometer scales, provide a common scale of measurement on which the preferences of the different interests can be compared. This can be done in two different modes. A between-interest decision conference can be convened, precisely as for the within-interest decision conferences, but involving representatives of each interest and with the different interests serving as "criteria". Alternatively, techniques of *multiple criteria decision making (MCDM)* theory can be used by analysts to identify the more promising potential options for consensus, for discussion and refinement. Certain of these options are discussed further in Section 5.4.

4. **Refinement and further iterations:** As indicated above, the expectation is that a single iteration of the above procedure will not generally provide immediate consensus on contentious policy issues. More normally, one iteration will lead to the elimination of certain options as clearly undesirable, leaving perhaps two or three promising alternatives. These then need to be refined, with the generation of sub-options resolving as yet undefined implementation details, and/or the generation of new policy scenarios combining features of some of the others. The whole process may then need to be repeated, with possibly enhanced levels of detail.

Practical details of the implementation of the decision conferences and of the SDAW software are included as Appendices A and B. In the next Section, we deal with an important philosophical issue which has arisen, *viz.* the extent to which the principles of SBPP can be used *proactively* by specific interest groups, either to formulate their own position better or actively to make policy proposals, rather than purely *reactively* in response to development proposals.

In a later chapter (Chapter 5) we shall deal with a variety of developments to the basic procedures which have been undertaken either to provide a better theoretical base or to address certain practical problems of implementation. Some experiences are described in Chapter 4.

## 2.2 Proactive and Reactive Implementations of Scenario-Based Policy Planning

The original development of SBPP was motivated largely by a desire to provide decision support and aid to the process of reaching consensus on regional water planning and development policies, and the procedures as outlined above were formulated in this context. One feature of this formulation was that at the time of the evaluations by the interest groups, the policy scenarios were assumed to have been defined (although the first report did include discussion of the manner in which policy alternatives might be generated in a systematic manner). In this sense the entire procedure may be seen as entirely *reactive.*

In many instances, there is an increasing requirement that different interest groups be *proactive* in the sense of providing a statement of needs from the water resource system. Thus for example, the Kruger National Park or similar bodies may be required to state what water quantity and quality requirements are necessary for the proper functioning of the ecosystems under their management. The intention seems to be to provide a benchmark against which the search for development policies can proceed, and to avoid the perception of "moving of the goalposts" which seems often to emanate from reactive approaches. Reactions to specific policy proposals tend easily to be couched in negative terms, which may be seen to be destructive even when reflecting genuine concerns. Attempts by policy planners to address the specific issues raised may, however, lead to further negative reactions, as other goals come into play. This should hardly be surprising in the light of Simon's concept of satisficing (Simon, 1976), but leads to frustrations and accusations of moving goal posts. The basic idea of SBPP was in part aimed at providing a sufficiently rich range of opportunities so that positive preferences can be expressed, but certainly a pro-active generation of scenarios by individual interest groups can also assist in formulating positive goals up front, as opposed to focussing on the negative.

It is, of course, also in each group's own interest to be proactive. By formulating specific but realistic requirements at an early stage, they can have a direct influence on policy development, and thus ensure that their concerns receive full consideration.

Now at first sight, it may seem that there is a fundamental conflict between the need to be proactive and the apparently reactive nature of SBPP. This, however, is very far from the truth, and in fact the use of SBPP concepts can make the proactive process much more effective by focussing on total system behaviour and response, rather than on a reductionistic evaluation of individual attributes and components of the system. The problem with complex systems is that it is very difficult to establish what are feasible requirements in a multidimensional sense. Thus demands on issues such as winter low flows, flood levels, total annual runoff and various quality measures may all individually appear to be realistic, in the sense of being within practically achievable levels. Yet, in their totality, these demands may be totally unrealistic in the sense that they can never be satisfied simultaneously. In this case, the up-front and proactive specification of requirements may turn out to be a useless exercise, and may in the end have little or no effect on policy decisions. (If no feasible policy can be found to satisfy the stated requirements, the concerns behind such requirements may be largely discounted in the final decision making.)

These problems have emerged particularly in the context of conservation needs. There has been a long-standing debate between the Department of Water Affairs and Forestry on the one hand, and participants in the Kruger National Park Rivers Research Programme, for example, as to what the water needs for conservation in the Kruger National Park are. This has led also to conflicts within the Kruger National Park Rivers Research Programme, as to whether and how they should respond. Some of the issues discussed have been (a) whether or not "up-front" requirements can be specified, or whether it is only possible to evaluate specific development proposals; (b) whether the evaluation of such needs is in practice too heavily focussed on physical or abiotic, rather than on the biotic processes; and the contrast between the needs of giving advice to policy planners in the light of current understanding, and those of developing greater (scientific) understanding.

It is at this point that SBPP has an important to play, in assisting groups to formulate realistic statements of requirements. Each group can for itself generate numbers of potential policy scenarios, and use whatever models, experimental data, survey data, etc. which it may have available at that time, to elaborate on the consequences of these scenarios to the extent justifiable. With this information, the SBPP process can be applied to establish some ranking of these scenarios, and an assessment of the relative merits (on the "thermometer scale") of these. In this process, conflicting criteria which may exist will be recognized explicitly. Expert judgements will be focussed precisely on the those issues which require the interpretation of as-yet incomplete information and understanding. For example, in the case of the conservation requirements, the various environmental scientists may be able to express clearly the relative merits of the different scenarios, which may largely be defined in physical or abiotic terms, from the point of view of the particular ecosystems (biotic requirements) with which they work. The demands of different ecosystems can, in turn, be seen as "criteria" according to which the scenarios are being compared. The aggregation steps of SBPP can thus be used to find the best compromises which exist.

Although the above procedures are focussed on the evaluation of a pre-defined set of policy scenarios (as a necessary discipline for the subjective evaluations), the process is not "reactive" as choice of the scenarios for consideration, and the refinements and modifications of these are entirely under the control of the group. At the end of the day, the group can specify realistic requirements, in the sure knowledge that these are achievable for the system as a whole. The thermometer scales can in fact provide measures of the relative costs to the overall system of falling short of these requirements. Such advice will then be much more valuable and useful to policy planners than a collection of non-negotiable demands which are not simultaneously achievable.

# Chapter 3

# Multi-Criteria Decision Analysis and Economic Evaluations[1]

*"It is absurd to expect that market equilibria will automatically coincide with ecological or demographic equilibria, or with a reasonably just distribution of wealth and income"* (Daly, 1976).

## 3.1 Introduction

The last few years have seen a mushrooming of interest and activity in resource and environmental economics in South Africa. The Department of Environmental Affairs and Tourism initiated a project in 1993, the aim of which is to replace, where possible, South Africa's 'command and control' form of environmental management with an environmental resource economic approach. Various well prepared and considered discussion documents have been published, and circulated for input by interested parties (DEA&T 1993, 1994 and 1995). The Water Research Commission has funded a project 'The Application of economics to water management in South Africa' (Mirrilees et al. 1994), and workshops have been held by world leaders such as David Pearce (July, 1996). Cost Benefit, or Benefit Cost Analyses (CBA or BCA) have been applied to catchment management (Conningarth Consultants, 1994) and project assessment (Leiman and van Horen, 1996). There have also been calls from other quarters for inclusion of sound economics into conservation in South Africa, and many are involved in evaluation of the various benefits associated with natural resources (e.g. Creemers, 1996, Findlay, 1996, Turpie, 1996, Turpie and Siegfried, 1996). Some of these studies have been based on the premise that conservation should 'pay for itself', which is not universally accepted, given that other socially beneficial functions, such as education, are largely state funded and few would argue otherwise. Others call for whole-scale change to reliance on 'market-forces' for certain resources (e.g. Walmsley, 1995), with all the faith of earlier adherents to laissez-faire economics, except that the faith has now been extended to include the belief that market forces will satisfy the sustainability objective.

[1]This chapter was written in collaboration with Mr A Leiman of the School of Economics at the University of Cape Town

Given the interest in, and popularity of environmental economics in its various forms, caution is necessary as to how the techniques are applied, and consideration needs to be given to their theoretical and practical limitations, as is repeatedly advised by practising environmental economists (e.g. DEA&T, 1993, p26), and alternative or complementary approaches should be applied where appropriate. For the purposes of this report, it is also useful to contrast the approaches of multiple criteria decision analysis with economics-based procedures such as benefit-cost analysis. This is not in any sense to suggest that one approach is superior to another, but rather to emphasize the need for these methodologies to be used in conjunction with each other, so that the strengths of one can complement the other.

In South Africa, as in some other third world countries, the political events of the past few years have seen major adjustments in the organisation and objectives of government and the bureaucracy. Project appraisal and decision-making has become increasingly open and transparent, involving public input and ongoing feedback at many levels. This has meant that bureaucrat or professional contractor controlled assessments, using tools such as benefit-cost analysis (BCA) and impact assessments (IA) may be less appropriate. The local development of the process of integrated environmental management to some extent anticipated the change of direction towards increased public participation: 'scoping' allows issues of importance to be identified by the public.

Experience in South Africa has indicated a number of the pitfalls facing those who endorse public participation in Less Developed Country (LDC) decision making. The government's effort to include public input and debate in all aspects of governance has introduced a whole suite of new problems and criticisms. Important decisions are delayed as they go to committee or judicial arbitration, often after time and money have been spent on specialist studies. Attempts at 'objective' decision making through the use of BCA or IA may be defeated by subsequent lobbying by special interest groups. Public participation has failed to improve decision making when it has been either unstructured or *ad hoc* in form.

In evaluating the merits of BCA and MCDA for the appraisal of a project in an LDC, the political and economic contexts are central. The features of LDCs which are of primary significance are distributional: skewed distributions of income, wealth and power and the absence of any built-in tendency for these to be levelled over time. There are other features which also need consideration. Economic dualism is important, especially where there is a subsistence sector. Economic growth or redistribution has implications for consumer preferences when incomes are rising off a low base. Impacts may appear non-commensurable due to these economic features or to cultural differences. Also relevant is the fact that many third world economies are agriculturally based, while also being in areas subject to unpredictable extremes of climate coupled with limited institutional and monetary ability to ameliorate their effects.

The political reality of life in LDCs is that certain policies which aim at strong sustainability: the low growth approach popularised by Mishan (1969) and the minimum throughput views of Georgescu-Roegen (1971) or Daly (1976), are simply unacceptable. On the other hand, the historic failure of 'trickle down' models means that policies which simply target growth without regard for its distribution may also encounter resistance, especially when (as in South Africa) there is a strong organised labour movement. Indeed, 'sustainability' may be predicated on political and economic factors as well as ecological ones.

Clearly any decision on the allocation of resources between projects will involve value judgements. The issue is whether it is better in an LDC to have them made explicitly by the affected stakeholder groups, as is the case with MCDA; or, as in BCA, by a theoretically objective third party who has evaluated their views? Can either of these approaches deal with the complexities of mixed data, different spatial and time scales and uncertainty? Can either approach ensure that societal values are appropriately considered and that decisions are not unduly influenced by powerful lobby groups?

In order to place the bulk of this chapter in context some economic history is given in Section 3.2, and the various paradigms or schools of thought followed from neo-classical economics to 'deep ecology' are discussed. The remainder of this chapter then more critically assesses the use of the environmental economics tool of BCA as it is used to make decisions between alternative projects and compares this to one particular form of multi-criteria decision analysis (MCDA). The chapter does not offer an experimental comparison of the differences between BCA and MCDA[2], but instead appraises the theoretical and practical foundations of these techniques within the context of LDCs and finally offers some thoughts on the issue of uncertainty in decision making. Critiques of the valuation techniques used in BCA, BCA itself and its neo-classical and welfare economic roots abound; this chapter cannot synthesise all of these but concentrates on a few issues.

## 3.2 Economic Background

For the remainder of this chapter the term environmental economics is broadly used to refer to the set of theories and tools used to:

1. Decide on 'appropriate' pollution levels and prescribe economic incentives to control these.

2. Ascribe monetary costs and benefits in order to help in the making of choices between projects and development options, as well as between policy alternatives.

3. Promote the efficient use of renewable and non-renewable natural resources.

[2]The terms BCA and MCDA are fairly broad and some clarification is necessary. A distinction is sometimes made between economic and social BCA. The former being used to assess economic efficiency using shadow prices and incorporating externalities, while the latter adjusts the results of the economic BCA to account for income distributional objectives, and for the higher value placed on income used for future savings. As the extension to social BCA is rarely made in practice, however, our use of the term BCA will be taken to include both unless otherwise specified. A multi-attribute utility or value theory based MCDA approach (MAUT or MAVT) is considered here, without any implication of its superiority over other approaches. This is extended to include input from various stakeholder viewpoints (e.g. French, Stewart and Scott 1995, Gregory et al. 1992), and an analysis of a discrete set of alternatives or scenarios (e.g. Stewart and Scott 1995).

The terms resource and environmental economics are often used interchangeably, but resource economics usually refers to the third aspect above.

Mainstream environmental economics analyses are rooted in neo-classical economic theory. Although not necessarily a problem in itself, this does imply acceptance of a particular set of assumptions and values. These need to be made more explicit when making use of its tools, as the implications may be far reaching. A brief history of the development of environmental economics therefore follows which was extracted mainly from Pearce and Turner (1990) and Samuelson and Nordhaus (1985).

Classical economists of the eighteenth and early nineteenth century (Smith, Malthus, Ricardo) viewed the long-term prospects of growth as limited by the availability of resources, particularly land. Later, Mill's contention was that technical progress could counteract the natural limits to growth, but that ultimately a steady state will still be reached. Marx, together with these classical economists, believed in the labour theory of value, which meant that prices were seen as a measure of labour costs. In contrast, the neo-classicists (late nineteenth century until the 1950s), perceived price to be a function of a commodity's scarcity. Thus, analyses of supply and demand, particularly using marginal analysis (which was not previously used), became the norm. Neo-classical capitalist economics states that the most allocatively efficient outcome will result from a market economy as long as conditions of perfect competition exist and there are no externalities (i.e. all social costs are included). It cannot, however, predict the distributive outcome and this will to a large extent depend on the original distribution. Since the 1950s Keynesian economics (a response to mass unemployment during the Depression), with its emphasis on government spending has grown alongside its neo-classical roots, adding macro-economic theory to the neo-classical micro-economics.

From the time that the neo-classical paradigm replaced the classical and until the 1970s, most economists believed (and some still do) that economic growth was sustainable indefinitely through an efficiently functioning price system. Technological change and substitution would compensate for the depletion of resources. With increasing environmental concerns since the 1970s, environmental economics has emerged as an economics sub-discipline which acknowledges environmental constraints as well as the First and Second Laws of Thermodynamics to various degrees. Neo-classical economics views man as a selfish, rational, utility maximiser who seeks to satisfy substitutable wants or preferences. The economic value of all things (marketable goods, unpriced environmental goods, concern for future generations) is determined according to the amount of personal utility yielded and the preferences of individuals are revealed by the choices which they make. Analyses within environmental economics are based on consumer (individual) preferences as expressed through prices paid or prices which the consumer is hypothetically willing to pay (for a gain) or accept (for a loss). There is much debate about whether individual preferences reflect the selfish utility maximiser met earlier or a mixture of selfish (private) and selfless (public or communal) preferences. The aggregate of individual preferences in either case is then taken to reflect societal preferences. Social desirability is then determined through the Pareto criterion: the welfare of at least one individual should be improved without diminishing anyone else's welfare.

Welfare economics sees this rational economic behaviour as socially desirable and governments' role as to intervene where market failures exist and collective welfare therefore not being maximised. Thus welfare economics is normative and rests on 'clearly stated value judgements about economic organisation, income distribution, or tax policy' (Samuelson and Nordhaus, 1985), as opposed to its purely descriptive or positivist neo-classical framework. Considering that the allocative efficiency achieved through market mechanisms will not be equitable, unless equitable distribution existed originally, collective action (perhaps through governments) can bring about improvements for everyone (i.e. push outwards the utility possibility frontier). Collective action could, through wars, strikes, or inappropriate government intervention have the opposite effect.

Environmental economics is thus derived from neo-classical welfare economics. The term *environmental economics* covers a range of philosophies: from those who believe it should retain a positivist, descriptive function (with normative judgements being made only at macro-economic level) to those who believe analyses need to include normative judgements. The range also includes the 'techno-optimists' who believe that growth can continue forever through technical innovation and substitution and, although peripherally, the 'eco-pessimists' who believe that only zero-growth options are sustainable and compatible with environmental protection. Associated with the 'eco-pessimists' is the so called 'deep ecology' view that ecosystems and species have intrinsic value and rights not only determinable by their utility to humans. There are those who believe that macro-economic policy interventions to constrain economic growth are necessary because of physical and social limits. The latter view means that although environmental economics normally falls under micro-economic studies, 'green accounting' indices such as Sustainable Economic Welfare instead of GNP or NNP are being promoted within macro-economic studies. However, environmental economics has primarily been seen as a means of 'getting the prices right', under the assumption that the market will 'sort things out'. The ecological economists (incorporating the more eco-pessimist and deep-ecology views) would prefer to see the creation of a completely new paradigm not rooted in the neo-classical framework. As ecological economists increasingly use interacting ecological and economic models and/or measures of values other than prices (such as energy), these tools may yet promote a more unifying theory than that provided by neo-classical economics.

The use of the environmental or welfare economic tool of benefit-cost analyses originated in the Flood Control Act of 1936 in the United States of America although much of the welfare economics now underlying BCA was developed later (Pearce, 1983). With the increasing emphasis on the value of the environment, BCA has been extended to try to include the values of goods not normally traded on markets through the use of, for example, contingent valuation, hedonic pricing and travel cost methods (e.g. Pearce and Turner, 1990). This bulk of this chapter considers some of the criticisms of BCA as a tool without particular consideration of these valuation techniques as used within BCA.

## 3.3 Theoretical Foundations of BCA and MCDA.

Both BCA and MCDA are rooted in utilitarian precepts aimed at the maximisation of social welfare. They are different responses to a 'simple' problem: when one party benefits, but another loses, i.e. a proposed course of action is not demonstrably better than the *status quo* according to the Pareto criterion, can we test to see if society as a whole is better off? In other words, if not every member of society benefits, and there are some whose welfare is diminished, how do we make the interpersonal utility comparisons needed to identify an outward shift of a welfare frontier.

### 3.3.1 The Hicks/Kaldor compensation principle

Confronted by the weakness of the Paretian test of optimality in decision making, Hicks and Kaldor separately argued for a compensation test. They suggested that social welfare could be said to increase if the party which gains from a project could satisfactorily compensate the losers, while retaining some net benefit. The reasons for the theoretical rejection of the compensation approach are well covered in the standard literature (e.g. Graaf, 1971. pp 60-91) but become especially compelling in a third world setting.

Firstly, the compensation principle requires only that compensation could be paid, not that it actually is paid (though this was not the case in Hicks's 1939 paper). It may consequently accentuate the asymmetry already found in the distributions of income and wealth.

Secondly, the compensation principle has the potential for a pro-rich bias. If one accepts that the marginal utility attached to an extra unit of income by 'the poor' is greater than that attached to it by 'the rich', then the prices the poor are willing and able to pay for a specified change in utility will be less than those which the rich would offer. The Hicks/Kaldor approach is, therefore, biased in favour of projects that benefit the already affluent who are both more willing and more able to 'compensate' those who have lost utility.

Thirdly, the implicit bias may be accentuated if the environment is used differently by rich and poor. For example, the environment may be a source of direct subsistence inputs to the rural poor, but of recreational benefits to an urbanised affluent elite. This is especially problematic when, as in most Southern African economies, the formal unemployment rate is over 35%. The cash value of the livelihood offered by the environment is then not the issue, rather it is the lack of any viable alternative. Where projects' impacts are borne by a non-monetised public, it may be significant that choices made in MCDA are not based on either willingness to pay (WTP) or to accept (WTA), but purely on preferences and the existence of mutually agreeable compromises.

As with BCA, value/utility based MCDA uses a compensatory approach[3] in attempting to find a set of Pareto optimal, efficient, or non-dominated alternatives. However, compensation takes place in the form of trade-offs within and between criteria (perhaps within stakeholder groups) and then through consensus seeking between stakeholder groups. This will be discussed in more detail later.

[3]Others such as ELECTRE are non-compensatory (Beinat, 1996)

### 3.3.2 Income distribution and equity

BCA attempts to achieve efficiency by mimicking a perfectly competitive market. However, even perfect laissez-faire cannot guarantee a general equilibrium that maximises aggregate social welfare, merely one that achieves efficiency given the current distribution. In addition, the second best theory (Lipsey and Lancaster, 1956) implies that the allocation of resources given by a project under perfect competition need not result in the greatest net improvement in overall welfare if there is at least one sector in the real world that is not itself perfectly competitive.

Equity can be introduced as an issue into BCA with the use of income distributional weights, though these are rarely used in practice[4]. Without them, however, BCA does not automatically lead to the maximisation of social welfare and may entrench existing inequalities, and without knowledge of difference in marginal utility of income between rich and poor, distributional weights are subjective. For this reason work such as Van Praag's attempts to measure the utility of income in a cardinal sense may be a prerequisite if distributional weights are to lead to social welfare optimisation (van Praag, 1978 and van Praag et al., 1982).

There are arguments that distributional aspects, as well as sustainability, should not be included in project level analysis (Leonard and Zeckhauser, 1986) as such a piecemeal approach does not lead to an efficient solution, and distributional issues are macro goals better dealt with through taxes. Their caveat to this argument is that income weights should be considered if the public legitimacy of the appraisal is consequently improved - this is certainly the case in new participatory democracies. More importantly, adjustment of project design *after* project appraisal, in order to improve distributional aspects, is less likely to be efficient (Maass, 1966). If growth and equity are seen as competing paradigms, then the introduction of equity as an explicit decision variable may be crucial to legitimise the decision making problem. In addition, there seems to be a growing belief in the virtue of BCA as an aid in the evaluation of macro policy innovations, some of which may have environmental consequences.

In contrast to BCA, MCDA can address equity issues directly by using improvement in income or non-income equity as project selection criteria, and by allowing stakeholders to participate in the process, permitting 'the poor' to address particular issues themselves. Moreover it can do so with utility measured on an interval rather than a monetary scale, thus avoiding biases caused by differences in marginal utility of income. For example, MACBETH attempts the construction of cardinal criteria functions from absolute semantic judgements (Bana e Costa et al., 1995), and the value/utility approach described here converts the levels of achievement of certain attributes (which could be income) to an interval scale of utility. This means that the interests of stakeholders are not only reflected by a survey of their WTP for various

[4]Although standard texts on BCA mention them (e.g. Pearce 1983) practical guides such as the South African Government handbook on BCA, (CEAS 1989) explicitly avoid them. In a debate on trends in American BCA at the January 1996 meeting of the American Economic Association, Maureen Cropper indicated that a committee representing the top practitioners in the country was split over whether or not income distributional impacts should even be raised as issues in BCAs and suggested that the latest consensus was that they be omitted. This has been a consistent feature of the American approach to BCA since its birth in pre-war flood control legislation.

goods and amenities, but by their direct input in the decision-making process. Although equity and sustainability issues are difficult to operationalise in a broad sense, measurable subcriteria may indicate at least relative movement towards these goals.

### 3.3.3 Utility maximisation

The simple notion of utility maximisation is one with which philosophers and, more recently economists, have found cause for concern, leading eventually to Sen's critique of the one-dimensionality of 'welfarism' (see Graaff, 1971; Sen, 1979). The simple utilitarian approach precludes the use of alternative or multiple objectives, such as sustainability and equity. At a practical level: weak sustainability - assuming substitutability - can be handled in a utilitarian framework and in BCA by asking: 'Will the total utility decrease over time as a consequence of a project?'. Strong sustainability - rejecting substitutability for some elements - cannot be effectively dealt with in terms of utilitarianism[5]. The use of some concept of satisficing using lower bounds, like safe minimum standards, for some elements may be considered.

Although utility maximisation also underlies value/utility based MCDA, the procedure as suggested here, does not aim at the maximisation of either one particular objective or of utility from one particular point of view. To this end, the criteria selected for assessment of project alternatives, may be grouped into broader criteria or objectives which reflect the achievements of the project in terms of the separate objectives of, say, growth, equity and sustainability (e.g. van Pelt, 1993, Faucheux and Froger, 1995). In cases where overall utility is a maximum for a certain project, but where for an interest group or for an important subcriterion, utility is at a minimum, this project would be rejected through the negotiation process. This may mean that most groups do not have their best option chosen, but it should avoid situations were any group's worst option is selected (Faucheux and Froger, 1995)[6]. The equivalent concept in the pure theory of corporate decision making is the long accepted 'satisficing' view (Cyert and March 1963, and Simon, 1967). However, a particular group's 'worst' option may be chosen if the decision maker, after negotiation gives greater weight to a particular stakeholder group or criterion (say equity instead of growth), or where, for example sustainability is used as a constraint. Ideally the groups concerned should agree to this arrangement for 'the greater good'.

There is a distinctly different emphasis encountered within the fields of environmental economics and MCDA with respect to utility theory. Since the beginning of the century, utility theory as taught in economics has been confined to ordinal utility, apart from some attempts, largely ignored by mainstream economics (e.g. van Praag et al., 1982). The use of cardinal utility is considered too complex, and the assumptions too restrictive to incorporate in economic analyses (Henderson and Quandt, 1980). Ordinal rankings obtained for individual preferences are considered a sufficient basis for choice, even though one may easily imagine situations where an

[5]Bentham wished to accept utility from any sentient being as part of the utilitarian objective, but the non-negotiables of hard sustainability need not involve sentient beings at all.

[6]A parallel to this is seen in Arrow's impossibility theorem which suggests that the democratic process, i.e. the aggregation of preference orders by voting schemes, need not provide the ideal response on each and every issue.

individual, faced with a complex choice (a change in career perhaps), may want to rate his/her strengths of feelings in terms of each of the considerations, rather than simply ranking the options on the basis of each (salary change, disruption, travel opportunities, work satisfaction, work environment). On the basis of the ordinal rankings of consumer preferences, the results of a BCA should therefore be a purely ordinal ranking of alternatives. Despite this, it appears that in BCA analyses, utility as expressed in monetary terms is often taken to reflect cardinal utility, and the relationship between money and utility is taken to be linear. If it is accepted that the marginal utility of money decreases with increasing income, adjustments to this effect are not made in BCAs. In contrast, the possibility exists within MCDA to use ordinal, cardinal, interval or ratio scale utility measurement, with full recognition of non-linearities, thus increasing its discriminatory powers, and improving the 'trade-off' capabilities of the approach.

## 3.4 Practical and Operational Differences

Some degree of reductionism in any approach to decision-making is clearly necessary, whatever form it may take, due to the limited human capacity for processing information simultaneously (Miller, 1956). The questions remain: how is the reduction of dimensions to occur (what algorithm, what process, what framework), who controls the reduction (an analyst, politicians, stakeholders, lobby groups), how far is it necessary or desirable to reduce the complexity of a problem. Pearce (1983) simplified this to two questions: whose preferences should count, and how should preferences be weighted.

BCA and MCDA show substantial overlap when the steps involved are simply listed (Table 3.1), but there is the potential for differences to arise from the earliest stages. These differences are mainly concerned with the degree of stakeholder involvement and the way in which multi-dimensionality is reduced (how and by whom).

- In order to deal with multiple dimensions, impacts or criteria need to be comparable (converted to a common scale), and one needs to be able to aggregate them to some extent.
- In order to make impacts comparable, one needs to be able to compare tangible with intangible, quantitative with qualitative.
- In order to aggregate one needs to be able to weight the amount each criteria should count in the aggregation.

Choice of an appropriate technique is difficult, particularly when, as in many LDCs, there are very skewed distributions of income, wealth and power, culturally non-homogeneous populations, and sectors of the population existing outside of the mainstream economy and depending directly on their immediate environment for most subsistence inputs. How multiple dimensions are dealt with and the degree of stakeholder involvement are discussed below with reference to a number of recent South African project appraisals which had apparently conflicting economic, social and environmental components.

