Final Preprint of paper published in *International Transactions in Operational Research* 20 (2013): 559–578. doi: 10.1111/itor.12016

Integrated Value Function Construction with Application to Impact Assessments

Theodor J Stewart* Ron Janssen†

Abstract

An integrated structure is provided for processing various forms of imprecise preference information in the context of multicriteria impact assessments. Linear programming formulations generate best-fit value function models and associated ranking of alternatives, both when preferences are over-determined (leading to potential inconsistencies) or when preference information is incomplete. In the latter context, the algorithm identifies a range of possible rank orders for the decision alternatives under consideration, consistent with the information provided. The approach is primarily aimed at structuring opinions of experts concerning the desirability of different actions in terms of technical aspects, intended as input into the final political decision making process. It is demonstrated that the approach described here can be implemented with modest levels of effort by the experts. Experiences are reported with the approach in the context of a problem of a soil sanitation problem in the Netherlands, in which experts expressed satisfaction with the resulting rank ordering of alternatives.

Keywords: Multiple criteria decision analysis; Environment

1 Introduction

Comparisons of policy alternatives in terms of their environmental impacts are almost inevitably multicriteria in nature. In other words, there exists no one alternative (not even the *status quo*) which is better than all others in terms of all impact measures, so that expert judgements or *value tradeoffs* are needed in order to assess whether losses in terms of one set of impacts are offset by gains elsewhere. Such considerations are central to the field of *multiple criteria decision making (MCDM)* or *multiple criteria decision analysis (MCDA)*. A review of this field may be found in Belton and Stewart [2002].

The problems which we address in this paper relate primarily to the more technical impact problems, where expert knowledge is used to construct a form of preference model, prior to interpretation at the final decision making level (where more qualitative and socio-political priorities may be taken into account). In this situation, relative levels of impact within each criterion and relative importance of different types of impact are based on explicit or implicit knowledge of the experts (and not, for example, on political priorities). A crucial limiting factor is the availability and cost of obtaining expert inputs. In applying the approach to repetitive classes of

*Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, South Africa
†Institute for Environmental Studies (IVM), Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

problems, it is thus beneficial to be able to reuse the preference model several times in distinct but related problems, in order to justify the effort invested in constructing the model. In other words, the outputs from the initial problem analysis and model construction is not only a ranking of alternatives in that case, but (arguably more importantly) also the preference model itself which can be applied in other contexts, perhaps with minor adaptation, in which somewhat different alternatives may need to be assessed. Forms of judgemental inputs from the experts would include both direct comparative assessment of options (in order to contextualize preferences) and more general value judgements such as relative tradeoffs, importance of criteria, etc.

The Dutch soil clean-up case study presented below fits precisely into the context described above. Similar situations with similar types of pollution and the same legal context are found all over the Netherlands. Once the preference model is constructed (largely from consideration of problems at one site), the intention is to have it applied for consistency at many other sites. Another example would be the clean-up off petrol stations. Once the model is constructed, as many as 1000 locations with similar conditions could be addressed, where it would not be feasible to engage experts separately for each station.

Many different schools of MCDA have evolved, such as goal or reference point [e.g. Lee and Olson, 1999, Wierzbicki, 1999] or outranking methods [e.g. Figueira et al., 2005b, Vincke, 1999] but are perhaps not all equally useful in formalizing expert judgements regarding tradeoffs between different impacts in the above explicit and reuseable sense. In our context, it is required of the experts in a particular field (e.g. regarding impacts of different pollutants in soil in a given region) to provide a clear set of motivated recommendations regarding preferences between different policy actions according to their field of expertise. The recommendations of such experts is not the final decision (which would typically be taken by political decision makers, taking cognizance of other issues and criteria), but needs to be fully informative to the decision makers regarding relative strengths of preference from their field of expertise. Within this context, we have adopted an additive value function model, which is a formalization of weighted scoring systems which are widely adopted in this context [Janssen, 2001], taking into account dangers such as over-linearization, is largely transparent to users and experts from many different backgrounds, and which clearly documents the rationale behind any rankings obtained.

There are strong axiomatic and practical operational foundations [Belton and Stewart, 2002, Section 4.2 and Chapter 5] underlying the use of additive value functions to model preferences, i.e. an expression of the form

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

where:

- a is a particualr policy alternative (to be compared with others);
- $V(a)$ is the overall score or "value" associated with a;
- $v_i(a)$ is the partial (standardized) score representing the value of alternative a in terms of criterion i (expressed on an interval scale of preference, which may require transformation from the natural performance measurement); and
- w_i is an importance weight associated with criterion i.

In many cases, especially with the impact studies described here, the partial score $v_i(a)$ is constructed indirectly in a two stage process as follows:

- A natural *measure of performance* of each alternative a with respect to criterion i, say $z_i(a)$, is established, typically a well-defined physical measure or *attribute*, such as ppm. of a specified contaminant, or as a categorical scale of (qualitative) performance. In principle, the methods described below will also work for performance measured ordinally (i.e. as rank orders of alternatives), but such measures are not easily exported to other contexts, which as described above is one of our key aims. For thhis reason, we exclude pur3ely ordinal measures from our discussion.

- This attribute measure will usually not satisfy the properties required of an additive value function as in (1), and needs to be transformed into a standardized value score by means of a *value function*, say $v_i(z_i)$. The important point to emphasize is that tradeoffs between different effects as measured by the attributes z_i are seldom constant over the full ranges of possible outcomes, so that the functions $v_i(z_i)$ cannot in general be assumed to be linear. In fact, we have demonstrated [e.g., Stewart, 1996] that inappropriate linearization can be a major source of bias in using models such as (1).

The assessments of the performance measures $z_i(a)$ are technical and outside of the scope of value measurement. The important observation for our purposes is that the partial scores $v_i(a)$ will be derived from applying the value function to the attribute measures, i.e. $v_i(a) = v_i(z_i(a))$.

In building the scoring model, it is sufficient to assess the *form of the partial value functions* $v_i(z_i)$ for any arbitrary measured level of performance, without reference to any specific alternative a. These value functions together with the importance weights w_i together define the *additive value function*:

$$V(\mathbf{z}) = \sum_{i=1}^{m} w_i v_i(z_i) \tag{2}$$

where $\mathbf{z}$ is a vector of performance measures. For any *particular alternative* a, it is then only necessary to obtain or to estimate the measures $\mathbf{z}(a)$, and to plug these into (2) to obtain $V(a) = V(\mathbf{z}(a))$.

It is in principle possible to identify the $V(\mathbf{z})$ function (i.e. the weights w_i and the partial value functions $v_i(z_i)$) before starting the evaluation and analysis of specific alternatives [see, for example, Belton and Stewart, 2002, Chapter 5], and in some cases of strategic decision making this may be desirable or even essential. Such strategic problems may need the values and tradeoffs to be documented *before* alternatives are evaluated, in order to avoid accusations that the decision models were tuned to favour a particular choice, and to allow regular inclusions during negotiations of new alternatives into the set under consideration.

Such *a priori* function evaluation can be highly demanding in time and effort, however, and will typically require consideration of many tradeoffs in the abstract. These features may render full prior assessment inappropriate in the context of using (2) for integration of expert assessments of impacts as described above. Within this context, there is need to develop a value function within tight time constraints, to a level of precision (but no more) than is needed to rank order alternatives under consideration, both currently and in the future.

