



## Multi-criteria decision analysis for use in transport decision making

Barfod, Michael Bruhn; Leleur, Steen

*Publication date:*  
2014

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Barfod, M. B., & Leleur, S. (Eds.) (2014). *Multi-criteria decision analysis for use in transport decision making*. (2 ed.) DTU Transport.

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Multi-criteria decision analysis for use in transport decision making

---



DTU Transport Compendium Series part 2

Version 2014

**Multi-criteria decision analysis for use in transport decision making**

DTU Transport Compendium Series part 2

Department of Transport, Technical University of Denmark

Second Edition, 2014

*Edited by: Michael Bruhn Barfod and Steen Leleur*

*Frontpage: Photo by Colourbox*

# Table of Contents

<b>Abbreviations .....</b>	<b>5</b>
<b>1 Introduction .....</b>	<b>7</b>
<b>2 Principles of multi-criteria decision analysis .....</b>	<b>9</b>
2.1 Schools of MCDA .....	10
2.2 Value measurement .....	10
2.3 Eliciting scores .....	12
2.4 Eliciting weights .....	17
2.5 The strengths of MCDA .....	21
2.6 The weaknesses of MCDA.....	21
2.7 MCDA versus CBA .....	22
<b>3 Main techniques .....</b>	<b>24</b>
3.1 The Simple Multi Attribute Rating Technique (SMART).....	24
3.2 The Analytic Hierarchy Process (AHP) .....	29
3.3 REMBRANDT.....	40
<b>4 Composite model for assessment.....</b>	<b>43</b>
4.1 General principles.....	43
4.2 The COSIMA approach.....	44
4.3 COSIMA calculation example .....	46
<b>5 Case examples.....</b>	<b>51</b>
5.1 Case 1: Composite decision support by combining CBA and MCDA .....	51
5.2 Case 2: Importance of criteria weights in composite appraisal of transport projects.....	60
<b>References.....</b>	<b>71</b>



# Abbreviations

Abbreviation	Full name
AHP	Analytic Hierarchy Process
BCR	Benefit-Cost Rate
CBA	Cost-Benefit Analysis
CI	Consistency Index
COSIMA	COmpoSite Model for Assessment
CR	Consistency Ratio
DMT	Danish Ministry of Transport
DSS	Decision Support System
IRR	Internal Rate of Return
MADM	Multi-Attribute Decision Making
MAUT	Multi-Attribute Utility Theory
MAVT	Multi-Attribute Value Theory
MCDA	Multi-Criteria Decision Analysis
NPV	Net Present Value
SMART	Simple Multi-Attribute Rating Technique
SMARTER	Simple Multi-Attribute Rating Technique Exploiting Ranks
REMBRANDT	Ratio Estimations in Magnitudes and/or deci-Bells to Rate Alternatives which are Non-DominaTed
ROC	Rank Order Centroid
ROD	Rank Order Distribution
RR	Rank Reciprocal
RS	Rank Sum
TV	Total Value
TRR	Total Rate of Return
VF	Value Function



# 1 Introduction

The most common methodology applied so far to the evaluation of transport systems has been conventional cost-benefit analysis (CBA) (Janic, 2003), which supported by traffic- and impact model calculations provides the decision-makers with a monetary assessment of the project's feasibility. A socio-economic analysis is in this respect a further development of the traditional CBA capturing the economic value of social benefits by translating social objectives into financial measures of benefits (Wright et al., 2009). Internationally seen there has been a growing awareness over the recent years that besides the social costs and benefits associated with transport other impacts that are more difficult to monetise should also have influence on the decision making process. This is in many developed countries realised in the transport planning, which takes into account a wide range of impacts of also a strategic character (van Exel et al., 2002). Accordingly, appraisal methodologies are undergoing substantial changes in order to deal with the developments (Vickerman, 2000) that are varying from country to country and leading to different approaches (Banister and Berechman, 2000). It is, however, commonly agreed that the final decision making concerning transport infrastructure projects in many cases will depend on other aspects besides the monetary ones assessed in a socio-economic analysis. Nevertheless, an assessment framework such as the Danish one (DMT, 2003) does not provide any specific guidelines on how to include the strategic impacts; it merely suggests describing the impacts verbally and keeping them in mind during the decision process.

A coherent, well-structured, flexible, straight forward evaluation method, taking into account all the requirements of a transport infrastructure project is for this reason required. An appropriate ex-ante evaluation method for such projects can be based on multi-criteria decision analysis (MCDA) (Tsamboulas, 2007. Vreeker et al. 2002), which in most cases can be combined with a CBA (Leleur, 2000). Scanning the literature (Belton and Stewart, 2002; Goodwin and Wright, 2009; Keeney and Raiffa, 1993; von Winterfeldt and Edwards, 1986) it is found that the use of MCDA in the decision process usually provides some or all of the following features:

1. Improvement of the satisfaction with the decision process
2. Improvement of the quality of the decision itself
3. Increased productivity of the decision-makers

MCDA can in this respect be seen as a tool for appraisal of different alternatives, when several points of view and priorities are taken into account to produce a common output. Hence, it is very useful during the formulation of a decision support system (DSS) designed to deal with complex issues. The literature on DSS is extensive, providing a sound basis for the methodologies employed and the mathematics involved. Moreover, there are numerous systems covering several disciplines, policy contexts and users' needs for specific application environments (Janic, 2003; Salling et al., 2007; Tsamboulas and Mikroudis, 2006). The use of DSS for solving MCDA problems has among others been treated by Barfod (2012), Chen et al. (2008) and Larichev et al. (2002), where it is shown that a DSS can effectively support a decision making process making use of appropriate MCDA methodologies.



## 2 Principles of multi-criteria decision analysis

In multi criteria decision analysis (MCDA), the relative values of different criteria are explicitly subjective – as opposed to the CBA where unit prices reflect some sort of objectivity. The methodological approach taken with MCDA can be indicated with the following quotation by the European Conference of Ministers of Transport (ECMT), Group of Experts (1981):

*Multi-criteria decision analysis is a fairly recent method for assessing and selecting projects exerting complex socio-economic effects. In this method, the individual assessment elements are taken separately and measured in the appropriate dimensions. .. the criteria will have to be weighted among each other because they are not of equal relevance. Determining the weights requires much responsibility and expertise from the decision-maker as the weights have considerable influence on the results of the assessment.*

MCDA stems from the field of operations research and its developers understand it as being different in evaluation approach compared with the economics-based CBA. Thus, in a comprehensive presentation of MCDA methods for regional planning from 1988, it is stated that (Seo and Sakawa, 1988):

*.. there exists the situation where the market price mechanism is not any longer well-functioning and for which alternative evaluation criteria have not yet been well established. The market price mechanism combined with the efficient allocation of resources has not worked as the proper evaluation index for planning. This problem is known as “market failure”. A major subject of MCDM (multi-criteria decision methods) research is thus to resolve the theoretical evaluation problem. ..this research ..highly intends to take problem-solving as well as problem-findings aspects into major consideration: thus this is an “engineering” ..approach in contrast to an “economics” approach..*

Hence, MCDA methods presuppose a preference structure giving preferences on the different criteria. It is emphasised that with an assumption of such a preference structure the methods depend very much on the personality of the decision-maker and the circumstances in which the decision process takes place. The purpose of a MCDA is therefore not searching for some kind of hidden truth – but rather to assist the decision-maker in mastering the (often complex) data involved and advance towards a solution (Gissel, 1999). Effects typically included in a MCDA are wider economic effects, landscape, mobility, network effects, etc.

The essence of a decision support analysis is to break down complicated decisions into smaller pieces that can be dealt with individually and then recombined in a logical way. For the MCDA methods there are basically three such distinct pieces: the set of possible alternatives, their characteristics (represented by a set of criteria), and the preference structure of the decision-maker(s) – reflected in criteria weights.

Generally, the alternatives and their criteria represent the objective part of the decision process, whereas the subjective part of the decision process lies in the preference structure. However, in the case where a given criterion cannot be quantified in an obvious way, the decision-maker and the analyst may be forced to make subjective assessments of the criteria scores, or they will have to find a surrogate measure that can function as a good proxy for the criteria. As discussed in Gissel (1999) the use of proxies should be preferred to subjective scores whenever possible. The principal argument for this is to restrict the

subjectivity in the decision process to elements for which a constructive exchange of (political) opinion can take place.

There exists a wide variety of MCDA methods representing a corresponding variety in methodological approaches. However, the fundamental structure of the methods is generally the same.

## 2.1 Schools of MCDA

The existing variety of MCDA methods can be categorized into two overall *schools*; The American school and the French school.

In the American school the view of the decision-maker is disaggregate in the sense that the decision-maker is assumed to have a complete preference system. This preference system enables him to express his preferences on all aspects of the decision problem, and it may be derived through asking the decision-maker relevant questions. Lootsma (1999) refers to this approach as *the normative approach*. Typical examples of this approach are the multi-attribute value theory (MAVT) and the multi-attribute utility theory (MAUT). The Simple Multi-Attribute Rating Technique (SMART) is a variant of the MAVT. This method will be explored further below as will the Analytic Hierarchy Process (AHP).

The French school covers the family of outranking methods. In the French school the existence of a well-ordered preference system is questioned, the view is more that of the decision-maker as a rational economic man (Gissel, 1999). The basic assumption is that the decision-maker explores the assertion that “alternative *i* is at least as good as alternative *k*”, and the only pre-existing preferences he has is an idea of the relative importance of the criteria. Lootsma (1999) refers to this approach as *the constructive approach*.

In this compendium the focus is on the American school of MCDA methods.

## 2.2 Value measurement

The purpose of value measurement is to produce a means of associating a real number with each alternative in an assessment, in order to construct a preference order of the alternatives consistent with decision-maker value judgments. In other words, it is desirable to associate a number or value,  $V(a)$ , with each alternative,  $a$ , in such a way that  $a$  is judged to be preferred to  $b$ , taking all criteria into account, if and only if  $V(a) > V(b)$ . This also implies indifference between  $a$  and  $b$  if and only if  $V(a) = V(b)$ . Note, that the preference order implied by any such value function must constitute a complete weak order or preorder (Belton and Stewart, 2002), i.e.:

*Preferences are complete:* For any pair of alternatives, either one is strictly preferred to the other or there is indifference between them.

*Preferences and indifferences are transitive:* For any three alternatives, e.g.  $a$ ,  $b$  and  $c$ , if  $a$  is preferred to  $b$ , and  $b$  is preferred to  $c$ , then  $a$  is preferred  $c$ , and similarly for indifference.

The value measurement approach thus constructs preferences which, in the first instance are required to be consistent with a relative strong set of axioms. However, it is important to note, that in practice value measurement will not be applied with such a literal and rigid view of these assumptions. The construction

of a particular value function does impose the discipline of coherence with these “rationality assumptions”, but the results of and the conclusions from the value function will be subjected to intensive sensitivity analyses. The end result will generally be much less rigidly precise than may be suggested by the axioms.

Within the value measurement approach, the first component of preference modelling (measuring the relative importance of achieving different performance levels for each identified criterion) is achieved by constructing “marginal” (or “partial”) value functions,  $v_i(a)$ , for each criterion. A fundamental property of the partial value function is that alternative  $a$  is preferred to alternative  $b$  in terms of criterion  $i$  if and only if  $v_i(a) > v_i(b)$ . Similarly, indifference between  $a$  and  $b$  in terms of this criterion exist if and only if  $v_i(a) = v_i(b)$ . Thus the partial value function satisfies the definition of a preference function (see Belton and Stewart, 2002, p. 83). However, the partial value functions will in addition need to model strength of preference in some sense, so that stronger properties than simple preservation of preference ordering will in general be needed.

Value function methods produce the assessments of the performance of alternatives against individual criteria, together with inter-criteria information reflecting the relative importance of the different criteria,  $w_i$ , to give an overall evaluation of each alternative indicative of the decision-makers preferences. The simplest and most widely used form of value function method is the additive model (Belton and Stewart, 2002, p. 86):

$$V(a) = \sum_{i=1}^m w_i v_i(a) \quad (2.1)$$

Considerably more complicated in appearance, but as easy to use, is the multiplicative model (Ibid, p. 94):

$$V(a) = \prod_{i=1}^m [v_i(a)]^{w_i} \quad (2.2)$$

In its analytical expansion the multiplicative model seems prohibitive compared to the additive model. However, it requires only the addition of a single parameter ( $w$ ), which defines all interaction terms. Therefore, the type of interaction it models is rather constrained (Von Winterfeldt and Edwards, 1986). Additive aggregation is the form that is most easily explained and understood by decision-makers from a wide variety of backgrounds, while not placing any substantially greater restrictions on the preference structures than more complicated aggregation formulae (Belton and Stewart, 2002, p.86).

In general the partial value functions should be standardised in a well-defined manner as will be described below. This is most easily done for criteria associated with measureable attributes, but it can be done quantitatively in other cases. Once an initial model structure like the above and a set of alternatives for evaluation have been defined, the next step will be to elicit the information required by the model. There are two types of information, sometimes referred to as intra-criterion information and inter-criterion information, or alternatively as scores and weights.

## 2.3 Eliciting scores

Scoring is the process of assessing a value derived by the decision-maker from the performance of alternatives against the relevant criteria. That is, the assessment of the partial value functions,  $v_i(a)$  in the above model. If the criteria are structured as a value tree then the alternatives must be scored against each of the bottom level criteria. These values need to be assessed on an interval scale of measurement, i.e. a scale on which the difference between points is the important factor. A ratio of values will only have meaning if the zero point on the scale is absolutely and unambiguously defined. Thus to construct a scale it is necessary to define two reference points and to allocate numerical values to these points. The minimum and maximum points on the scale can be defined in a number of ways, e.g. 0 and 100, but it is important to distinguish between a local and a global scale:

A **local scale** is defined by the set of alternatives that are under consideration. The alternative which does best on a particular criterion is assigned a score of 100 and the one that does least well is assigned a score of 0. All other alternatives will receive intermediate scores which reflect their performance relative to the end points. The use of local scales permits a relative quick assessment of values and can be very useful for initial “roughing out” of a problem, or if operating under time constraints.

A **global scale** is defined by reference to the wider set of possibilities. The end points may be defined by the ideal and the worst conceivable performance on the particular criterion, or by the best and worst performance that can realistically occur. The definition of a global scale requires more work than a local scale. However, it has the advantages that it is more general than a local scale and that it can be defined before consideration of specific alternatives. This also means that it is possible to define criteria weights before consideration of alternatives.

Valid partial value functions can be based on either local or global scales. The important point is that all following analysis, including assessment of the weights ( $w_i$ ), must be consistent with the chosen scaling. Once the reference points of the scale have been determined consideration must be given to how other scores are to be assessed. This can be done in one of the following three ways (Belton and Stewart, 2002):

1. **Definition of a partial value function.** This relates to performance in terms of a measurable attribute reflecting the criterion of interest.
2. **Construction of a qualitative value scale.** In this case, the performance of the alternatives can be assessed by reference to descriptive pointers, or word models to which appropriate values are assigned.
3. **Direct rating of the alternatives.** In this case, no attempt is made to define a scale which characterises performance independently of the alternatives being evaluated. The decision-maker simply specifies a number, or identifies the position on a visual analogue scale, which reflects the value of an alternative in relation to the specified reference points.

### 2.3.1 Definition of a partial value function

The first step in defining a value function is to identify a measurable attribute scale which is closely related to the decision-makers values. If it is not possible to identify an appropriate quantitative scale, or if such

scales as are available are only remotely related to the decision-makers values then it will be necessary to construct a value scale (this will be described in the next section). The value function reflects the decision-makers preferences for different levels of achievement on the measurable scale. Such a function can be assessed directly or by using indirect questioning. Direct assessment will often utilise a visual representation.

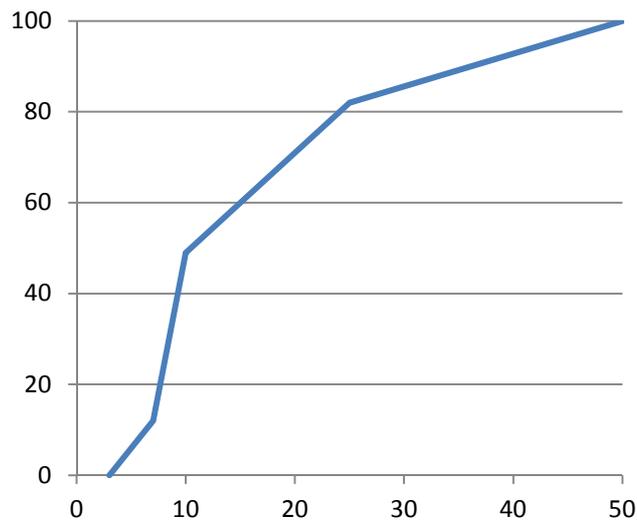
When making direct assessment of a value function the decision-maker should begin by determining whether:

- The value function is monotonically increasing against the natural scale, i.e. the highest value of the attribute is the most preferred and the lowest value the least preferred.
- The value function is monotonically decreasing against the natural scale, i.e. the lowest value of the attribute is the most preferred and the highest value the least preferred. This is e.g. the case with cost criteria.
- The value function is non-monotonic, i.e. an intermediate point on the scale defines the most preferred or least preferred point.

Von Winterfeldt and Edwards (1986) suggest that if the value tree has been well structured then the value functions should be regular in form, i.e. no discontinuities. They go further to argue that all value functions should be linear or close to linear and suggest that the analyst should consider restructuring a value tree to replace non-monotonic value functions by one or more monotonic functions. Whilst Belton and Stewart (2002) agree that an extremely non-linear value function, in particular a non-monotonic function, may indicate a need to revisit the definition of criteria, they caution against over-simplification of the problem by inappropriate use of linear value functions. Experimental simulations of Stewart (1993, 1996) suggest that the results of analyses can be sensitive to such assumptions. Thereby, the default assumption of linearity, which is often made, may generate misleading answers.

Indirect assessment methods assume that the value function is monotonically increasing or decreasing over the range of attribute measurement considered. The end points of the scale must, as previously mentioned, be defined first. Thereafter, two methods of assessment are widely used, namely the bisection and the difference methods.

Using the *Bisection method* the decision-maker is asked to identify the point on the attribute scale which is halfway, in value terms between the two endpoints. To help the decision-maker identify the midpoint value it may be helpful to begin by considering the midpoint on the objective scale and then pose a question regarding which of the two half's increase is the most valuable. The considered point can then be moved towards the most preferred half and the question repeated until the midpoint is identified. The next step would then be to find the midpoints between the two endpoints and the previous found midpoint. It is generally accepted that 5 points (2 endpoints and 3 "midpoints") give sufficient information to enable the analyst to sketch in the value function, see Figure 2.1.



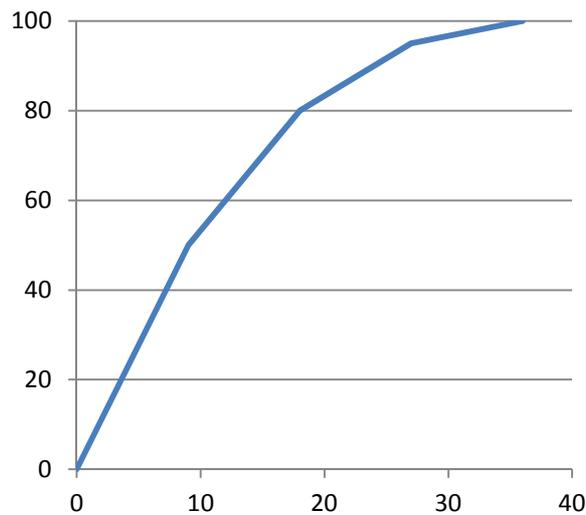
**Figure 2.1:** Example of a value function sketched using the bisection method

*Difference methods* could be viewed as a collection of methods rather than a single one, but all of them require the decision-maker to consider increments on the objectively measured scale and to relate these to difference in values. In the first approach as described by Watson and Buede (1987), the attribute scale is divided into, e.g., four equal intervals. To illustrate this approach, consider an example where a new department of a company has to hire people (besides the management) in order to make the department run. For the criterion “number of people” the minimum number is 0, and the maximum 36. Since preference is for more people, an increase in the number results in an increase in value. The decision maker is asked to rank order the specified differences according to increase in associated value. For example, is the increase in value which occurs in going from 0 to 9 greater than, equal to or less than the increase in value achieved in going from 9 to 18? Suppose the information from the decision maker is as given below in Table 2.1.

**Table 2.1:** Intervals on the criterion “number of people”

Increase in number of people		Increase in value
From	To	
0	9	1 = greatest
9	18	2
18	27	3
27	36	4

The ranking gives an idea of the shape of the value function. In this example the increase in value is greatest for low numbers of people, suggesting a concave, increasing value function. The curve could be sketched directly on the basis of this information, as illustrated in Figure 2.2, or may be further refined by asking the decision maker to assess the relative magnitude of value increases.



**Figure 2.2:** Example of a value function sketched using the difference method

Another approach described by von Winterfeldt and Edwards (1986) is to begin by defining a unit level on the attribute scale (between one tenth and one fifth of the difference between the minimum and maximum points is suggested). Consider again the criterion “number of people”, measured as above. The minimum and maximum points on this scale is 0 and 36 people, thus let the specified unit be equal to 4 people (close to one tenth of the range). To assess the value function using this method we would first ask: What is the number of people,  $P$ , such that an increase from 4 to  $P$  people results in the same increase in value as an increase from 0 to 4 people? Suppose the decision maker suggest that  $P$  should be 9. We next pose the question: What is the value of  $P$  such that an increase from 9 to  $P$  people is equal in value to the increase from 4 to 9? The decision maker responds that it would be necessary to double the number of people in order to achieve the same increase in value. The additional value of 9 extra people then diminishes further, the increase from 18 to 36 perhaps equating in value to the increase from 9 to 18, but beyond 36 extra people do not add value. These responses give rise to a value function, specified in Table 2.2, which is very similar in shape to that defined using the previous method.