Table 3.1: The usual stages of a BCA and a value/utility based MCDA

| BCA | MCDA |
|---|---|
| 1. Define the set of project alternatives | 1. Identify the decision makers and stakeholder groups, alternatives, goals and objectives |
| 2. Assess the impacts of each alternative (Perform an IA) | 2. Stakeholder groups identify criteria with which they will judge the performance of the alternatives |
| 3. Order these in terms of time | 3. Identify context specific ranges of the criteria |
| 4. Monetise the impacts | 4. Determine non-linearities of the criteria |
| 5. Adjust monetary values for shadow prices, wages, exchange rates etc. | 5. Scores are given to each alternative based on each criterion separately (separate stakeholder groups) on an interval scale indicating strengths of preference for each alternative |
| 6. Weight impacts for income distribution | 6. The criteria are scaled so as to make them commensurate and so as to indicate the trade-offs between them (otherwise known as weights) |
| 7. Convert the stream of weighted costs and benefits into a single net present value for each alternative | 7. Lower level criteria are aggregated (within stakeholder groups) |
| 8. Perform a sensitivity analysis | 8. Sensitivity analyses (within and between stakeholder groups) |
| | 9. Consensus seeking -joint workshops |

### 3.4.1 Coping with multiple dimensions

BCA essentially reduces multi-dimensional problems to a single dimension (i.e. NPV of the stream of benefits and costs). Approaches which keep the multi-dimensionality more to the fore are possible within a BCA framework through extensive sensitivity analyses. The interpretation of the sensitivity analysis is left to the final decision maker, who must decide (a) which outcomes are most likely to reflect 'reality' and (b) which income weights and discount rates most accurately reflect societal preferences (or government objectives). Where the project shows conflict between objectives, the ratio of costs to benefits should not be the only consideration for the decision maker, our suggestion being that 'other considerations' need to be treated in a more consistent way. Experience in South Africa has shown that once information from BCAs or IAs, gathered at much cost, has been presented to the decision maker, it may become irrelevant to the final decision as s/he is subjected to such intensive lobbying by pressure groups. This may have been the case when the BCA and IA of the proposed St Lucia mining project were made public (§3.4.2). Intense reaction to the project from a vocal minority caused the issue to go to arbitration, and may have influenced the final decision. MCDA cannot avoid the possibility of subsequent pressure influencing decisions, but the involvement of stakeholders in the actual evaluation of projects should reduce this problem, as acceptance of the final outcome is more likely. Thus, it seems sensible that, where recognisable groups are likely to react strongly for and against various options, they participate in the decision taking process and thus legitimise it rather than enter into the debate once the specialist report is a public document.

In contrast to the BCA approach, MCDA does not limit the number and nature of objectives and criteria or dimensions chosen (van Pelt, 1993). The process of selecting and assessing criteria in MCDA is the focus of attention, context dependent and iterative. The criteria and objectives may be organised hierarchically (in a value tree) with the top level being broad-level societal objectives (such as improving the quality of life in a catchment area), the intermediate levels may identify the means to this end (perhaps in three groups: economic criteria, equity criteria, sustainability criteria) (e.g. van Pelt, 1993 and Faucheux and Froger, 1995) and the lower levels, grouped under these, being measurable attributes by which the alternatives can be judged. It is assumed that there is no *ex ante* hierarchical system of objectives, criteria and weights which should be applied in any project selection procedure, but that they will depend on the context of the decision. This context dependence is sometimes seen as a shortcoming of MCDA, but in fact the reverse may be true.

The lower level measurable attributes or criteria may be measured on a natural scale (e.g. money) or a constructed scale (e.g. biological degradation). The scales may then be converted to a value or utility function which relates the level of the criterion to the value of that criterion (as described in Section 2 and Appendix A). Economic impacts can also be scaled in this way, thus incorporating the effects of different marginal utilities of income of rich and poor. Criteria can then be aggregated to whatever level seems sensible. Criteria from different stakeholder groups could be combined into the super-criterion groups measuring overall objectives (such as sustainability and equity), or a stakeholder group could aggregate all criteria to obtain an overall ranking of alternatives for the group (Section 2 and Appendix A). The approach followed would depend on the context.

### Monetary and non-monetary scales

All impacts in a BCA need to be converted into monetary units (Step 4, Table 3.1). Where these are easily determined, BCA may be an appropriate decision tool. However, projects may have impacts on the environment and on society which, although discernible are not easily monetised, and so may be omitted, or wrongly valued (Schulze, 1994). In environmental economics the total economic value of the environment is made up of the sum of 'actual use value', 'option to use value' and 'existence value' (Pearce, 1990). Ordinarily, a consultant trying to derive monetary values, would use real markets or create proxy and surrogate markets. Besides the technical problems discussed in the literature, developing world settings may render the use of monetary scales as well as these approaches inappropriate, particularly as the environment is a direct source of subsistence, but goods are not traded in a formal market setting.

Use value is usually a measure of the recreational benefits of an environment: where people are directly dependent on the environment for factors of production, the travel cost and hedonic pricing[7] techniques are inappropriate. If, on the other hand, resources such as food were valued at market value they would be undervalued if life depends on them. Additionally, where a project will directly affect the way of life of a community, any market approach will fail. This may be illustrated by the proposed damming of Epupa falls on the Cunene river in Namibia: the Himba people will lose much of their land, and as a result change from a pastoralist society to itinerant labourers. A market approach would have to attach a pecuniary value to something unknown by the people concerned.

Although there is much debate about what constitutes 'existence value', it can be said to stem from three possible sources: (a) many people value the concept of 'wilderness', 'pristine' or 'unspoilt' areas and thus the existence of these areas, (b) all people depend on the functional integrity of ecosystems for their survival and ecosystems or species may have intrinsic value not stemming from human preferences.

The Contingent Valuation Method (CVM)[8] is the only technique which can be used to estimate existence value and option value, although such studies will tend to ascribe more value to the spectacular or charismatic than the functional (Vatn and Bromley, 1994). In addition, they have been devised mainly for assessment of recreational value, and are of little use in assessing the value of ecosystem services (Angelsen and Sumaila, 1995) and resources upon which people may be directly dependent. There is no demand curve for environmental functions so value cannot be attached to their loss or gain, and so, especially in LDCs it is likely that the mone-

[7]The travel cost technique determines consumer surplus for, say, a nature park by looking at the number of visitors to an area who travel from different distances. The hedonic pricing technique regresses distance from an amenity against housing prices to determine the value of that amenity.

[8]CVM finds the value of, say, a nature reserve threatened by a development, by surveying a representative sample of visitors and non-visitors of various income groups and asking them how much they would be willing to pay towards a fund to help to conserve the area, or alternatively how much compensation they would be willing to accept if the development were to go ahead.

tary numeraire used in BCA will obscure the inclusion of environmental services in project assessment (van Pelt, 1993). A suggested solution to the problem of finding monetary values for environmental functions, are so-called shadow projects (van Pelt, 1993). There are obvious problems attached to this concept: are they, or will they ever be implemented and if so who will pay for them. More importantly, there are not many ecologists who would be sanguine about man's ability to create functioning ecosystems, even if they did not need to exactly replicate whatever they replaced (van Pelt, 1993)[9]. Furthermore, the requirement of a monetary scale implies that 'the environment' or other factors need to be considered as tradable commodities. It is difficult for most non-economists to see 'the environment' in these terms (Vatn and Bromley, 1994) and this problem may be exacerbated in the developing world.

Other valuation methods, such as the Effect on Production, Replacement Costs, and Preventative Expenditure have apparently usefully been used in project appraisal in LDCs (Angelsen and Sumaila, 1995) to capture costs associated with environmental degradation to local populations. However, in the developing world, certain resources are likely to be undervalued simply because no market has hitherto existed. Additionally, lack of data may limit the use of these approaches. Ecological economics attempts to redress the lack of appropriate recognition given to human dependence on functioning by placing economic activity within an ecological framework, and by using energy as a common measure. Again, lack of data on an appropriate scale may be a limiting factor, particularly in LDCs.

Furthermore, there are other intangibles, besides environmental values, which may be important to those affected by decisions which may not lend themselves to monetisation. The 'disruption of traditional lifestyles' mentioned earlier, cannot purely be seen in economic terms, and attempts to construct markets for these will meet with the same resistance and arguments raised against valuing the environment. The monetary scale may also be an obstacle to addressing equity issues within BCA as discussed Section 3.3.

Some have claimed that the use of monetary units does not imply any bias towards goods sold in a market (Pearce 1983, Angelsen and Sumaila 1995). That money is commonly used as a measure of value, however, does not indicate that it is a desirable, sensible, or possible measure in all cases (Vatn and Bromley, 1994). There is an unwillingness on the part of many economists to accept that conversion to a monetary scale is problematic (e.g. Pearce 1983, 1990, Angelsen and Sumaila 1995) or that it may produce biased or incorrect results. The latter may occur even in a 'fully monetised' society, where using WTP produced rank reversal in choices as compared to using other scales, including simple rating on a scale of 1 to 10 (Vatn and Bromley, 1994). For many people a statement of preferences on an interval scale (as in MCDA) may be simpler and more intuitive than answering the question: 'What would you be willing to pay for your most preferred or least preferred alternative'. It is thus surely an advantage that monetary units are not required in MCDA.

These problems and the emotive issues associated with monetisation are avoided

[9]Although not actual shadow projects, the rehabilitation of areas affected by a project after closure are often included in project costs. This was the case with the St Lucia mining scheme, where in fact, restoration may have improved on present conditions, by replacing alien pine with the indigenous dune forest of the area.

by MCDA which can use non-monetary scales as well as both quantitative and qualitative data to obtain preferences. Nevertheless, all the diverse impacts need to be comparable in some way. Rather than using money to achieve commensurability, value/utility based MCDA uses interval scales and swing weights. This is achieved in two stages. Firstly, alternatives are rated on an interval scale (from 0 to 100, say), with gaps indicating strengths of preferences. This may be done either by direct rating on a "thermometer scale", or by examining and comparing increments in some underlying attribute value (say $z_i$) in order to construct an approximate value function, $v_i(z_i)$ for each criterion. Secondly, weights are assigned to each criterion, in a "swing-weighting" sense: in this approach, the relative worths of "swings" in values across the full ranges of possibilities for each criterion are directly compared. The resulting ratios define weights, which have the dual function of rescaling criteria to be commensurate, and of indicating the relative importance of each criterion (or in other words indicating the trade-offs between criteria).

An advantage from the perspective of LDCs is that the use of a preference rather than a monetary scale means that choices are not limited by a group's ability to pay (and no adjustment need be made). For example, using MCDA, the ideal alternative in terms of environmental conservation, is based simply on the fact that it is the ideal, not on whether people want to, or what they are able to pay (however much adjusted), for this ideal situation. Although the value functions derived may ultimately be translatable into monetary values, these are not artificially expressed *a priori*, and do not assume complete knowledge and perfect competition (Vatn and Bromley, 1994).

### Trade-offs between different dimensions

Conventionally, BCA does not attempt to make explicit trade-offs between the dimensions of a problem. Although the income distributional weights (sometimes) used in BCA theoretically promote equity, implying some trade-off between this dimension and others, the trade-off is not explicit and the procedure usually in the hands of the analyst. Moreover, as we have seen, the functioning of the compensation system does not automatically lead to the maximisation of social welfare and may entrench existing inequalities, and if a weighting system is not used, the implication is either that the current distribution of income is ideal, or that appropriate macro-level intervention will occur. The use of different discount rates as a means of changing the importance given to future effects also involves an implicit trade-off, although it is often unclear what effect changes in the discount rate have on intergenerational equity or on the promotion or rejection of environmentally damaging projects. Low discount rates do not necessarily favour future generations or sustainability, as high discount rates which penalise expensive projects may also promote intergenerational equity and long-term environmental quality. However, the use of high positive discount rates which is standard for BCA in LDCs, will skew development towards projects that offer shorter term gains. MCDA may, of course, lead to similar results if long-term issues are not included in the criteria used to assess alternatives. The point, however, is that with careful use of MCDA, the relative weighting of future outcomes is not constrained to be of the geometric discounting form.

In the MCDA approach outlined here, the trade-offs between different stakeholders and criteria are a focus of attention, through assessment of criteria ranges, non-linearities and swing weights. A compensatory approach is used, which assumes that losses in one criterion can be compensated for by gains in another criterion within any one group. The obvious difference is that gains to one group of stakeholders are not assumed to compensate for losses to another stakeholder group. Where many stakeholders are involved, this becomes more complex. Ranking of alternatives will primarily be done within stakeholder groups: the aggregation of scores across stakeholder groups requires interpersonal comparisons of preferences and values and the allocation of weights to stakeholder groups. This is precisely the problem which led to the emergence of the Hicks/Kaldor compensation approach within BCA. It is suggested that intergroup 'compensation' would occur through negotiation and compromise in MCDA, which may, as an inevitable result, mean the creation of new alternatives or variations on alternatives. This process relies on the assumption that people are willing to compromise to a certain extent and be willing to accept second best, as discussed in reference to utility maximisation. In various practical examples (e.g. Section 4.3 and Stewart and Scott, 1995), participants have been willing to compromise, but it is possible that where issues become politicised or emotional this may be difficult.

As a further stage in an MCDA, sensitivity analyses can be done; aggregating the different stakeholder scores, to see what weights would have to be attached to the different stakeholders for them to each achieve their highest ranked option. Similarly, if the criteria were grouped into super-criteria such as equity, sustainability and economic growth, a sensitivity analysis could show the weights required for these objectives to change project selection. This parallels the extensive sensitivity analyses suggested for BCA. These sensitivity analyses, together with the selection and analysis of criteria (ranges, non-linearities, weights), should give the decision maker and stakeholders an idea of the real impact of moving from a group's best option to their second best, increasing the likelihood of finding compromises. Moreover, the use of acceptable and unacceptable trade-offs rather than income distributional weights takes the decision out of the hands of the analyst and places it into those of the communities involved. The potential pro-rich bias inherent in the Hicks/Kaldor compensation approach is avoided, as all stakeholders are represented (rich and poor), the criteria chosen are those which reflect their values (in a non-monetary sense) and preferences are not governed by ability to pay.

### Data requirements

Perhaps one of the most attractive features of many MCDA approaches is that the choice of impacts included in an analysis is not bounded; both quantitative and qualitative data and both tangible and intangible factors can properly be included. In other words: MCDA is applicable in both data-rich and data-poor situations. However, it does require that intangible and qualitative data are quantified in so far as they are converted to an interval scale. The process of conversion of data, of whatever form, to an interval scale, helps to clarify and separate facts from values (Gregory et al., 1992). The inclusion of intangibles does not mean a move to non-rigorous techniques. Complex hydrological models may be required to determine the impacts of increased afforestation on downstream flows, but if exact effects on instream biota are unknown an indication can be given as to which alternatives are 'much better' or 'much worse'. Similarly, if the 'impact on traditional lifestyles' of a particular project is important to those concerned, this can be included without the need to be monetised (although a proxy attribute like 'number of people affected' may be used instead). In LDCs, where it is more likely that certain data will be scarce and expensive to obtain in the short term, decision making with MCDA is less likely to be constrained than with BCA. MCDA thus provides for a form of "capacity building", in the sense that different peoples and communities are empowered to participate meaningfully in the planning process.

### 3.4.2 Participatory decision making

It is important in new democracies that decision-making is transparent to those affected as well as participatory. In applying (or perhaps misusing) BCA, it may often happen that participation is limited to analyst surveys of people's wiilingness to pay through CVMs, travel cost or other techniques. WTP is then taken to reflect individual preferences and choices and the aggregate to reflect societal preferences. The BCA process may also be found difficult to understand by unsophisticated or disadvantaged groups, especially when, for example, discount rates and income weightings are determined by ethical criteria (in texts, by analyst, by government) (Dubourg and Pearce, 1994). The ranking of alternatives may be sensitive to these, but the implications of income and discount rate adjustments may be unclear to those affected (Munda et al. 1994a). This then can lead to a serious lack of transparency.

MCDA is inherently participatory and transparent in a number of senses. *Firstly*, the involvement of the public is extended in MCDA as representatives of groups become more intimately involved at all stages, rather than perhaps only as part of a scoping exercise which involves public input to highlight issues of concern. In MCDA, the issues are not simply identified by the public, they are also evaluated by the stakeholder group concerned who may be IA specialists, BCA analysts, community leaders, or environmentalists. *Secondly*, the involvement of stakeholders is more structured than in a simple scoping exercise, as the range of values and importance of these impacts are assessed by the groups within an established but flexible framework. Thus the stakeholders (lay or specialist) are informed by the process itself. *Thirdly*, the range of alternatives considered may be expanded by suggestions from stakeholders and not bounded by the options originally tabled (Gregory et al. 1992, Munasinghe, 1993), thus encouraging creativity in problem solving. There is a greater risk under more technical economic or impact assessments, that all interested and affected parties are not identified, or that the full set of potentially viable alternatives may not be identified or tested. The following examples illustrate some of these problems.

St Lucia is in an area of South Africa where unemployment and poverty are endemic. It is one of the two most important coastal wetlands in South Africa for waterbirds (Turpie, 1995), an important nursery area for marine stocks (S. Lamberth, Zoology Dept, University of Cape Town, pers. comm.), a Ramsar site and contains some of the little remaining indigenous dune forest in South Africa, although the area has been heavily afforested with pine and gum. A BCA was conducted when a proposal was put forward to mine the deposits of titanium and other minerals along the eastern dunes. The environmental IA recommended that the area should be declared a National Park and should acquire World Heritage status. The area was subsequently placed on the Montreux record due to the threat of mining. In contrast, the BCA indicated that mining should go ahead. Sociological considerations were addressed in the IA (Miltz et al. 1992), but the process did not include any consultation with the community which had been removed from the area by the previous government for the purposes of expanding forestry in the area. The preferences of this community, who had some claim to the area, were thus not considered. When the issue went to an arbitration board, it was decided that no conclusion could be reached (either in favour of mining or in favour of the area remaining under the control of Natal Parks Board for tourism purposes) until consultation had occurred. Perhaps in response to lobbying by other groups, the new government subsequently decided against mining, although still without consulting the community. If formally involved, the needs of the community would have had to be weighed up against the needs of others (tourism, mining and the environment). This would have made the problem more complex, but this could have occurred within a rational framework.

Another example was seen in the IA performed for the Saldanha Steel Project (CSIR, 1995). The site selected for a rolled steel mill by the company concerned was on the shores of Saldanha Bay which leads to Langebaan Lagoon. The lagoon is an important area for waterbirds including large numbers of migratory waders, a Ramsar site, supports a small but fairly lucrative aquaculture industry and attracts some, mainly local, tourists. The area has a high unemployment rate and was previously designated a growth node. Although the issue of alternative sites was raised by interested and affected parties during the scoping stage of the IA, the problem was never satisfactorily addressed as the terms of reference of the consultants hired to do the IA excluded the assessment of other sites. Ironically, after arbitration by a judicial commission, it was an alternative site (2 km inland) which was finally selected, although it had never been evaluated in the IA process. An unfortunate consequence of the fact that the performance of IAs has become an industry in itself, is that the process has become bureaucratic, and terms of reference accepted by consultants have become inflexible.

Of course, there are problems in responding to the imperative of participatory or democratic process. If there are too many stakeholder groups the process may become cumbersome, inefficient or costly. Where stakeholder groups are easily defined and coalesce into a manageable number, the process may still flounder if groups deliberately misrepresent their beliefs in order to enforce a particular solution. In practice such strategic behaviour appears not to be the norm even when interests are widely divergent (Stewart, 1994), and MCDA has proved useful in consensus building in a number of cases (e.g. Stewart and Scott, 1995).

## 3.5 Uncertainty

Perhaps one of the most important issues with regards to decision making in project selection is uncertainty. Uncertainty as to values (prices) scores, weights and ranges elicited can be dealt with by using adequate sensitivity analyses. More fundamental is uncertainty about long term environmental consequences, future political, social and economic conditions, and the dynamic relationships between these, as well as their impacts on preferences. Since decisions may be environmentally/socially detrimental and irreversible, decision makers may wish to minimise the potential for loss of future opportunities (Vatn and Bromley, 1994) or to 'avoid regret'. The following section deals with approaches to uncertainty about the consequences of decisions, about future states and changing preferences over time in LDCs.

### 3.5.1 Changing preferences

With the expectation of rapidly changing income distribution there is an implication of changing preferences over time even in the short term. We do not know how our tastes will change with a changing world, let alone how future generations tastes will change. BCA is unable to satisfactorily handle the dynamics of changing preferences, changes in the nature of the world, the perception of problems and adjusted expectations. This problem arises because people's tastes may change with their incomes, so that even if compensation does occur, people may feel after the event that they preferred their original pre-project situation (Scitovsky, 1941). This becomes particularly problematic when attempts are made to value the environmental consequences of a project. Valuation of impacts obtained *ex ante* may differ from that which the same respondents would provide given the distribution of income after the project (i.e. em ex post). This problem is in a sense, philosophical and is certainly not avoided by MCDA, nor any other approach.

### 3.5.2 External uncertainties

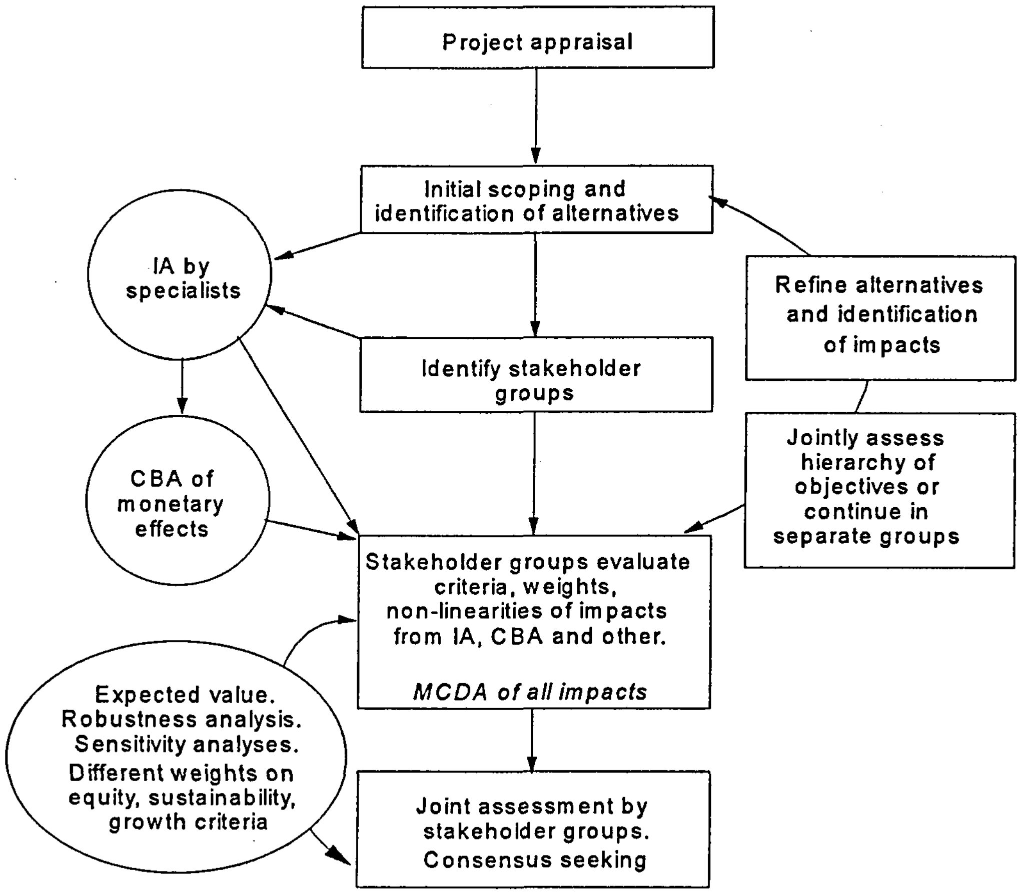
Uncertainty with regard to the value of environmental amenities or the consequences of decisions, is considered in environmental economics through quasi-option values which represent the value of preserving future use or existence given some expectation of an increase in knowledge about the environment in question (Pearce and Turner, 1990). Given uncertainty with regards to future preferences quasi-option value may also be seen as the value of forgone natural assets in the event of a shift in public tastes, or the cost society is willing to incur to acquire the scientific information needed for more accurate assessment of a project's net benefits. Quasi-option values therefore are related to the rate of change of income expected in the near future: this itself is a source of uncertainty in LDCs, as is the distribution of that income. If the benefits of growth are not expected to accrue to the poor but to the already affluent, the resulting quasi-option values will be reduced. Quantifying quasi-option values is clearly difficult given the uncertainties of the economic future of many LDCs.

The standard approach to uncertainty in MCDA would be to use expected utility analysis. This has limitations as one is expected to either know the probabilities of future outcomes or to act as if one knew them (Faucheux and Froger, 1995). In many of the situations with which we are concerned, probabilities of future states are unknowable. In the first world, sectoral interrelationships are easily shown using input-output matrices. Tracing the effect of a project through the economy is certainly possible, although it has been suggested that, due to effects of scale, first world countries may have reached a state of 'endemic unpredictability' (Rosenhead, 1989). In third world economies where much activity is outside the formal economy and involves close and reciprocal links with the environment, assessing the impacts of a project is less straightforward. Such difficulties are exacerbated by high levels of intra- and interannual climatic variation as seen in much of Africa. Decisions may be made on the basis of present or mean conditions without due regard for effects of high variance. However, combining several different approaches to the exploration of options may provide a reasonable basis for making decisions which 'avoid regret'. Three possible complementary approaches are suggested (Figure 3.1), all of which can be incorporated into our original MCDA framework (Table 3.1).

Although limited in cases of strong uncertainty, exploring preferences using expected utility functions will provide invaluable insights. Research has suggested that a few 'scenarios' can be used to represented uncertainty rather than requiring a full probability distribution or an exhaustive set of all outcomes (Stewart, 1995). This immediately makes the problem more manageable. Decision trees or influence diagrams (Gregory et al. 1992) may be useful in the construction of links between present decisions and possible future events, and in the determination of expected values through probabilities. Although probabilities may be difficult to estimate, sensitivity analyses can reasonably easily be implemented to assess the range of possible values, while outcomes under each scenario can be treated as decision criteria in their own right. Fuzzy approaches (e.g. Munda et al. 1994b) may also be used, but the benefits of this approach appear to be outweighed by its complexity. Any MCDA technique used in workshop settings, where people of many different educational backgrounds may be participating, needs to be relatively easy to understand as well as to use. The fuzzy sets approach risks becoming as problematic as BCA, where the processes and adjustments and assumptions are well hidden from the user.

A useful extension or complement to this standard approach would be to use some of the 'soft' operations research techniques such as strategic choice and 'robustness analysis' (see Chapter 6, or Rosenhead, 1989) or the ideas of adaptive management suggested by Walters (1986). In robustness analysis, a set of decision 'packages' are identified and the pathways resulting from these and sequential decisions explored. Decision packages which produce acceptable or desirable results in a number of different scenarios, or which are less likely to result in undesirable outcomes, are preferred to others; thereby increasing intertemporal flexibility of decisions. In the end the decision taken will be a 'gamble' whatever the process used (Walters, 1986), but with some exploration of the alternatives, the actors should be more comfortable that it is a reasonable gamble. Stakeholders may suggest that rather than taking a gamble with uncertain consequences, we should 'wait and see' or 'do more research' (Walters, 1986); this is equivalent to the quasi-option approach. However, in many cases decisions need to be made as a matter of urgency for other reasons. In LDCs this may be the need to 'deliver' expressed earlier, combined with a real need for income growth amongst the poor. It must also be recognised that, in some cases, more research will not necessarily provide the answers, as ecological realities remain complex and consequences may be unknowable in any near term future. Postponing decisions until research projects find 'absolute truth' is not usually a viable option. The view of the scientist as the disinterested specialist who should not become involved in public

Figure 3.1: Examples of combining IA, BCA, MCDA and approaches to uncertainty in a decision making framework



debate except to offer objective and well researched facts is admirable. However, the extension of this to suggest that until facts are known no opinion should be given, or to censure any action taken without them, is counter-productive. It must be accepted that uncertainty will always be with us, and that there is no perfect solution which will remove all risk from decision-making. Making no decisions at all will certainly not improve the status of the poor nor that of the environment.

