A systematic approach to build evaluation indices for environmental decision making with active public involvement

Andrea Nardini
Centro EULA and Faculty of Engineering
Universidad de Concepción, Casilla 156-C, Concepción, Chile
(now at Ambiente Italia, via C. Poerio 39, 20129 Milano, Italy; Tel.+39-2-277441; Fax.+39-2-27744222: e-mail: nardini@galactica.it)

Abstract

Decision making in the environmental arena is always a Multicriteria problem. The relevant information should therefore be represented holistically through an *evaluation matrix*. This task calls for using evaluation indices. Furthermore, the problem is characterised by conflicts of interest and, consequently, by the need for public involvement and negotiation. Hence, such indices should be easily understandable, and should make evident what are the impacts on the interests at stake. The indices proposed in the literature show, however, a number of structural weaknesses when they are used to face this task. More in general, there is still a gap between the essentially technical phase of Impact Assessment and the actual evaluation phase, where ethical and moral value judgements about the relative importance of the different impacts are involved and should be made explicit. With this in mind, a set of 'requirements' that evaluation indices should fulfil in order to support active public involvement in the decision making process are proposed and discussed in relation to a case study.

(Keywords: decision-making; evaluation indices; Multicriteria Analysis; Environmental Impact Assessment)

INTRODUCTION

Environmental decision making requires to trade-off between different, incommensurable effects. It also involves conflicts of interest among different interest groups of people (*groups* in what follows), and therefore negotiation aimed at conflict resolution is required. Decision making is a process: the terms of the problem are gradually refined, additional information is collected, processed and analysed, people are involved, and so on. Nevertheless, there is always one (or more) step(s) in which an evaluation is required. The evaluation-negotiation task cannot be correctly performed without a clear, summarised and holistic view of the problem at hand. Multicriteria Analysis (MCA) provides a framework which can greatly help in this respect [Nijkamp and Beinat, 1997]. Accordingly, the information should be organised in an *evaluation matrix* (also called *effect table*), which is a two-entry table where each column corresponds to one specific project alternative (a*lternative* in what follows), and each row corresponds to one of the relevant effects (positive or negative). The information contained in each cell is a representation of the effect assessed. This representation should be easily understandable by the people involved. Moreover, it should make evident the relative performance of the *alternatives*, in terms of impact on the interests at stake. This goal can hardly be reached if not through suitable *evaluation indices*.

The indices topic may seem an old-fashioned one. Indeed, a great deal of research has been carried out on the subject of environmental indices since the '70s (O'Connor [1972], Dee et al. [1972], Liu [1975], Inhaber [1976], Ott [1978]), or even earlier [Horton, 1965]. This theme deserves, however, a renewed interest because at least of three reasons:

i) The analytical tools which support Impact Assessment have been developed to a degree which allows analysts deal with, and make predictions of, extensive and qualitative information; however, a synthesis is required in order to present such information in a concise and consistent form;

ii) The need for a holistic approach to evaluation based on a Multicriteria framework (which in turn requires evaluation indices) is being increasingly recognised [Nijkamp and Beinat, 1997; Gregory et al., 1992; Renn et al., 1993];

iii) The analysts have, in a certain sense, reduced their aims, by leaving a greater scope to the sectorial experts in the form of ‘descriptive cause-effect models’ (formalised through evaluation indices [Beinat, 1995]); this implies to recognise the need for a separate and explicit elicitation of subjective preferences to carry out the evaluation-negotiation task.

Many advocate that adopting evaluation indices rises the risk of loss of information, or mis-representation [see for instance Andrews, 1974; Eisel and Gaudette, 1974; Elliott, 1981]. Nevertheless, as it will be apparent then, many of these problems arise because indices are not designed expressly to deal with the evaluation-negotiation problem outlined above; they are rather diagnostic or classifying tools, and are not sufficiently structured.

This paper is an attempt to identify the requirements that any index should fulfil to be effective in a decision making process in which the public is actively involved and conflict issues explicitly addressed. It is intended to advance the way in which indices can be used in impact assessment and evaluation, and to resolve some of the conceptual difficulties surrounding the way in which MCA is applied.

The reader would probably expect now an analysis of indices from the literature to identify common weaknesses and then infer some guidelines. This is actually the natural process undertaken, however limited space does not allow to reproduce it here.

I therefore adopt a somehow axiomatic approach: let me first state the requirements that, once fulfilled, guarantee that an index is suitable in the sense specified above.

While discussing such requirements the underlying motivations will be apparent.

Afterwards, an example case-study is presented that should clarify the concepts. Finally, a brief review of some classic indices from the literature will be undertaken to show the relevancy of the requirements proposed.

Although this is unlikely to be a classical structure for a paper, it is one that complies with space limits and, at a time, provides the ideas I am willing to communicate in the most structured form I found. Further motivations and details for the framework proposed here are presented in Nardini [1997a,b; 1998].

REQUIREMENTS THAT EVALUATION INDICES SHOULD FULFIL

A - Impacts and indices. Only the effects to which a judgement of ‘good’ or ‘bad’ can be directly assigned are to be considered in the evaluation of a set of project *alternatives*. These are ‘fundamental effects’ in the sense that someone's satisfaction (*satisfaction* in the following) is being modified. Let me denote such type of modification an *impact*. There are, of course, other effects (impacts) but these are just intermediate relationships among the causes and the fundamental effects, and do not have to be considered in the final evaluation because no value judgement can be associated with them. There should be, then, an index for each positive or negative *impact* identified as relevant. They should be, however, as few as possible.

B - Index and *satisfaction*. Each index should be a measure of the (relative) *satisfaction* that a given *group* associates with a given *impact*. There are not ‘objective indices’ because each index necessarily contains the subjective value judgements of the corresponding *group* (through its representative, Delphi sessions, or other options). Hence, the index is always a Value Function (or Utility Function when uncertainty is explicitly considered) in the sense specified by Keeney and Raiffa [1976] (see also the Appendix), i.e. it is a mathematical representation of human preferences. As a consequence, any index is positively oriented, i.e. the larger its numerical value, the higher is the *satisfaction*.

For each *impact* there is a set of attributes (and corresponding indicators) which are relevant for the determination of the associated level of *satisfaction*. These are called, in the Keeney and Raiffa [1976] terminology, *natural* attributes (possibly ‘constructed’). For instance, *natural* (constructed) attributes for the *satisfaction* associated with the impact on ‘human relationships’, could be: the degree of brotherhood, the frequency with which people meet at private houses, the likelihood of robberies, the frequency of social events, etc.

It is however quite common practice to adopt instead *proxy* attributes, i.e. attributes which lie at an upper level in the cause-effect network. For example, most frequently used attributes in environmental problems are concentrations of the air pollutants, like NOx, SO2, dusts, etc. There is a very good reason to use *proxy* attributes: the relationship (the ‘model’) between causes (the *proxy* attributes) and effects (the *natural* attributes describing, for instance, health conditions) is very often unknown, because empirical evidence or its satisfactory interpretation (a mathematical model, or a clear verbal description) is lacking. In all these cases, the only piece of factual (objective) information available is that described by the *proxy* attributes.