**Table 2.2:** Value function for the criterion “number of people”

Number of people	Value (per units defined above)	Value (0 to 100 scale)
0	0	0
4	1	25
9	2	50
18	3	75
36	4	100

The measurement scales used for the assessment of value functions for the two examples above arise naturally in the given context. Von Winterfeldt and Edwards (1986, p. 221) comment that:

*“A natural scale which is linear in value is obviously the most economical device for communicating value relevant information”*

However, in some instances a simple natural scale may not exist and it becomes necessary to construct an appropriate measurement scale.

### **2.3.2 Construction of qualitative value scales**

Often it is not possible to find a measurable attribute which captures a criterion. In such circumstances it is necessary to construct an appropriate qualitative scale. As discussed in the previous section, it is necessary to define at least two points on the scale (often taken as the end points). Intermediate points may also be defined. An example of such a scale in regular use is the Beaufort for measuring the speed of wind (Belton and Stewart, 2002). Points on the scale are defined descriptively and draw on multiple concepts in the definition. An alternative approach to defining a scale could be to associate specific alternatives, with which the decision makers are familiar, with points on the scale.

Qualitative scales should have the following characteristics (Belton and Stewart, 2002, pp. 128-129):

- *Operational*: allow the decision makers to rate alternatives not used in the definition of the scale
- *Reliable*: two independent ratings of an alternative should lead to the same score
- *Value relevant*: relates to the decision makers' objective
- *Justifiable*: an independent observer could be convinced that the scale is reasonable

The approach described above directly assigns values to the qualitative statements. The MACBETH system by Bana e Costa and Vansnick (1994) can, as an example, be used to build a value scale from a category scale by a process of pair wise comparisons requesting ordinal judgements about preference differences. The output of the MACBETH system is a range of values associated with each category, consistent with the judgments input to the analysis. The decision maker may choose to work with the midpoints of these intervals as the corresponding value scale, or may wish to further refine the input judgments to arrive at a tighter definition of values. It is possible that the initial judgments are ordinal inconsistent, in which case the method highlights inconsistencies and suggest revisions which would move towards consistency.

### **2.3.3 Direct rating**

Direct rating can be viewed as the construction of a value scale, but defining only the end points of the scale. A local or global scale can be used, the former creating minimal work for the decision makers. If using a local scale, the alternative which performs best of those under consideration is given the highest score, usually 100, and the alternative which performs least well (not necessarily badly in any absolute sense) is given a score of 0. All other alternatives are positioned directly on the scale to reflect their performance relative to the two end points. Although no attempt is made to relate performance to a measurable scale, the positioning of alternatives can generate extensive discussion, yielding rich information on the decision makers' values. Ideally this information should be recorded for future reference. A disadvantage of using a local scale is that if new alternatives are introduced into the evaluation this may necessitate the revision of scales, something which has consequences for the weighting of criteria.

## Direct rating by pair wise comparisons

The use of pair wise comparisons is implicit in all scoring procedures as scores are assessed relative to reference points rather than in an absolute sense. Furthermore, in order to check consistency of judgments a facilitator may incorporate questioning procedures which make explicit pair wise comparisons between alternatives. However, even if explicit, such comparisons tend to be ad-hoc and do not consider all possible comparisons. A systematic pair wise comparison approach is one of the cornerstones of the Analytic Hierarchy Process (AHP) by Saaty (1977). The AHP employs a procedure for direct rating which requires the decision maker to consider all possible pairs of alternatives with respect to each criterion in turn, to determine which of the pair is preferred and to specify the strength of preference according to a semantic scale (or associated numeric 1-9 scale). However, the AHP treats responses as ratio judgments of preferences, which is not consistent with the value function approach. The underlying mathematics is easily modifiable to be consistent with difference measurement. The MACBETH approach mentioned before, which is founded on difference measurement and also based on pair wise comparisons, can be used to derive direct ratings. An additional approach, which can be used for the purpose of deriving direct ratings, is REMBRANDT (Lootsma, 1992). The approach is also based on pair wise comparisons and overcomes some of the problems with the underlying mathematics of AHP by using a logarithmic scale and the geometric mean method, see Olson et al (1995).

One of the potential drawbacks of pair wise comparison methods is the large number of judgments required of the decision maker ( $n(n-1)/2$ ) for each criterion, where  $n$  is the number of alternatives). Nevertheless, the approach is powerful and can be effectively utilised if decision makers find the direct rating procedure difficult. With some pair wise comparison approaches it is not necessary to compare all possible pairs and considerable work has been done to derive appropriate sampling procedures (Belton and Stewart, 2002).

## 2.4 Eliciting weights

It is clear that in any evaluation not all criteria carry the same weight, thus it is desirable to incorporate an assessment of the relative importance of criteria. This aspect of analysis has been the focus of extensive debate (Belton and Stewart, 2002). Decision makers are able and willing to respond to questions like: “what is most important to you when choosing a new car, safety or image?” Furthermore, they are able and willing to respond to questions asking them to rate the relative importance of safety and image against a numerical or verbal scale. The AHP is, as mentioned earlier, founded on such questions. However, it has been argued (Ibid.) that the responses to such questions are essentially meaningless. The questions are open to many different interpretations, people do not respond to them in a consistent manner and responses do not relate to the way in which weights are used in the synthesis of information. The weights which are used to reflect the relative importance of criteria in a multi-attribute value function are, however, well defined. The weight assigned to a criterion is essentially a scaling factor which relates scores on that criterion to scores on all other criteria. Thus if criterion A has weight which is twice that of criterion B this should be interpreted as the decision maker values 10 value points on criterion A the same as 20 value points on criterion B and would be willing to trade one for the other. These weights are often referred to as swing weights to distinguish them from less well defined concept of importance weights. Thus the notion of swing weights captures both the psychological concept of “importance” and the extent

to which the measurement scale adopted in practice discriminates between alternatives. One of the commonest errors in naive scoring models is to assume that weights are independent of the measurement scales used. It is clear from the algebraic structure of (2.1), however, that the effect of the weight parameter  $w_i$  is directly connected to the scaling used for  $v_i(a)$ , so that the two are intimately connected.

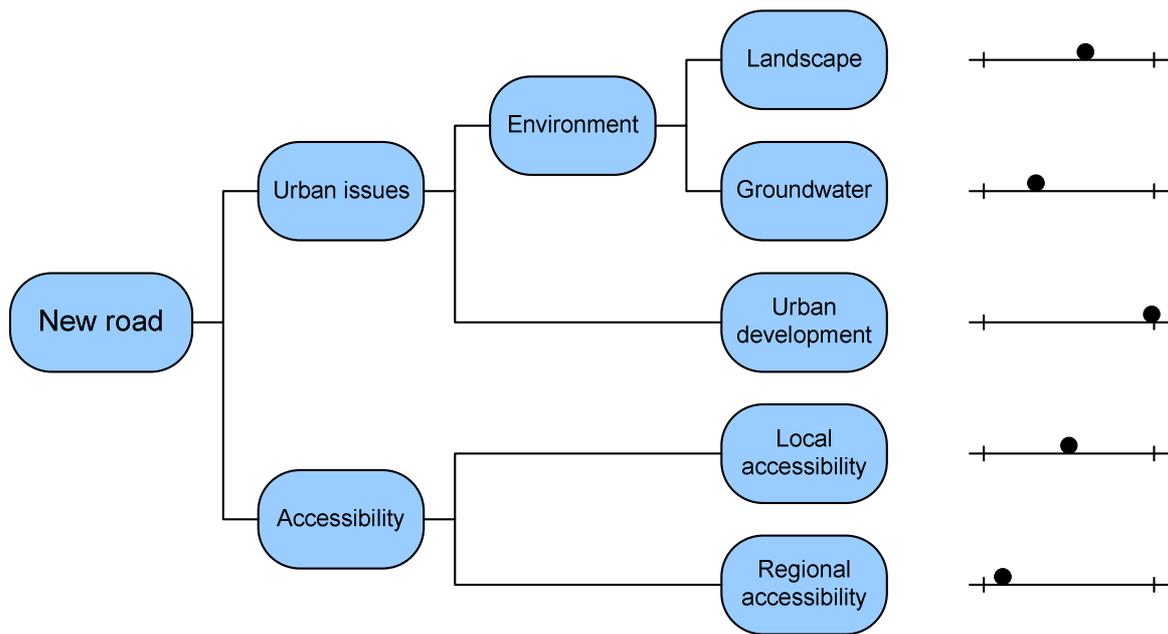
### 2.4.1 Swing weights

The swing which is usually considered is that from the worst value to the best value on each criterion. If the value tree is small, then the decision maker may be asked to consider all bottom level criteria simultaneously and to assess which swing gives the greatest increase in overall value; this criterion will have the highest weight. The process is repeated on the remaining set of criteria, and so on, until the order of benefit resulting from a swing from worst to best on each criterion has been determined, thereby defining a ranking of the criteria weights. To assign values to the weights the decision maker must assess the relative value of the swings. For example, if a swing from worst to best on the most highly weighted criterion is assigned a value of 100, what is the relative value of a swing from worst to best on the second ranked criterion? It is important to remember that these weights are dependent on the scales being used for scoring as well as the intrinsic importance of the criteria. This means that it is not possible to assign swing weights until the scales for each criterion have been defined. If an intrinsically important criterion does not differentiate much between the options – that is, if the minimum and maximum points on the value scale correspond to similar levels of performance – then that criterion may be ranked quite low.

Note that, although it has been customary to derive swing weights by reference to swings over the whole range of value measurement, it is perfectly valid to use any two reference points on the criteria scales. Bana e Costa and Vansnick (1994) have used definitions of “neutral” and “good” as reference points. Decision makers may feel more at ease comparing swings standardised in this way rather than swings between extreme points, particularly if the degree of differentiation differs substantially across criteria (for example, on one criterion the “worst” and “best” reference points may both represent very acceptable performance, discriminating little between the alternatives under consideration, whereas on another “worst” may be truly awful, whereas “best” is extremely good).

Having established a rank order for the criteria weights, the next step is to assign values to them. Once again, there are a number of ways of doing this. The decision maker could be asked directly to compare each of the criteria in turn with the most highly ranked criterion. For each criterion the decision maker is asked to assess the increase in overall value resulting from an increase from a score of 0 to 100 on the most highly ranked criterion.

Decision makers are generally comfortable working with visual analogue and may be willing to assess the relative magnitude of the swing weights directly using this means, as illustrated in Figure 2.3. These provide a means for communicating a good sense of the magnitude of judgments whilst removing the need for numerical precision. However, it is important that this degree of imprecision is not forgotten when information is aggregated.



**Figure 2.3:** Swing weights – visual analogue scale (adapted from Belton and Stewart (2002))

## 2.4.2 Normalisation

The weights implied by the visual representation in Figure 2.3 may be translated into numerical values, as shown in Table 2.3 below. The second column of the table lists the weights as they are displayed in Figure 2.3, i.e. standardised with the largest weight set to 1. It is usual, although not essential, to normalise weights to sum to 1 or 100, as shown in the third column of Table 2.3. Such normalisation allows the decision makers to interpret for example the weight of landscape in Table 2.3 as constituting 24% of the total importance weight. This seems often to be a useful interpretation. However, in specific cases decision makers may find it more intuitive to specify a reference criterion whose units are weighted 1 and against which all other criteria are compared, as shown by the original weights with urban development as the reference criterion.

**Table 2.3:** Swing weights – original and normalised values

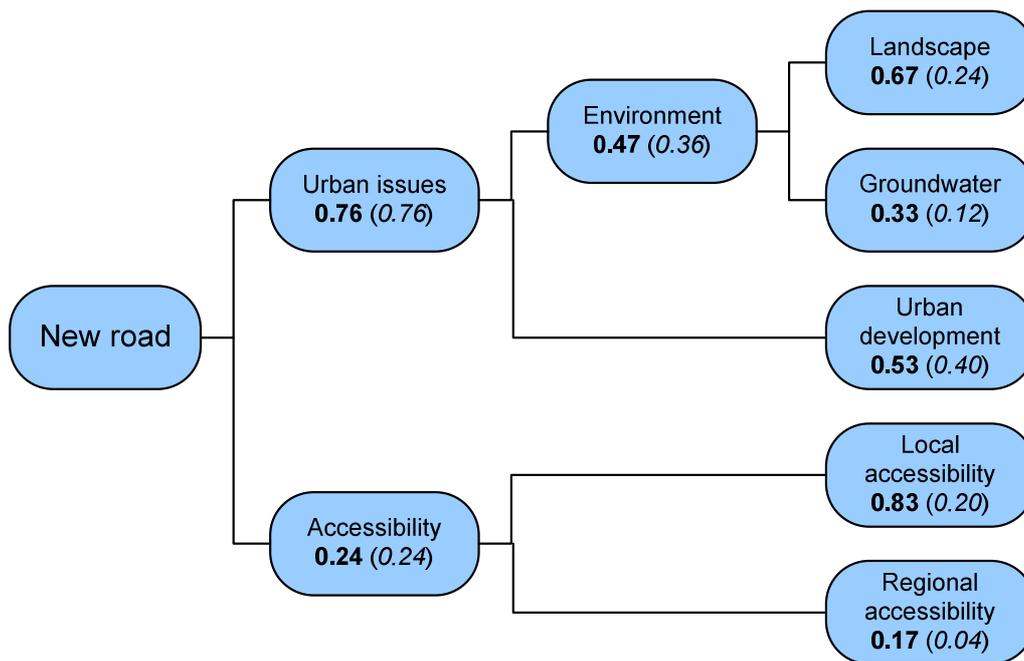
Criterion	Original weights	Normalised weights
Landscape	0.6	0.24
Groundwater	0.3	0.12
Urban development	1.0	0.40
Local accessibility	0.5	0.20
Regional accessibility	0.1	0.04

### 2.4.3 Weights in value trees

When the problem is structured as a multi-level value tree consideration has to be given to weights at different levels of the tree. It is useful to define relative weights and cumulative weights. Relative weights are assessed within families of criteria – i.e. criteria sharing the same parent – the weights within each family being normalised to sum to 1 (or 100). The cumulative weight of a criterion is the product of its relative weight in comparison with its siblings and the relative weights of its parent, parent’s parent, and so on to the top of the tree.

By definition, the cumulative weights of all bottom-level criteria (leaves on the tree) sum to 1 (or 100) – thus the normalised weights shown in Table 2.3 are cumulative weights. The cumulative weight of a parent criterion is the total of the cumulative weights of its descendants.

As illustrated for the example problem, if the value tree does not have too many leaves, then weights can be assessed by directly comparing all bottom-level criteria to give the cumulative weights. Weights at higher levels of the tree are then determined by adding the cumulative weights of all members of a family to give the cumulative weight of the parent. Relative weights are determined by normalising the cumulative weights of family members to sum to 1. Relative and cumulative weights for the example problem are illustrated in Figure 2.4.



**Figure 2.4:** Relative weights (in bold) and cumulative weights (in italics)

For larger models it is easier to begin by assessing relative weights within families of criteria. Weights at higher levels of the value tree can be assessed top-down or bottom-up. The top-down approach would assess relative weights within families of criteria by working from the top of the tree downwards. However, the analyst must be aware of the difficulty of interpreting weights at higher levels of a value tree – the weight of a higher level criterion is the sum of the cumulative weights of all its sub-criteria. Thus, in comparing two higher level criteria the decision maker should be thinking in terms of a swing from 0 to 100

on all sub-criteria of the two higher level criteria. If the top-down approach is used it is important to carry out cross family checks on the cumulative weights of bottom level criteria.

The bottom-down approach begins by assessing relative weights within families which contains only bottom level criteria and then carrying out cross family comparisons using one criterion from each family (perhaps the most highly weighted criterion in each family) and comparisons with any unitary bottom level criteria. This process would eventually give the cumulative weights of the bottom level criteria which can be aggregated to higher levels as described before.

## **2.5 The strengths of MCDA**

The strengths of the MCDA can be summarised in the following based on MOTOS (2007):

- MCDA overcomes most measurement problems
- Participation of the decision-maker(s)
- MCDA can be accommodated to address equity concerns

Firstly, MCDA overcomes the difficulty of translating all values into monetary units by using subjective weights. In addition, both qualitative and quantitative indicators can be used depending on the criteria, and the time and resources available. That is, if an impact cannot be quantified (be it because of scarcity of time or resources) it may instead be represented by some sort of indicator (a proxy or a subjective score).

Secondly, stakeholders can be involved throughout the decision making process to determine the alternatives and criteria, the criteria weights, and to score and determine the best solution. The technique offers – in fact it often requires – a more participatory approach as it takes decisions out of the hands of analysts and puts it with those stakeholders involved.

Thirdly, MCDA can address equity concerns by incorporating equity criteria into the analysis, and it allows each individual the same representation, unlike the market (i.e., CBA) where those with greater money exert greater influence.

MCDA may be seen as an extension of the CBA. In practice, the CBA represents only a part of the decision making basis; other non-monetised impacts represent another, and the final choice is based on a weighing of these different parts. Using only CBA this weighing may be completely opaque. The MCDA methods offer a tool for approaching the subjectivities in the decision process and for reaching a decision in a methodological and transparent way. In MCDA, both the monetised impacts of the CBA as well as more strategic impacts can be accommodated in one approach.

## **2.6 The weaknesses of MCDA**

Of course, there are also problems associated with the use of MCDA. In short these may be presented as the following based on MOTOS (2007):

- The method can give no “absolute” measure of “goodness” – it is a tool for comparative evaluation only.

- The participatory nature of the MCDA makes it both time and resource intensive.
- Difficulties in deriving criteria weights.

Firstly, the method is a tool for comparative evaluation only, as it gives no absolute measure of the “goodness” of the project or policy as does the CBA with its socio-economic profitability. This makes MCDA a tool for deciding between options and not for a “go/no-go” decision.

Secondly, the active involvement of the decision-maker(s) makes the MCDA both time and resource intensive as it requires much from decision-makers. However, in our opinion, this should not necessarily be seen as a drawback of the method. Decisions are seldom objective, and if subjective judgments are present these should be transparent – at least to the decision-maker himself. MCDA offers a tool for obtaining this insight.

Thirdly, there can be problems with the elicitation of criteria weights. The analyst should be very aware that the derivation of weights is a fundamental and critical step in the MCDA. Different MCDA methods may require different types of weights derived in different ways. In the weights derivation process, the decision-maker should be helped to understand the meaning and importance of his stated weights to increase the understanding and acceptance of the MCDA with the decision-maker. This understanding and acceptance is crucial for the applicability of the method.

## **2.7 MCDA versus CBA**

The overall feature of the cost-benefit analysis (CBA) is that of comparing costs and benefits with all such elements measured on the same scale; that of monetary units. This means that all relevant impacts have to be assigned a unit price. The basic principle underlying the CBA is that of maximizing the net socio-economic benefit of the project, which may be seen as society’s welfare gain. As society in some sense consists of the sum of its individuals, it is natural to see the social change in welfare from a given investment as the aggregate value of the individual utility gains and losses. This means that there is an underlying assumption that social decisions can and should be founded on the aggregation of individuals’ willingness to pay (Grant-Muller et al., 2001).

The appealing features of the CBA are quite convincing. Firstly, the CBA provides a methodological tool for comparing projects and/or alternatives, which makes it a powerful decision support tool in the planning process. Secondly, the CBA converts all social implications into an absolute monetary measure of the social profitability. It is desirable to be able to sum up all aspects of the decision problem in one simple value. Thirdly, values on cost and benefit elements are consistent between investments and over time. This means that social profitabilities can be compared across projects and at different points in time. Fourthly, the CBA requires the collection of detailed information of financial as well as social costs and benefits. This gathering of information improves the basis on which the decision is made.

There are of course also problems associated with the CBA method. Firstly, it is difficult to maintain consistency between the theoretical assumptions of the method and the practical application of it, due to the fact that there may be problems involved when estimating unit prices for non-marketed impacts such as travel time savings, emissions, safety, etc. In practice, therefore, compromises are often made on the valuation of such non-marketed impacts, implying that the resulting unit prices are inherently of a

subjective nature – without such subjectivities being visible in the evaluation. This is a problem with the CBA method since the presentation of a single evaluation measure thus implies a “false air of objectivity”. Secondly, there are impacts for which it is difficult or even impossible to estimate unit prices. These are especially impacts of a more long-term and/or strategic nature. Thirdly, an important philosophical and moral problem in the evaluation of long term impacts is that of the *present* generation valuing an impact which they may not live to experience. This means that they are valuing such impacts on behalf of the *future* generation(s). The final problem with CBA to be mentioned here is that although the method rests on the aggregation of individuals’ willingness to pay, no actual payment takes place and no actual redistribution of money results. Hence, the socio-economic optimum resulting from the CBA could be argued on equity grounds as being somewhat hypothetical.

In multi-attribute decision aid methods (a family of multi-criteria methods, MCDA) the subjectivity in the decision making process is explicitly recognized as such methods require decision makers to express their specific preferences. Multi-attribute methods consist of three basic elements: the set of alternatives; the set of attributes describing the alternatives; and the preference structure of the decision maker(s). It is this last element which makes the subjectivity in evaluation an explicit part of the method.