The third approach is linked to earlier suggestions where broad, unmeasurable or non-operational goals (such as growth, efficiency, equity and sustainability), are used to group intermediate and lower level attributes or criteria which are measurable, and which in some way contribute to the overall objective (Faucheux and Froger, 1995 and van Pelt, 1993). BCA may be reasonably used for the analysis of aspects of growth, efficiency, equity and sustainability, but, as discussed, runs the risk of bias if used alone, particularly in LDCs. Using MCDA, we may not be able to operationalise 'sustainable' as an attribute, but we can estimate (a) whether one alternative is 'more sustainable' than another, and we can know that (b) critical levels of certain attributes of an ecosystem (ecological sustainability) or an economy will probably cause irreversible harm (such as species extinction). Safe minimum standards may be used or the precautionary principle invoked (Faucheux and Froger, 1995 and Vatn and Bromley, 1994) keeping in mind that it is the value of deviations from the standards which are of concern rather than the physical amount (Dubourg and Pearce, 1994). In practical terms, this approach has been used in South Africa to determine the timing and duration of critical low flows and high flows for rivers with modified flow regimes (King and Tharme, 1994). Whatever data is available from the hydrological record is combined with 'groundtruthing' where critical reaches are identified and expert opinion used to determine the impacts of reduction below critical low flows or loss of floods, the process taking into account the inherently unpredictable and highly variable flow regime of most South African rivers. These ideas have been incorporated into an MCDA assessment of development plans for the Sabie River catchment (Stewart and Scott, 1995), the process continuing with an assessment of other users' needs. The river catchment is characterised by very high commercial afforestation in its upper reaches, high abstraction for irrigation of commercial farms as well as the use of instream flows for the domestic use of a fairly dense rural subsistence community. The river flows, eventually, through the Kruger National Park to Mozambique. The example thus encompasses many of the problems of decision making in LDCs which we have mentioned. There are multiple users with very different social and economic status, the river itself is acknowledged as a user, the river flowing through the Kruger Park must supply water to game and vegetation while being part of the expected 'unspoilt' environment of tourists and then must supply water to another country, with even poorer rural communities.

## 3.6 Concluding Remarks

Many have agreed that it is difficult to find monetary values for certain project impacts (Pearce 1983, Angelsen and Sumaila 1995, Munasinghe 1993, Vatn and Bromley 1994, van Pelt 1993, Munda et al. 1994a, Gregory et al. 1992, etc.) and it may seem unnecessary to belabour the point. Some have suggested that these impacts may well be considered using MCDA (Gregory et al. 1992, van Pelt 1993, Munasinghe 1993, Munda et al. 1994b, etc.). This chapter has supported the latter view, and extended it to suggest that some monetary impacts may be better handled in an MCDA framework. Additional arguments make the use of MCDA more attractive.

The theoretical and practical consideration of equity and sustainability within BCA has proved problematic. While not capable of solving all the problems associated with operationalising these concepts, MCDA approaches offer some advantages (Gregory et al. 1992, van Pelt 1993, Munasinghe 1993, Munda et al. 1994b, Faucheux and Froger, 1995, etc.), perhaps through the use of a hierarchical system of objectives, goals and criteria. Only the lower level criteria need to measurable, whether quantitatively or qualitatively. These criteria (e.g. the number of corridors between plantations) contribute in some way to the achievement of the higher level objectives (e.g. sustainability). While BCA may be appropriately used to determine certain growth and equity considerations, intangible and qualitative data can more easily be included in MCDA and effects on particular groups more explicitly assessed, while the flexibility of MCDA also allows the use of a variety of approaches to uncertainly. Thus a narrow and exclusive approach to project appraisal which necessarily limits creativity without ensuring rigour is avoided.

While the above comments apply to any project appraisal, they are more persuasive in LDCs, where impacts are often felt by those to whom any small change in income or quality of life is highly significant, due either to a low income base or to direct dependence on the environment.

# Chapter 4

# Case Studies

In this chapter we summarize the results of the application of the procedures in three practical cases in which we were involved, and seek to extract general principles for future implementation from the experiences reported. These case studies provide the motivation for the more technical development of procedures as discussed in Chapters 5 and 6.

## 4.1 Kruger National Park Rivers Research Programme

In August 1993, the Decision Support Systems sub-program of the Kruger National Park Rivers Research Group initiated a project entitled, "A prototype decision support system for the management of the natural environment of river systems". The aim of this project was "to develop a structured process for providing, in an efficient and cost-effective manner, the information required to improve the quality and usefulness of responses to enquiries from researchers, resource managers and stake-holders, with regard to the management of the natural environment of river systems."

This provided an ideal opportunity for development of procedures and subsequent testing in workshops of Scenario Based Policy Planning, and we undertook to work on the DSS as an integral part of the present project.

It was envisaged that the devised DSS could be used to both respond to management proposals and to proactively investigate "what if" questions. The DSS itself is fully documented in a separate report (MacKay, 1994) and only the sections relevant to policy evaluation will be dealt with here. A conceptual DSS was developed during workshops comprising members of the river research community from a variety of different technical backgrounds. Two workshops were held in Pretoria during August and September 1993. The time between workshops was used to develop and produce information to support the processes within the workshops.

*First workshop*

During the first workshop (August 1993) opportunity was provided for collaborating researchers to present their approaches to tackling different problem areas and to inform the group of how and when the methods could be used effectively. The workshop then debated the various methods (and dearth of them in some problem areas) and devised a DSS protocol which was basically a coherent and consistent approach to tackling research questions.

The various steps of this process were identified as follows :

(1) Information transfer from managers (project initiation)

(2) Scoping and initial assessment of proposals that come out of (1)

(3) Prediction of changes that proposals will bring about

(4) Evaluation of change and its acceptability to researchers and stakeholders.

*Second workshop*

At the second workshop (September 1993), five hypothetical scenarios were chosen for evaluation. Three of these were based on the supposition that the Madras Dam was to be built on the Sabie River near the Western boundary of the Kruger National Park, and differed according to the release policies being contemplated, viz :

(i) Guaranteed minimum dry-season flow rate of 1 cubic metre/sec in Sabie River (Scenario 1).

(ii) Abstraction such that flow in Sabie River is always 30% of naturalised flow, as determined by patterns of flow measured in the relatively undisturbed Mac-Mac River (Scenario 2).

(iii) Abstraction such that present day flows are maintained (Scenario 3).

In addition, two further scenarios were included for comparison, viz.:

(iv) Maximum possible abstraction for irrigation, or "worst case" scenario (Scenario 4).

(v) Abstraction such that naturalised (pre-development) flows are maintained, or "ideal case" (Scenario 5).

The above scenarios were evaluated by the group using the SDAW interactive software (see Appendices A and B) to facilitate the sessions. The procedure required participants to:

- Conceptualise a hierarchy of goals/criteria for the system under consideration (Sabie River);
- Evaluate the consequences of each proposed policy in terms of each (sub) goal/criterion; and
- Assess the importance (weight) of each (sub) goal/criterion.

The "Evaluation Phase" of the proposed DSS was initially perceived as the final stage of the process. However, when the time came for workshop participants to evaluate the proposed water policies (with the support of the SDAW methodology and software) the participants were greatly surprised to find that they had difficulty in defining the criteria whereby they would make their assessments. The 'Key Issues' that had been discussed at length in the 'Scoping Phase' of the project had not materialised as independent criteria whereby objective comparisons could be made. Nor had these 'decision criteria' emerged during the 'Predictive Phase' of the project, despite extensive discussions around the hydrology, hydraulics, geomorphology, water chemistry, water quality and ecosystem response of the system under consideration.

What emerged was the fact that the issues that had been crucial to the discussions that had preceded the formal 'Evaluation Phase' were in fact a tangle of sub-criteria and overlapping criteria which had never been fully separated or defined. Consequently, conflicting views amongst the group of participants prevented a convergence of the process of evaluation. There was a realisation amongst participants that the 'Evaluation' step is not merely a tool which is applied at the end of the DSS which provides a quantitative assessment of the proposals. The 'Evaluation' is a process which begins somewhere early on in the course of events with an articulation of the criteria whereby policies will be judged. It is at this stage of the process that the real conflict within a group of participants is made explicit by specifying those criteria around which the divergence of opinion can be expressed or resolved.

It is the acceptance of this process as part and parcel of the whole DSS (not a final appendage) that is fundamental to the success of the entire operation. The participants at the workshops agreed that the difficulties they experienced were evidence of a need for a structured decision support environment.

Participants also expressed the need to take the process a step further and judge the proposals not only in a comparative sense but also in terms of overall acceptability. This may require further development of the methodology/procedures or it may be able to be accommodated if the full iterative SBPP procedure is applied whereby there is not only a decision made as to the acceptability of a particular proposal but there is also a search for a policy more likely to be acceptable to the group in question.

In the end, participants assessed the scenarios in terms of the four criteria indicated in Figure 4.1, where the heights of the bars indicate the importance weights attached to each criterion. After aggregation of the resulting assessments using the weights in Figure 4.1, the resulting overall group evaluation of the scenarios is as represented by the thermometer scale displayed in Figure 4.2. This group evaluation process produced a powerful message in terms of relative levels of environmental disturbance. As can be seen from Figure 4.2, the use of the interval scale in the SBPP methodology led to the conclusion that the best management plan on offer (Scenario 2) would cause the same degree of environmental upheaval relative to present day conditions (Scenario 3) as the destruction involved in going from pristine (pre-development) conditions (Scenario 5) to the present day level of development. The ability to make this sort of statement in effect provides environmentalists with a "currency" within which to convey their attitudes/ opinions about proposed strategies.

The evaluation of the above policy proposals was repeated at the 3rd Annual Workshop of the KNP Rivers Research group (September 1993) as a means of re-

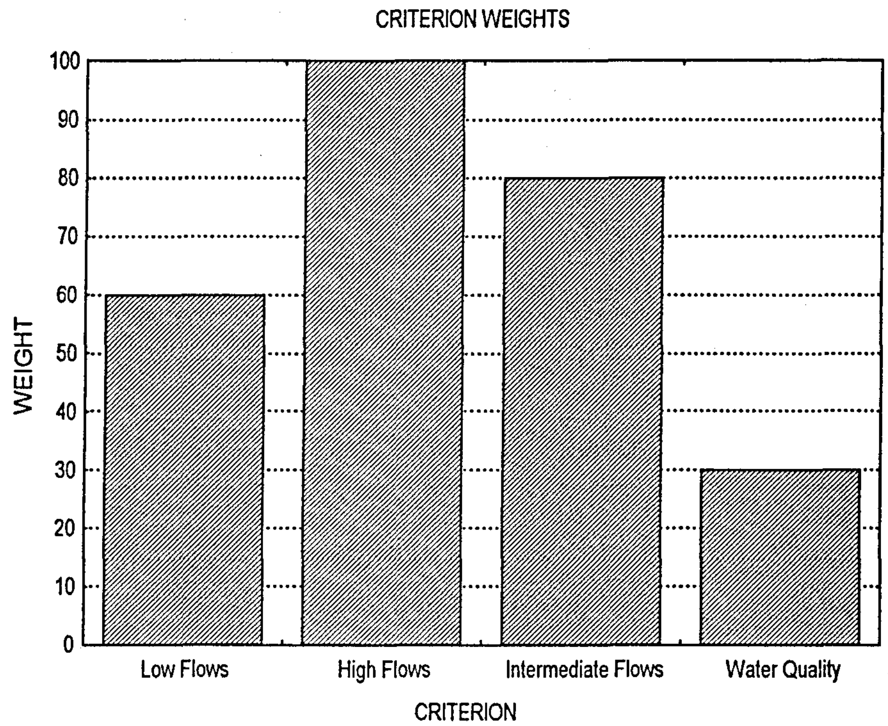


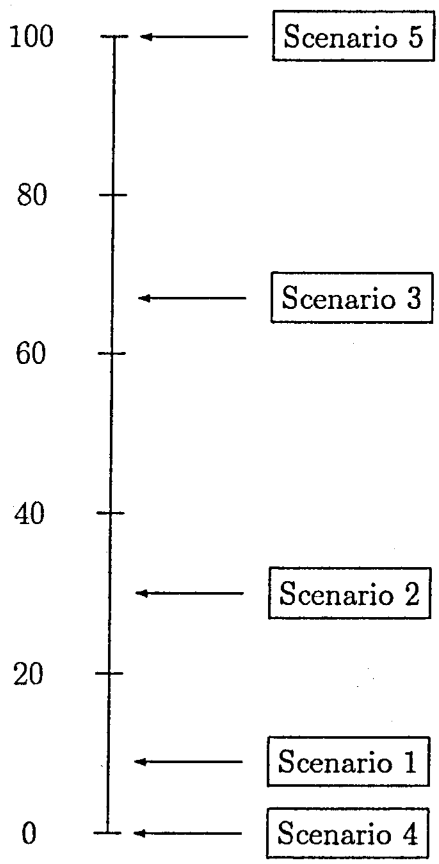

Figure 4.1: Weights attached to criteria for Kruger National Park rivers

porting back to the Research Group the progress with the DSS to date. Feedback from participants at all of these workshops was used to develop research proposals to further the decision support process. Thus, motivation was put forward for further research into the evaluation of proposed water management policies, specifically looking at supporting the following processes:

- the selection of and agreement on criteria or sets of criteria by which change is evaluated;
- the assignment of weights to criteria by individuals or groups of individuals; and
- the achievement of consensus amongst a group of individuals as to the acceptability of change.

Subsequent to these proposals being accepted at the 3rd Annual Workshop, the Decision Support Program of the KNPRR group underwent a change of management. This appeared to result in a shift in emphasis from more holistic decision tools (such as the SBPP approach) to more technically precise and detailed models of the ecosystem as aids to decision making. Consequently the above proposals as to how the Decision Support as to how the process could be taken further have yet to be implemented.

Figure 4.2: Overall thermometer scales for Kruger Rivers scenarios



## 4.2 Physical Importance Rating of Estuaries

### 4.2.1 Background

A number of different techniques were used in combination in order to derive a system for rating the physical importance of estuaries. This formed part of a larger study, run by the Consortium for Estuarine Research and Management (CERM) to determine the relative importance or value of South African estuaries in terms of both environmental and social value. The importance rating decision support system had previously been organised into three main value categories; rarity value, biological value and physical value. These values were to be determined by various indices and scores, which had been organised by CERM during several workshops with relevant experts (see CERM, 1995). Apparent duplication of values, and inconsistencies in the scoring systems had been identified by CERM and a workshop was held (10 August, 1995) in order to identify any further problems and to reorganise the physical importance rating section into a workable framework.

It was clear from preliminary discussions that, although some of the criteria which had been identified previously were relevant, substantial clarification of certain concepts was necessary, and this implied reworking the framework as well as the individual criteria. The approach adopted was to apply concepts of scenario based policy planning to the construction of indices, as is discussed in more detail in Section 5.3.

It was emphasised at various stages of the workshop that the physical component index should only measure issues that are substantially missing from consideration in the other indices, or that need to be re-considered in a different context for particular reasons. It was apparent that some criteria or indices were measuring conservation 'importance' while others were measuring conservation 'status'. A similar debate has been ongoing in rivers research and other disciplines, and it is generally considered important to keep these issues separate (e.g. O'Keeffe 1995). After some discussion, it emerged that the concern for the physical component covered two more-or-less distinct issues, viz.:

1 **The 'uniqueness' of the physical nature of the estuary**, which might not necessarily be reflected in other rarity measures (such as specific flora and fauna): It was recommended, however, that these issues preferably be included in the rarity index, by taking into account the physical size of the estuary, and how many other estuaries occur in the same combination of climatic zone and type of estuary. It was also clear that if all South African estuaries were classified into types and climatic zones (as suggested in CERM, 1995), a rating system would be worthwhile for only two groups of estuaries as the other groups consisted of only a few estuaries each. The two groups thus considered in the physical importance rating are the *naturally permanently open* group and the *naturally temporarily closed* group.

2 **The extent to which the estuary currently represents natural estuarine conditions**: This index would of necessity measure degradation or deterioration of the physical system from that which would be expected to occur (therefore in essence measuring conservation status), and the rate at which such

degradation was continuing. The work described here focused on the construction of an index relating to this issue.

### 4.2.2 Criteria

Four criteria of evaluation were initially identified as relatively independent components of the physical index. These were:

- Degree of siltation
- Extent to which the estuary is open to the sea relative to its natural condition
- Water quality
- Hydrodynamics

Subsequently, and in the light of attempts made to construct and to score scenarios, the last of the above was subdivided into *salinity* and *freshwater and seawater flushing*, where in both cases the issue was the extent of change from naturally occurring conditions.

This provided five criteria, and the workshop defined numbers of different scenarios for each of these, along the lines developed in Section 5.3. Details of the scenarios are discussed later in §4.2.4. Scores associated with each of these scenarios, representing levels of achievement for each criterion, were assessed on a 0-100 "thermometer" scale, for inclusion in the indexing system.

### 4.2.3 Questionnaire for weight assessment and conjoint scaling

Time available at the workshop did not permit the assessment of weights on the criteria, or the use of conjoint scaling for consistency checks (see Section 5.3, and the participants agreed to provide the required assessment for this phase by means of a questionnaire, to be completed after the workshop.

The first part of the questionnaire was aimed at establishing importance weights, using the "swing weighting" concept. The relative weights were indirectly determined in the following way (with reference to Table 4.1):

1 Six hypothetical scenarios were defined. The first of these (Scenario A) had all criteria at a medium or slightly worse than medium level. The remaining five scenarios each had 4 of the 5 criteria as in Scenario A, but one criterion set at its best level. The scenarios were ranked by the participants in order of importance from the physical importance rating point of view (Table 4.1: Rank order).

2 A percentage score of 100 was given to the scenario that ranked first, while Scenario A was clearly the worst and given a percentage score of 0. Percentage scores for the remaining scenarios were allocated by respondents, to indicate the relative degree to which each scenario was better than A (Table 4.1: Relative weight).

| Scenario | Criterion set to best level | Rank order | Score of scenario | Value range | Normalized weight |
|---|---|---|---|---|---|
| A | | 6 | 0 | | |
| B | Siltation | 2 | 90 | 75 | 0.259 |
| C | Openness to sea | 1 | 100 | 65 | 0.333 |
| D | Water quality | 3.5 | 50 | 60 | 0.180 |
| E | Salinity | 3.5 | 50 | 66.67 | 0.162 |
| F | Flushing | 5 | 30 | 100 | 0.065 |

Table 4.1: Results from the swing weighting questionnaire

| | | |
|---|---|---|
| Siltation | vs. | Open to sea |
| Open to sea | vs. | No. of undesirable salinity aspects |
| No. of satisfactory water quality measures | vs. | No. of undesirable salinity aspects |
| No. of undesirable salinity aspects | vs. | Siltation |

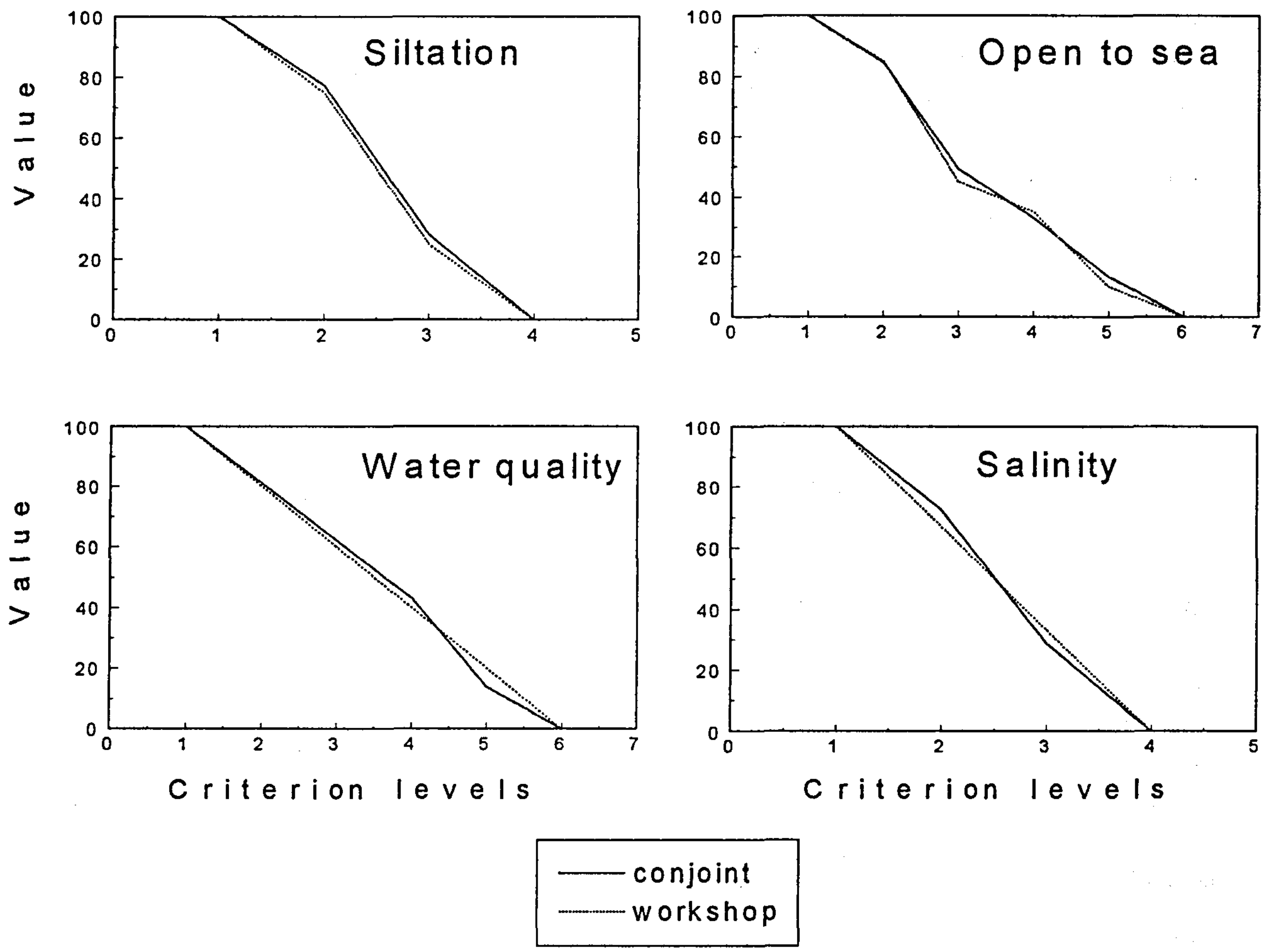
Table 4.2: Pairs of criteria tested in the conjoint scaling exercise.

3 The range of thermometer scale values (as determined at the preceding workshop) between the levels in Scenario A and the best levels varied between criteria. In order to determine relative weights, the percentage score for each of the scenarios B–F had to be divided by the range of values for the associated criterion. For example, the weight for the first criterion (siltation) was obtained by dividing the score of Scenario B (90) by the range for the criterion (75) to obtain the value 1.2. The resulting weights for the five criteria (1.2, 1.54, 0.83, 0.75 and 0.3) were divided by their sum (4.62) to get the normalized weights shown in the last column of Table 4.1.

The second part of the questionnaire was a conjoint scaling exercise (Section 5.3), in which pairs of performance levels for two criteria at a time were directly ranked. Each table in the questionnaire related to a specific pair of criteria, as indicated in Table 4.2. Respondents were required to rank order as far as possible, pairs of performance levels for the two criteria, arranged respectively as the rows and columns of a table. As an illustration of this, the results for the comparison of performance levels for siltation and openness to sea are recorded in Table 4.3. The meanings of the level codes are explained in §4.2.4. What this indicates, however, is that respondents would prefer the combination of second best level on siltation and best level on openness to sea, to *vice versa.*

As described in Section 5.3 and in Appendix C.1, the ranks of pairs of levels for criteria as in Table 4.3 can be used as a consistency check on the thermometer score scales and on the weightings, and as a means to adjust these scores and weights to achieve greater consistency. In this case the degree of consistency was extremely good, as indicated in Figure 4.3, which displays the scoring functions derived from the workshop and conjoint scaling respectively, for four of the criteria. (The last criterion is binary, and does not lend itself to display in this manner.) Overall, therefore, this provided an excellent confirmation of the validity of the scores.

Figure 4.3: Comparison of scoring functions obtained from the workshop and conjoint scaling exercises respectively.



| Siltation | Open to sea | | | | | |
|---|---|---|---|---|---|---|
| Levels | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 | 2.6 |
| 1.1 | *1* | *3* | *5* | *7* | | |
| 1.2 | *2* | *4* | *6* | | *17* | |
| 1.3 | | | | *16* | *20* | *21* |
| 1.4 | *15* | *18* | *19* | *22* | *23* | *24* |

Table 4.3: Ranks (in italics) of performance levels for the two criteria 'Siltation'(criterion 1) and 'Open to sea' (criterion 2). See §4.2.4 for explanation of level codes

### 4.2.4 Final scoring system

The total score for an estuary was thus constructed as the weighted sum of the scores for each criteria (which does assume additive independence between the criteria). In view of the high degree of consistency between the workshop and conjoint scaling scores, it was decided that the conjoint scaling scores be used, where feasible, as these incorporate both sets of information. The resultant weights, expressed in percentage form were as follows:

| Criterion | Percentage weight |
|---|---|
| Siltation | 26 |
| Openness to sea | 33 |
| Water quality | 19 |
| Salinity | 15 |
| Fresh- and seawater flushing | 7 |

In principle, therefore, the score allocated to an estuary is given by $[26v_1 + 33v_2 + 19v_3 + 15v_4 + 7v_5]/100$, where $v_1, \ldots, v_5$ are the scores associated with this estuary on the five criteria. In practice, it is more convenient to combine the weights and the scores, into a single system, i.e. scores $u_i$, say, defined by $w_i v_i/100$, where $w_i$ are the percentage scores as above. Thus, for example, an estuary which gets the best score ($v_1 = 100$) on siltation, will gain a score of 26 for this criterion (to be added to the scores for all other criteria). In the following paragraphs and tables, we summarize the levels (scenarios) defined, and associated weighted scores to be used for each criterion.

#### Criterion 1: Degree of Siltation

Levels (scenarios) and associated scores were as follows:

| Level | Scenario | Score |
|---|---|---|
| 1.1 | Little or no erosion evident in the catchment area | 26 |
| 1.2 | Detectable erosion, should not lead to serious filling of the estuary within the foreseeable future | 20 |
| 1.3 | Serious levels of erosion, good chance (50/50 odds) that the estuary may be filled with sediment, or substantially reduced in size, within 50 years | 7 |
| 1.4 | Extremely high levels of erosion, highly likely that the estuary will be filled with sediment in 50 years | 0 |

### Criterion 2: Openness to sea

In this case a matrix of scenarios was created, based on a comparison of assessed natural and current conditions. Five nominal states were identified, corresponding to proportions of the year that the estuary should be, or currently is open, namely, 0%, 25%, 50%, 75% and 100%. This gives, principle, a total of 25 scenarios to be scored. The conjoint scaling and swing-weighting exercises were however based on a selection of six of these (labelled 2.1,...,2.6), as indicated in the table below. The scores for all combinations (based on conjoint scaling where available, and on the original workshop scores otherwise) were as follows:

| Natural state | Current state | | | | |
|---|---|---|---|---|---|
| | 100% | 75% | 50% | 25% | 0% |
| 100% | 33 *(2.1)* | 11 *(2.4)* | 4 | 2 | 0 |
| 75% | 27 *(2.2)* | 33 | 16 *(2.3)* | 4 *(2.5)* | 0 |
| 50% | 23 | 27 | 33 | 13 | 0 *(2.6)* |
| 25% | 0 | 23 | 27 | 33 | 0 |
| 0% | 0 | 0 | 23 | 27 | 33 |

### Criterion 3: Water quality

Scoring for this criterion was based on the number of positive responses to the following questions:

1 Do suspended solids exceed 10% of ambient?

2 Are organic toxins present?

3 Is the DO in surface water within the range 5 to 6 mg.$l^{-1}$ ?

4 Is the estuary eutrophic (is there excessive algal growth)?

5 Are faecal coliforms within limits?

It was originally suggested at the workshop that there was no reason to consider any of these conditions more or less important than any of the others, and so the score was based purely on the number of conditions that were found to be satisfactory, giving 6 levels of performance as in the following table:

| Level code | No. of satisfactory quality items | Score |
|---|---|---|
| 3.1 | 5/5 | 19 |
| 3.2 | 4/5 | 15 |
| 3.3 | 3/5 | 12 |
| 3.4 | 2/5 | 8 |
| 3.5 | 1/5 | 3 |
| 3.6 | 0/5 | 0 |

It was realized during completion of the questionnaire, however, that the relative importance of different conditions would depend on the use made of the estuary. For this reason, the way in which this criterion is scored may need further attention.

### Criterion 4: Salinity

Scoring here was based on the number of the following aspects in which salinity conditions were considered to be unsatisfactory:

1 The freshwater/brackish component (salinity < 10 ppt) has altered in volume, or the frequency and duration of its occurrence has increased/decreased, by more than 30%.

2 The frequency and duration of occurrence of hypersalinities (salinity > 35 ppt) or the magnitude of hypersaline maxima have increased by more than 10%.