Nevertheless, people do not understand technical information so to be able to associate value judgement with it. Then experts can and should be called to implement the 'model' (an example can be found in Keeney et al. [1984]). This 'expert based model' (which is in itself a Value Function) should, however, just be an assessment tools, but it should not cover the whole evaluation index. In this latter case, in fact, the expert would have to represent both the 'model' and, implicitly, the subjective value judgements of the *group*. This is often difficult and even questionable because anytime something is implicit, there is a loss of transparency. The result is that the index does not represent *satisfaction*, but mixes the two types of subjective judgements in an unclear fashion.

The use of an 'expert based model' implies a drawback. Even if there may be agreement about the factual information at the level of the *proxy* attributes, the agreement is likely to reduce at the level of the *natural* attributes, because subjective judgements of the expert, his/her personal training, skill and non-formalised knowledge are involved. A countermeasure is certainly offered by working with panels of experts instead of single experts. An interesting methodology to deal with this topic is presented in Beinat [1995, chap.6].

C - Communicability. Associated with the numerical values of an evaluation index, there must be a representation of the information corresponding to the *impact* described. This representation must be such to allow each *group* to perform its own preferential ranking of the *alternatives* from the particular point of view of that *impact*. This information can be visual or verbal. In a near future, however, multimedia and virtual reality will certainly support also other types of information, like sounds and, why not, tactile and gustatory-odorous sensations. For instance, to communicate the *satisfaction* for use-value of a water supply service, a numerical description can be given of the price of water, and a verbal one for the frequency and intensity of water shortages; in the future, also the taste of water could perhaps be provided.

Also *communicability*, as well as the previous requirement, calls for using only *natural* attributes. In fact, with reference to the case of the *proxy* attributes 'concentrations of contaminants', although it is clear that lower concentrations are better, a non technically expert person (e.g. the people from the *group*) is not able to evaluate how important is a change in these attributes, when compared to a change in the attributes characterising other *impact*s, as, for instance, an income increase. It is thus impossible to trade-off among *impact*s on a conscious and transparent base. The problem also exists when *proxy* indicators refer directly to standards (as is typical in pollution topics): when concentrations are lower than the admitted standard, it is easy to conclude that ‘there is no problem’; however, when they pass the standards, the intuitive interpretation is just one of ‘fear’, which is counterproductive for negotiation purposes.

D - Intensive and extensive dimensions. The *satisfaction* depends on *intensive* as well as on *extensive variables*. To see this, consider the following example: it seems obvious that the forest ecosystem A (‘good’ quality) is preferred to the forest ecosystem B (‘medium’ quality) from the point of view, say, of the existence and option value. If, however, we know that ecosystem A is just one square kilometre in size, while the B is 100 times bigger, our preferences might change. Indeed, we could be willing to dedicate the available budget to the conservation of B instead of A (it is assumed, of course, that no intermediate options are possible). Another example: a noisy industrial plant can be located in site A (generating a ‘subtle-to-medium’ noise effect to the local community A), or in site B (generating a ‘medium’ noise effect to the local community B). It seems natural to prefer site A from the point of view of the value associated with a nice, silent surrounding. If, however, we consider that community A is 10 times smaller than the B, we might change our judgement.

In both cases, the corresponding index, to be correct, should incorporate both quality (ecosystem quality, and noise, respectively) and extension (ecosystem area, and number of people in the community, respectively).

E - Spatial and temporal variability. The evaluation index should be given a structure such that the spatial and temporal variability of the attributes considered are taken into account (this issue is further discussed in Nardini 98). For example, an energy plan *alternative*, at a national level, could affect (because of thermo/hydro power plants) a set of different rivers. If a comparison among different energy plan *alternatives* has to be carried out from the point of view of, say, recreational value (among others), then the corresponding evaluation index should spatially aggregate the elements (river stretches) contributing to that type of *satisfaction*.

Spatial variability also exists at smaller scales and must be handled. For instance, water quality in a river generally varies greatly from section to section. It can be good on the average, but very bad downstream of a sewerage system discharge. Which quality to consider or how to average among the different values, depends on what the index is intended to measure. This must clearly be decided and specified.

Analogous macro and micro-scale reasonings hold for the temporal dimension. For instance, dissolved oxygen (DO) can be a relevant *proxy* indicator to characterise a river from the point of view of, say, the productive use-value (fishing). However, DO may be very high during the day (even over-saturated), and very low (close to zero) during the night, because of eutrofication. Then, a suitable aggregation of these values must be considered for the evaluation index. (In this specific case the worst value is perhaps sufficient if the fish species of interest is very sensitive to low DO concentrations.) On the other hand, an example of temporal aggregation on a macro-scale is the very common time discounting applied to a stream of future economic benefits [see for instance Dixon and Hufschmidt, 1986].

F - Internal coherence. Each numerical value of an evaluation index is associated with a particular degree of *satisfaction*. This corresponds to a given *situation* described by the natural indicators (d_1,d_2,..,d_n). It is possible that different *situations* produce the same numerical value of the index. This must imply that such situations are considered indifferent in terms of *satisfaction*, that is, the involved person (*group*) does not prefer one to the other. In addition, if *situation* A is preferred to *situation* B, then the numerical value of the index must be greater for *situation* A than for the B (because it is positively oriented). I propose to qualify an index as *internally coherent* if the above holds for all the possible *situations*.

To verify that an index accomplishes with this property, the idea is as follows. First, select a sufficiently numerous set of (real or fictitious) *situations*. Then, ask the affected *group* to perform a preferential ordering of such *situations*. This is the product of an holistic, intuitive-based judgement, and is considered the ‘true ordering’. If the index is perfect, then the ordering it provides should coincide with the ‘true’ one, otherwise the index must be refined, or changed. When many attributes are involved, however, it is very difficult, or just impossible, that a person produces a consistent, holistic ordering. For this reason, according to the method proposed among others by Beinat [1995, chap.5], it is recommendable to compare and order *situations* which differ only because of two attributes at a time, the others being constants. Of course, many sub-sets of this type can be prepared, by changing the pair of attributes considered.

Notice that if an index is *internally coherent* it means that, although based on subjective value judgements, its structure is not arbitrary, but actually represents the preference structure of the *group* involved.

G - Predictability. It must be possible to predict the numerical values that the index would assume, given future *situations* produced by the implementation of the *alternatives*. Otherwise, it cannot be used in the evaluation. The structure of the index must therefore be such that the required information can be collected at reasonable cost and in a reasonable time. It is useless to adopt very refined structures if it is impossible to obtain the information required. The frontier between refinement of the index structure and quantity of information required is, however, a matter of trade-off to be defined during the decision process itself.

H - Summarisation. The index should summarise the final effect of both direct and indirect impacts (causal factors) on the component to which a *satisfaction* is assigned. For instance, the realisation of a road crossing on a pristine ecosystem certainly will generate a (negative) direct impact on it. An indirect impact might then be generated by the consequent settlement of people along the road. From the point of view of the ecosystem quality, to be measured by an index, what counts is the combined effect of the two impacts.

I - Usefulness. This is perhaps an obvious requirement: the numerical value of the index must vary among the *alternatives*, otherwise it means that they are completely indifferent from the point of view of that *impact* and the index would be useless. Less obvious is perhaps the fact that any attribute which, although relevant to determine the *satisfaction* associated with a given *impact*, does not change among the *alternatives* considered can be omitted from the formal structure of the index. It must, however, be mentally considered when different *impact*s are compared (i.e. when judgements about their relative importance are expressed). For instance, suppose that the *impact* ‘recreation’ of a set of *alternatives* refer to the same water body (e.g. ‘lake Known’) and would affect only its water quality. Then, there is no need (contrary to requirement d) to explicitly include the extensive dimension into the index structure, because it does not vary among the *alternatives*. The judgement of relative importance of this *impact* with respect to others may depend, however, on the extension (for a given quality -including visual and cultural attributes- a large lake is likely to count more than a small one, although exceptions are well possible). But this, as the name suggests, is mentally .. known.