MCDA may be seen as an extension of the CBA. In practice, the CBA represents only a part of the decision making basis; other non-monetised impacts represent another, and the final choice is based on a weighing of these different parts. Using only CBA this weighing may be completely opaque. The MCDA methods offer a tool for approaching the subjectivities in the decision process and for reaching a decision in a methodological and transparent way. In MCDA both the monetised impacts of the CBA as well as more strategic impacts can be accommodated in *one* approach. This also implies that equity considerations can be explicitly accounted for in the MCDA. The applicability of the two methods in different planning situations can be illustrated by Table 2.4.

**Table 2.4:** Applicability of CBA and MCDA under various conditions

	<b>Most important effects can be quantified</b>	<b>Few effects can be quantified</b>
No significant effects of strategic / political nature	CBA	MCDA
One or more effects of strategic / political nature	CBA + MCDA	MCDA

## 3 Main techniques

The following sections will elaborate on the two main techniques SMART and AHP as well as some of their applications.

### 3.1 The Simple Multi Attribute Rating Technique (SMART)

The SMART technique is based on a linear additive model. This means that an overall value of a given alternative is calculated as the total sum of the performance score (value) of each criterion (attribute) multiplied with the weight of that criterion as noted in (2.1) in Section 2.2.

The main stages in the analysis are (adapted from Olson (1996)):

- **Stage 1:** *Identify the decision-maker(s)*
- **Stage 2:** *Identify the issue of issues:* Utility depends on the context and purpose of the decision
- **Stage 3:** *Identify the alternatives:* This step would identify the outcomes of possible actions, a data gathering process.
- **Stage 4:** *Identify the criteria:* It is important to limit the dimensions of value. This can be accomplished by restating and combining criteria, or by omitting less important criteria. It has been argued that it was not necessary to have a complete list of criteria. Fifteen were considered too many, and eight was considered sufficiently large. If the weight for a particular criterion is quite low, that criterion need not be included. There is no precise range of the number of criteria appropriate for decisions.
- **Stage 5:** *Assign values for each criteria:* For decisions made by one person, this step is fairly straightforward. Ranking is a decision task that is easier than developing weights, for instance. This task is usually more difficult in group environments. However, groups including diverse opinions can result in a more thorough analysis of relative importance, as all sides of the issue are more likely to be voiced. An initial discussion could provide all group members with a common information base. This could be followed by identification of individual judgments of relative ranking.
- **Stage 6:** *Determine the weight of each of the criteria:* The most important dimension would be assigned an importance of 100. The next-most-important dimension is assigned a number reflecting the ratio of relative importance to the most important dimension. This process is continued, checking implied ratios as each new judgment is made. Since this requires a growing number of comparisons there is a very practical need to limit the number of dimensions (objectives). It is expected that different individuals in the group would have different relative ratings.
- **Stage 7:** *Calculate a weighted average of the values assigned to each alternative:* This step allows normalization of the relative importance into weights summing to 1.
- **Stage 8:** *Make a provisional decision*

- **Stage 9: Perform sensitivity analysis**

In SMART, ratings of alternatives are assigned directly, in the natural scales of the criteria. For instance, when assessing the criterion "cost" for the choice between different road layouts, a natural scale would be a range between the most expensive and the cheapest road layout. In order to keep the weighting of the criteria and the rating of the alternatives as separate as possible, the different scales of criteria need to be converted into a common internal scale. In SMART, this is done mathematically by the decision-maker by means of a Value Function. As mentioned in the previous Section 2.2 the simplest and most widely used form of a value function method is the additive model, which in the most simple cases can be applied using a linear scale (e.g. going from 0 to 100).

### 3.1.1 SMART Exploiting Ranks (SMARTER)

The assessment of value functions and swing weights in SMART can sometimes be a difficult task, and decision-makers may not always be confident about it. Because of this, Edwards and Barron have suggested a simplified form of SMART named SMARTER (SMART Exploiting Ranks) (Roberts and Goodwin, 2002). Using the SMARTER technique the decision-makers places the criteria into an importance order: for example 'Criterion 1 is more important than Criterion 2, which is more important than Criterion 3, which is more important Criterion 4' and so on,  $C_1 \geq C_2 \geq C_3 \geq C_4 \dots$ . SMARTER then assigns surrogate weights according to the Rank Order Distribution method or one of the similar methods which are described below.

Barron and Barret (1996) believe that generated weights may be more precise than weights produced by the decision-makers who may be more comfortable and confident with a simple ranking of the importance of each criterion swing, especially if it represents the considered outcome of a group of decision-makers. Therefore a number of methods that enable the ranking to be translated into 'surrogate' weights representing an approximation of the 'true' weights have been developed. A few of these methods are described below. Here  $W_j > 0$  are weights reflecting the relative importance of the ranges of the criteria values, where  $\sum_{j=1}^n w_j = 1$ ,  $i = 1, \dots, n$  is the rank of the criteria, and  $n$  is the number of criteria in the decision problem.

*Rank order centroid (ROC) weights:* The ROC weights are defined by (Roberts and Goodwin, 2002):

$$w_i(ROC) = 1/n \sum_{j=1}^n 1/j, i = 1, \dots, n \quad (3.1)$$

*Rank sum (RS) weights:* The RS weights are the individual ranks normalized by dividing by the sum of the ranks. The RS weights are defined by (Ibid):

$$w_i(RS) = (n + 1 - i) / n(n + 1)/2, i = 1, \dots, n \quad (3.2)$$

*Rank reciprocal (RR) weights:* This method uses the reciprocal of the ranks which are normalized by dividing each term by the sum of the reciprocals. The RR weights are defined by (Ibid):

$$w_i(RR) = \frac{1}{i} / \sum_{j=1}^n \frac{1}{j}, \text{ rank } i = 1, \dots, n, \quad \text{option } j = 1, \dots, n \quad (3.3)$$

For each of these methods, the corresponding weights for each rank, for numbers of criteria ranging from  $n = 2 - 10$  are listed in Table 3.1 - Table 3.3.

**Table 3.1:** (ROC) weights (Roberts and Goodwin, 2002)

Criteria									
Rank	2	3	4	5	6	7	8	9	10
1	0.7500	0.6111	0.5208	0.4567	0.4083	0.3704	0.3397	0.3143	0.2929
2	0.2500	0.2778	0.2708	0.2567	0.2417	0.2276	0.2147	0.2032	0.1929
3		0.1111	0.1458	0.1567	0.1583	0.1561	0.1522	0.1477	0.1429
4			0.0625	0.0900	0.1028	0.1085	0.1106	0.1106	0.1096
5				0.0400	0.0611	0.0728	0.0793	0.0828	0.0846
6					0.0278	0.0442	0.0543	0.0606	0.0646
7						0.0204	0.0334	0.0421	0.0479
8							0.0156	0.0262	0.0336
9								0.0123	0.0211
10									0.0100

**Table 3.2:** (RS) weights (Roberts and Goodwin, 2002)

Criteria									
Rank	2	3	4	5	6	7	8	9	10
1	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222	0.2000	0.1818
2	0.3333	0.3333	0.3000	0.2667	0.2381	0.2143	0.1944	0.1778	0.1636
3		0.1667	0.2000	0.2000	0.1905	0.1786	0.1667	0.1556	0.1455
4			0.1000	0.1333	0.1429	0.1429	0.1389	0.1333	0.1273
5				0.0667	0.0952	0.1071	0.1111	0.1111	0.1091
6					0.0476	0.0714	0.0833	0.0889	0.0909
7						0.0357	0.0556	0.0667	0.0727
8							0.0278	0.0444	0.0545
9								0.0222	0.0364
10									0.0182

**Table 3.3:** (RR) weights (Roberts and Goodwin, 2002)

Rank	Criteria								
	2	3	4	5	6	7	8	9	10
1	0.6667	0.5455	0.4800	0.4379	0.4082	0.3857	0.3679	0.3535	0.3414
2	0.3333	0.2727	0.2400	0.2190	0.2041	0.1928	0.1840	0.1767	0.1707
3		0.1818	0.1600	0.1460	0.1361	0.1286	0.1226	0.1178	0.1138
4			0.1200	0.1095	0.1020	0.0964	0.0920	0.0884	0.0854
5				0.0876	0.0816	0.0771	0.0736	0.0707	0.0682
6					0.0680	0.0643	0.0613	0.0589	0.0569
7						0.0551	0.0525	0.0505	0.0488
8							0.0460	0.0442	0.0427
9								0.0393	0.0379
10									0.0341

*Rank order distribution* (ROD) is a weight approximation method that assumes that valid weights can be elicited through direct rating. In the direct rating method the most important criterion is assigned a weight of 100 and the importance of the other criteria is then assessed relative to this benchmark. The ‘raw’ weights, ( $w_i^*$ ) obtained are then normalized to sum to 1. Assuming that all criteria have some importance, this means that the ranges of the possible ‘raw’ weights will be:

$$w_1^* = 100, \quad 0 < w_2^* \leq 100, \quad 0 < w_3^* \leq w_2^*$$

And in general:

$$0 < w_i^* \leq w_{i-1}^* \text{ (where } i \neq 1)$$

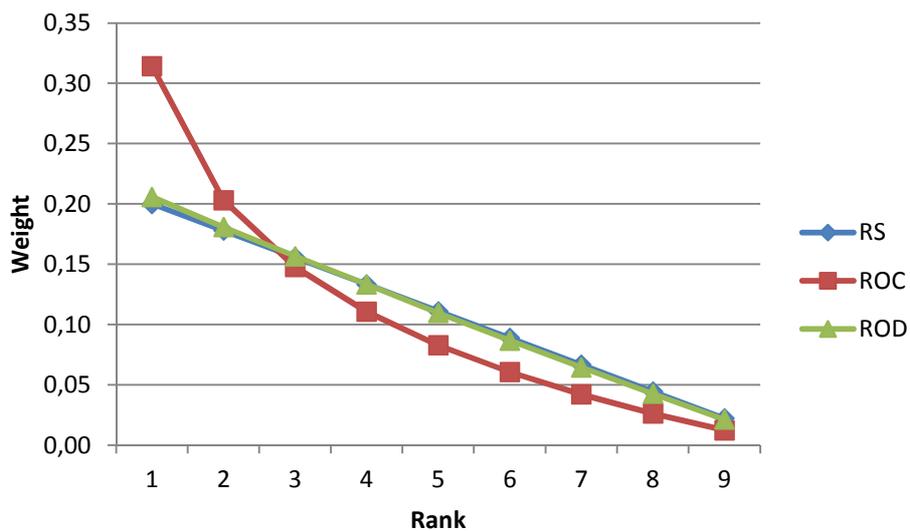
These ranges can be approximated by representing all of the inequalities by less-than-or-equal-to expressions. The uncertainty about the ‘true’ weights can then be represented by assuming uniform distribution for them. To determine ROD weights for general problems it is needed to consider the probability distributions for the normalised weights that follow from the assumptions about the distributions of the raw weights. For  $n > 2$  the density functions are a series of piecewise equations.

The means of each rank order distribution (ROD) for  $n = 2$  to 10 have been found mathematically and are displayed in Table 3.4. For further information about the calculations behind see Roberts and Goodwin (2002).

**Table 3.4:** ROD weights (Roberts and Goodwin, 2002)

Attributes									
Rank	2	3	4	5	6	7	8	9	10
1	0.6932	0.5232	0.4180	0.3471	0.2966	0.2590	0.2292	0.2058	0.1867
2	0.3068	0.3240	0.2986	0.2686	0.2410	0.2174	0.1977	0.1808	0.1667
3		0.1528	0.1912	0.1955	0.1884	0.1781	0.1672	0.1565	0.1466
4			0.0922	0.1269	0.1387	0.1406	0.1375	0.1332	0.1271
5				0.0619	0.0908	0.1038	0.1084	0.1095	0.1081
6					0.0445	0.0679	0.0805	0.0867	0.0893
7						0.0334	0.0531	0.0644	0.0709
8							0.0263	0.0425	0.0527
9								0.0211	0.0349
10									0.0173

A graphical comparison of the ROD, ROC and RS weights for 9 criteria can be seen in Figure 3.1 (Roberts and Goodwin, 2002).



**Figure 3.1:** Comparison of weights for 9 attributes (Roberts and Goodwin, 2002)

There is a very close match between the ROD and RS weights. This matching is found whatever the number of criteria. Indeed, in general, the ROD weights tend towards the RS weights as the number of criteria increases. Thus, given that ROD weights are difficult to calculate when the number of attributes is large, a practical solution is to use RS weights for large criteria problems. The ROC weights depart markedly from both the RS and ROD weights.

The figure also demonstrates another benefit of using ROD instead of ROC weights. ROC weights are 'extreme' in that the ratio of the highest to the lowest weights is so large that the lowest ranked criterion will only have a very marginal influence on the decision. In practice, criteria with a relative importance as low as this, would usually be eliminated from the decision model. The use of ROD weights goes some way

to reducing this extreme value problem. However, it can be argued that the inclusion of criteria with very low weights, e.g. 0.02, does not contribute in any way to the overall result and therefore should be omitted from the analysis. For a discussion of this see Barfod et al. (2011).

### **3.1.2 Pros and cons of SMART**

*Pros:* The structure of the SMART method is similar to that of the traditional CBA in that the total “value” is calculated as a weighted sum of the impact scores. In the CBA the unit prices act as weights and the “impacts scores” are the quantified (not normalized) CBA impacts. This close relationship to the well-accepted CBA method is appealing and makes the method easier to grasp for the decision maker.

*Cons:* In a screening phase where some poorly performing alternatives are rejected leaving a subset of alternatives to be considered in more detail the SMART method is not always the right choice. This is because, as noted by Hobbs and Meier (2000), SMART tends to oversimplify the problem if used as a screening method as the top few alternatives are often very similar. Rather different weight profiles should be used and alternatives that perform well under each different weight profile should be picked out for further analysis. This also helps identify the most “robust” alternatives. The SMART method has rather high demands on the level of detail in input data. Value functions need to be assessed for each of the lowest-level attributes, and weights should be given as trade-off

In SMART analysis the direct rating method of selecting raw weights is normally used as it is cognitively simpler and therefore is assumed to yield more consistent and accurate judgments from the decision-maker. These raw weights are then normalised and this normalisation process yields different theoretical distributions for the ranks. The means of these distributions are the ROD weights.

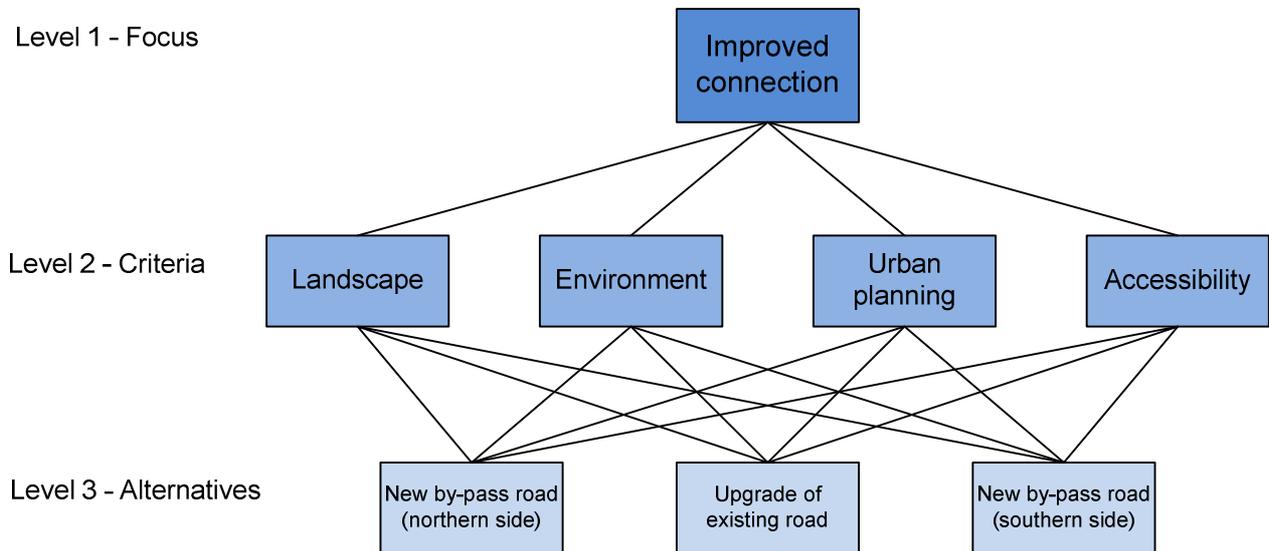
The formulae for the distribution of the ROD weights become progressively more complex as the number of criteria increase. Since the RS weights are so easy to calculate and closely match the ROD weights for higher numbers of criteria it is recommended to use RS weights when working with problems involving large numbers of criteria, and in cases where it can be assumed that the appropriate alternative method for eliciting the ‘true’ weights would have been the direct rating method.

## **3.2 The Analytic Hierarchy Process (AHP)**

The Analytic Hierarchy Process (AHP), developed by Saaty (1977), is essentially the formalisation of our intuitive understanding of a complex problem using a hierarchical structure as stated in (Hwang and Yoon, 1995). The AHP offers an alternative approach to SMART when a decision maker is faced with a problem involving multiple objectives. The method widely applied to decision problems in areas such as economics and planning, and because the AHP involves a relative complex mathematical procedure, user friendly computer software, such as Expert Choice, has been developed to support the method.

The crux of the AHP is to enable a decision maker to structure a Multi-Attribute Decision Making (MADM) problem visually in form of an attribute hierarchy. An attribute hierarchy has at least three levels: the focus or the overall goal of the problem on the top level, multiple criteria that define alternatives in the middle level, and competing alternatives in the bottom level. When criteria are highly abstract such as e.g. “well-being”, sub-criteria (or sub-sub-criteria) are generated subsequently through a multilevel hierarchy.

For example consider the problem of choosing which alternative to construct in a case, where an old road connection through a city has run out of capacity and a new improved connection is needed. The decision makers can choose between: A new by-pass road on the northern side of the city ( $A_1$ ), an upgrade of the existing connection ( $A_2$ ), or a new by-pass road on the southern side of the city ( $A_3$ ). Figure 3.2 shows the generated decision criteria by means of a hierarchical structure.



**Figure 3.2:** A hierarchy for choice of an improved connection

At level 1 the focus is overall an improved road connection. Level 2 comprises the criteria that contribute to the decision making: Landscape (L), environment (E), urban planning (UP) and accessibility (AC). Level 3 consist of the three solution possibilities:  $A_1$ ,  $A_2$  and  $A_3$ . It is obvious that each criterion in level 2 should contribute differently to the focus. The decision can be made on the relative importance among four criteria by pair-wise comparisons, due to the fact that pair-wise comparisons are much easier to make than a comparison of four criteria simultaneously.

In order to help the decision maker to assess the pair-wise comparisons, Saaty created a nine point intensity scale of importance between two elements (Saaty, 2001). The suggested numbers to express degree of preference between the two elements A and B are shown in Table 3.5.

To decide the relative weightings between  $n$  alternatives, it is in principle only necessary to perform  $n-1$  assessments. By performing a complete set of full pair-wise comparisons more information than necessary is collected, but a more varied evaluation is obtained, and if one or more answers are inaccurate the other answers will compensate the inaccuracy. The number of judgments,  $J$ , that have to be made in a full pair-wise comparison can be determined by (Belton and Stewart, 2002):

$$J = \frac{n \cdot (n-1)}{2} \quad (3.4)$$

**Table 3.5:** The fundamental scale for pair-wise comparisons (Saaty, 2001)

Intensity of importance	Definition	Explanation
1	Same	Neither of the two alternatives is preferable over the other
3	Weak	One alternative is preferred slightly over the other
5	Clear	One alternative is preferred clearly over the other
7	Strong	One alternative is preferred strongly over the other
9	Very Strong	One alternative is preferred very strongly over the other
2, 4, 6, 8	Compromise	Can be used for graduation between evaluation
Reciprocals of above	If activity <i>i</i> has one of the above nonzero numbers assigned to it when compared with activity <i>j</i> , then <i>j</i> has the reciprocal value when compared with <i>i</i>	A comparison mandated by choosing the smaller element as the unit to estimate the larger one as a multiple of that unit

In the road problem, there are four criteria in level 2. The decision maker then makes six pair-wise judgments among four criteria with respect to level 1 ( $4(4-1)/2=6$ ):

$$\begin{array}{lll}
 (L : E) = (7 : 1) & (L : UP) = (1 : 1) & (L : AC) = (7 : 1) \\
 (E : UP) = (1 : 3) & (E : AC) = (2 : 1) & (UP : AC) = (5 : 1)
 \end{array}$$

This information can be concisely contained in a so-called comparison matrix whose element at row *i* and column *j* is the ratio of row *i* and column *j* (Hwang and Yoon, 1995). The comparison matrix *A*, as introduced by Saaty, is seen below:

$$A = \begin{bmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \dots & \dots & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \dots & \dots & \dots & \frac{w_2}{w_n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & \dots & \dots & \frac{w_n}{w_n} \end{bmatrix} \tag{3.5}$$

Where  $w_1, w_2, \dots, w_n$  is the weights obtained by the comparisons. Applied on the case example that is:

$$\begin{matrix} L \\ E \\ UP \\ AC \end{matrix} \begin{bmatrix} 1 & L/E & L/UP & L/AC \\ E/L & 1 & E/UP & E/AC \\ UP/L & UP/L & 1 & UP/AC \\ AC/L & AC/L & AC/UP & 1 \end{bmatrix} = \begin{matrix} L \\ E \\ UP \\ AC \end{matrix} \begin{bmatrix} 1 & 7 & 1 & 7 \\ 1/7 & 1 & 1/3 & 2 \\ 1 & 3 & 1 & 5 \\ 1/7 & 1/2 & 1/5 & 1 \end{bmatrix}$$

The next step for the decision maker is to make pair-wise comparisons of the three alternatives in level 3 with respect to four criteria in level 2:

For L:

$$\begin{matrix} A_1 \\ A_2 \\ A_3 \end{matrix} \begin{bmatrix} A_1 & A_2 & A_3 \\ 1 & 1/3 & 2 \\ 3 & 1 & 5 \\ 1/2 & 1/5 & 1 \end{bmatrix}$$

For E:

$$\begin{matrix} A_1 \\ A_2 \\ A_3 \end{matrix} \begin{bmatrix} A_1 & A_2 & A_3 \\ 1 & 3 & 1/5 \\ 1/3 & 1 & 1/7 \\ 5 & 7 & 1 \end{bmatrix}$$

For UP:

$$\begin{matrix} A_1 \\ A_2 \\ A_3 \end{matrix} \begin{bmatrix} A_1 & A_2 & A_3 \\ 1 & 1/5 & 2 \\ 5 & 1 & 7 \\ 1/2 & 1/7 & 1 \end{bmatrix}$$

For AC:

$$\begin{matrix} A_1 \\ A_2 \\ A_3 \end{matrix} \begin{bmatrix} A_1 & A_2 & A_3 \\ 1 & 1/3 & 1/5 \\ 3 & 1 & 1/3 \\ 5 & 3 & 1 \end{bmatrix}$$

After the construction of the pair-wise comparison matrix, the next step is to retrieve the weights of each element in the matrix. There are several methods for retrieving these weights: the originally introduced eigenvector method (Hwang and Yoon, 1981), and the later introduced geometric mean method (Saaty, 2001).