As an alternative to complete prior evaluation of $V(\mathbf{z})$, a number of methods (see, for example, [Stewart, 1999] for a review) have been developed for progressive and interactive elicitation of just sufficient information regarding the form of $V(\mathbf{z})$ to allow the required comparative evaluation to be done. Some of the approaches that have been adopted include:

- Use of ordinal or imprecise preference statements only, e.g. a is preferred to b; or the strength of preference for a over b is greater than that for c over d [Cook and Kress, 1991, Salo and Hämäläinen, 1992], or the PRIME approach of Salo and Hämäläinen [2001];

- Use of categorical semantic scales [Bana e Costa et al., 2005, Bana e Costa and Vansnick, 1999];

- Inferring value function models or implied rank orderings which give maximal consistency with stated preference orders, either within a subset of the real alternatives, as in the UTA methodology [Jacquet-Lagrèze, 1990, Jacquet-Lagrèze and Siskos, 2001] and later extensions such as ordinal regression (UTA^{GMS}) [Greco et al., 2008] and GRIP [Figueira et al., 2009], or as conjoint scaling within hypothetically constructed sets, which may also be viewed as an extension of UTA.

In seeking to achieve efficient use of the experts' time in the context of impact assessment as described above, it is necessary (a) to integrate various forms of input from experts, including imprecise ratio judgements as in PRIME and ordinal judgements as in UTA and extensions and conjoint scaling, and (b) to generate an explicit representation of the constructed value model, probably in imprecise terms. The purpose of the present paper is thus to synthesize concepts from various of the above approaches, and to apply the resultant integrated methodology specifically to a practical problem of impact assessment. Although individually each of the individual elements have been described by the authors cited, we are not aware of an integration of these ideas into a unified software framework for the purposes identified above.

The objectives of the paper are thus:

1. To provide a framework for incorporating a number of progressive preference modelling approaches into a single modelling approach capturing information in different forms depending on the preferences of the experts, with the ultimate goal of recording the resultant model (expressd as a piecewise linear approximation) for use in future impact studies (a framework which we have implemented in the DEFINITE software [Janssen and van Herwijnen, 2011]); and
2. To illustrate this integrated approach by means of an application to a soil sanitation problem in the Netherlands.

2 Additive Piecewise Linear Value Functions

As indicated in the previous section, a number of approaches have been developed for incorporating partial and/or imprecise information into the assessment of value functions and for the determination of implied preference orders on the set of alternatives which need to be evaluated. Many of these approaches do make prior assumptions about the functions $v_i(z_i)$ and/or approximate them in some way, such as by piecewise linear functions used in UTA [Jacquet-Lagrèze and Siskos, 2001] as we adopt below.

The ordinal regression and related approaches [Greco et al., 2008, Figueira et al., 2009] do not require any explicit functional approximation. In effect, these methods correspond to our formulation of piecewise liner models described below, but in which the "breakpoints" defined there correspond to performance on each of the alternatives (i.e. ordinal performance evaluation only). These approaches do not, however, permit the incorporation of direct prior information (typically imprecise) on the value function structures. Furthermore, these methods aim primarily at providing pairwise comparison of alternatives directly, and no explicit value function is recovered, even in approximate form, for future use (as we have argued is necessary for the impact study context). We have thus retained a piecewise linear structure applied to cardinal or categorical scales. There is relatively loss of fidelity in such an approximation, as experiments have shown [Stewart, 1993, 1996] that as few as four segments in a piecewise linear approximation are sufficient to provide a level of accuracy commensurate with usual judgemental imprecisions.

The approximation of the partial value functions $v_i(z_i)$ in piecewise linear form (defined more precisely below) allows for a wide range of different forms of value or preference inputs and a rich range of preference models, while also retaining simplicity of presentation. In the case of ordinal preference information (rank ordering of a subset of real or hypothetical alternatives), this corresponds to the UTA approach. Other forms of imprecise preference ratios as discussed by Salo and Hämäläinen [2001] (PRIME) are equally well implemented within the same framework. As indicated in the previous paragraph, in practice we would use no more than 4 linear segments as approximation for each $v_i(z_i)$. However, for purposes of describing the approach, we allow in principle an arbitrary number of segments ν_i for each attribute.

Let us now outline the piecewise linear model in the form which we have implemented. Suppose that we have an MCDM problem in which the criteria of relevance to the evaluation of alternatives have been represented in terms of m attributes (performance measures), say $z_1, \ldots, z_m$. These attributes may be defined on a cardinal or a categorical scale. For ease of discussion, and with no loss in generality, we shall suppose that each attribute is defined in the direction of increasing preference (i.e. such that larger values are preferred to smaller). Let the best (largest) and worst (smallest) values for z_i that need explicitly to be taken into consideration be defined respectively by z_i^* and z_i^0 respectively.

Complete *a priori* assessment of $V(\mathbf{z})$ is conventionally separated into two phases, i.e. evaluation of the standardized partial value functions $v_i(z_i)$, i.e. scaled in some consistent manner (e.g. such that $v_i(z_i^0) = 0$ and $v_i(z_i^*) = 100$ for each i), followed by assessment of the weight parameters w_i which are then a measure of the relative importance of an improvement in z_i (or "swing") from z_i^0 to z_i^*.

For purposes of interactive methods of assessment, however, the separation into two phases is not strictly necessary, and it is sufficient to convenient to simplify (2) as follows:

$$V(\mathbf{z}) = \sum_{i=1}^{m} u_i(z_i) \tag{3}$$

where we fix $u_i(z_i^0) = 0$, and set $u_i(z_i^*)$ proportional to relative importance (so that in effect $u_i(z_i) = w_i v_i(z_i)$ for each i. An overall scaling of the function must also be specified, for example:

$$V(\mathbf{z}^*) = \sum_{i=1}^{m} u_i(z_i^*) = 100 \qquad \text{(say).} \tag{4}$$

In the case of categorical scales, it is only necessary to associate the $u_i(z_i)$ with each of the defined categories, which we shall denote by $z_i^0, z_i^1, \ldots, z_i^{\nu_i}$ defining $\nu_i + 1$ categories (for convenience of consistency with the piecewise linear forms). For performance defined on cardinal scales, the functions $u_i(z_i)$ need to be defined for all outcomes on a specified range, and will typically be continuous. It is such continuous functions which we now approximate in piecewise linear form, as illustrated in Figure 1 for four segments (and which also introduces the parameters which are defined below).

A piecewise linear approximation to $u_i(z_i)$ based on ν_i segments (where $\nu_i = 4$ in Figure 1) is defined as follows:

- The breakpoints (points at which one linear segment meets the next), which are defined by $z_i^0, z_i^1, \ldots, z_i^{\nu_i}$ where $z_i^{\nu_i} = z_i^*$ by definition: It is assumed that the breakpoints are chosen by the user, based on the user's physical knowledge of the attribute, before the elicitation of values commences.

- The value increments across each segment, say $d_{i1}, d_{i2}, \ldots, d_{i\nu_i}$: These are the parameters which are to be progressively estimated.