3 Vertical differences in salinity (surface to bottom) vary in magnitude or the frequency and duration of occurrence by more than 10%.

The scores allocated were as follows:

| Level code | No. of unsatisfactory salinity items | Score |
|---|---|---|
| 3.1 | 0/3 | 15 |
| 3.2 | 1/3 | 11 |
| 3.3 | 2/3 | 4 |
| 3.4 | 3/3 | 0 |

### Criterion 5: Freshwater and Seawater flushing

This criterion is based on the answer to the single question: "Has the occurrence and magnitude of freshwater flooding of the estuary altered from the ideal (natural situation) to the extent that the dominance of freshwater flushing has been partially or fully replaced by seawater flushing?". This is a simple binary criterion, in which a score of 7 (the weight of the criterion) is gained for a "No" answer, and a score of 0 otherwise.

### 4.2.5 Conclusions

The work described in this section made use of a number of different techniques. The initial decision conference clarified many of the conceptual problems, reorganised the criteria, defined the criteria levels, and scored these levels. The subsequent questionnaires allowed the weights and final scoring system to be determined by use of the linear programming formulation described in Appendix C.1.

It was clear that the workshop helped to clarify many issues for those present, and the resulting criteria and scoring system appeared to be generally acceptable. The conjoint scaling exercise had the additional benefit of highlighting violations of the assumptions of additive independence necessary for using additive scores. It seems probable that this is not a serious problem, but did suggest that the water quality scoring system may need to be revised at a later date.

Subsequent to this workshop CERM has published an interim report (CERM, 1996), which included the above scoring system in the overall estuarine importance rating scheme.

## 4.3 Forestry Land Use (Maclear District)

### 4.3.1 Introduction

The work described in this section arose from plans to extend commercial forests in the Maclear district in the northern part of the Eastern Cape province. The natural vegetation of the area is afro-montane grassland, and cattle grazing was previously the dominant land-use. Commercial forestry began in the district in 1989, subsequent to the area being designated a potential forestry area, and expanded rapidly to the present levels of around 32 000 ha. A prolonged drought meant that many farms were on the market at the time of initial forestry expansion in the area.

The Maclear district, about 200 000 ha in extent, presently consists of about 58% natural pasture (as opposed to 83% in 1988), crop and small scale forestry take up a further 13% and commercial forestry the remaining 29%, although only about 55% of this is actually planted. North East Cape Forest's (Mondi) original permit was for 56 000 ha, while private farmers also held permits for additional plantations. Mondi would ideally like to expand to about 55 000 ha in the area. Mondi hopes that enough land will be afforested in the Transkei, under a community forestry system in partnership with Mondi, to enable them to pursue their 'ideal' scenario of around 100 000 ha under afforestation to feed a large pulp-mill. Initiatives were already underway to negotiate with the tribal authorities and civic organisations in these areas.

In 1990 the World Wild Fund for Nature (WWF), identified conservation priorities throughout the world. In Africa the three most transformed regions were identified as afromontane grasslands, tropical forests and wetlands. Afromontane grasslands throughout Africa, as far as Ethiopia, are threatened by both agriculture and forestry, and in some areas (for example, the Eastern Transvaal) they have more or less disappeared. WWF therefore helped to fund a project aimed at the conservation evaluation of afromontane grasslands, which form one of South Africa's seven floristic regions.

As South Africa is a signatory to the Convention on Biological Diversity (1992), it is committed to maintenance of its biodiversity. The Maclear district is at the southern end of the Eastern Mountain 'hotspot' of plant diversity, one of eight recognised for southern Africa. About 30% of the plant species are endemic and about 5% of the 'hotspot' is formally conserved, almost exclusively in the Natal Drakensberg (Cowling and Hilton-Taylor, 1994). There is presently no formally conserved area which is representative of the northern part of the Eastern Cape, and the Maclear area may have high conservation potential as parts are as yet relatively untransformed, but threatened by forestry expansion, increased crop farming and bad grazing practices.

The Department of Agriculture has identified two relatively homogeneous farming areas in the Maclear district. Firstly, the Drakensberg Highland Sourveld (Themeda and some Themeda-Festuca veld) which has rainfall in the range of 800 to 1400 mm, shallow soils and is primarily suitable for cattle grazing. Important wetland and sponge areas occur here which feed streams and are used for stock water and some irrigation. Secondly, the Elliot/Maclear High Potential Sourveld, which is lower lying with rainfall in the range 700 to 1000 mm. The deeper soils mean that some of this land is suitable for cultivation of maize and potatoes, and it is also suitable for cattle. It is thought that afromontane grasslands evolved over millennia in conjunction with a frequent fire regime and grasslands are presently managed under this assumption, and because the grass becomes less nutritious and unpalatable to grazers as it senesces in autumn.

With the change from predominantly cattle grazing to commercial forestry over the last seven years, the economic and social structure of the area has changed considerably. This needs to be considered in the context of the fact that the Eastern Cape has one of the highest unemployment figures in the country at around 45% (CSS, 1994). Also relevant is the fact that the relatively wealthy farms of the Maclear area, are bordered on the eastern side by the Transkei, where population pressures and cultural factors have led to massive overgrazing. In addition, economic pressures have led to a trend in recent years for farmers in this and adjacent areas to move into towns, with only the larger farms remaining 'viable'.

### 4.3.2 Workshops

In order to help in the formulation of an integrated land-use decision making process, three workshops have been held to assess land-use and specifically afforestation decisions in the Maclear magisterial district. Each workshop, or decision conference, brought together a small group of people, who were representative of a range of interests and who could bring to the workshops the relevant expertise and information. The workshops aimed at establishing a generic framework for land-use decision-making, using the Maclear area as an example, using the concepts described in Section 2 and Appendix A as guideline.

The first workshop was necessarily exploratory, while the second, using development scenarios, built on some of the insights gained in the first, but, due to time constraints went no further than definitions of alternative courses of action and of criteria. During the third workshop a new set of scenarios was evaluated on the basis of the criteria selected at the second workshop, and a more complete analysis was carried out (apart from a few criteria for which information was still unavailable). The following subsection describes the three workshops in more detail.

### 4.3.3 The first workshop (27 & 28 November, 1996)

The first workshop was held in the town of Ugie in the Maclear district. The participants included representatives from North East Cape Forests (Mondi), Department of Agriculture, Department of Nature Conservation and Forestry (Univ. of Stellenbosch) and the CSIR. A representative from the Department of Nature Conservation (Eastern Cape) was unable to attend. After some discussion the participants at this workshop agreed to concentrate on the micro-scale, or the farm level. This was for a number of reasons: this is the level at which permits are presently assigned, the level for which wildlife indices are available (Armstrong et al. 1994), and because farms are the unit of exchange.

The participants identified the main criteria and subcriteria for each, as well as ranges of possible values of the subcriteria at the farm level Much of the value assessments were carried out according to the "mid-value splitting" technique (see Appendix A), and the workshop participants appeared able to identify the non-linear relationships between the level of a criterion and the value of that level without much difficulty. Distinct non-linearities were revealed for some criteria (e.g. numbers of animal endemic species), while for others, because the range of values on a farm would be so small, the relationship was essentially linear (e.g. percentage reduction in low flows, number of new employment opportunities).

An attempt was made to assess the priorities or weights which could be applied to the different criteria, but this proved difficult. Within the conservation category, the first priority criterion was either the number of red-data species or the total number of plant species (a surrogate for ecological integrity). Which conservation criterion was the main priority would depend to some extent on which red-data species were concerned. The most important criteria from a forestry point of view was the mean annual increment (MAI). However, the relative importance of MAI, harvesting costs and distance to tarred roads would depend on the distance to the tarred road (i.e. if distance was less than 10 km, then MAI and harvesting costs were first and second priority, if the distance were more than 50 km then the actual distance would be of first priority). The first priority of non-conservation and non-forestry criteria was either the reduction in low flows or employment generated. Which criterion was considered to have highest priority when all criteria were considered, depended to a fairly large extent on the interest group. The suggested top priorities were: number of red data species, MAI, employment generated and reduction in low flows.

Perhaps the most useful result of this first workshop was the conclusion that although information was necessary at the site level, it was not possible to make coherent decisions at this level, and that incremental decisions made without reference to a larger scale would result in sub-optimal allocation of land to particular land-uses. It was therefore agreed that the subsequent workshop should concentrate on the meso-scale, which for our purposes means the Maclear district. Although the conclusion of the workshop was that decision frameworks were needed on larger scales, the workshop produced some interesting and useful results and helped to inform the subsequent discussion of larger-scale planning.

### 4.3.4 The second workshop (14 & 15 May, 1996)

The inclusion of a larger scale of reference meant that there were more role-players involved in the second workshop. Participants included representatives from North East Cape Forests (Mondi), Department of Agriculture (Directorate of Land-Use Planning), Department of Water Affairs and Forestry, Department of Nature Conservation, Department of Nature Conservation and Forestry (University of Stellenbosch), the CSIR and the mayor of Maclear. With the decision to concentrate on the meso-scale (the Maclear district boundaries), came a decision to use a scenario based approach, i.e. SBPP.

#### Scenarios

Before the workshop, representatives from Mondi and the Department of Nature Conservation (University of Stellenbosch) helped to develop scenarios to represent a range of possible situations. The conservation scenarios were based on the combination of land-types (as classified by Armstrong et al., 1994) which would most efficiently account for total species and afromontane endemic species (based on grasshoppers, butterflies, small mammals and birds). It was considered that ranges adequate for certain birds would provide enough habitat for larger mammals and that the ranges required for plants would be similar to those required for insects and smaller mammals. The forestry information indicated likely forestry expansion from the present situation, through a 'consolidation' phase to the maximum likely afforestation in the area. This information was combined into five scenarios, ordered from a maximum amount of natural grassland available for conservation, to a maximum amount of afforestation (implying the least amount of land for conservation). The crucial assumption was made that the natural pasture in the area was well-managed and therefore fulfilling a conservation role.

The workshop participants were provided with a short description of the five potential development scenarios for the district and maps showing the land-types and forestry owned areas. Scenarios were also formulated for the area prior to any afforestation but these are not included. Although more data were to become available at a later stage, at the time of this workshop, the scenarios were found to be specified at an inadequate level of detail for assessment on the basis of separate criteria. As one of the reasons for the revision of the present permit system is because it is time-consuming and expensive, and as different amounts of data are available for different areas, any decision aiding process will need to be flexible in terms of data requirements.

#### Criteria

Once the scenarios had been discussed to some extent, use was made of the GroupSystems software (Ventana Corporation, 1994) in order to generate ideas referring to the question *"What points of view, interests or other issues need to be taken into consideration when selecting between the forestry development scenarios?"*. The

GroupSystems software allows participants to enter their ideas at their own terminals simultaneously and anonymously into a list of ideas contributed by the rest of the participants. The complete list of ideas appears on all the participants' screens, who can further comment on any of them. Subsequent to the generation of ideas, and based on them, an initial set of criteria were developed. (See later in Figure 4.4 for a somewhat refined version of these criteria.)

The criteria which were identified grouped naturally into the three broad objectives of equity, economic growth and sustainability which have been suggested for project or policy level planning (e.g. Faucheux and Froger, 1995 and van Pelt, 1993). While economic growth used to be approximated by measures such as GDP, it is now acknowledged that these measures are inadequate and they need to include measures of quality of life such as education, nutrition, basic freedoms and spiritual welfare (Perrings, 1994), and these aspects also emerged clearly from the workshop.

The criteria developed in the workshop appear to be useful indicators of the importance, the benefits, the losses or the costs of a change in land-use from the point of view of a particular category. However, many of the criteria still needed to be made operational, either through the definition of quantitative measures (such as Rand per cubic metre of water, for example) or through the redefinition of the criterion to ensure unambiguity at the stage of ranking scenarios.

Between this workshop and the next, the criteria were further classified and redefined with the aim of achieving the required operational definitions. The resultant value tree is displayed in Figure 4.4, which was accepted by all participants at the start of the third workshop.

### 4.3.5 The third workshop (28 & 29 October, 1996)

#### Background work

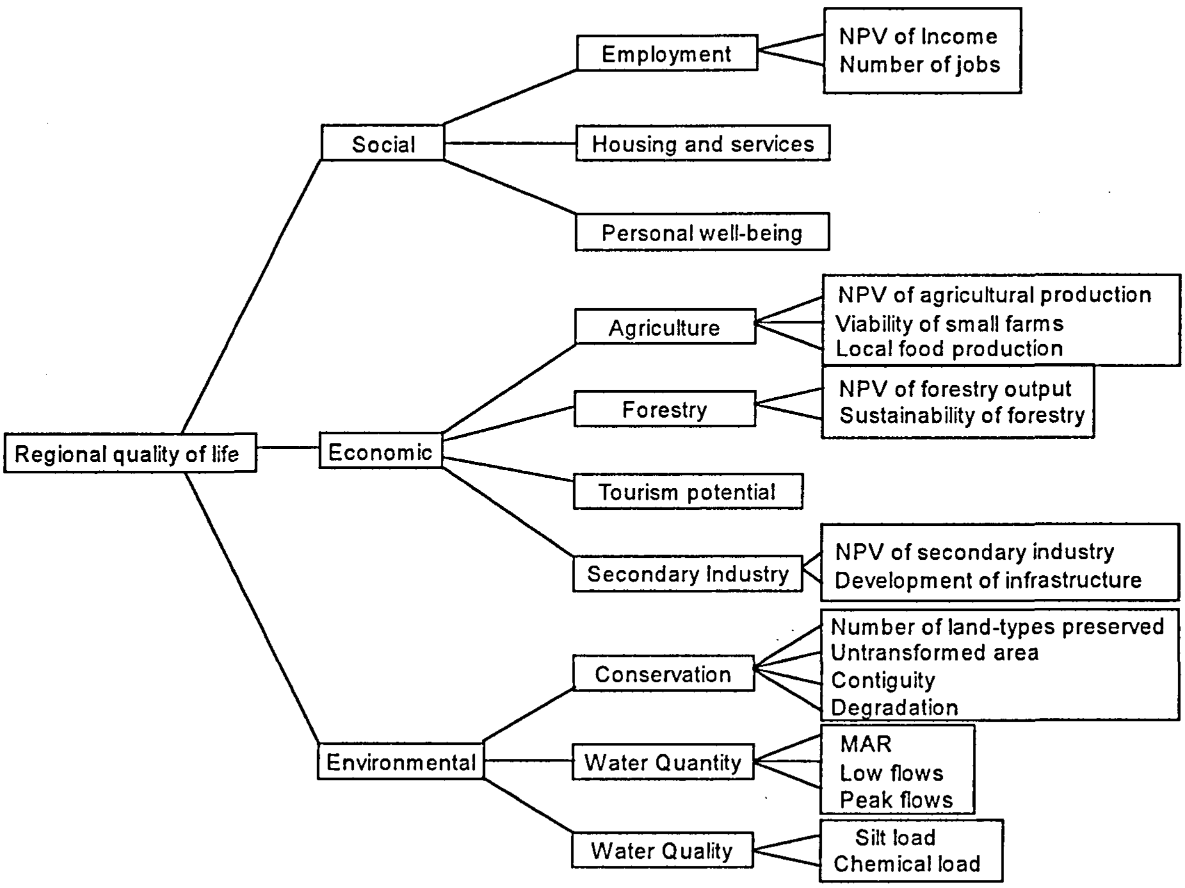
The Maclear district was visited prior to the third workshop in order to gather some of the relevant information for the specification of the scenarios. Meetings were held with the provincial Department of Agriculture (Directorate of Land-use Planning), Mondi, and the mayors and town clerks of the two towns in the area (Maclear and Ugie). A limited number of farmers and individuals from the towns were also interviewed. Mondi was able to supply details of exactly which farms had options to purchase pending, and where, besides these, they would still like to expand within the district. Besides this information, Mondi supplied detailed maps of the actual arrangement of plantation blocks within their presently afforested farms. This allowed for the updating and more detailed specification of six development scenarios, which can briefly be described as follows.

**Scenario 1:** Approximate Status Quo: Approximately 35 000 ha of afforestation, with one timber saw mill to be constructed in the region.

**Scenario 2:** Expansion of forestry to take up existing options on farms (expanding forestry to 44 000 ha); saw mill to be constructed.

**Scenario 3:** Afforestation of all suitable land except for a small ecologically sensitive area to the north of the region (giving 50 000 ha of forests); saw mill to be

Figure 4.4: Value tree for the 3rd forestry workshop



constructed.

**Scenario 4:** Maximum afforestation (to 53 000 ha) plus the saw mill

**Scenario 5:** Afforestation as in scenario 4, but with further afforestation in the Transkei, with both forests feeding a pulp mill set up in the Maclear district.

**Scenario 6:** Afforestation as in scenario 4, but with timber transported to a pulp mill in another region.

Although one of the consequences of Scenarios 5 and 6 would be forestry expansion into the Transkei, impacts of this expansion was not directly assessed, although cognisance was taken of this in the relevant scenarios.

### Data organisation

Although it was hoped to enter the relevant information into a GIS format, the problem obviously lending itself to spatial analysis, this was not possible as insufficient data was available in this form. As much information as was possible was, therefore, entered into an Excel spreadsheet. Each farm and its surface area was linked to an associated land type, based on altitude, rainfall and underlying geology. Each land type was in turn linked to a list of associated grasshopper, butterfly and bird species (Armstrong et al. 1994). The average pay and average number of workers per hectare, on forestry and agricultural land, was determined from the Central Statistical Services data for the district, and from Mondi's employment records, and these were linked to farm areas. The farms presently afforested were designated as such and those that would be afforested in the other scenarios were entered into the 'scenario' function of the software package. This meant that in order to change from one scenario to another, one needs simply to click on the new scenario within this function. Graphics and statistics are then appropriately updated. The link of each farm to a land type and the associated species data, meant that the number of species represented at least a certain number of times, in at least a minimum area could be calculated. The number of times each land type occurred in an unafforested state was calculated, as was the total afforested and unafforested area of each land type. Most of the unafforested land remains as grassland and is used for grazing, providing some conservation value. However, some land-use information was available for the catchments in the area, which gave the percentage crop cover (crop-lands providing no or little conservation value) and this could be used to modify the calculations to give an idea of untransformed area available for 'conservation'. A recently completed assessment of the hydrological impacts of present and expanded (up to about 43 000 ha) forestry levels was used to supply information regarding impacts on flows in the quaternary catchments in the district (Forsyth et al., 1996).

### Workshop

The third workshop, which lasted just over a day, took place in the Maclear district, in October, 1996. The scenarios and brief summaries of relevant information from the Excel spreadsheet were presented to the participants and source material was made available. The value tree obtained from the previous workshop was presented to the participants (Figure 4.4), using the VISA software (Belton and Vickers, 1990). The scenarios, data and value trees were further discussed and clarified by the participants.

Some of the criteria related easily to natural quantitative attributes for which ranges of values could be assessed, and value functions derived using thermometer scales or mid-value splitting (Appendix A). In a few cases, however, although much of the information was in principle available, this process still proved difficult and/or controversial. For example, the conventional wisdom in the forestry industry seemed to be that a pulp-mill would employ directly and indirectly a large number of people (up to 3000), implying a very high multiplier effect, but this was difficult to verify from the available data. The only pulp-mill employment figures available were for direct employment, or figures aggregated for the whole country for all pulp-, paper- and board-mills, which were difficult to interpret.

Due to time constraints and problems relating to some of the economic information, relative weights of the criteria could not be determined at this workshop.

### 4.3.6 Recent Developments

Before the first workshop, the Department of Water Affairs and Forestry published the green paper entitled: "Towards a Policy for Sustainable Forest Management in South Africa: A Discussion Paper". Although the three workshops and the approach used had not specifically been linked to government policy, the parallels between the SBPP approach, the meso-scale of decision-making and the stated government policy are striking. These parallels include the need for broad consultation, the acknowledgement that full scale EIAs are expensive, that district level planning is required within a hierarchy of larger and smaller scale planning, and that development plans need to be considered on the basis of social, environmental and economic criteria.

Since then the document 'Sustainable Forest Development in South Africa: The Policy of the Government of National Unity' (the 'White Paper') has been published which sets the policy to be put into a National Forestry Action Programme by April 1997.

The CSIR has set up a baseline for the hydrological modelling of the quaternary catchments in the Maclear district. The availability of a calibrated hydrological model for the area means that there are potentially reliable data available to inform the criteria relating to water quantity, and that the long term hydrological impacts of afforestation can be included. At the same time, the Department of Water Affairs and Forestry in conjunction with CSIR and the Department of Nature Conservation is developing a Strategic Environmental Assessment procedure (similar to Integrated Environmental Management or IEM) aimed at improving decision making in the forestry industry.

### 4.3.7 Discussion and suggestions for future work

The iterative and participatory nature of the process allowed the participants to significantly increase their levels of understanding of the issues at stake, as well as of their own preferences and values. To a certain extent it also helped to separate value judgements and emotional issues from the facts which may or may not have supported them. The discussions also allowed for an understanding of what was an appropriate scale for decision-making in the area, with the acknowledgement that larger and smaller scale decisions would still need to be made. The scenarios developed by the time of the third workshop were at a sufficient level of detail for people to find realistic and accessible, while the criteria where also much clarified by this stage. The final stages of this process are still on-going with some research needed on multiplier effects. Once this is available, the relative weights of the various criteria can be assessed and the process 'completed'.

The workshops and background work leading up to these helped to clarify the advantages and disadvantages of different approaches. Difficulties encountered while developing scenarios and criteria helped to isolate the causes of these difficulties which could, perhaps, be avoided in future. These are summarised below.

**Use of GroupSystems software versus open debate.** The GroupSystems software was used twice within the second workshop to get as wide a range of input in as short a time as possible. The advantages of this approach were that: many ideas could rapidly be generated, the anonymity of the system probably promoted honesty among participants, the generation of ideas and the process were not dominated by those with stronger personalities or more confidence, ideas could easily be organised into groups or categorised and direct confrontation could be avoided. The disadvantages of using this system were that: those with little or no knowledge of computers are at a disadvantage, the misperception is created that the ideas generated and then categorised have been explored, time is required to explore these ideas subsequently. In contrast open debate can easily degenerate into two-way disagreements, but with proper facilitation more understanding of the issues may be reached than may occur electronically.

**Scenario Development.** Scenarios need to include an appropriate level of detail. By the time of the third workshop, enough detail was available, although certain aspects still require further research. Flow and influence diagrams can helped to clarify these for participants, while also indicating the different spatial and temporal scales involved. It was clear that the various assumptions regarding management practices (in both forestry and agriculture) needed to be included in the scenarios. However, including the full range of these may have meant that too many scenarios were created, so a realistic balance needs to be found. Ideally predictions regarding demographic changes as well as other changes such as from grazing to potato farming, increases in small-scale farming also need to be included. For the most part this type of information was unavailable. The emphasis on detail needs to be balanced be the realisation that for other areas where this approach may be useful much less information may be available. The time horizon used by implication in this study was the length of one saw-log rotation (about 25 years), however most ecological process will take place on a much longer scale.

The differences between the attributes of a scenario, the impacts of a scenario and the criteria used to assess the scenario need to be clarified. Cause and effect relationships need to be explored, as well as the links between these and the criteria used. An example of this is seen in the criterion 'viability of small-scale farming'. The redistribution of land to small-scale farmers is a socio-political objective, and the scenarios as presently defined may allow this to varying degrees. For this reason, it is a valid criterion. At the same time the existence of small-scale farming in the district will have impacts on a number of categories (e.g. water demands, soil erosion, the conservation value of grazing lands). Therefore the potential extent of small-scale farming should ideally be included in the scenario descriptions as an attribute, to allow for the measurement of these impacts.

**Criteria and the assessment of scenarios.** The scenarios and relevant available information were entered into an Excel spreadsheet for the third workshop. This allowed for the interactive updating of scenarios and assessment of impacts. The use of GIS for this purpose would have the added advantage of allowing spatial analyses such as for the 'contiguity' criterion within the conservation category.

The criteria defined across various categories need to be consistent as far as possible. For example, various measures of income suggested for the forestry, agriculture and tourism categories, were replaced by Net Present Value, while some were removed to avoid double counting of issues. Reference may need to be made to larger and smaller scales to verify decisions. Information is needed on the micro (farm) scale and even finer levels, while the district level decisions will need to be appropriate when considered within a provincial or national (macro) level. The Strategic Environmental Assessment procedure may well provide a more macro level decision being informed by and informing the district level process described here.

# Chapter 5

# Development of Scenario Based Policy Planning Procedures

## 5.1 Use of Ordinal Information

Although the use of the "thermometer scales" is very helpful in conveying information about strengths of preference, there may be occasions in which participants in the process find the assessment of gaps difficult to do, or to find consensus on. In this and the next section, we discuss the modifications needed to the procedure in order to allow for imprecisions in assessing the scores. We start by examining the situation in which participants can do no more than rank order the policy scenario options in terms of each identified criterion. This may apply either within one group, in which conflicting criteria are identified, or between groups in the search for an overall evaluation of options (with the interests of each group viewed as "criteria").

Suppose that $N$ policy scenarios have to assessed in terms of $n$ criteria or interests. Let $r(i,j)$ be the rank position of scenario $j$ according to criterion or interest $i$, ranking from best to worst in the sense that $r(i,j) = 1$ if policy scenario $j$ is adjudged the best option in terms of criterion $i$, $r(i,j) = 2$ if it is second best, etc., with $r(i,j) = N$ indicating the worst option. (Ranking in the opposite direction is also possible.) The work of Arrow which has been termed the "impossibility theorem" (e.g. Keeney and Raiffa, 1993, Section 10.2) has demonstrated the fundamental limitations of basing any overall score for each option purely on the ranks. Nevertheless, one commonly used approach to seeking a consensus ranking of the options is to form the "Borda score" for each option $j$, defined by the sum of the rank positions:

$$B_j = \sum_{i=1}^{n} r(i,j).$$

In this approach, the option minimizing the Borda score $B_j$ would be selected as the "best" in some sense. A generalization of the Borda score would be to weight the criteria differently, i.e. to use:

$$B_j = \sum_{i=1}^{n} w_i r(i,j).$$

A fundamental problem with this approach is the underlying assumption that a gain of one rank position for a particular criterion has the same overall benefit, no matter whether it is from worst to second worst, from second best to best, or any other intermediate shift. It is difficult to verify the validity of this assumption in any specific case, but it seems unlikely to hold generally under all circumstances. In order to avoid this problem, let us start with the weaker assumption that for each criterion there exist values (cardinal rather than purely ordinal), which may be unobserved, or even unobservable, associated with each rank level for each criterion. Thus let $u_{ir}$ be the "value" of achieving a rank level (as defined above) of $r$ for criterion or interest $i$. Although the values themselves may not be known, we can certainly assert that:

$$u_{i1} > u_{i,2} > \cdots > u_{iN}.$$

For practical implementation of the linear programming models described below, it is convenient to rewrite the above in the following form:

$$u_{ir} - u_{i,r+1} \geq \epsilon \quad \text{for } r = 1, \ldots, N-1 \tag{5.1}$$

where $\epsilon > 0$ is some form of discrimination threshold.

Analysis based on this formulation was suggested by Cook and Kress (1991), and much of the proposals and examples in this Section and the next can be seen as variations (generally simplifications) of their approach.

Subject to standard assumptions concerning preferential independence of the user groups, there exist values for the $u_{ir}$, satisfying the above constraints, such that it is legitimate to form a total score for each scenario $j$ ($j = 1, 2, \ldots, N$) of the form:

$$V_j = \sum_{i=1}^{n} u_{i,r(i,j)} \tag{5.2}$$

where the scaling of the scores for each user group may in principle differ to reflect different importance weightings.

The crucial question is then: for any given policy scenario $k$, does there exist a set of values for the $u_{ir}$ satisfying (5.1), and such that $V_k \geq V_j$ for all $j \neq k$? If the answer is in the positive, then this implies that the optimality of $k$ cannot be excluded on the basis of the given rank orders. The question is answered by maximizing $V_k$ subject to (5.2) (for all $j$) and (5.1) (for all $i$ and $r$). In fact we can maximize the extent to which scenario $k$ outperforms all others by maximizing the difference between $V_k$ and the average of the $V_j$'s for all other $j$, i.e. by maximizing:

$$V_k - \frac{1}{N-1} \sum_{j \neq k} V_j \tag{5.3}$$

This is a simple linear programming problem, but the problem as posed in this way is incomplete, as it will allow (for example) one user group to totally dominate all others (by making it's utility measures much larger than for all other user groups, thus giving it considerably more weight than all others). Additional constraints are thus necessary to ensure more balanced measures: three possible linear programming formulations (at least) emerge which appear to be of interest. Full details are given in Appendix C.2, but in outline these can be described as follows.