The geometric mean method with calculations regarding the case example is introduced below as this method is the most suitable for calculations in hand. The eigenvector method is described next, but only for a small numerical example as this method uses more demanding calculations that normally will be carried through in a software program such as Expert Choice.

### 3.2.1 The eigenvector method

The first step in the eigenvector method is to reduce the pair-wise comparison matrix to a comparison vector, i.e. a set of scores (or partial values) representing the relative performance of each alternative. The values in the pair-wise comparison matrix are interpreted as ratios of these underlying scores.

Saaty originally introduced a method of scaling ratios using the principle eigenvector of a positive pair-wise comparison matrix. The method presumes that matrix  $A$  is (Hwang and Yoon, 1981):

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ a_{n1} & a_{n2} & \cdot & \cdot & \cdot & a_{nn} \end{bmatrix} = \begin{bmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdot & \cdot & \cdot & \frac{w_1}{w_n} \\ \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdot & \cdot & \cdot & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \cdot & \cdot & \cdot & \frac{w_2}{w_n} \\ \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdot & \cdot & \cdot & \frac{w_1}{w_n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdot & \cdot & \cdot & \frac{w_n}{w_n} \\ \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdot & \cdot & \cdot & \frac{w_1}{w_n} \end{bmatrix} \quad (3.6)$$

This is a reciprocal matrix (as before), which has all positive elements and has the reciprocal property:

$$a_{ij} = \frac{1}{a_{ji}} \quad (3.7)$$

and

$$a_{ij} = \frac{a_{ik}}{a_{jk}} \quad (3.8)$$

Multiplying A by  $\underline{w} = (w_1, w_2, \dots, w_n)^T$  yields

$$A \cdot \underline{w} = \begin{bmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdot & \cdot & \cdot & \frac{w_1}{w_n} \\ \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdot & \cdot & \cdot & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \cdot & \cdot & \cdot & \frac{w_2}{w_n} \\ \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdot & \cdot & \cdot & \frac{w_1}{w_n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdot & \cdot & \cdot & \frac{w_n}{w_n} \\ \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdot & \cdot & \cdot & \frac{w_1}{w_n} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{bmatrix} = n \cdot \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{bmatrix} = n \cdot \underline{w} \quad (3.9)$$

or

$$(A - nI) \cdot \underline{w} = 0 \quad (3.10)$$

Due to the consistency of (4.10), the system of homogeneous linear equations (4.11) has only trivial solutions. In general the precise values of  $w_i/w_j$  are unknown and must be estimated, so in other words, human judgments cannot be so accurate that (4.10) can be satisfied completely. In any matrix small perturbations in the coefficients imply small perturbations in the eigenvalues (Hwang and Yoon, 1981). If we define  $A^-$  as the decision makers estimate of A and  $\underline{w}^-$  is corresponding to  $A^-$ , then

$$A^- \underline{w}^- = \lambda_{\max}^- \underline{w}^- \quad (3.11)$$

Where  $\lambda_{\max}^-$  is the largest eigenvalue of  $A^-$ .  $\underline{w}^-$  can be obtained by solving the system of linear equations. In order to show the steps in the computation of weights, a numerical example is reviewed.

The following example is taken from (Hwang and Yoon, 1981). The positive pair-wise comparison matrix is given:

$$A = \begin{bmatrix} 1 & 1/3 & 1/2 \\ 3 & 1 & 3 \\ 2 & 1/3 & 1 \end{bmatrix}$$

The determinant of  $(A - \lambda \cdot I)$  is then set to zero:

$$\det(A - \lambda \cdot I) = \begin{bmatrix} 1-\lambda & 1/3 & 1/2 \\ 3 & 1-\lambda & 3 \\ 2 & 1/3 & 1-\lambda \end{bmatrix} = 0$$

The largest eigenvalue of A,  $\lambda_{\max}$ , is 3.0536, and we have:

$$\begin{bmatrix} -2.0536 & 1/3 & 1/2 \\ 3 & -2.0536 & 3 \\ 2 & 1/3 & -2.0536 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = 0$$

The solution of the homogeneous system of linear equations, where it is assumed that  $\sum_{i=1}^3 w_i = 1$ , gives:

$$\underline{w}^T = \begin{bmatrix} 0.1571 \\ 0.5936 \\ 0.2493 \end{bmatrix}$$

### 3.2.2 The geometric mean method

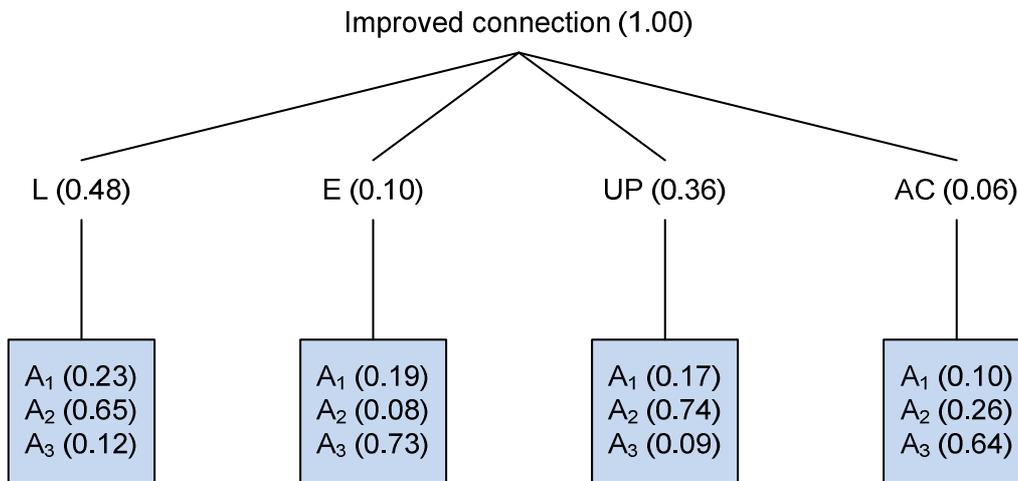
For an approximation method that provides sufficiently close results in most situations, (Saaty, 2001) suggest the geometric mean of a row: Multiply the  $n$  elements in each row, take the  $n$ th root, and prepare a new column for the resulting numbers, then normalise the new column (i.e., divide each number by the sum of the numbers). The weights for the four criteria in the case example are shown below.

$$\begin{array}{l} L \\ E \\ UP \\ AC \\ \text{sum} \end{array} \begin{bmatrix} (1 \cdot 7 \cdot 1 \cdot 7)^{1/4} = 2.65 \\ (1/7 \cdot 1 \cdot 1/3 \cdot 2)^{1/4} = 0.56 \\ (1 \cdot 3 \cdot 1 \cdot 5)^{1/4} = 1.97 \\ (1/7 \cdot 1/2 \cdot 1/5 \cdot 1)^{1/4} = 0.35 \\ 5.53 \end{bmatrix} = \begin{bmatrix} 0.48 \\ 0.10 \\ 0.36 \\ 0.06 \\ 1.00 \end{bmatrix}$$

Similarly, the relative contributions (i.e., weights) among three alternatives towards the four criteria are computed below.

$$\begin{array}{c}
 L \quad E \quad UP \quad AC \\
 A_1 \begin{bmatrix} 0.23 & 0.19 & 0.17 & 0.10 \end{bmatrix} \\
 A_2 \begin{bmatrix} 0.65 & 0.08 & 0.74 & 0.26 \end{bmatrix} \\
 A_3 \begin{bmatrix} 0.12 & 0.73 & 0.09 & 0.64 \end{bmatrix}
 \end{array}$$

The final stage of the AHP is to compute the contribution of each alternative to the overall goal (i.e., improved connection) by aggregating the resulting weights vertically. The overall priority for each alternative is obtained by summing the product of the criteria weight and the contribution of the alternative, with respect to that criterion. Refer to Figure 3.3 for the road choice problem.



**Figure 3.3:** Priorities for each hierarchical level

The computation of the overall priority for alternative A1 is as follows

$$0.48 \cdot (0.23) + 0.10 \cdot (0.19) + 0.36 \cdot (0.17) + 0.06 \cdot (0.10) = 0.1966$$

Similarly, they are 0.6020 and 0.2014 for  $A_2$  and  $A_3$  respectively. Therefore the decision maker's choice would be to make an upgrade of the existing road ( $A_2$ ), if the decision was only to be based on these criteria.

### 3.2.3 Consistency

The AHP allows inconsistency, but provides a measure of the inconsistency in each set of judgments. This measure is an important by-product of the process of deriving priorities based on pair-wise comparisons. It is natural for people to want to be consistent, as being consistent is often thought of as a prerequisite to clear thinking. However the real world is hardly ever perfectly consistent and we can learn new things only by allowing for some inconsistency with what we already know.

Some causes for inconsistency are listed below:

- *Lack of information.* If the decision maker has little or no information about the factors being compared, then the judgments will appear to be random and a high consistency ratio will result. It

is useful to find out that a lack of information exists, although sometimes decision maker might be willing to proceed without immediately spending time and money gathering additional information in order to ascertain if the additional information is likely to have a significant impact on the decision.

- *Lack of concentration.* Lack of concentration during the judgment process can happen if the decision makers become fatigued or are not really interested in the decision. It is the decision analyst job to prevent this from happening.
- *Real world is not always consistent.* The real world is rarely perfectly consistent and is sometimes fairly inconsistent. Football is a good example: It is not uncommon for team A to defeat team B, after which team B defeats team C, after which team C defeats team A. Inconsistencies such as this may be explained as being due to random fluctuations, or to underlying causes (such as match-ups of personnel), or to a combination. Regardless of the reasons, real world inconsistencies do exist and thus will appear in our judgments.
- *Inadequate model structure.* A final cause of inconsistency is “inadequate” model structure. Ideally one would structure a complex decision in a hierarchical fashion such that factors at any level are comparable, within an order of magnitude or so, of other factors at that level. Practical considerations might preclude such a structuring and it is still possible to get meaningful results. Suppose for example, several items that differed by as much as two orders of magnitude were compared. One might erroneously conclude that the AHP scale is incapable of capturing the differences since the scale ranges from 1 to 9. However, because the resulting priorities are based on second, third and higher order dominances, AHP can produce priorities far beyond any order of magnitude. A higher than usual inconsistency ratio will result because of the extreme judgments necessary.

It is important that a low inconsistency ratio does not become the goal of the decision making process. A low inconsistency ratio is necessary but not sufficient for a good decision. It is possible to be perfectly consistent but consistently wrong. It is more important to be accurate than consistent.

### 3.2.4 Consistency ratio

The consistency ratio is computed from the eigenvalue,  $\lambda_{\max}$ , which will often turn out to be larger than the value describing a fully consistent matrix. In order to provide a measure of severity of this deviation, Saaty defined a measure of consistency, or consistency index (CI) by (Belton and Stewart, 2002):

$$CI = \frac{\text{Principal eigenvalue} - \text{size of matrix}}{\text{size of matrix} - 1} = \frac{\lambda_{\max} - n}{n - 1} \quad (3.12)$$

The consistency index is compared to a value derived by generating random reciprocal matrices of the same size, to give a consistency ratio (CR) which is meant to have the same interpretation no matter the size of the matrix. The comparative values from random matrices are as follows in Table 3.6 for  $3 \leq n \leq 10$  (Belton and Stewart, 2002, p. 156):

**Table 3.6:** Comparative values

Size of matrix	3	4	5	6	7	8	9	10
Comparative value	0.52	0.89	1.11	1.25	1.35	1.40	1.45	1.49

In the small example concerning the eigenvalue method the largest eigenvalue was 3.0536, the consistency index is thus  $(3.0536-3)/2=0.0268$ . Since  $n=3$ , the consistency ratio is  $0.0268/0.52=0.05$ . A consistency ratio of 0.1 or less is generally stated to be acceptable.

As seen the AHP method is fundamentally different to other assessment methods in many respects. This section is considering strengths of the AHP, but also looks on some of the criticisms which have been made of the technique.

### 3.2.5 Relative strengths of AHP

**Formal structuring of the problem.** Like SMART and other decision analysis techniques the AHP provides a formal structure to problems. This allows complex problems to be decomposed into sets of simpler judgments and provides a documented rationale for the choice of a particular option.

**Simplicity of Pair-wise comparisons.** The use of pair-wise comparisons means that the decision maker can focus, in turn, on each small part of the problem. Only two attributes or options have to be considered at any one time so that the decision maker's judgmental task is simplified. Verbal comparisons are also likely to be preferred by decision makers who have difficulty in expressing their judgments numerically.

**Redundancy allows consistency to be checked.** The AHP requires more judgments to be made by the decision maker than is needed to establish a set of weights. For example, if a decision maker indicates that attribute A is twice as important as B, and B, in turn, is three times as important as C, then it can be inferred that A is six times more important than C. However, by also asking the decision maker to compare A with C it is possible to check the consistency of the judgments. It is considered to be good practice in decision analysis to obtain an input for a decision model by asking for it in several ways and then asking the decision maker to reflect on any inconsistencies in the judgments put forward. In the AHP this is carried out automatically (Goodwin and Wright, 2009).

**Versatility.** The wide range of applications of the AHP is evidence of its versatility. In addition to judgments about importance and preference, the AHP also allows judgments about the relative likelihood of events to be made. This has allowed it to be applied to problems involving uncertainty and also to be used in forecasting. AHP models have also been used to construct scenarios by taking into account the likely behaviour and relative importance of key actors and their interaction with political, technological, environmental, economic and social factors (Goodwin and Wright, 2009).

### 3.2.6 Criticism of AHP

**Conversion from verbal to numeric scale.** Decision makers using the verbal method of comparison will have their judgments automatically converted to the numerical scale, but the correspondence between the two

scales is based on untested assumptions. If you indicate that A is weakly more important than B the AHP will assume that you consider A to be three times more important, but this may not be the case.

**Inconsistencies imposed by 1 to 9 scale.** In some problems the restriction of pair-wise comparisons to a 1 to 9 scale is bound to force inconsistencies on the decision maker. For example if A is considered to be 5 times more important than B, and B is 5 times more important than C, then to be consistent A should be judged to be 25 times more important than C, however this is not possible.

**Meaningfulness of responses to questions.** Unlike SMART, weights are elicited in the AHP without reference to the scales on which attributes are measured. For example, a person using SMART to choose a house might be asked to compare the value of reducing the daily journey to work from 100 kilometers to 10 kilometers with the value of increasing the number of bedrooms in the house from two to four. Implicit in this type of comparison is the notion of a trade-off or exchange: 90 fewer kilometers may only be half as valuable as two extra bedrooms. AHP questions, which simply ask for the relative importance of attributes without reference to their scales, are therefore less well defined, if they are meaningful at all. This fuzziness may mean that the questions are interpreted in different, and possibly erroneous, ways by decision makers.

**New alternatives can reverse the rank of existing alternatives.** This issue, which is related to the last point, has attracted much attention. Suppose that you are using the AHP to choose a location for a new company and the weights you obtained from the method give the following order of preference: 1. London, 2. Paris and 3. Rome. However before making the decision you discover that a site in Berlin is also worth considering, so you repeat the AHP to include this new option. Even though you leave the relative importance of the attributes unchanged, the new analysis gives the following rankings: 1. Paris, 2. London, 3. Berlin and 4. Rome, so the rank of Paris and London have been reversed, which does not seem to be intuitive reasonable. (Goodwin and Wright, 2009) claims that this arises from the way in which the AHP normalizes the weights to sum to 1, and that this is consistent with a definition of weights which is at variance with that used in SMART. Most decision makers would consider SMART to be the reasonable one.

**Number of comparisons required may be large.** While the redundancy built into the AHP is an advantage, it may also require a large number of judgments from the decision maker. Consider, for example, an office location problem which involves 8 alternatives and 8 attributes, this would involve 224 pair-wise comparisons of importance or preference. This requirement to answer a large number of questions can reduce the attraction of the AHP in the eyes of potential users, even though the questions themselves are considered to be easy.

**The axioms of the method.** In (Goodwin and Wright, 2009) it is argued, that the SMART method is well founded on a set of axioms, that is, a set of rules which are intended to provide the basis for a rational decision making. The clarity and intuitive meaning of these axioms allows their appeal, as rules for rational behaviour to be debated and empirically tested. In contrast to this (Goodwin and Wright, 2009) argues that the axioms of the AHP are not founded on testable descriptions of rational behaviour.

### 3.2.7 Consistency checks

It is good practice to carry out more than the minimum number of comparisons necessary to specify the set of criteria weights, thus building in a check of consistency of the decision makers' judgments. As with

scores, the assessment of weights is also implicitly a process of pair wise comparisons. This may be formalised by specifying a reference criterion against which all others are compared (requiring the minimum number of comparisons), or each criterion may be compared with every other one giving full specification (requiring  $n(n-1)/2$  comparisons) as in the AHP approach. Alternatively something between these two extremes may be sought by judicious choice of criteria to be compared.

### 3.2.8 Working with weak information

The process of determining values for criteria weights calls for a lot of hard thinking on the part of the decision maker. Questions such as those described above are difficult to answer. Depending on the circumstances of the decision, an alternative way of proceeding might be to use the rank order of to give simple initial estimates of criteria weights and then to use this as starting point for extensive sensitivity analysis. This may show that the preferred alternative is insensitive to changes in weights which preserve rank order – in which case it would not be necessary to specify more precise values. Or it may indicate that attention should be focussed on the weight assigned to a specific criterion.

Some writers (e.g. Edwards and Barron (1994) discussing their SMARTER approach) suggest that, in the presence of only ordinal information, an initial analysis can be carried out using weights which are in some sense most central in the region defined by  $w_1 > w_2 > w_3 > \dots > w_n > 0$ . One possibility is to estimate weights by the centroid, i.e. the arithmetical average of the extreme points of the region. When normalising the weights to sum to 1, it is easily confirmed that the  $n$  extreme points are:  $(1, 0, 0, \dots, 0)$ ,  $(1/2, 1/2, 0, \dots, 0)$ ,  $(1/3, 1/3, 1/3, 0, \dots, 0)$ , ...,  $(1/n, 1/n, \dots, 1/n)$ . However, the use of the centroid to generate a set of weights is not entirely satisfactory, as it leads to rather extreme values. For example, with  $n = 3$  criteria, the centroid weights are 0.611, 0.278 and 0.111 respectively, so that the ratio of  $w_1$  to  $w_3$  is 5.5, so that the third criteria will only have a very marginal influence on the outcome, contrary to what appears to be meant by the inclusion of all 3 criteria in the analysis. The situation becomes more extreme as  $n$  increases; with  $n = 5$ , the centroid weights are 0.457, 0.257, 0.157, 0.090 and 0.039. The ratio from  $w_1$  to  $w_5$  is now over 11:1, so that the fifth criterion has an almost vanishingly small influence. Experience seems to suggest, that at weight ratio of 5 or more indicates an almost absolute dominance of one criterion over another, so that larger ratios should not really be expected amongst criteria which have been retained in the value tree.

Two possibilities for ameliorating the extreme effects of using the centroid may be (Belton and Stewart, 2002):

- Inclusion of a further constraint to the effect that no criterion  $i$  will have been included in the model if the ratio of  $w_1$  (the largest weight) to  $w_i$  exceeds some factor  $R$ : For example with  $R = 9$ , the extreme points of the feasible weight region for  $n = 3$  will be:  $(9/11, 1/11, 1/11)$ ,  $(9/19, 9/19, 1/19)$  and  $(1/3, 1/3, 1/3)$ . The centroid weights based on these extremes are 0.541, 0.299 and 0.159.
- Assume a geometrically decreasing set of weights, with each  $w_i$  being a constant proportion  $r$  of the next most important weight  $w_{i-1}$ : The centroid weights do in fact decrease in an approximately geometric fashion, but with a high rate of decrease (with a value of  $r$  around 0.4 – 0.45 for  $n = 3$ , and around 0.5 – 0.55 for  $n = 5$ ). Some anecdotal experience suggests a rather less dramatic rate of

reduction in weights, certainly well above 0.5. For example, even with  $r$  increased up to 0.6, the estimated weights for  $n = 3$  are much more moderate, namely: 0.510, 0.306 and 0.184.