J - Uncertainty. The effects deriving from the implementation of a project are always affected by uncertainty. This may originate from: i) the stochasticity of the exogenous inputs driving the socio-economic-environmental system (e.g. the prices of agricultural products; the occurrence of volcanic eruptions; etc.); ii) errors or incompleteness in the information concerning the current state of the system (e.g. lack of a census of the endemic species of a river); iii) errors or imprecisions in the 'models' used to estimate the future *impacts* (e.g. badly calibrated mathematical models; wrong assumptions; misunderstanding of the results; mistakes in expert based assessments; etc.); iv) difficulty in expressing precise judgements when assessing the *satisfaction* associated with different *situations*, or in the expert-based translation of the *proxy* attributes into *natural* attributes.

Uncertainty can be considered at two different levels: a) that of the Multicriteria Analysis (MCA) technique adopted to support the preferential ranking of the *alternatives*; or b) that of the evaluation indices (objectives).

At the first (MCA) level, three tools are available: i) MultiAttribute Utility Theory (*MAUT*) [Keeney and Raiffa, 1976]; ii) *scenario analysis*; iii) *sensitivity analysis*.

The following discussion may be skipped by those readers not directly interested in the issue of how evaluation indices can be linked to MCA techniques in the evaluation phase. Logically, however, this is, to me, the right place to deal with these issues.

MAUT basically is still the only theoretically rigorous approach explicitly designed to solve a Multicriteria problem with uncertainty. Conceptually, it can tackle the first three types of uncertainty specified above. Namely, *MAUT* supplies 'the best' solution. For this, it supports the definition and construction of the MultiAttribute Utility Function (UF) $u(d_1,d_2,..,d_n)$, which is a function of all the indicators $(d_1,d_2,..,d_n)$, and incorporates the subjective preferences of who is making the decision, with respect to both the attributes and uncertainty (the Utility Function is a VF extended to the uncertain case). The theoretical developments show that the 'best solution' turns out to be the one to which the maximum expected value of the UF corresponds. Nevertheless, *MAUT* is hardly applicable to the type of problem here considered because, even if the public were actively involved in the construction of such a UF (what is probably impossible because of the lack of the necessary theoretical background and because of lack of time), no function could avoid the existence of interest conflicts! Furthermore, the determination of the involved multidimensional probability distribution (required to compute the expected value) is always a very difficult task, sometimes theoretically refutable, and certainly not transparent and understandable to common people, and thus questionable.

Scenario-analysis is a much less structured, but much simpler approach. A *scenario* is a guessed pattern of the uncertain, un-controllable exogenous inputs affecting the socio-economic-environmental system. The idea is to repeat the MCA evaluation for a number of different possible *scenarios*, and finally draw a synthetic conclusion by applying some empirical or formalised rule (e.g. the ‘worst case’ or ‘least regret’ rules: see French [1988], and also Loucks et al. [1981]). This tool is explicitly suited for the first type of uncertainty (see above).

Finally, *sensitivity analysis* (see for instance Janssen [1992]) can and should be used to assess how relevant are the consequences of lack of information, or of possible errors in assessments and/or predictions (i.e. the last three types of uncertainty).

The second level at which uncertainty can be dealt with is that of the evaluation indices. The idea is to transform each uncertain evaluation index into a deterministic equivalent. The rational for this is twofold: On the one hand, people find extremely much easier to compare deterministic things, instead than uncertain objects; therefore, to be able to substitute each original uncertain *situation* -very difficult to be represented- with a deterministic equivalent *situation* is a very attractive idea. On the other hand, *scenario* and *sensitivity analysis* can in any case be applied, so that no insight is lost.

To transform the uncertain index into a deterministic equivalent three practical possibilities can be adopted:

1- ‘Naive approach’: uncertainty is ignored, so that the *indicator*s are dealt with as if they were deterministic, and so is the index. This is perhaps the most commonly adopted approach to address the third type of uncertainty (typically the mean values are directly used for the *indicator*s and uncertainty consideration is delegated to the sensitivity analysis only)[1].

2- ‘Wald-Laplace hybrid approach’: The Wald criterion [French, 1988], also known as the *worst case* criterion, corresponds to a strong risk-averse attitude. It is applicable in a ‘strict uncertainty’ framework in which it is assumed that the probabilities of the uncertain outcomes are not known. It implies that the *indicator*s can assume values only in given uncertainty ranges, which is exactly the type of information that a person is spontaneously used to handle. The Laplace criterion [o.c.] is the classic, commonly adopted criterion, of considering the expected value of the outcome, which gives an idea of the *average*

---

[1] Some of the assessment techniques available to build VFs [Beinat, 1995] deal with the ‘imprecise’ information which characterizes ‘expert based models’. Through such techniques, the second and especially the fourth type of uncertainty is incorporated into the evaluation index. This approach can be considered a variant of the ‘naif’ approach because again the actitude against risk is not explicitly taken into consideration.

performance. The *worst case* performance corresponds to the value $\underline{v}$ that the uncertain evaluation index assumes given the worst possible combination of values of the *indicator*s in their uncertainty ranges. The *average performance* corresponds to the expected value $\bar{v}=E[v]$ of the index (E[] denotes expectation). When making a decision in presence of uncertainty, it is sometimes sufficient to integrate the Wald criterion with the Laplace criterion. A simple, but meaningful possibility for the sought deterministic-equivalent evaluation index is then to build a VF of the type $v(\underline{v}, \bar{v})$, which aggregates the *worst case* performance and the *average performance* according to the subjective risk-attitude specific to the person involved (expert or *group* representative).

Furthermore, by playing then with the *internal coherence* property of the VF, it may be possible to find a fictitious indifferent *situation* characterised by the fact that its average performance coincides with the worst case performance (i.e. $\underline{v}=\bar{v}$), while the overall value of the VF is maintained. In such a way the original uncertain *situation* is transformed in a (fictitious) deterministic equivalent.

By accepting increasing degrees of empiricism, other variants are possible. For instance, instead of the expected value E[v] (often impossible to compute because the required probability distribution is not known), a surrogate can be used, like the value which the index assumes when the *indicator*s are given their mean (or average) values.

Or, finally, the sought deterministic index can be built directly as a VF whose arguments are the average and the worst case values, respectively, assumed by the *indicator*s. This approach was adopted, for instance, by Castro [1996].

3- 'MAUT approach': this is the most theoretically sound approach that can be adopted, but also the most difficult to apply. According to this approach, the evaluation index must be built as a (MultiAttribute) Utility Function $u(d_1,d_2,..,d_n)$. Then, from MAUT, it is known that the 'utility' U (that is, the *satisfaction*) associated with the original uncertain *situation* is given by the expected value of such a Utility Function; that is, in formulas: $U = E[u(d_1,d_2,..,d_n)]$. The interesting thing is that the original uncertain *situation* can now be instantly substituted with the new fictitious deterministic *situation* C defined by a particular set of values $(d_1^c,d_2^c,..,d_n^c)$ of the *indicator*s such that the corresponding deterministic Utility equates the expected Utility of the uncertain *situation*; i.e., such that: $u(d_1^c,d_2^c,..,d_n^c) = E[u(d_1,d_2,..,d_n)]$. Of course, all this is possible only if the multidimensional probability distribution of the *indicator*s values exists and is known.

