ASPECTS OF MCDA CLASSIFICATION AND SORTING METHODS

by

RENÉE KOEN

submitted in part fulfilment of the requirements
for the degree of

MASTER OF SCIENCE

in the subject

OPERATIONS RESEARCH

at the

UNIVERSITY OF SOUTH AFRICA (UNISA)

SUPERVISOR: PROF CJ SWANEPOEL (UNISA)

JOINT SUPERVISOR: DR CD ELPHINSTONE (CSIR)

NOVEMBER 2008

Contents

Acronyms

AHP	Analytic Hierarchy Process	75
CSIR	Council for Scientific and Industrial Research	62
DOJCD	Department of Justice and Constitutional Development	62
DME	Department of Minerals and Energy	63
GDP	Gross Domestic Product	53
HDI	Human Development Index	53
IRIS	Interactive Robustness analysis and parameter Inference Software	35
LEI	Life Expectancy Index	53
LP	Linear Programming	38
MAUT	Multi-Attribute Utility Theory	11
MAVT	Multi-Attribute Value Theory	11
MCAP	Multicriterion Aggregation Procedures	27
MCC	Multiple Criteria Classification	39
MCDA	Multi-Criteria Decision Analysis	5
MCDM	Multi-Criteria Decision Making	5
MCSP	Multiple Criteria Sorting Problems	16
MH DIS	Multi-Group Hierarchical Discrimination	39
OR	Operations Research	7
ORCLASS	Ordinal Classification method	38
ROSE	Rough Sets Data Explorer	37
SMAA	Stochastic Multiobjective Acceptability Analysis	37
SMART	Simple Multiple Attribute Rating Technique	39
UTADIS	UTilitès Additives DIScriminantes	38

Chapter 1

Introduction

This document addresses certain practical and theoretical issues of classification and sorting within the field of Multi-Criteria Decision Analysis (MCDA).

It should be noted that the acronymn MCDA is often translated by experts in the area as Multi-Criteria Decision Aid in order to stress the fact that it provides an aid to the decision-maker. Such "aid" includes the analysis required for modelling or solving a problem, but is seen as a more encompassing term that includes the formulation of a problem and the creation of alternatives, even if such formulation and alternative creation does not lead to further analysis of the problem. (See [53] in this regard.) Furthermore, the field is also often termed Multi-Criteria Decision Making (MCDM), but this could be seen as a narrower term, as methods directly for a decision-maker and excluding the role of a facilitator (see [3], for instance). For the purpose of this document, the acronymn MCDA will be used, but it will be understood to cover all fields usually associated with MCDM and MCDA by the different groups of experts.

The document consists of this introduction, five main chapters and a brief concluding chapter.

The chapter following this introduction provides an overview of MCDA, discussing the main aims, definitions, and foundational concepts. This is followed by a chapter looking in more depth at some of the issues within MCDA that one would have to understand in order to use MCDA methods in practice, and focuses specifically on the subcategory of MCDA classification and sorting methods. Chapters 4 and 5 build on chapters 2 and 3 in order to investigate two main issues, namely choosing an appropriate method and constructing measurements.

Chapter 6 provides a post-analysis of two practical case studies, which contrasts the approach followed in the case studies with the MCDA approach presented in chapters 2 – 5. The two case studies gave rise to questions about the application of MCDA in practice, and specifically about application of MCDA classification and sorting methods. They involved projects carried out within the public sector in South Africa, i.e. problems that had to be solved for specific government departments. Chapter 6 describes the case studies, provides conclusions on the implementation of MCDA methods on the practical problems, and offers general conclusions about the application of classification and sorting MCDA methods within the public sector.

The document ends with a few concluding remarks in chapter 7.

Chapter 2

Foundations of MCDA

The field of Multi-Criteria Decision Analysis (MCDA) focuses on problems in which there are more than one criterion that have an impact on the problem or decision at hand. Usually, these criteria are in conflict with one another, or one criterion may place a restriction on the optimal achievement of another criterion. Daellenbach [16] provides insight on why other, more traditional, Operations Research (OR) procedures are not able to provide a satisfactory solution to problems of this nature:

> The traditional way of modelling multiple objectives is to optimize what is considered to be the most important objective, while meeting minimal performance targets on all other objectives. For instance, for most issues involving safety, such as the operation of various means of transport or plants, like a nuclear power station, an acceptable 'solution' is obtained by minimising operating costs, subject to meeting certain safety standards. Usually, cost minimisation is considered to be the most important objective, while safety and other objectives are subordinate to it.
>
> In this approach the lesser objectives are replaced by minimal performance targets that have to be met, i.e., by firm surrogate constraints. Therefore, they restrict, in fact, dictate the best level of achievement that is possible for the most important objective. In other words, this approach guarantees that the targets on the lesser objectives are satisfied, before it allows any look at the most important objective. So, by a rather ironic twist, the most important objective becomes subordinate to the less important objectives.
>
> Mind you, when dealing with safety issues this may be all the better. However, there are many situations where this inadvertent reversal of priorities is more questionable. In particular, these minimal performance targets on the lesser objectives may involve a considerable degree of arbitrariness, nor do the performance targets chosen have the inviolate nature of physical constraints. They are often the result of a policy decision. They reflect what is seen as a reasonable or a desirable level of achievement – both rather fuzzy and highly subjective notions.

He goes on to explain that in the MCDA context, an 'optimal' solution cannot exist when there is more than one objective to consider, since the optimum (maximum or minimum) only has a real meaning in terms of one objective. In MCDA problems, one faces the situation that one solution may do well in terms of some of the objectives or criteria, while another solution may do well on a(nother) set of objectives and criteria. In MCDA problems one does not search for an optimal solution, but rather for the 'most preferred' solution.

One should of course understand that any problem or decision situation involving more than one criterion for consideration would not necessarily qualify as an MCDA problem. Belton and Stewart [3] indicate that in order to warrant a formal MCDA modelling or analysis procedure, a problem must have substantial consequences, have impacts that are long-term or affect many people, and be of a nature in which mistakes may not easily be remedied. Furthermore, MCDA problems typically have a large amount of information of a complex and conflicting nature that must be organised in order to support the decision. Some of the information could also be in the form of personal views and opinions, and these could change over time. They state that:

> One of the principal aims of MCDA approaches is to help decision makers organise and synthesise such information in a way that leads them to be confident about making a decision, minimising the potential for post-decision regret by being satisfied that all criteria or factors have properly been taken into account. *Thus, we use the expression MCDA as an umbrella term to describe a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter.* Decisions matter when the level of conflict between criteria, or of conflict between different stakeholders regarding what criteria are relevant and the importance of the criteria, assumes such importance that intuitive "gut-feel" decision-making is no longer satisfactory.

It should be pointed out that MCDA problems should not be equated to complex problems. Many very complex problems can be formulated, modelled and solved with OR techniques other than the classical MCDA techniques. MCDA problems typically are quite complex, but the distinguishing characteristic is the fact that various conflicting criteria and the interactions between them have to be modelled explicitly in order to gain an understanding of the problem or to provide a solution to the problem.

2.1 Foundational concepts

In order to understand the MCDA field, it is necessary to understand the various terminologies and concepts that are being used in this field. This subsection therefore provides a brief summary of what is meant by certain terms and why these terms represent foundational concepts that underlie the field of MCDA.

It should firstly be noted that it is usually assumed that some sort of "problem" or "decision-making situation" drives the need for an(y) OR solution. Since MCDA represents a class of OR methods, this would also be a requirement underlying an MCDA approach. Such a "problem" is often initiated from the environment, for instance action by a business competitor or the introduction of new policies or legislation from local or national government. However, it is also possible that the "problem" is more of an opportunity that is identified and initiated by the decision-maker, as is pointed out by Keeney [30]. However, one could argue that this issue is not specific to MCDA, and, although obviously important in understanding the situation and the focus of the decision-maker(s), not exclusively part of the MCDA field.

There are other concepts that, although also of importance within other fields, take on a specific meaning within the MCDA field. Such foundational concepts specific to MCDA, or of specific meaning within MCDA, include the following:

- The first aspect to note is that the decision-maker has to choose between a number of *alternatives*, also referred to as *actions, options, strategies,* or *plans*. The crux of the MCDA problem is usually that there is no clear and easy way to select among the possible alternatives, therefore requiring some supporting methodology from within the MCDA field. Note that the problem may be to identify the 'best' alternative, or it may be to find a group of 'best' alternatives, or it may even require a two-step process in which one first selects a 'best' group and from that group eventually select a single 'best' alternative. Another type of problem situation may require that different alternatives be ranked from 'best' to 'worst' in some acceptable way. And, finally, note that within the context of decision aid, the aim may in fact be just to clarify what is meant by a 'best' alternative in order to improve choices on the matter. It should also be pointed out that all alternatives need not be defined or available at the decision time. MCDA support may in fact be required to identify them or to provide guidelines for their identification. Also, in most MCDA methods the alternatives are assumed to be independent in the statistical sense, so that picking one alternative would not immediately imply also having to pick another alternative at the same time.

- Usually, the way to define the goodness or attractiveness of an alternative or group of alternatives, are to relate them to certain *criteria*, also called *attributes* or *key factors*. A very important point is that MCDA methods are usually required if these criteria are in conflict, i.e., when no alternative performs 'best' in terms of all criteria, and alternative(s) that measure well on some of the criteria may measure badly in terms of another group of criteria. Again, it should be pointed out that all the criteria to be used for measurement of alternatives need not necessarily be identified or clearly specified at the time when a decision needs to be made.

- The decision-maker(s) may want to satisfy certain *objectives* or *values* or further certain important *goals* through the decision. The MCDA process needs to account for these objectives through the criteria (i.e. the objectives must drive criteria measurements) in order to provide an acceptable choice or ranking of alternatives. This can provide a particular challenge, because the decision-maker may not want to disclose all relevant objectives (especially if there are areas of conflict or political maneuvering within the group of decision-makers), and may also may not even be aware of certain underlying objectives that are perhaps taken for granted.

- A very important concept, if not the most important concept, within the MCDA field is the issue of reflecting the *preferences* of the decision-maker correctly. These preferences must give guidance to the relative importance attached to the various criteria, given the underlying objectives and values, in order to select, classify or rank the alternatives successfully. An important concept is that MCDA methods allow different types of preference relationships to be expressed by the decision-maker, for example, whether one alternative or criterion is strictly preferred to another, or whether the decision-maker feels indifferent (unable to express a clear preference) between them.

- Some alternatives may have *consequences* attached to them. In modelling preferences, attention should be paid to assessing the consequences of the alternatives, if applicable.

- Since the criteria that measure the alternatives are often in conflict with one another, it is necessary to determine *trade-offs* between them. It is important that these trade-offs be quantified correctly to correspond to the preferences of the decision-maker. Trade-offs may also be referred to as *inter-criteria comparisons* or *compensation* as in Gitouni and Martel [27] and Bouyssou and Vansnick [6]. Although Gitouni and Martel acknowledge that this concept is not yet well defined, they do mention the following categories of compensation:

  - **Compensatory** methods require that trade-offs (or compensation) can be made between criteria, so that an improvement in one criterion can be counter-balanced with a decline in performance on another criterion.
  - In **non-compensatory** methods no trade-offs between criteria are allowed, for instance when the decision-maker indicates that criteria are so important that trade-offs between them cannot be considered.
  - For **partially compensatory** methods, which include most of the MCDA methods, some form of trade-off can be accepted between criteria, and the major problem is to evaluate the degree of compensation between criteria.

- Although the issue of a *hierarchy* is known in many fields, within the MCDA field it is often used to provide a certain amount of ordering, comparison, or grouping, specifically within criteria. For instance, criteria are often presented in a hierachical or tree structure to indicate how they are related to each other. For instance, when one criterion is defined by certain sub-criteria, the criterion could be represented within a hierarchy as a node which branches into the sub-criteria nodes "further down".

- Decision problems often involve an element of *uncertainty*. This could affect a MCDA problem situation in more than one way. It could refer to the fact that criteria, preferences or trade-offs cannot be exactly measured or quantified. This type of uncertainty therefore has an impact in terms of measurements used and could even impact on whether a specific MCDA method (for instance one requiring exact quantification) is applicable to the problem situation. Another type of uncertainty could be due to the fact that a decision must be made at a certain point in time, but that this decision is influenced by something that will happen in future and the outcome of this future event is not known at the time when the decision must be made. This brings a *stochastic* element to many decisions, and the presence of such a stochastic component needs to be considered in the MCDA method that supports the decision.

- The concept of *utility* was developed within the economic and decision sciences, and is also used within certain MCDA methods, in order to combine both the subjective assessment by a decision-maker of the "value" or preference attached to a criterion, as well as the likelihood of it happening (see p. 110 in [30]). The von Neumann-Morgenstern utility function was designed as a way in which to model the preferences of a decision-maker(s), especially in the presence of uncertainty or risk. For example, a decision-maker would "value" a potential loss or income differently depending on the size, i.e. how much money it represents, *as well as* on how likely it is believed to occur. Clemen indicates that a *ordinal utility function* can be used for decisions in which there is not a large degree of uncertainty, while *cardinal utility functions* should be used when a decision is made under uncertainty ([15] on p. 552). Utility measurement is one of the central concepts underlying Multi-Attribute Utility Theory (MAUT), while the related Multi-Attribute Value Theory (MAVT) makes use of similar preference value functions, but do not incorporate utility.

- Of importance within the MCDA field is the way in which or the extent to which criteria can be expressed in a quantitative way. As in many other fields, in MCDA it is recognised that certain criteria could be measured as a 'continuous' variable, while others can only be expressed as 'discrete' (categorical) values. Many MCDA methods recognise the fact that some criteria are of a qualitative rather than quantitative nature and needs to be incorporated into the model in that way.

- Although a certain element of *subjectivity* is involved in many different types of modelling, it is of particular importance in the field of MCDA. Not only is the emphasis of MCDA methods on the explicit expression of the highly subjective preferences, values and criteria of the decision-maker(s), but the decision itself usually has subjective elements. The decision is taken within a very specific time frame, and is based on the sometimes limited information available at that specific point in time. There may also be certain risks and uncertainties that influence how the decision is perceived. The problem context provides a specific subjective element to the decision process.

- The concept of *weights* is also interpreted slightly differently in MCDA than in other fields. While the mathematical expressions of, for instance, preference structures appear to be weighted sums, the interpretation of the weights are much more subjective that just "the importance" indicated by the weights. The weights are in fact subjective expressions of trade-offs which could roughly be equated to an expression of the importance of one compared to another. However, Belton and Stewart [3, pp. 82-83] warn that the preferences may be very non-linear (for example, expressing values before and after a threshold value is reached) and Keeney [30, pp. 147-148] indicates that they are linked to the decision-maker's preference given a certain time period and other contextual elements. He uses the following example to explain the issue:

> For instance, a respondent might state that pollutant concentrations are three times as important as costs. While the sentiment of this statement may make sense, it is completely useless for understanding values or for building a model of values. Does it mean, for example, that lowering pollutant concentrations in a metropolitan area by one part per billion would be worth the cost of $2 billion? The likely answer is "of course not". Indeed, this answer would probably come from the respondent who had just stated that pollutant concentrations were three times as important as costs. When asked to clarify the apparent discrepancy, he or she would naturally state that the decrease in air pollution was very small, only one part per billion, and that the cost was a very large $2 billion. It is necessary to know how much the change in air pollution concentrations will be and how much the costs of regulation will be in order to logically discuss and quantify the relative importance of the two objectives.

It should also be noted that the interpretation of "weights" could differ between different MCDA techniques, being seen as close to trade-offs within some techniques such as MAUT, but seen more as strength of evidence in the outranking techniques [3, p. 110, 114].

2.2 Definitions

Within the MCDA field, there are many words that may be used to define certain aspects. For the purpose of this document, certain terminologies will be used, and will be defined in the following way:

- The *decision problem* refers to the issue requiring the systematic methodology of an MCDA solution. Note that the word 'problem' is used not to indicate that the situation is a negative one creating a problematic situation, but is merely a term used to encompass situations that may be seen as problems, challenges or opportunities by the decision-maker(s).

- When referring to the *decision-maker*, this term is taken to mean the individual or group who experiences ("owns") the decision problem. Although in a practical problem there may be people working to find an MCDA solution on behalf of those who actually experience the problem, in this document there is not such a fine distinction in terms of the various groupings within the decision-making group, and 'the decision-maker' will refer to anyone who provides context to the decision problem or the subjective preference information that will be required for a solution.

- The word *alternative* will be used to refer to a possible solution (alternative, option, strategy or plan) to the MCDA decision problem.
- The term *criteria* will be used to refer to the criteria, attributes or key factors against which the alternatives will be measured in order to find the solution.
- The term *values* will refer to underlying values, objectives or goals of the decision-maker that influence their choice in terms of alternatives, criteria or preferences.
- The word *preferences* will be used to refer to the quantified preference values that are used in the MCDA methods to model the subjective goals and values of the decision-maker, given the subjective context of the specific decision-maker group, the time at which the decision must be made and the limited extent and type of information available. Note that the term *preferences* does not simply apply to compensatory preferences that allow trade-offs, but is intended to encompass non-compensatory and partially compensatory preference structures as well.
- The term *measurements* will be used to refer to the way in which criteria or preferences are "quantified", even if such measurements are of a qualitative nature.

2.3 Limitations of this study

This document aims to provide details that affect *the practice of applying* MCDA methods (specifically MCDA classification and sorting methods), measurement issues within MCDA problems, and the way to choose an appropriate method to provide an MCDA solution. Only brief references are made to aspects of the decision problem that affect the context of the MCDA decision problem or the implementation of the solution.

This study therefore does not specifically address the issues of problem formulation for MCDA problems or conflict resolution within the decision-making team. Obviously, this means that the more behavioural issues regarding how decisions are made or influences by group dynamics or facilitation of a group are not discussed in detail. Also, although the normal OR cycle of problem formulation, solution, and implementation is assumed to underlie the MCDA field (as well as any other OR field), issues of implementing a solution, providing recommendations and revising the solution based on feedback from the implementation process are not within the main focus of this document.

Finally, note that there is also no scope within this document to address issues regarding how a decision-maker without knowledge of OR or MCDA must use MCDA. Furthermore, although brief mention is made of available software that apply MCDA classification and sorting methods, there is no scope to discuss details of obtaining, using or trouble-shooting any software packages that implement MCDA methods.

Chapter 3

Important concepts in MCDA classification and sorting problems

MCDA has become a very wide field of study, with many different methods and applications reported in scientific literature. Some of the reasons motivating the development of different methods are due to specific requirements from the practical problem context, or due to various experts taking different views on the same problem, or due to computational issues affecting solution algorithms. While the exact reasons for differences between MCDA methods or experts are not considered in this document, it is considered important for this study to understand more about the specific areas of difference between methods, and the concepts behind those differences. This chapter therefore contains a brief summary of some of the concepts that distinguish between different MCDA methods, with a brief explanation of why these issues are important.

Furthermore, since this study focuses on classification and sorting methods, this chapter provides an indication of to what extent the concepts of importance to general MCDA methods also impact on classification and sorting problems, as well as how classification and sorting methods have some specific issues of importance that differ from general MCDA methods.

3.1 Decision problematique

Although from a theoretical point of view the type of MCDA problem is perhaps not of great concern, in the practical application of these methods the **type of problem** often plays a pivotal role in the approach and methodology to be used.

One of the first references to the different kinds of analyses which may be conducted with MCDA methods was provided by Bernard Roy in 1985 in the French book entitled *Methodologie Multicritere d'Aide a la Decision* (quoted by [60], [3] and [27]). This included the following four types:

1. to identify the best alternative or select a limited set of the best alterna-

tives, namely the **choice problematique** (denoted by α)

2. to classify or sort the alternatives into predefined homogeneous groups, namely the **sorting problematique** (denoted by β)

3. to construct a rank ordering of the alternatives from the best to the worst ones, namely the **ranking problematique** (denoted by γ)

4. to identify the major distinguishing features of the alternatives and perform their description based on these features, namely the **description problematique** (denoted by δ)

Belton and Stewart ([3]) suggested the addition of two other problematiques, namely the so-called *design problematique* which would deal with the search or identification of (new) alternatives, and the *portfolio problematique* which would involve determining a subset of alternatives that is not only a group of 'best' alternatives, but also capitalise on interactions and synergies between the various selected alternatives. Guitouni and Martel ([27]) also quote Carlos Bana e Costa as suggesting problematiques such as *the choice of k from n*, *the successive choice problematique*, and so on. Zopounidis and Doumpos ([60]) suggest that the sorting (β) problematique be further broken down into a distinction between **sorting** that would place the alternatives in groups that are ordered in some way and **classification** which does not require any ordering of the categories.

It is of course possible to provide many extensions to Roy's list of problematiques, either through a new type of problem that may be addressed with MCDA methods or a finer distinction of different "subproblems" within a specific problematique.