### 3.2.9 Overall evaluation

The overall evaluation of an alternative is determined by first multiplying its value score on each bottom-level criterion by the cumulative weight of that criterion and then adding the resultant values. If the values relating to individual criteria have been assessed on a 0 to 100 scale and the weights are normalised to sum to 1 then the overall values will lie on a 0 to 100 scale. If the criteria are structured as a value tree then it is also informative to determining scores at intermediate levels of the tree. In the example from earlier this allows the alternatives to be compared, for example, on accessibility.

However, the determination of an overall value should by no means be viewed as the end of the analysis, but simply another step in furthering understanding and promoting discussion about the problem. Although the underlying model is simple and static that should not be a limitation in its use. It provides a powerful vehicle for reflecting back to decision makers the information they have provided, the judgments they have made, and an initial attempt at synthesising these. The extent to which the model will be a successful catalyst for discussion of the problem and for learning about one's own and other's values, depends on the effectiveness with which feedback can be provided. Simple, static visual displays are an effective means of reflecting back information provided and well-designed visual interactive interfaces provide a powerful vehicle for exploring the implications of uncertainty about values.

In exploring the model decision makers should test the overall evaluation and partial aggregations of information against their intuitive judgment. Are the results in keeping with intuition? If not, why not? Could some values have been wrongly assessed? Is there an aspect of performance which is not captured in the model? Is the additive model inappropriate? Or does the model cause the decision makers to revise their intuitive judgments? The aim of the analysis should be to arrive at convergence between the results of the model and the decision makers' intuition.

Decision makers should look not only at the overall evaluation of alternatives, but at their profiles. How is an alternative's overall value made up? Is it a good "all-rounder" or does it have certain strengths and weaknesses? Alternatives with similar overall scores can have very different profiles. Is there any dominating, or dominated alternatives? One option dominates another if it does at least as well on all criteria relevant to the decision. In simple terms, if there is an option which dominates all others it should be preferred, or if an option is dominated by another it should not be a candidate for choice. However, rather than acting as rigid guidelines, these concepts should be used as catalysts for further thought and learning about the problem situation.

## 3.3 REMBRANDT

The presented AHP method has – as mentioned – been criticised because of weaknesses in the theoretical basis (Belton and Stewart, 2002). An improved theoretical model is illustrated by the REMBRANDT method (Ratio Estimation in Magnitudes or deci-Bells to Rate Alternatives which are Non-Dominated) by Lootsma (1992). The technique is briefly described in the following.

The REMBRANDT technique is intended to adjust for three contended flaws in AHP. First, direct rating is on a geometric scale. Second, the weights are calculated by the geometric mean method. Third, aggregation of scores by arithmetic mean is replaced by the product of alternative relative scores weighted by the power of weights obtained from analysis of hierarchical elements above the alternatives.

Both methods consist of so-called preference scales, illustrated opposite each other in Table 3.7. In connection with the specific pair-wise comparison, the user should only concentrate on an "explanation scale", ie. the verbal scale, while the two numerical scales only are of technical interest for later input into the mathematical model.

**Table 3.7:** AHP and REMBRANDT scales

Preference	Explanation	AHP	REMBRANDT
Same	Neither of the two alternatives is preferable over the other	1	0
Weak	One alternative is preferred slightly over the other	3	2
Clear	One alternative is preferred clearly over the other	5	4
Strong	One alternative is preferred strongly over the other	7	6
Very Strong	One alternative is preferred very strongly over the other	9	8
Compromise	Can be used for graduation between evaluation	2, 4, 6, 8	1, 3, 5, 7

There is as mentioned mainly three criticisms of the AHP method as the REMBRANDT method tries to correct. The first is related to the scale in AHP, where 1 represents two similar values, 3 means that the first object is slightly better than the second object, 5 indicates the clear benefits, 7 a strong advantage and 9 a very strong advantage. Lootsma (1999) has based on a number of examples and reflections adjusted the numerical scale of REMBRANDT, so it was more convenient for subsequent calculations.

The second point that REMBRANDT tries to improve, is the calculation of scores. AHP uses a method which has the disadvantage that if one later in the process adds a new alternative, it may reverse the existing ranking of alternatives (known as 'rank reversal of alternatives'). REMBRANDT uses logarithmic regression and the method does not have this error incorporated. The scale, which REMBRANDT uses, goes from 0 to 8. For a more detailed technical analysis refers to Olson et al. (1995).

The third and last point which the REMBRANDT method tries to improve compared with AHP is the way the individual scores are aggregated. The AHP uses a method based on eigenvectors calculation and summation of scores multiplied by the criteria weights, while REMBRANDT calculates the value of an alternative by using a geometric mean calculation (multiplying scores uplifted with criteria weights).

To illustrate the principle of REMBRANDT a small calculation example is shown (Olson et al., 1995). There is a decision problem involving three alternatives (A, B and C) and four criteria (W, X, Y and Z). The criteria weights are already set to:

$$(0,493 ; 0,246 ; 0,174 ; 0,087)$$

Scores for each alternative under each criterion is calculated using the following transformation:  $e^{\ln(2) \delta(jk)}$ . It is noted that when REMBRANDT is used to determine criteria weights the transformation  $e^{\ln(v2) \delta(jk)}$  is used (Lootsma, 1999). After determining the scores for each criterion all four criteria are aggregated

Pair-wise comparison:

Transformation:

Geo. mean:

**W:**

	A	B	C
A	0	4	6
B	-4	0	4
C	-6	-4	0

	A	B	C
A	1	16	64
B	0.0625	1	16
C	0.015625	0.0625	1

10.08
1
0.0992

**X:**

	A	B	C
A	0	-2	1
B	2	0	4
C	-1	-4	0

	A	B	C
A	1	0.25	2
B	4	1	16
C	0,5	0.0625	1

0.7937
4.0
0.3150

**Y:**

	A	B	C
A	0	0	-4
B	0	0	-3
C	4	3	0

	A	B	C
A	1	1	0.0625
B	1	1	0.125
C	16	8	1

0.3969
0.5
5.0397

**Z:**

	A	B	C
A	0	1	-1
B	-1	0	-2
C	1	2	0

	A	B	C
A	1	2	0.5
B	0.5	1	0.25
C	2	4	1

1
0.5
2

The values of the four criteria using multiplicative aggregation become:

$$\begin{aligned}
 \text{A: } & 10.08^{0.493} * 0.7937^{0.246} * 0.3969^{0.174} * 1^{0.087} = 2.513 \sim 0.624 \\
 \text{B: } & 1^{0.493} * 4^{0.246} * 0.5^{0.174} * 0.5^{0.087} = 1.174 \sim 0.292 \\
 \text{C: } & 0.0992^{0.493} * 0.315^{0.246} * 5.0397^{0.174} * 2^{0.087} = 0.339 \sim 0.084
 \end{aligned}$$

When all criteria are considered for all alternatives, the final values for the value function can be determined as described in the main text.

## 4 Composite model for assessment

In the appraisal and planning of transport infrastructure projects the examination should be based on all relevant impacts, which are depending on the type and size of the project addressed. Some of these impacts can be assessed monetarily and are thereby possible to include in a conventional CBA. However, no valid assessment knowledge exists for impacts such as urban development, landscape, etc. These impacts are denominated as non-monetary impacts or strategic impacts and have to be assessed by use of a MCDA.

Earlier research on composite decision support systems (DSS) within transport planning has mainly concentrated on incorporating the CBA in the MCDA. Here the European Commission's fourth framework project EUNET (EUNET/SASI, 2001), which has developed a methodology dealing with the combination of CBA and MCDA, can be mentioned. The EUNET framework applies scores to the investment criterion, e.g. the benefit cost rates (BCR), thus, it treats the rates as any other criterion in the MCDA. Exactly which criteria to include in the framework is a matter of judgment depending, among other factors, on the reliability of the data and the preferences stated by the decision-makers and/or stakeholders in the decision process. Another similar inclusive approach is proposed in (Sayers et al., 2003) for transport project appraisal in the UK. Different methodological frameworks are used varying from country to country; however, it is roughly possible to divide them into two main categories: CBA-based and MCDA-based frameworks. Among the CBA-based frameworks the Danish and German can be mentioned, while e.g. French and Dutch frameworks are based on the use of MCDA (for further information about European frameworks see e.g. (Banister and Berechman, 2000) and (EUNET/SASI 2001). Reviews of transport appraisal methodologies and their premises and results can be found in for example (Hayashi and Morisugi, 2000) and (Mackie and Preston, 1998) listing also various sources of error and bias in them.

The COSIMA DSS presented here provides a theoretical and practical methodology for adding non-monetary MCDA-criteria to the monetary CBA-impacts. Unlike previous attempts this DSS is based on the argument that the MCDA-criteria can be added to the CBA-impacts – if value functions can be computed for the MCDA-criteria using a weighting procedure describing the importance of each criterion. Hence, the COSIMA approach is based on the theoretical valid and widely used methodology of additive value functions (see e.g. (Keeney and Raiffa, 1993) or (von Winterfeldt and Edwards, 1986)).

### 4.1 General principles

The idea behind composite modelling assessment (COSIMA) is to extend conventional CBA into a more comprehensive type of analysis, as often demanded by decision-makers, by including "missing" decision criteria of relevance for the actual assessment task. The missing criteria often address issues that have been difficult to assess by the conventional CBA but hold a potential of improving actual decision support from the assessment if treated properly. In COSIMA the added criteria will be referred to as the MCDA part of the COSIMA analysis.

#### 4.1.1 Assumptions

There may be four alternatives between which the alternative to be implemented is to be found. To find the alternative to be implemented among the four it is necessary that the phases with formulation and

design are implemented in a way so that relevant considerations are reflected in the set of the four alternatives. This process has led to four alternatives and not just one, which reflect an uncertainty in the decision making which is expressed by necessitate a choice between four alternatives. An assumption is also that this choice made by one or more decision-makers can be qualified through the use of decision support.

The four alternatives are pre-tested using a cost-benefit analysis. With four results for example in terms of BCRs the election issue could be solved through the alternative with the highest BCR was chosen as it from a relative point emerges as the best of the four alternatives. At the same time it may be assumed that it also satisfies an absolute requirement with regard to socio-economic viability, namely the BCR is greater than 1 equivalent to a positive net present value.

COSIMA is based on that, at least hypothetically the alternative with highest BCR is chosen. If decision-makers feel that the CBA has provided the necessary foundation the decision process ends here. However, if decision-makers are unsure whether the necessary conditions have been present, this can be further developed through the MCDA as a supplement to the CBA. The following will be based on whether it is justified that the alternative with the highest BCR should be substituted with any of the other three alternatives. It is important to maintain that it is precisely this situation which is being treated, as this has an impact on both the process as on methodology.

## 4.2 The COSIMA approach

The COSIMA approach consists of a CBA part and a MCDA part and the result of the COSIMA assessment is expressed as a total value (TV) based on both parts. This model set-up emphasizes that the MCDA part should be truly additive to the CBA part. For this reason a project alternative,  $A_k$ , is better represented for the decision making by the  $TV(A_k)$  than by e.g. the net present value (NPV) derived from the CBA. The principle in COSIMA can be expressed by (Leleur et al., 2007):

$$TV(A_k) = CBA(A_k) + MCA(A_k) \quad (4.1)$$

The formulation of COSIMA introduced by (4.1) thus resembles CBA but the assessment principles made use of in the MCDA part, generally based on decision-maker involvement, justifies the notation as multi-criteria analysis. It can be noted on the basis of (5.1) that in a situation where the investment in  $A_k$  equal to the investment costs  $C_k$  is not feasible seen from CBA (i.e.  $CBA(A_k) \leq C_k$ ), then the investment can be justified by the wider COSIMA examination if  $TV(A_k) > C_k$ . This can also be expressed as  $TRR(A_k) > 1$  where TRR expresses the total rate of return.

In a COSIMA analysis where  $A_k$  denominates the project alternative it has been found to be convenient to express the feasibility by the total rate of return  $TRR(A_k)$  from the investment  $C_k$  which leads to (4.2) below (Ibid.):

$$TRR(A_k) = \frac{TV(A_k)}{C_k} = \frac{1}{C_k} \cdot \left( \sum_{i=1}^I V_{CBA}(X_{ik}) + \alpha \cdot \left[ \sum_{j=1}^J w(j) \cdot V_{MCA}(X_{jk}) \right] \right) \quad (4.2)$$

where:

- $C_k$  are the total costs of alternative  $A_k$
- $V_{CBA}(X_{ik})$ , is the value in monetary units for the CBA effect  $i$  for alternative  $k$  for altogether  $I$  CBA impacts
- $V_{MCDA}(X_{jk})$  is a value-function for MCDA criterion  $j$  for alternative  $k$  for altogether  $J$  MCDA criteria
- $w(j)$  a weight that expresses the importance of criterion  $J$
- $\alpha$  is a calibration factor that expresses the model set-up's trade-off between the CBA and the MCDA part

The general COSIMA principles are presented by (4.1) and (4.2). It can be realized that with sufficient information about the MCDA part, (4.2) can be specified into a CBA. This would for example be the situation if a conventional CBA is carried out and afterwards supplemented with some extra criteria which can be specified fully by impact models that lead to net effects which can be given satisfactory unit prices similar to the assessment in the CBA part. However, this will most often not be possible as the MCDA part in general is "less known" than the CBA part. The purpose of COSIMA is to handle such a situation in a comprehensive and transparent way ensured through the determination of appropriate values for  $\alpha$  and  $w(j)$  for the  $J$  MCDA criteria and appropriate value functions  $V_{MCDA}(X_{jk})$ . The latter supplement the determination of  $V_{CBA}$  that, however, can be derived from a CBA manual relevant for the actual assessment case e.g. DMT (2003).

#### 4.2.1 Calibration of the COSIMA DSS

Regarding the  $\alpha$ -indicator, that expresses the balance between the CBA and MCDA parts in the model set-up, it should be noted that the CBA calculation remains unchanged in COSIMA, but that different  $\alpha$ -values will change the MCDA's influence on the TRR. In practice it has been found convenient to express  $\alpha$  based on a MCDA%, which reflects the relative weight of the MCDA-part compared to the CBA-part. The value of  $\alpha = \alpha(\text{MCDA}\%)$  is then set by determining  $\text{MCDA}\% = 100 \cdot \sum_j B_j / [\sum_i B_i + \sum_j B_j]$ , where  $B_i = \sum_{\kappa \in K} (b_{i\kappa})$  and  $B_j = \sum_{\kappa \in K} (b_{j\kappa})$  represent the value elements for the individual CBA-impact  $i$  and MCDA-criterion  $j$  summed over the  $\kappa$  alternatives (the alternatives  $A_\kappa$  chosen for calibration of the model). Thus  $\sum_i B_i$  and  $\sum_j B_j$  are summations of the  $I$  CBA-impacts and the  $J$  MCDA-criteria, and  $B_i$  and  $B_j$  the results of the  $b_{i\kappa}$  and  $b_{j\kappa}$  summations of the alternatives, where some if not all are selected for the model calibration (Leleur, 2008).

The calculations in the COSIMA DSS use a parameter for the calibration named  $UP_j$  which functions as a shadow price per index value for each of the  $J$  MCDA-criteria in order to produce the  $b_{jk}$  values. These benefit values obtained are determined by  $b_{jk} = VF_j(Y_{jk}) \cdot UP_j$  where the shadow price,  $UP_j$ , is a function of the  $\alpha$ -indicator (MCDA%), the criteria weights ( $w_j$ ), the sum of benefits from the CBA for the alternatives used for calibration ( $\sum_i \sum_{\kappa \in K} (b_{i\kappa})$ ) and the sum of VF-scores for the alternatives used for calibration ( $\sum_{\kappa} VF_j(Y_{jk})$ ). In the procedure  $\alpha(\text{MCDA}\%)$  and  $w_j$  determine a fraction of  $\sum_i \sum_{\kappa \in K} (b_{i\kappa})$  that by unit scaling leads to the  $J$  unit prices that are used for calculating the  $TRR(A_k)$ . It should be noted that TRR-values are also calculated for alternatives not used in the calibration set. Changes in the set of alternatives  $A_\kappa$  behind the calibration will influence the  $UP_j$  values and thereby the TRR-values. This pool dependence is of great interest for the decision analyst who is formulating the model set-up. The alternatives should in this respect be scrutinised so that the calibration pool only consist of alternatives that are possible solutions (Leleur, 2008).

### 4.3 COSIMA calculation example

The COSIMA calculations can be illustrated with a simple calculation example based on Hiselius et al. (2009).

At a present alternative survey four alternatives A1, A2, A3 and A4 are available. Using a national cost-benefit manual and its fixed unit price values the total benefits are calculated: B1, B2, B3 and B4, which by dividing with the observed total expenditure C1 C2, C3 and C4 leads to BCRs for the four alternatives, see Table 4.1 below.

**Table 4.1:** Benefit-cost data for the calculation example

Alternative	A1	A2	A3	A4
B1..B4	110	160	165	120
C1..C4	70	80	120	65
BCR	1.57	2.00	1.38	1.85

If there is agreement among the decision-makers - after the content of the cost-benefit analysis is reviewed - that decision-making is complete, a decision to choose A2 can be taken, since this alternative has the highest BCR value = 2.

If the CBA is insufficient, you can further involve new criteria to evaluate them by a multi-criteria analysis (MCDA) and finally perform a composite analysis after the COSIMA principles. The procedure is as follows:

1. First a number of criteria are described. In this example k1, k2, k3 and k4. This should be done so overlap with components of the cost-benefit analysis is avoided.
2. Next, the four criteria are rated and weighted. Rates means that each alternative for each criterion is assigned a value (score), which lies between 0 and 100. The value 0 is given to the alternative that is performing worst under the given criterion and 100 to the alternative which is performing the best. The two remaining alternatives will have values between 0 and 100. The approach to this consists of a pair-wise comparison of all four alternatives under each of the four criteria k1, k2, k3 and k4. For each of these criteria is it with four alternatives needed  $(4 \cdot 3) / 2 = 6$  pair-wise comparisons. More technically, by using the MCDA method REMBRANDT, where the results are translated into a value function. It gives the following results in Table 4.2:

**Table 4.2:** Value function scores for the calculation example

Criteria /Alternative	A1	A2	A3	A4
k1	25	100	0	45
k2	0	75	60	100
k3	0	26	100	35
k4	100	68	35	0

3. Since the criteria usually do not get assigned equal importance by the decision-makers, the criteria are assigned the weights K1, K2, K3 and K4. This can be done directly or using the ranking criteria of importance (ROD technique) or swing weight (SW). The result, where the weights summarize to 1, is for example: (K1, K2, K3, K4) = (0.20, 0.55, 0.10, 0.15).

4. In the last part of the calculation the CBA and MCDA is linked together, which is done by decision makers provides an MCDA %. At a high MCDA % the MCDA dominates the final result, while a low MCDA % means that it will be the CBA and the detected BCR-values that dominates.

The decision-makers are asked what effect the MCDA should have and may answer 50%. Since this is a relative percentage, it means that the CBA % is 50 as well and MCDA and CBA must therefore count the same in the overall analysis. Based on the choice of A2 with the highest BCR the MCDA now should "count the same". Benefit value B2 was found to 160 which mean that the MCDA part of A2 should sum to 160. Now p1 can be determined in the following manner, as p2, p3 and p4 is expressed by p1 and the available criteria weights:

$$100 \cdot p_1 + 75 \cdot p_2 + 26 \cdot p_3 + 68 \cdot p_4 = 160 \quad \Rightarrow$$

$$100 \cdot \frac{0.20}{0.20} \cdot p_1 + 75 \cdot \frac{0.55}{0.20} \cdot p_1 + 26 \cdot \frac{0.10}{0.20} \cdot p_1 + 68 \cdot \frac{0.15}{0.20} \cdot p_1 = 160$$

Hereby the set of prices is determined:

$$p_1 = 0.43$$

$$p_2 = \frac{0.55}{0.20} \cdot 0.43 = 1.19$$

$$p_3 = \frac{0.10}{0.20} \cdot 0.43 = 0.22$$

$$p_4 = \frac{0.15}{0.20} \cdot 0.43 = 0.32$$

With this set of prices the following values of total rate (total rate of return TRR) is given, which is expressing the overall attractiveness of an alternative from CBA and MCDA:

$$TRR(A1) = \frac{110 + (25 \cdot 0.43 + 0 \cdot 1.19 + 0 \cdot 0.22 + 100 \cdot 0.32)}{70} = 2.18$$

$$TRR(A2) = 4.00$$

$$TRR(A3) = 2.25$$

$$TRR(A4) = 4.09$$

From this it is seen that A4 is the most attractive alternative. In Table 4.3 the results are shown together:

**Table 4.3:** The results of the calculation example

	A1	A2	A3	A4	Method	Unit
Costs	70	80	120	65	CBA + NM	Mkr
Benefits	110	160	165	120	CBA + NM	Mkr
BCR	1.57	2.00	1.38	1.85	CBA + NM	
k1	11	43	0	19	MCDA	Evaluation Mkr
k2	0	89	71	119	MCDA	Evaluation Mkr
k3	0	6	22	8	MCDA	Evaluation Mkr
k4	32	22	11	0	MCDA	Evaluation Mkr
Total MCDA	43	160	104	146	MCDA	Evaluation Mkr
Total value	153	320	269	266	CBA + NM + MCDA	Mkr + Evaluation Mkr = attractiveness Mkr
Total rate	2.18	4.00	2.25	4.09		

The result of the above example is based on the use of CBA carried out by use of a national manual (NM) and MCDA. CBA + NM produce a monetary result, which is validated from a socio-economic thinking and common use. MCDA produces a result which is based on preferences indicated in the decision and the result is in principle only valid from this point of view.