This is actually the technique applied in the Arbia case study (see later) for the Hydraulic risk *impact*. That case was not particularly complex because just one attribute is involved (scalar UF).

Notice that the main critique about the applicability of MAUT (see above) can be overcome now at the level of the evaluation indices. In particular, there is now only one 'decision maker' (the representative of the *group*) and there are not interest conflicts, by definition of *group*.

It can also be observed that, when MAUT is adopted for each evaluation index, the ranking of the *alternatives* given by the criterion of the weighted sum of the deterministic equivalent objectives (which is the principal although not unique ranking criterion in Multicriteria Analysis; see for instance Janssen [1992]) is substantially coincident with the ranking produced by MAUT if applied at the level of the overall problem (there may be a difference according to the way in which weights are determined). To see this, first note that adopting the weighted sum as a ranking criterion is equivalent to assuming that the overall Utility Function has an additive structure. This, according to Keeney [1992], is actually the case when the considered impacts are 'fundamental', i.e. it is possible to directly attach to them a *satisfaction*; and this is, by definition, our case. Notice, then, that the expected value of the additive Utility Function (a weighted sum of the stochastic objectives) is equal (being expectation a linear operator) to the sum of the expected values of the scalar functions, i.e. it is equal to the weighted sum of the deterministic equivalent objectives.

AN EXAMPLE

A very short example of an environmental decision making problem is given now to illustrate the application of the concepts presented above. Given the limits of space it is impossible to describe it with sufficient details; these can be found in Nardini and Bacci [1999]. The object is the planning of a small-scale flood-control work on the Arbia river (Siena, Italy) aimed at the protection of the town of Taverne d'Arbia. The work is an earth-bank with a fixed pre-calibrated out-flow mouth that allows small floods to pass unmodified, while floods with return periods higher than 5 years are partially and temporarily stored upstream to reduce flood peaks downstream. Four project *alternatives* were initially considered, depending on the location, all characterised by the same cost (1 billion Lira). The do-nothing, no-cost ('Alt.0') *alternative* was also included. *Alternative* A (not presented in the *evaluation matrix* in Fig.1) was later discarded because clearly dominated. In Fig.1, each impact is represented through one or more indicators (the meaning of which is reminded in the third column of the matrix and briefly discussed next).

For any given impact, the numerical value of the corresponding evaluation index (denominated *objectives J*) is also shown. This allows an immediate preferential ranking of the *alternatives* (1 stands for the best situation and 0 for the worst one amongst the *alternatives*). For the curiosity of the reader, *alternative* D was the one finally selected.

Fig.1 - An example of *evaluation matrix* (Arbia project, Siena, Italy)

Impact on	Type of *value* (V) Interest *group* (G)	Explanation (attributes)	ALT 0	ALT B	ALT C	ALT D
Hydraulic risk	V: economic, safety G: residents	**water height h_n [cm] over threshold occurring once every 10 years**	**42**	**33**	**32**	**33**
objective	J_R	(normalised)	0	0.9523	1	0.9157
Strategic	V: economic, safety G: residents	**access to future funding for further hydraulic works**	**no**	**yes**	**yes**	**yes**
objective	J_S	(normalised)	0	1	1	1
Ecosystem (*a*=20.15 ha)	V: existence, recreation, aesthetic G: environ.lists everybody	**‘long term state of the ecosystem’ with:** • ***Non-naturality of veget. species*** (negatively oriented) • ***N. of rare species***	• **74%** • **12**	• **55%** • **12**	• **57%** • **12**	• **55%** • **20**
objective	J_E	(normalised)	0	0.7796	0.6992	1
Gravel and sand processing	V: use of transport infrastructure G: processing firm	**accessibility to the work area**	**normal**	**obstacolated**	**normal**	**normal**
objective	J_L	(normalised)	1	0	1	1
Sight	V: aesthetic, recreational, existence G: residents, people on the way, ev.body	**Main sight* of the earth-bank site (units affected)** • **B (*a*=9.7)** • **C (*a*=14.2)** • **D (*a*=4.8)**	• **7** • **4** • **3**	• **6^** • **4** • **3**	• **7** • **3^** • **3**	• **7** • **4** • **1^**
objective	J_P	(normalised)	1	0	0.2764	0.9917
Damage to farmers	V: economic G: farmers	**Mean annual damage [ML/year]**	**0.9**	**2.1**	**2.3**	**4.2**
objective	J_A	(normalised)	1	0.6509	0.5877	0
Cost	V: economic G: Siena Province tax-payers	**Investment [billion Lira]**	**0**	**1.0**	**1.0**	**1.0**
objective	J_C	(normalised)	1	0	0	0

*The reader should imagine a picture (with or without project depending on the case) instead of the image code (the ^ refers to a modification of the project occurred during the DM process): space does not allow the complete presentation. In parenthesis is indicated the extensive dimension of the impact, i.e. the number ‘*a*’ of conventional units involved (1 unit=100 m of country road from which the work is visible).

****For the editor: in case space limits can be overcome I could provide the pictures mentioned. That might help****

An explanation of the meaning of the impacts and indices (see Fig.1) is now briefly provided. To clarify the terminology, notice that to each *alternative* some *impacts* are associated which imply a modified level of the relevant *satisfaction* component as seen by the involved interest group; this level is measured by an *objective* that is an ad hoc evaluation index intended to fulfil the Requirements (and built according to the framework presented in Nardini 1998).

Hydraulic risk: The index adopted is a deterministic equivalent of the stochastic variable 'river water height'. The 'threshold' is the height under which there are no flood damage in the town of Taverne d'Arbia. Events with return periods different than 10 years may well occur, but the 10-years-event is used as a reference to compare the *alternatives*. All other events are incorporated, together with the risk-averse attitude of the involved people, because the index is the expected value of a suitable Utility Function. It can be noted that the benefit of the project (about 10 cm reduction) is very limited. This was a quite surprising result that only the adoption of an indicator (water height h_h) clearly associated with the (economic&safety) *values* at stake made evident.

Strategic: If the project is cancelled, future funding for further works is very unlikely to be obtained.

Ecosystem: The 'long term state of the ecosystem' affected by the work, and/or modified as part of an environmental enhancement accompanying the project, was characterised by a set of nine indicators (specified in the original report). These characterise the ecosystem from the point of view of the existence and recreational *values*. The original complex situations have then been substituted with simpler, but completely equivalent fictitious situations, by taking advantage of the property of internal coherence discussed above. These simpler situations are characterised just by the two indicators shown[2].

Gravel-sand processing: Nearby site B there are some equipment for the processing of gravel and sand of a local firm. The realisation of the earth-bank would imply a modification of the access road with consequent disturb to the transit of the vehicles of the firm.

Sight: This refers to the visual effect that the people passing on the roads from which the work can be seen, as well as the people resident in houses located close to the sites, would experience. The evaluation index (objective) takes into account both the subjective appreciation of the sights (see the note at the foot of the matrix) and indirectly the number (a) of people involved in each *alternative.*

Agriculture: This is the economic *value* of the harvest which is expected to be lost on the average because of the flooding of the agricultural land upstream of the earth-bank.