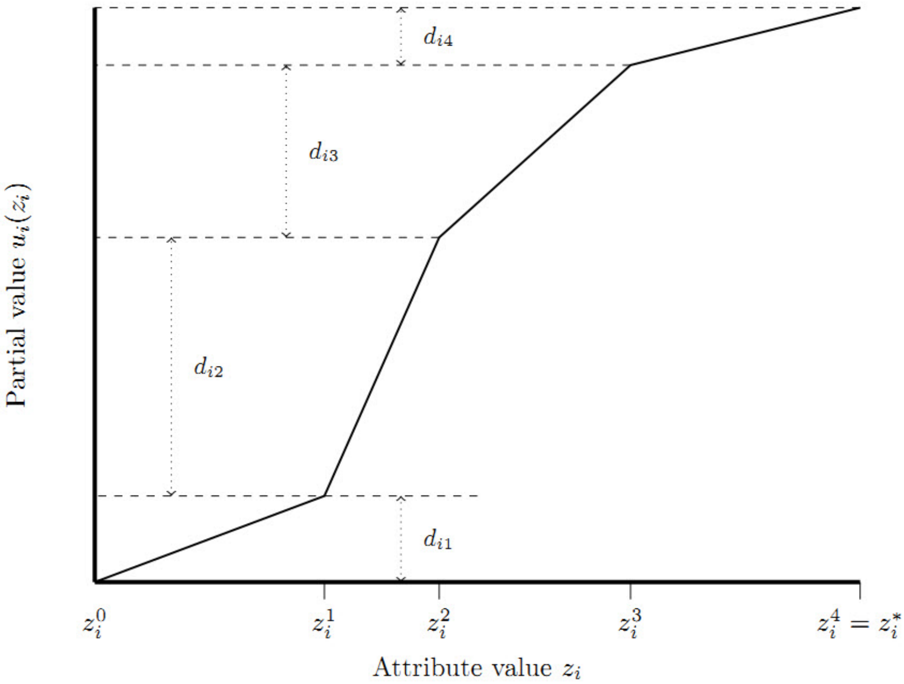

Figure 1: Illustration of piecewise linear approximation

By the above definitions, $u_i(z_i^r) = \sum_{k=1}^{r} d_{ik}$ at each of the break points or categories $r = 1, \ldots, \nu_i$. Values for $u_(z_i)$ at other values of continuous attributes z_i are linearly interpolated between the breakpoints at either side, i.e. for $z_i^{r-1} < z_i \leq z_i^r$ we have $u_i(z_i) \approx \sum_{k=1}^{r-1} d_{ik} + d_{ir}(z_i - z_i^{r-1})/(z_i^r - z_i^{r-1})$. The key point to note is that for a fixed set of breakpoints, the approximated $u_i(z_i)$ is linear in the parameters to be estimated (d_{ik}). The value structure for categorical scales is defined in the same way by the increments d_{ik}, except that now no interpolation is needed.

Finally, the scaling defined by (4) becomes a further linear constraint on the d_{ik} parameters given by $\sum_{i=1}^{m} \sum_{k=1}^{\nu_i} d_{ik} = 100$.

In the next section, we discuss how different types of preference information can be exploited to obtain estimates for the d_{ik}.

3 Forms of Imprecise or Incomplete Information

The basis of indirect assessment is to allow users to input preference information in a simple and easily understood manner. It is accepted that such information may be somewhat imprecise, but the principle is that many pieces of imprecise information can eventually establish a relatively much more precise preference model. We recognize at least the following types of inputs that users might be prepared to provide:

- *Interval estimates of relative values*, e.g. that the value of a stated improvement on one attribute is between 3 and 5 times that of another interval of improvement in the same or another attribute.
- *Semantic scales*, in which verbal descriptions are provided for the worth of specified value

gains, e.g. that a given level of improvement on one criterion is of little importance, important, very important, etc.

- *Ordinal value statements*, e.g. that a particular level of improvement on one criterion is more (or less) important than another.
- *Ordinal preference statements* to the effect that one (real or hypothetical alternative) is preferred to another.
- *Specification of functional shape:* In some instances, the user may be able to state that the marginal value function needs to have a particular shape, such as *concave* (decreasing rates of marginal gain per unit increase in the attribute value), *convex* (increasing rates of marginal gain), or perhaps even *sigmoidal* (early increases in rates of marginal gain, which later fall off after a threshold is passed).

In the next few paragraphs, we shall elaborate on each of these modes of stating values or preferences, and indicate the implication of each in terms of the piecewise linear additive value function model. Individually, each of these forms of inputs have been discussed by other authors (see references given below), but as commented above, we seek to provide a simple integrated framework within which users can make use of any desired combinations of such inputs, from which underlying value functions can be derived.

3.1 Interval estimates of relative values

This form of input is related to the ideas introduced by Salo and Hämäläinen [1992]. Consider potential increments, say from z_i^a to z_i^b on criterion i (where $z_i^a < z_i^b$) and from z_j^c to z_j^d on criterion j (where $z_j^c < z_j^d$). The points z_i^a, z_i^b, z_j^c and z_j^d need not correspond to any of the breakpoints. The user is asked to compare the relative value gains of these two increments in a ratio sense, but need only state an interval in which this ratio lies (i.e. not necessarily a precise value). A statement that the ratio lies between α and β, say, implies that:

$$\alpha \leq \frac{u_i(z_i^b) - u_i(z_i^a)}{u_j(z_j^d) - u_j(z_j^c)} \leq \beta$$

or more conveniently:

$$\alpha \left[u_j(z_j^d) - u_j(z_j^c)\right] \leq u_i(z_i^b) - u_i(z_i^a) \leq \beta \left[u_j(z_j^d) - u_j(z_j^c)\right] \tag{5}$$

which is a pair of linear inequalities. Replacement of the $u_i(z_i)$ terms by the piecewise approximations thus generates a pair of linear inequalities in the d_{ik} and d_{jk} sets of parameters. In other words, both inequalities in (5) can be represented as an expression of the form:

$$\sum_{i=1}^{m} \sum_{k=1}^{\nu_i} \theta_{ik} d_{ik} \geq 0. \tag{6}$$

In practice, such interval estimates may usually be restricted to two special cases, as follows.

Performance levels within one criterion, i.e. $i = j$. In this case the $u_i(z_i)$ terms may be replaced by the standardized partial values $v_i(z_i)$, since the weight parameters w_i will cancel. If we restrict attention to increments from the worst case levels (i.e. $z_i^a = z_i^c = z_i^0$), and set $z_i^d = z_i^*$ (the best available performance), then (using the conventional scaling) (5) simplifies to:

$$100\alpha \leq v_i(z_i^b) \leq 100\beta \tag{7}$$

which forms an envelope around the "true value" of the partial value function at the point z_i^b. We shall refer to the resulting band of uncertainty around the partial value function as a *value region* (as will be illustrated in the case study of Section 5).

Importance weights (trade-offs between criteria). This option requires $i \neq j$, and is most easily understood when $z_i^a = z_i^0$, $z_i^b = z_i^*$, $z_j^c = z_j^0$ and $z_j^d = z_j^*$, so that the two increments represent "swings" (for example, Section 5.4 of [Belton and Stewart, 2002]) on the two criteria. For this special set of choices:

$$\frac{u_i(z_i^*) - u_i(z_i^0)}{u_j(z_j^*) - u_j(z_j^0)} = \frac{w_i}{w_j}$$

so that the interval estimates imply $w_i \geq \alpha w_j$ and $w_i \leq \beta w_j$.