By comparing the MCDA results with CBA results “evaluation Mkr” are calibrated, as seen in Table 4.3 and as explained in the above example. The total value for an examined alternative is found by adding Mkr (found by CBA) by 'evaluation Mkr' (found by MCDA). Mixing of Mkr and the imaginary 'evaluation Mkr' are then expressed by the unit 'attractiveness Mkr', which not only are socio-economically based as it now also includes 'evaluation Mkr' based on specific MCDA preferences. The final result of the analysis is presented by the total rate, where the result in 'attractiveness Mkr' is weighed against the costs for the alternative concerned, see Table 4.3. In the table you can see for each MCDA criterion, which values (in 'evaluation Mkr') the method assigns each alternative. It is noted that because of the approach of the method the worst alternative is always given the value 0.

By the COSIMA calculations it is possible to base the decision of the choice of alternative on the socioeconomic BCR contribution attributed to an additional, wider social evaluated value. This combined result expresses the overall attractiveness of a given alternative. As mentioned, the CBA result is valid from a socio-economic evaluation, while MCDA in principle only is valid from the decision conference proceedings, etc.

The values in Table 4.3 are developed on the basis of the phrase "influence of the MCDA must be 50%". How can this be interpreted further? The basis for choosing between A1, A2, A3 and A4 is a CBA, which shows that the A2 due to the highest BCR (= 2.00) is the best choice. This BCR-value for A2 is given by C2 = 80 and B2 = 160. A balance between CBA and MCDA must be arranged so that the MCDA criteria indirect priced also contributes with 160, which has just been completed in the calculation example. As a result of fixed alternative-scores and criteria weights the unique total rates are determined for all four alternatives, and A4 stands as the most attractive alternative. From a calibration point of view is the principle, however, that CBA and MCDA, based on A2 both counts by 50%.

### **4.3.1 Pros and cons of COSIMA**

The COSIMA model is straightforward in its design and application as it simply “adds to” (and does not hide or change) CBA information. Through its incorporation of relevant MCDA criteria COSIMA may be seen as an MCDA extension of the traditional CBA and may therefore by administrative units be perceived as less of a “black box” than other types of current MCDA decision aid tools.

However, COSIMA depends on an interpretation of the unit prices of the MCDA criteria. What does it mean that given MCDA impacts must contribute with a certain percentage of the total amount of benefits associated with the reference project(s) – which indirect trade-offs lie behind such percentages? The method assumes that the users can express their preferences by the percentage measures as described above.



## 5 Case examples

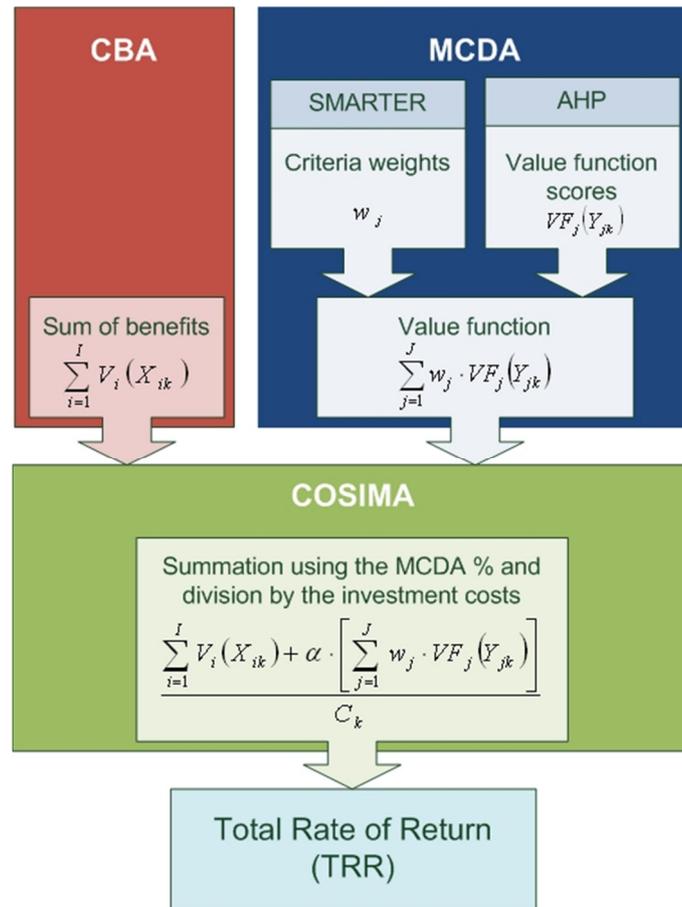
The following sections present two different case examples addressing the combined use of CBA and MCDA. The first case example (case 1) shows the use of the COSIMA DSS in a decision situation while the second example (case 2) addresses how the uncertainties related to criteria weights can be handled using robustness analysis.

### 5.1 Case 1: Composite decision support by combining CBA and MCDA

This example concerns composite decision support based on combining CBA with MCDA for the assessment of economic as well as strategic impacts within transport projects. The COSIMA DSS ensures that the assessment is conducted in a systematic, transparent and explicit way. The approach is presented through a case study concerning alternatives for the construction of a new fixed link between the city of Frederikssund and the peninsula of Hornsherred on the north-eastern part of Zealand, Denmark (Barfod, 2006).

Additive value functions and the AHP are well established and widely applied methods for MCDA (Belton and Stewart, 2002; Hwang and Yoon, 1995; Saaty 2001; Tsamboulas 2007). Consequently, they seem appropriate to use in the COSIMA DSS. Moreover, in terms of transparency, the additive model appears most favourable, since it is considered to be able to cope with almost any problem (Belton and Stewart, 2002; Jiménez et al. 2003). It is, however, commonly known that the main drawback of these methods is the assignment of criteria weights, since individuals are determining these weights. On the other hand, the performance (scores) of alternatives for each criterion is determined objectively, even if artificial scales are used for non-quantifiable criteria. However, the AHP contributes to overcoming this disadvantage by deriving weights in a quasi-independent manner, using pair wise comparisons that make it difficult to promote open biases towards specific criteria. Thus, AHP is a common method used for prioritisation when having a wide variety of choices. More specifically, with regard to the application of the DSS for the case study, the group that assigned the weights was composed of key stakeholders involved in the project.

Figure 5.1 depicts an overview of the methodological approach resulting in the TRR incorporating the CBA, the MCDA applying the AHP and SMARTER techniques and the summation in COSIMA using the MCDA % ( $\alpha$ ).



**Figure 5.1:** Overview of the steps in the COSIMA DSS

It is important to note that the TRR-value due to the theoretical differences between CBA and MCDA has no economic argument like e.g. the BCR. However, the TRR makes it possible to incorporate and hereby retain the information from e.g. the BCR in its original form. Additionally, the TRR delivers information concerning the expression of the alternatives performance in relation to the MCDA-criteria – all included in one single rate. Hence, the COSIMA approach provides the decision-makers with a composite measure of attractiveness.

### 5.1.1 The case

The case study considered concerns the city of Frederikssund which is situated at Roskilde Fjord in the northern part of Zealand, Denmark. The fjord is separating a peninsula, Hornsherred, from Frederikssund and the rest of the Greater Copenhagen area. Currently, the only connection across the fjord is a bridge featuring only one lane in each direction. This is creating a huge bottleneck for the traffic in the area around the city of Frederikssund. Moreover, the Danish government has current plans for the construction of a new motorway between Copenhagen and Frederikssund; this will only lead to a further increase of the traffic problems in the area. Several preliminary studies with the purpose of finding a solution to the traffic problems have been conducted by the municipality of Frederikssund in cooperation with the Danish Ministry of Transport (Barfod, 2006). According to (Barfod, 2006) only four alternative solutions seem realistic for relieving the current traffic situation in Frederikssund:

1. An upgrade of the existing road through Frederikssund and the construction of a new bridge parallel to the old bridge
2. A new high-level bridge and a new by-pass road south of Frederikssund
3. A short tunnel with embankments and a new by-pass road south of Frederikssund
4. A long tunnel without embankments and a new by-pass road south of Frederikssund

According to the characteristics of the above mentioned alternatives different impacts will be derived from each alternative implying different investment costs and layouts. The primary stakeholders behind the project, the Region and the municipalities in the area, have formulated the goal and objective of the project as follows (Barfod, 2006):

*Improve the traffic flow across the fjord ... the solution should take great considerations as concerns the environment in the form of traffic derived consequences (e.g. noise and air pollution) and furthermore consequences derived from the construction itself (e.g. nature and landscape).*

This statement calls for a broader type of appraisal than a conventional CBA, which as mentioned in Denmark only includes the first type of consequences, however, supplemented with a verbal description of the last type. For this reason the comprehensive approach of the COSIMA DSS embracing both monetary and more strategic consequences is applied to the case study in order to produce informative decision support to the decision-makers. In order to make the appraisal of the alternatives as comprehensive as possible, representatives for key players involved in the decision process are gathered to systematically discuss and analyse the issues at a so-called decision conference as described by (Phillips, 1984; Phillips, 2006). The objective of such a decision conference is to constructively deal with the conflicting issues at hand so that a common understanding of the issue can be achieved (Mustajoki et al. 2007). The COSIMA DSS is consequently used to model the viewpoints of the participants and to evaluate the alternatives in an auditable manner.

### **5.1.2 The socio-economic analysis**

The first step in the COSIMA analysis is to conduct a socio-economic CBA (to derive:  $V_i(X_{ik})$  from (4.2)). This CBA is carried out in a model named the CBA-TEN model in accordance with the Danish Manual for Socio-Economic Appraisal (DMT, 2003). Thus, the CBA includes an assessment of the principal items: Construction and maintenance costs, travel time savings and other user benefits, accident savings and other external effects, taxation, scrap value, and finally tax distortion. A traffic- and impact model is set up in order to calculate the impacts derived from each project alternative and the construction and maintenance costs are estimated (Barfod, 2006). All impacts are then entered into the CBA-TEN model, where forecasting is added according to assumptions about the future development in traffic. The various economic parameters necessary for conducting the CBA are set in accordance with Danish standards (DMT, 2003) and will not be treated further in this paper. Finally, the feasibility of the alternatives is described by three different investment criteria in the model: The BCR, the internal rate of return (IRR) and the net present value (NPV). For elaborating description of the investment criteria see e.g. (Leleur, 2000). The results for all the four alternatives are shown in Table 5.1.

**Table 5.1:** Investment criteria and the sum of benefits ( $V_i(X_{ik})$ ) for the four alternatives concerning the case

	High-level bridge	Short tunnel	Long tunnel	Upgrade
BCR	1.63	0.99	0.25	0.36
IRR [%]	8.84	5.95	2.06	2.76
NPV [mil DKK]	415.5	-10.2	-1,869.3	-235.6
$V_i(X_{ik})$ [mil DKK]	1076	965	607	133

The CBA results clearly show that a high-level bridge is the only feasible alternative if the decision is to be based solely on monetary impacts. The three other alternatives are not feasible according to the investment criteria; the short and the long tunnels because of their high construction costs and the upgrade because of its less significant user benefits.

### 5.1.3 The MCDA

The second step of the COSIMA analysis, thus, comprises a MCDA in order to determine scores for the alternatives and weights for the criteria, i.e.  $VF_j(Y_{jk})$  and  $w_j$  from (4.2).

Applying creative techniques such as brainstorm at the decision conference the group decided to include four different MCDA-criteria for the case study, covering the ‘missing’ effects of the CBA. Special attention was made in the definition of the criteria in order to avoid double counting. The criteria definitions are depicted in Table 5.2.

**Table 5.2:** The criteria to be assessed by the MCDA

	Definition
Accessibility	The criterion ranges from local accessibility through regional accessibility to public accessibility and favours alternatives that contribute to improve the overall accessibility in the transport network.
Urban development	The criterion favours alternatives that contribute to develop the existing parts of the city considered plus the opportunity to expand the city and develop new parts.
Landscape	The criterion covers the alternatives visual impact on the landscape and favours those alternatives, which have the least negative impact on this plus on recreational areas and areas worthy of preservation.
Environment	The criterion covers the environmental issues that are not treated in the CBA, i.e. maritime conditions in the fjord plus plant and animal life in and around the fjord.

### Scoring of alternatives

In order of determining the impact of the alternatives within the MCDA-criteria (the value function scores  $VF_j(Y_{jk})$  from (4.2)) an appropriate assessment technique (MCDA method) has to be chosen. Several assessment techniques are as mentioned available for the purpose of determining the value function scores ( $VF_j(Y_{jk})$ ). However, as one of the purposes set out for the COSIMA DSS is to assure to be transparent and easily understandable for decision-makers the well-established and widely applied pair wise comparison

technique AHP by Saaty (2001) is used. It should be noted, that even though the AHP technique is considered to be transparent and appropriate for the current case study, other more or less complicated case studies may call for other techniques.

Using the AHP technique, the decision-makers are involved in direct ratings via pair wise comparisons (Belton and Stewart, 2002; Saaty, 2001) of the alternatives within each of the criteria. For each comparison the decision-makers have to state the strength of their preference for one alternative over another according to the semantic scale that spans from equal preference to absolute preference (1 to 9 on the numerical scale) (Saaty, 1977). Next, the scale values obtained by the pair wise comparisons are for each criterion implemented in a comparison matrix, and normalised scores (AHP scores) for the alternatives are derived using the geometric mean method (Barzilai et al. 1987; Hwang and Yoon, 1995). These normalised scores are computed into value-function (VF) scores utilising a local scale from 0-100, where the score 0 is describing the worst performing alternative and the score 100 is describing the best performing alternative (Belton and Stewart, 2002). All other alternatives will receive linear intermediate scores relative to the two end points. It is noted that the use of a local scale limits the appraisal to concern only the relationship between the already identified alternatives; no absolute measure of their performance is obtained. A local scale is considered to be a useful solution when only dealing with alternatives for one project, i.e. dealing with a closed system. The local scale would, however, not be adequate if it should be possible to include more alternatives at a later stage in the appraisal. In such a case it would be necessary to revise the scale or use a global scale taking the extreme endpoints into account (for more information about local versus global scales see e.g. (Belton and Stewart, 2002)). The COSIMA DSS is customised to the specific problem at hand and thereby assumes that no other alternatives than those identified at the preliminary stage, will be relevant for the appraisal, hence the local scale is appropriate to use.

As a result of the choice of the AHP technique the participants at the decision conference were faced with full pair wise comparisons of the four alternatives within the four criteria comprising a total of 24 comparisons. In order to assure the validity and reliability of the comparisons, time was allocated for a thorough discussion of each comparison and the rationale were recorded in an assessment protocol. Efforts were in this respect made for the participants to reach consensus on each of the comparisons before moving on to the next. In the cases where it was not possible to agree upon the comparisons the different viewpoints were noted with a view to a later sensitivity analysis if felt needed by the participants. In addition to these efforts a consistency check of the comparisons were made and the participants were notified and asked to revise one or more comparisons if the inconsistency index exceeded 0.1 (Goodwin and Wright, 2009).

Next, the AHP-scores derived from the input of the participants are computed into VF-scores. The VF-scores are for each of the four project alternatives shown in Table 5.3. Considering these scores special attention should be paid if nearly identical AHP-scores are derived and succeeding transformed into very varying VF-scores. If this is the case, the criterion from the sample should be assigned with a low weight or maybe even omitted as it does not contribute to the segregation between alternatives and consequently it will not be relevant to include in the appraisal. The decision analyst should after the completion of the pair wise comparisons study the VF-scores, perform a check-up with regard to the above and notify the participants if the issue above is relevant.

**Table 5.3:** Value function scores describing the alternatives performance within the criteria

	High-level bridge	Short tunnel	Long tunnel	Upgrade
Accessibility	100	100	100	0
Urban development	100	100	100	0
Landscape	0	7	34	100
Environment	8	0	46	100

Table 5.3 depicts the VF-scores for the four alternatives within the four types of MCDA-criteria. It is noted that three of the alternatives, i.e. the High-level bridge, the Short tunnel and the Long tunnel, have identical VF-scores for “accessibility” and “urban development” namely 100 while the Upgrade’s VF-scores are 0 for both criteria. This is due to the former’s identical alignments, which have much larger impact within the two criteria than the latter. Hence, the large span between 0 and 100 seems reasonable within the criteria. Within the “landscape” and “environment” criteria the alternatives differentiate more between each other and further investigation is not needed. Clearly, the Long tunnel alternative has the overall best performance within the four criteria. It is for this reason the most attractive alternative if the importance of each of the four criteria is weighted equally. However, the decision-makers would most often feel that some criteria are more important than other. Thus, a weighting procedure describing the importance of each criterion is assigned.

### Weighting of criteria

For the determination of the criteria weights ( $w_j$ ) the SMARTER technique (Goodwin and Wright, 2009) using ROD weights (Roberts and Goodwin, 2002) is applied to the COSIMA DSS. The ROD-weights are surrogate weights which provide an approximation to unrestricted original weights. Surrogate weights based on rankings have been proposed as a method for avoiding the difficulties associated with the elicitation of weights in MCDA (Belton and Stewart, 2002). The decision making process is thereby simplified as the technique only requires an importance ranking of the four MCDA-criteria. Thus, no specification of the weights is needed from the decision-makers as these are determined by the ROD technique and assigned to the criteria according to the ranking. Hence, the participants at the decision conference were faced with the task of ranking the criteria in order of importance. Different viewpoints were expressed by the participants; however, it was possible through discussion to reach consensus in the group. The ranking agreed upon is in correspondence with the SMARTER technique assigned with the ROD-weights as shown in Table 5.4.

**Table 5.4:** ROD weights assigned to the criteria according to the level of importance

Rank	Criteria	ROD-weight
1	Accessibility	0.42
2	Environment	0.30
3	Landscape	0.19
4	Urban development	0.09

### 5.1.4 Combining CBA and MCDA

The last parameter in (4.2) – the  $\alpha$  indicator – is determined as a balance (weight) between the CBA and the MCDA and expressed by the MCDA%. The calibration of the model was made using all four alternatives as all were regarded to be serious contenders for the final choice; the High-level bridge due to its performance within both the CBA and MCDA, and the three other alternatives due to their performance within the MCDA. As noted, the CBA calculation result remains unchanged at all stages in the composite appraisal, but different values of  $\alpha$  (the MCDA%) will change the weight of the MCDA on the TRR. The MCDA% will always be less than 100 as the CBA outcome always will have influence on the TRR i.e. the CBA result cannot be omitted from the appraisal. Figure 5.2 discloses the result covering the agreed ranking of the MCDA-criteria. Note that MCDA% values larger than 80 are not shown on the figure as no changes in the ranking between the alternatives take place based on these values.

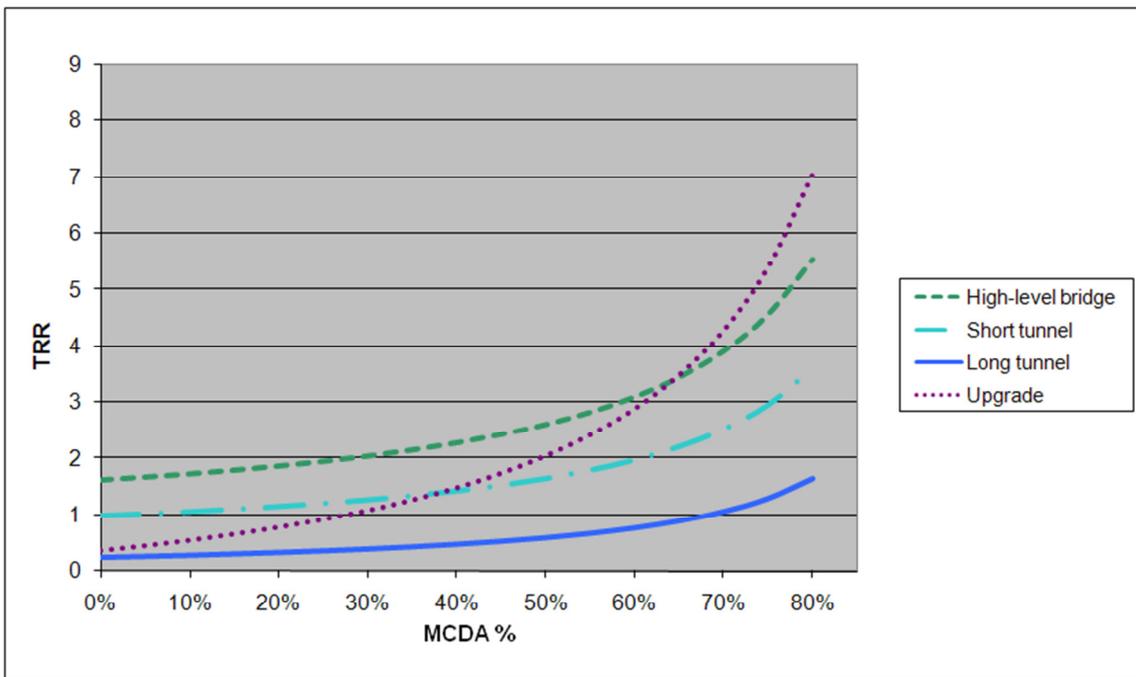


Figure 5.2: TRR values for the four alternatives as a function of the MCDA%

Figure 5.2 depicts that the TRR-values increase for all four alternatives as the MCDA adds a higher importance. However, it can be seen that the TRR for the upgrade of the existing connection is increasing more rapidly than the other alternatives which is due to lower construction costs for this alternative, compared with the three others. The increase, however, is also very dependent on which MCDA criteria that are considered the most important ones. The ranking of criteria in Table 5.4 shows accessibility and environment as the most important criteria. As a result of this the alternatives with high scores on these criteria, will also have the largest increase in their TRR-value and vice versa. The TRR-values point out different alternatives as the most attractive depending on which MCDA% that is considered. However, it is not always beneficial to let the decision-makers base their decision on an interval result, e.g. from 0 to 80 %. In most cases the decision-makers should agree upon a specific MCDA% or a short interval (e.g. 30 – 50 % MCDA) before deciding about the project. Practical experience so far points to MCDA% values in the range between 20 to 30 % (Barfod, 2006). Furthermore, it seems that high MCDA% values are most likely to be adopted when appraising larger and more complex strategic infrastructure projects e.g. the Fehmarn

Belt fixed link between Germany and Denmark. The MCDA% to base the decision on will also vary depending on society's economy, current political tendencies and the type of project being appraised.