In practice the importance of paying attention to a list of problematiques may not be to get an extensive list of possible types. The importance of such a list lies in the guidance it can give in terms of an appropriate (or not appropriate) technique to be used to address a specific problem type. Mousseau *et al* brings this point across quite strongly in [39] when they say the following:

> Among these problem statements, a major distinction concerns *relative vs absolute judgement* of alternatives. This distinction refers to the way alternatives are considered and to the type of result expected from the analysis. ... Choice (selecting a subset A^* of the best alternatives from A) or ranking (definition of a preference order on A) are typical examples of comparative judgements. The presence (or absence) of an alternative a_k in the set of best alternatives A^* result from the comparison of a_k to the other alternatives. Similarly, the position of an alternative in the preference order depends on its comparison to the others. ... The sorting problem statement refers to absolute judgements. It consists in assigning each alternative to one of the categories which are pre-defined by some norms corresponding to vectors of scores on particular criteria, called profiles, either separating the categories or playing the role of central reference points in the categories. The assignment of an alternative a_k results from the intrinsic evaluation of a_k on all criteria with respect to the profiles defining the categories (the assignment of a_k to

> a specific category does not influence the category to which another alternative should be assigned).

He also distinguishes two categories of sorting (β) problems, namely *ordered* Multiple Criteria Sorting Problems (MCSP) and *nominal* MCSP where the distinction lies in whether the categories into which the sorting is done have a specific logical order attached to them or not. Araz and Ozkaran ([2]) emphasises the ordering of categories by calling classification *nominal classification* and sorting *ordinal sorting* problems.

The main difference between sorting and classification, according to Greco *et al* [26], is that sorting requires the addition of user preferences to the classification and that there should be consistency within the preferences. For instance, suppose two companies A and B are compared on various criteria in order to determine the bankruptcy risk, and company B has a worse debt ratio than that of company A, but evaluations of these companies on all other criteria are equal. Then they must not only be placed in different categories according to risk, but the category in which A is placed should be *preferred to* ("better than") the category in which B is placed. Also note that the ordering may be based on subjective ordering by the decision-maker. For instance, there is no intrinsic ordering between a certain model car that is only available in green and another model that is only available in red, since green is not intrinsically better or worse than red. However, a decision-maker may prefer red to green and this may then provide a subjective ranking. It is therefore possible that sorting problems require more preference modelling than classification problems.

It can be concluded that it is quite important, especially from a practical point of view, to know the problematique – if not by the exact classification, then at least to know if this is a problem in which one would be doing relative comparisons or absolute comparisons of alternatives on the basis of criteria. Also, it is important to know how much and the type of subjective preference judgements to include in the modelling. This could affect both the method used and the general methodology involved in eliciting preferences.

A further insight into this aspect is provided by Belton and Stewart [3] in terms of the type of decision modelling problem and how this may affect the choice of an appropriate MCDA method.

> Where the terms of reference and culture of a group are such that a clear recommendation for action is required, especially when such a recommendation requires clear justification in an open forum, then it is our view that the value function methods ... are particularly well-suited. Where terms of reference and/or culture are, on the other hand, such that the primary aim is to provide succinct summaries to higher level decision-makers concerning the pros and cons of alternative courses of action, then the outranking methods ... may be particularly useful. For a relatively homogeneneous group seeking a rapid solution to their own satisfaction, not requiring justification to outsiders, the interactive and aspiration level methods ... are useful tools.

It seems that the problematique as well as the aims of the decision-making needs to be carefully considered in applying MCDA methods in practice.

3.2 Important general issues within MCDA problems

MCDA problems and methods differ in various ways, and it is important to have some understanding of how they differ and what this means.

3.2.1 Interpretation of weights

As has been indicated in the previous chapter, the interpretation of preference weighting may differ between the different types of MCDA methods. This is, however, not just a theoretical difference. As indicated in [3]:

> One of the critical challenges facing the analyst in guiding decision-makers through the process of MCDA is to ensure that the clients correctly understand and interpret the meaning of the weights when they are asked to provide numerical inputs ... weight parameters may have widely differing interpretations for different methodologies and different decision contexts, so that we must be wary of any approach to assessment of weights which do not take such differences into account. ... The overall responsibility of the decision analysis or facilitator, particularly with value measurement and outranking methods, is thus to ensure:
>
> - that the decision-makers are fully informed concerning the meaning of weights within the context of the model being used
> - that questions concerning importance weights are formulated in such a way that mismatch between responses and model requirements are avoided
> - that the potential effects of the cognitive biases arising from problem framing and hierarchical structuring of criteria ... are fully explored as part of the sensitivity analysis

Before using a specific MCDA method, one should pay attention to the role played by the criteria weights in the method. Not only does this affect the solution method, it also affects the way information should be obtained from the decision-maker.

3.2.2 Preference articulation

Closely linked to the issue of how preference weights will be used in a method is the way in which decision-maker preferences are extracted for MCDA modelling. According to [27], some examples of ways to extract preference information from decision-makers include:

- tradeoffs: the decision-maker has to go through a process of indicating the amount of criterion c_i that could be sacrificed in order to obtain a unit increase in criterion c_j for all criteria, i.e. for all values of i and j.

- lotteries: the decision-maker is asked to indicate what fixed amount they would prefer as opposed to the opportunity to 'play' a lottery in which they stand to win a certain amount with a specific probability. This information supports the construction of a utility (value) function indicating the decision-maker's view of risks, and therefore this method is usually applicable to decisions made under conditions of uncertainty and not always generally applicable to any MCDA problem situation.

- direct rating: the decision-maker provides information on the alternative that rates best or worst on a specific criterion, and then to rate all alternatives on a scale between the best and the worst alternatives.

- pairwise comparisons: the decision-maker is asked to compare alternatives to each other in a pairwise way, indicating the relative preference of individual alternatives in some quantitative or preference structure description.

3.2.3 Preference structures

Key to being able to express decision-maker preferences in the model, is the preference relations that can be included in the model in order to capture or express views / values / preferences as closely as possible. Guitouni and Martel ([27]) list the following as possible preference relationships:

- Strict preference (P): a is strictly preferred to b. This case should be well understood to mean that there is enough evidence to conclude that a is really better or more preferred than b.

- Indifference (I): a is indifferent to b, which means that either there is no difference between a and b or the difference is too small to be considered a real distinction between the two.

- Weak preference (Q): the hesitation between the indifference and preference situations, not sure whether a is preferred to b or not.

- Incomparability (R): the hesitation between whether a is preferred to b or b is preferred to a (between aPb and bPa). Usually this happens if a is better than b according to some criteria and b is better than a on some other criteria, but these criteria are not compensatory (comparable).

- Outranking relation (S): when there is a strong reason to believe that with respect to all the n criteria an alternative a is at least as good as b without any reasons to absolutely prevent this conclusion.

A simple explanation of the difference between indifference and incomparability is provided by Belton and Stewart ([3, p. 108]), as illustrated in table 3.1. In this example, there is a very small difference between alternatives 1 and 3, and so a decision-maker may want to indicate that he/she is indifferent between these two options. However, alternative 1 scores highly on criteria A and B, and low on C and D, while alternative 2 has almost the opposite score pattern: low on A and B, and high on C and D. The decision-maker may well want to consider alternatives 1 and 2 as incomparable.

Table 3.1: Three alternatives scored on 4 criteria using a scale of 1 to 9

Alternative	A	B	C	D
1	9	9	2	2
2	1	1	9	9
3	8	8	2	2

Note that not all MCDA methods allow all of these relationships to be expressed. Most methods allow for strict or weak preferences to be implicitly expressed in terms of the attached weights, and some methods allow both preferences and indifference to be modelled. Outranking methods allow all of the types listed above to be explicitly modelled.

3.2.4 Aggregation procedures

Many different procedures have been suggested (and sometimes hotly debated) for aggregating criteria scores and weights within MCDA methods. Guitouni and Martel ([27]) distinguish between three such procedures:

- Single synthesizing criterion approach without incomparability: in these types of methods, compensation between criteria and comparability between alternatives are assumed, and the aim is to determine an aggregating function V in which alternative a_i is evaluated in terms of all criteria g_j in order to obtain a single score g such that $g(a_i) = V[g_1(a_i), g_2(a_i), \ldots, g_n(a_i)]$.
- Outranking synthesizing approach: in this approach, non-compensatory criteria are allowed, and the problem of incomparability between alternatives is addressed by allowing more preference structures to be used than just the preference and indifference relationships allowed in the single synthesising approach. There are different methods according to the different preference structures accommodated as well as other assumptions such as transitivity of preferences (see subsection below for a brief discussion of transitivity).
- Interactive local judgements with trial-and-error approach: this approach is also called interactive methods by others such as Belton and Stewart ([3]) who explain that these methods are usually carried out in three iterative steps. Firstly, some method is used to generate a small set of feasible alternatives. Secondly, if the decision-maker is happy with this solution, the process stops, otherwise the decision-maker is asked for information that would improve the process of selecting alternatives. In the third step this information is used to update the method of selecting alternatives for consideration, and the process returns to step 1.

It should be noted that each of these three main approaches can be further broken down into the various mathematical methods (algorithms) used to accomplish the aggregation.

3.2.5 Transitivity of preferences

Transitivity refers to the property that if the decision-maker prefers alternative A to alternative B and alternative B to alternative C, then one can assume that the decision-maker will prefer alternative A to alternative C. Simon French indicates that one can usually expect a rational decision-maker to have transitive preferences ([23]) and that if intransitivity of preferences is pointed out to the decision-maker, he or she will revise the preferences in order to make the preferences transitive. Alternatively, intransitive preferences could be an indication that there are some criteria affecting the decision that have not been specified in the preference structure, and therefore that additional modelling may be required.

However, Roy and Vincke [48] point out that there could be a few decision situations in which transitivity of preferences cannot be maintained. For instance, there may not be enough information about the alternatives to make the decision-maker sure of the exact preferences. Furthermore, there could be decision situations in which the preferences of the "real" decision-maker is modelled by a person other than the decision-maker and this modeller may not have complete information on the preferences. Also, they indicate that there may be a group of decision-makers with conflicting views instead of a single decision-maker and such a group could in fact display intransitive preferences.

Therefore, although transitivity of preferences may generally be a realistic expectation, and even a requirement, in any rational decision-making problem, there could well be situations in which decision-makers do not have transitive preferences, and this may not always be due to inconsistency on the part of the decision-maker.

3.2.6 Data types

Some methods are designed to provide solutions aimed specifically at problems in which data for either criteria measurements, preference weights or alternative scores are of a particular data type. For instance, methods based on mathematical programming techniques, such as goal programming, can only be used if measurements are quantitative (continuous), while other methods may specifically apply to qualitative data. In the article by Hogarth and Karelaia ([29]), a discussion is provided of simplified methods that can be used if the criteria are measured as binary variables, i.e. if they can take on one of two values, such as yes/no, present/absent, acceptable/not acceptable.

3.2.7 Completeness of preference structures

A very difficult aspect concerns knowing whether the preference structure is complete (i.e. whether the criteria are exhaustive enough to do a valid assessment). Mostly, this is a question of judgement by the participants of whether all aspects affecting the decision are contained in the structure.

Also, there are practical ways in which one can determine areas in which the preference structure may be lacking. For instance, there may be conflict or argument about a certain criterion, and such conflict could be an indication that the criterion is seen as measuring different things by different decision-makers. In such a case, one may want to break down this criterion further, or investigate underlying value systems leading to this criterion in more depth so that the reasons for the conflicting opinions may be highlighted. Or, the analysis may produce results that the decision-makers consider to be counter-intuitive or illogical. This would also point out that there are errors or deficiencies in the preference structure.

There are also some more objective guidelines to be followed. Engelbrecht [21] quotes a way in which the "depth" of a value tree may be used as a guideline for the completeness of the tree. This is maybe a very rough guideline, but could be useful for comparing sections of a tree for possible areas in which expansion or pruning may be required. If the value tree shows *quality of life* and *economic measures* as the two main criteria, but *quality of life* is evaluated only in terms of two sub-criteria, while economic measures is evaluated by 10 sub-criteria in 3 different levels of the tree, one may want to investigate the appropriateness of the *quality of life* sub-criteria. Another way is to assess whether the preference structure allows for a comprehensive number of **preference relations** to be indicated: is the decision-maker allowed to express indifference between criteria, if necessary?

A classic reference with regard to the comprehensiveness of a model is the paper by Phillips ([44]). According to Phillips, most analysts distinguish between two types of decision models, namely a descriptive model or a normative model. In broad terms, a descriptive model would describe the process that was followed to reach a decision, while a normative model would prescribe the best process to follow. Phillips indicate, however, that neither of these two types of models fully describe a model that explores the judgement applied during a decision, and the type of "social reality" or "socially shared view" created during this modelling. He suggests that a new type or class of model, namely a **requisite model**, be used to describe this kind of modelling. A model can be classified as a requisite model when everything required to solve the problem is represented in the model or can be simulated by it, and such a requisite model is usually created by the interaction between specialists who understand the modelling techniques and methods and problem-owners who add the value judgements and problem content. Typically, the problem owners are a group of people who bring their own views, objectives and conflicts into the modelling process. The process of putting together a requisite model uses any unease among the problem owners about the results of the current model as a signal that the model does not yet contain all elements, or do not yet represent all elements correctly, and therefore is not yet sufficient to solve the problem. Note that Phillips therefore also does not provide a simple and easy measure to apply to establish whether the model is complete, since he provides the satisfaction of the problem owners with the model as a guideline to whether the model is complete. Furthermore, Phillips also points out that a requisite model (i.e. a model that the problem owners feel represents the problem adequately) is a very conditional model, since any changes to the group of problem owners or the environment of the problem / problem owners, may create a new sense of unease with the model and therefore require further modelling efforts.

3.3 Issues specific to classification and sorting MCDA methods

It should firstly be pointed out that some issues studied within the general framework of all MCDA problems may be of less importance within the classification and sorting problematique — even though this may not be immediately evident from literature sources.

Firstly, it was mentioned above that this problematique does not require that one finds an overall best solution or set of solutions. This has the practical impact that the issue of 'dominance' (Pareto optimality), which is a very important concept within the general MCDA framework, takes on a different slant when applied to MCDA classification and sorting problems. In general MCDA problems, dominance is used to remove alternatives from consideration. In a problem where the aim is to select the best alternative, an alternative that is dominated by another alternative cannot be chosen as the best. However, in sorting and classification problems, alternatives are not directly compared to each other and dominance could provide information other than that required for eliminating alternatives. Greco, Matarazzo and Slowinski ([26]) indicate that if the dominated alternative is placed in a specific category, then it implies that the dominating alternative must be placed either in the same or in a "higher" category. The presence of dominated alternatives can either be used as information for assigning similar items together in one category, such as when the difference between dominating and dominated alternatives are small, or for guidelines on placing alternatives in different categories when such differences are substantial.

Secondly, where the fact that a decision-maker is indifferent between a few alternatives can create a challenge if one of these must be chosen as the best option, such indifference can be very useful in classification and sorting problems. Indifference can provide important information in terms of assigning alternatives into groups with similar characteristics. It should also be pointed out that classification and sorting methods can be distinguished from ranking methods in that similar alternatives can be assigned to a category without a need for specifying the order in which they should be listed in the category. For ranking problems the alternatives could first be assigned to ordered categories, but then there should be an additional indication of the position (ranking) that the alternative takes on within its category.

Thirdly, the fact that many alternatives are not comparable (compensatory) can create problems within choice or ranking problems. In choice or ranking MCDA problems non-compensation can make it difficult to decide whether one alternative is in fact preferable to another one. In terms of classification and sorting MCDA problems it would be possible to place such non-compensatory alternatives into different categories.

Within the sorting problematique, the main difference between methods for classification and those for sorting have to do with the categories into which alternatives need to be placed. In classification, these categories do not have to represent any specific order, while for sorting, the categories have some ordering implicit to the categories relative to each other. Perny (in [42]) provides the following three examples of sorting problems in which the categories have an ordering attached to them:

- the credit rating attached to persons applying for a financial loan
- evaluation of the efficiency of a number of manufacturing plants
- interpreting the marks of students in terms of their results (pass, fail, pass with distinction, etc.)

It should be clear that a student categorised into a "pass" category should have obtained "better" marks than one categorised into the "fail" category. Examples of classification problems in which there may not be any ordering, or not a complete ordering, between categories include the following:

- medical diagnosis
- pattern recognition
- assigning students to different groups for practical sessions

The fact that categories are not ordered does not mean that the categories are meaningless. Although limited, these examples give some idea as to the usefulness of classification into categories that are distinguished from each other even though not ordered from good to bad.

Why should the distinction between classification and sorting methods be important? An example of categorising students provided by Greco *et al* in [26], shown in table 3.2, illustrates the implication of an ordering of the categories 'good' and 'bad'. Students were given scores on three subjects (Mathematics, Physics and Literature) as well as an overall score. It can be seen that students

Table 3.2: Example of an MCDA classification problem

Student	Mathematics	Physics	Literature	Overall class
1	good	good	bad	good
2	medium	bad	bad	bad
3	medium	bad	bad	good
4	bad	bad	bad	bad
5	medium	good	good	bad
6	good	bad	good	good

2 and 3 received exactly the same scores on the three subjects, but their overall scores differed. This could mean that for these two students the three subjects did not provide enough information in terms of assignment to the correct overall category, or it could mean that there is an inconsistency in the assignment. An example such as this one would be a challenge to assign, irrespective of whether the problem is considered a classification or a sorting problem, since the information seems to be either insufficient or incorrect. However, a case like that of student 5 would be picked up as an inconsistency **only** if this is considered a sorting problem and **not** if it is considered a classification problem. In a sorting problem it could be picked up that student 5 obtained better marks on all three

subjects than student 3, but that student 3 ended up in a better overall category. This seems to indicate an inconsistency in the assignment. However, if this was simply a classification method, then ‘good’ would not be interpreted as **better** than ‘bad’, simply as **different from** ‘bad’ and this would not be picked up as an inconsistency. This example should therefore illustrate that the ordering of input information and ordering of final categories should have an impact on the interpretation of data and the algorithms used to do the assignment to the categories. It also illustrates that one may place different demands in terms of input information consistency (and output possibilities) between classification and sorting problems.

It should perhaps be pointed out that preference information such as criteria weights or preferences between alternatives are as important in MCDA classification and sorting problems as in other MCDA problems. As Perny points out in [42], one may have a fixed set of alternatives, scored against an agreed set of criteria and a pre-defined set of categories into which to assign the alternatives at the start of the problem. However, this information may not be sufficient in terms of doing the actual assignment to categories, since the criteria may conflict with each other and the scores themselves may not provide a clear assignment. The MCDA principle of bringing in the subjective preferences of the decision-maker in order to obtain the correct assignment, within the required context of time, problem and decision-making group, is as relevant in the classification and sorting problematique as in any other type of MCDA problem.

Chapter 4

Classifying MCDA methods

An important issue within the practical application of MCDA methods lies in how to select an appropriate MCDA method to use within a relevant practical problem context. Before one can select an appropriate method it is necessary to understand what characteristics of the various methods to consider when making such a selection. This implies that existing methods must be classified in some way to inform the process of method selection. This chapter considers the issue of classification of MCDA methods, with specific attention to MCDA sorting and classification methods, in some depth.

The idea of putting together a classification system to guide a user in selecting the appropriate MCDA method to use is not a new one. Li [34] refers to MacCrimmon in 1973 as being the first to work on ways to select the best method. Both Guitouni and Martel ([27]) and Ozernoy ([41]) indicate that there is no one 'best' MCDA method that will work equally well in all decision situations and that there is a need to determine what the most appropriate MCDA method is for any particular problem situation. Li ([34]) says with regard to MCDA methods that:

> Currently, over 70 decision-making methods have been proposed with the intention of facilitating the decision-making process, and have already been applied to deal with different decision problems. With the complexity of the decision problem and the demand for more capable methods increasing, new methods keep emerging. Paradoxically, these numerous methods don't ease the decision problem as they are expected to do, but complicate the problem because one has to determine which method is appropriate before he/she can proceed, considering the fact that the use of an inappropriate method may create misleading solutions to the decision-making problem.

This problem of a diversity of MCDA methods to choose from is echoed by Bouyssou *et al* [5] (as quoted by Guitouni and Martel [27]), while Hazelrigg [28] points out how using the wrong method could lead to wrong conclusions.

Hazelrigg provides the following example of three customers (I, II and III) rating products over three attributes (A, B, C). Each of the attributes can be offered in two ways (A_1 and A_2, B_1 and B_2, C_1 and C_2) and the customer had to indicate preferences in terms of the attribute instances. The preferences indicated by the customers are presented in table 4.1. If a method of selecting the best product

Table 4.1: Product design problem illustrated by Hazelrigg

	Attributes					
	A		**B**		**C**	
Customers	A_1	A_2	B_1	B_2	C_1	C_2
I	Hate	Great	Great	OK	Great	OK
II	Great	OK	Hate	Great	Great	OK
III	Great	OK	Great	OK	Hate	Great

design works in a way that combines the most preferred attribute instance, then the product designed on the basis of these preferences will be $A_1B_1C_1$, whereas if a method was used that selected the best product design as that having no 'hate' instance, but only having 'great' or 'OK' preferences, then the product has to have a design consisting of $A_2B_2C_2$. Using two different methods will result in obtaining two diametrically opposed product designs.