Table 5.1: Usergroup preference rankings for Wilderness

| WILDERNESS | SCENARIOS Ranked from best to worst | | | | | | |
|---|---|---|---|---|---|---|---|
| FORESTRY (A) | 2 | 3 | 5 | 6 | 1 | 4 | 7 |
| AGRICULTURE (B) | 2 | 3 | 7 | 1 | 5 | 6 | 4 |
| DOMESTIC | 2 | 1 | 5 | 4 | 3 | 7 | 6 |
| TOURISM (D) | 5 | 4 | 2 | 1 | 6 | 3 | 7 |
| ENVIRONMENT (E) | 1 | 2 | 4 | 3 | 5 | 6 | 7 |

**LP1: Equal weights; Arbitrarily allocated "gaps"** In other words, although the constraints (5.1) are required to hold, the differences between successive values $u_{ir}$ and $u_{i,r+1}$ may otherwise take on any arbitrary positive values. In thermometer scale terms, the scenarios may be arbitrarily spread across the scale, provided only that the rank ordering remains as given. This formulation thus simply maximizes (5.3), subject to the constraints (5.1) and (5.2), and some normalization.

**LP2: Equal weights; Decreasing "gaps" with increasing ranks** This formulation allows for an additional set of constraints to allow for the common case in which the importance placed on the distinction between best and second best options may be perceived to be of a much higher order than the distinction between worst and second worst, say.

**LP3: Variable weights; Decreasing "gaps" with increasing ranks** This formulation is a generalization of LP2, to allow some variability in the weights placed on each user group.

In order to illustrate the uses of the above procedures, it is useful to re-examine a study carried out by Fijen and Kapp (1995) on water management strategies for the Wilderness, Swartvlei and Groenvlei Lakes catchment. The permission of Mr Fijen for our use of the details of their study is gratefully acknowledged.

Seven detailed development scenarios for the Wilderness and Swartvlei catchments were developed, and their consequences in terms of a number of attributes were evaluated. In the motivating the proposed strategy, Fijen and Kapp describe the results of the ranking of these 7 development scenarios by each of five user groups in the Wilderness and Swartvlei catchments respectively. No further scoring of the scenarios (to establish the perceived "gaps" between scenarios by the various groups) was undertaken. Their results are summarized in Tables 5.1 and 5.2 respectively. Overall "consensus" rankings for preferences, in each of the two catchments separately, were obtained by use of a Borda score. In each catchment, scenario 2 ("natural mouth, abstractions: present or slight increase") was identified as a clear winner, a result which was not affected by differential weighting of user groups.

All three of the LP forms defined above were run for both the Wilderness and Swartvlei preferences, using $\epsilon = 1$ (i.e. 1% of the allowable range, although in fact

Table 5.2: Usergroup preference rankings for Swartvlei

| SWARTVLEI | SCENARIOS Ranked from best to worst | | | | | | |
|---|---|---|---|---|---|---|---|
| FORESTRY (A) | 2 | 3 | 5 | 6 | 1 | 4 | 7 |
| AGRICULTURE (B) | 2 | 3 | 7 | 1 | 5 | 6 | 4 |
| DOMESTIC (C) | 5 | 4 | 6 | 2 | 1 | 3 | 7 |
| TOURISM (D) | 5 | 4 | 2 | 1 | 6 | 3 | 7 |
| ENVIRONMENT (E) | 1 | 2 | 4 | 3 | 5 | 6 | 7 |

the results were insensitive to choice of $\epsilon$). Tables 5.3 to 5.6 summarize the results for LP1 and LP2 in terms of the "consensus" rank orders obtained when maximizing the position of each scenario in turn relative to the others. The results of LP3 are not shown, as 20% variations in weights applied to different interests had virtually no effect on the first three rank positions (i.e. differential weights on the user groups have little or no influence on the outcome).

It is instructive to compare the results in Tables 5.3 to 5.6 with the conclusion reached in the original report that 2 was clearly the preferred choice.

For the Wilderness catchment, it is seen from Table 5.4 that only scenario 2 can be made top-ranked, no matter how the individual utilities are manipulated, *provided that* the conditions of LP2 apply. However, if the individual utilities are free to be chosen in any way, subject only to 5.1, then scenarios 1 and 5 can also be forced into the top rank. For example, scenario 1 would be top ranked if each user group scored the scenarios as in Table 5.7, which is fully consistent with the rank orderings in Table 5.1.

The primary feature of Table 5.7 is the existence of a strong threshold for each usergroup, dividing the scenarios essentially into two groups. For example, with these scores, the Forestry group would be saying that scenarios 1, 2, 3, 5 and 6 are all reasonably acceptable (even though there is a preference ordering between them), but that scenarios 4 and 7 are completely unacceptable. Since no attempt had been

Table 5.3: Rank positions for Wilderness scenarios based on LP1

| **Rank** | **Scenario being maximized** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Position** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **1st** | 1 | 2 | 2 | 2 | 5 | 2 | 2 |
| **2nd** | 2 | 1 | 3 | 4 | 2 | 5 | 1 |
| **3rd** | 5 | 5 | 1 | 1 | 1 | 6 | 5 |
| **4th** | 3 | 4 | 4 | 5 | 3 | 1 | 3 |
| **5th** | 4 | 3 | 5 | 3 | 4 | 3 | 7 |
| **6th** | 6 | 6 | 6 | 6 | 7 | 4 | 4 |
| **7th** | 7 | 7 | 7 | 7 | 6 | 7 | 6 |

Table 5.4: Rank positions for Wilderness scenarios based on LP2

| Rank Position | Scenario being maximized | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1st | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 2nd | 1 | 1 | 3 | 1 | 5 | 5 | 1 |
| 3rd | 5 | 4 | 5 | 5 | 1 | 1 | 5 |
| 4th | 3 | 5 | 1 | 4 | 3 | 3 | 3 |
| 5th | 4 | 3 | 4 | 3 | 4 | 6 | 7 |
| 6th | 7 | 6 | 7 | 7 | 6 | 4 | 4 |
| 7th | 6 | 7 | 6 | 6 | 7 | 7 | 6 |

Table 5.5: Rank positions for Swartvlei scenarios based on LP1

| Rank Position | Scenario being maximized | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1st | 1 | 2 | 2 | 4 | 5 | 5 | 2 |
| 2nd | 2 | 5 | 3 | 2 | 2 | 6 | 5 |
| 3rd | 5 | 4 | 1 | 5 | 3 | 2 | 1 |
| 4th | 3 | 1 | 4 | 1 | 1 | 1 | 3 |
| 5th | 4 | 6 | 5 | 3 | 4 | 3 | 7 |
| 6th | 6 | 3 | 6 | 6 | 7 | 4 | 4 |
| 7th | 7 | 7 | 7 | 7 | 6 | 7 | 6 |

Table 5.6: Rank positions for Swartvlei scenarios based on LP2

| Rank Position | Scenario being maximized | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1st | 2 | 2 | 2 | 2 | 5 | 5 | 5 |
| 2nd | 5 | 5 | 5 | 5 | 2 | 2 | 2 |
| 3rd | 1 | 4 | 3 | 4 | 1 | 1 | 1 |
| 4th | 3 | 1 | 1 | 1 | 3 | 4 | 3 |
| 5th | 4 | 6 | 4 | 3 | 6 | 6 | 7 |
| 6th | 7 | 3 | 6 | 6 | 4 | 3 | 4 |
| 7th | 6 | 7 | 7 | 7 | 7 | 7 | 6 |

Table 5.7: Hypothetical usergroup scores for Wilderness scenarios

| | Scenario No. | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| FORESTRY (A) | 80 | 100 | 95 | 5 | 90 | 85 | 0 |
| AGRICULTURE (B) | 85 | 100 | 95 | 0 | 10 | 5 | 90 |
| DOMESTIC (C) | 95 | 100 | 10 | 15 | 20 | 0 | 5 |
| TOURISM (D) | 85 | 90 | 5 | 95 | 100 | 10 | 0 |
| ENVIRONMENT (E) | 100 | 25 | 15 | 20 | 10 | 5 | 0 |

made to assess group values on a cardinal scale, one cannot immediately reject the possibility of values and thresholds such as those of Table 5.7, and thus the possibility that scenario 1 (or 5) may truly give maximum social benefit. In this particular example, this is not a serious problem, as scenario 2 never ranks worse than second, and remains an excellent compromise no matter what the underlying values. But the example still illustrates the need for care in simply basing recommendations on ranks, without making some effort to establish the existence or otherwise of anomalously large "gaps" between rank-ordered alternatives.

In the case of the scenarios for Swartvlei, the situation is less clear-cut than for Wilderness. Even with the use of LP2 (decreasing gaps with increasing ranks), both scenarios 2 and 5 can be argued to be possible value maximizing options (with the other always second), while for LP1, scenarios 1, 2, 4 and 5 can be argued to be potentially optimal options (while scenario 2 can be ranked as low as 3rd, behind scenarios 5 and 6). We cannot, of course, determine here how plausible or otherwise the cardinal scores yielding other selections than scenario 2 are. And in the context of this example, it may well be that scenario 2 remains the best overall choice. The example (which is real and not contrived) does however give a clear warning that the use of rank order information only can lead to highly misleading results, especially where uneven gaps occur for some or all interests or criteria (for example, when alternatives are perceived to cluster into two groups, one largely satisfactory, and the other highly unsatisfactory).

To a large extent this case study confirms the results of much other research we have conducted (and which is briefly summarized in §5.4.2), namely that the most sensitive feature of any scoring system is the relative magnitudes of the "gaps" between rank ordered items, as perceived according to different interests or criteria. In contrast, results tend to be quite robust to choice of weights applied to different interests and criteria. This has emerged again from the above example.

## 5.2 Use of Imprecise Value Judgements

The discussion of the previous section assumed that no more information was available than rank orderings of the scenarios according to each criterion. We now consider how the ideas may be extended if some imprecise scoring of scenarios and/or imprecise weighting of criteria are available. We shall use the same notation and nomenclature as in the previous section.

We shall assume that the rank orderings according to each criterion are relatively well-agreed. Where substantial disagreements arise even on the simple ordering of alternative policy scenarios, this is indicative of either or both of:

- the existence of different sub-criteria, representing different points of view of values, which need to be made explicit, at which point the scoring would be done separately for each of these;
- different understanding of the consequences of the alternative policies, which needs to be resolved by additional data gathering, modelling, etc.

Now let $v_{ir}$ be the value of the $r$-th ranking scenario according to criterion $i$ on a standardized (typically 0–100) scale; then $u_{ir} = w_i v_{ir}$, where $w_i$, is the importance weighting for criterion $i$. We suppose that the $v_{ir}$ values are not specified precisely, perhaps by indicating intervals rather than a single point on the relevant thermometer scale, but that they are consistent with the agreed rank-ordering, i.e.:

$$v_{i1} \geq v_{i2} \geq \cdots \geq u_{iN}.$$

One manner in which the imprecise judgements may usefully be expressed in a manner consistent with the above rank orderings, is to link the interval estimates for scores to bounds on successive score differences of the form:

$$a_{ir} \leq v_{ir} - v_{i,r+1} \leq b_{ir}$$

which can be expressed in terms of the $u_{ir}$ as:

$$a_{ir} w_i \leq u_{ir} - u_{i,r+1} \leq b_{ir} w_i.$$

An alternative representation of imprecise scores may be to provide some statement as to the relative sizes of the gaps between successive rank positions on one criterion. For example, for two rank positions, say $r$ and $s$, it may be stated that:

$$u_{ir} - u_{i,r+1} \geq \alpha(u_{is} - u_{i,s+1}).$$

This states that the gap between the scenarios ranking $r$ and $r+1$ on criterion $i$ is at least $\alpha$ times as great as the gap between the scenarios ranking $s$ and $s+1$ on this criterion. Note that the same ratios would apply when working in terms of the standardized scores ($v_{ir}$), as when working in terms of the weighted scores ($u_{ir}$).

The scaling of the weights is arbitrary, but it is conventional to standardize them to sum to 1, and we shall adopt this convention. As with the scores, the assessment of the weights may be imprecise, and we shall assume that for certain pairs of criteria (say criteria $i$ and $k$), bounds of the form:

$$\rho_{ik} \leq \frac{w_i}{w_k} \leq R_{ik}$$

provide the imprecise assessments.

As in the case of purely rank based assessments, it is possible to find appropriate scores and weights, consistent with the available imprecise information, that will provide the highest aggregate score of the form:

$$V_j = \sum_{i=1}^{n} u_{i,r(i,j)}$$

for any alternative scenario relative to all others. This once again requires the solution of a linear programming problem, details of which are provided in Appendix C.3. By doing this for each scenario in turn, a range of overall (aggregate) rank positions that can arise for each scenario is easily calculated.

In general, the values for $V_j$ will not stretch across the full 0–100 scale, unless the scenarios include hypothetical cases which are respectively best and worst on all criteria simultaneously. Nevertheless, the resultant values are easily re-standardized to a 0–100 scale, and in this way it is also possible to identify a range of scores that each scenario can achieve on the aggregate scale when the inputs are imprecise.

The above concepts may be illustrated by a simple example. Suppose that we have a situation in which six scenarios are to be evaluated in terms of three scenarios. If precise scores were assessed for each criterion, then something of the following form may be obtained:

| Rank | Criterion 1 | | Criterion 2 | | Criterion 3 | |
|---|---|---|---|---|---|---|
| | Scenario | Score | Scenario | Score | Scenario | Score |
| 1 | 3 | 100 | 5 | 100 | 6 | 100 |
| 2 | 2 | 90 | 1 | 80 | 5 | 95 |
| 3 | 6 | 75 | 4 | 70 | 2 | 85 |
| 4 | 4 | 60 | 3 | 65 | 3 | 55 |
| 5 | 1 | 20 | 2 | 10 | 1 | 40 |
| 6 | 5 | 0 | 6 | 0 | 4 | 0 |

Note that this is a non-trivial problem, as no scenario dominates any other in the sense of being better on all three criteria, i.e. all six scenarios are "Pareto optimal". With weights of 0.5, 0.3 and 0.2 respectively on the criteria, the standardized aggregate rank orders and scores for the scenarios are:

| Scenario | Rank | Score |
|---|---|---|
| 1 | 6 | 0 |
| 2 | 2 | 59.7 |
| 3 | 1 | 100 |
| 4 | 4 | 23.4 |
| 5 | 5 | 18.2 |
| 6 | 3 | 40.3 |

Let us suppose, however, that only imprecise statements are given. Firstly, suppose that the weights are identified only to the extent of the following bounds:

$$0.5 \leq \frac{w_2}{w_1} \leq 0.7 \quad ; \quad 0.25 \leq \frac{w_3}{w_1} \leq 0.5$$

Then assume that ranges have been inferred for all successive gaps, as follows:

**Criterion 1:**

| Value difference | Bounds |
|---|---|
| $v_{11} - v_{12}$ | 5–15 |
| $v_{12} - v_{13}$ | 10–20 |
| $v_{13} - v_{14}$ | 10–20 |
| $v_{14} - v_{15}$ | 30–50 |
| $v_{15} - v_{16}$ | 10–30 |

**Criterion 2:**

| Value difference | Bounds |
|---|---|
| $v_{21} - v_{22}$ | 10–30 |
| $v_{22} - v_{23}$ | 5–15 |
| $v_{23} - v_{24}$ | 0–10 |
| $v_{24} - v_{25}$ | 30–70 |
| $v_{25} - v_{26}$ | 5–15 |

**Criterion 3:**

| Value difference | Bounds |
|---|---|
| $v_{31} - v_{32}$ | 0–10 |
| $v_{32} - v_{33}$ | 5–15 |
| $v_{33} - v_{34}$ | 20–40 |
| $v_{34} - v_{35}$ | 10–20 |
| $v_{35} - v_{36}$ | 25–50 |

Applying the linear programming model of Appendix C.3, the following ranges of rank positions and scores are obtained:

| Scenario | Range of ranks | Range of scores |
|---|---|---|
| 1 | 4–6 | 0–13 |
| 2 | 1–3 | 23–100 |
| 3 | 1–1 | 100–100 |
| 4 | 2–5 | 15–61 |
| 5 | 3–6 | 0–58 |
| 6 | 2–5 | 4–94 |

Note that scenario 3 is best no matter what the choice of weights and scores in the stated ranges. The imprecision in inputs has left us with a high degree of uncertainty regarding the relative positions especially of scenarios 4 and 6. If the results of this analysis have to form part of the inputs to a higher level of aggregation, then these uncertainties would have to be resolved, either by tightening up on some of the inputs, or by direct holistic judgement at this stage.

A rather weaker form of information (i.e. greater degree of imprecision) is simply to recognize the four largest "gaps" which exist in the interval scales. This may lead to assertions of the following form (again, given purely as a numerical example):

(i) $v_{14} - v_{15}$ is at least twice as large as any of the other gaps on the scale for criterion 1;

(ii) $v_{24} - v_{25}$ is at least three times as large as any of the other gaps on the scale for criterion 2;

(iii) $v_{33} - v_{34}$ and $v_{35} - v_{36}$ are at least twice as large as any of the other gaps on the scale for criterion 1;

Applying the linear programming approach of Appendix C.3 with this information on scores, in place of the range information previously used, yields the following rank and value ranges for each scenario:

| Scenario | Range of ranks | Range of scores |
|---|---|---|
| 1 | 3–6 | 0–21 |
| 2 | 1–3 | 24–100 |
| 3 | 1–3 | 86–100 |
| 4 | 1–5 | 4–100 |
| 5 | 3–5 | 7–68 |
| 6 | 1–6 | 0–100 |

This yields similar, but (not surprisingly) less precise information than before. The position of scenario 3 is less unambiguous as regards ranks, although it's standardized aggregate score is always high. The position of scenario 6 is essentially undefined, implying that more precise value judgements are going to be essential.

It should be noted that the results of this and the previous section have been derived for aggregation at one level in hierarchy of criteria which may in principle exist. This may, for example, only apply to the aggregation of the conservation criteria into an overall assessment of the impacts of the alternative policy scenarios on conservation interests. Before these interests can be aggregated with other relevant interests (social, economic, political) in the full SBPP sense, it would be necessary to reach a final consensus on the rank orderings at least. This would require either a firming up of the input assessments (giving narrower ranges or more precise scores), or the exercise of holistic judgement. The decision support tools developed here are never more than an *aid* to decision making; they cannot make the decisions!

## 5.3 Application to the Development of Indices

The original concept of scenario-based policy planning was directed towards facilitating choice between competing policy options, from which one would ultimately have to be selected for implementation. In a number of cases, however, it appeared that there was a need to use similar concepts at what might be called a "pre-decision" stage, i.e. at a preliminary evaluation stage at which certain areas for action are simply being prioritized in some sense, preliminary to a later decision stage at which issues such as resource constraints will be taken into consideration. One example of this is the case study on prioritization of the conservation importance of estuaries described in section 4.2. In this case, the aim was to develop a procedure for creating an index of importance of estuaries from the point of view of conservation. The idea would be for the relevant experts to rate all estuaries according to this indexing system, which would become an input into later decision making regarding regional development, when conservation would need to be evaluated alongside other societal goals. (As an aside, it is worth commenting that the recent scheme developed by the Water Research Commission for evaluating research project proposals, as described in Offringa and de Wet, 1996, is in fact also an indexing scheme for prioritization, as described here.)

The general principles and underlying assumptions, of scenario-based policy planning, as summarized in Chapter 2, do still apply. The primary difference is that in place of a small number of comprehensive scenarios defining complete policies, we now need "mini-scenarios" to describe levels of achievement on each identified interest or criterion separately, from an ideal (best achievable) to a worst possible case. In applying the results to any particular instance, the items being evaluated will be matched to the appropriate scenarios for each criterion. While scoring systems are often developed on rather ad hoc bases, the use of the formality of the decision analysis principles underlying SBPP does help to ensure that the resultant scores have theoretical validity. Ad hoc approaches can easily overlook some of the key assumptions which are essential to justify, for example, simple additive scoring systems.

The first step, as in SBPP, is to identify clearly an appropriate set of criteria for evaluation. As before, and as well described for example by Keeney and Raiffa (1993), Section 2.4, these need to be:

**Complete** in the sense of comprehensively covering all relevant interests;

**Independent** in the sense that it must be possible to define levels of performance and their relative values on one criterion without reference to what performance is achieved on other criteria;

**Non-redundant** i.e. avoiding double counting of issues; and

**Of minimum size** i.e. avoiding a multiplication of trivial concerns which confound interpretation but add little to the prioritization.

This process is identical to that in scenario based policy planning.

Once the criteria have been established and agreed to, it is necessary to formulate a small number (preferably between 4 and 7) of descriptions of levels of performance on each criterion which might be encountered (the "mini-scenarios"). As with policy scenarios, these should not be over-burdened with detail, but should include sufficient detail to be unambiguously understood by interested and affected parties, and to allow them to be scored in the thermometer scale sense.

If the scores from different criteria are to be additively aggregated into an overall index score (which is the simplest and most transparent approach to use), then the scores themselves need to be on an interval scale of preference, i.e. gaps of the same size between different levels of achievement on one criterion must have the same impact or importance (or trade-offs with other criteria) no matter at what level it occurs. One of the key reasons for using the thermometer scale approach is that it focuses attention directly on to the gaps between alternative outcomes, and thus encourages a natural thinking in an interval scale mode.

In many cases, it may be desirable to formulate the performance levels in such a way that the gaps between them, as obtained from the interval scale, are equal. This would suggest an iterative approach, in which initial performance level definitions are assessed on a thermometer scale, and then re-formulated in such a way as to even out the gap sizes. This may well facilitate the later use and interpretation of the resultant scoring system, but is not strictly necessary from a theoretical point of view. In the case study of section 4.2 this was not in fact done.

The recommendation is that the scoring of performance levels for a single criterion be done initially on the 0–100 scale as previously suggested. At this stage the emphasis is on relative values within one criterion (comparing like with like), and not on the relative importances of the different criteria. Once this scoring is complete, attention can be given to defining importance weights. These should ideally be interpreted in a "swing weight" sense, as described in Appendix A.2.3. A variation of the approach as described in the Appendix is useful in the present context. We can start from a hypothetical case in which all performances (i.e. for all criteria) are at some set levels (generally medium to low, but at still realistic levels, in the scale). Consideration is then given to the relative worths of shifting performance on one criterion only from this set level to the ideal level for that criterion. The criterion for which this worth is adjudged the greatest takes on the maximum weight. The relative worths of the other shifts define the weights of the associated criteria relative to the most important, and this can be assessed either by holistic judgement or by consideration of trade-offs.

This swing-weighting approach is perhaps best explained by reference to a simple example. Suppose we have three criteria only, and that 6 equally spaced performance levels, defined as levels 1–6, are defined for each, which by definition then have scores of 100, 80, 60, 40, 20 and 0 respectively (i.e. level 1 being the ideal). Those providing the value judgements may then be asked to consider the following four hypothetical cases:

| Case | Performance levels for criterion | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 0 | 4 | 4 | 4 |
| 1 | 1 | 4 | 4 |
| 2 | 4 | 1 | 4 |
| 3 | 4 | 4 | 1 |

If, say, case 2 is adjudged to have the greatest worth, then clearly criterion 2 has the greatest weight. If the participants holistically judge the worths of shifting from case 0 to cases 1 and 3 respectively are 66% and 50% of the worth of shifting from case 0 to case 2, then these immediately define holistically assessed importance weights for criteria 1 and 3 respectively. Alternatively, participants may be asked what level for criterion 2 (the most important) in case 1 would be necessary to ensure that cases 1 and 2 were equally desirable. Suppose that the answer was level 3 (rather than 4), i.e. that a case with levels of 1, 3 and 4 respectively was of the same desirability as case 2. If $w_1$ and $w_2$ were the importance weights for criteria 1 and 2, then with the scores given we have that:

$$100w_1 + 60w_2 = 40w_1 + 100w_2$$

from which it would be calculated that $w_1/w_2 = 40/60 = 2/3$.

The final scoring system is constructed from the scores and the weights. Let $v_{ij}$ be the score associated with the $j$-th level for criterion $i$, and $w_i$ the importance weight associated with criterion $i$. In principle the overall score for a case in which the performance level of criterion $i$ is $\ell(i)$ is given by:

$$V = \sum_i w_i v_{i\ell(i)}.$$

In practice, it is simpler to report the weighted score for each level on each attribute directly, namely:

$$u_{ij} = w_i v_{ij}$$

which would normally be rounded to the nearest integer. If the weights $w_i$ are standardized to sum to 1, then:

1 the maximum score on criterion $i$ will be $100w_i$;

2 the score of an item evaluated as being at the best level on all criteria will be 100.

When an index scoring system is constructed in the above manner, it is possible to obtain a check on consistency, and in fact to refine the scoring system, by a method related to *conjoint scaling* which is widely used in marketing studies. Taking any pair of criteria, a table is set up with rows representing levels of one criterion, and columns the levels of the other. If the first level is the ideal in each case, the pair of levels at the upper left hand corner of the table is the best combination, and the bottom right the worst. Participants are asked to perform a partial ranking of the remaining cells in the table in two ways. Firstly, start from the upper left corner, and move down or across to identify the 2nd best, 3rd best, etc. combinations, proceeding as far as participants are able. Then the process is repeated from the bottom right, identifying 2nd worst, 3rd worst, etc. The following table is based on the case study of section 4.2, and illustrates the results of such an exercise for two criteria with 4 and 6 levels respectively, giving 24 possible combinations. The entries in the table are the rank orders, and the empty cells are where the participants were no longer able to provide rank orderings.

| Levels for | Levels for second criterion | | | | | |
|---|---|---|---|---|---|---|
| first criterion | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 1 | 3 | 5 | 7 | | |
| 2 | 2 | 4 | 6 | | | |
| 3 | | | | 16 | 20 | 21 |
| 4 | 15 | 18 | 19 | 22 | 23 | 24 |

By evaluating the sum of the scores for the two criteria at each combination of levels, it is trivial to check whether these are consistent with the rank orders given in the table. This can be repeated for a connected sequence of pairs of criteria (e.g. 1st and 2nd, 2nd and 3rd, etc., ending with the last and the first again), in order to provide a comprehensive consistency check. The check can be carried out simultaneously for all pairs of criteria, and a minimal adjustment of the scores found which will force consistency with the rank orders in the table, by solving the linear programming problem formulated in Appendix C.1.

## 5.4 Simulation Studies for Refining and Testing SBPP Procedures

One difficulty for research into decision support procedures is that the "correct" answer is unknowable, and one cannot, therefore, directly assess the quality of solutions obtained. Furthermore, the same strategic decision will not be repeated many times, and thus statistical tests of differences between the quality of results obtained with and without the use of decision support are also not possible. In some research elsewhere, students (typically MBA or similar students) have been presented with certain decision problems, and divided into groups, with each group allocated to a different decision support procedure. While this does allow for statistical testing, the commitment of the subjects to the problem must always be questioned. The first report (Stewart et al., 1993) did record some measures of user satisfaction with the approach adopted, but even this is of limited value in attempting a scientific validation of the approach.

For the above reasons, we have adopted a simulation approach. Hypothetical decision contexts, defined in terms of alternatives (differing on a range of attributes) and ideal preference structures for different group interests or criteria (represented by utility functions within a specified class) are randomly generated on the computer. The procedures are then applied to this context, building in non-idealities (such as imprecision in stating preferences or scores). By having knowledge of the "true" preferences which exist, it is possible to evaluate the extent to which the procedures facilitate the finding of the best compromises, and/or the extent to which the procedures may bias results inadvertently in one or other direction. Two sets of such studies were conducted during the course of the project, and are described below.

### 5.4.1 Evaluation of the overall scenario based policy planning procedures

The work described in this section is based on the MSc dissertation of Heynes (1995). The purpose of this work was to establish the extent to which the basic scenario-based policy planning procedures, including the use the thermometer scale scoring, is likely to converge towards a solution which can justifiably be viewed as a good consensus or compromise. The SBPP procedure is defined to be iterative. An initial set of policy scenarios is generated for evaluation, and in the light of the evaluations of these by different interests, a number of the original policy scenarios will be discarded, and a new set created (largely as refinements and/or variations of the retained scenarios). The process may then be repeated, in principle, as often as needed to reach consensus.