The participants at the decision conference were asked to express their preferences with regard to the CBA/MCDA weighting and there was agreement that CBA was the most important part of the appraisal as the project is not regarded to be one of the large strategic infrastructure projects mentioned before. After a discussion a MCDA% set to be 30 was chosen, and with this CBA/MCDA balance applied the high-level bridge is still clearly the most attractive alternative. Thus, this alternative appears as the most robust choice based on the comprehensive appraisal.

## 5.1.5 Results

Summarising the calculations and the process concerning the case it is noted that the analysis is based on the use of CBA and MCDA. The CBA produces results that can be measured in a monetary unit – here million Danish Kroner (m DKK). The MCDA on the other hand produces results that by comparison with the CBA results can be calibrated to 'assessment m DKK'. In order to obtain a total value for an examined alternative m DKK and 'assessment m DKK' is added. This mix between m DKK and the fictitious 'assessment m DKK' is expressed by the unit 'attractiveness m DKK'. The result can also be presented as a total rate of return (TRR), where the result in 'attractiveness m DKK' is divided by the investment costs, see Table 5.5. In this context it should be noted that the process based on input (scoring of alternatives, determination of criteria-weights and balancing the CBA and MCDA) makes it possible to provide a transparent evaluation process involving the decision-makers.

**Table 5.5:** Results of the composite COSIMA analysis using a MCDA% set to be 30

	High-level bridge	Short tunnel	Long tunnel	Upgrade	Method	Unit
Investment costs	661	975	2477	369	CBA	m DKK
Total benefits	1076	965	607	133	CBA	m DKK
<b>BCR</b>	<b>1.63</b>	<b>0.99</b>	<b>0.25</b>	<b>0.36</b>	CBA	
Accessibility	165	165	165	0	MCDA	Assessment m DKK
Urban development	101	101	101	0	MCDA	Assessment m DKK
Landscape	0	8	42	122	MCDA	Assessment m DKK
Environment	11	0	66	144	MCDA	Assessment m DKK
Total MCDA	277	274	374	266	MCDA	Assessment m DKK
<b>Total value</b>	<b>1353</b>	<b>1239</b>	<b>981</b>	<b>399</b>	CBA+MCDA	m DKK + 'Assessment m DKK' = 'Attractiveness m DKK'
<b>Total rate</b>	<b>2.05</b>	<b>1.27</b>	<b>0.40</b>	<b>1.08</b>		

Thus, the result of the COSIMA analysis is that using decision-maker involvement it is possible to apply values to the MCDA-criteria which are comparable to the monetary values from the CBA. The results depicted in Table 5.5 indicate the 'gain' by choosing an alternative which performs well within the MCDA

instead of the alternative which performs the worst. In strategic terms the decision-makers in the case study would achieve most from the investment by choosing the Long tunnel alternative (MCDA alone). However, overall (CBA + MCDA) the High-level bridge alternative will continue to be the most attractive.

As mentioned the existing assessment framework in Denmark does not attempt to incorporate the strategic issues (the MCDA-criteria) of a decision problem into appraisals of transport infrastructure projects. Using the COSIMA DSS the decision-makers are provided with a result that contains a level of information which comprises both the CBA and MCDA expressed in an easy accessible way. Generally, decision-makers are used to make decisions on the basis of a BCR and are hence comfortable with this type of expression. The new feature in the COSIMA DSS is that the MCDA part is converted to the same scale as the CBA part providing the decision-makers with an indication of the value of the strategic issues based on their own preferences expressed as a total rate of return. The TRR result will most likely vary based on who is stating the preferences, however, by assuring diversity in the assessment group the result becomes valid to a wide audience.

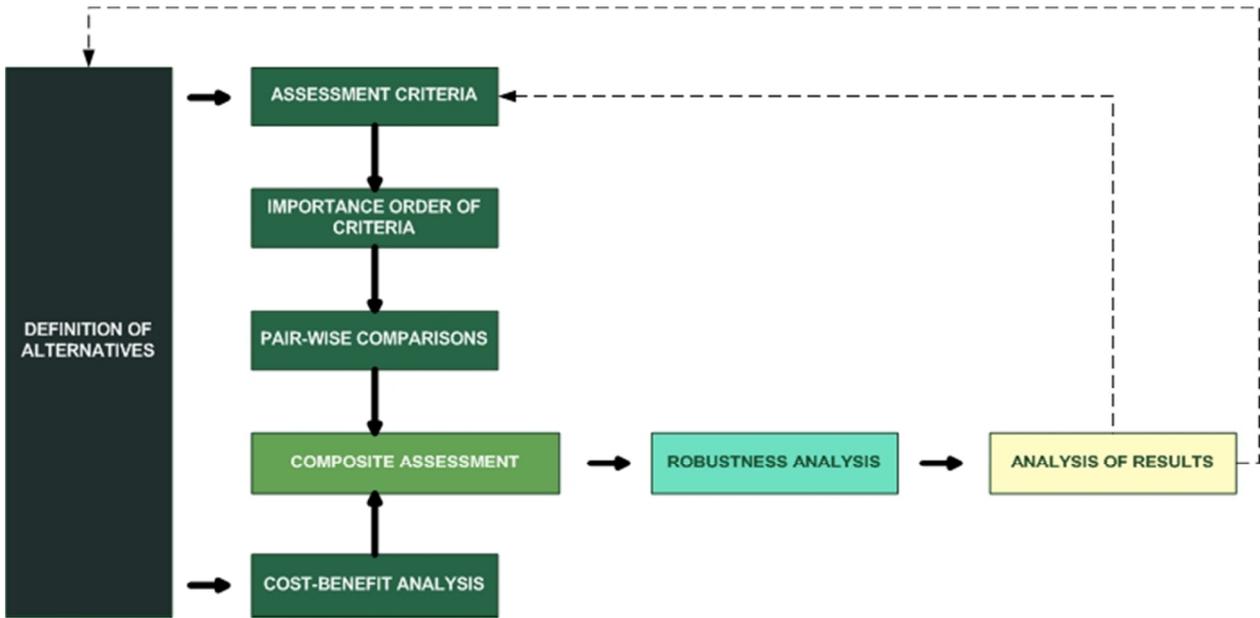
## **5.2 Case 2: Importance of criteria weights in composite appraisal of transport projects**

This example concerns the use of robustness analysis which can be applied to the COSIMA methodology.

In order to include the non-quantifiable, or non-monetised, impacts into the appraisal it is necessary for the involved stakeholders to discuss and prioritise the different impacts. For many decision makers and stakeholders the task of setting the criteria weights for several criteria can be very difficult (Alfares and Duffuaa, 2008b). To overcome this, the proposed method uses importance weights based on rankings of the criteria, by the use of ROD weights (Roberts and Goodwin, 2002). This reduces the problem to assigning a rank order value for each criterion; the ROD methodology then converts the ordinal ranking into a numerical weight. It should be noted that there exists several different criteria rank-weight methods (see (Belton and Stewart, 2002) and (Alfares and Duffuaa, 2008a)). The COSIMA methodology uses an indicator which expresses the trade-off between the CBA and MCDA part, resulting in a total rate expressing the desirability of each alternative. However, it should be mentioned that the proposed procedure for estimating the importance of criteria weights is not limited to the ROD and COSIMA methods described above. The REMBRANDT technique is utilised to determine how the alternatives perform on each criteria (Olson et al., 1995). This technique is a variant of the well-known AHP and uses pair-wise comparisons.

In order to capture the stakeholders' preferences a decision conference can be conducted with the purpose of finding a ranking of the desirability of each alternative seen from the involved stakeholders' point of view (Edwards et al., 2007; Phillips and Bana e Costa, 2005). The decision conference should be tailored in order to capture all the different perspectives of the specific decision problem and to give input to the COSIMA model. One of the advantages with involving the stakeholders (and the decision makers) are that many stakeholders tend to reject decision analysis that appear to provide the "right" answer through a single number coming out an economic model. They want the flexibility of implementing their own value judgments in the decision (DeCorla-Souza et al., 1997). Furthermore, a decision made during a decision conference has a higher probability of acceptance than a decision made based on a complex decision process (Goodwin and Wright, 2004).

The framework of the proposed assessment methodology is presented in Figure 5.3.



**Figure 5.3:** Framework for the assessment methodology

The basis of the framework is a set of predefined alternatives. All of these alternatives undergo a CBA. Furthermore a set of assessment criteria are defined. The criteria are ranked after importance and each of the criterion is assessed by the use of pair-wise comparisons. By the use of the COSIMA methodology a composite assessment is made and thereafter the robustness analysis is performed. At the end the result of the total analysis is presented for the decision makers. If the decision makers feel the need for it further iterations of the framework can be conducted, where changes can be made in assessment criteria, ranking of the criteria or/and the pair-wise comparisons. If a new alternative is introduced at a later point the decision makers have to carry out the framework from the beginning again, including a cost benefit analysis for the new alternative.

### 5.2.1 Problem definition

This example considers a deterministic MCDA problem with  $k$  alternatives and  $j$  decision criteria. The criterion weights reflect the relative importance of each criterion. These weights are normalised by making their sum equal to 1 ( $\sum_{j=1}^n w_j = 1$ ). Given the specific performance value  $V_{MCA}(X_{jk})$  of each alternative  $k$  ( $k = 1, 2, \dots, m$ ) in terms of each criterion  $j$  ( $j = 1, 2, \dots, n$ ), the overall performance of each alternative  $k$  can be calculated as follows:

$$P(A_k) = \sum_{j=1}^n w(j) \cdot V_{MCA}(X_{jk}) \quad (5.1)$$

It is assumed that input can be obtained from several individuals, where each individual  $i$  may rank the criteria after importance. Thus, a list of  $n_i$  prioritised (ranked) criteria is given by each individual  $i$ , who gives each criterion  $j$  a rank  $r_{ij}$ . Furthermore, it is assumed that a group of individuals can (partly) agree on a list of prioritised criteria. The rank is assumed to be inversely related to weight (rank 1 denotes the highest weight). The problem is to determine to what extent the ranking of criteria influences on the overall

performance of each alternative. The aim is to determine, on the basis of ROD weights, how robust the alternatives are for changes in the ranking of criteria.

### 5.2.2 Case study 1: Ostlänken

For testing the proposed methodology, two case studies have been examined. The case studies were conducted under the project “An overall evaluation of public transport investments” (Hiselius et al., 2009). For both cases a decision conference was conducted to analyse the stakeholders’ preferences. Both of the case studies are appraisal of public transport systems.

A new high speed railway line is planned in Sweden connecting Stockholm with Göteborg and Malmö. The case study concentrates on the link between Norrköping and Backeby. 4 different corridors have been identified and the decision problem is to choose between one of these 4 corridors.

However, each of the four alternatives has both strengths and weaknesses regarding multiple objectives which the railway link seeks to meet. A CBA had been conducted a prior to the decision conference. The CBA treated the following impacts: investment costs, maintenance costs, operation costs, changes in revenue and travel time. The results of the CBA are shown in Table 5.6. All of the proposed alternatives have a BCR higher than 1, and thereby the all are economically feasible alternatives.

**Table 5.6:** CBA results for planning corridors for Ostlänken

Corridor	Description	BCR
Red	Follows the existing highway	2
Blue, short tunnel	Follows the existing railway, with a short tunnel going into Norrköping	1.77
Blue, long tunnel	Follows the existing railway, with a long tunnel going into Norrköping	1.54
Green	Goes through untouched nature landscape	1.45

The decision conference was held with delegates from the Swedish railway agency and representatives from the municipality of Norrköping and took place in Norrköping. Because of constraints on the available time for the conference it was chosen to use already well documented material as background for the decision conference. The background material used was the Environmental Impact Assessment (EIA) which treats a number of criteria and assesses the alternatives impacts on each criterion. Before the conference it was decided to work with criteria found in the EIA and use the assessments made in the report as a starting point for the discussion.

**Table 5.7:** Criteria from the EIA report

Criteria	Description
City and scenery impression	Visual intrusion
Cultural environment	Cultural historic development and important cultural environment
Nature	Protection and preservation of important nature and biological diversity
Recreation and outdoor life	Outdoor activities and recreational areas
Health	Barrier effect, noise, vibrations, electromagnetic fields and air quality
Natural resources	Raw material, water and climate
Risk and safety	Risk for accidents on and around the railway, transportation of dangerous goods
Disruption (construction period)	Maintaining the surroundings other functions during the construction

After a presentation of the CBA and the criteria from the EIA report (Table 5.7), the delegates were asked to rank the criteria after importance. After some discussion they were able to agree on a ranking of the criteria (Table 5.8).

**Table 5.8:** Criteria weights

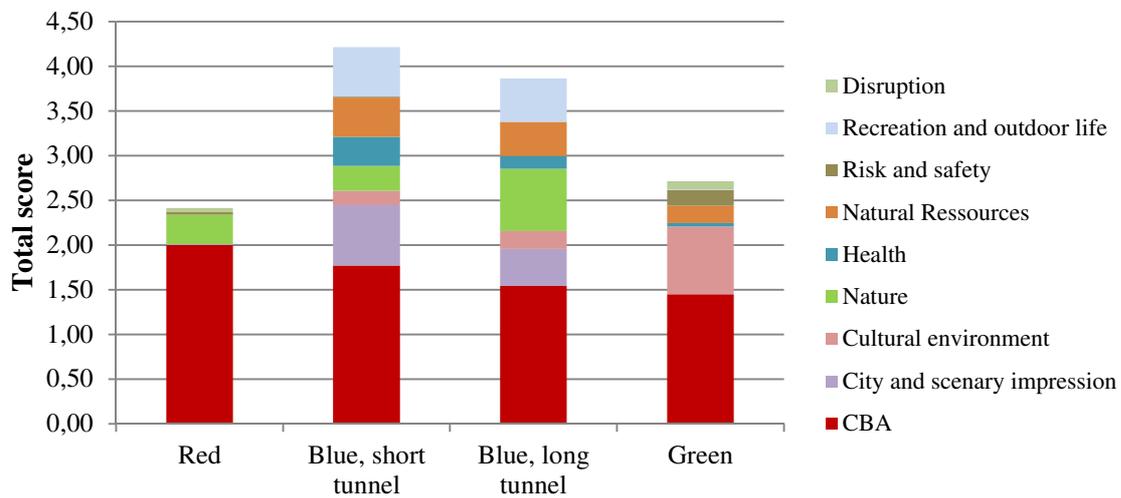
Criteria	Importance order	ROD weight
Town and landscape scenery	1	0.23
Cultural environment	2	0.20
Nature	3	0.17
Recreation and outdoor life	4	0.14
Health	5	0.11
Natural resources	6	0.08
Risk and safety	7	0.05
Disruption (construction period)	8	0.02

For later sensitivity analysis each of the delegate's individual ranking of criteria was also recorded. This step is also made to ensure fairness, so each set of weights proposed is later applied (Labbouz et al., 2008). Then the pair-wise comparisons of the alternatives with respect to each of the criteria were made and a linear value function was used to transform the comparisons to a performance score on a scale from 0 (worst) to 100 (best). At this stage of the decision conference a CBA and MCDA score have been made available and the COSIMA methodology is then used to combine these two scores into an overall desirability score for each alternative. In order to combine the two scores a calibration indicator has to be set (Salling et al., 2007). The delegates of the decision conference agreed to set the indicator to 50 % resulting in an even

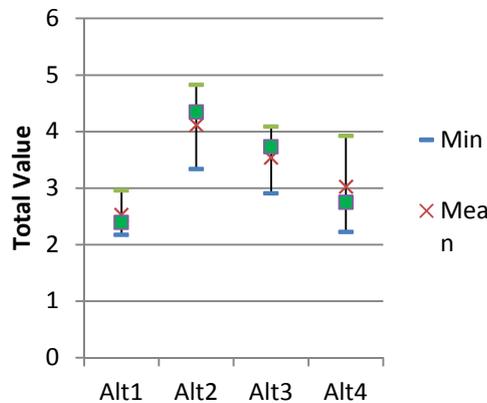
weight on the CBA and MCDA score. The resulting total scores of the alternatives at a calibration indicator set to 50 % can be seen in Figure 5.4.

As seen on the figure above the ranking of the alternatives are: Blue, short (4.21), Blue, long (3.86), Green (2.71) and Red (2.41). The ranking of alternatives, however, does not provide any information on how likely it is that a reversal of the rankings of the alternatives will occur with a change in the criteria weights (Hyde et al., 2005). Therefore an analysis of how robust the decision analyse is for changes in the criteria importance order would be useful.

Figure 5.5 shows the total score for each of the alternatives with the calibration indicator set to 50 % for all possible rankings of the criteria and thereby all possible criteria weights. The points named 'Actual' shows the total score when the criteria weights determined at the decision conference are used. The figure also shows how sensitive the different alternatives are for changes in the criteria weights with alternative 1 being the most robust alternative and alternative 4 the most sensitive alternative for changes in the criteria weights (the widest span between min and max).



**Figure 5.4:** Total scores for the 4 alternatives with a calibration indicator set to 50 %



**Figure 5.5:** Desirability scores for the alternatives at a calibration indicator set to 50 %

The results from the robustness analysis shows that alternative 2 (the Blue short tunnel corridor) is most likely to be the most desirable alternative, but alternative 3 and 4 can also compete when certain criteria weights are chosen. Therefore a further examination of these 3 alternatives is recommended and maybe a more constrained robustness analysis should be conducted with new input from the delegates of the decision conference.

### 5.2.3 Case study 2: Malmö light rail

A new public transport system in the Swedish town Malmö, located in the southern part of Sweden, is to be implemented. The new transport system consists of several light railway lines, which are planned to be introduced in different stages and in the first stage 3 different lines have been chosen to be further examined. The 3 lines run in different parts of the city and therefore differ in which impacts they result in. As a preliminary analysis the municipality of Malmö had a CBA conducted (Hiselius et al., 2009). The CBA was carried out following the national Swedish standard for transport project appraisal and the impacts included in the analysis are: investments costs, maintenance, operation costs, revenue, travel time, impacts on other modes and land value.

**Table 5.9:** Results from the CBA of the 3 light railway lines.

Line	BCR
A	2.26
B	3.28
C	1.96

All of the three alternatives have a BCR over 1 and thereby economically feasible alternatives. In addition the municipality also conducted a wider analysis of the 3 lines covering impacts which are not included in the CBA. As a consequence all these impacts could not be assessed in a monetary way. This wider analysis of the 3 alternatives was performed by a working group formed by the municipality and consisted of 10 different stakeholders. However, the municipality had difficulties in making an overall assessment of the 3 lines taking all the impacts into consideration. Consequently, there was a need for a more comprehensive assessment and a decision conference was held to undertake this assessment. All the 10 participants at the decision conference were stakeholders of the project and member of the before mentioned working group.

One of the main focus areas of the decision conference was to clarify which additional criteria were to be treated in the MCDA. The working group had produced a larger list of criteria which during the first steps of the decision conference was narrowed down to 8 criteria in a process where care was taken to ensure that there were no overlap with the CBA, the criteria were defined to be mutually preference independent and the group had a common understanding of each criterion. With 10 participants at the decision conference it took some discussion before a ranking of the criteria after importance could be agreed on. The group decided on the following criteria to be implemented in the analysis (in preference order): capacity, urban renewal, social integration, compact city, environment, corresponding public transport, impact on car traffic and finally impact potential.

As with case study 1 pair-wise comparison was made for each of the 8 criteria resulting in a desirability score for each alternative.

In order to test the results robustness for different ranking of the criteria all possible combinations of how the 8 criteria can be ranked were calculated. The 8 criteria can be ranked in 40320 different ways and the group consensus on the ranking was only one of all these combinations. The simulation was carried out with the same pair-wise comparisons and same CBA information. Furthermore, all combinations were calculated with the calibration indicator in the range from 0 % to 80 %. A simulation result for alternative B, the alternative with the highest total score, is illustrated in Figure 5.6.