There has been a number of researchers who have done work in the area of classification of MCDA methods with the view of providing guidance on what method to use in a given problem. Some of this work is discussed in the following section.

4.1 Previous work on classification of MCDA methods

The literature references that were found on the classification of MCDA methods followed different approaches. However, there does seem to be a few broad groups that can be identified within these approaches.

A first approach is to provide a classification of methods based on a **qualitative grouping (verbal description)**. This approach was followed by Mendoza and Martins [37] and, although the direct reference was not found, seems to have been the approach by J. Deason in a PhD dissertation at the University of Virginia in 1984, as well as by M. Gershon in a PhD dissertation at the University of Arizona in 1981 (both of these as quoted in Al-Shemmeri *et al* [1].)

The *model selection paradigm* of Deason is quoted as based on a set of 'decision situation descriptors' which characterise multi-criteria decision situations, as listed in table 4.2. These descriptors are used to reduce the large number of available techniques to a smaller subset of appropriate techniques.

The method used by Gershon involved selecting from a total of 27 criteria (listed in table 4.3) the ones that are relevant to the problem in order to determine which method would be most relevant for the solution of the problem. The quotation by Al-Shemmeri *et al* [1] does not give many details, but it seems that Gershon's method may be used by an analyst to select the most appropriate

Table 4.2: Deason's decision situation descriptors

A	Finite set of discrete alternatives
B	Continuous alternatives
C	Ordinal attributes
D	Ordinal ranking of alternatives sought
E	Cardinal ranking of alternatives sought
F	Portfolio of discrete alternatives sought
G	Single stage decision problem
H	Multi stage decision problem with changing preferences
I	Large number of objectives or discrete alternatives
J	Need for highly refined solution
K	Decision maker reluctant to express preference explicitly
L	Decision maker experiences difficulty in conceptualising hypothetical trade-offs or goal levels
M	Decision maker preferences for marginal rates of substitution among objectives not independent of absolute levels of objective attainment
N	Need for decision maker understanding of method
O	Limited time with decision maker available

technique from a group of techniques known to the analyst.

In contrast to the amount of detail used in the methods of Deason and Gershon, Mendoza and Martins ([37]) used the high-level classification of methods suggested in [3], namely classifying methods into *'value measurement models'*, *'goal, aspiration or reference level models'* and *'outranking models'*.

A second approach that was followed in some of the references is to **compile a tree structure** in which the user will travel down a specific path, guided by questions at each node that provide the logical splits in the path, and to end up at a selection of the most appropriate method. This type of approach was apparently used in the earlier work by MacCrimmon in 1973 and Hwang and Yoon in 1981 (as quoted extensively by Li in [34]). More recently Evans [22], Sen and Yang (1998) as quoted by Li [34], and Guitouni and Martel [27] also provided classifications based on a hierarchical tree structure. A tree structure provides a clear and easy to follow path that explains the way a method can be selected. See, for example, an extract of the tree model by Guitouni and Martel in figure 4.1 on page 29. This extract illustrates the logic of their Multicriterion Aggregation Procedures (MCAP) classification method.

It should be noted that even though Guitouni and Martel give a so-called typological tree as partly illustrated in figure 4.1, they link the tree to a set of 7 guidelines as indicated by table 4.4 (slightly edited from the original). The article also gives comparison tables that provide details on both the guidelines and the information required at the nodes of the typological tree. It could be argued that the comparison tables in combination with the guidelines follows

Table 4.3: Gershon's model choice criteria and evaluations

Mandatory binary criteria (which are used to delete techniques for further consideration if not appropriate)	1. Handle qualitative criteria 2. Handle discrete sets 3. Handle continuous sets 4. Handle dynamic problems 5. Handle stochastic problems
Non-mandatory binary criteria (which do not necessarily delete techniques for further consideration)	6. Comparison to goal point 7. Comparison to aspiration level 8. Direct comparison 9. Strongly efficient solution 10. Complete ranking 11. Cardinal ranking 12. Ability to handle integer values
Technique dependent data (used to rate techniques subjectively on a 1-10 scale)	13. Computer time required 14. Implementation time required 15. Interaction time required 16. Decision maker's awareness 17. Consistency of results 18. Robustness of results 19. Handle group decision makers
Application dependent criteria (problem-specific criteria against which techniques are rated on a 1-10 subjective scale)	20. Number of objectives 21. Number of systems 22. Number of constraints 23. Number of variables 24. Decision maker's level of knowledge 25. Time available for interaction 26. Desire for interaction 27. Confidence in preference structure

a very similar approach to that of the verbal descriptor type of classification systems described above, with the typological tree prescribing the process of using the classification and not used as the only means of classification.

A third approach is to use **knowledge engineering / expert systems** as done by Ozernoy [41], Poh [45] and Li [34]. This involves developing a knowledge base of MCDA methods and their characteristics which is interrogated by some kind of if-then-else logic or inference engine providing the backbone to the selection of a method. Although it is possible to use a tree structure as the logical structure behind an expert system, it would be typical to have a more sophisticated logical structure. An expert system also has the built-in capacity to provide a trail of the logic that was used to reach the final decision so that the user can see the reasons why a certain method was selected as the 'best' to use. This can enhance learning and also help to establish the confidence of the user in the selection process.

As indicated by both Ozernoy [41] and Poh [45], there is quite a large range of different software packages available to implement MCDA methods, but this

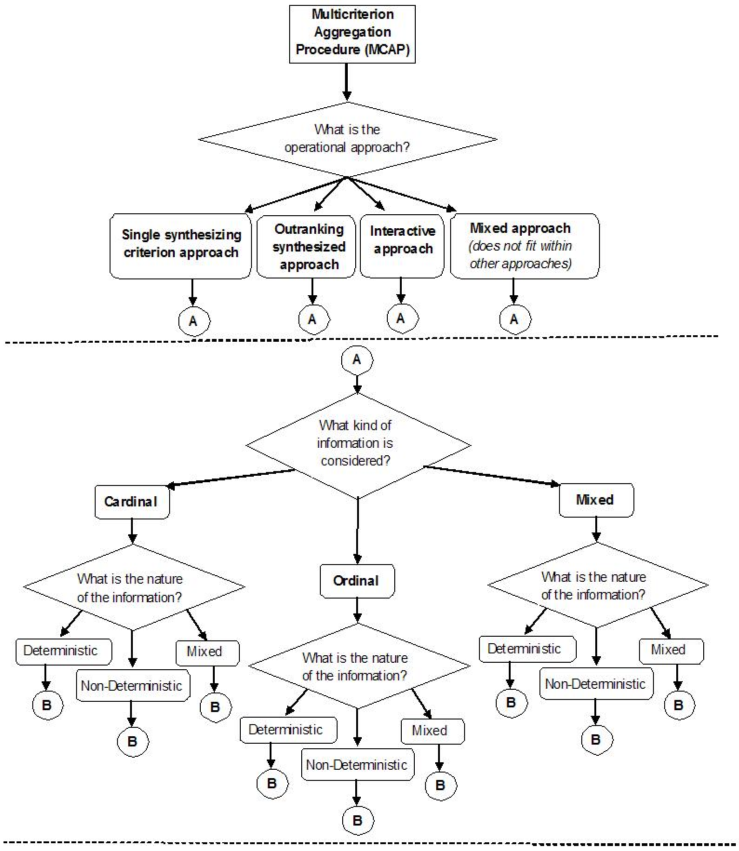

Figure 4.1: MCAP Tree model

software is difficult to apply for a user not familiar with all the different methods. They both believe that if an 'expert' classification system could be provided with such software, then the user can use this classification system to select the best method to use and then go on to use this method through the software. It would not only mean that the software that implement MCDA methods will now be more accessible to a user that may not be familiar with all the different methods, but also that the user will use the software in a more intelligent and appropriate way. They argue that it will help to ensure the successful application of software if the user is able to select the appropriate method to use, given the problem situation, and then be led through the method with the implementation software. The argument is that it will be in the interest of MCDA to have software that makes successful implementation of MCDA methods accessible to users who may have the problems but may not necessarily have the knowledge of appropriate solution methods. Of course, one may question whether they are right in believing that it will be an advantage to have users using MCDA methods without really understanding the methods and their underlying assumptions well, but there is validity in the argument that providing 'expert' advice on MCDA methods could be useful to users of those methods.

Table 4.4: Guidelines for selecting MCDA methods: Guitouni and Martel

Guideline 1	Determine the stakeholders of the decision process
Guideline 2	Consider the 'cognition' (way of thinking) of the decision-maker
Guideline 3	Determine the decision problematique pursued
Guideline 4	Choose the multicriterion aggregation procedure (MCAP) that can handle the input information and deliver the required information
Guideline 5	Consider the degree of compensation allowed by the MCAP
Guideline 6	Consider the fundamental hypothesis of the MCAP
Guideline 7	Consider the decision support system (software) available for the MCAP

Although Li [34] also developed his classification to use as a first step in a solution process, his aim was not in the first instance to provide an interface for software, but rather to provide a solid basis for the difficult area of engineering design. By applying a classification structure to select the most appropriate method for the engineering concept selection (the starting point of the engineering design process), he hoped to reduce the extent of the design process and provide a better foundation for the success of the total design process.

4.2 Discussion of the previous work on classification of methods

The verbal description type of classifications seemed to have been more predominant in earlier work in the classification field. While it seems as if the methods developed consisted of a very thorough overview of the main characteristics, there seems to be quite a few drawbacks to this classification method, including the following:

- A verbal method does not seem to have a clear and easy-to-follow structure: one has to read through the whole list of descriptions to receive guidance.
- There could be a problem with the level of technical language used in descriptions, since the person compiling the verbal descriptions often resort to technical language in order to keep the descriptors concise. Of course this helps the person using the list to read through all details faster, but it may make the meaning of some of the individual descriptors unclear and therefore more difficult to use.
- It is difficult to retain an overview of the whole list. This makes it difficult to judge if the list of descriptors is comprehensive and complete, especially with regard to new information or methods that become available over time.

The use of a tree method has an advantage in that it provides more structure and a better overview of the classification method, but it also has its own drawbacks. For instance, it seems to operate on a very linear approach to the classification (you cannot move down two "branches" at the same time, so a clear choice must be made at each node). Also, it seems to be built on the premise that there will always just be one 'best' method that will provide a solution within the problem context.

The expert systems approach, which seemed to be the preferred method in more recent works, appear to be the best compromise in terms of both providing a comprehensive base of information as well as providing structure and multiple "roads" to determining an appropriate method. However, an expert system would have to be kept up to date to remain useful, and this may not be easy to do — especially not for a person wanting to use an expert system developed by someone else.

4.3 Needs and requirements for a practical classification of MCDA methods

It is not realistic to expect that an OR practitioner who has some knowledge of MCDA principles would have such a well-grounded knowledge of MCDA methods that he/she would be able to identify the appropriate technique to use for *any* MCDA decision problem. In fact, Ozernoy [41] says that more often than not a practitioner will be familiar with a specific technique and apply it to a problem more due to familiarity with the technique than because they are sure it is the best way to provide a solution. This may in many cases not be a problem if the familiar technique is able to deal effectively with the demand of the problem. Even if there exists a technique that would be slightly better suited to the problem, one may get still get satisfactory results from using the 'familiar' technique if it is roughly relevant. However, it is possible that the results obtained from the 'familiar' technique may not be valid, or it may happen that during the course of the application of the technique one becomes aware of certain drawbacks to the technique when it is too late to revert to a more suitable method.

It is therefore possible to mention two important considerations in terms of a classification system of MCDA problems that will be useful to a practitioner. Firstly, such a system must supply information on a large number of techniques so as to be a comprehensive reference for those practitioners who are very familiar with a few methods and may not know of the relative advantages offered by other techniques. Secondly, information about techniques should be provided in such a way that the practitioner can at the outset of a project get an idea of which methods would be potentially suitable or completely unsuitable. A third, more subtle, consideration may be added to this last point. While much of the existing work on classification tries to find the 'best' method for a situation, it may in fact be a disadvantage to be led to the one best solution. It may be more useful to provide a practitioner with a (short)list of methods that will be suitable within the decision problem so that the practitioner can choose a method that is convenient from this list. For instance, a convenient method may be a method with which the practitioner is familiar, or a method preferred by the client, or a method for which there is software available. The practitioner can use a convenient method, yet have the assurance that the method is suitable and not *only* convenient, or can do further investigation and method selection within the subset of suitable methods.

It would be ideal to have the following three characteristics:

1. A comprehensive list of characteristics that describe the features of different MCDA methods and a very complete list of available techniques assessed against those features.
2. A clear and easy structure that would find methods in a way that is both fast and comprehensive. Note that this structure need not be graphical, but could be verbal.
3. Access to both the information base and structure logic in order to allow for easy updates to either of these.

The information and characteristics considered in the paper by Gitouni and Martel [27] seems to be a good starting point for base information, but it is not structured well. It is proposed that this be taken as a starting point, but that information and structure be added. Specifically, it is proposed that structure is added in terms of providing a better grouping of features and a list of questions to start the process.

It should be noted, however, that a limitation of any selection process used to find an appropriate method is that it **cannot** deal with all possible extensions of the various methods and all possible combinations of methods. While it would not be unusual for the 'best' method to apply within a practical problem to be a combination of various techniques (for instance, see comments by [37], [51], and [38] in this regard), it would be very difficult for a selection process to support the provision of such a combination.

4.4 Selecting appropriate MCDA classification and sorting methods

In order to consider ways of supporting the selection of appropriate MCDA methods, the problem of selecting an appropriate MCDA classification and / or sorting method was studied. Firstly, an understanding was gained of the various existing methods and how they differ from one another. This enabled an understanding of the issues to consider in distinguishing between the different methods. Secondly, this understanding was compiled into a number of specific aspects to consider in terms of selecting an appropriate method.

The details of these two processes are provided in the subsections that follow.

4.4.1 Understanding features of existing MCDA classification and sorting methods

From the literature a number of methods have been identified that are either aimed at, or could be used for, classification and/or sorting. The methods that have been found in the literature, as well as a short description of each one and a reference to more complete descriptions, are provided below:

- **Conjunctive and disjunctive methods:** The conjunctive and disjunctive methods of classification are discussed in [57] as methods to classify alternatives into binary categories (for example, yes / no, or acceptable / unacceptable) when measured against non-compensatory criteria. It should be noted that in a practical problem the one category may be preferred above the other, which would imply that these methods could be used for either sorting or classification. Also, these two methods require criteria measures that are either numerical (continuous quantitative) scores or ordinally arranged qualitative values. Both methods require the decision-maker to specify a performance level on each criterion that would distinguish between the categories (such as performance levels that would provide a cutoff between unacceptable and acceptable performance). The conjunctive method would require that an alternative meets or exceeds the performance level on all criteria, while the disjunctive method would require an alternative to meet or exceed performance on at least one criterion. It is also possible to use both of these methods together, whereby an alternative would be acceptable if it meets the required performance level on all of a subset of criteria (conjunctively) but exceeds performance levels on some of the criteria on a different subset of criteria (disjunctively). An example of a combined conjunctive / disjunctive method would be the evaluation of candidates for a position advertised as follows:
  A suitable candidate is sought for a position as a software engineer. The candidate must have completed at least a BSc (Hons) degree in Computer Science, with an MSc or MEng degree an advantage. At least 2 years of practical experience in C++ programming is required, as well as proven experience in 3D modelling packages such as 3DStudio, GMax or Blender.
  The conjunctive part is the minimum levels of qualification AND years of C++ experience that the candidate MUST have (note that in this case the minimum 'levels' imply an ordinal measure associated with any specific qualification), while the requirement that the candidate must have experience in at least one of three 3D software packages is a disjunctive rule (here a 'has experience'/'does not have experience' would imply an ordinal measure, since 'has experience' would be considered better than 'does not have experience'). A candidate with no experience in any of these packages will not qualify for this position, but the candidate does not have to have had experience in all three packages to qualify for the position.

- **Fuzzy conjunctive/disjunctive method:** A reference to a fuzzy conjunctive/disjunctive method that can be used for MCDA classification problems is provided by [27]. In [27] the method is described to be used for fuzzy data for which there is not a precise matching between the alternatives and the criteria specifications. The degree of matching between alternatives and criteria are calculated and used for classification purposes. The method is discussed in Dubois *et al* ([19]) as a weighted fuzzy pattern matching in which a datapoint can be matched against a required pattern. Since it is not a method that was intentionally designed for MCDA classification or sorting, one would have to interpret the degree of matching value (a value between 0 and 1, with 0 indicating no match and 1 indicating perfect match) in terms of the categories that one wants to use for classification.

- **Lexicographic methods:** This is a very simple method that can be used in situations where criteria are non-compensatory, but have a clear importance structure. The name of the method is derived from the word 'lexicography' that describes the way words are ordered in a dictionary. This method is discussed and illustrated briefly in [57, pp 22 – 25]. The method is used to solve an MCDA choice problem in the following way: all alternatives are evaluated against the most important criterion, and if one alternative is the best when measured against this criterion, it is selected. If there is no clear choice in terms of the most important criterion, then this smaller group of alternatives judged as 'best' (or equally good) under the most important criterion are then compared in terms of the next most important criterion, and so on until the best alternative have been chosen or all alternatives have been considered. If all alternatives are considered and no clear choice has been made, then the remaining group of alternatives are considered to be equally good or acceptable choices. Although the brief discussion in this reference does not elaborate on all possible uses of the lexicographic method, it seems possible to use the same importance order of criteria to subset the alternatives into different categories rather than simply choosing one best alternative. This seems especially true for the extension called lexicographic semi-order method (it is referred to by [57] as an extension to the lexicographic method and is described by Luce in *Econometrica* in 1956 and Tversky in *Psychological Review* in 1969). In the lexicographic semi-order method an alternative is selected as best only if it is significantly better than others on a criterion, where the significance of the difference is judged in terms of a quantitative threshold value. Such threshold values could lead to a number of alternatives being considered equal.

- **ELECTRE TRI:** This method is one of the group of ELECTRE outranking methods. Within this group, it has been specifically designed to deal with MCDA classification and sorting problems, whereas the other methods were designed more directly for choice or ranking problems. A great deal of literature can be found that refers to the theory or application of ELECTRE methods and the distinctions between the different methods. For instance, [47] indicates that the main aim of these methods is to derive an outranking relation that captures the preferences of the decision-maker as closely as possible. Such an outranking relationship does not just distinguish between 'preference' and 'indifference', but sees strict preference and absolute indifference almost as the extremes on a scale, and allows for preference expressions that fall somewhere between these two extremes. In [60] the outranking relationship is described as allowing the conclusion that alternative A outranks B (i.e. is preferred to) alternative B if there is enough evidence to conclude that A is at least as good as B (this is called the *concordance* relationship), and there is no reason to conclude that A is not better than B (called the *discordance* relationship). The ELECTRE TRI method requires that the decision-maker specifies categories into which the alternatives must be classified, and also provides quantitative 'profiles' as cutoff points between categories. Such cutoffs need not be specified exactly, but could be specified as thresholds that allow for, for instance, indifference between two categories. The ELECTRE TRI method classifies all alternatives according to the profiles of boundary values attached to the different categories, by using an outranking relation to compare the alternatives to those profile boundary values. The outranking relation can be used in various ways, for example, an alternative may be classified into a certain category if the alternative outranks the category's lower boundary value. Although it has been directly used on practical problems, the difficulty of obtaining all the preference parameters required for the ELECTRE TRI method to work has led to extensions in which the preference information is derived implicitly, by asking the decision-maker to compare a few 'reference' alternatives and deriving the preference model from that process, such as described by the article by Mousseau *et al* ([39]).

Note that the ELECTRE TRI method seems limited to a few alternatives and criteria, because the preference modelling is so intense and the methods are quite sophisticated. However, through the available software that supports the execution of the method, the method can be applied to more alternatives and criteria. The 'ELECTRE Assistant' extension described in [39] is also available for the ELECTRE TRI software.

- **IRIS:** Interactive Robustness analysis and parameter Inference Software (IRIS) was designed to implement the ELECTRE TRI method to solve a classification problem, but also to extend the method to allow a decision-maker more flexibility in terms of developing and testing the set of criteria and parameters to be used in the classification of alternatives. The aim is also to do a check on the criteria provided by the decision-maker in order to ensure consistency of preference and parameter information — it may therefore also be used to extract preferences of a group of decision-makers to reach a set of agreed criteria and parameters. For a discussion of the software and underlying theory, see [18].