3.2 Use of semantic scales

This form of input has been discussed in Bana e Costa and Vansnick [1999] and Bana e Costa et al. [2005], and involves similar questions to those regarding relative values as described above. Now, instead of comparing two increments with each other, a single increment is placed into one of a number of importance classes, which may be described in terms of semantic labels such as (for example): *weakly important*, *important*, *strongly important*, and *extremely important*. Let us denote the ordered categories by $C_1, C_2, \ldots, C_S$, say, ordered from least to most important. Suppose then that the importance of the gain from z_i^a to z_i^b on attribute i is classified into category C_s, while the importance of the gain from z_j^c to z_j^d on attribute j is classified into category C_t, where $t > s$. Such a situation implies that $u_i(z_i^b) - u_i(z_i^a) < u_j(z_j^d) - u_j(z_j^c)$. Once again, this expression may be translated into a linear inequality in the d_{ik} and d_{jk} as given by (6), except that the inequality is now required to be strict. One such inequality will be generated for every pair of value increments which are classified into different importance categories.

A further set of inequalities could in principle be generated by arguing that the differences between the magnitudes of $u_i(z_i^b) - u_i(z_i^a)$ and of $u_j(z_j^d) - u_j(z_j^c)$ (i.e. the magnitude of $u_i(z_i^b) - u_i(z_i^a) - u_j(z_j^d) + u_j(z_j^c)$) should be larger for those pairs of increments which have larger distinctions between categories. At the present time, however, such considerations have not been built into the DEFINITE software.

3.3 Ordinal value statements

Such statements may be viewed as a simplification of the previous two types of preference information. Instead of providing numerical bounds on the ratios of value differences or providing categorical classifications, the user may simply state that the importance of the increment from z_i^a to z_i^b is greater than the importance of the gain from z_j^c to z_j^d. Such ordinal statements lead directly to the same form of inequality as identified in the use of semantic scales.

3.4 Ordinal preference statements

In this form of interaction, the user states directly which of two alternative outcomes, say vectors $\mathbf{z}(a)$ and $\mathbf{z}(b)$, is preferred. If the former is preferred, this implies that $\sum_{i=1}^{m}[u_i(z_i(a)) - u_i(z_i(b))] > 0$ which again generates a (strict) linear inequality of the form given by (6).

The alternatives selected for such pairwise comparisons may arise in two ways:

- Comparisons may be made between elements of a subset of the real alternatives under consideration, which is the basic approach taken in the UTA method [Jacquet-Lagrèze, 1990, Jacquet-Lagrèze and Siskos, 2001].

- Alternatives for comparison may be hypothetically constructed, typically such that the alternatives being compared differ on two attributes only, allowing in effect a direct assessment of tradeoffs. For the implementation in DEFINITE, three levels of performance are selected for each attribute, namely z_i^0, z_i^* and an intermediate value, say z_i^M, chosen such $u_i(z_i^M) \approx 0.5u_i(z_i^*)$. Pairs of attributes are taken in turn. For each such pair, say attributes i and j, the user is asked to rank order the nine pairs of performance levels, i.e. (z_i^0, z_j^0), (z_i^0, z_j^M), $\ldots (z_i^*, z_j^*)$, assuming that the alternatives have identical performance in terms of the remaining $m-2$ attributes. By definition, (z_i^0, z_j^0) must be the lowest ranking and (z_i^*, z_j^*) the top ranking hypothetical alternatives, but the other rank positions provide the required ordinal preference statements. This approach is sometimes termed **conjoint scaling**.

3.5 Shape of the marginal value function

A concave value function will require that $d_{ik} < d_{i,k-1}$ for $k = 2, \ldots, \nu_i$. These inequalities are reversed for a convex value function. When $\nu_i \leq 4$, a sigmoidal value function is equally easily described by two inequalities, namely: $d_{i2} > d_{i,1}$ and $d_{i\nu_i} < d_{i,\nu_i-1}$.

4 Model Fitting and Use

In the previous Section, we have described a various forms of value judgement that could be made by experts. Choice of which to use would depend both on the personality of the users and on the problem context, and in any one instance a combination of different inputs may be employed. We now describe how a best fit value function model and associated rank orders of alternatives can be derived from any such combination of inputs.

It has been noted that a variety of interactions with the user all generate linear inequalities which can be expressed in the form

$$\sum_{i=1}^{m} \sum_{k=1}^{\nu_i} \theta_{ikp} d_{ik} \geq (\text{or } >)\, 0 \tag{8}$$

for $p = 1, 2, \ldots, P$, where P is the total number of inequalities generated by the different user inputs. For the moment we shall not distinguish between the $\geq$ and strict $>$ cases, but we shall return to this point later.

The decision aiding approach derived from such imprecise or incomplete information is to extract the available information concerning the set of parameter values d_{ik} ($i = 1, 2, \ldots, m$; $k = 1, 2, \ldots, \nu_i$), given the set of inequalities in (8). The aims are to find the best possible fit and/or a range of plausible fits to (8), amongst parameter values satisfying the scaling constraint (4). The resulting value function (or range of plausible value functions) can then be applied to the actual set of alternatives under consideration. This may lead either to acceptance of the resulting optimal solution, or to a recognition that the information is still insufficiently complete or precise, so that further preference information will need to be entered.

We recognize that two situations may arise:

1. The parameter values may be *under-determined* by (8) in the sense that multiple feasible solutions exist (within a convex space);

2. The parameter values may be *over-determined* in the sense no feasible solution exists.

Our approach is, however, to adopt the same model fitting procedure for both situations (and in fact for the boundary case of one unique solution, although this would be so rare an event as to be uninteresting), so that we do not have to distinguish between the cases. The principle is as follows:

1. In the case of under-determination, we seek what Zionts and Wallenius [1983] termed the "middlemost" solution, i.e. the solution maximizing the minimum slack across all inequalities in (8). Note that this approach now eliminates the distinction between the $\geq$ and $>$ constraints. The range of feasible solutions is also minimized in the manner described below.

2. In the case of over-determination, we seek the solution which minimizes the maximum constraint violation in (8).

In order the apply the above principles, it is necessary to scale the inequalities in (8) so that deviations are directly comparable. For this purpose, the θ_{ikp} are scaled for each p to unit Euclidean norm, i.e. such that $\sum_{i=1}^{m} \sum_{k=1}^{\nu_i} \theta_{ikp}^2 = 1$. It would in principle be possible to rescale the normalized θ_{ikp} by multiplying through by a positive constant in order to force deviations on certain constraints to have greater importance than others, but this refinement has not at this stage been implemented.

Best-fit value model. Define $\delta_p = \sum_{i=1}^{m} \sum_{k=1}^{\nu_i} \theta_{ikp} d_{ik}$, which we shall denote as the feasibility gap. A negative value of δ_p indicates that the current solution does not satisfy the p-th constraint, while a positive value indicates the extent to which the inequality is satisfied. Further define $\Delta = \min_{p=1,2,\ldots,P} \delta_p$, which is negative if the current solution (set of d_{ik} values) violates one or more constraints in (8), and is positive when all inequalities are satisfied. The required best model fit according to the above principles is found by solving the following linear programming problem:

$$\left.\begin{array}{lcll} \text{Maximize} & & \Delta + \epsilon \sum_{p=1}^{P} \delta_p & \\ \text{Subject to:} & & & \\ \sum_{i=1}^{m} \sum_{k=1}^{\nu_i} \theta_{ikp} d_{ik} & = & \delta_p & \text{for } p = 1, \ldots, P \\ \Delta & \leq & \delta_p & \text{for } p = 1, \ldots, P \\ d_{ik} & \geq & 0 & \forall i, k \end{array}\right\} \quad (9)$$

where $\epsilon > 0$ is a small positive constant. The introduction of the $\epsilon \sum_{p=1}^{P} \delta_p$ term ensures that the solution is Pareto optimal in the sense that there is no distinctly different feasible solution to (9) with feasibility gaps $\delta'_1, \delta'_2, \ldots, \delta'_P$ such that $\delta'_p \geq \delta_p$ for $p = 1, 2, \ldots, P$.