Cost: The *alternatives* were designed under a budget constraint of 1 billion Lira.

---

[2] The *Non-naturality of the vegetational species* is a measure of how many 'non natural' species are present and it is defined over the cardinal scale: very low =presence of 'exogenous species' (anthropocore, exotic, or indicating sinanthropic environments) over 80%; low=presence equal 70%; medium= presence equal 60%; high=presence lower than 50%. A situation is therefore more desirable when the value of this indicator is lower.

Let us now see what is the role played by the Requirements in this case study.

Impacts and indices; index and satisfaction; communicability

The impacts were identified by the team supported by the information gathered through a public involvement process. The few impacts considered correspond to relevant components of the quality of life of the groups of people somehow affected. In the matrix (Fig.1), the *groups* affected are clearly shown as well as the *values* at stake, i.e. the reasons why the *satisfaction* of such *groups* would be modified. The only exception perhaps is the 'Strategic' objective; the satisfaction associated with it is actually of the same type involved by the 'Hydraulic risk', although displaced in the future. (Further comments are presented below at the Uncertainty section.)

Attention was paid to adopt, as far as possible, *natural* attributes. Water height used in the Hydraulic risk impact is not, however, really a natural indicator. Economic losses, or at least water height and associated urbanised area being flooded, would prove better choices. But the necessary deeper analysis was, as often is the case, incompatible with the budget and time limits of the planning phase.

The reaction of the people involved during the public meetings set up ensured that communicability was achieved to a sufficient degree.

Intensive and extensive dimensions

In the Arbia case study this type of problem appeared both for the *impacts* ‘ecosystem’ and ‘sight’.

In the case of 'sight' the relevancy of this issue may not appear evident. To see this, note that considering only the preferential ordering of the visual aspects of the sites would be a mistake. In fact, the number of people (a) affected changes significantly (although not dramatically). Therefore, a given (negatively) altered sight produces a worse final *satisfaction* (index value) when associated with a higher a, than with a lower a.

Spatial and temporal variability

For the *impact* ‘ecosystem’ the micro-scale temporal variability was handled by assessing and predicting the values of the indicators with reference to only one given period in the year-cycle (when the field surveys were carried out). That is, the variability during the year was ignored, by relying on the natural yearly periodicity of natural phenomena. The macro-scale time variability was handled, according to a strict sustainability approach, by referring only to the long term conditions, that is, by conceptually setting the discount rate to zero.

The spatial variability was considered when relevant, i.e. for 'ecosystem' and 'sight'. In fact, the representation given for each of such impacts is the result of a spatial aggregation over the sites involved. The methodology adopted to carry out this delicate operation cannot be explained here for limits of space (see Nardini and Bacci [1999] for details), but in simple words it consists of a corrected weighted averaging based on the extension, in the case of 'ecosystems', or number of people, in the case of 'sight'.

Internal coherence

All indices are Value Functions and therefore, by construction, they are *internally coherent*, at least with a reasonable approximation. In particular, the index ‘ecosystem’ was built on the basis of a complex questionnaire posed to a set of sectoral experts (a related experience is described in Senior [1996]). Then, by exploiting *internal coherence*, the original *situations* described by nine indicators were transformed into fictitious *situations* differing among them because of two indicators only (see the *evaluation matrix* in Fig.1), the others being held constant (including the extension). These new *situations* are equivalent to the original ones in the sense that the index values (and thus the ordering) are maintained, and so is the judgement about the *satisfaction*.

Predictability

The indices are predictable with the techniques (e.g. hydrological models), experts and information available. Uncertainty is however high (an analysis was undertaken in Nardini and Bacci [1999]). Nevertheless, what counts most is not absolute precision, but rather relative precision among the *alternatives*, and this is certainly much higher.

Summarisation

This requirement does not add much in this case because the analysis was limited to the most relevant impacts and these were all direct.

Usefulness

For the impact ‘ecosystem’, in the Arbia case study, it can be noted that the extension is not explicitly considered amongst the indicators (Fig.1). This is because after the spatial aggregation (see point E above), the area of the global, equivalent ecosystem was the same for all the *alternatives*. Therefore, it is just reminded as a constant attribute (see the first cell at the left in the ecosystem row).

Uncertainty

This was considered explicitly for the 'Hydraulic risk' through a MAUT approach: the index value is indeed the expected value of a Utility Function of the stochastic variable 'water height'. Rigorously speaking, one should have incorporated in the probability distribution of such variable also the uncertainty associated with possible future funding, while dropping the 'Strategic' objective. The approach adopted is a common 'escamotage'; it however introduces some degree of confusion because it is not clearly stated what is its contribution in terms of the only things with which a satisfaction judgement can be associated.

For the others, uncertainty was relegated to sensitivity analysis.

ANALYSIS OF SOME CLASSIC INDICES FROM THE LITERATURE

It is interesting to analyse how a typical environmental index from the literature performs with respect to the requirements. It must be noted that such indices are in general designed as diagnostic or management tools to describe the present *situation*, or to classify different items in order. They are not explicitly designed for evaluation in the sense specified in the Introduction (i.e. comparison and trading-off among the *satisfaction* corresponding to future *situations*). The following should not therefore be interpreted as a critique, but as a means of clarifying the meaning and importance of the requirements.

Brown et al. [1970] proposed a general Water Quality Index (WQI), known as the National Sanitation Foundation index. It is widely used in the United States, and is based on nine of the most common variables (DO, faecal coliforms, pH, BOD5, temperature T, nitrates, turbidity, suspended solids and phosphates) and two groups of toxic substances. Each variable i (i=1,2,..,n) is measured by a corresponding indicator d^i; this is translated in a quality score through a suitable Value Function $v_i(d^i)$. In Fig.2 the VF (called ‘quality score function’) for the faecal coliforms is shown as an example. For the toxicants the quality score function is more schematic: it is unity if the toxicant concentration is lower than a threshold, while it is zero if the threshold is passed. The index is defined as a weighted sum of these functions; i.e.: $WQI=w_1v_1(d_1)+..+ w_nv_n(d_n)$ where w_i is the 'weight' for variable i. A modified version exists which is defined as the product of the n functions each one elevated at the power of w_i .

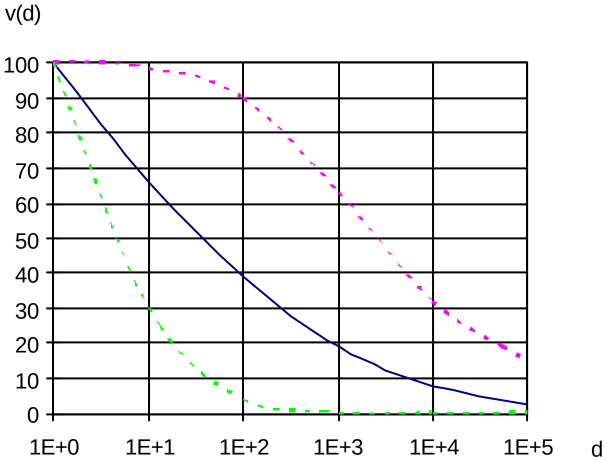

Fig.2 - Example of scalar Value Function (from Brown et al., 1970). Indicator d denotes the faecal coliforms concentration. The dotted lines correspond to the expert judgements deviating most from the average (solid line).