- **Multi-criteria filtering methods:** Another method within the general group of Outranking methods is the so-called multi-criteria filtering methods as described by Perny in [42]. The method can be applied to both classification and sorting problems, which Perny calls, respectively, *filtering by indifference* and *filtering by preference*. The methods are applied to problems with a 'reference set' of representative alternatives and a pre-defined number of categories, and requires that weights, preference and indifference thresholds, as well as veto criteria be linked to all criteria. From this information, concordance, discordance and preference indices can be calculated, from which in turn membership indices are calculated to be used as a basis for classification or sorting.

- **PROMETHEE-related methods:** The PROMETHEE group of methods, described in terms of its general principles in [7], also include some classification and sorting methods. The PROMETHEE approach also uses outranking relationships and pairwise comparisons, like the ELECTRE methods, but replaces the concordance and discordance relationships of

ELECTRE with a *preference relationship* between alternatives for each criterion. The parameters required for these preference relationships should be easier for decision-makers to supply since they have a practical meaning, in contrast to the ELECTRE parameters that include some practical and some technical parameters — the technical parameters could be difficult for decision-makers to supply since it is not easy to understand their impact on the solution. A sorting method that has been developed based on the PROMETHEE approach is the **FlowSort** method as described in [40]. FlowSort determines reference profiles that can be used to quantify the comparisons between the different categories. Alternatives are assigned to a category based on a comparison of the alternative to the reference profiles of all the categories. The method has the advantage of being able to derive the reference profiles of the categories according to boundary information (i.e. upper and lower boundary limits) or centroid information for these categories. Another sorting method based on PROMETHEE is **PROMSORT** as described in [2]. PROMSORT also can use both "limit profiles" and "reference alternatives" as a basis for sorting, allows the decision-maker to define a pessimistic or optimistic point of view, and guarantees that categories will be ordered. Araz and Ozkarahan point out, however, that even though most of the parameters are meaningful values and can logically be determined in any decision-making situation, all these parameters must be supplied by the decision-maker(s): category profile values, weights, thresholds and a cut-off point for the 'distance' from the category profile limits that will determine into which category the alternative will be sorted.

- **Classification by Discordant Collective Preferences:** Details of this method is provided in [43]. The main aim of this method is to combine the preferences of a group of decision-makers in such a way that information of the individual preferences will not be lost. It pays specific attention to the fact that decision-makers do not just differ in opinions, but could provide inconsistent opinions (i.e. *discordant* preference data) due to errors or inconsistent or intransitive preferences. This method therefore keeps information on such discordances so that they can be highlighted and questioned. This is accomplished by using the mathematics associated with multisets in order to combine a set of group preferences in such a way that they are not aggregated into one single set. The author argues that a multisets approach not only keeps more information, but is also simpler to use when compared with some of the other more complex mathematical methods found in MCDA methods. The output is a set of decision rules (of the format IF ... THEN CLASSIFY AS ...) extracted from the combined preferences. Inconsistencies in preferences would be visible in the decision rules and can therefore be queried or accepted. Note that the term 'discordant' in the name of the method refers to the fact that the preference data can be discordant (showing errors or inconsistencies), and should not be confused with the "discordance" measure or indices calculated in some of the Outranking methods.

- **Rough sets theory:** This method, described in more detail in [26], was initially developed to focus on classification problems, but has also been extended to be applied to sorting problems. It assumes that categories are pre-defined and criteria to be used for classifying into the categories are specified. The method assumes that the decision-maker may have difficulty expressing his/her preferences directly, but may prefer to give sample classifications of a subset of alternatives that will implicitly contain preference information. The methods works at using such sample classifications to derive a set of IF ... THEN decision rules that explains the assignment process and can be used to assign remaining alternatives. The method has the advantage that criteria for the quality of the classification or sorting can be provided and that it can identify when certain criteria are not contributing additional information to the assignment process. The method can also handle the classification of alternatives even though information on some of the criteria are missing for certain alternatives, provided that not all alternatives have missing information. The author points out that they have found the output in terms of decision rules to be a useful format. Decision-makers they interacted with have found that it clarifies the assignment process in a way they find comfortable and understandable, and the decision-makers found it easy to pinpoint inconsistencies through incorrect or inconclusive decision rule specifications. Although not specifically mentioned in the article, the discussion of the method implies that it can be applied to non-compensatory or compensatory criteria and could include both qualitative and quantitative criteria as input data. There is software available to support the rough sets approach, for example the Rough Sets Data Explorer (ROSE) software available at http://www-idss.cs.put.poznan.pl/site/rose.html.

- **Stochastic Multiobjective Acceptability Analysis (SMAA):** This technique is perhaps not a true MCDA technique, but more of a way to test the sensitity of parameters. However, in [31], it is mentioned that it can be useful to explore the decision space in a (choice or ranking) situation in which a group of decision-makers may be able to measure alternatives against certain criteria, but would be unwilling to express their preference weights on the criteria either implicitly or explicitly. It is therefore included here for the sake of completeness. In order to use this method, scores for each alternative should be available for each criterion, and there should be an agreed way of calculating an overall score from the criteria scores (for example, an additive method). The first step consists of calculating the set of different weight vectors for each alternative that would make the overall score of that alternative greater than or equal to the overall scores of all the other alternatives. This is called the *set of favourable weight vectors* for the alternative and these sets are used to determine a range of measures for each alternative. These measures include an *acceptability index* which is a quantitative measure of the proportion of weight vectors which would result in that alternative having the highest score, as a proportion of all the possible weight vectors. An alternative with a zero acceptability index would never have a highest overall score, no matter what set of weights are selected. The set of *central weights* are also calculated for each alternative: it gives the 'typical' set of weights that would result in that alternative having the highest overall score. A variation of the SMAA method, called SMAA-TRI has been developed for classification problems, as described in [52]. This method is designed to use the classification method of ELECTRE TRI, but to use an iterative procedure to assess the influence of parameter values on the classification of alternatives. Acceptability indices are calculated for all pairs of alternatives and categories. This helps to provide insight into the impact of the choice of parameter values for ELECTRE TRI and also assists in finding classification solutions that are robust and therefore not as sensitive to changes to parameters. Since the SMAA methods are fairly computationally intensive, its application is potentially limited to "small" problems with not many alternatives or criteria.

- **Ordinal Classification method (ORCLASS):** This method, discussed in the article by Larichev and Moskovich ([32]) is specifically designed for sorting problems, since it relies on ordinal criteria information and ordinal categories. The method was designed to limit the number of alternatives that has to be classified directly. In fact, the method can select alternatives that have to be classified by the decision-maker from the total group of alternatives, based on an assessment of the information added by the classification of each specific alternative. The method also picks up inconsistencies or errors of the decision-maker in terms of the ordinal information, and provides some analysis of the eventual classification. Since the method is limited to ordinal data, it can be assumed that it only allows data that is qualitative and has an explicit ordering. Note that the method assumes that the decision-maker's preferences are represented by the criteria used as well as the ordering provided to criteria scores (or implicitly by the criteria score method used), and that weights are not assigned to the criteria themselves.

- **Goal programming methods:** While Goal Programming is a well-known technique within the MCDA field, it is typically applied to choice problems rather than classification problems. It has the advantage that it can deal with a large number of alternatives, but usually requires that the criteria be only quantitative data. There has been some work on methods using a Linear Programming (LP) formulation of classifying problems into categories, specifically classifying into two categories. Examples of such work can be found in [25], [14] and [50], however, it seems to be limited mostly to the classification of alternatives into one of two categories and seems to require quantitative or at least ordinal qualitative data as criteria measurements. The article [50] refers to the RAGNU software that has been developed to support classification into two groups using Mathematical Programming models and solution methods, but the software itself could not be located.

- **UTilitès Additives DIScriminantes (UTADIS) and extensions**: A description of this method is provided in [58] and this was used as a basis for this brief summary. The UTADIS method uses as input the pre-defined categories, the criteria used for classification into these categories and a collection of 'reference' alternatives that provide the basis for discrimination. No information is required in terms of criteria weights. It is not clear from the stated reference whether the method is intended for sorting or classification problems, but the example provided seems to be a sorting problem due to the implicit ordering of the categories and the criteria. The aim of the method is to derive an additive model with estimated parameters that serve as weights representing decision-maker preferences. The model is solved with linear programming techniques, based on the data provided on the group of reference alternatives. Since linear programming is used, the input data (criteria measurement scores) must be continuous quantitative data. The output is a parameterised model for classifying both the reference set, as well as any "new" alternatives, into the defined categories, in such a way that classification errors are minimised. Three variants of the general method has been proposed that focus on different aspects of the classification error:

1. UTADIS I that focuses on making sure alternatives are classified 'far' away from the cutoff values (*utility thresholds*) distinguishing between categories.
2. UTADIS II that aims to minimise the *number* of wrongly classified alternatives, as opposed to the *size* of the total misclassification errors. (This means that the method would prefer a model with one big classification error rather than a few small classification errors.)
3. UTADIS III which combines the objectives of I and II, aiming to ensure both a low number number of wrongly classified alternatives as well as alternatives classified well inside the derived category cutoff values.

Note that the article indicates that the UTADIS method can be applied using the PREFDIS software, but the software could not be located. Furthermore, the Multi-Group Hierarchical Discrimination (MH DIS) method described in [59] is also aligned with the UTADIS methods, and Zopounidis and Doumpos list some applications in which it performed better than the UTADIS methods.

- **Multiple Criteria Classification (MCC) method:** This method was developed to address classification problems specifically and is not designed to be applicable to sorting problems. The article by Chen *et al* ([13]) describes the motivation and theoretical underpinnings of the method. In order to apply this method to a classification problem, the alternatives to be classified must be known, the categories into which they must be assigned must be defined in terms of criteria to be used and the associated weights of the criteria for each category must be supplied. Each alternative must be scored in terms of all the criteria identified for all the categories. The main advantage of this method is that it can allow for situations where either all available alternatives need not be assigned into any category (the so-called *incomplete coverage* of the classification) and also for situations in which an alternative may be assigned to more than one category at the same time (the so-called *overlap* in categories). The reference article illustrates how the Simple Multiple Attribute Rating Technique (SMART) method can be used to do a classification of alternatives and how different classifications can be obtained by specifying a minimum number of alternatives to be assigned to a specific category (as such, the reference implies that the method applies to compensatory criteria only, since SMART is an additive weights method applied to compensatory criteria choice problems). The method prescribes an analysis procedure in the form of a flowchart, of which the output is a classification matrix indicating the assignment of an alternative to a category. Information about which alternative is classified into which category and how many alternatives are classified into a specific category is easy to obtain from the solution matrix: the matrix has one row for each alternative and one column for each category, and indicates with a 1 if the alternative is assigned to the category and a 0 if not assigned to the category. A possible criticism of the reference article is the fact that the analysis procedure allows for an iterative post-criteria assessment as well as an iterative post-optimisation assessment, but while two quantitative measures are suggested for the post-criteria assessment, no such quantitative measures are provided for the post-optimisation assessment, implying that the assessment of the final solution is a very qualitative one.

- **MCDA clustering:** In an article by Malakooti and Yang ([35]), a technique for clustering multiple criteria alternatives is discussed. This method requires the decision-maker to use reference alternatives and pairwise comparisons to classify alternatives into groups. A set of criteria and their measurement levels to use as the basis for assignment is built up through the pairwise comparisons and alternatives are classified based on their distance from these criteria measurements. Since this is a clustering method, the categories into which the alternatives have to be classified does not have to be pre-specified.

- **Discriminant analysis**: This is a well-known multivariate statistical method for classification of objects into pre-defined categories. Detailed discussion of this technique can be found in various statistical text books, such as a very accessible description in chapter 7 of [36]. It aims to find an additive function that can best separate alternatives into different categories, trying therefore to provide a description emphasising the differences between the different categories. It is a method that is aimed at classification rather than sorting. As pointed out in [58], however, discriminant analysis requires that the criteria used for classification shlould be normally distributed, as well as that the within-category covariance matrix is the same for all categories. Although the method can be used on data that do not meet these requirements, the significance tests associated with the method cannot be expected to work well, so that final conclusions about the validity or optimality of the classification cannot be made. This could be a drawback in practice, especially if the criteria represent qualitative information that are typically not normally distributed.

From the discussion of the various methods it can be noted how similar the requirements are for the different methods. Some of these similarities are because the methods are linked in some way (at least philosophically), but some of them relate to specific issues within the field of MCDA classification and sorting.

It should be noted that most of the methods assume that the categories are pre-defined, but this is an underlying principle of these types of problems — they are classification and not clustering methods (of the type discussed in [17]), so that the aim is not to find groupings within a set of alternatives, but to assign them into known groups. However, what is interesting is the fact that most of the methods are solution methods, with aspects about designing criteria, determining criteria measurements, assigning weights and resolving areas of conflict between decision-makers not being paid much attention in the literature references. Also, all of the classification and sorting methods assume a stable context with no uncertainty or multi-stage decision-making specifically modelled.

It can also be noted that most of the references do not discuss how sensitivity analysis is supported by the various methods, but this could partly be due to the fact that there is not a great deal of attention paid to determining a measure of the quality of the sorting and classification. (Of course, there are some exceptions such as the UTADIS, MCC and rough sets approaches.) Since there is no easy way to determine how good a sorting and classification result is, it is problematic to follow a structured process of sensitivity analysis, and any scrutiny of the results or sensitivity analysis seems to be mostly judgemental and qualitative.

4.4.2 Choosing appropriate MCDA classification or sorting techniques

It was decided to select appropriate classification or sorting techniques by asking a few simple questions. Based on the answers to these questions, techniques can be selected that would potentially be appropriate for the specific context. A spreadsheet was compiled to test whether a selection process based on such key questions could work.

A list of questions were compiled, using the literature references discussed earlier in this chapter, specifically [27], as a starting point. This list of questions was then tested on the listed classification and sorting techniques. The questions were implemented in a spreadsheet containing the following worksheets:

1. The complete set of information to be used in terms of available methods and how they measure against different criteria. The aim of this sheet was to provide an accessible way to update / maintain data or to obtain detailed information about methods.
2. A simplified structure that helped to select a method, i.e. the information in the detailed sheet in a shorter and more accessible way.
3. A list of questions in simple, everyday language that were based on the simplified structure but would be an easier method to use for practitioners or operations researchers who are not very familiar with all MCDA details and terminologies.

Even though the spreadsheet tested this way of selecting appropriate methods in a very rudimentary way, it was found that using the list of simple questions could work and could provide a practical way of distinguishing between the various methods. The questions that were found to be useful for identifying an appropriate method are listed in Table 4.5. The table lists the questions, and the possible answers to the questions, that were used in the spreadsheet.

An issue not addressed by the questions in table 4.5 relates to limitations in the number of criteria and / or alternatives that each method can handle. This is an aspect not discussed in very much detail in any of the literature references, and detailed investigation of each method under various conditions and for various problem situations would be required to provide estimates of such limitations (if any). Since such a detailed investigation is outside the scope of this study, it is assumed that none of the methods have restrictions in terms of the number of alternatives or criteria measurements that can be handled. Also, all the methods assume a stable environment and no conflicts between the decision-maker(s), therefore a distinction between the methods in terms of this issue has not been included in the table.

Table 4.5: Choosing an appropriate classification / sorting technique

Topic	Question	Possible answers
Decision-maker	*Agreement required in group	*yes OR no
Problem	*Problematique *Number of categories	*classification OR sorting *two (binary problem) OR no limit on number
Input data	*Preference information required *Can classify alternatives with "missing" criteria data *Criteria measure type *Criteria data assumptions	*reference group of alternatives selected by decision-maker OR reference group of alternatives selected by method OR pairwise comparison of alternatives OR criteria structure OR criteria structure and parameters *yes OR no *qualitative categorical OR qualitative ordinal OR quantitative OR quantitative or ordinal OR any type *normally distributed OR non-parametric
Criteria	*Requires weights *Compensatory criteria	*yes OR no *compensation possible OR non-compensation OR compensatory or non-compensatory
Method	*Approach *Complexity *Checks consistency of preferences *Finds "irrelevant" criteria *Allows some alternatives not to be classified *Allows alternatives to be classified into more than one category	*classification rules OR outranking relationships, classification matrix OR list of classified alternatives OR optimised model *complex mathematics OR not complex *yes OR no OR not completely *yes OR no *yes OR no *yes OR no
Outputs	*Measures classification quality *Provides classification summary	*yes OR no *yes OR no
Software	*Software available	*yes OR no OR can be done on spreadsheet

Chapter 5

Assessment of criteria and criteria measurement

The field of MCDA, as any other field within OR or general quantitative modelling, faces certain difficulties in the modelling process. For instance, how does one strike the correct balance between a *simplified* version of the situation being modelled and including *sufficient detail* for a realistic representation of the situation being modelled? Or, should one place more emphasis on a *description* of the problem situation, or on the development of *solution algorithms* to address the problem situation?

Furthermore, modelling and measurement within the MCDA context faces complications that may not be present in other types of quantitative modelling, namely that the aim is to model the subjective **preferences** of a decision-maker. Modelling such preferences is different from modelling a "real system", as pointed out by Brownlow and Watson ([9]) and Belton and Stewart ([3] [pp. 80 – 81]). When a process or system that can be observed "in the real world" is modelled, it is possible to test and validate the performance of the model (at least to some extent) against the system that it is supposed to represent. However, the preferences of a decision-maker or group of decision-makers do not "exist" in the same way as a practical process or system exists. In fact, the modelling may involve constructing a model of preferences that a decision-maker may not be aware of him/herself, or may have known about but never had to express in a clear, consistent and logical way. It is therefore not possible to measure how well the model is constructed by comparing it to any existing "real" system. Measurement of the correctness of an MCDA model often requires a qualitative assessment of the extent to which the decision-maker is satisfied with the model, or has gained insight and guidance from the model.

A further complication in MCDA modelling is the intent to model the **subjective** nature of the preferences of the decision-maker. Such subjectivity is not just an unfortunate by-product of MCDA modelling, but is crucial to the problem context being modelled. For instance, in the classical MCDA problem of choosing which car to buy (the example provided in [4] and [20], amongst others), one should realise that there is no one correct answer to the question, i.e. there is no car that will always be the correct choice no matter who the decision-maker. This is because the 'correct' answer will depend on the individual decision-maker. Whereas a roomy station wagon with good petrol consumption and resale value may be the best choice for me, the best choice for you may be a yellow sports car. Because my environment and related point of view is different from yours, our subjective preferences and requirements in terms of selecting a car does not just happen to be different, they **should** be different.

Finally, it should be noted that preference modelling is mostly subject to uncertainty and is limited to a specific time **context**. As indicated by [24], one may not always have an exact subjective judgement or preference, and, furthermore, one's preferences may change over time due to improved knowledge or changing circumstances. For instance, predicting the likelihood of a future event is difficult, and such a prediction will almost definitely change over time. Also, decision-makers within a company may have a certain current preference based on the company's current profitability as well as the current external economic conditions, but these preferences could change given a more or less positive environment.

In summary, when studying the requirements for constructing measurements in the MCDA field, it is important to take cognisance of general modelling principles, as well as issues that are particular to the MCDA field. It is important to know that it may not be possible to validate MCDA measurements against the "reality" it aims to measure, and that MCDA modelling has to capture both the subjectivity as well as the context, in terms of time and environmental conditions, of decision-maker preferences.

5.1 Constructing criteria measurements

In MCDA modelling there are three main challenges, namely: identifying alternatives, quantifying preference information, and determining and measuring appropriate criteria. While not underestimating the importance of the first two of these challenges, this section deals specifically with the measurement of criteria as a crucial aspect of the MCDA modelling. Even if the purpose of the MCDA modelling is more descriptive than analytical, the process of attaching measurements to criteria may help to refine the definition, meaning and impact of the elements included in the model: areas of ambiguity may be seen and discussion initiated to resolve such ambiguity (also see comments by [3] and [30] in this regard).

As indicated above, MCDA modelling has to meet some particular demands, in addition to the general problems experienced in quantitative modelling. Since criteria measurement is an important part of this modelling process, this principle also applies to the selection or construction of criteria. This section focuses specifically on criteria measurement: how to ensure that the criteria used within the model are appropriate representations of the context of the problem and are able to capture the subjective preferences of decision-makers.

There are, firstly, some general guidelines in terms of measurement that are applicable to all quantitative measurements, such as the principles provided by Raisbeck [46] when describing the way to derive measures of effectiveness. He states that there are four ways in which such measures may be constructed:

1. From analysis: If the aspect that is to be represented by the measurement is well understood, or has been analysed on the basis of data, this analysis or understanding can form the basis of the measurement.