Characterizing the range of plausible solutions. If the optimal solution to (9) has $\Delta < 0$, there does not exist a model satisfying all constraints, but the solution represents a "best fit" model, which in most practical situations would be unique. The resulting model will then imply a unique ordering of alternatives which are most consistent with the stated preferences of the user.

When the solution has $\Delta > 0$, however, there will generally exist a range of models satisfying the inequalities given by (8). Each of these feasible models may result in a different ordering of the alternatives, and all such orderings would be consistent with the preference information. A first step is to establish for each alternative a whether it is potentially optimal for any value function consistent with the provided preference information. This can be done by solving the following LP (still making use of the piecewise linear approximation to $V(\mathbf{z})$):

$$\left.\begin{array}{lcll}
\text{Maximize} & & E + \epsilon \sum_{b \neq a} e_b & \\
\text{Subject to:} & & & \\
V(\mathbf{z}^a) - V(\mathbf{z}^b) & \geq & e_b & \forall b \neq a \\
E & \leq & e_b & \forall b \neq a \\
\sum_{i=1}^{m} \sum_{k=1}^{\nu_i} \theta_{ikp} d_{ik} & \geq & 0 & \text{for } p = 1, \ldots, P \\
d_{ik} & \geq & 0 & \forall i, k
\end{array}\right\} \quad (10)$$

The variable e_b measures the extent to which alternative a is preferred to b according to the model defined by the d_{ik}, while $E = \min_{b \neq a} e_b$. A negative value for E implies that a can never be first ranked in any model consistent with the preference inputs, while a positive value indicates that a is potentially optimal.

Solving (10) for each alternative in turn thus provides up to n (=number of alternatives) different feasible models, each of which implies a rank ordering of the alternatives (although some may be duplicated). A table of these resultant rank orders thus provides a representation of possible rank orders consistent with the preference information provided, which provides the decision maker with a simplified picture of potentially optimal selections.

This approach has been implemented in the DEFINITE software mentioned below. Extensions that could be considered would include Monte Carlo generation of d_{ij} values from the feasible set, somewhat allied to the SMAA methods summarized for example in Lahdelma et al. [1998], and a ranking approach developed in Kadzinski et al. [2013] which can extend (10) to identify not merely whether a is potentially optimal, but also the highest ranking position that a can take on. This latter extension is achieved by solving the following model:

$$\left.\begin{array}{lcll}
\text{Minimize} & & 1 + \sum_{a \neq b} \lambda_{ab} & \\
\text{Subject to:} & & & \\
V(\mathbf{z}^a) - V(\mathbf{z}^b) + 100\lambda_{ab} & \geq & 0 & \forall b \neq a \\
\sum_{i=1}^{m} \sum_{k=1}^{\nu_i} \theta_{ikp} d_{ik} & \geq & 0 & \text{for } p = 1, \ldots, P \\
d_{ik} & \geq & 0 & \forall i, k \\
\lambda_{ab} & \text{binary} & & \forall a \neq b
\end{array}\right\} \quad (11)$$

The procedures described above are (with the exception of the last-mentioned potential extensions) integrated into the DEFINITE software [Janssen and van Herwijnen, 2011] as an optional procedure called *EValue*. EValue can be used as an alternative to direct specification of value functions by the user. The user provides inputs such as the shape of the value function, value regions, ordinal weights, and rankings of different combinations of two attribute scores for the conjoint scaling approach). The resultant best fit value functions and weights can further be used directly in one of the other multicriteria methods available in DEFINITE.

We now illustrate the above procedures by means of a case study.

5 Case Study: Soil Sanitation Decision Problem

The assessment procedure described in the previous sections is now applied to a soil sanitation problem in the Netherlands. The sanitation problem fits into the context of a relatively technical decision problem in which expert assessment is required. The underlying mechanisms relating pollution to effects will generally be well understood by such experts, but in most cases are not adequately modelled in a formal sense. Expert judgements are therefore important [Janssen,

Table 1: Effectiveness (residual concentrations) of 11 clean-up alternatives

Pollutant	Units	Clean-up Alternatives										
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
Zinc	mg/kg	1685	0	1800	1876	1876	235	235	131	1876	1866	1743
Cadmium	mg/kg	19	0	20	17	14	1	1	1	21	20	17
Mineral oil	mg/kg	135	0	521	37	12	75	104	45	297	297	178

2001]. As similar assessments would ultimately be required for a number of sites, there was value in establishing appropriate value functions that can be reused.

The polluted site discussed here is a former industrial area, with high concentrations of zinc, mineral oil and cadmium. Because legal pollution standards are exceeded the site needs to be cleaned up. A number of alternatives are available to clean-up the site, the effectiveness of each of which is summarized in Table 1. The most important attributes according to which these alternatives are to be evaluated are the residual concentrations of zinc, cadmium and mineral oil. Other important issues include costs of clean-up and the disamenity to people living near the site during the clean-up, but for this paper we will focus on the assessment of the value functions for the three pollutants, as these are where the expert judgements are applied.

Performance of the alternatives for these three attributes is presented in Table 1. It can be seen that only complete removal of the soil (alternative A2) results in totally completely clean soil. All others result in residual concentrations of the pollutants that are above 0 but within legal standards.

After a few test sessions, assessment sessions were organized with five experts. These experts were all directly involved with decision-making on soil sanitation problems and could be seen to reflect the state of knowledge in this field. The experts were provided with background information on the site and were also given an introduction to the concepts used in the assessment procedure. Each session took between two and three hours. This resulted in five sets of value functions and weights.

In this section we work systematically through the steps of the assessment procedure. For purposes of presentation, we report on the process for one expert only. In the conclusions we will comment on the differences that arose between the experts.