From the point of view of the requirements, it is perhaps evident that requirement B (*satisfaction*) is not fulfilled because the index represents a mixture of use-values (e.g. potable; recreational), and health, aesthetics, existence and option values, what implies a trade-off fixed *a priori*. Furthermore, the implicit subjective value judgements are not those characteristic of the affected *group*, but those of the experts who were involved in its construction. Finally, it is built on *proxy* attributes, hence value judgements are mixed with the experts' subjective technical judgements in an unclear fashion. The communicability (requirement C) of the index is low because it is not based on *natural* attributes: A verbal description is associated only with the whole ranges of the possible values. The index considers only an intensive dimension (requirement D). The index refers to a point in space and time; spatial and temporal variability are not explicitly considered (requirement E). They can be managed only by averaging the values assumed by the indicators. The same holds for uncertainty (requirement J). Requirement F (*internal coherence*) is perhaps the most critical point, as it is practically always the case. In fact, firstly, it is not explicitly checked that it holds at the level of the 'expert based model', i.e. the subjective translation of the *proxy* indicators into *natural* indicators implicitly carried out by the experts; second, the *internal coherence* should hold with respect to the value judgements of the *group*s affected, while in the index these judgements are delegated to the experts; third, it would be hard to check it directly because *communicability* is low. *Predictability* (requirement G) is in principle possible: mathematical models [e.g. Jorgensenn, 1994, Rinaldi et al., 1979; Thomann and Mueller, 1987; Willis and Yeh, 1987] are the most spontaneous tool utilizable. However, the effort required to quantitatively predict the value of all the variables included in the index is considerable, if not prohibitive (see Nardini et al. [1990] for BOD-DO river quality models). *Summarisation* (requirement H) can be fulfilled if the models adopted to predict the value of the indicators (concentrations) do consider all the relevant direct and indirect causes.

Similar considerations hold for many other environmental indices available in the literature (see the WQI of House and Ellis [1987]; or that of Smith [1989]; or Bhargava's [1983]; or also the air quality index of Fensterstock [1969]; the PINDEX of Babcock [1970], the ORAQI of Thomas et al. [1971], the PSI described in Ott [1978]; or also the environmental quality index of Dee et al. [1972 and 1973], Kimball [1972], Kreisel [1984]; or, the biological indices described in Metcalfe [1989]).

In conclusion, such indices are quite far from the ideal model proposed. Therefore, they would not be really suitable to support an MCA-based evaluation. This is, in my view, the main reason why evaluation indices are often criticised and consequently one of the reasons why Multicriteria Analysis not so widely applied as it could.

CONCLUSIONS

In this paper it was noted that any decision making problem with environmental concern is always a Multicriteria problem, in which interest conflicts are a constitutive element. The evaluation-negotiation should be based on a clear, summarised, and holistic view of the problem at hand. This view can hardly be achieved if not through a synthetic representation of the relevant information based on suitable *evaluation indices*.

Classic indices from the literature show a series of weaknesses that challenge their suitability for the DM problem considered here. The conclusion from this observation should not be, however, that indices are not useful or that they are 'dangerous'; rather, it should be concluded that indices can be used under the condition that they fulfil a well specified set of requirements. For this reason, I proposed a set of basic requirements and discussed them with reference to a simple case study. The key role and usefulness of evaluation indices can now be synthesised as follows.

Perhaps their main usefulness lies in supporting the statement of the decisional problem. In fact, the process of index-building ensures that the attributes really relevant to represent the involved *satisfaction* component (and associated values) have been considered. Otherwise, the index could not be *internally coherent*. But all this is possible only if the *group* clearly identifies its relevant *impact*s and corresponding values at stake, what in general may not be a banal task. In this sense, the index-building process produces a significant educational effect. At the same time, it forces to distinguish among factual information, subjective technical judgements (experts), and ethical-moral value judgements, so contributing to assigning the right responsibilities to the right persons.

The numerical values of the indices, in conjunction with their intuitive representation, allow get a very synthetic, holistic view of the decision problem at hand. It could be advocated that to get such a representation it is not necessary to construct indices. However, if one agrees with the index requirements that I propose, and therefore processes the information accordingly, then s/he is doing nothing else than actually constructing evaluation indices.

The index can be used to represent in a simpler way the relevant information. In fact, thanks to the *internal coherence* property, it is possible to transform a given *situation* (corresponding to the implementation of a given *alternative*) into another which is judged indifferent from the point of view of the *satisfaction*, but is characterised by different ‘active’ attributes (i.e. attributes whose *indicator*s assume different values among the *alternatives*). In particular, only the attributes easiest to be represented can be left active, while the others can be set at the same constant value among the *alternatives*.

Multicriteria Analysis is highly simplified because it is now just a matter of trading-off among conflicting *satisfaction* components, clearly represented by numerical indices (objectives) which are positively oriented and resume the information conveyed by the (*natural*) attributes in an *internally coherent* fashion. There is, in other words, a clear separation between the essentially technical task (involving analyst, experts and *group*s) of building a correct measure of the *satisfaction*, from the ethical-moral task of trading-off and negotiating among conflicting interests. This two-step approach is, to me, fundamental to successfully take advantage of the MCA potentialities in the decision problems here considered where public involvement is given an active role.

Notice, furthermore, that some of the MCA techniques (e.g. ELECTRE [Roy, 1978]) are designed to work directly on the (qualitative) attributes, without using more structured indices. However, *a priori* there is not a rule which unequivocally establishes the number of attributes to be adopted in order to characterise a given *impact*. An increase in the number of attributes could of course modify (improve) the degree of refinement in the description of the *impact*. If the attributes are not previously aggregated into evaluation indices, it is possible, however, that different number of attributes lead to different rankings of the *alternatives*, not because the relative importance assigned to that *impact* is judged higher, but just because each attribute in any case receives a part of the final score (apart when very particular decision rules are adopted). This is very likely to lead to incorrect decisions.

Of course, a detailed, clear documentation of the index-building process must accompany the evaluation matrix and be available to all interested parties.

Many of the concepts developed here for building evaluation indices also apply to building 'expert based models' which translate *proxy* attributes into *natural* ones. The advantage of adopting a more formalised approach with respect to empirical and common approaches is twofold: i) the experts inevitably get more insight into the problem at hand because the object of their evaluation is clarified and they are pushed to improve their coherence both when the index is being built, and *a posteriori*, when its *internal coherence* is checked [see for instance Beinat, 1995, and Senior, 1996]; ii) an agreed index allows the extrapolation of the experts' knowledge to other cases without requiring new and time-consuming and difficult-to-organise working sessions.

Much can be done to improve the process of index construction. In particular, multimedia and virtual reality can greatly improve the communicability of the impacts.

ACKNOWLEDGEMENTS

This paper is the fruit of research project n.93.91.15-1.1 of the Dirección de Investigación of the Universidad de Concepción (Chile), that I would like to thank. The research would have not been possible without the technical and moral support of my friends Carmen Teresa Senior and Juan Carlos Castro whom I want to warmly thank.