In order to limit the extent of the investigation, the simulations reported here were based on a double iteration only, i.e. two rounds of evaluations. In doing so, the effects of a variety of choices for some of the unspecified aspects of the procedures described in Chapter 2 could be evaluated to "fine-tune" the process. In particular, the following aspects were systematically evaluated:

- The number of scenarios presented for evaluation by the interest groups (i.e. the size of the "foreground set" as defined in Stewart et al., 1993);
- The number of scenarios eliminated at the end of the first round; and
- The procedure(s) used for selecting the scenarios to be eliminated in the first round, and for generating a final proposed "best" compromise at the end of the process: procedures considered were (i) weighted averages of the thermometer scale scores; (ii) sums of the rank positions of each scenario according to each interest or criteria; (iii) worst-case scores, i.e. the lowest score obtained across all interests for each criterion; and (iv) a simple form of outranking procedure based on the ELECTRE approach (cf. Stewart et al., 1993, for a description).

Two sets of studies were conducted, as follows:

1 **Variations of the Sabie-Sand case study:** This was based on the case study reported in Stewart et al. (1993), and detailed in Appendix B of that report. A background set of 20 scenarios was to be evaluated by four key interest groups: forestry, irrigators, rural communities and conservation. The scenarios were based on various options for four key policy elements (change in level of afforestation, cuts in irrigation, size of dam construction, and proportion of population provided with water in taps), and information was provided on costs and effects on mean annual runoff, peak and low flows.

The assumption for the simulation was that the interests of the four groups were related to the scenario attributes in the following manner:

**Forestry:** Maximize afforestation and dam size; Minimize costs

**Irrigators:** Minimize cuts in irrigation; Maximize water to rural communities

**Rural Communities:** Maximize water to communities and minimum flow levels; minimize floods (peak flows)

**Conservation:** Maximize flows (mean annual, low and peak)

These assumptions are somewhat arbitrary, and were selected only to provide a basis for simulation studies (i.e. *IF* preferences are of this form, how well would the procedures perform?)

Interests of each group were assumed to be representable by additive value functions with the marginal values for each supposed attribute of interest being of an exponential form. The parameters of these functions (weights and exponential parameters) were generated randomly for each simulation run carried out. In this way, the simulation model "knows" precisely what the preferences of each interest group are in an ideal sense, even though in practice not even the groups themselves would have this level of understanding of their preferences. (In fact, the assumed value functions should be seen as hypothetical ideals towards which the groups are striving, but which are being developed in a learning process as the process proceeds.)

2 **Hypothetical problem setting:** This can be viewed as a generalization of the previous problem. The initial (or background) set of 25 scenarios was based on four policy elements, arranged in an "experimental design" as suggested by Stewart et al. (1993). Randomly generated models produced values for a fixed number of attributes. A pre-specified number of interests were defined, whose interests were related to the attributes and policy elements in a random manner. For any one such randomly generated problem setting, the remainder of the simulation proceeded as in the previous case. This approach enabled us to evaluate the effects of issues such as the complexity of the preference structures for each interest (i.e. the number of criteria under consideration) and the number of interested parties.

For each iteration of the simulation process, we have then (in either of the problem settings) a "known" set of preferences for each group on a cardinal utility scale. As a benchmark, it is then possible to construct the Nash solution to the problem. This is a solution derived from game theory (Nash, 1950, 1951), and establishes the solution which is optimal according to a number of equity axioms, when the interests of each party are representable by cardinal utilities. (In practice, of course, this is not very helpful, as cardinal utilities for each group are very difficult to assess with sufficient accuracy, and with broad acceptability to each group, to be used as an automatic consensus-forming algorithm.)

Simulation of the process was according to the following steps:

(i) A pre-specified number ($t$) of scenarios are selected from the background set, according the procedure described in Stewart et al. (1993): in essence weights ($w_i$) for each attribute are generated randomly, and the scenario ($j$) maximizing a function of the form:

$$\max_i w_i z_{ij} + \epsilon \sum_i w_i z_{ij}$$

is selected, where $z_{ij}$ is the value of attribute $i$ for scenario $j$, and all $z_{ij}$ are defined in a maximizing sense. This is repeated many times, and the $t$ scenarios which are selected most frequently are chosen to form the foreground set.

(ii) Scores for each scenario according to each interest are set according to the value functions, corrupted by random "noise". These scores are combined to produce an aggregate ordering according to one of the aggregation methods described above. The $K$ lowest ranking scenarios are eliminated, where the number $K$ is pre-specified in the simulation run.

(iii) Step 1 is repeated, but with constraints on the weights generated to ensure that the eliminated scenarios cannot recur.

(iv) Step 2 is repeated, but only to select a final best solution.

The final solution obtained in each simulation was compared with the Nash solution. Simulation runs differed according to the ranking procedures used, the values of $t$ and $K$, and (for second problem setting only) the problem size and structure. Results from the different simulation runs were compared according to the frequency with which the Nash solution was obtained, the average rank order of the solution generated according to the Nash criterion, and the sensitivity to the random noise in the inputs.

In the case of the problem setting based on the Sabie-Sand case study, it was found that when inputs were precise (i.e. no "noise"), the best performance was obtained when using one of maximizing average score, maximizing minimum score, or outranking (ELECTRE) for elimination of scenarios in the first step; and one of the first two of these methods for the final selection. Typically, the procedure ended by selecting the first (about 2 times out of 3) or second ranking scenario out of the 20 available scenarios, according to the Nash criterion described above. Maximizing the minimum scores was also robust to quite substantial imprecisions in inputs (deviations of the order of 10% of the range of scores), whereas the other options, especially outranking, were more sensitive. It is interesting to note that even with this level of imprecision, the methods based on the assessed scores perform better than purely rank-based methods. This is consistent with the findings in §5.1.

Generalization to the wider class of problems included in the hypothetical problem settings produced very similar results, except that the outranking methods produced somewhat erratic results (picking up the optimum solution more often than any other methods, i.e. 68% of the time, but generating rather poor answers the rest of the time). Maximizing average score also produced the Nash solution nearly 2 times out of 3, but with occasionally poor results. Maximizing minimum score was generally the most consistent approach, identifying the Nash solution about half the time, but tending otherwise to select second or third best (out of 25 in this case). Of course, no decision support system can ever produce *the best* social welfare solution. The DSS can only produce a final shortlist for debate, and in this sense the procedure of selecting scenarios on the basis of maximizing minimum scores appears to do well in generating a shortlist of 2 or 3 final options.

As in the previous case, maximizing minimum scores is also robust to imprecisions in the scoring. All results were largely unaffected by choice of the number of scenarios selected for evaluation at a time ($t$), and the number eliminated after the first round ($K$). As the number of interest groups increased, quality of the solution deteriorated for all procedures except for maximizing the minimum scores, which once again remained robust.

The overall conclusions from this study were thus the following:

1 The use of scoring on thermometer scales to represent interests of each group provides a firm and justifiable basis for seeking compromise solutions, capable of generating results which are compatible with theoretical best (fairest) compromises according to the Nash criterion (see above).

2 In order to approach these Nash-optimum solutions, it is recommended that aggregation of scores across interests be carried out by examining the worst case (minimum) scores in each case. This is sometimes termed the "Tchebycheff"

criterion.

3 Use of this Tchebycheff criterion leads to results which are robust to imprecisions in inputs and to varying numbers of interest groups.

### 5.4.2 Evaluation of additive procedures for aggregating preferences

This work is fully documented in a separate paper (Stewart, 1996), and only a summary is provided here. The question investigated was the extent to which additive aggregation of scores across criteria is sensitive to the underlying assumptions. We have generally recommended this form of aggregation across the subcriteria within single interest groups, and have often it used across interests (results of the previous sub-section notwithstanding). The use of weighted sums of scores is theoretically justified, provided that a number of axioms concerning the underlying preference structures are satisfied. Key amongst these axioms are the following:

(a) The scores must lie on an interval scale of preferences, i.e. equal increments on this scale must have the same incremental value to the interests being evaluated (measured perhaps by willingness to trade-off against some fixed currency), irrespective of the baseline.

(b) The marginal values must also satisfy "additive independence", i.e. the value gained by a fixed increment in score on one criterion, when performance levels of all other criteria are unchanged, should not depend on the fixed levels of the other criteria.

(c) Irrespective of how the weights are actually assessed, they must be interpretable in trade-off terms. Weights are often assessed in practice by holistic subjective judgement, rather than by the explicit specification of tradeoffs. This does require that the scale of measurement of the value function be clearly and unambiguously identified, for example by careful definition of the end points of the scale (*cf.* the "swing weight" concept, as described by von Winterfeldt and Edwards, 1986, p 275).

While these assumptions are easily stated in theoretical terms, it is difficult to assess to what extent they apply in practice. The manner in which the problem is structured, and in which questions are asked, can potentially invalidate one or more of the axioms, and thus the use of additive aggregations. Some problems which can arise are the following:

(i) Over-smoothing or linearization of scores or value functions, when in fact there may be threshold levels below which the relevant parties would be very resistant to compromise;

(ii) Reference point effects: It is known that people will view certain results as gains and others as losses, relative to a perceived reference level, and that they will tend to be risk averse above this level and risk seeking below it. If the mode of questioning temporarily shifts the reference level, results may be very misleading in the long run.

(iii) Variable and imprecise preference statements: Responses may be given in the pressure of the workshop setting which are at variance with long run goals.

(iv) Criteria may be either omitted entirely, or may be defined in a manner which does not satisfy the additive independence property as defined above.

In view of the other desirable properties of additive aggregation (its simplicity and transparency), it was considered useful to conduct a systematic study of which of the above problems are most influential on the results obtained, so that care can be taken on these points at least.

Once again, a simulation approach was adopted. Hypothetical decision scenarios, differing on a pre-specified number of attributes, were randomly generated in the computer. To avoid triviality, these scenarios were generated in such a manner that none were dominated, in the sense of being worse than another on all criteria simultaneously. Different classes of distributions of attribute values across the alternative scenarios were used in different simulation runs.

Then value functions, relating scores on each criterion to the underlying attribute values, were also generated randomly. All of these allowed for a reference point, as described above, below which value functions are convex (risk seeking), and above which they are concave. Simulation runs differed according to the positions of the reference point in the ranges of attribute values, and the steepness of the curve when approaching resistance thresholds.

Non-idealities and imprecisions were introduced by (a) randomly excluding criteria; (b) randomly creating new criteria as mixtures of the true criteria (to destroy independence); (c) randomly perturbing reference points and responses to simulated questions; and (d) modelling value functions in piecewise linear form, and adjusting the number of "pieces" in order to create greater or lesser amounts of artificial smoothing of responses. Detailed results of the simulations may be found in Stewart (1996), but the key findings can be described as follows:

1 Linearization of value judgements causes by far the most serious problems, and can lead to results which are almost arbitrary. This is a major concern, as the method can give the impression of being scientific and objective, when in fact it is essentially meaningless. In other words, great care has to be taken in constructing scores in such a way that the increments do truly represent strengths of preference.

2 Introduction of a relatively modest degree of non-independence of criteria has as large an influence on the results as eliminating (say) one out of seven criteria. This implies that as much effort needs to go into defining criteria properly (to ensure close to "additive independence") as into ensuring a comprehensive list of criteria. In some follow up studies, there has been preliminary evidence that the use of goal programming concepts (as opposed to the additive aggregation of scores), of which the "max-min" procedures of the previous section can be seen as an example, is much more robust to independence violations. This work is continuing.

3 Results are generally robust to distributions of attribute values, to types of preference structures and to imprecisions in input, which is encouraging for the use of additive value function methods.

The overall conclusion is thus that the use of additive value functions in the analysis of multi-criteria decision making problems give consistent and reliable results provided:

- non-linearities in the marginal value functions are adequately captured (by using interpolation between 3 or 4 points at least); and
- due care is taken in ensuring that the modelled criteria are close to additively independent.

# Chapter 6

# Use of "Soft" Problem Structuring Methods in Scenario Based Policy Planning Procedures

## 6.1 Introduction

Scenario based policy planning, as described in Chapter 2 and Appendix A, includes three key phases in which group judgements are crucial. These are (i) the identification of the "policy elements" out of which scenarios are to be constructed; (ii) the identification of relevant stakeholders, interests and affected parties; and (iii) the generation of criteria by which interest groups are to assess and to evaluate the policy scenarios. Little detail has been provided in the previous report (Stewart et al., 1993) or this, concerning the dynamics of these group processes themselves. In practice, much of what has been done has been based on the use of nominal group or brainstorming techniques (e.g. Delbecq et al., 1975), including electronic versions (group systems software) as mentioned in Section 4.3.

There has been a trend in the operations research/ management science fields of recent years to include what is sometimes termed "soft" systems or "soft" OR approaches, particularly at the phases of problem structuring. On *prima facie* grounds, it would seem that the advantages of these approaches should apply equally well to scenario based policy planning, and they do in fact appear to be especially relevant to the group evaluation phases as described in the previous paragraph. At this stage, we have not yet undertaken substantial research into the role of these problem structuring methods in SBPP, but in this Chapter we shall nevertheless outline three of these methods and the potential points in the SBPP process where they may usefully be applied. Details of the integration between SBPP and problem structuring methods will be addressed in follow-up research.

## 6.2 Strategic options development analysis (SODA)

The SODA approach, based on the concept of "cognitive mapping", has been developed by Eden (1986, 1989) and his co-workers. It uses the traditional model building and analysis skills of operational research, but in a non-mathematical manner, to assist groups involved with "messy" problems.

Cognitive Mapping is a technique designed to help structure, organise and analyse data and perceptions, and to produce a formal representation of how individuals or groups think about the issues or situations. A cognitive map is simply a network of ideas linked by arrows. In the first instance, the analyst interviews interested or affected individuals, and encourages them to express concerns, desires and action proposals in a relatively unstructured manner. The analyst then crystallizes the points raised into a number of concepts, usually in an antithetic form (e.g. natural river conditions *rather than* reduced flows), and starts to link them by arrows showing positive or negative influences (e.g. poorly managed forestry negatively influences natural river conditions). As the map with these concepts and the links between them evolves, it becomes possible to identify:

- which concepts represent policy actions (or policy elements), aims and objectives, or intermediate processes; and
- the existence of feedback loops, conflicts and dilemmas.

The process of setting up cognitive maps as described above is greatly facilitated by software called *Graphics COPE* which has been developed in conjunction with Eden's SODA (Ackerman et al., 1995). With this software, concepts are entered anywhere on the computer screen, and links inserted by "dragging" the mouse from one to the other. Menu options allow the user to zoom in on to subsets of concepts, and/or to explore all direct and indirect links (and hence also feedback loops) relating to a specific concept. Clusters within the map can be identified, each representing a problem arena within which there will be problem related goals (criteria in our terminology) at the 'head' of the cluster, strategic options within the cluster, and options at the 'tail' of the cluster. Each cluster will probably be linked to other clusters, in the sense that the goals of one problem lead to the options of another, and the options within one problem are consequences of the goals of a subordinate problem.

The procedure recommended by the developers of SODA involves a number of stages. Initially, members of each interested party would be interviewed in their own environment, aiming at about 45 min for each interview. From this will be constructed cognitive maps for each interviewee, which are discussed with each as soon as possible (a few days later?) after completion of the initial interviews, to ensure that their views have been faithfully captured.

The analyst would then merge the individual maps into an aggregate map termed a "strategic map". The aim is to use this as a facilitative device, to encourage some form of co-ownership of the ideas and perceptions. A workshop of about one day, and involving all players, is convened to work through the strategic map. This might initially be restricted to members of one particular interested group, after which the process may be repeated by merging maps of each group, and convening a workshop of representatives of all parties.

The timetable suggested in the standard SODA framework suggests that the experience of the developers may have been in the context of single organizational structures (in which the whole task can be carried out within days). It is not clear exactly how this can be carried over to contexts such as national or regional water resources planning, in which different parties may be widely dispersed geographically and culturally. This is clearly a topic for continuing research.

Links between SODA and the use of the Graphics COPE software on the one hand, and SBPP on the other, will need to be one of the aims of on-going action research. At the time of writing, some preliminary studies are underway in the UCT group on these links, but in the context of lobster resource management rather than water resources. The initial view is that SODA and Graphics COPE would be a useful additional tool at the early stages of a problem, when divergent thinking is to be encouraged. We have emphasized the need to ensure that the policy scenarios as initially formulated, should be inclusive enough to encourage all parties to remain part of the process. If that is not done, then there may be a strong tendency towards polarization of interests for and against the (limited range of) options on the table. Furthermore, all texts on decision analysis routinely emphasize the need to ensure that the set of criteria chosen for evaluation are "complete", i.e. not omitting important interests or considerations. The use of SODA, insofar as it achieves its aims of creating a shared perception and ownership of the problem, could well counter any sense of polarization, and ensure completeness. Certainly, it seems that the approach can facilitate the identification of:

- A rich set of *policy elements*, which in turn can stimulate the creative identification of policy scenarios;
- A complete set of criteria and interests; and
- A shared understanding of the impacts and consequences of policy actions, which would assist in the understanding of the subsequent evaluation of scenarios by different interests.

## 6.3 Soft systems methodology (SSM)

The soft systems methodology, see for example Checkland (1989), is (like SODA) aimed at obtaining a general overview of problem structures and interrelationships, as they are perceived by different actors. Use is also made of graphical devices (concepts and arrows) to represent these perceptions. The approach is, however, perhaps somewhat more structured than the SODA methodology, which may make SSM more suited to direct use for group work in decision conferences.

Checkland (1989) distinguishes SSM from what he terms "hard" systems engineering. He sees the latter as an approach to satisfying a well-defined need, and where the objectives of the system to be designed are givens. SSM, in contrast, is viewed as a learning system, or a means of enquiry. The methodology is aimed at creating descriptions of human activity in relation to different images of the world (or "*Weltanschauung*"), recognizing that in complex, human-related systems (which water resources planning is) there may be many conflicting perceptions, not only of purpose (goals), but even of the dynamics of the system.

In many senses, the aims of SSM and SODA are similar, both seeking to describe different world views. SODA, however, is a largely unstructured approach in which as much as hundreds of concepts, and many more linkages, may emerge, to get as complete a view as possible. SSM, on the other hand, imposes a more structured discipline on to the enquiry process, and seeks to limit the concepts examined at any one time to the "magical number 7 plus or minus 2" (Miller, 1956). Central to the structure of the SSM approach is the seeking of a succinct "root definition" to describe the system under consideration, its purpose and dynamics. It is suggested that the root definition give attention to six key elements relating to the "purposeful activities" of the system, identified by a mnemonic "CATWOE", defined as follows:

**Customer:** The victim or beneficiary of the activity;

**Actors:** Those who carry out the activity;

**Transformation process:** The transformations of inputs into outputs resulting from the activity;

**Weltanshauung:** The world view under which the definition is constructed;

**Owner:** The person or group able to stop the activity; and

**Environmental constraints:** Constraints on freedom of action accepted as given at the current stage of thinking.

Checkland (1989) suggests that a crisp (three or four line) definition of the system, in which purposeful activity has to be undertaken, be constructed from the above six elements. Thereafter, conceptual models relating the activities in the system are to be constructed, using the terms in the root definition. As in SODA, use is made of diagrams, in which the activities in the conceptual models are linked to each other by arrows representing influences. The ideal is to keep the models simple, consistent with the "seven plus or minus two" concept, although it is possible to create a hierarchy of conceptual models, in which an activity at one level can be decomposed into a full conceptual model at the next. The conceptual models can be used as the basis for constructing more formal models (e.g. systems dynamics models), but are also useful in their own right in simply understanding the system, and/or conveying perceptions of the system between parties.

A further useful concept emerging from the SSM approach is that of "monitoring and control", i.e. answering the question: "How could the system fail?". In this sense, Checkland suggests explicit consideration of the effectiveness ("Is this the right thing to do?"), efficacy ("Does it work?") and efficiency ("What resources are used?") of proposed activities.

Within the context of SBPP, the SSM approach holds potential as a means of creating group focus and shared perceptions, prior to moving to detailed implementation of the decision analysis phases. The concept of "ownership" may be difficult to define in the context of water resources planning (is it the people, their political representatives, or state structures such as the Department of Water Affairs and Forestry?); but the identification of customers and actors may facilitate identifying the interested and affected parties. The conceptual models would facilitate the construction of policy scenarios, while consideration of effectiveness, efficacy and efficiency will assist groups in establishing the criteria by which policy scenarios are to be compared. As with SODA, however, the links between SSM and SBPP need to explored further in continuing action research.

## 6.4 Strategic choice

This is a methodology developed by Friend (1987, 1989). While the previous two techniques were relevant particularly to the divergent, or idea generating and learning phases of policy formulation, the strategic choice approach of Friend is perhaps more oriented towards the analytical phase, and in particular to identification of operational criteria for assessment and of needs for further research in the construction of scenarios.

A fundamental concept within the strategic choice is that of the characterization of three broad categories of uncertainty which need to be identified by participants in the planning process. Each type of uncertainty calls for a correspondingly different type of response. These categories can be described as follows:

**Uncertainties pertaining to the working Environment (UE)** This relates to the context (environment) within which the consequences of policy actions will be played out. In principle, such uncertainty can be dealt with by responses of a relatively technical nature, by undertaking relevant surveys or research, and/or by development and running of relevant models (e.g. hydrological, forecasting or economic models).

**Uncertainties pertaining to guiding Values (UV)** This refers to questions as to what criteria or interests are relevant to the choices which have to be made, and the relative trade-offs between them. By focussing on the uncertainties in definition of the criteria, groups may be led to seek consensus between themselves by further facilitated consultation, and/or to seek policy guidance from a political authority, as to what issues or interests need to be taken into consideration, and the relative importance of each.

**Uncertainties pertaining to Related decision fields (UR)** This is the kind of uncertainty that calls for a response in the form of exploration of the structural relationships between the decision currently in view and others with which it appears to be interconnected. The result may be to expand the agenda of decision, to broaden the definition of the scope of the policy decisions and/or to introduce additional policy elements. The aim is ultimately to achieve, in the terminology of SBPP, a richer range of scenarios, to create a greater sense of ownership amongst interested parties.

By focussing in a structured manner on the existence of these three types of uncertainty, the first aim is to achieve a consensus and shared vision amongst participants that all three kinds of uncertainty are significant and should be vigorously addressed. Once the uncertainties are identified in this manner, it becomes easier to face the important practical choices as to how much to invest in responses to uncertainty (e.g. research, modelling, further group involvement).

It is interesting to record here the identification of four complementary modes of decision making activity which are defined in strategic choice (Friend, 1989). These correspond very well to the SBPP process as we have identified it, even though (it must be conceded) the SBPP process was developed in ignorance of the work of Friend. These four modes are as follows:

1 **The shaping mode:** When functioning in this mode, decision makers address concerns about structure of the set of decision problems which they face. They may debate what ways problems should be formulated, and how far one decision should be seen as linked to another. They may be consider whether their current focus should be enlarged or, conversely, whether related problems should be broken down into more manageable parts. This can be viewed as the *idea generation* phase.

2 **The design mode:** When functioning in this mode, the decision makers address concerns about what courses of action are possible in relation to their current view of the problem shape. Debates about the constraints of either a technical or policy nature surface which might restrict their ability to combine options for dealing with different parts of the problem in particular ways. This can be viewed as the *scenario development* phase.

3 **The comparing mode:** When functioning in this mode, the decision makers address concerns about the ways in which the consequences or other implications of different courses of action should be compered. They consider economic, social and other criteria, and debate in which ways assessments of consequences can be made. This is the mode in which uncertainties of the three types UE, UV and UR tend to come most clearly to the surface - though they can arise when working in any of the other modes as well. This can be viewed as the phase of *criteria identification* and *evaluation of consequences.*

4 **The choosing mode:** When functioning in this mode, the decision makers address concerns to do with incremental commitment to actions over time. From a strategic choice perspective, this means not only considering whether there are particular commitments to substantive action that could be undertaken straight away, but also thinking about ways in which the future process might be managed. This can be viewed as the *assessment* and *implementation* phases.

In contrast to SODA, which in the first instance involves evaluating the perceptions of individuals, the strategic choice methodology is in principle set up to work with groups. As with SODA, strategic choice is supported by commercially available software. Initial evaluation of the software, however, reveals rather complex and "busy" screens, and it is difficult to envisage using this in groups, especially when these groups involve widely divergent cultures and backgrounds. It appears, therefore, that the strategic choice methodology may contain elements which can be incorporated into SBPP, but the details will require further research. It seems, likely, furthermore, that appropriate software may still need to be developed for local implementation in water resources planning.

# Chapter 7

# Conclusions and Recommendations for Future Research

In this report, taken in conjunction with our previous Water Research Commission report (Stewart et al., 1993), we have developed the concept of scenario based policy planning (SBPP) as a means for decision support for water resources management (or, in fact, for any other natural resource management). This has been done by:

- Studying the implementation of SBPP in a number of different contexts;
- Linking SBPP to other management science and decision support methodologies such as cost benefit analysis (Chapter 3), environmental impact assessment (Chapters 3 and 4), and "soft" operational research methods (Chapter 6); and
- Developing and refining the SBPP procedures in the light of the previous two points.

The evidence quoted in the previous report, and the case studies described in the present report, present ample justification for the claim that the use of multicriteria decision analysis techniques together with the policy scenario framework which we have developed can facilitate policy decisions for water resource management in the following ways:

▷ Providing a means whereby different interested parties and role players can be involved in the policy making process at all stages;

▷ Providing a mechanism whereby different interests (tangible and intangible, qualitative and quantitative) can be compared and aggregated in a common and justifiable currency;

▷ Ensuring that each interest is taken into consideration in a transparent manner;

▷ Focussing on the human judgements that are critical to the decision making, while identifying directions of policy development which are clearly unsatisfactory or offer the most potential for compromise or consensus;

▷ Providing a mechanism whereby different interest groups can explore their own preferences, in order to be proactive but realistic in demands they make.

In short: We believe that the use of SBPP in water resources planning is effective efficacious and efficient. Furthermore, the procedure is adaptable to a variety of types of input from role players, and to different forms of problem setting (e.g. the creation of indexing or scoring systems, as well as for explicit policy formulation).

In order to realize the potential benefits of SBPP in full, two important thrusts will be necessary, namely *technology transfer* and *research and development*, as we now discuss briefly.

## Technology transfer

The primary need to be addressed is that for education and training in the techniques and tools described in this report. People are (often with justification) reticent to adopt what may be seen to be revolutionary new approaches, and may often feel threatened. In fact, SBPP should largely be seen as complementary to other approaches such as environmental impact assessments and benefit cost analyses. It is a means by which expertise from different areas (scientific, social and political) can be brought together, and the insights shared. Those involved with policy formulation and assessment (both those responsible for the final recommendations to political leadership, and those representing specific areas of expertise and/or interests) need to be re-assured that SBPP is a means to address their concerns and to ease their task.

An important need is therefore to design and to present a series of short familiarization courses, on the aims and mechanisms of scenario based policy planning, and more comprehensive training courses on using the approach and related software. Only in this way can the above concerns be addressed effectively.

In parallel with this education and training, there is also a more general need to make the approaches more widely known amongst stakeholders generally in the water community. This may be achieved by the production and publication of a non-technical information document. This needs to be done in the first instance by professional technical writers, backed up by the research team at UCT.

## Research and development needs

No procedure is ever complete: there is always room for refinement. In the light of the experience over the past few years, three broad areas of research have been identified, and the aim is to pursue all of these vigorously in a follow-up project. These areas of research are as follows.

1 **More "action research" in monitoring implementation of the procedures, and in taking the procedures to the communities:** The project team needs still to be involved in many more case studies, both to refine the procedures further and to develop meaningful training courses as indicated above. It has been disappointing that the team had so few opportunities for substantial involvement in real case studies. Discussions were held with those involved in a number of other water policy projects over the past years (in addition to those described in Chapter 4), but for a variety of reasons there was reticence to involve our group. To ensure the

full realization of the benefits of SBPP, this situation will have to change, and the team plans to continue the search for suitable projects for the action research, possibly using the training courses to establish the necessary contacts.

2 **Development of the use of "soft" problem structuring methods:** This has been discussed in Chapter 6. A number of these problem structuring methods have been reported in the literature, and have been widely used in some contexts, especially in the UK. It seems, however, that they do need some adaptation for use in conjunction with SBPP for water resources planning, especially where this involves widely divergent groups. Such methods hold the potential, in particular, for both defining scenarios and facilitating identification of criteria. The use of these approaches in conjunction with modern technologies such as electronic meeting rooms appears also to deserve further research.