The assessment procedure involves the following steps:

1. Assessment of value regions (cf. (7))
2. Assessment of weights
3. Indirect assessment (conjoint scaling)
4. Application of the LP algorithm for estimation of the value functions and weights.
5. Evaluation of results (value functions, weights and ranking)

If the expert is satisfied at the evaluation step, then the process stops. Otherwise it is repeated from step 1. An expert may go through these steps several times, as typically a number of rounds were needed before a satisfactory result was achieved. In our case study it took between three and five rounds depending on the expert. As indicated above, the assessment procedure is integrated in the software package DEFINITE [Janssen and van Herwijnen, 2011], which has the advantage that the results can be used directly in the evaluation. Figure 2 shows the interface of the assessment procedure. The menu on top indicates the steps: value regions, weights, indirect assessment and computation. The tabular portion of the interface shows for each of the three

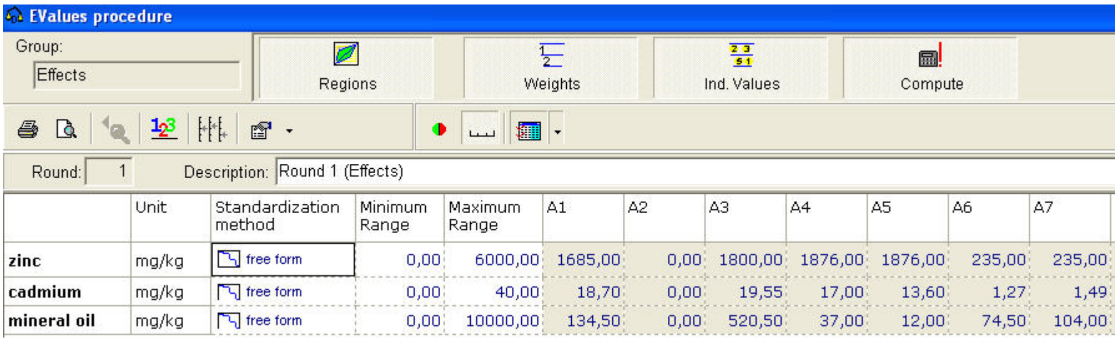

Figure 2: EVALUE interface in DEFINITE

attributes a possible selection of a shape for the marginal value function and the range of values specified.

In the following paragraphs, we describe details of the interactions at each step of the procedure.

Step 1: Assessment of value region. In this first step the expert is asked to specify a value region for each attribute, as described for the first of the two inputs in subsection 3.1. A value region represents the range of uncertainty in the marginal value function for each attribute. The region can be wide if the expert is uncertain about the position of the value function or narrow if the position is relatively clear. If the expert has no idea about the position the value region covers the entire figure.

Figure 3 shows the value regions as specified by expert 1 for the cadmium pollutant attribute, measured in mg/kg. The shape of the region is determined at five levels (breakpoints) of the concentration of cadmium. The most important concentration is the legal standard that determines if a soil needs to be cleaned up, which for cadmium is 20mg/kg. The upper limit of the range is set at twice this value, i.e. at 40mg/kg. Since total clean-up is technically feasible the lower end of the range is set at 0 mg/kg. The intermediate concentration levels of 1 and 5 mg/kg are linked to long and short-term policy goals for cadmium. The line showing the values of the alternatives sits exactly in the middle of the value region. The eleven alternatives are shown for reference purposes.

Step 2: Assessment of weights. The expert now assesses the attribute weights in the sense described in subsection 3.1, but using only ordinal preferences (subsection 3.3). To do this the maximum and minimum pollution levels before and after clean up were presented to the experts. Next they were asked which pollutant they would reduce from maximum to minimum first, then which second, etc. Although this proved difficult for the experts, as they indicated that it was not possible to clean up one pollutant without cleaning up the other, expert 1 nevertheless indicated that getting rid of Cadmium was his first priority, but that he was indifferent about the order of Zinc and Mineral oil, i.e. $w_{\text{Cadmium}} > w_{\text{Zinc}} \sim w_{\text{Mineral oil}}$.

Step 3: Conjoint scaling. In this step the experts are asked to order hypothetical alternatives constructed from variations in two pollutants at the time (as described as the second approach in subsection 3.4). This proved to be a much easier task for the expert as it conforms more to their usual practice. The pollutants are compared in three pairs: (1) Cadmium-Zinc, (2) Cadmium-Mineral Oil, and finally (3) Zinc-Mineral Oil. For illustration, the comparisons of hypothetical alternatives involving three levels each of Cadmium