2. From statute: There may be prescriptions based on legal or policy requirements that could be used as a measurement.

3. By consensus: Within a certain industry, there may be a consensus prescription in terms of measuring something.

4. As prescribed by client/sponsor: It may be that the decision-maker or sponsore requires a certain measure to be used.

As Raisbeck also illustrates, there may be problems with measures that are prescribed by the client or by consensus in that the measurement may be manipulated without actual improvement in the underlying system, or that it is not an entirely relevant measure. Also, measures derived from statute may in fact be open to interpretation. For instance, the effectiveness of a public utility may, according to statute, be measured in terms of profitability. However, whether profitability is defined as return on capital investment or as profit as a percentage of payroll or sales could make a difference in how these public utilities are measured — and in how they decide to operate in order to improve the measurement.

Whether one constructs a new criterion for measurement in one of these four ways, or uses an existing criterion, attention must be paid to the content of the measurement to make sure that it is useful within the project context. This is even more crucial when such a measurement is to be used as a criterion within an MCDA model. Not only must one consider whether the measurement is an appropriate representation of the aspect that has to be measured, but one must also ensure that it is a correct representation within the added context of the MCDA problem situation. Is the criterion able to measure what the problem requires and can it be used to express the preferences of the decision-maker adequately?

The difference between general measurements and those used in MCDA models would mostly be due to requirements that the measures be appropriate given the context of time, environment and decision-maker subjectivity.

Useful hints on the structuring of criteria for use within MCDA methods are provided by Buede ([10]) as well as Brownlow and Watson ([9]). Both of these articles favour the construction of *value attribute hierarchies*. Brownlow and Watson motivates this by the fact that people tend to not make good decisions based on large quantities of information, or tend to want to simplify a complex problem in some way in order to make a decision. Such simplicification may disregard important pieces of information. By presenting information in a hierarchy, large amounts of information are **structured** in a way that makes the information seem simple and accessible, but at the same time prevents important pieces of information from being excluded. In addition, a hierarchical presentation has the following advantages:

1. It improves clarity: a hierarchy provides a decomposition of the problem and linkages between issues in a clear and simple way that most decision-makers find easy to understand.

2. It can integrate work from different groups: Often a problem consists of a few distinct areas, and different groups of people may have expertise within these areas. For instance, information about costs of certain options would be provided by one part of the organisation, information about technical product quality by another group and information about staff impacts by yet another group. The hierarchy makes it clear how the information by the different groups link together and also allows the different groups to work separately on refining their own aspects while maintaining the overall context.

3. It allows trade-offs to be made clear: The level at which certain criteria appear often provides a clue on trade-offs. An example of this is that it is typically easier to make trade-offs between the more technical criteria at the bottom of the hierarchy than between the more strategic criteria at the top of the hierarchy.

Buede suggests using either a *top-down* (objective driven) or *bottom-up* (alternative driven) approach to the construction of criteria hierarchies. With the top-down approach, one starts at the "top" of the hierarchy, with the main objectives. A process of identifying subobjectives linked to main objectives are followed and attributes are linked to all objectives, sub-objectives, sub-sub-objectives, and so on. This approach is ideal for situations where the problem is of a more strategic nature and the decision-maker needs help in defining the problem more than in judging a set of (readily available) alternatives. The bottom-up approach, by contrast, is more suitable to tactical or operational problems in which some alternatives have already been identified and the decision-makers feel they have a relative good understanding of the problem. It involves starting from the identified group of alternatives, grouping them together in terms of decision-maker preferences and building up a hierarchy representing these groups and the measurements attached to them. The bottom-up approach has the advantage that it is clear when the hierarchy is completed, while the "end" of the top-down approach is not as clear. Buede stresses, however, that although these general approaches may help to guide the construction of criteria hierarchies, care should be taken to express the correct problem context and that there are no "value structure templates" that can be used repeatedly on different practical problems.

Even more detailed and useful hints for MCDA criteria are provided by Keeney ([30]) in chapters 4 and 5 of his book. He indicates that there may be three types of ways to measure criteria, namely as natural measurements, proxy measurements or constructed measurements (note that Keeney calls these 'attributes' but, as explained in chapter 2, the more general term 'criteria' will be used in this document). The following provides a short summary of what is meant by these three types of measures:

- Natural measurements refer to measures that are obviously attached to a criterion. For instance, if the criterion is to minimise costs, then the natural measurement for this criterion is "costs measured in rands/dollars/euros". Obviously, if a criterion has a natural measurement that can be attached to it, this should be used even though it may not be very straightforward to obtain the measure. For instance, if the criterion was to minimise manufacturing costs, one might have to develop a definition covering the calculation of such costs, specifying over which period it is measured, whether it includes both direct and indirect costs, and so on.

- Proxy measurements refer to measures that do not directly measure the criterion, but that measure a related aspect which can be taken as a surrogate for a direct or natural measure. For instance, Keeney uses the example of damage to buildings in a city due to acid rain. Although the aim might be to have a measure of the amount of damage done to buildings by acid rain, there is no natural measure to be used for such damage. In such a case, a proxy measure might be to use the concentrations of sulphur dioxide in rainwater (which forms acid rain) to indicate the *potential for damage*. The main advantage of using proxy measurements is that they may consist of existing information or information that would be possible to measure in situations when natural measurements would be difficult, costly or even impossible to obtain. The disadvantages of proxy measures all centre around the issue that the proxy measure is a not direct measure of the aspect one wants to measure. Sometimes there is not such a strong direct relationship between the proxy and the 'real' measure, or other aspects may be included in the proxy measure that is not very relevant to the context. Also, it may be problematic to determine the weight that should be attached to a proxy measure, when the weight attached to a natural measure would be easier to determine.

- Constructed measurements refer to a measure designed for a criterion which does not have a natural or proxy measure. Although one can "construct" a natural or proxy criterion, in general a natural or proxy measure can be used for more than one purpose, and is not only relevant to the problem situation being modelled. The constructed measurement, as Keeney defines it, is specifically designed as a measure within the context of the problem situation and may not be generally applicable to other situations. For example, Keeney mentions a constructed measurement that they defined to measure the objective "maximise public receptivity to the proposed power plant site" for a project to assess potential sites for power plants. The measurement consisted of a verbal scale, translated into categories ranging from 1 to -3, that describe public perceptions in a way that was useful to the project. A constructed measurement therefore may have an in-built subjectivity related to, or even limited to, the problem context under study.

Constructed measurements could be particularly useful in an MCDA problem, since criteria often do not have any natural measurements or obvious proxy measurements, or there may be natural or proxy measures but data for these are not available. Constructed measures could also have the advantage that they fit very well into the context of the problem. Not only will a constructed measure provide a direct measure of the relevant physical entity, but it will contain the appropriate subjective interpretation within the model context. It should be noted, however, that constructed measures should conform to what Keeney calls **measurability**, **operationality** and **understandability** in order to be used in an appropriate way in the MCDA modelling process. The requirements for measurability, operationality and understandability lead to the following set of recommendations for constructed measurements (note that some of these could apply equally to natural and proxy measures even though they are presented here in terms of constructed measurements):

- The aim should be to have a *direct* measure of a criterion. It is better to use a more rough (qualitative) direct measure than to have a very accurately recorded measure that only describes part of the criterion or only indirectly affects the specific criterion. Even a simple dichotomous measure (such as yes/no or acceptable/not acceptable), when measured directly in terms of the criterion, would be better than a measure that cannot be directly interpreted in terms of the criterion.

- It is preferred that measures for different criteria should be *independent* in the statistical sense. If two criteria measurements are dependent in the sense that more of one can compensate for less of another one in order to achieve the same effect, it could potentially be better to combine those two measurements into one single measure that describes the overall effect of interactions between them rather than to have two separate measurements of which interactions have to be modelled additionally in some way.

- The measurement should be constructed in such a way that preference judgements can *clearly be concluded* on the basis of the measurement. For instance, suppose that the criterion to be measured is in terms of the levels of a certain chemical pollutant in a water source. A simple measure such as parts per million could be used to measure this criterion, but will it provide the necessary information? If certain concentration levels (cutoff points) are known to cause different health or environmental impacts, then the measurement must be done at a resolution (i.e. in units and range of values) that will allow interpretation of the measure in terms of those cutoff points. Information about the impacts of a measurement, as well as the value, preference and trade-off judgements to be made on the basis of a measurement are required if the measure is to be useful in terms of MCDA modelling.

- The measurement should be defined *unambiguously*, which means that a person providing a value for the measurement and a person(s) interpreting the value of the measurement must get the same information and impression from the measurement. Specifically, Keeney advises against turning an actual measurement (such as number of fatalities recorded) into categories such as 'none', 'low', 'moderate', 'high' if such categories have no clear definition or interpretation within the context of the problem. Such a re-categorisation simply results in loss of information and does not add value to the MCDA model.

- In defining the measurement, care should be taken to include *complete measurement details* in the definition. Defining a measure as "pollution levels" is not sufficient, since details must also be provided in terms of where, when, by which method and in which units such a measurement should be recorded.

- A measurement does not need to be provided as a single value, but can be given in terms of a *probability distribution.* If there is uncertainty about the measure, especially if future values of the measurement need to be taken into account, then it is better to model the values as a probability distribution than, for instance, as categories like 'probable', 'not probable', etc. Possible consequences of a decision could also be measured with a probability distribution.

It is quite important that criteria measurements not only meet strict requirements in terms of measurement accuracy, but that they should also contribute to the information content, interpretability and subjective decision-making context of the MCDA problem being modelled.

In cases where the values and judgements of the decision-maker are difficult to articulate, a useful way of identifying and constructing qualitative scales for measuring criteria is provided by the method of repertory grids as described by Eden and Jones ([20]). Repertory grids are well-known in the psychological field, and are used to capture so-called 'constructs' people use to interpret their world. Eden and Jones demonstrate how this can be used to determine elements that the decision-maker considers meaningful to him/her/them and that is relevant to the decision problem at hand. The repertory grid method consists of presenting the alternatives of the MCDA problem to the decision-maker in groups of three, and asking the decision-maker to motivate similarities and dissimilarities between the alternatives. Descriptions of such similarities/dissimilarities then lead to identification of the elements that the decision-maker uses in order to make judgements about alternatives. It also provides some rough ideas on how these elements are measured, since the decision-maker has to give some kind of scale of similarity and dissimilarity used to compare the three alternatives to each other. Although Eden and Jones provide an explanation of how the resulting grid is used for analysis, the description is not complete, and it seems as if the main usefulness of the repertory grid approach is to determine criteria that must be considered. It seems as though it provides a useful starting point from which a more formal preference structure (hierarchy and weights) and analysis of alternatives can be derived.

5.2 Handling uncertainty in measurement

In a typical MCDA problem, uncertainties of various kinds can impact on the modelling process in general, and specifically on the construction and validity of measurements. An article by French ([24]) lists the following possible areas of uncertainty that could affect modelling and/or analysis:

1. Uncertainties expressed during **modelling** could include the following aspects:

    - Uncertainty about what might happen or what can be done

- Uncertainty about meaning (ambiguity)
- Uncertainty about related decisions

2. There could be uncertainties expressed during **exploration** of models such as:

   - Uncertainty arising from physical randomness or lack of knowledge
   - Uncertainty about the evaluation of future beliefs and preferences
   - Uncertainty about judgements, e.g. of belief and preference
   - Uncertainty about the accuracy of calculations

3. Uncertainties which typically are expressed during **interpretation** of model results include these ones:

   - Uncertainty about the appropriateness of a descriptive model
   - Uncertainty about the appropriateness of a normative (analytical or prescriptive) model
   - Uncertainty about the depth to which to conduct an analysis

French also suggests ways of dealing with these listed uncertainties. Some of the uncertainties can be dealt with in an appropriate model, such as uncertainty about future events that could impact on a decision, which can be handled with probability or utility theory. This type of model may be be mathematically complex, but can be done. Also, fuzzy theory or qualitative measurements can be used to capture measurements that are not possible to determine precisely.

Dealing with uncertainties about calculation methods (such as a complex algorithm), is more problematic. Although taking care in the construction of measurements and checking the effect of calculations on the values of such measurements may help in some way, the MCDA practitioner should keep in mind that there can often be uncertainty about a solution, even if such a solution was carried out with a sophisticated computer algorithm.

In terms of uncertainty about related decisions, there could be more than one way to deal with this issue. If further decisions that follow the current one have to be made by the *same* group of decision-makers, one can consider constructing a model that includes the current and follow-up decisions in order to find a solution that will support both current and future decisions (for example, using dynamic programming). If any follow-up decisions are *not under the control* of the group taking the current decisions, then one has two possible options. Such follow-up decisions may be modelled by game theory where the different role players' actions can be represented. Alternatively, one can investigate ways of reducing uncertainty about future actions, for instance by negotiating contracts in which different parties agree to particular actions. This last example, namely a negotiation, indicates that creative solutions which extend beyond the simple modelling and analysis of the MCDA problem itself and into the environment of the problem situation could be useful when faced with uncertainty or conflict.

He also argues that uncertainties which have their roots in ambiguity or uncertainty of consequences need not be modelled as uncertainties, but must in fact be investigated in more depth. Ambiguities in definitions need to be resolved in terms of what it means or is taken to mean within the particular problem context. Uncertainty around the impact of certain new regulations or legislation may be investigated in more depth: what possible impacts could there be and how severely would the impact be on the current decision?

Although it may not be possible to model uncertainty about quantitatively expressed judgements directly (for example, should this criterion have 70% or 75% of the weight?), it is better if this is not considered as impacting the measuring and modelling *per se*. It is preferable to construct a model using a stated judgemental value, analyse the model and then do sensitivity testing on the effect of changing the value. Also, if there is uncertainty about whether a specific normative model is appropriate, then sensitivity analysis may have to go as far as testing the impact of using certain models as opposed to others.

To some extent sensitivity analysis may also be used when there is uncertainty about future changes in preferences or knowledge, since it could help to provide confidence in the validity of a current solution. However, French does not consider sensitivity analysis as a complete solution to this problem, and in fact considers this issue as requiring further research.

Finally, one may have to realise that there are certain types of uncertainties that cannot or should not be modelled. If the decision-makers are uncertain about whether they have considered all possible solutions or taken into account all environmental factors or measured all relevant criteria, or if they wonder whether the descriptive model presents an accurate picture of the reality, or if they worry that a very different normative model should perhaps have been used, they may just have to accept that such uncertainties will always exist. Also, a decision about when to stop further refinements to the model, or when to use more in-depth measures may be impossible to determine in any objective way. To a certain extent, it is only by relying on the confidence and satisfaction of the decision-makers that one can address such uncertainty. If the decision-makers feel that they are comfortable with the solution, have learned from the process of constructing the model, or have covered as much as they can think of, then the modelling has been successful.

One should realise that all models and measurements, including those used in MCDA modelling and measurement, are limited in what they can achieve and one should be careful of "selling" MCDA as the answer to all problems and the way to deliver the best and only solution.

5.3 Validation and final choice of criteria measurements

Once an initial set of criteria have been compiled as an outcome of the problem formulation process, the set of criteria should be assessed against certain requirements in order to be refined into a final set. Again it is important to assess the measurements both in terms of general modelling principles as well as against the specific needs of MCDA modelling.

5.3.1 Assessing the validity of measurements

The first step in validating the set of criteria measurements is to assess them mathematically, i.e. in terms of what exactly each one measures.

For instance, examples are provided by Bouyssou *et al* in [4] of how the construction of measures can limit their use in certain decision situations. They discuss the use of weighted sum measures in chapter 2 and also list the following aspects to consider in assessing a constructed indicator/index in chapter 3:

- Scale normalisation: Whenever a measure is 'normalised', i.e. adjusted to fit within a useful range (for instance to take on a value of between 0 and 1), attention should be paid to the way that normalisation was done. An example is given of the Life Expectancy Index (LEI) used within the Human Development Index (HDI) in which a country's life expectancy is normalised against a minimum of 25 and a maximum of 85 — but these minimum and maximum values seem to be arbitrarily chosen according to the perceived maximum and minimum values at a certain point in time. Unfortunately, it has since happened that Rwanda recorded a life expectancy value of lower than the minimum of 25, which would result in a meaningless LEI figure.

- Scale construction: There are many ways of combining different measures into into a single index, and often the measures are manipulated before combination. Attention should be paid to whether this manipulation is meaningful and to whether it has the potential to be distorted — both in the manipulation of the individual measures and in the way the measures are combined after such manipulation to form the index. For example, a linear combination of measures that could take on values over different scales may create a bias in the index, but normalisation of the measures before combining them into an index may not necessarily solve the problem. Normalisation of such measures would imply that a greater improvement is required in a measure that can fall within a wide range than for a measure that can fall within a narrow range in order to create an improvement in the overall index. In fact, Yoon and Hwang ([57]), on the same topic, suggests that it may be better to aggregate such measures by multiplying them, instead of normalising and then adding them, to create the index.

- Compensation: In combining a number of measures into an overall index, one usually creates a situation in which a low value for one measure may be compensated for by a high value in another measure. If this is done intentionally because the different measures are comparable and do in practice compensate for each other, such an index can be useful. However, there may be situations in which measures that should not be directly comparable or compensatory are combined into an index. Again, the HDI provides a useful example. The HDI consists of components for life expectancy, literacy and Gross Domestic Product (GDP), and one may argue as to whether they are in fact compensatory in the sense of how far a reduction in life expectancy can be "traded" for increases in GDP.

- Dimensional independence: When it makes sense to allow compensation between measures within an index, one should ensure that the index reflects the correct balance in the amount of improvement/reduction in each of the different measures that provide the compensation. For instance, should one measure be required to double in order to make up for a one unit reduction in another measure, or should similar sized changes be required for compensation? Such a balance in the compensation units is called dimensional independence and should be done to validate an index. Note, however, that dimensional independence is often closely linked to the scale normalisation or scale construction issues discussed above.

- Aggregation issues: It should be noted that an aggregation of averages or sums must be taken as a theoretical value and that an index compiled from such aggregated measures may be difficult to interpret directly. Again taking life expectancy as an example, the fact that the average life expectancy of a country has improved does not mean that the life expectancy of every citizen of the country at the individual level has improved equally. One should ensure that improvement in components of the index affects the value of the index, but at the same time one should be careful in interpreting the value of the overall index directly in terms of changes in the underlying measures making up the index.

- Monotonicity: An index may be compiled from different measures in ways that are not directly linked to the individual measures themselves. An example is the ATO air quality index used in France to report on the status of air pollution on any specific day. This index combines a qualitative interpretation of the status of four different pollutants into one index. This means that an increase or decrease in the underlying concentrations of these pollutants are not directly translated into a similar sized change in the overall index. Such an index construction can have certain advantages, but must be interpreted in a different way to an index that reflects changes in the underlying measures in a direct monotonic way.

In summary, one should understand what an index measures, how it is constructed and what the valid range of values for the index and its underlying measures can be in order to use and interpret the measure in the correct way. Care should be taken to assess the mathematical validity of criteria measurements to ensure that it works correctly and consistently measures what one expects it to measure.

5.3.2 Ensuring validity in the MCDA problem context

The requirements specific to MCDA modelling listed below are based on the discussion on pp. 55–59 of Belton and Stewart ([3]).

- *Value relevance:* Is it possible to link all criteria measurements directly to the objectives? For instance, in assessing which car to buy, "size" may be listed as a criterion. However, it is not clear how size should be interpreted in terms of the problem: does size refer to the amount of luggage space, or the number or passengers that can be carried, or the perceived status of the car?

- *Understandability:* The meaning of the measurement attached to a criterion should be clear, and all decision-makers involved must share the same understanding of each criterion measurement. It is important to ensure that the criteria themselves do not lead to confusion and conflict. However, it should be noted that ensuring correct understanding and agreement will not always be possible during the establishment of criteria, since it may be possible that differences in understanding only emerge once criteria are assessed. If necessary, criteria must be revised at any time when such differences in understanding becomes apparent.

- *Measurability:* It must be possible to assess and compare the performance of alternatives against the various criteria. This does not necessarily imply quantitative measurement, but it does mean that such measures should be carried out consistently for all alternatives. This may have an impact on the level of decomposition required from the criteria. As for the previous point, it may only be once one starts using the criteria measurements that such deficiencies in the criteria become clear.

- *Non-redundancy:* Care should be taken to avoid redundancy, namely the fact that more than one criterion measures the same factor. During the process of generating criteria a certain amount of redundancy may be necessary (and is in fact encouraged by Keeney [30, chapter 3]), but once a set of criteria is being reduced to the final set, such redundancy must be removed since it could lead to the importance of the factor in the assessment being artificially increased. It may be difficult to recognise redundancy in the criterion, since two criteria that seem the same could in fact be measuring different aspects. An example of this is given in [11] where a bank's decision to relocate one of its head office functions is described. The top level of the criteria hierarchy was split into staff acceptability and bank acceptability and both of these had unemployment as a sub-criterion. However, in this example, for staff acceptability the unemployment value had to be minimised, while for bank acceptability unemployment would mean the ability to recruit new staff easily and should therefore be at high levels. The 'redundancy' therefore was required in order to represent the actual decision problem correctly. Belton and Stewart recommend using correlation coefficients or repertory grids, as explained in [20], to help locate such redundant criteria.