3 **Refinement of software, and incorporation of "multi-media" approaches:** A clearly important area for research is the representation of policy scenarios in a manner which is accessible to the widest possible range of actors and affected parties. Participants need to be able to "see" and to "feel" the impacts of each scenario, and also to be able to change the scenarios on-line, to understand the effects of these changes. Only then can the thermometer scale and similar assessments be properly informed. Developments in the computer field, including the use of images and sound, and GIS systems, appear to hold considerable potential in this regard. Although the current project team will not easily become experts in this fields, they can and should seek co-operative research with other relevant groups, to incorporate their research into the practice of scenario based policy planning.

At the same time, effort needs to be placed on the development of a decision support system shell which can easily be incorporated into other software developments, to make the basic principles of SBPP available to other projects. Already, other software systems (such as GIS systems) often do include value assessment and aggregation "decision support" modules, but many of these appear naive, and/or to violate known principles of value measurement. There is therefore a need to provide alternatives.

# Bibliography

Abu-Taleb M F, Mareschal B (1995). Water resources planning in the Middle East: Application of the PROMETHEE V multicriteria method. *European Journal of Operational Research*, 81, 500–511.

Ackerman, F, Eden, C and Cropper, S (1995), *Getting started with Cognitive Mapping*, Graphics COPE package, Banxia Software, Glasgow.

Angelsen A, and Sumaila U R (1995). *Hard methods for soft policies: Environmental and social cost-benefit analysis.* Chr. Michelsen Institute for Development Studies and Human Rights, Bergen. WP 1.

Armstrong A J, van Hensbergen H J and Geertsma H (1994). Evaluation of afforestable montane grasslands for wildlife conservation in the North-Eastern Cape, South Africa. Part 1. Methods. *South African Forestry Journal*, 171, 7–20.

Arrow, K J (1951). Little's Critique of Welfare Economics. *American Economic Review*, 41, 923–934.

Bana e Costa, C A, Vansnick J-C (1994). A theoretical framework for Measuring Attractiveness by a Categorical Based Evaluation TecHnique (MACBETH). *Proceedings of the XIth International Conference on MCDA*, Coimbra, Portugal.

Bana e Costa, C A, Stewart, T J and Vansnick, J-C (1995). Multicriteria decision analysis: Some thoughts based on the tutorial and discussion sessions of the ESIGMA meetings. *Proceedings of the 14th European Conference on Operational Research*, Jerusalem, Israel (and to appear in the *European Journal of Operational Research*).

Barzilai, J, Cook, W D and Golany, B (1987) Consistent weights for judgements matrices of the relative importance of alternatives. *Operations Research Letters*, 6, 131–134

Barzilai, J and Golany, B (1994) AHP rank reversal, normalization and aggregation rules. *INFOR*, 32, 57–63

Beck, M P and Lin, B W (1983). Some heuristics for the consensus ranking problem. *Computers and Operations Research*, 10, 1–7.

Belton, V and Vickers, S (1990). Use of a simple multi-attribute value function incorporating visual interactive sensitivity analysis for multiple criteria decision making. In: Bana e Costa (Ed.), *Readings in Multiple Criteria Decision Aid*, Springer, Berlin.

Brown, S, Schreier, H, Thompson, W A and Vertinsky, I (1994). Linking Multiple accounts with GIS as decision support system to resolve forestry/wildlife conflicts. *Journal of Environmental Management*, 42, 349–364.

Central Economic Advisory Services (1989). Cost-Benefit Analysis in South Africa.

CERM (1995). *A co-ordinated research programme on decision support for conservation and management of estuaries.* Draft interim report of the Consortium for Estuarine Research and Management', PO Box 100767 Scottsville, 3209.

CERM (1996). *A co-ordinated research programme on decision support for conservation and management of estuaries,* Draft final report of the Consortium for Estuarine Research and Management', PO Box 100767 Scottsville, 3209.

Chang, N-B, Wen, C G and Wu, S L (1995). Optimal management of environmental and land resources in a reservoir watershed by multiobjective programming. *Journal of Environmental Management,* 44, 145–161.

Checkland, P (1989). Soft systems methodology. In: Rosenhead, J (Ed.) *Rational Analysis for a Problematic World,* Wiley & Sons., Chichester.

Cocks, D and Ive, J (1996). Mediation support for forest land allocation: The SIROMED system. *Environmental Management,* 20, 41–52.

Connigarth Consultants (1994). *Sabie/Sand River water resource development: Macroeconomic and cost-benefit analyses with reference to the proposed Injaka Dam.* Conningarth Consultants, Consulting Economists, PO Box. 109, Cedarville, 4720.

Cook, W D and Kress, M (1985). Ordinal rankings with intensity of preference. *Management Science,* 31, 191–198.

Cook, W D and Kress, M (1991). A multiple criteria decision model with ordinal preference data. *European Journal of Operational Research,* 54, 191–198.

Cowling, R M and Hilton-Taylor, C (1994). Patterns of plant diversity and endemism in southern Africa: an overview. *Strelitzia,* 1, 31–52.

Creemers, G (1996). Economic assessment of protected area: guidelines for their assessment. Preliminary comments on the draft report prepared by the IUCN task force on economic benefits of protected areas. In prep.

CSIR (1995). *Saldanha Steel Project, 1994. Final environmental impact report.* CSIR Environmental Services. CSIR report EMAS-C94017C.

Cyert, R M and March J E (1963). *A behavioural theory of the firm.* Prentice Hall, Englewood Cliffs, N. J.

Daly, H E (1976). The steady-state economy. In: Wilson, K D (Ed.) *Prospects for growth.* Praeger, New York, pp 263–281.

De Graaff, J and Kuyvenhoven, A (1996). Economic evaluation of watershed development. *Inaugural conference of the European branch of the International Society for Ecological Economics,* 23–25 May, 1996. Paris.

Delbecq, A L, van de Ven, A H and Gustafson, D H (1975). *Group Techniques for Program Planning,* Scott, Foresman & Company, Glenview, Illinois.

Dubourg, W R and Pearce D W (1994). Paradigms for environmental choice: sustainability versus optimality. *Proceedings of the International symposium on Models of Sustainable Development. Exclusive or Complementary Approaches of Sustainability?*

Eden, C (1985). Perish the Thought, *Journal of the Operational Research Society*, 36, 809–819

Eden, C (1986). Problem solving or problem finishing?, In: Jackson M and Keys, P (Eds) *New Directions in Management Science*, Gower, Aldershot.

Eden, C (1989). Using Cognitive Mapping for Strategic Options Development Analysis (SODA), In: Rosenhead, J (Ed.) *Rational Analysis for a Problematic World*, Wiley & Sons., Chichester.

Faith, D P and Walker, P A (1996). Integrating conservation and development: Effective trade-offs between biodiversity and cost in the selection of protected areas. *Biodiversity and Conservation*, 5, 431–446.

Faucheux, S and Froger, G (1995). Decision-making under environmental uncertainty. *Ecological Economics*, 15, 29–42.

Fijen, A P M and Kapp, J F (1995). *Wilderness, Swartvlei and Groenvlei lakes catchment, water management strategy. Proposed water management strategy, objectives and goals.* Department of Water Affairs and Forestry, Pretoria (Draft Report)

Findlay, K (1996). The value of whale watching to coastal communities: The Hermanus case study. Paper presented at the South African Wildlife Management Association Conference, University of Cape Town.

Forsyth, G G, Versfeld D B, Chapman, R A and Fowles, B K (1996). *The hydrological implications of afforestation in the North-Eastern Cape.* Water, Environment and Forestry Technology, Council for Scientific and Industrial Research, Stellenbosch.

French, S (undated manuscript). *Multi-attribute value analysis and decision conferencing.* Division of Operational Research and Information Systems, School of Computer Studies, University of Leeds.

Friend, J K and Hickling, A (1987) *Planning Under Pressure: the Strategic Choice Approach*, Pergamon, Oxford.

Friend, J K (1989). The Strategic Choice Approach. In: Rosenhead, J (Ed.) *Rational Analysis for a Problematic World*, Wiley & Sons., Chichester.

Georgescu-Roegen, N (1971). *The entropy Law and the Economics Process.* Cambridge, Mass., Harvard University Press.

Graaff, J de V (1971). *Theoretical Welfare Economics.* Cambridge University Press, Cambridge.

Gregory, R, Keeney, R and von Winterfeldt, D (1992). Adapting the environmental impact statement process to inform decisionmakers. *Journal of Policy Analysis and Management*, 11, 58–75.

Gregory, R and Keeney, R (1994) Creating policy alternatives using stakeholder values. *Management Science*, 40, 1035–1048

Heynes, W G (1995). *Selection of multicriteria decision making methodologies in scenario based planning.* MSc dissertation, Department of Statistical Sciences, University of Cape Town.

Hicks (1939). The foundations of welfare economics. *Economic Journal*, 49, 696–712.

Hobbs, B F, Chankong, V, Hamadeh, W and Stakhiv, E Z (1992). Does choice of multicriteria method matter? An experiment in water resources planning. *Water Resources Research*, 28, 1767–1779.

Keeney, R L and Raiffa, H (1993). *Decisions with multiple objectives: Preferences and value trade-offs.* Cambridge University Press.

King and Tharme (1994). *Assessment of the instream flow incremental methodology and initial development of alternative instream flow methodologies for South Africa.* WRC Report No 295/1/94.

Larichev, O I, Moshkovich, H M (1995). ZAPROS-LM - A method and system for ordering multiattribute alternatives. *European Journal of Operational Research*, 82, 503–521.

Leiman, A and van Horen, C (1996). Proposed changes to Saldanha Bay oil transfer operations: Cost-benefit Analysis. Unpubl. Report, 46 p.

Leonard, H B and Zeckhauser, R J (1986). Cost-benefit analysis applied to risks: its philosophy and legitimacy. In: Maclean, D (Ed.) *Values at risk.* Rowman & Allanheld Publishers, USA.

Lewis, H S and Butler, T W (1993). An interactive framework for multi-person, multi-objective decisions. *Decision Sciences*, 24, 1–22.

Lipsey, R G and Lancaster, K (1956). The general theory of second best. *Review of Economic Studies*, 24, 11-32.

Loomes, G and Sugden, R (1982). Regret theory: an alternative theory of rational choice under uncertainty. *Economics Journal*, 92, 805-824.

Lootsma, F A (1989). Conflict resolution via pairwise comparison of concessions. *European Journal of Operational Research*, 40, 109–116.

Lootsma, F A (1993). Scale sensitivity in the multiplicative AHP and SMART. *Journal of Multi-criteria Decision Analysis*, 2, 87–110.

Maass, A (1966) Benefit-cost analysis: its relevance to public investment decisions. *Quarterly Journal of Economics*, 208–226

Maguire, L A and Boiney, L G (1994). Resolving environmental disputes: a framework incorporating decision analysis and dispute resolution techniques. *Journal of Environmental Management* 42, 31–48.

Martunnen, M and Hämäläinen, R P (1995). Decision analysis in environmental impact assessment. *European Journal of Operational Research*, 87, 551–563.

McDaniels, T L (1995). Using judgement in resource management: A multiple objective analysis of a fisheries management decision. *Operations Research*, 43, 415–426.

McKay, H (1994) *A prototype decision support system for the Kruger National Park Rivers Research Programme.* Research Report, Watertek, CSIR.

Miller, G A (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 63, 81–97.

Miltz, D, Mirrilees, R I, Williams, C J and Willekens, M (1992). *Eastern shores of Lake St. Lucia: Kingsa / Tojan lease area.* Environmental Impact Assessment: Economic Valuation.

Mishan, E J (1969). *The costs of economic growth.* Penguin, Harmondsworth.

Mousseau, V (1993) *Problémes liés à l'èvaluation de l'importance relative des critères en aide multicritère à la décision: réflexions théoriques, expérimentations et implémentations informatiques.* Thèse de Doctorat, Université Paris-Dauphine

Munasinghe, M (1993). Environmental economics and sustainable development. *World Bank environmental paper number 3.* The World Bank, Washington D.C.

Munda, G, Nijkamp, P and Rietveld, P (1994a). Monetary and non-monetary evaluation methods in sustainable development planning. In: *Proceedings of the International symposium: Models of sustainable development. Exclusive or complementary approaches of sustainability?*, 2, 887–898.

Munda, G, Nijkamp, P and Rietveld, P (1994b). Qualitative multicriteria evaluation for environmental management. *Ecological Economics*, 10, 97–112.

Nash, J F (1950). The bargaining problem. *Econometrica*, 18, 155–162

Nash, J F (1951). Non-cooperative games. *Annals of Mathematics*, 54, 286-295

Offringa, G and de Wet, G (1996). A quality of life-based decision support model for the determination of water research priorities. *Water SA*, 22, 359–372.

O'Keeffe, J (1995). Southern hemisphere methods of assessing conservation status for natural freshwaters. Paper delivered at the conference *Freshwater quality: defining the indefinable?*, University of Stirling, Sep 6–7.

Olson, D L, Fliedner, G and Currie, K (1995). Comparison of the REMBRANDT system with analytic hierarchy process. *European Journal of Operational Research*, 82, 522–539.

Pearce, D W (1983). *Cost-Benefit analysis* (2nd edition). Macmillan, London.

Perrings, C (1994). Biotic diversity, sustainable development, and natural capital. In: Jansson, Hammer, Folke and Costanza (Eds.) *Investing in Natural Capital.* Island Press, Washington.

Ralls, K and Starfield, A (1995). Choosing a management strategy: Two structured decision-making models for evaluating the predictions of stochastic simulation models. *Conservation Biology*, 9, 175–181.

Rosenhead, J (1989). *Rational analysis in a problematic world.* John Wiley and Sons, New York, NY, USA.

Saaty, T L (1980). *The analytic hierarchy process.* McGraw-Hill, New York.

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

Salo, A A and Hämäläinen, R P (1992). Preference assessment by imprecise ratio statements. *Operations Research*, 40, 1053–1061.

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

Schulze, P C (1994). Cost-benefit analyses and environmental policies. *Ecological Economics*, 9, 197–199.

Scitovsky, T de (1941). A note on welfare propositions in Economics. *Review of Economic Studies*, 9, 77-88.

Sen, A K (1979). Personal utilities and public judgements: or what's wrong with welfare economics?. *Economic Journal*, 89, p 537.

Simon, H A (1969). Theories of decision making in economics. *American Economic Review*, 57, 1–33.

Simon, H A (1976) *Administrative Behavior*, 3rd edn., The Free Press, New York

Smith, P G R and Theberge, J B (1987). Evaluating natural areas using multiple criteria: Theory and practice. *Environmental Management*, 11, 447–460.

Smith, P N (1994). Applications of fuzzy sets in the environmental evaluation of projects. *Journal of Environmental Management*, 42, 365–388.

Stansbury, J, Woldt, W, Bogardi, I and Bleed, A (1991). Decision support system for water transfer evaluation. *Water Resources Research*, 27, 443–451.

Stewart, T J (1992). A critical survey on the status of multiple criteria decision making theory and practice. *OMEGA The International Journal of Management Science*, 20, 569–586

Stewart, T J, Scott, L and Iloni, K (1993). *Scenario Based Multicriteria policy planning for water management in South Africa.* Water Research Commission Report No 296/1/93.

Stewart, T J (1994). Scenario analysis and multicriteria decision making. In: *Proceedings of the 11th International Conference on Multicriteria Decision Making*, Coimbra, Portugal.

Stewart, T J (1995). Simplified approaches for multicriteria decision making under uncertainty. *Journal of Multi-Criteria Decision Analysis*, 4, 246–258.

Stewart, T J and Scott, L (1995). A scenario-based framework for multicriteria decision analysis in water resources planning. *Water Resources Research*, 31, 2835–2843.

Stewart, T J (1996). Robustness of additive value functions in MCDM. *Journal of Multi-Criteria Decision Analysis*, 5, 301–309.

Tecle, A (1992). Selecting a multicriterion decision making technique for watershed resources management. *Water Resources Bulletin*, 28, 129–140.

Turpie, J K (1995). Prioritizing South African estuaries for conservation: A practical example using waterbirds. *Biological Conservation*, 4, 35–37.

Turpie, J K (1996). A preliminary economic assessment of de Hoop Nature Reserve. Unpubl. Report: 93 p.

Turpie, J K and Siegfried, W R (1996). The conservation-economic imperative: Securing the future of protected areas in South Africa. *African Environment and Wildlife* 74, 175–185.

van Huylenbroeck, G (1995). The conflict analysis method: Bridging the gap between ELECTRE, PROMETHEE and ORESTE. *European Journal of Operational Research*, 82, 490–502.

van Pelt, M J F (1993). Ecologically sustainable development project appraisal in developing countries. *Ecological Economics*, 7, 19–42.

van Praag, B M S (1978). The perception of income inequality. In: Shorrocks, A F and Krelle, W (Eds.), *Personal Income Distribution*, North Holland, Amsterdam.

van Praag, B M S, Hagenaars, A J M and van Weered, H (1982). Poverty in Europe. *Review of Income and Wealth*, 28, 345–359.

Varis, O, Kuikka, S, and Taskinen, A (1994). Modelling for water quality decisions: Uncertainty and subjectivity in information, in objectives and in model structure. *Ecological Modelling*, 74, 91–101.

Vatn, A and Bromley, D W (1994). Choices without prices without apologies. *Journal of Environmental Economics and Management*, 26, 129–148.

Ventana Corporation (1994). *Group Systems for Windows: Quick Reference*. Ventana Corporation, Tucson, Arizona

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

Walters, C (1986). *Adaptive management of renewable resources*. Macmillan Publishing Company, New York.

Xiang, W-N, Gross, M, Fabos, J Gy and MacDougall, E B (1992). A fuzzy-group multicriteria decisionmaking model and its application to land-use planning. *Environment and Planning B: Planning and Design*, 19, 61–84.

# Appendix A

# Manual for Implementing Scenario Based Policy Planning in Decision Conferences and Workshops

The basic concepts in scenario-based policy planning (SBPP) have been described in Chapter 2. This Appendix is provided as a manual for implementing SBPP in decision conferences or workshops, In it, we describe the implementation of SBPP in terms of three key phases, *viz.*: establishing the policy scenarios to be evaluated; running workshops with groups representing different interests, in order to establish the value scales; and identifying alternative policy directions which have the potential to be broadly acceptable.

## A.1 Before the Workshop: Defining Policy Scenarios

In the SBPP workshops, different interest groups will express their values and objectives by ordering specific policy alternatives along a number of preference axes, as described in the next Section. We re-emphasize that these alternatives will not in general describe the policy options in absolute detail, but will be expressed as *policy scenarios* in which requisite detail is given to allow the ordering to be done meaningfully, even if not with absolute precision. In Section A.4 we discuss the iterative nature of the process, in which increasing levels of detail will be sought as the range of alternatives converges to some sort of consensus, but for now we will focus on a single pass through process.

The set of alternative policy scenarios needs therefore to be identified *a priori*, and described in the level of detail requisite to the current considerations. This description will need to include both the key defining characteristics of each scenario (cf. comment on "policy elements" below), and the implications of each alternative in terms which are both meaningful to participants in the process and more-or-less objectively verifiable in the light of available information (i.e. these implications should not include the value judgements which are the prerogative of the interest groups themselves). The assessment of these "objective" implications will typically require *inter alia* the use of a variety of systems and models and/or public scoping.

A number of points do need to be borne in mind when selecting the policy scenarios for consideration, as follows.

1 The number of scenarios under consideration must be manageably small, i.e. participants must be able to keep all in mind at the same time, suggesting (per Miller, 1956) approximately $7 \pm 2$ scenarios. This requirement will conflict to some extent with the next, and some systematic procedures for selecting a set of scenarios satisfying both requirements are discussed by Stewart et al. (1993).

2 It is important (at early stages of evaluation at least) to maintain as wide a coverage of options as possible. If some interest groups feel that there is nothing for them in the options available, they are likely to become disenchanted with the process. One means of ensuring adequate coverage is to start with a list of all *policy elements*, i.e. the various components of management policy which might be applied in the context under consideration. For example, in the management of a river basin, these elements might include land use restrictions, abstraction restrictions, construction of reservoirs, interbasin transfers and the relocation of certain activities. By examining the various combinations of options for each element, there is less danger of overlooking potential policy directions. Once again, details of the process of evaluating combinations can be found in Stewart et al. (1993).

   Of course, as the process of evaluation and consultation proceeds, it is inevitable that certain policy directions (perhaps favoured by certain interest groups) will eventually have to be dropped from consideration if found to be unacceptable. This will lead to a contraction of the coverage, as convergence to a broadly acceptable compromise is achieved. As this happens, the remaining scenarios will increasingly be refined, with various sub-options emerging, all with greater precision of detail.

3 In spite of best efforts to meet the preceding requirement, it may still be true that some interest groups find little to choose between the alternatives (which may be from their perspective either all equally good or equally bad). It may be useful therefore to include two hypothetical scenarios as benchmarks, to represent respectively an ideal that might realistically be hoped for and a worst fears outcome for each interest group. To make this meaningful, each group would need to specify what in fact constitutes these ideal and worst cases respectively. The actual policy scenarios would then be compared not only against each other, but also against these two hypothetical benchmarks. There are pros and cons in introducing such hypothetical scenarios, and research will be continuing to provide better guidelines as to when this approach should or should not be employed.

# A.2 Assessing Criteria and Values for an Interest Group

## A.2.1 Preliminaries and Identification of Criteria

Suppose now that a set of policy scenarios have been defined as in Section A.1, and that descriptions of each to a requisite level of detail (including outputs from relevant systems models) have been made available. The next step is to involve representatives of each significant interest group in decision workshops, to provide some evaluation of the scenarios (in the sense to be described below). Initial workshops may involve relatively homogeneous interests (e.g. a "conservation" interest), but at some stage representatives of quite divergent groups may need to be brought together in a single decision conference.

Each workshop should preferably involve a group of between 3 and 8 participants. Once the group becomes too large, there is a tendency for a smaller sub-set of the group to dominate proceedings, and it may be more effective to divide into smaller workshops with a report-back session later.

It is essential to ensure that participants do have an understanding of the policy scenarios and their implications. For more sophisticated groups, they should be prepared to "do their homework", *i.e.* to study the documentation available before the workshop. For less sophisticated groups, it may be necessary to spend as much as a day or more taking them through the options, using as much visual aid as is feasible, for example using the power of GIS etc. The evaluation phase can only start once this has been done.

The first step towards evaluating the scenarios is to identify the criteria of concern to the group. It is useful to approach this step in a variety of ways. One is simply to spend a period of free brainstorming, aimed at getting issues of concern out on the table. Another is to ask what features might favour one scenario over another, or what characteristics might differentiate one pair of scenarios from another scenario. It is important at this stage to obtain as complete as possible a picture of what the concerns and values of the group are. Once a reasonably complete set of criteria are identified in this way, the facilitator should try to group these into 3–7 primary but relatively independent points of view. For example, in one workshop involving the forestry industry, the three criteria which emerged were company profits, company image and regional economic development. A very useful discussion of the processes used for eliciting criteria from amongst divergent interest groups (in the context of the development of mining activities in Malaysia) is given by Gregory and Keeney (1994).

The group should then take some time to describe their perceptions of (a) a realistically best achievable outcome from the points of view of each criterion in turn, and (b) realistic worst fears in terms of each criterion. These constitute the ideal and worst-case benchmark scenarios, as discussed in the previous Section, and may or may not be one of the actual policy scenarios under consideration. We have emphasized the need for *realism* in this context, as unrealistically extreme benchmarks will have the effect of making all real alternatives virtually indistinguishable from one another (on the scales to be introduced in the next sub-section).

### A.2.2 Scoring of Scenarios on each Criterion

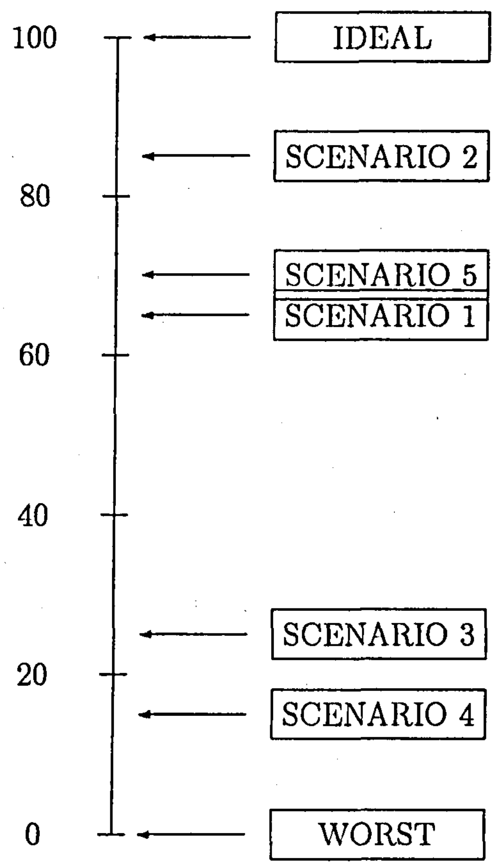
At this stage, the group is invited to arrange the policy scenarios (the actual scenarios plus the ideal and worst-case scenarios) from the points of view of each criterion taken in turn, along a "thermometer scale" with the ideal scenario at the top (score of 100) and the worst-case scenario at the bottom (score of 0). The other scenarios are placed in between, in such a way that the gaps between the alternatives represent the perceived relative magnitudes of the differences between them in terms of the criterion under consideration. In other words, if the gap between two scenarios A and B on one particular criterion is given as 10 units (with A preferred to B) and that between B and C is given as 20 units (with B preferred to C), then the implication is that the gain (in terms of this particular criterion) in moving from scenario C to scenario B is twice that of moving from B to A. This evaluation can be done in one or both of two ways, as described in the following paragraphs.

The first possibility, which is in our experience the approach most frequently found to be useful, is simply and directly to place the scenarios along a thermometer scale, as illustrated in Figure A.1. In this figure, five scenarios are being compared in terms of a single criterion, and "ideal" and "worst" cases are included for comparison. The scenarios indicated as SCEN3 and SCEN4 are viewed as being particularly poor options in terms of this criterion. There is then a substantial "gap", with the other 3 scenarios being considerably more acceptable: the gain in being able to replace scenario SCEN3 by SCEN1 is close to double the perceived gain in moving from the worst conceivable situation to SCEN3. It is further noted that SCEN5, although relatively good in comparison with SCEN3 and SCEN4, is still as much worse than SCEN2, as SCEN2 is relative to the ideal.

It may be helpful to let the group experiment on some hypothetical problem (e.g. rating local restaurants by quality of service or ambience), in order to gain some familiarity with the use of the thermometer scale, before progressing to the real assessments. Groups do seem to be able to fit into this mode of thinking quite quickly, however.