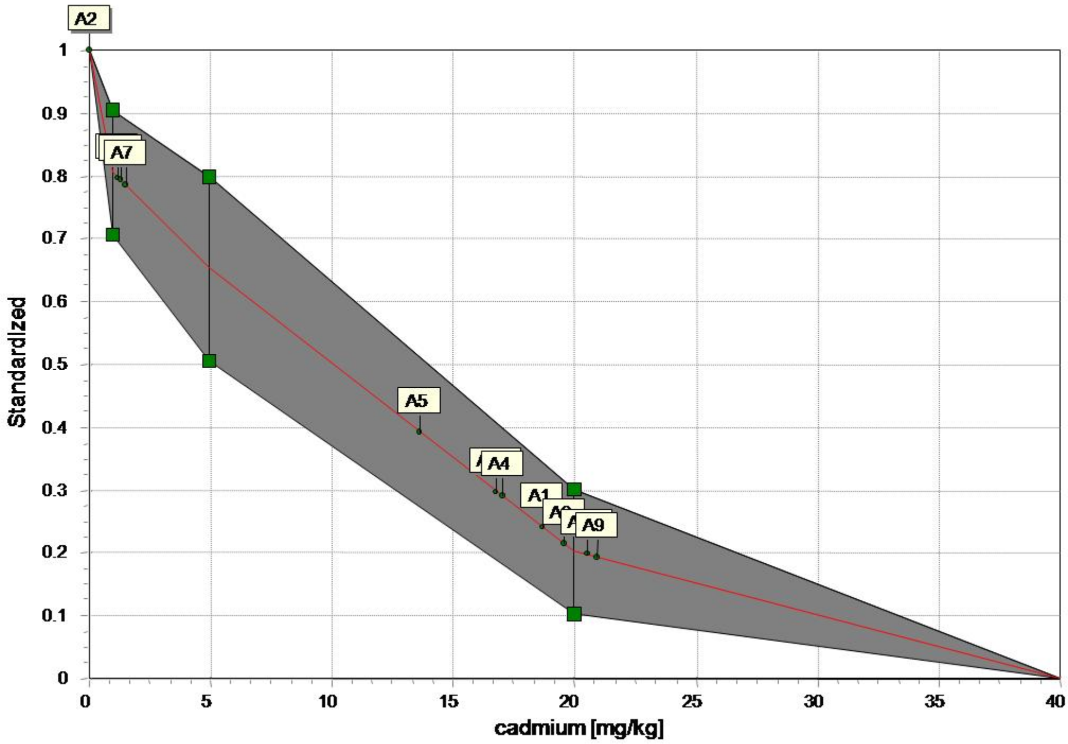

Figure 3: Value region for Cadmium

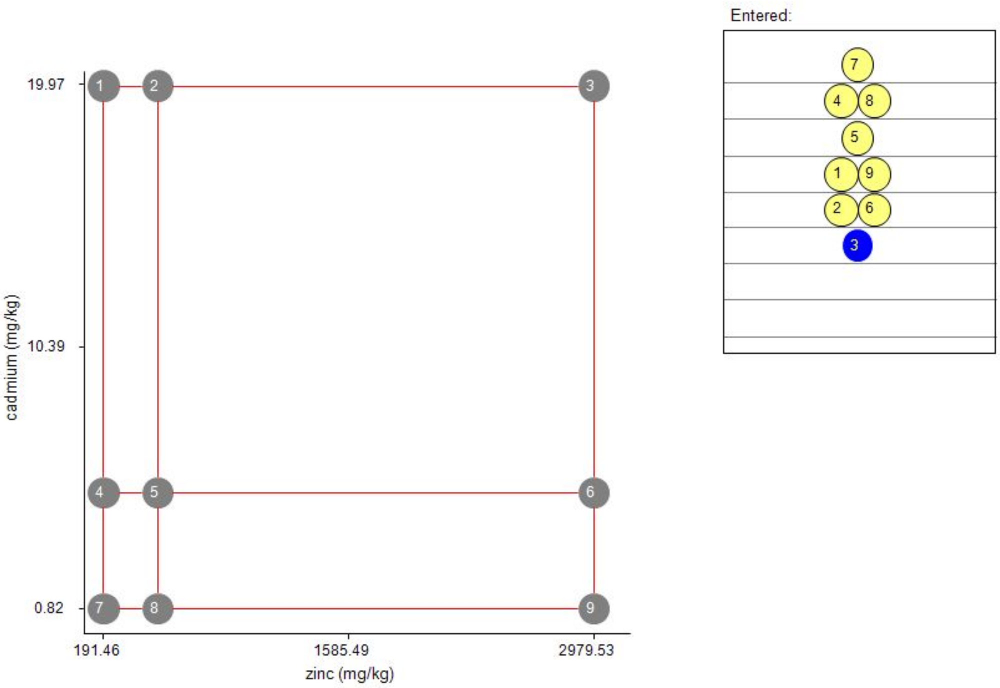

Figure 4: Input to conjoint scaling for Cadmium and Zinc

and Zinc contamination (giving nine hypothetical sites) is shown in Figure 4. Each of these hypothetical sites are assumed to exhibit the same level of mineral oil concentration.

The expert is asked to rank the nine hypothetical sites according to their environmental quality. Site 7 is obviously the best site with the lowest concentration for both pollutants. The first non-trivial choice which needs to be made by the expert is thus between alternatives 4 and 8 (or he could indicate that they are about equally ranked). The routine leads the expert systematically through all relevant pairwise judgements that need to be made. The ranking provided by Expert 1 is displayed to the right of Figure 4, showing that he considered Sites 4 and 8 of equal quality, followed by 5, etc. Note that due to the monotonicity of the value functions the sites are *a priori* partially ordered. Experts proved to be relatively confident about the reliability of these orderings.

Step 4: Estimation of value functions and weights. At this point, the information contained in the value regions, weights and conjoint scaling inputs are combined as described in Section 4, in order to generate marginal value functions and weights that are as consistent as possible with the information.

As illustration, Figure 5 shows the value region as provided by the expert in the case of Cadmium pollution. The bold line contained in the region is the best fit marginal value function for this attribute, as determined by the linear programming model of Section 4.

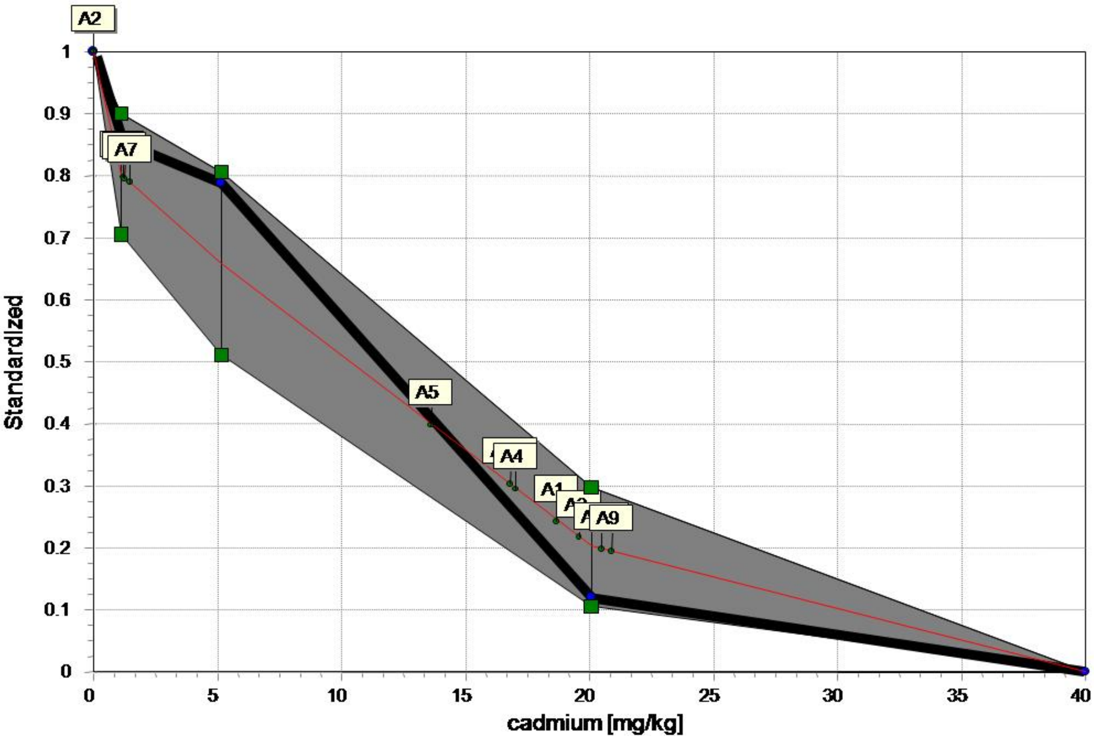

Figure 5: User defined value regions and computed marginal value function for Cadmium

The same LP model produced best-fit weights given by $w_{\text{Cadmium}} = 0.5$, $w_{\text{Mineral oil}} = 0.39$ and $w_{\text{Zinc}} = 0.1$, which preserves the strict ordering provided by the expert (i.e. w_{Cadmium} greater than the other two weights), but provides a definite separation of the other two weights (based essentially on results from the conjoint scaling step).

It is informative also to compare the ordering provided by the expert for the nine hypothetical Cadmium-Zinc combinations, with the ordering implied by the best fit model from the LP. These two rank orders are displayed in Figure 6, with the ordering provided by the expert on the left, and the ordering implied by the LP results on the right. There is some discrepancy between the rank positions of hypothetical alternatives 4, 5 and 9 (although the computed values, also displayed, are quite close). The (minor) discrepancies reveal the existence of some conflict between the different types of input provided by the expert.

Step 5: Evaluation of results. In this step the expert evaluates value functions, weights and orderings computed. If the expert is satisfied with the computed value model the session stops. If not the expert can reassess the value regions, weights or orderings in the light of the results obtained. In this example, however, the expert proved to be satisfied with the result. Once he is satisfied, the resulting overall value function can be used to rank the real clean-up alternatives. This ranking is shown in the second column of Table 2.