- *Judgemental independence / preferential independence:* It is usually advisable to avoid the type of criteria preferences where the preference information of certain criteria are dependent on the stated preference for another criterion. For instance, suppose that, for the problem of choosing a car, one has to consider the criteria of price, fuel economy and four-wheel drive ability. If one considers price to be more important than fuel economy for a vehicle that has four-wheel drive ability, but considers fuel economy more important than price in judging vehicles without this ability, these criteria are not judgementally independent. It may be necessary to revise or combine criteria to improve independence. (Once again, such independence may only be discovered when one starts using the criteria measurements, and may not always be anticipated at the time of constructing the criteria.)

- *Balancing completeness and conciseness:* It is of course ideal not to leave out any criteria that could impact on the decision problem, but it is also recommended to keep the criteria as few and as simple in structure as possible. Many literature references discuss the issue of how difficult it is to determine when the modelling process is complete (refer, for instance, to Belton and Stewart in [3], Buede in [10], Brownlow and Watson in [9] and the work of Phillips in [44] on 'requisite' models), but conclude that the modelling usually stops when the decision-makers feel that they have confidence in the model and that most of the important factors have been covered.

- *Operationality:* It is important that the model must be usable and not place excessive demands on the decision-makers in terms of information requirements. The context in which the model will be used must give guidance on this aspect. For instance, if an important decision has to be taken in terms of capital expenditure that will cost millions of rands, then a model requiring a great deal of time, effort and information in its assessment may be justified.

- *Simplicity versus complexity:* As with any other type of model, one prefers an MCDA model to be as simple as possible. The aim is to capture the essence of the problem without including too much unnecessary details. However, it is difficult to judge whether the final model has, in fact, achieved this.

As mentioned in the last three guidelines, above, it can be difficult to decide when a set of criteria is sufficient both to describe the decision problem as well as to allow the correct assessment of alternatives based on these criteria. General modelling principles, "common sense", principles of logic, and the co-operation of decision-makers, both in developing and in using the criteria, would be required.

5.4 Possible pitfalls in extracting preferences

While there are mathematical and contextual issues that complicate MCDA modelling, there are also many pitfalls associated with behavioural or psychological aspects affecting decision-making which could affect the successful modelling of MCDA problems. Although it is not within the scope of this document to cover **all** psychological and behavioural issues that could affect the decision-making process or the process of extracting preference information, some known pitfalls could be mentioned.

The first aspect deals directly with pitfalls occurring when decision-makers have to provide judgemental weights. One should keep in mind that methods which may feel *comfortable* to decision-makers may in fact *not lead to correct* modelling information. In the article by Von Nitzsch and Weber ([55]), the authors point out findings by themselves and others that decision-makers may feel comfortable with simple and direct ways of indicating preference weightings on criteria and alternatives, but that these more simple measures may not produce correct weights according to the mathematical principles underlying the various methods. Specifically, they mention the requirement of the MAUT theory that the weights of criteria have to be linked to the value range of the criteria. However, the simpler methods of assigning weights (direct or ratio assessments) do not take such ranges into account correctly, since the weights often stay the same even if the range varies. This means that, for instance, a person wanting to decide whether to take on a specific job offer will attach the same importance to salary, irrespective of whether the different salary alternatives range from R40 000 to R60 000, or from R45 000 to R55 000 or from R30 000 to R70 000. Not only does this seem illogical, it violates the mathematical properties required by MAUT. Belton and Stewart explain this phenomenon [3, p. 115] by indicating their belief that the weights decision-makers assign may be based on previous experience of similar problem situations rather than on the information offered in terms of the specific problem area under consideration. This may be less of an issue if the currently considered problem is very similar to problems experienced by the decision-maker in the past, but obviously will create problems if the new situation is very different from those experienced previously.

It may therefore be worthwhile to consider whether the preferences supplied by decision-makers are in fact linked to the current problem context or not. This can be done by carefully facilitating the process of obtaining the preference weights or asking for the context of past problems and pointing out the differences between that and the current context. However, Von Nitzsch and Weber also point out that one may want to look at the method used for extracting weights. Specifically, they say the following, **"Methods that do not incorporate ranges when weight judgements are derived might lead to biased weights."** Either one should use simple methods in such a way that ranges are emphasised, or one should consider the more sophisticated methods such as swing weights, conjoint analysis or tradeoff methods that specifically emphasise the range of values which are applicable to the present context.

There are also certain aspects which could affect interacting weighting structures, specifically hierarchical structures. Belton and Stewart [3, p. 117] refer to various studies that have found that decision-makers tend to give larger weights to a criterion if the weight of the criterion is calculated from the aggregation of the weights of its subcriteria than if the weight of the criterion is assessed directly without taking into account any disaggregation. Also, they relate studies that have shown that criteria at 'higher' levels in an hierarchy tend to be given higher weights than those perceived to be on a 'lower' level. Both of these aspects indicate that one should realise that methods which prescribe where certain criteria should be situated on the hierarchy may have effects in terms of the possible psychological weighting it may enforce at the same time. Also, one should perform checks to ensure that the disaggregation or break-down of a criterion into sub-criteria did not have an implicit effect on the weight provided. Such a check could for instance be done by confirming the weights when moving both up and down through the preference structure.

It is especially important to consider possible impacts on the weights whenever a structure is enhanced with more details, either through new information coming to light, or during the continual refinement of the structure during the modelling process. After any changes are made to the structure, it may be worthwhile to confirm changes (or not) made to the concurrent weights.

A third pitfall that has received attention in the literature is the possibility of manipulation of weights due to the way a problem is described or presented. Two examples of these are:

- Prospect theory, commonly referred to as **problem framing** and originating from the works of Tversky and Kahneman, as cited and discussed by Van Schie and Van der Pligt [54]. It has been found that when decision-makers are presented with a risky situation that is 'framed' in a positive way, they tend to view it more positively than a situation that is 'framed' more negatively. Note that the risk parameters (probabilities) are not distorted during framing. It is simply that the so-called 'reference point' from which the problem is described allows some issues to take on a more positive than negative slant. The following example shows how such a framing can be done. It states the problem in the positive framing with the negative framing added in brackets to show the difference.

  > *Imagine an outbreak of a disease which is expected to kill 600 people. Traditionally this disease has been combatted by the use of vaccine A. Results of this treatment can be estimated with great certainty. Another option is to use a newly developed vaccine B. One has to choose between the two vaccines. If vaccine A is adopted, 300 people will be saved (will die). If vaccine B is adopted, there is a 0.5 probability that 600 people will be saved (nobody will die) and a 0.5 probability that no people will be saved (600 people will die). Which vaccine would you opt for?*

  Belton and Stewart ([3], p. 117) point out that framing may affect weight assessments. A criterion may be given a higher weight if it is framed in terms of losses compared to another criterion (i.e. risk aversion or avoidance encouraged), while it may be assigned a lower weight if it is framed in terms of gains (encouraging risk seeking behaviour).

- In **outcome salience** the positive or negative outcomes of a risky option is given prominence ('salience') so that decision-makers are tempted to view it more positively and therefore give it more weight or more acceptance. Although outcome salience is related to problem framing in terms of manipulating the positive or negative in order to change the psychological perceptions of a decision-maker, there is a subtle difference. With problem framing, all aspects are provided, but the positive or negative may be stressed with words. In outcome salience, on the other hand, only the negative or positive side is provided, and the other part of the information is omitted, even though one may infer that information. The above example would be stated as follows with outcome salience, again giving the positive slant in the discussion with the negative slant in brackets.

  > *Imagine an outbreak of an unusual disease; 600 people are being infected. Traditionally this disease has been combatted by the use of vaccine A. Using this vaccine it is certain that half of the people will survive and half of the people will die. Recently*

> *a new vaccine has been developed. The results of this vaccine are still uncertain. This vaccine will either be effective for all infected people or not at all. The predictions are that there is a 50% chance that all 600 people will be saved (will die). Would you opt for this new vaccine?*

It should be pointed out that it is not always possible to predict which way, if at all, decision-maker preferences will be influenced through problem framing. An example is provided by Li and Adams ([33]) of how one decision is affected, but a follow-up decision is not as strongly affected.

While it is of course possible that either problem framing or outcome salience can be used purposefully to manipulate the preference weights provided by decision-makers, and such intentional practice must be discouraged on ethical grounds, a bigger danger is perhaps that it may unintentionally be used and its impact not realised. Whenever weights are determined on very risky or emotive issues, it may be a good idea to keep in mind the possible impacts of the way the issue is described. When obtaining weights on such impacts, one could first state the problem in a positive way and then check if there are doubts among the decision-makers about those weights when the same problem is stated with a negative slant. If it looks as though weights could be affected by the framing or emphasis on the positive, attention must be paid to investigating this aspect in more depth, and exploring the perceptions of decision-makers about it, before weights are fixed.

Although it is important to consider psychological/behavioural issues that could affect individual decision-makers, as in the issues mentioned up to now in this section, one should also consider the effect of issues that arise due to differences between the analyst and the decision-maker(s). The article by Wenstøp and Carlsen ([56]) provide a rather negative picture in which the analysts criticise the government MCDA process, which they consider too motivated by political aims, and contrast it with their own process which they describe as more scientific and rational. The article is useful in showing how models and results can be dramatically different from two different perspectives. In my personal opinion, however, I do not agree with this article's diametrically opposed 'rational' *versus* 'irrational' approach — playing off the consultants' views against that of the client seems dangerous and not very productive.

In my opinion, a much better approach is suggested in the article by Brown [8] in which he points out the dangers that could arise when the priorities of the analyst/consultant (who he calls the 'decision aider') differs from the priorities of the decision-maker (who he calls the 'decider'). The sketch in figure 5.1 shows his view of a causal scheme that illustrates the effects of various role players in producing a useful decision aid solution. Although there are many aspects that impact on whether the decision aid produces useful results that are eventually adopted by the decision-maker, he points out certain problems that can be directly related to priorities of the analyst (those listed in block 2, labeled 'aider priorities') which differed from that of the decision-maker.

a. Intellectual comfort: If the analyst puts too much emphasis on his/her own intellectual comfort, it could result in the analyst preferring to study

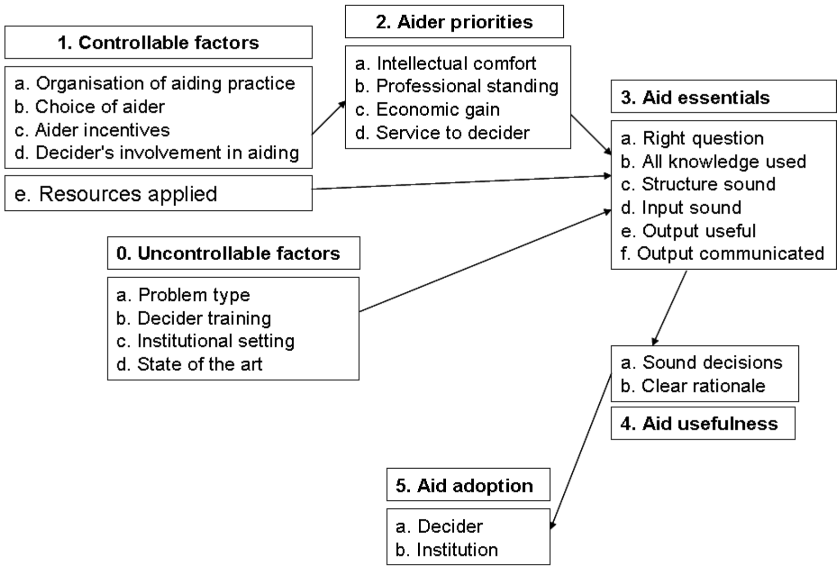

Figure 5.1: Producing a useful decision aid solution

a (sub)problem that is more interesting or complex rather than to study the actual problem that is central to the needs of the decision-maker. It can also happen that the analyst has some preference for a specific method or approach due to his/her intellectual orientation, and may choose to apply this preferred method whether directly applicable or not. Another potential problem is that the analyst may be too preoccupied with the model and analysis procedure so that he/she does not pay enough attention to important practical aspects, such as the suitability and/or quality of the input data used in the analysis. Furthermore, it may happen that the analyst regards it as a higher priority to obtain a solution that is publishable or acceptable in the scientific community than to provide the client with an acceptable solution. This could motivate the analyst to delay providing the results of the analysis to the client until the analyst is satisfied that it would stand up to the most rigourous scrutiny by the scientific community — while this sounds commendable, it could mean that the results are out of date, in an inappropriate format, or virtually meaningless by the time the client receives them.

b. Professional standing: Although the fact that an analyst considers his/her own professional standing important could ensure that they provide a good and reliable service to the client, too much emphasis on the professional standing of the analyst within a specific professional community could impact negatively on an MCDA problem. Brown mentions the example of engineers called in as analysts to assess the risks of a nuclear plant. The engineers did not want to question the method prescribed by the engineering industry for conducting such an analysis for fear of losing their standing in the engineering profession. Unfortunately, however, this generally accepted method was **not** the most appropriate method to be used in the specific problem context. Other potential problems could be that the analyst will shy away from assessing the problem with different methods, or from questioning and checking inputs and outputs, or from acknowledging areas of uncertainty or conflicting results because of fears that this will make them appear as if they do not know their job.

c. Economic gain: It should be obvious that the analyst who pays too much attention to payment for his/her services may provide the answer that the person responsible for the payment wants to obtain. However, Brown also mentions the more hidden problem of the analyst receiving instructions from a middle-man who pays for the service and then provides the results to the decision-maker, instead of the analyst receiving instructions directly from the final decision-maker. If there is a conflict between the aims of the decision-maker and the aims of the middle-man, the analyst may end up providing more support to the middle-man who pays than the decision-maker who is not directly involved in payment. Also, the analyst may gear outputs in terms of the needs of the middle-man rather than the needs of the decision-maker if it is the middle-man that has to approve the outputs before payment. An obvious solution to this problem is to make sure the middle-man does not have full control over payment and sign-off, but that the decision-maker is involved in these aspects on a continual basis.

d. Service to decision-maker: Often, the real problem is that this aspect is not the analyst's main priority, but that it is outweighed by one of the other aspects (a — c) on the above list. This could result in various problems: not asking the right questions, not providing the correct outputs or not communicating the outputs in an appropriate way, not providing outputs in time, not willing to rework results if errors in communication or practical usage of outputs come to light, and so on. As Brown states: "Nothing less that total aider obsession with usefulness may be called for."

It is important that the analyst should always query their own priorities and take steps to ensure that it does not negatively impact on the way an MCDA problem is addressed.

Finally, there are some practical issues to remember in terms of the ethics, tasks and responsibilities of facilitating an MCDA process that are discussed in some detail by Belton and Stewart in chapter 9 (pp. 261-292), but that are not repeated here.

Chapter 6

Application: Two case studies

The initial interest in MCDA classification and sorting methods originated from two practical problem situations, described below. These problem situations arose within the course of projects relating to measurement of performance and priorities which were carried out for certain departments within the South African national government by a research team from the Council for Scientific and Industrial Research (CSIR).

The projects presented some problem aspects that could not be addressed with general MCDA approaches. In order to complete the projects successfully, the CSIR team used some of the principles of MCDA, but had to adapt them to fit the needs of the projects.

Although the project team and the government clients were satisfied with the solutions which were developed, the team members decided to carry out post-assessments of the projects in order to confirm that the logical and workable solutions were indeed supported by all published scientific studies relevant to the field. It was therefore decided to study applicable methodology and prescriptions from MCDA literature in order to see whether additional insights or potential areas of improvement could be identified.

6.1 Description of practical case studies

The CSIR team was contracted by the national Department of Justice and Constitutional Development (DOJCD) to assist with the creation of a data analysis capacity within the department, initially called the Operations Centre ("Ops Room"). One aspect of the analysis work required the monitoring of court performance. South Africa has a high crime rate, and the ability of courts to deal with crime is continually under the spotlight of the public and the media. The department was therefore interested in finding a way to pick out "best" and "worst" performers amongst the criminal courts. The "best" performers would be used as role models while the "worst" performers would be assisted to improve performance. Many agencies working within the court system were using different indicators as criteria of performance, but which of these criteria should be used by DOJCD to classify court performance? And how should the indicators be combined to reflect the complexity of court performance? Such performance criteria would have to be applied to a large number of individual courts: depending on what is counted as a court (in terms of temporary or permanent sites) there were at the time more than 600 district/magistrates/lower courts in the country, roughly 200 regional courts, and 11 high courts.

In a second project, the CSIR team worked for the national Department of Minerals and Energy (DME). This project required the study of various aspects regarding the new Mining Act which places strict requirements for rehabilitation of mining sites on companies, from the time when such a company applies for mining rights on a certain site. The specific problem issue that arose was in terms of unrehabilitated, derelict mines that currently exist, since the rehabilitation of these mines became the responsibility of DME under the new Mining Act. There are many of these derelict mines: although final numbers had not been compiled at the start of the project, it was estimated that there could be a few of thousand of them. Therefore, DME's budget constraints would not allow rehabilitation of all of them at once, and it was considered important to be able to rank the mines in some way to ensure that those derelict mines posing a big danger to the surrounding environment or communities are prioritised for rehabilitation.

Although these two particular problems are quite diverse on first sight, they share some common traits, as follows:

- The first task was to devise a satisfactory group of categories representing the full scope of, respectively, performance levels or rehabilitation priorities.
- The second, and main, task was to find the best method by which to classify all courts or derelict mines into the different categories.
- The classification needed to take into account multiple, conflicting objectives. For example, in the court environment good performance would mean resolving cases speedily (the so-called principle of "justice delayed is justice denied" from the point of view of the victims of crime), but not so fast that human rights or judicial principles are violated. For the derelict mines, there were the conflicting objectives of protecting people, animals and the natural environment from harm.
- There were many "alternatives" that had to be classified, namely the various courts and the large number of derelict mines.
- Each individual classification did not carry with it a high cost or benefit. It was only the total classification that would be beneficial, not the individual placements.
- The group of "decision-makers" were within a national government department. Therefore, they had to take decisions on behalf of the general public, and not only had to consider their own personal views and preferences.

- If possible, a final solution had to take into account expert opinions or scientific knowledge in the various fields. How should expert opinions and knowledge be incorporated into the classification or measurement?

- The final solution had to be defendable and transparent to the public. Although some members of the public may have technical experience in the specific fields of justice or mining, it had to be assumed that the final solution will have to be easily understood by a general public who will not want a very mathematically complex answer. Specifically, if reports are given in the media about the problems at a specific derelict mine, it should be possible to provide a simple and acceptable response to defend the classification of that mine.

6.2 Chosen solution method

The practical solutions that were developed on these two projects, although differing very much in the exact details required by the application area, shared the following characteristics:

- The first step was to define what has to be measured and what the aim would be of the measurement. For instance, in the DOJCD example, it was decided not to focus only on identifying "best" and "worst" courts, but in giving a general performance measure that would help *all* courts enhance their operations in a meaningful way. Initially, the DME requested a score for each mine that would indicate a risk priority percentage, but the recommendation by the CSIR team was made to rather place mines in groups describing a more qualitative risk status. Such a status would be more meaningful than a score which would still need further interpretation (for example, does a score of 80% represent a high risk or not?).

- Then the process required exact definition of the categories in which the alternatives had to be grouped. In both these projects a very vague statement about the eventual grouping that was required had to be translated into a more specific, usable group of categories. It should be noted that, in both of these projects, the eventual categories did not turn out to be ordinal. While there may have been some order on the extreme ends of the categories (i.e. there was a category that came out as either relatively "best" or "worst" in terms of performance or priority), but the "middle" groups were not strictly ordered relative to each other.

- The next step was to formulate reliable indicators that would measure what was required, yet could be obtained from data readily available or practical to collect. In many cases, a composite indicator had to be derived to represent a specific logical criterion.

- The final step was to construct a hierarchical pattern of indicators that would prescribe the process to follow in order to classify each "alternative" into a specific "category". Note that the hierarchy did not always prescribe a strict analysis sequence, but also allowed for the ability to "fast track" some classifications. For instance, a first question was asked that would place the alternative into a specific category immediately if the answer was "yes", but would require more questions if the answer was "no" in order to place the alternative.

6.3 Contrasting case studies with MCDA methodology

When the problem situations and proposed solutions for these two practical projects were considered in the light of the information provided in the previous chapters of this document, a number of insights were gained. These insights are discussed in more detail below, distinguishing between insights in terms of the *measurements used* (contrasting the proposed solutions to the guidelines in chapter 5), and in terms of the *methodologies employed* (as compared to the discussion in chapter 4).

6.3.1 Reviewing measurements used in case studies

In both of the practical projects, a great deal of attention was paid to the development of indicators that would eventually form the basis of the classification — what in MCDA classification methodology be seen as the criteria measurements.

Court performance

In the example of the criminal court classification, the project team initially reviewed the indicators used by different stakeholders.