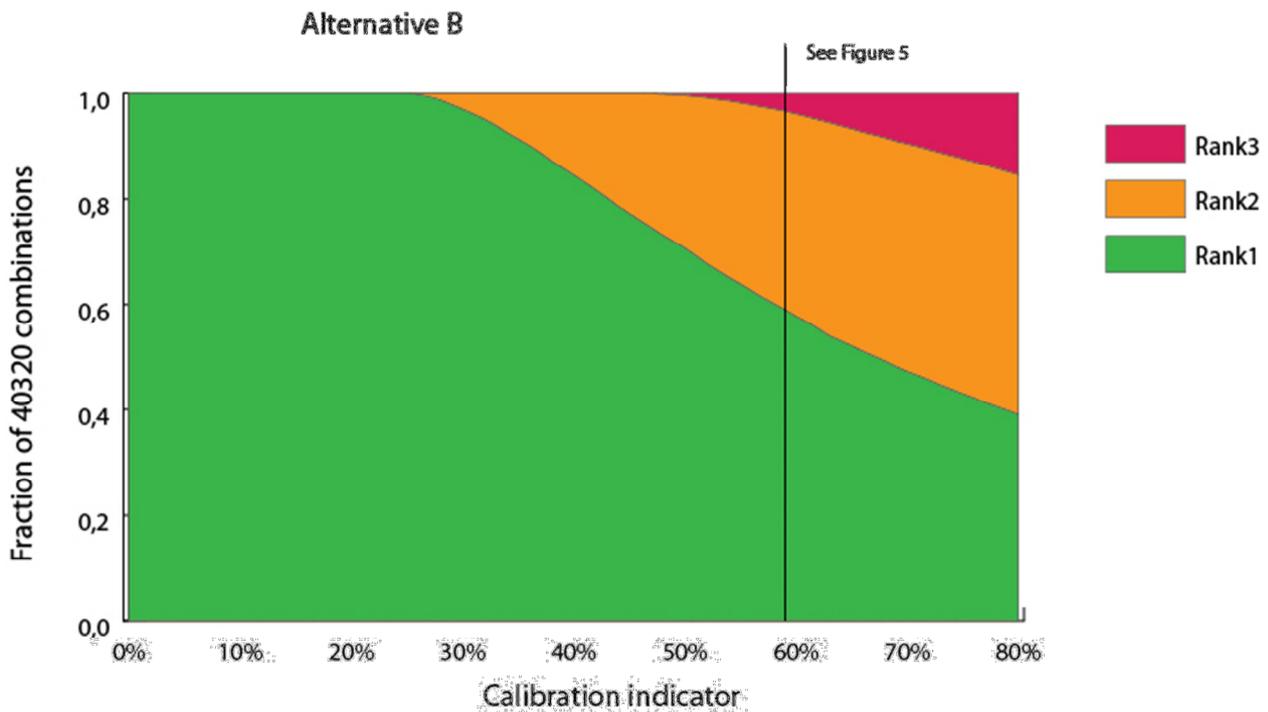
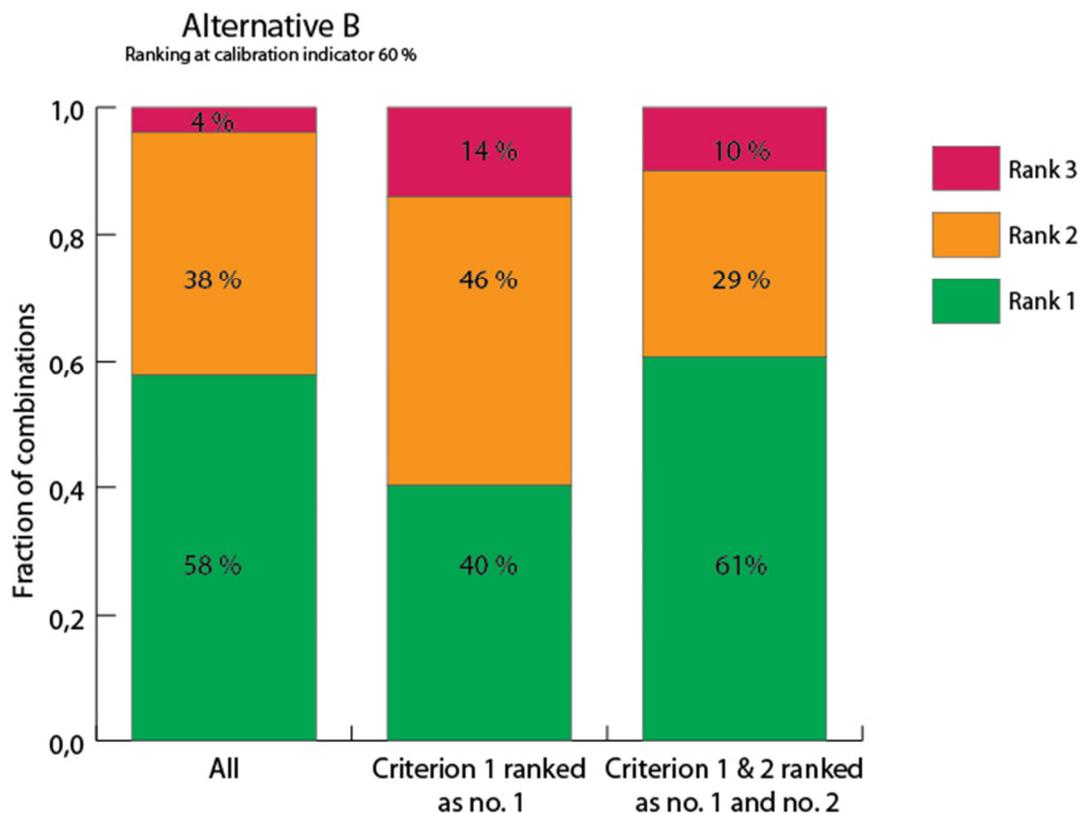


Figure 5.6: Simulation results for alternative B

The simulation results are illustrated in a manner where the decision makers easily can see how robust the total score of the alternative is for changes in the ranking, and thereby the weighting of the criteria. For alternative B it can be seen that in the range of a calibration indicator on 0 % to approximately 25 %, it does not matter how the criteria are ranked after importance – alternative B will be ranked as number 1 among the alternatives in all cases. In the case of a calibration indicator set to 60, which was chosen by the group, alternative B is ranked as the most attractive alternative in approximately 58 % of all possible rankings. The group did choose a ranking which put alternative B as the highest ranked alternative, but other rankings of the criterion would put either alternative A or C as the most desirable alternative. Similar illustrations could also have been made for the other alternatives.

However, the decision conference revealed a tendency in the ranking of the criteria, and this information could be used in the simulation. All the participants' individual ranking of the criteria was recorded and the all ranked the same criterion as the most important, so a simulation with this criterion fixed as ranked number 1 could be made. Furthermore, there was also a tendency in which criterion was ranked as second most important. These tendencies can be simulated and the results can either confirm that the decision

makers (in this case the participants of the decision conference) have a robust result that points out a certain alternative as being the most desirable or that the result is sensitive to changes in the ranking of the criteria. Figure 5.7 illustrates the results from 3 different simulations at a calibration indicator at 60 %. The first bar shows the simulation results with all combinations, where in 58 % of all the combinations alternative B would be ranked as number 1. The next bar is the result from a simulation with the capacity criterion fixed as ranked number 1. It can be seen that there is a lower fraction of all combinations giving alternative B the highest rank than in the simulation with all combinations. The reason for this result is that alternative B has the lowest score among the alternatives for criterion 1 and ranking this criterion as number 1 is then a disadvantage for alternative B. The next bar shows the results for the simulation with capacity and urban renewal criteria fixed as ranked as number 1 and 2. This simulation shows the highest fraction of all possible combinations giving alternative B the highest rank (61 %). A reason for this can again be found in how the alternative B is scoring on the criterion. Alternative B has the highest score on urban renewal criterion, together with alternative A, this gives alternative B an advantage when this criterion is fixed as the next most important criterion.



**Figure 5.7:** Constrained simulation approach with one or two criterion rankings fixed

This approach provides an opportunity to deal with disagreement among the participant about the ranking of the criteria. If the group have difficulties in ranking one or more criteria, a simulation can be made based on keeping all the agreed upon rankings fixed.

## 5.2.4 Discussion

The output of the decision conference is a set of criteria and criteria weights together with assessments of the alternatives with respect to each criterion. The set of criteria and the assessments of alternatives can be seen as the more objective part of the decision problem and defining the decision space, in fact this step can be assessed by experts who have no what so ever interest in the project assessed (Hyde et al., 2005). However the setting of the criteria weights opens up the analysis to a more certain amount of subjectivity (Zietsman et al., 2006) and each setting defines a specific subset of the decision space.

For each examined alternative an interval score is calculated by changing the calibration indicator used by the COSIMA methodology to combine the CBA and MCDA. It is seen as important information for the decision makers to know how robust the calculated ranking of the alternatives are to changes in the criteria weights. This is done by examining all possible rank orders of the criteria, and thereby all possible criteria weights, for the calibration indicator determined by the decision makers. If more than one decision maker is attending the decision conference, both the calibration indicator for the group of decision makers and for each individual can be examined together. The whole decision space which the decision conference has established is then examined, taking into account both the joint and the individual decision on calibration indicator.

This work proposes two variants of a robustness analysis using the rank order of the criteria to estimate the robustness of the decision analysis. In contrast to other techniques for robustness analysis of the weights, the proposed methodology changes all of the weights at once and not just a single weight. The first variant of the robustness analysis methodology is illustrated in case study 1, while case study 2 widens the analysis by taking into account all calibration indicators. When applying the later proposed robustness analysis a clear picture of the total decision space is made available for the involved stakeholders together with the subset the decision conference established. If the group of stakeholders only can agree on one or more criteria weights, a constrained approach is available for examining the robustness of the criteria weights. The approach gives the stakeholders a possibility to lock some of the criteria weights and only examine alternative criteria weights for the remaining criteria. The outcome of the procedure for examining the importance of criteria weights is a graphic illustration of how the project performs as a function of the calibration indicator.

In case study 2 the delegates were asked to rank the criteria in the group before the individual ranking. It can be discussed if this order has some impact on the way the different stakeholders ranked the criteria. The reason for this approach is that the main goal of the decision conference was to reach a consensus about the ranking, and not to specify the individual preferences. However, the correlation between the group ranking and the individual rankings is so significant that it is obvious that the group ranking have had some impact on the individual rankings. Therefore, it is recommended that the order is changed.

To what a degree of consensus among the decision makers is successful? - could be a valid question for this approach. The answer would be that this is to be decided by the involved decision makers. Complete consensus would clearly make the decision making more straight-forward, but not necessarily a goal in itself. In many cases complete consensus is not possible to reach and in these cases the proposed methodology is a valuable tool. This is because the methodology can illustrate if this missing consensus have any critical influence on the ranking of alternatives. It should be mentioned that the use of a multi

criteria decision model, as proposed in this paper, would increase the probability of reaching consensus by encouraging communication from involved individuals (Hall, 2008).

The time available for the two decision conferences were only 4 hours for each of them. This time constraint had influence on how much material it was possible to cover. As a consequence of this each decision conference was tailored in order to be able to cover the most critical aspect of the decision problem. Case study 1's main concern was the importance order of the criteria and the pair-wise comparisons of the alternatives with respect to the already described criteria. For case study 2 the main concern was to establish well described criteria, which did not overlap with the CBA and the importance order of these criteria. In spite of this sparse time for conducting the decision conferences the delegates expressed that they found the process very useful and gave them a good input to their further work with the project. As one of the delegates expressed: "The time available was enough to conduct a useful analysis and a lot more time would not necessarily be a lot better"



# References

- Alfares, H.K. and Duffuaa, S.O. (2008a). Assigning cardinal weights in multi-criteria decision making based on ordinal ranking. *Journal of Multi-Criteria Decision Analysis*, vol. 15, no. 5, pp. 125-133.
- Alfares, H.K. and Duffuaa, S.O. (2008b). DETERMINING AGGREGATE CRITERIA WEIGHTS FROM CRITERIA RANKINGS BY A GROUP OF DECISION MAKERS. *International Journal of Information Technology and Decision Making*, vol. 7, no. 4, pp. 769.
- Bana e Costa, C.A. and Vansnick, J.C. (1994) MACBETH – an interactive path towards the construction of cardinal value functions. *International Transactions in Operational Research* 1:489-500.
- Banister, D. and Berechman J. (2000). *Transport Investment and Economic Development*. UCL Press, London, UK.
- Barfod, M.B. (2012). Optimising Transport Decision Making using Customised Decision Models and Decision Conferences. PhD thesis, Department of Transport, Technical University of Denmark.
- Barfod, M.B., Salling, K.B., and Leleur, S. (2011). Composite decision support by combining cost-benefit and multi-criteria decision analysis. *Decision Support Systems*, 51, 167-175.
- Barfod, M.B. (2006). Appraisal of alternatives for a new Hornsherred connection, Master Thesis, Center for Traffic and Transport, Technical University of Denmark (in Danish).
- Barron, F.H, and Barret, B.E. (1996). Decision quality using ranked attributes weights. *Management Science* 42(11): 1515-1523.
- Barzilai, J., Cook, W.D. and Golany B. (1987). Consistent Weights for Judgment Matrices of the Relative Importance of Alternatives. *Operations Research Letters* 6 (3), pp. 131-134.
- Belton, V. and Stewart, J.T. (2002), *Multiple criteria decision analysis: an integrated approach*, Kluwer Academic Publishers, London.
- Chen, Y., Kilgour, D.M., and Hipel, K.W. (2008). Screening in multiple criteria decision analysis. *Decision Support Systems*, 45, 278-290.
- DMT, Danish Ministry of Transport (2003). Manual for Socio Economic Appraisal: Applied Methodology in the Transport Sector (in Danish). Danish Ministry of Transport, Copenhagen.
- DeCorla- Souza, P., Everett, J., Gardner, B. and Culp, M. (1997). Total cost analysis: An alternative to benefit-cost analysis in evaluating transportation alternatives. *Transportation*, vol. 24, no. 2, pp. 107-123.
- Edwards, W., Miles, R.F. and Von Winterfeldt, D. (2007). *Advances in decision analysis: from foundations to applications*. Cambridge University Press, Cambridge.

- Edwards, W. and Barron, F. H. (1994). SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement. *Organizational Behavior and Human Decision Processes* 60, pp. 306-325, Elsevier.
- EUNET/SASI (2001), Final Report – Executive Summary. 4th RTD Framework Programme of the European Commission.
- European Conference of Ministers of Transport, ECMT group of experts (1981). Exchange of Information Criteria applied to Transport Infrastructure Projects, pp. 16-23.
- van Exel, J., Rienstra, S., Gommers, M., Pearman, A. and Tsamboulas, D., (2002). EU involvement in TEN development: network effects and European value added. *Transport Policy* 9, pp. 299-311.
- Gissel, S. (1999). *Decision aid methods in rail infrastructure planning*. Ph.D. thesis, Report 4, Technical University of Denmark, Department of Planning.
- Goodwin, P. and Wright, G. (2009). *Decision Analysis for Management Judgment*. Fourth edition, WILEY, Great Britain.
- Grant-Muller. S., Mackie, P., Nellthorp, J. and Pearman, A. (2001). Economic Appraisal of European Transport Projects – the state of the art revisited. *Transport Reviews*, 21.
- Hall, D.J. (2008). Decision Makers and Their need for Support. in *Handbook on Decision Support Systems 1*, eds. F. Burstein and C.W. Holsapple, Springer-Verlag, Berlin Heidelberg, pp. 83.
- Hayashi, Y. and Morisugi, H. (2000). International comparison of background concept and methodology of transport project appraisal. *Transport Policy* 7, pp. 73-88.
- Hiselius, L.W., Barfod, M.B., Leleur, S., Jeppesen, S.L., Jensen, A.V. and Hjalte, K. (2009) *Helhetsorienterad utvärdering av kollektivtrafikåtgärder*. Lunds Universitet, Tekniska Högskolan i Lund.
- Hobbs, B. F. and P. Meier (2000). *Energy decisions and the environment – a guide to the use of multicriteria methods*. Kluwer Academic Publishers.
- Hwang, C.L. and Yoon K. (1995), *Multi Attribute Decision Making – An Introduction*, Sage University Papers, Sage Publications.
- Hwang, C. L. and Yoon, K. (1981). Multiple Attribute Decision Making – Methods and Applications, *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, pp 40-51.
- Hyde, K.M., Maier, H.R. and Colby, C.B. (2005). A distance-based uncertainty analysis approach to multi-criteria decision analysis for water resource decision making. *Journal of environmental management*, vol. 77, no. 4, pp. 278-290.
- Janic, M. (2003). Multi-criteria Evaluation of High-speed Rail, Transrapid Maglev and Air Passenger Transport in Europe. *Transportation Planning and Technology* 26 (6), pp. 491-512.

- Jiménez, A., Ríos-Insua, S. and Mateos, A. (2003). A decision support system for multiattribute utility evaluation based on imprecise assignments. *Decision Support Systems* 36, pp. 65-79.
- Keeney, R.L., and Raiffa, H. (1993). *Decisions with Multiple Objectives – Preferences and Value Tradeoffs*. Cambridge University Press.
- Labbouz, S., Roy, B., Diab, Y. and Christen, M. (2008). Implementing a public transport line: multi-criteria decision-making methods that facilitate concertation. *Operational Research*, vol. 8, no. 1, pp. 5-31.
- Larichev, O.I, Kortnev, A.V. and Kochin, D.Y (2002). Decision support system for classification of a finite set of multicriteria alternatives. *Decision Support Systems* 33 (2002), pp. 13-21.
- Leleur, S. (2000). *Road Infrastructure Planning: A Decision-Oriented Approach*. Second edition, Polyteknisk Press, Denmark.
- Leleur, S., Petersen, N.B., and Barfod, M.B. (2007). The COSIMA Approach to Transport Decision Making – Combining Cost-benefit and Multi-criteria Analysis for Comprehensive Project Appraisal. In K-S. Kim (Ed.), *Improving Public Investment Management for Large-Scale Government Projects: Focusing on the Feasibility Studies – Lessons and Challenges* (pp. 100-122). KDI.
- Leleur, S. (2008). *Systemic Planning*. Second edition, Polyteknisk Press.
- Lootsma, F. A. (1999). *Multi-criteria decision analysis via ratio and difference judgement*, Kluwer Academic Publishers.
- Lootsma, F.A. (1992). The REMBRANDT system for multi-criteria decision analysis via pair wise comparisons or direct rating. Report 92-05, Faculteit der Technische Wiskunde en Informatica, Delft University of Technology, Netherlands.
- Mackie, P. and Preston, J. (1998). Twenty-one sources of error and bias in transport project appraisal. *Transport Policy* 5, pp. 1-7.
- MOTOS (2007). *Transport Modelling: Towards Operational Standards in Europe*, CTT, DTU, TREN/06/FP6SSP/S07.56151/022670.
- Mustajoki, J., Hämäläinen, R. P., and Sinkko, K. (2007). Interactive computer support in decision conferencing: Two cases on off-site nuclear emergency management. *Decision Support Systems*, 42, 2247-2260.
- Olson, D. L. (1996). *Decision Aids for Selection Problems*. Springer Series in Operations Research.
- Olson, D.L., Fliedner, G. and Currie, K. (1995). Comparison of the REMBRANDT system with analytic hierarchy process. *European Journal of Operational Research*, vol. 82, pp. 522-539.
- Phillips, L.D. (1984). A theory of requisite decision models. *Acta Psychologica*, 56, 29-48.
- Phillips, L.D. (2007). Decision conferencing. In W. Edwards, R.F. Miles, and D. von Winterfeldt (Eds.), *Advances in Decision Analysis* (pp. 375-399). Cambridge University Press.

- Phillips, L.D. and Bana e Costa, C. A. (2005). *Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing*, The London School of Economics and Political Science, London.
- Roberts, R. and Goodwin, P. (2002). Weights Approximations in Multi-attribute Decision Models, *Journal of multi-criteria decision analysis* 11: 291-303.
- Saaty, T. L. (1977). Scenarios and priorities in transport planning: Application to the Sudan. *Transport Research*, 11, 343-350.
- Saaty, T. L. (2001). *Decision Making for Leaders: The Analytical Hierarchy Process for Decisions in a Complex World*. The Analytical Hierarchy Process Series.
- Salling, K.B., Leleur, S. and Jensen, A.V. (2007). Modelling decision support and uncertainty for large transport infrastructure projects: The CLG-DSS model of the Øresund Fixed Link. *Decision Support Systems*, vol. 43, no. 4, pp. 1539-1547.
- Sayers, T.M., Jessop, A.T. and Hills, P.J. (2003). Multi-criteria evaluation of transport options – flexible, transparent and user-friendly. *Transport Policy* 10, pp. 95-105.
- Seo, F- and Sakawa, M. (1988). *Multiple Criteria Decision Analysis in Regional Planning*. D. Reidel Publishing Company, p. xiii.
- Stewart, T. J. (1993). Use of Piecewise Linear Value Functions in Interactive Multicriteria Decision Support: A Monte Carlo Study. *Management Science* 39, pp. 1369-1381, Informs.
- Stewart, T. J. (1996). Robustness of Additive Value Function Methods in MCDM. *Journal of Multi-Criteria Decision Analysis* 5, pp. 301-309, Wiley.
- Tsamboulas, D.A. and Mikroudis, G.K. (2006). TRANS-POL: A mediator between transportation models and decision makers' policies. *Decision Support Systems* 42, pp. 879-897.
- Tsamboulas, D.A. (2007). A tool for prioritizing multinational transport infrastructure investments. *Transport Policy* 14, pp. 11-26.
- Vickerman, R. (2000). Evaluation methodologies for transport projects in the United Kingdom. *Transport Policy* 7, pp. 7-16.
- von Winterfeldt, D., and Edwards, W. (2007). Defining a Decision Analytic Structure. In W. Edwards, R.F. Miles, and D. von Winterfeldt (Eds.), *Advances in Decision Analysis* (pp. 81-103). Cambridge University Press.
- von Winterfeldt, D., and Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge University Press
- Vreeker, R., Nijkamp, P. and Welle, C.T. (2002). A multicriteria decision support methodology for evaluating airport expansion plans. *Transportation Research Part D* 7, pp. 27-47.

Watson, S. R. and Buede, D. M. (1987). *Decision Synthesis: The Principles and Practise of Decision Analysis*. Cambridge University Press New York, United States of America.

Wright, S., Nelson, J.D., Cooper, J.M. and Murphy, S. (2009). An evaluation of the transport to employment (T2E) scheme in Highland Scotland using social return on investment (SROI). *Journal of Transport Geography* 17, pp. 457-467.

Zietsman, J., Rilett, L.R. and Kim, S. (2006). Transportation corridor decision-making with multi-attribute utility theory. *International Journal of Management and Decision Making*, vol. 7, no. 2-3, pp. 254.