APPENDIX - AN OUTLOOK OF VALUE FUNCTIONS

Being the Value Function (VF) concept so fundamental in the context of evaluation indices, in this section, for ease of the readers not used to this topic, a non technical idea of what it is and some fundamental comments are provided. A complete and interesting text on the argument is that of Beinat [1995], where many assessment techniques to build VFs are discussed, interesting observations presented, and a rich list of references is provided.

The VF, as presented by Keeney and Raiffa [1976], is a mathematical tool that can be used to solve a Multicriteria Analysis problem, i.e. preferentially rank the *alternatives*. The same concept can, however, be adopted to preferentially order different *situations* where each *situation* is characterised by specific values of the n indicators ($d_1,d_2,..,d_n$) adopted. This is the case of interest for building evaluation indices.

A VF v can be scalar (i.e. a function v(d) of only one argument d; also called 'marginal' VF), or MultiAttribute (i.e. a function $v(d_1,d_2,..,d_n)$ of the multiple arguments $d_1,d_2,..,d_n$). Let me write in general v(A) to denote the value assumed by the VF in correspondence with the *situation* A. The basic properties which characterise a VF are the following: if *situation* A is preferred to *situation* B, then $v(A) > v(B)$; if A is judged indifferent with respect to B (i.e. generates the same *satisfaction*), then $v(A) = v(B)$; finally, if A is not preferred to B, then $v(B)>=v(A)$.

A VF can be either *ordinal* or *cardinal*. An *ordinal* VF allows performing a preferential ordering of given *situations*, according to the above rules. However, it does not represent the strength of preferences, i.e. does not say how much a *situation* is preferred to another; furthermore, it is not unique. A *cardinal* VF also represents the strengths of preferences and, when exists, it is unique. French [1988] proved that if three rather loose axioms of rational choice hold, then an *ordinal* VF exists. The idea expressed by these axioms is, loosely speaking, the following: given any pair of *situations*, an individual either prefers one of them or is indifferent between the two; furthermore, preference and indifference are transitive. Analogously, Dyer and Sarin [1979] and French [1988] stated the axioms under which a *cardinal* VF exists. The idea is, again loosely speaking, that it is possible to express judgements of the type: 'A is preferred to B more than C is preferred to D'.

It is quite common the case in which it is possible to express judgements of this last type. This, however, may require a certain effort. Not always, hence, the inquired person (the expert or the representative of the *group*) is willing to make the effort, or can dedicate the necessary time to the assessment. Therefore, the information obtained, which is of a qualitative type, would often allow only the assessment of an *ordinal* VF. Nevertheless, it is rather general practice to assess instead the *cardinal* VF which better fits the qualitative answers. In fact, the assessment techniques for *cardinal* VF are better developed and relatively easy to apply. In any case, the *a posteriori* checking of *internal coherence* can ensure the correctness of this procedure. In the following, therefore, reference is made only to *cardinal* VFs. An example of a scalar *cardinal* VF is given in Fig.2.

It is important to note that a VF can be built only once the numerical ranges of its *indicator*s are specified. Indeed, the maximum numerical value that the VF can assume (e.g. 1 if it is normalised) corresponds to the 'best *situation*' considered, which is characterised by the set of best numerical values of all the indicators. Analogously, the VF assumes its worst value (zero) in correspondence of the 'worst *situation*', characterised by the set of worst values of the indicators. All the intermediate values that the VF can assume do have meaning only as a comparisons between a given *situation* and these two extreme best and worst *situations*. This relative nature of VFs is certainly a weakness. However, it is also a strength because, by requiring only relative preference judgements, it allows VFs to be used also in the case of 'intangible' *impact*s (e.g. psychological ones).

Keeney and Raiffa [1976] demonstrated that, when certain conditions hold, then the VF assumes a particular mathematical structure. They proved this in relation to three 'canonical' structures. The most common one is the *additive* structure, which is the one adopted, for instance, in the first version of the NSF index of Brown et al. [1970] analysed above. Other two 'canonical' structures exists: the *multilinear* and the *multiplicative.* Many evaluation indices, however, were proposed which do not have any of these canonical structures. The fact that a rigorous theory does not exist which states under which conditions the VF does assume the particular non-canonical structure adopted does not imply that this is wrong. It just does not support the fact that it is the right one for the case at hand. It is always possible, however, to empirically check *a posteriori* the *internal coherence* requirement: if it holds, then the structure is all right.

As already said, this *a posteriori* checking should be always performed to ensure the correctness of the built VF.

Castro [1996], for instance, in order to assess the impact of the realisation of a water reservoir (on a Chilean river) over the health of a local community whose diet depends on trout, successfully experimented with an additive-multiplicative structure to represent the 'expert based model' of the relationship between the couple 'water temperature T and suspended solids concentration SS' and the local trout productivity. The rational for this model was the assumption that T and SS were the only *indicator*s reliably predictable in a quantitative fashion. The mathematical structure is very intuitive; it is additive when the *indicator* values belong to an intermediate interval; when any *indicator* lies outside of the corresponding interval, then the index assumes the value corresponding to the same *additive* structure multiplied by a an exponential-decay function. In this way, it is possible to represent the concept that a *situation* characterised by a very bad value of one *indicator* should be judged globally very bad, no matter the value assumed by the other. That is, the marginal rate of substitution among the scalar VFs is not constant. This structure is in particular much less 'drastic' than the 'worst case' structure adopted, for instance, by Smith [1989].

It is worth noting [Keeney, 1992] that when the evaluation index is built on the base of *natural* attributes it almost always has an *additive* structure. On the contrary, when it is built on the base of *proxy* attributes this structure may not be adequate, so that fulfilling *internal coherence* can be a problem. This is because the value judgements incorporated into the VF necessarily refer, as already noted, to the *natural* attributes, and these are bound to the *proxy* attributes by synergetic or antagonic non-linear relationships. Also the use of standards into the index structure can produce problems related to *internal coherence*. These problems originate from the fact that one would expect that when one of the indicators passes its corresponding threshold (standard), then the evaluation index also passes a 'global standard' threshold; on the other hand, when the index passes this 'global standard', at least one of the indicators should pass its corresponding standard. Tomlinson [1988] showed that some of the most used VF mathematical structures (among which the *additive*) do not show the desired behaviour just described. There are, indeed, 'ambiguous regions' where the index passes the 'global standard' but none of the indicators passes its own; and 'regions of eclipses' where one or more of the indicators passes their own standards, while the index keeps lower than the global standard. This problem, however, vanishes when *natural* attributes are used, as it should be.

REFERENCES

Andrews R. (1974), ‘Comments on 'An Environmental Evaluation System for Water Resources Planning' by N. Dee et al.’ Water Resources Research, V.10(2).

Babcock L.R. (1970), ‘A combined pollution index for measurement of total air pollution’ J. of Air Pollution Control Association, V.20, n.10, pp.653-659.

Beinat E. (1995), Multiattribute value functions for environmental management (Book n.103 of the Tinbergen Inst. Res. Series, Free University, Amsterdam).

Beinat E., et al. (1994), ‘Value functions for environmental pollutants: a technique for enhancing the assessment of expert judgements’,Environmental Monitoring and Assessment 30, pp9-23.

Bhargava D.S. (1983), ‘Use of a water quality index for a river classification and zoning of Ganga river’ Environmental Pollution (series B), Vol.6, pp.51-67.

Brown, R. M., N. I. McClelland, R. A. Deininger, and R. G. Tozer (1970), ‘A water quality index. Do we dare?’ Water Sewage Works, October, pp.339-343.