Although initially it seemed as if performance is not really measured, or is simply measured on the basis of anecdotal information or media reports, after further investigation a few quantitative performance measures that were used by different stakeholders were found. The two main indicators were that of monthly average court hours (calculated as the average number of hours per day that the court managed to be in session in a month, i.e. total hours per month divided by total number of days in session) and the conviction rate. Closer inspection of both of these measures showed that they could not provide a complete picture of performance.

The measure of average court hours has merit in the sense that no progress can be made on court cases if the court is not in session, so that courts recording low average court hours compared to targets could well be assumed to be performing badly. However, the measure had two main shortcomings. The first is that the average court hours of all courts were compared to one national target, without recognition that some of the smaller courts have fewer and less complicated cases, and therefore may in fact get through their case load in less court time. Secondly, while it may be true that courts that do not spend long enough time in session cannot perform well, the converse is not necessarily true. It cannot be assumed that a court that spends 'enough' time in session is in fact performing adequately, since they may be spending their time inefficiently.

The second indicator, that of the conviction rate, presented a different concern. The conviction rate is calculated as the percentage of convictions (guilty verdicts) out of the number of court cases that are finalised, and every court's conviction rate is again compared to a national target. The problem is that the conviction rate focuses on one specific aspect of court performance only, and may miss other important aspects. There are many ways to finalise a court case, namely by withdrawing a case, by referring the case to another court (typically higher or more specialised) or by reaching a guilty or not guilty verdict. One way to improve the conviction rate without doing much to improve performance is to increase the number of withdrawals, i.e. to pursue only 'easy' cases into verdict and finding some reason to withdraw more 'difficult' cases. (In fact, in practice this would not only **not** improve performance, it would almost definitely be seen as declining court performance.)

It was therefore clear that the current indicators were not sufficient, and that attention would have to be paid to developing more appropriate indicators to be used as performance criteria to measure criminal court performance. In order to develop such criteria, the team had to consider four important aspects:

1. An understanding of the system would be required in order to measure the correct aspects. It was necessary to describe and analyse the underlying processes at work so that one could derive measurements at the appropriate position(s) in the process. For instance, it was clear that one had to monitor both withdrawals and convictions, and not only convictions, in order to be sure that improvements in the indicators in fact signify improvements in performance.

2. Consideration had to be given to what exactly a measurement records and how this links to performance. Although this sounds like a point similar to the one above, there is a substantial difference. For example, one would think that a good performance indicator would be the average time it takes to finalise a case. As indicated above, there is a saying of "justice delayed is justice denied". It is in fact true that if a case drags on for too long, witnesses may start forgetting details, and victims of crime can feel that they are not receiving closure. However, some cases are complicated and require detailed investigation and evidence presentation, and there may be very good reasons for delays in certain cases. So while a case that proceeds speedily can be an indication of good performance by the court, a case that proceeds too fast can be an indication of sloppy work or violation of the rights of the accused to a fair trial. It was therefore decided that while the time required for a case may present potential for use as a performance indicator, it could not be used as an unambiguous measure of performance.

3. It was preferable to base performance measures on data that can readily be collected. If such a performance indicator was to be used to track performance over time, then it should be easy to collect on a regular and sustainable basis. This basically translated into using data that was already collected as the basis for performance indicator measures. An example of this is that it would be nice to track performance according to the type of court case, for example, murder cases, or robbery cases, or shoplifting cases. However, not only was such data not recorded at the time, it would also have been fairly difficult to recommend that such

data had to be collected. The reason for this difficulty is that court cases typically consist of more than one accused, more than one charge or more than one charge per accused. A person charged with murder may also (as part of the same case) be charged with possession of an illegal firearm, assault, and so on. It is much easier to collect data about numbers of cases than about the content of cases, and it is also much easier to interpret numbers of cases than content of cases in terms of performance.

4. Since courts differ in size and function, it was important to find a way to standardise data for performance measures to be comparable for all courts. To use averages for such standardisation would not be acceptable, since case numbers per court showed a distribution that was very skewed in the statistical sense, making an average per court an inappropriate measure. It was therefore decided to try to use performance measures that consisted of ratios or percentages as a way to incorporate standardisation. For instance, it was decided to develop the so-called "clearance rate" to measure whether courts are managing to deal with their case loads. The clearance rate was calculated as the number of cases finalised in any month compared to the number of new cases entering the court roll in the same month. This measure was considered to be quite a powerful indicator of performance, since a court that does not finalise a high proportion of incoming cases, such as when inputs into the court by far exceeds outputs, cannot be performing well. Also, this measure is standardised automatically, since it does not just compare number of finalised cases per court, but balances finalised cases against new cases and therefore corrects for the size of the court.

The performance indicators that were recommended for use in monitoring the performance of criminal courts was therefore decided to be the clearance rate (number of finalised cases divided by number of incoming cases), the conviction rate, the withdrawal rate (number of withdrawn cases divided by number of finalised cases) and average court hours per month. It is interesting that the process used to decide on which indicators to use, as described in some detail above, seemed to correspond into the guidelines provided in chapter 5. The team started by trying to find measures that Raisbeck ([46]) calls *from statute* or *by consensus* and scrutinising the measures to ensure that they measure the correct aspects. Then, when such measures were found to be inadequate, additional measures were developed *from analysis*. Attention was also paid to using appropriate ways of *normalisation*, in line with the issues that was pointed out as important by Bouyssou *et al* (([4]). Finally, the set of recommended indicators also met both the requirements of *operationality* and *balancing completeness and conciseness* aspects mentioned by Belton and Stewart ([3]). It can be concluded, therefore, that the measurements developed for use in the court performance case study held up to what is generally prescribed in the referenced literature as sound principles for the development of indicators to be used as MCDA criteria measurements.

Derelict mines

In the derelict mines case study, the main problem was how to integrate quantitative scientific measures, scientific expert views and community values into a measurement system that would provide a single answer in terms of the risk status of a mine. There were some risks that had been studied well by the scientific community and were well understood, such as the impact of mined interval thickness and overburden thickness in underground cavities on the possibility of subsidence. In other aspects, scientific expertise could provide opinions but not exact measurements. Such aspects included the impact of different types of mining activities on the environment, for instance experts knew that gold or coal mining presented higher risks of pollution than diamond mining, but it was difficult to capture this impression as a quantitative measure. Then there were other aspects that related more to community values and behaviour than scientific study, such as the dangers presented to the community from open pits (children or animals could fall in and be injured). For most of these aspects, indicators could be constructed either as scientific measures or as simple qualitative measures.

However, constructing an indicator to measure risk was more problematic. The reason why this was problematic is because risk is related to the combination of a few factors. For instance, a certain mine may be polluting a groundwater source quite severely, but there is no community using the groundwater source for either household or farming. Another mine does not pollute water sources, but causes air pollution, and this air pollution affects a nearby community quite severely because it is close to the community and on most days the wind blows in the direction from the mine to the community. Surely this second mine must have a higher risk factor?

The indicator suggested by the team to calculate risk factors was a composite indicator in which a measure of **hazard** (how much pollution or threat) is multiplied with a measure of **'receptor' effect / usage** (human or environmental) as well as a measure of the **strength of a pathway** between the hazard and the receptor. A high risk factor would be obtained if there was a big hazard, linked via a strong pathway, to high usage by a receptor. If the hazard itself was less dangerous, or there was a smaller community or less usage, or if there was no real pathway between the hazard and the receptor, then the risk factor would be much lower. The team determined that a multiplicative indicator worked much better for a meaningful aggregation into a composite indicator than, for instance, an additive one. After consideration of the literature reference on composite indicators by Bouyssou *et al*, it seems that this suggested indicator was also acceptable in terms of the *scale construction* and *aggregation issues* mentioned.

The second difficulty in the development of measures for this case study, was in the development of the categories to be used for the classification of the need for rehabilitation of a specific mine. Eventually, it was proposed that the categories were designed in the form of a matrix rather than a single measure, as indicated in table 6.1. The table defines the category in terms of a combination of "impact" and "risk", where the risk is determined as the potential harm (as described by the composite / multiplicative risk factor indicator discussed in the previous paragraph) and the impact is determined from how extensive the harm could be (for example, how badly it impacts on health, water quality levels, air quality levels, etc.). In order to attach some indication of *the order in which* *rehabilitation has to take place*, as related to the category into which the mine was classified, a number was attached to the category names. Note that this number was suggested by the project team in order to assist in the interpretation of the categories, and was added in a purely judgmental way. This number is also indicated on the matrix: category 1 refers to mines which must be urgently rehabilitated, while category 9 would include mines which can either wait until last or perhaps not even need any rehabilitation at all. As it can be seen from table 6.1, the category number does not necessarily follow a regular pattern across the matrix. Category 1 and 2 follow each other horizontally, but is then followed by category 4 and not by category 3 as perhaps would be expected. If one takes the view that problems that already exist take priority over problems that could occur in the near future, a category 3 mine that is categorised as a current problem (even though it is exacerbated rather than caused by the mine) should get priority over a category 4 mine that is categorised as a "problem waiting to happen".

This matrix could be seen as an example of a constructed measure according to Keeney's definition ([30]), since it has specific relevance to this problem context and has been developed because there were no readily available natural or proxy measures to use to describe the rehabilitation category.

Table 6.1: Rehabilitation classification for derelict, ownerless mines

		Impact		
		High	**Medium**	**Low**
	High	(1) Should be rehabilitated immediately	(2) Problem mine to be rehabilitated soon	(4) Problem waiting to happen
Risk	**Medium**	(3) Problem, but not only caused by mine	(5) Moderate problem	(6) Moderate future problem
	Low	(7) Mine not the problem	(8) Moderate, stable problem	(9) Stable, no problem

As an aside it should be mentioned that the intention was to categorise the mines only and not to provide a complete input into the actual process of rehabilitation. Also, potential cost or difficulty of rehabilitation was not brought into the classification at all. Planning the implementation of the rehabilitation may therefore require information in addition to the rehabilitation category, but the scope of the modelling could not allow the inclusion of such aspects.

Once clarity was achieved in terms of the matrix used to describe the categories into which the classification had to be done, as well as on the measurements to use for the classification, the next challenge was to classify the mines into these categories. Therefore, the next step related to the classification methodology.

The classification methodologies used in the two case studies, as compared to methods suggested by MCDA classification literature, are discussed in the next subsection.

6.3.2 Use of classification methods in the case studies

It was interesting to compare the methods used in the case studies to the more formal MCDA literature.

Court performance

In the final recommendation for the court performance classification, the four performance indicators were arranged into a hierarchical representation. However, this hierarchy did not so much represent a value tree as a lexicographic ordering of the importance of the criteria in terms of determining a performance classification. In fact, the method used for classification seems quite closely aligned to a lexicographic approach, even though no literature references were found of anyone using a lexicographic method for MCDA classification problems.

The classification method used first looked at classifying courts according to their clearance rate. It was recommended that the most important factor should be whether a court manages to keep up with its workload, i.e. whether the court is able to match outputs to inputs. This led to a first classification into preliminary categories. Then, in a second classification, each of these categories were further subdivided according to a combination of conviction rate and withdrawal rate. Finally, average court hours were used to provide a ranking of courts within each of the categories.

The method classified courts into nine performance categories. The 'best' courts had a high clearance rate coupled with a high conviction rate and a low withdrawal rate, whereas the 'worst' courts had a low clearance rate and a low conviction rate. Note that if the clearance rate was low and the conviction rate was low, the withdrawal rate could be disregarded in the classification, since the information on the clearance rate and the conviction rate was sufficient to provide the classification. In general a low conviction rate was seen as an indication of bad performance, but if the conviction rate was high, performance was considered better if the high conviction rate could be obtained in combination with a low withdrawal rate. The lexicographic nature of the classification method can therefore be seen both in the fact that the most important criterion was considered first as well as due to the fact that it was often possible to classify a court based on two of the criteria only. The classification method is illustrated in figure 6.1, with the 'BEST' and 'WORST' categories marked with a label and a dotted line.

In terms of a formal distinction between classification and sorting methods, as is often found in the MCDA literature, this method provided a classification rather than a sorting solution, since the categories were not completely ordered from 'best' to 'worst' performance. The overall best and worst categories were of course ordered, but there was no clear ordering between, for instance, a court obtaining a high clearance rate and a low conviction rate as opposed to a court having a medium clearance rate coupled with a high conviction rate and a high withdrawal rate. However, the advantage of using classification was that it could not only provide an indication of how courts are performing, but also of what problems they need to address to improve performance. If, for instance, the court had no problems with regard to its clearance rate (i.e. the clearance

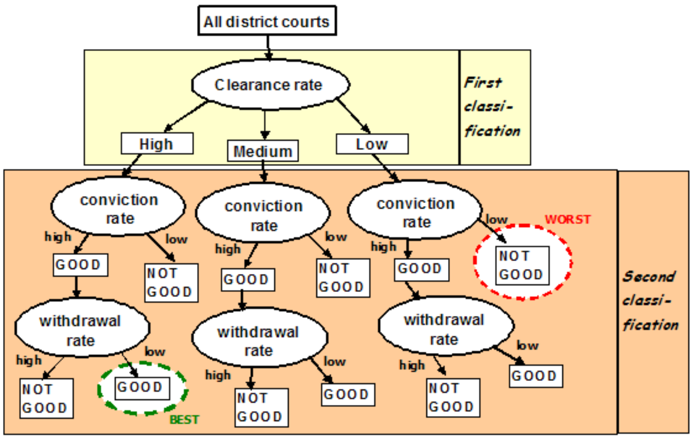

Figure 6.1: Criminal court performance model

rate was high or medium), but experienced problems in terms of withdrawal rates that are too high, that would require a different intervention to improve performance than a court that is struggling with a low clearance rate.

It may, as a final comment, be noted that the problem context was specifically limited in terms of court performance to look at general process flows and to disregard the actual quality of verdicts. This was because there were acknowledged processes for peer review and appeals in place to monitor the quality of verdicts, whereas there were no proper measures of performance that could be used to guide resource allocation and other types of process planning within the court environment.

Derelict mines

For the mine rehabilitation classification, a very rough hierarchical structure of indicators was also used, but this hierarchy was closer to that of a top-down attribute hierarchy described by Buede than that used for the court performance case study.

Every mine was classified into an overall rehabilitation category. This could be done in one of two ways:

1. Firstly, the overall category was determined in terms of the matrix classification explained above. This overall measure was not measured directly, but was in fact aggregated from similar classifications on each of four different ‘exposure pathways’. These four exposure pathways were surface water, groundwater, air and direct access (the physical holes or structures on land). For each of the four pathways, a classification was obtained according to the matrix measure. The classifications on the four exposure pathways were then aggregated into the overall index by taking the worst case accross all exposure pathways: if a mine was placed as category 1 (should be rehabilitated immediately) on any exposure pathway, it would

be ranked overall into category 1. Note that the classification within each exposure pathway was based on the risks and impacts relevant to that pathway, with the risks and impacts in turn calculated from scientific models and expert opinions (differing for the different pathways).

2. Secondly, the method allowed for overriding concerns to place a mine immediately into category 1. An example of an overriding concern would be if there had been media reports of persons or animals being hurt on or by something originating from the mining site. It would be prudent to address a known danger situation as a category 1 mine to be rehabilitated immediately.

The classification method was therefore a hybrid model combining lexicographic, disjunctive and attribute integration rules. A lexicographic-type rule was used to incorporate the use of overriding factors, while the aggregation of the classifications on each exposure pathway could be seen as a type of disjunctive rule and the calculation of the classification within each exposure pathway as a more traditional hierarchical attribute aggregation procedure.

It should be noted, however, that there was an explicit decision not to attach preference weights to any individual measure (which in essence gave equal weights to all measures aggregated into a combined measure), since no stakeholder was willing to compare, for instance, loss of human life to loss of environmental diversity, or seriousness of water pollution compared to seriousness of air pollution. It may be possible to provide such weights at a specific local or regional level, but since this model had to be acceptable at a national level, it was considered too difficult to obtain consistent national weight values. The methodology therefore had to be adapted not only to cater for different types of criteria data, but also to ensure that the assumption of equal weights was correctly carried through the classification process.

It could further be mentioned that the methodology also prescribed default values to be used in the case of missing values so that a mine may be initially classified based on very little information, allowing the classification to be updated as more and more information became available. This was possible due to the scientific information about certain mining activities being enough for a rough first cut at classification. For example, any asbestos mine could almost automatically be classified into a high priority rehabilitation class, and the model allowed such a type of classification.

The method is illustrated in 6.2 below. The proposed method for classification of mines was found acceptable by the client, and has been used as the backbone for the development of a system to collect the required information and to carry out a classification of mines. Even though the classification model was conceptually quite complex to design, it was found quite easy to apply in software.

Critical assessment of both methods

While the criteria measure development on both projects could be confirmed as acceptable based on the referenced literature, there was a less clear-cut answer

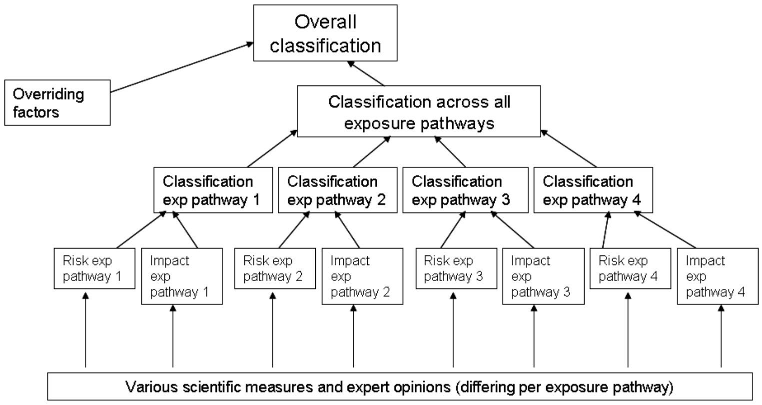

Figure 6.2: Mining classification model

in terms of the comparison of the case study methodologies compared to the literature on MCDA classification methods.

On the positive side, consideration of the methods described in the literature did not point to any substantial omissions in the methodology, and could also not point to an alternative methodology that could work better. Specifically, the fact that preference weights were not incorporated in the case study classification models was an initial concern, since it as felt that most MCDA methods seem to put a great deal of emphasis on the incorporation of weights. One of the reasons for not wanting to assign explicit preference weights to the different risk aspects in the mining classification example, was due to experts feeling uncomfortable comparing criteria they considered incomparable. In the court classification case, there was an implicit understanding that clearance rate should be the most important criterion, but a reluctance to give a quantitative weight to express such an importance. Not including explicit weights in the solution models therefore seemed to be a potential criticism of the case study models.

However, this was not found to be such a problem in terms of using MCDA classification models. It was found that for some of the recognised MCDA classification and sorting methods one does not have to specify explict weights. For example, the ORCLASS method seems to assume that the ordering of the criteria measures express an implicit preference structure and the method does not require additional specification of weights. Similarly, the rough sets method does not require any explicit preference weighting of criteria, but instead models preferences via the criteria levels leading to a specific classification. One could also argue that although explicit weights were not assigned in the model, the logical structure of the models ensured the application of a consistent implicit weighting system. It would be possible to attach explicit mathematical weight values that would be equivalent, if required, but there was no need to do so, especially since the intended users of the models preferred not to have any explicit quantification of the weights. Also, from the literature it seems that issues of non-compensatory criteria are common. So the fact that weights were not explicitly assigned was not necessarily a criticism — although it could be asked whether the solution methods dealt correctly with the non-compensatory criteria.

On the negative side, there are a few criticisms that one could bring against the case study methodologies. Many of the MCDA methods described in the literature provides some form of summary of the eventual classification, as well as some measure of the quality of the classification. Not enough attention was paid to this aspect in the case studies. Also, there may not have been enough scrutiny of the way in which criteria were aggregated: while using a matrix or a ratio to determine a criteria measurement may be logically satisfactory, neither of these are linear measures. More effort should have been spent in studying the measures to ensure that they remain valid over the entire range of data values. Also, not enough attention was paid to the independence of criteria. Since an important concept in MCDA modelling is that criteria should be preferentially independent from one another, the solution models should have been scrutinised to look for possible dependencies between the various criteria.