As there were inconsistencies in the inputs provided by Expert 1, the linear programming model generates a unique best-fit model and associated ranking of alternatives. Such inconsistencies typically arise when preference information from the expert is over-specified. As described in Section 4, the model also deals with estimation when there are no inconsistencies in the user inputs (typically when the preference information provided is relatively

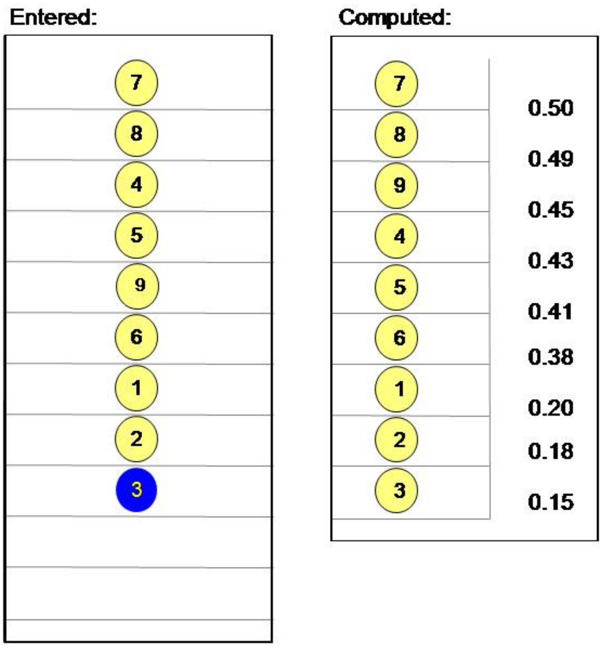

Figure 6: User specified and computed orderings of alternatives

sparse). In order to illustrate the model outputs in such an underspecified context, the EVALUES procedure was re-applied with the omission of the conjoint scaling inputs for the Cadmium-Mineral oil pairs. The remaining inputs exhibit no inconsistencies, and the LP generates four sets of rankings of the alternatives that are consistent with these inputs, as shown in the last four columns of Table 2. Note, however, that these four rank orders place the top five alternatives in the same order (which is in fact also consistent with the best fit to the complete information). Thus the expert might well be satisfied to accept the model at this point.

It is perhaps not surprising that none of the four rankings based on reduced information are precisely equivalent to the best fit ranking based on the full information. By definition, the full information contains inputs which are inconsistent with that retained in the reduced set of inputs. As the LP model takes these inputs into account, it is to be expected that some shift in rank orders will be generated.

6 Conclusions

The LP model for integration of different types of preference information was easily implemented with standard LP software. The user interface was satisfactory to the users (experts), who were happy to participate in the process. The rank ordering of alternatives generated by the estimated value function model(s) proved broadly satisfactory to the experts. It is important to brief experts well about the exact meaning of the questions asked and the concepts used, but no serious barriers to understanding were experienced. The availability of multiple rounds of assessment to allow experimentation and of different modes for providing preference information make it possible to achieve detailed results without a need to ask for unnecessarily detailed input from the experts.

Table 2: Computed rankings of alternatives

Rank Position	Best Fit Ranking	Consistent rankings with incomplete information Set 1	Set 2	Set 3	Set 4
1	A2	A2	A2	A2	A2
2	A8	A8	A8	A8	A8
3	A6	A6	A6	A6	A6
4	A7	A7	A7	A7	A7
5	A5	A5	A5	A5	A5
6	A4	A11	A11	A4	A11
7	A11	A4	A4	A11	A4
8	A1	A1	A1	A1	A1
9	A3	A10	A3	A10	A10
10	A10	A3	A10	A3	A9
11	A9	A9	A9	A9	A3

Each assessments session took about two to three hours including the introduction, which is considerably less than the time typically required for a full traditional assessment as described in Belton and Stewart [2002], Chapter 5. It is to be expected that the assessment will be used mainly for major projects or for projects as in our clean up example that will be repeated several times under comparable circumstances.

A critical element in the procedure is the correct interpretation of the questions asked. Different interpretations may lead to very different answers. In our case study, the input from one of the experts was found to differ substantially from that of the other four. Discussions with this expert revealed that he had used a different reference model. Rather than the environmental quality *per se*, this expert judged the quality of the clean-up job in a technical or professional sense. He stated, for example, that anyone who leaves more than 1000mg/kg mineral oil in the soil has done an incompetent job, deserving a value of 0. This experience demonstrates how important it is that all parties share the same frame of reference.

Although the initial value regions specified by the different experts differed substantially, the final resulting models were remarkably similar. Future research might nevertheless address how models derived in this was from different experts might be combined into a single overall model. An important question would be whether separate LP model fitting exercises should be carried for each expert, with aggregation of the final models, or whether the inputs of each expert should be included in a single LP structure to provide a model which is most consistent with all inputs. The structure of Section 4 could easily be adapted to achieve this end.

Acknowledgements

The contribution of Theodor Stewart was partially supported by a grant from the National Research Foundation of South Africa.

References

C. A. Bana e Costa and J. C. Vansnick. The MACBETH approach: basic ideas, software and an application. In N. Meskens and M. Roubens, editors, *Advances in Decision Analysis*, pages 131–157. Kluwer Academic Publishers, Dordrecht, 1999.

C. A. Bana e Costa, J. De Corte, and J. Vansnick. On the mathematical foundations of MACBETH. In Figueira et al. [2005a], chapter 10, pages 409–442.

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

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

J. Figueira, S. Greco, and M. Ehrgott, editors. *Multiple Criteria Decision Analysis – State of the Art Annotated Surveys*. International Series in Operations Research and Management Science Volume 76. Springer, New York, 2005a.

J. Figueira, V. Mousseau, and B. Roy. ELECTRE methods. In Figueira et al. [2005a], chapter 4, pages 133–162.

J. R. Figueira, S. Greco, and R. Słowiński. Building a set of additive value functions representing a reference preorder and intensities of preference: GRIP method. *European Journal of Operational Research*, 195:460–486, 2009.

T. Gal, T. J. Stewart, and T. Hanne, editors. *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications*. Kluwer Academic Publishers, Boston, 1999.

S. Greco, V. Mousseau, and R. Słowiński. Ordinal regression revisied: multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research*, 191:416–436, 2008.

E. Jacquet-Lagrèze. Interactive assessment of preferences using holistic judgements: the PREFCALC system. In C. A. Bana e Costa, editor, *Readings in Multiple Criteria Decision Aid*, pages 335–350. Springer-Verlag, Berlin, 1990.

E. Jacquet-Lagrèze and Y. Siskos. Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research*, 130:233–245, 2001.

R. Janssen. On the use of multi-criteria analysis in environmental impact assessment in th eNetherlands. *Journal of Multi-Criteria Decision Analysis*, 10:101–109, 2001.

R. Janssen and M. van Herwijnen. *DEFINITE 3.1. A system to support decisions on a finite set of alternatives*. Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, 2011. Software package and user manual.

M. Kadzinski, S. Greco, and R. Slowinski. RUTA: A framework for assessing and selecting additive value functions on the basis of rank related requirements. *Omega*, 41:735–751, 2013. doi: 10.1016/j.omega.2012.10.002.

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

S. M. Lee and D. L. Olson. Goal programming. In Gal et al. [1999], chapter 8.

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

A. A. Salo and R. P. Hämäläinen. Preference ratios in multiattribute evaluation (PRIME) – elicitation and decision procedures under incomplete information. *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans*, 31:533–545, 2001.

T. J. Stewart. Use of piecewise linear value functions in interactive multicriteria decision support: A monte carlo study. *Management Science*, 39:1369–1381, 1993.

T. J. Stewart. Robustness of additive value function methods in MCDM. *Journal of Multi-Criteria Decision Analysis*, 5:301–309, 1996.

T. J. Stewart. Concepts of interactive programming. In Gal et al. [1999], chapter 10.

P. Vincke. Outranking approach. In Gal et al. [1999], chapter 11.

A. P. Wierzbicki. Reference point approaches. In Gal et al. [1999], chapter 9.

S. Zionts and J. Wallenius. An interactive multiple objective linear programming method for a class of underlying nonlinear utility functions. *Management Science*, 29:519–529, 1983.