The alternative to direct assessment on a thermometer scale is appropriate when the criterion of interest is clearly and unambiguously related to a single measurable attribute of the scenarios (e.g. minimum flow levels in a river). There is often a danger at this point, in that participants may feel that the only objective approach is to use the attribute values directly as the scores on the thermometer scale (after re-standardizing to the 0–100 interval). There is no rational or scientific reason for such an assumption, and in fact the simulations described in §5.4.2 of the main report reveal that such assumptions can be one of the major reasons for the failure of quantitative decision analysis. In this case a device called "mid-value" splitting can be used. Participants are asked to successively split intervals on the natural attribute scale, in such a way that the benefits of moving from the lower to the upper points of each interval are adjudged to be of equal value. This is illustrated in Figure A.2. Initially, the group would have agreed that the increment in attribute value from the origin to that for point 2, and the increment from this level to the maximum on the attribute scale, were of equal worth. For this reason, the value on the vertical scale for this mid-value point in the range of attributes is given as 50. The process was repeated

Figure A.1: Illustration of the thermometer scale approach



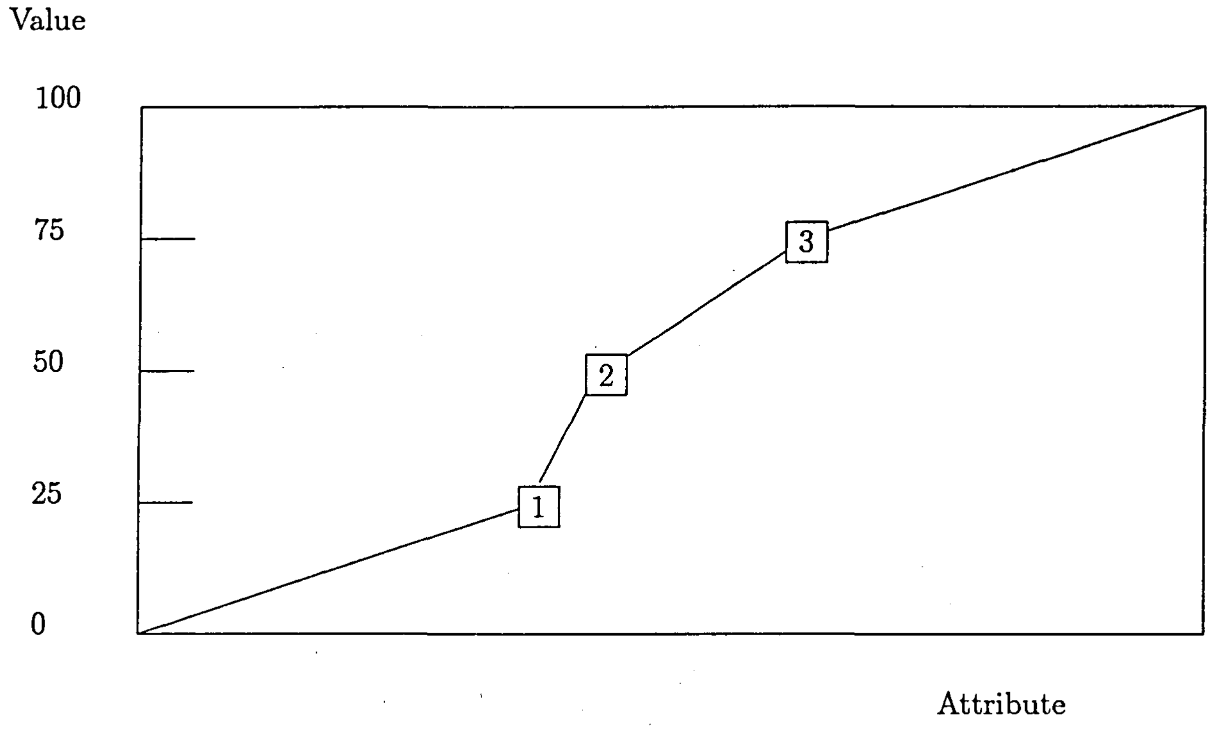



Figure A.2: Illustration of mid-value splitting

for the two sub-intervals thus identified, to give the attribute values corresponding to 25 and 75 respectively (points 1 and 3). One could draw in a smooth curve for interpolation purposes, but the piecewise linear interpolation has been shown in the simulations of §5.4.2 to be entirely adequate.

The construction of these scales in an interactive graphical manner is facilitated by the use of suitable software such as VISA ("Visual Interactive Sensitivity Analysis", Belton and Vickers, 1990) or SDAW which is described in the next Appendix. VISA allows the construction of thermometer scales directly, or indirectly by mid-vale splitting. SDAW, at this stage, only provides the direct thermometer scale option, and mid-value splitting has to be done off-line.

There will always be some difference of group opinion as to precisely where each scenario should be located on the thermometer scale, but within reason this is not a problem. Results from the full analysis (still to be described below) tend to be relatively robust to the precise scores, as long as the ordering is well-established, and it is in any case relatively simple to experiment with changes to the scores, so as to evaluate their impact. (The VISA software provides a particularly convenient means of achieving such sensitivity studies.) It can nevertheless happen that a major divergence of opinion in the group emerges. This may be as a result of incompletely scenarios (indicating a need for further research), or could suggest that the criterion is still multi-dimensional in some sense, i.e. contains multiple points of view. By questioning different participants as to why they prefer certain alternatives over others, the facilitator needs to identify these points of view. Once identified, these can be taken into account in one of two ways:

1 Replace the criterion by two or more other criteria which more accurately reflect the points of view. This is the appropriate response if it becomes clear that the original criterion was confounding two or more fundamentally different points of view. For example, "human needs" may have been initially chosen as a criterion, but it might later transpire that there is a very fundamental divergence between the "needs" of rural populations and those of urban dwellers.

2 Hierarchically subdivide the criterion into two or more sub-criteria. This is appropriate if it appears that the criterion itself is still well-defined, but that it includes certain internal conflicts which still need to be resolved. For example, a criterion of environmental quality might need to be sub-divided into water quality, air quality and aesthetics, all of which contribute to the environment, but may be differently affected by each of the policy scenarios. In this case, what has been done in effect is to view the criteria for the moment as sub-interests, each of which need to be evaluated in the above manner by identification of component criteria, etc.

There is often no clear reason to select one or other of the above options. The facilitator should, however, attempt to restrict the number of criteria at one hierarchical level to about 7; if criteria are proliferating, it would be better to re-group them, and to view the criteria in one group as sub-criteria of one broad consideration. For example, one might group all economic criteria into a single "economic benefit" criterion, but with different aspects (e.g. capital costs, running costs, regional development) viewed as sub-criteria.

### A.2.3 Obtaining an Aggregate Group Ranking of Scenarios

The output from the previous Section will be a set of "thermometer scales", one for each criterion which had been identified by the group. The final step is to aggregate these into a single scoring of the alternative scenarios which represents the considered preferences of the interest group. Note that if one or more of the criteria have been sub-divided into sub-criteria, then it is first necessary to aggregate these into a single scoring for the parent criterion. The procedures are, however, precisely the same as those for aggregating the criteria into an overall interest score, and we will discuss these procedure in this latter context.

Initially, it is useful simply to display the component thermometer scales for each criterion side-by-side. An aid to visualization is to join up the points representing each alternative policy scenario, to yield what has been termed a *value path.* This is illustrated in Figure A.3 for a hypothetical case in which the interests of National Parks might have been decomposed into four criteria, *viz.* effects on ecotourism, water quality, protection afforded to endangered species, and maintenance of biodiversity. Note for example that the alternative labelled "SCEN1" is clearly a broadly unacceptable option, while "SCEN4" and "SCEN5" are potentially good compromises as they are never very bad in terms of any criterion.

A valuable tool at this stage is to construct a weighted average, or *additive aggregation*, of the individual thermometer scales. As an example, consider the SCEN3

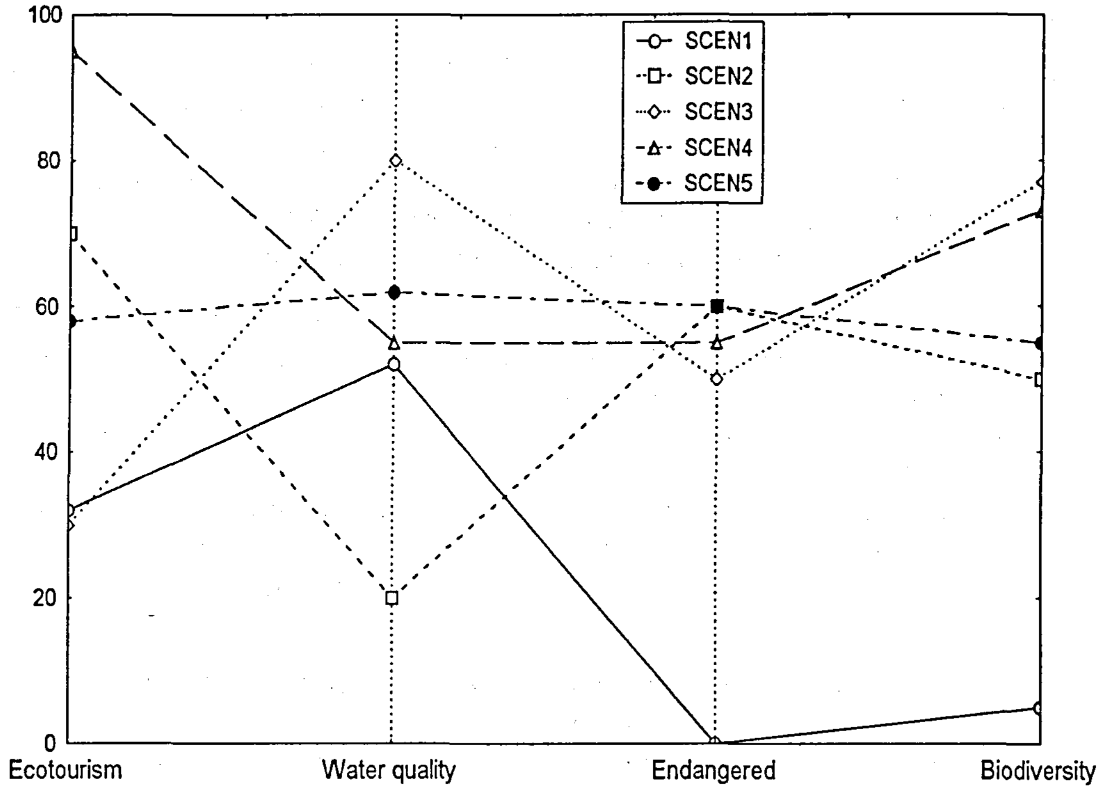



Figure A.3: Example of "Value Path"

alternative in Figure A.3, which has scores of 30, 80, 53 and 75 respectively on the four criteria. If relative importance weights were associated with the criteria in the ratios 60:80:60:100, the average score for SCEN3 would be $(60 \times 30 + 80 \times 80 + 60 \times 53 + 100 \times 75) \div 300 = 62.9$. Once these calculations have been carried out for every alternative scenario, this gives a set of aggregate scores for each scenario (which can, if desired, be re-standardized to a 0–100 scale).

Some care must, however, be exercised in assessing the weights to be used in this aggregation step, as in a mathematical sense the weights represent relative trade-offs between the scores for the different criteria, on the scale of measurement represented by the value functions or thermometer scales. When asked to assess weights directly, people quite generally appear not to take cognizance of the actual ranges of possibilities available, i.e. they do not naturally seem to think in trade-off terms (e.g. Mousseau, 1993). The elicitation of weights needs to compensate for this. In some cases, after explanations, groups may be able to express trade-offs directly. Frequently, however, it is useful to use an approach called "swing weighting". Participants are asked first to consider a hypothetical scenario in which all criteria are at some relatively poor level, for example all at their worst levels (i.e. 0 on every scale). They are then asked which criterion they would most like to switch to its best level (i.e. 100 on the scale), if this can be done for one and only one criterion. If (say) criterion $i$ is identified as the most desirable "swing", then it must have the largest weight, which can be set to some arbitrary value (e.g. 100). Finally, and for each other criterion, say $k$, participants may be asked either to assess directly the relative worth of switching $k$ to its best level rather than $i$, or to state how much they would be willing to give up again on $i$ (as measured on the thermometer scale), in order to switch $k$ to its best level. In this latter case, if the reponse is that $i$ can be dropped by $x$ units below the ideal 100 on the thermometer scale, then the ratio of weights, say $w_k/w_i$ must be $x/100$ (which is of necessity and by definition less than 1). A variation of this approach is illustrated in Section 5.3 in the context of the "conjoint scaling" described there.

The above aggregation calculations are done automatically by software such as VISA or SDAW, which also provide simple facilities for varying the weights interactively. Once the calculations have been done for all alternative scenarios, we have a suggested composite scale on which the overall interests of the group can be represented. This will, of course, be dependent upon the precise relative weights attributed to each criterion, which is why the sensitivity analyses provided by VISA and SDAW are important. Surprisingly, perhaps, the rank ordering of the alternatives on the composite scale tends to be quite robust to choice of weights in most instances, however.

## A.3 After the Workshop: Potential Compromises and Consensus Scenarios

The workshops for each interested group thus generate thermometer scales for each group, which provide a direct and visual comparison of the impacts of the different policy options on each group, especially if a "value path" such as illustrated in Figure A.3 is used (with interest groups such as conservation, forestry industry and rural communities replacing the criteria). This form of display may be of use either to planners or consultants in proposing compromise solutions, or in public forums in which representatives of different groups might attempt to reach a consensus directly.

Certain additional aids can assist in identifying potential compromise solutions. The use of weighted averages of the group thermometer scales, coupled to sensitivity analysis, as described in the previous Section, can again be used, except that this time the weights are to be placed on the different interest groups themselves. While this may not always be a politically acceptable thing to do, it may be valuable to establish, for example, that certain alternatives can never be optimal, no matter what weights are used, or may be optimal only if extremely lop-sided weights are used. Thus the use of weighted averages of scores may still provide valuable insights, even if it may be difficult to associate precise weights with each group. Once again the facilities provided by SDAW and VISA for interactively varying the weights is particularly useful.

Another simple aid to discussion is to identify the "max-min" solution, *i.e.* the alternative which has the largest minimum score taken across all interests (cf. §5.4.1). For example, if in Figure A.3, the scales referred to different interest groups rather than to specific criteria, then the alternatives labelled SCEN4 and SCEN5 are approximately equally max-min solutions, as neither score less than 55 by any interest.

An alternative use of software such as SDAW or VISA is suggested in the paper by Gregory and Keeney (1994). Instead of using the interest group thermometer scales directly, it may be useful to analyse the criteria identified by the different groups. In some cases, essentially the same criteria may appear in the evaluations of different groups. It may then be useful to construct a list of all criteria which have been deemed to be relevant to the current policy decisions, and to re-assess (using relevant experts) the alternatives according to each of these criteria. It would be hoped that these re-assessments bear some resemblance to the assessments done within the groups on the same criteria. (If there is major divergence, this would indicate very different *perceptions* of impacts, *i.e.* not only differences in values but differences in belief, which may need to be resolved separately, possibly by additional research.) These criteria may be aggregated in much the same way as we have previously discussed.

## A.4 Postscript: On-going Iterations

It is important to realize that the procedures described above are but *one iteration* in what will always be an on-going process. The set of alternative policy scenarios is never complete; other variations can always be synthesized or entirely new alternatives created. The results of one workshop, run according to the guidelines above, will indicate the extent to which consensus can be reached in terms of the options currently under consideration. If a clear "winner" emerges, which is broadly acceptable to all parties, then this is the end of the story. More commonly, however, it may be more helpful to seek not a single winner, but a shortlist of 2 or 3 options worthy of further consideration. In this way clearly *unacceptable* alternatives have been eliminated, and further attention can be focussed on the apparently more desirable options and variations thereof. Difficulties of choosing between the remaining alternatives may be due either to substantial inter-group conflicts which remain, or to the possibility that distinguishing between these options depends on details of implementation which have yet to be specified.

Where the result is a shortlist rather than a single winner emerges, the results of the workshops can also assist in creating or defining new alternative policy scenarios. In cases where none of the retained options are satisfactory to all interest groups, it may be possible to suggest some "mixing and matching" of options in the shortlist, to create alternatives that have the potential to better meet the aspirations of all groups (*i.e.* to reduce inter-group conflict). Where the problem is one of adequately distinguishing between alternatives, the records of the workshops may reveal those uncertainties which existed in the minds of participants, and which hindered precise evaluation. On the basis of such considerations, a new set of policy scenarios may be proposed, centred around the shortlist emanating from the workshops, but including additional variants and/or refinements of detail. The entire process can then be repeated, and indeed needs to be repeated, until such time as an acceptable alternative emerges.

# Appendix B

# User Manual for SDAW Software

In the text, reference has been made to the commercially available software, marketed under the name VISA (Visual Interactive Sensitivity Analysis), which is available (at time of writing) in a Windows 3.1 version. Details are obtainable from Visual Thinking International Ltd., 141 St James Rd., Glasgow G4 0LT, Scotland (Fax: +44-141-552-3886). SDAW is a simple alternative, operating under DOS, and which has been developed for use in the Water Research Commission project at UCT. This software is available *without guarantees* but also without charge to others in the South African Water Community. This Appendix is a brief description of the use of this software.

The program is supplied on a diskette containing the DOS program SDAW.EXE plus an illustrative file EXAMPLE.DWS which can be accessed by SDAW (see file options below).

The intention of SDAW is to facilitate the process of selection of one from a number of alternative courses of action (or policy, or action, scenarios), in such a way that a broadly acceptable compromise between different goals and/or interests is achieved. It is possible that some of the "alternatives" may be hypothetical ideals, introduced as benchmarks even if not in fact feasible courses of action. The approach adopted is based on hierarchical decision analysis. The apex of the hierarchy is defined to be "overall benefit" - to society in some or other sense. This is broken down into "criteria" (or perhaps sub-interests), each of which constitutes an "interest" in its own right at the next level of the hierarchy, and may therefore in turn be subdivided into further criteria. An operational basis for deciding whether or not to sub-divide an interest further is quite simple: If those involved in assessing the alternatives from the point of view of the interest under consideration are able more-or-less to achieve consensus as to the preference ordering of the alternatives from this point of view, then further sub-division is unnecessary; on the other hand, any substantial lack of consensus is an indication of the existence of different points of view, or criteria. In the EXAMPLE file, which relates to river basin planning in the Eastern Transvaal, overall benefit is sub-divided into four criteria, viz. Conservation, Forestry, Rural and Irrigators. Each of these is sub-divided once more into further constituent criteria.

Different interests may well involve different actors or groups of people. The evaluations of alternatives according to single points of view, and the aggregation of points of view at different levels in the hierarchy, may well be carried out by different groups. One of the primary aims of SDAW is to provide a means to a common currency whereby these different groups can communicate their preferences and values.

For each interest, the alternatives are compared with each other by scoring them on a "thermometer" scale, with the least desirable alternative (from the point of view of this interest) having a score of 0 and the most desirable a score of 100. Positions of the other alternatives are meant to be such that the gaps between them represent the relative gains in desirability when moving from one alternative to the next most preferred. So if 4 alternatives are scored respectively as 0, 25, 80 and 100, this implies not only the preference ordering, but also that replacing the second alternative by the third is relatively much more desirable than moving either from the first to the second, or from the third to the fourth. In other words, from this point of view, both of the 3rd and 4th alternatives are quite acceptable, while neither of the first two are. The scoring on the thermometer scales can either be done directly, or indirectly by aggregation of the scores for all the constituent criteria of the current interest. Direct scoring allows the user to move alternatives at will along the thermometer scale. In the case of indirect scoring (aggregation), however, the user associates weights with each component criterion (sub-interest), and an aggregate score is calculated as the weighted sum of scores over all such criteria (re-standardized to lie between 0 and 100 again). The user can interactively vary the weights to observe the effect on the implied ordering of alternatives.

The primary display shown by the software consists of three windows, viz. the current interest being evaluated (which is initially set as "overall benefit"), the criteria into which this interest is being decomposed, and the alternatives being evaluated. The software is mainly menu driven. The main menu items are now briefly described:

**Alterns:** Add or delete alternatives for evaluation

**Criteria:** Add or delete criteria, i.e. component sub-interests of the current interest

**Change Level:** This allows the user to change to a new interest to be considered as the "current" interest. One can move (1) up to the parent interest for which the current interest is a criterion; (2) adjacent to a sibling interest of the current (that is to a criterion sharing the same parent interest); or (3) down to make one of the criteria the current interest.

**Evaluation:** This allows direct or indirect evaluation of alternatives as described above. Of course, indirect evaluation is only possible if alternatives have been evaluated in terms of all component criteria. In the case of direct evaluation, a best and worst alternative must first be selected (although earlier selections are remembered), before proceeding to scoring on the thermometer scale. Once the scale itself is displayed, alternatives may be selected by clicking or by using arrows and enter key; scores may then be adjusted by arrow keys or mouse.

For indirect scoring (aggregation) the user will be asked on first entry to this module which criterion is most important to the current interest. Thereafter a bar chart is shown with weights for each criterion, and a thermometer scale showing the implied aggregate scores. The bar for any weight (except the most important which is used as a reference) can be selected (by tab and enter or by clicking), and the weight adjusted by arrow keys. Any criterion can be made "most important" by selecting it and pressing F2. On pressing F1, a summary display of all the thermometer scale scores for each component criterion is shown.

**File:** Usual file operations, plus an option to clear the current problem definition (in order to define a new problem).

**Quit**

The software can be supplied for the cost of diskettes and mailing, to anyone in the water community in South Africa. In case of problems in using the software, please contact us at telephone number (021)-650-3224/3219, or fax (021)-689-7578.

# Appendix C

# Linear Programming Formulations

A number of the methods described in this report include one or more linear programming (LP) formulations to compute the full answers and/or as consistency checks. The full set of formulations are collected together in this Appendix for easy reference.

## C.1 Conjoint scaling for consistency checks on scores

This linear programming formulation was introduced at the end of section 5.3 as a means of simultaneously evaluating the consistency between directly assessed scores (from "thermometer scales" and "swing weights") and preference rank orders on combinations of levels for pairs of criteria. Suppose that we have $n$ criteria, and that the number of levels defined for criterion $i$ is $m(i)$. Let $\tilde{u}_{ij}$ be the directly assessed score (after multiplying by the criterion weight, as in section 5.3), and let $u_{ij}$ be the consistent weights computed in the LP. Four sets of constraints are defined as follows.

**Normalization:** Standardize so that the maximum possible score is 100 (say):

$$\sum_{i=1}^{n} u_{i1} = 100 \tag{C.1}$$

**Consistency with rank orders:** Consider a table giving rankings for pairs of levels in a conjoint scaling exercise, as illustrated by the table on page 92. For the pair of levels indicated as having a rank of $r$, let $p(r)$ be the performance level on the first criterion and $q(r)$ the performance level on the second criterion. Thus, by definition $(p(1) = 1;\ q(1) = 1)$, while for the pair of criteria illustrated in the table on page 92, $(p(2) = 2;\ q(2) = 1)$, $(p(3) = 1;\ q(3) = 2)$, ..., $(p(6) = 2;\ q(6) = 3)$, ..., $(p(23) = 4;\ q(23) = 5)$. Consider now any pair of criteria, say $i$ and $k$; for each case in which rank orders (on pairs of levels) $r$ and $r+1$ have been given in the table, we would require for absolute consistency that:

$$u_{i,p(r)} + u_{k,q(r)} > u_{i,p(r+1)} + u_{k,q(r+1)}.$$

There is, however, no guarantee that consistent values, can be found. Thus **for each $i$, $k$ and $r$ for which the required ranking information is provided,** we introduce constraints of the form:

$$u_{i,p(r)} + u_{k,q(r)} - u_{i,p(r+1)} - u_{k,q(r+1)} + \delta_{ijr} \geq 0 \tag{C.2}$$

where $\delta_{ijr}$ is a measure of inconsistency for this one comparison.

**Deviation from directly assessed scores:** Ideally, we would wish to have $u_{ij} = \tilde{u}_{ij}$ in each case, but once again this may be infeasible, and we thus include constraints of the form:

$$u_{ij} - d_{ij}^{+} + d_{ij}^{-} = \tilde{u}_{ij} \tag{C.3}$$

for all $i$ and $j$. The variables $d_{ij}^{+}$ and $d_{ij}^{-}$ represent deviations above and below the directly assessed scores.

**Definition of the maximum deviation:** A weighted maximum deviation, $\Delta$, is defined by the following three sets of inequalities:

$$\Delta - \beta\delta_{ijr} \geq 0 \tag{C.4}$$

**for all $i$, $j$ and $r$ for which this is defined;**

$$\Delta - d_{ij}^{+} \geq 0 \tag{C.5}$$

**for all $i$ and $j$**; and

$$\Delta - d_{ij}^{-} \geq 0 \tag{C.6}$$

**for all $i$ and $j$.**

The parameter $\beta$ is an optional weighting factor to attribute different levels of importance to consistency with rank orders and to consistency with the directly assessed values respectively. In our studies, we have found that $\beta = 0.1$ yields useful results.

**Non-negativity:** By definition, we set $u_{i,m(i)} = 0$ for all $i$. All other variables are non-negative.

The best fit scores are then obtained by **minimizing:**

$$\Delta + \epsilon \sum_{i,j,r} \delta_{ijr}$$

subject to the numbered and non-negativity constraints above. The summation term is over all $i$, $j$ and $r$ for which the deviations are defined. The $\epsilon$ term is a small positive constant; we have generally used $\epsilon = 0.01$.

## C.2 LP Formulations for pure rank assessments

In this section we document the linear programming formulations relevant to the use of ordinal information on preferences only, as described in §5.1. As in that section, we define $u_{ir}$ to be the weighted value of the $r$-th ranked alternative according to criterion (or interest) $i$ (weighted by the importance of $i$). By definition, if there are

$N$ alternative scenarios, we have $u_{iN} = 0$ (which need not, therefore, be explicitly included in the LP formulation) and $u_{i1} = w_i$, i.e. the weight of criterion $i$.

The **objective function** to be *maximized* for each criterion $k$ in turn, is given by:

$$V_j - \frac{1}{N-1}\sum_{j \neq k} V_j \tag{C.7}$$

The following constraints are common to all three of the LP formulations in this section.

**Ordering:**

$$u_{ir} - u_{i,r+1} \geq \epsilon \tag{C.8}$$

for $i = 1, 2, \ldots, n$, $r = 1, 2, \ldots, N-1$, and for some specified number $\epsilon > 0$.

**Value definition:**

$$V_j - \sum_{i=1}^{n} u_{ir(i,j)} = 0 \tag{C.9}$$

for each alternative scenario $j$, where for the case $r(i,j) = N$, $u_{iN}$ is treated as zero.

With this background, three alternative formulations were stated as follows.

**LP1: Equal weights; Arbitrarily allocated "gaps"**

This formulation simply maximizes (C.7), subject to the constraints (C.8) and (C.9), to $u_{i1} = 100$ (to create equal weights), and non-negativity of $u_{ir}$ for $r = 2, \ldots, N-1$.

**LP2: Equal weights; Decreasing "gaps" with increasing ranks**

In many cases, the importance placed on the distinction between best and second best options may be perceived to be of a much higher order than the distinction between worst and second worst, say. With this in mind, the formulation of LP2 is as for LP1, but with the addition of an extra set of constraints to ensure that the value "gap" between the first and second placed scenarios for any user group is larger than the gap between the second and third placed scenarios for the same user group, etc. In formal terms, the constraints C.8 are replace by the following for all except the case of $(r = N-1)$:

$$u_{i,r-1} - u_{ir} \geq u_{ir} - u_{i,r+1} + \epsilon$$

for $r = 2, \ldots, N-1$. This is better represented in standard fashion by:

$$u_{i,r11} - 2u_{ir} + u_{i,r+1} \geq \epsilon \tag{C.10}$$

for $r = 2, \ldots, N-1$.

**LP3: Variable weights; Decreasing "gaps" with increasing ranks**

This formulation is a relaxation of LP2, to allow some variability in the weights on each user group. The only change from LP2 is that the separate normalization for each $i$ (i.e. $u_{i1} = 100$) is replaced by:

$$u_{i1} \leq 100$$

and by constraints of the following from for each *pair* of criteria or interests, say $i$ and $\ell$:

$$u_{i1} - \beta u_{\ell 1} \geq 0 \tag{C.11}$$

$$u_{\ell 1} - \beta u_{i1} \geq 0 \tag{C.12}$$

for a specified $0 < \beta < 1$. The parameter $\beta$ represents the proportionate inter-criterion variation in weights which is allowed: the weight on any one user group cannot be less than a proportion $\beta$ of the weight of any other.

## C.3 LP formulation for imprecise value judgements

In this section we state the linear programming formulations required to implement the treatment of imprecise value judgements, as described in §5.2. The formulation is in fact precisely as for LP1 in the previous section, with the addition of one or more constraints of the following three generic forms:

**Weight ranges:** For any pair of criteria, say $i$ and $\ell$, a user might state that the ratio of importance weights lies in a range such as the following:

$$\rho_{i\ell} \leq \frac{w_i}{w_\ell} \leq R_{i\ell}$$

This translates into the following pair of constraints:

$$u_{i1} - \rho_{i\ell} u_{\ell 1} \geq 0 \tag{C.13}$$

$$u_{i1} - R_{i\ell} u_{\ell 1} \leq 0 \tag{C.14}$$

**Bounds on gaps:** The gap between the $r$-th and $(r+1)$-th ranking alternatives in terms of scenario $i$, expressed on a standardized scale (taken here to be 0–100), may be stated to be between to numbers $a_{ir}$ and $b_{ir}$, where of necessity $0 \leq a_{ir} \leq b_{ir} \leq 100$. This translates into a pair of constraints:

$$u_{ir} - u_{i,r+1} \geq a_{ir} \tag{C.15}$$

$$u_{ir} - u_{i,r+1} \leq b_{ir} \tag{C.16}$$

**Gap ratios:** The gap between the $r$-th and $(r+1)$-th ranking alternatives in terms of scenario $i$ may be stated to be at least $\alpha_{irs}$ times as great as the gap between the $s$-th and $s+1$-th ranking alternatives in terms of same scenario. This translates into the constraint:

$$u_{ir} - u_{i,r+1} - \alpha_{irs} u_{is} + \alpha_{irs} u_{i,s+1} \geq 0 \tag{C.17}$$

A number of instances of each of the above types of imprecision may be included in any one problem setting.