Castro J. (1996), Multicriteria Analysis in the Environmental Field (Civil Engineering degree Thesis, Dep. of Civil Engineering, Universidad de Concepción, Concepción, Chile; in spanish).

Dee N., J.K. Baker, N.L. Drobny, K.M. Duke and D.C. Fahringer (1972), Environmental Evaluation System for Water Resources Planning (Report n.208822, Battelle Columbus Laboratories, Columbus, Ohio).

Dee N., J.Baker, N. Drobny, and K. Duke (1973), ‘An Environmental Evaluation System for Water Resources Planning’ Water Resources Research, V.9(3).

Dixon, J. and M. Hufschmidt (1986). Economic valuation techniques for the environment. Baltimore, Maryland: The Johns Hopkins University Press.

Dyer J.S., and R.K. Sarin (1979), ‘Measurable multiattribute value functions’ Operations Research, V.27, n.4, pp.810-822.

Eisel L., and J. Gaudett (1974), ‘Comments on 'An Environmental Evaluation System for Water Resources Planning' by N. Dee et al.’ Water Resources Research, V.10(2).

Elliott M. (1981), ‘Pulling the pieces together: amalgametion in Environmental Impact Assessment’ Environmental Impact Assessment Review, V.2(1).

Fensterstock J.C., K. Goodman, G.M. Duggan, and W.S. Baker (1969), ‘The development and utilization of an air quality index’. Paper n.69-73 presented at the 62nd Annual Meeting of the Air Pollution Control Association, New York, NY, June 1969.

French S. (1988), Decision Theory. An Introduction to the Mathematics of Rationality (Ellis Horwood, Chichester, UK).

Gregory R., R. Keeney, and D. von Winterfeldt (1992). Adapting the Environmental Impact Statement process to inform Decision Makers. *J. of Policy Analysis and Management, V.11(1).*

Horton R.K. (1965), ‘An Index Number System for Rating Water Quality’ J. of Water Pollution Control Federation, V.37, n.3, pp.300-305.

House M.A., and J.B. Ellis (1987), ‘The development of wtare quality indices for operationa management’ Water science & Technology, V.19, pp.145-154.

Inhaber H. (1976), Environmental Indices (John Wiley & Sons, London).

Janssen R. (1992), Multiobjective Decision Support for Environmental Management (Kluwer Academic Publishers).

Jorgensenn S.E. (1994), Fundamentals of Ecological Modeling (II Edition, Elsevier).

Kimball T.L. (1972), ‘Why Environmental Quality Indices?’, in W.A. Thomas (ed.) *indicator*s of Environmental Quality (Plenum Press, New York).

Keeney, R. (1992), Value Focused Thinking (Harvard University Press, Cambridge, Massachusetts).

Keeney R., and H. Raiffa (1976), Decisions with Multiple objectives: preferences and value tradeoffs (Wiley, N.Y).

Keeney R., R. Sarin, and R. Winkler (1984), ‘Analysis of alternative National Ambient Carbon Monoxide Standards’ Management Science, V.30, n.4, 518-528.

Kreisel W.E. (1984), ‘Representation of the Environmental Quality Profile of a Metropolitan Area’ Environmental Monitoring and Assessment, V.4, pp.15-33.

Loucks D.P., Stedinger J.R., and D.A. Haith (1981), Water Resource Systems Planning and Analysis (Prentice-Hall, Inc., Englewood Cliffs, New Jersey).

Liu B.C. (1975), Quality of life indicators in U.S. Metropolitan areas, 1970: a comprehensive assessment (Report prepared by Midwest Research Inst. for the U.S. Environmental Protection Agency, Washington, D.C).

Metcalfe J.L. (1989), ‘Biological Water Quality Assessment of Running Waters Based on Macroinvertebrate Communities: History and Present Status in Europe’ Environmental Pollution (series A), V.60, pp.101-139.

Nardini A. (1997a). Integrating Environmental Impact Assessment, Cost-Benefit Analysis, Multicriteria Analysis: a step forward through the value-focused identification of impacts. Proceedings of the IV InterAmerican Congress on Environment, Caracas (December 1997).

Nardini A (1997b). ‘A proposal for integrating Environmental Impact Assessment, Cost-Benefit Analysis and Multicriteria Analysis’. Project Appraisal, Vo.12, n.3, Sept. 97.

Nardini A. (1998). Improving decision-making for land-use management: key ideas for an integrated approach built on an MCA-based negotiation forum. In Multicriteria evaluation in land-use management: methodologies and case studies. E. Beinat and P. Nijkamp (eds.), Kluwer Academic Press.

Nardini A. and M. Bacci (1999). Participatory evaluation -based on Multicriteria Analysis- of a flood-control earth-bank on the river Arbia (Siena, Italy). Being published by Regione Toscana, Firenze, Italy.

Nardini A., R. Soncini-Sessa, and M. Bacci (1990), River pollution: realization and use of mathematical models (Marsilio ed., Venezia, Italia; in Italian).

Beinat E. and P. Nijkamp (eds.) (1997). Multicriteria evaluation in land-use management: methodologies and case studies. Kluwer Academic Press.

O'Connor M.F. (1972), The application of Multi-attribute scaling procedures to the development of indices of water quality (Ph.D. Dissertation, University of Michigan).

Ott W.R. (1978), Environmental Indices. Theory and practice (Ann Arbor, Michigan, Ann Arbor Science Publishers Inc).

Renn, O., Webler, T., Rakel, H., Dienel, P. and B. Johnson (1993). Public participation in decision making: A three-step procedure. *Policy Analysis, 26, 189-214.*

Rinaldi S., R. Soncini-Sessa, H. Stehfest, and H. Tamura (1979), Modeling and Control of River Quality (Mc Graw-Hill).

Roy B. (1978), ELECTRE III: un algorithme de classement fondé sur une representation floue de preference en presence de critères multiples. Cahiers Centre de Etudes Recherche Operationelle, LAMSADE, Paris, 20(1), 3-24.

Senior C. (1996), A Proposal for Indices to Evaluate the Impact of Large Hydraulic Works on the Biological Environment. Case Study: Laja-Diguillín project (PhD Thesis, Centro EULA, Concepción, Chile; in spanish).

Smith D.G. (1989), ‘A new form of water quality index for the great lakes’ Water science & Technology, V.21, n.2, pp.123-127.

Thomann R., and J. Mueller (1987), Principles of Surface Water Quality Modeling and Control (Harper & Row Publishers, New York).

Thomas W.A., L.R. Babcock, and W.B. Schults (1971), Oak Ridge air quality index (Report n.ORNL-NSF-EP-8, Oak Ridge National Laboratory, Oak Ridge, TN, September 1971).

Tomlinson P. (1988), ‘Conceptual problems in the use of environmental indicators in EIA’, in. P. Schmidt di Friedberg (ed.), Gli indicatori ambientali (Franco Angeli, Milano, Italia; pp.749-770).

Willis R., and W.W. Yeh (1987), Groundwater systems planning and mangement (Prentice-Hall Inc., Englewood Cliffs, New Jersey).

FIGURE CAPTIONS

Fig.1 - An example of *evaluation matrix* (Arbia project, Siena, Italy)

Fig.2 - Example of a quality score function v(d) (slightly modified from Brown et al. [1970]; d: faecal coliforms/100 ml logarithmic scale).