In summary, if the problems examined in the case studies had to be contrasted against the method proposed in chapter 4 for selecting an MCDA classification method, there does not seem to be an method that would be guaranteed to work better. For the court performance example, it could have been worthwhile to consider some of the methods that can be used for non-compensatory criteria, such as the outranking methods. However, it cannot be guaranteed that these methods would have worked better in the given problem context. People working in the Justice environment are not very comfortable with mathematical concepts, and the mathematical complexity of the outranking methods as well as the large number of quantitative parameters required for the ELECTRE group of methods may have been a drawback. Also, it seems as if outranking methods are usually applied to methods with a relatively small number of alternatives, and this could be a drawback in this situation in which there were so many alternatives. In future, however, one may want to consider these methods for a similar type of classification, especially given the fact that there were only a few criteria to be used for classification. For the derelict mine example, an appropriate alternative method is not as clear-cut. The main problem with this practical case study is that all the alternatives were not known at the time of deriving the classification model, and criteria scores were not available for all alternatives either. In fact, very little data was available per mine at the time and the model had to, amongst other things, provide a guideline for what information to collect. So, possible methods had to be limited to those that used criteria as inputs and had to exclude any methods that required information on alternatives or reference groups of alternatives. Furthermore, only methods that allow non-compensatory criteria, that do not require preference weights on criteria and are not limited to quantitative data for inputs could be considered. There are not many methods that meet all of these requirements, except the fairly simple lexicographic and conjunctive/disjunctive methods. If the situation was that a more complete list of derelict and ownerless mines were available, and that more data was available per mine, the rough sets approach may have been a useful method to apply, since it would potentially have been able to sort through the various scientific criteria for the most effective set of indicators to use. Also, if more data were available it would perhaps have been possible to use different methods and to compare them to see which would give the most appropriate results.

It seems as if a more in-depth study of MCDA classification and sorting would not have found methods that would definitely have been superior to the methods that were in fact used. The methods seemed to work well in practice, to meet the approval of the clients, and to provide easy-to-implement classification frameworks. From this comparison to literature it seems as if the methods may benefit from some refinement, but could be considered as scientifically valid when compared to the existing body of knowledge.

6.4 Issues specific to public sector problem environments

Any MCDA solution aims to be appropriate within the specific context of the problem situation, and care should be taken to reflect the complexities and subjectivity of the decision-making process. Modelling subjective preferences of decision-makers is challenging in any context, but when the context of the problem is within the public sector, there are additional issues to consider.

Belton and Stewart ([3] in pp. 59-60) indicate the importance of taking account of the views of stakeholders in terms of taking decisions, but also point out the importance of considering the power of the stakeholder to disrupt or sabotage a decision. This is specifically relevant in the public sector, where many of the decisions are motivated on the basis of political issues or have to address the needs of a wide variety of stakeholders.

An interesting article by Carlsson and Walden, [12], discusses a practical problem in which a difficult decision had to be taken about the siting of a ice hockey stadium. The Analytic Hierarchy Process (AHP) MCDA method was used to provide a rational framework for the ranking of potential sites. Although the authors pointed out some flaws in the method, the team of administrators that participated in the application of the method expressed their satisfaction that the process allowed them to think through and discuss all the issues in a rational way, and also gave them an opportunity to consider all relevant aspects to the decision-making problem. They were also satisfied, after a process of sensitivity analysis, that the ranking given to the various sites were correct, and that the site given a rank of 1 (most preferred) was indeed the best site. However, when the adminstrators put the decision and their recommendations to the political decision-makers, the politicians decided to choose the site ranked third by the AHP process rather than the site ranked first. The authors expressed their initial surprise at the decision, since the site ranked first was superior to the site ranked third in many respects. However, in investigating the reasons for choosing the site, they found that it was a purely political decision in which the one political party knowingly chose a site that was not 'best' in order to embarrass their political opponents and to put themselves in a better bargaining position in terms of another decision that they had to make subsequently.

Brownlow and Watson([9]) list two main aspects to take note of in the public policy context. Firstly, the aim of MCDA support should not be to help make quick decisions, but should rather be to find ways to present large amounts of technical information in a way that is both summarised and rational, yet easy to understand. Scientific information must be incorporated into a decision in such a way that the the political decision-makers can combine their judgement with the scientific information to support the validity of the decision-making. A second aspect is that one can expect to find many different stakeholders that have an interest or a stake in a decision problem. These stakeholders may all approach the problem from different points of view, and there may even be different points of view within the different stakeholder groups. The different points of view will be a source of conflict in the decision-making process and could influence how they perceive the relative importance of the various criteria, what they consider appropriate ways to measure criteria and the scores criteria attain on these measures.

Scott ([49]) points out the responsibility of local government to take cognisance of the needs and priorities of the communities they serve. This is of course also true of the public sector at provincial and national levels. Not only must government take into account constitutional legal requirements and official policies, they also need to represent the values and needs of the communities they represent.

6.5 Conclusion about MCDA classification and sorting for public sector problems

One of the reasons why it was felt that a post-analysis of the problem case studies would be useful, was because the main characteristics of the case studies were very similar although the details of the problems seemed to differ. It was felt that there could be similar problem situations within the public sector of developing countries in general, and South Africa in particular. It was therefore considered potentially useful, even important, to gain better insight into possible solutions for such problems.

One of the main challenges that have to be addressed by the South African government is how to determine the priorities of reversing developmental backlogs compared to that of investing in new developments to support future growth. Since addressing the historical backlogs and promoting future growth could be seen as two conflicting objectives, it could be argued that the use of MCDA methods, which were designed to cope with such decision-making situations, could be of value. It has been argued by many that MCDA methods can support the development of public policy in many ways (for instance, see [49]). Firstly, MCDA methods provide systematic and rational ways to structure problems and preferences, and could therefore support strategic and developmental planning. Furthermore, by providing better structure and more transparency about strategic and planning decisions, actions at an operational level can be better aligned with strategic plans — it seems that many of the good policies adopted at a strategic level are not properly translated into implementation plans.

However, this study not only showed that there may be specific advantages to using general MCDA methods, but also that the use of MCDA classification and sorting methods may be particularly useful in public sector problems. This could be concluded both practically, from observing the usefulness of classification methods when applied to the two case studies described above, as well as theoretically, from the extensive study of MCDA classification and sorting methods discussed in this document.

The following are seen as advantages that could stem from the use of systematic and rational ways to classify alternatives with multiple objectives, using MCDA classification and sorting methods:

- The MCDA classification and sorting methods seem to have less stringent requirements than the MCDA methods used for ranking or choice. As mentioned in chapter 3, both ranking and choice require relative comparisons between alternatives, while classification and sorting methods use absolute comparisons of alternatives to the criteria structures. Classification and sorting methods can therefore be used when, for instance, all alternatives are not know yet, or there are too many alternatives to allow direct comparisons between them, or decision-makers feel uncomfortable about pairwise comparisons between alternatives.

- One of the important aspects in terms of decision-making in the public sector, as mentioned above, is that there are typically many stakeholder groups interested in a specific decision. Such stakeholder groups may be within the decision-making team, or could be groups that would require the justification of a classification or sorting procedure once it was completed. As mentioned in terms of the derelict mine case study, if an accident happened at the site of a specific mine, the community may demand that the DME explain why the mine was not rehabilitated in order to prevent such an accident, and then DME would have to justify why that mine was placed in a specific risk category. Classification and sorting methods have the advantage over choice or ranking MCDA methods in that there does not have to be such an exact distinction between alternatives and that the requirement for distinguishing between alternatives does not have to be such an exact quantification. If two alternatives are incomparable, classification methods will put them in the same category and no further distinctions will be required between those alternatives. However, in a choice or ranking problem, there would be a need to do more analysis in terms of which of the two are preferred in some way to the other. This fact of less stringent requirements in classification and sorting methods as opposed to choice and ranking problems have the following advantages in terms of public sector problems:

  - Choice of methods: whereas the simpler methods (conjunctive / disjunctive, lexicographic, etc.) are often not sufficient in order to obtain the solution to a choice or ranking problem, they may be entirely sufficient for classification problems.
  - Mathematical complexity: the mathematical complexity of classification methods may be less than that required for other types of MCDA

problems. Not only is this an advantage when dealing with problems in the public sector that require politicians as decision-makers, but it also is helpful having an method that can more readily be explained to and accepted by the public and other stakeholders.

  - It is possible to use more implicit methods of weighting in classification methods than in other types of MCDA problems. An example would be using an ordinal measure or category to implicitly indicate preference order, even if the preference order is not quantified. In the public domain, decision-makers may be hesitant to give weights to non-compensatory criteria, which would be a problem in using most choice or ranking methods, but would not be such a serious drawback when using some of the classification or sorting methods. Also, many of the classification and sorting methods make it possible to deduce the preference ordering of a decision-maker from a few classification examples and still do a complete classification.

- The measurements used in MCDA classification methods are not so different from those required by other types of MCDA methods, but again there may be less stringent requirements on the measurements in classification as opposed to other MCDA methods. This means that it may be easier to:

  - incorporate data and scientific results.
  - include vague qualitative criteria in order to measure all aspects rather than leave out an aspect that is difficult to measure or to give a preference weighting.

One can perhaps go so far as to say that it would be a good idea to consider changing a "difficult" problem that is presented as a choice or ranking MCDA problem into a classification or sorting problem, as was done for the two case studies mentioned. Or, it may be an option to do a classification process as a first step to identify a subset of alternatives to consider in terms of choice or ranking in more detail. (Classification or sorting methods could, for instance, be used to first group alternatives into one of two categories called 'acceptable solutions' or 'unacceptable solutions'. This will then reduce the number of alternatives to consider further in terms of finding the best one, and thereby simplify the choice or ranking problem.)

Although MCDA classification and sorting methods seem to have potential for wide use within public sector problems, this study shows that there may be aspects of these methods that could be further refined to be of even more use. It has been mentioned that sensitivity testing and ways to measure the quality of the classification have not received much attention in the literature. Within other MCDA methods there are many references indicating how sensitivity analysis had been indispensable in the decision-aiding and decision-modelling processes, especially in confirming areas of which the decision-makers were uncertain, and sensitivity analysis must certainly be incorporated into classification and sorting methods to a greater extent than seems to be current practice.

Finally, a drawback of the currently published set of classification and sorting methods is that they generally assume agreement between decision-makers, especially about the criteria to use and the ordering within the criteria. It is conceivable that all public sector problems would be affected, at least to some extent, by serious conflicts of opinion between the various decision-makers and stakeholders. It would therefore be necessary, in the practical application of MCDA classification and sorting methods to public sector problems, to ensure that specific attention is paid to resolving issues of conflict, even though this is a topic on which guidelines are not readily available in the current literature.

Chapter 7

Final remarks and conclusions

The first conclusion from this study is that MCDA classification and sorting methods form a valid group of methods within the general MCDA field. Although perhaps not the most well-known subfield within the general MCDA field, many literature references have been found that describe the theory and application of classification and sorting methods. In contrasting existing MCDA classification and sorting methods to other types of MCDA methods, the classification and sorting methods are mostly simpler, more accessible in terms of mathematical complexity, and have less stringent requirements than other types of MCDA methods.

The study investigated the various ways in which the problem of selecting an appropriate MCDA method have been addressed, and put forward some recommendations for ways of choosing a method or group of potential methods. In terms of selecting an appropriate MCDA classification and sorting method, it was found to be easier than for other types of MCDA methods, since there are fewer methods, and there are fewer requirements in terms of mathematical assumptions and algorithms.

It is important to take cognisance of general literature in terms of quantitative modelling and general principles of MCDA modelling when putting together the measurements to use in MCDA classification and sorting problems. Although the use of measurements may be simpler in some of the classification and sorting methods, most of the time the same requirements will be applicable to the measurements used in MCDA classification and sorting problems than for measures used in other types of MCDA methods. One must, for instance, not forget to check important issues such as independence between criteria measures and the validity of non-linear measurements.

When studying the use of MCDA classification and sorting methods in two case studies, the classification and sorting methods were found to be extremely useful. This underlines the conclusion that these methods must not be underestimated in terms of their applicability to practical problems. It is foreseen that MCDA classification and sorting methods could be especially applicable in public service problems in which there are typically difficulties with non-comparability of alternatives and criteria, as well as a need for methods that are simpler and less complex mathematically that would be easy to justify within a political domain or to community stakeholders. These methods would provide enough rational structure and scientific validity to provide a valid basis for decision-making and therefore to provide proper decision support, but their simplicity could improve transparancy. It may also be easier to gain acceptance for the solutions offered by these methods and to extract the information required in order to apply these methods.

Bibliography

[1] T. Al-Shemmeri, B. Al-Kloub, and A. Pearman. Model choice in multicriteria decision aid. *European Journal of Operational Research*, 97:550–560, 1997.

[2] C. Araz and I. Ozkarahan. Supplier evaluation and management system for strategic sourcing based on a new multicriteria sorting procedure. *International Journal of Production Economics*, 106:585–606, 2007.

[3] V. Belton and T. J. Stewart. *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers, 2002.

[4] D. Bouyssou, T. Marchant, M. Pirlot, P. Perny, A. Tsoukias, and P. Vincke. *Evaluation and Decision Models: A Critical Perspective*. Kluwer Academic Publishers, Dordrecht, 2000.

[5] D. Bouyssou, P. Perny, M. Pirlot, A. Tsoukias, and P. Vincke. A manifesto for the new MCDA era. *Journal of Multi-Criteria Decision Analysis*, 2:125–127, 1993.

[6] D. Bouyssou and J.-C. Vansnick. Noncompensatory and generalised noncompensatory preference structures. *Theory and Decision*, 21:251–266, 1986.

[7] J. Brans and P. Vincke. A preference ranking organisation method. *Management Science*, 31(6):647–656, June 1985.

[8] R. Brown. The operation was a success but the patient died: aider priorities influence decision aid usefulness. *Interfaces*, 35(6):511–521, 2005.

[9] S. Brownlow and S. Watson. Structuring multi-attribute value hierarchies. *Journal of the Operational Research Society*, 38(4):309–317, 1987.

[10] D. M. Buede. Structuring value attributes. *Interfaces*, 16(2):52–62, 1986.

[11] N. Butterworth. Giving up "the smoke": A major institution investigates alternatives to being sited in the city. *Journal of the Operational Research Society*, 40(8):711–717, 1989.

[12] C. Carlsson and P. Walden. AHP in political group decisions: a study in the art of possibilities. *Interfaces*, 25(4):14–29, 1995.

[13] Y. Chen, D. M. Kilgour, and K. W. Hipel. MCC with applications in water resource planning. *Computers and Operations Research*, 33:3301–3323, 2006.

[14] E. U. Choo and W. C. Wedley. Optimal criterion weights in repetitive multicriteria decision-making. *The Journal of the Operational Research Society*, 36(11), 1985.

[15] R. T. Clemen. *Making Hard Decisions: An Introduction to Decision Analysis (Second Edition).* Duxbury Press, 1996.

[16] H. G. Daellenbach. *Systems and Decision-Making: A Management Science Approach.* John Wiley & Sons, 1994.

[17] Y. De Smet and L. Montano Guzmàn. Towards multicriteria clustering: An extension of the k-means algorithm. *European Journal of Operational Research*, 158:390–398, 2004.

[18] L. C. Dias and V. Mousseau. IRIS: A DSS for multiple criteria sorting problems. *Journal of Multi-Criteria Decision Analysis*, 12:285–298, 2003.

[19] D. Dubois, H. Prade, and C. Testemale. Weighted fuzzy pattern matching. *Fuzzy sets and systems*, 28:313–331, 1988.

[20] C. Eden and S. Jones. Using repertory grids for problem construction. *Journal of the Operational Research Society*, 35(9):779–790, 1984.

[21] G. Engelbrecht. *Introduction to Multi-criteria Decision Analysis.* Not yet published.

[22] G. W. Evans. An overview of techniques for solving multiobjective mathematical programs. *Management Science*, 30(11):1268–1282, 1984.

[23] S. French. *Readings in Decision Analysis.* CRC Press, 1989.

[24] S. French. Uncertainty and imprecision: Modelling and analysis. *Journal of the Operational Research Society*, 46:70–79, 1995.

[25] W. Gochet, A. Stam, V. Srinivasan, and S. Chen. Multigroup discriminant analysis using linear programming. *Operations Research*, 45(2):213–225, 1997.

[26] S. Greco, B. Matarazzo, and R. Slowinski. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129:1–47, 2001.

[27] A. Guitouni and J.-M. Martel. Tentative guidelines to help choosing an appropriate MCDA method. *European Journal of Operational Research*, 109:501–521, 1998.

[28] G. A. Hazelrigg. Validation of engineering design alternative selection methods. *Engineering Optimization*, 35(2):103–120, 2003.

[29] R. M. Hogarth and N. Karelaia. Simple models for multiattribute choice with many alternatives: when it does and does not pay to face trade-offs with binary attributes. *Management Science*, 51(12):1860–1872, 2005.

[30] R. L. Keeney. *Value-Focused Thinking.* Harvard University Press, 1992.

[31] R. Lahdelma, J. Hokkanen, and P. Salminen. SMAA - stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, 106:137–143, 1998.

[32] O. Larichev and H. Moshkovich. A approach to ordinal classification problems. *International Transactions in Operational Research*, 1(3):375–385, 1994.

[33] S. Li and A. S. Adams. Is there something more important behind framing? *Organizational Behaviour and Human Decision Processes*, 62(2):216–219, 1995.

[34] Y. Li. *An intelligent, knowledge-based multiple criteria decision making advisor for systems design.* PhD thesis, Scool of Aerospace Engineering, Georgia Institute of Technology, 2007.

[35] B. Malakooti and Z. Yang. Clustering and group selection of multiple criteria alternatives with application to space-based networks. *IEEE Transactions on Systems, Man and Cybernetics part B: Cybernetics*, 33(9):0, 2003.

[36] B. F. Manly. *Multivariate Statistical Methods: A Primer.* Chapman and Hall, 1986.

[37] G. Mendoza and H. Martins. Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest and Ecology Management*, 230:1–22, 2006.

[38] J. Mingers. Variety is the spice of life: Combining soft and hard OR/MS methods. *International Transactions in Operational Research*, 7:673–591, 2000.

[39] V. Mousseau, R. Slowinski, and P. Zielniewicz. A user-oriented implementation of the ELECTRE-TRI method integrating preference elicitation support. *Computers and Operations Research*, 27:757–777, 2000.

[40] P. Nemery and C. Lamboray. Flowsort: A flow-based sorting method with limiting and central profiles. *TOP (Official Journal of the Spanish Society of Statistics and Operations Research)*, to appear.

[41] V. M. Ozernoy. Choosing the "best" multiple criteria decision-making method. *INFOR*, 30(2):159–171, 1992.

[42] P. Perny. Multicriteria filtering methods based on concordance and non-discordance principles. *Annals of Operations Research*, 80:137–165, 1998.

[43] A. B. Petrovsky. Multiple criteria decision making: discordant preferences and problem description. *Journal of Systems Science and Systems Engineering*, 16(1):22–33, 2007.

[44] L. Phillips. A theory of requisite decision models. *Acta Psychologica*, 56:29–48, 1984.

[45] K. Poh. A knowledge-based guidance system for multi-attribute decision-making. *Artificial Intelligence in Engineering*, 12:315–326, 1998.

[46] G. Raisbeck. How the choice of measure of effectiveness constrains operational effectiveness. *Interfaces*, 9(4):85–93, 1979.

[47] B. Roy. The Outranking approach and the foundations of ELECTRE methods. *Theory and Decisions*, 31:49–73, 1991.

[48] B. Roy and P. Vincke. Relational systems of preference with one or more pseudo-criteria: Some new concepts and results. *Management Science*, 30(11), 1984.

[49] L. Scott. Unpacking developmental local government using Soft Systems Methodology and MCDA tools. *ORiON*, 21(2):173–195, 2005.

[50] A. Stam and D. R. Ungar. RAGNU: A microcomputer package for two-group mathematical programming-based nonparametric classification. *European Journal of Operational Research*, 86:374–388, 1995.

[51] T. J. Stewart and F. B. Losa. Towards reconciling outranking and value measurement practice. *European Journal of Operational Research*, 145:645–659, 2003.

[52] T. Tervonen, J. R. Figueira, R. Lahdelma, J. A. Dias, and P. Salminen. A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research*, In Press, Corrected Proof, Available online 17 September 2007, 2007.

[53] A. Tsoukias. On the concept of decision aiding process. *Annals of Operations Research*, 154:3–27, 2007.

[54] E. C. van Schie and J. van der Pligt. Influencing risk preference in decision making: The effects of framing and salience. *Organizational Behaviour and Human Decision Processes*, 63(3):264–275, 1995.

[55] R. von Nitzsch and M. Weber. The effect of attribute ranges on weights in multiattribute utility measurements. *Management Science*, 39(8):937–943, 1993.

[56] F. E. Wenstop and A. J. Carlsen. Ranking hydroelectric power projects with multicriteria decision analysis. *Interfaces*, 18(4):36–48, 1988.

[57] K. P. Yoon and C.-L. Hwang. *Multiple Attribute Decision-Making: An Introduction.* Sage Publications Inc., 1995.

[58] C. Zopounidis and M. Doumpos. Business failure prediction using the UTADIS multicriteria analysis method. *Journal of the Operational Research Society*, 50(11):1138–1148, 1999.

[59] C. Zopounidis and M. Doumpos. Building additive utilities for multi-group hierarchical discrimination: The MH DIS method. *Optimisation Methods and Software*, 14:219–240, 2000.

[60] C. Zopounidis and M. Doumpos. Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research*, 138:229–246, 2002.