Manuscript forthcoming in Environment, Development and Sustainability

Version November 2003

"Measuring Sustainability": A Multi-Criterion Framework

Giuseppe Munda
Universitat Autonoma de Barcelona
Dept. of Economics and Economic History
Edifici B
and Institute for Environmental Sciences and Technologies
08193 Bellaterra (Barcelona)
Spain
giuseppe.munda@uab.es

Abstract. So far, the elementary question of whether one country's or region's economy is moving towards sustainability or away from it cannot be answered with unanimous consensus on the "measuring rod(s)" to be employed. The main assumption of this article is that sustainability assessment needs a set of multidimensional indicators. From this assumption a question arises: how could such indicators be aggregated? Often, some indicators improve while others deteriorate. For instance, when incomes grow, SO_2 might go down while CO_2 increases. It has to be noted that this is the classical conflictual situation studied in multi-criteria decision theory. The use of a multi-criterion framework for making operational the "measuring of sustainability" is discussed here by means of illustrative examples and more formal arguments.

KEY WORDS: COMPOSITE INDICATORS, RANKING PROCEDURES, SUSTAINABILITY BENCHMARKING, SOCIAL MULTI-CRITERIA EVALUATION

1. A Short Introduction to Multi-Criteria Evaluation

The *discrete multi-criterion problem* can be described in the following way: A is a finite set of N feasible actions (or alternatives); M is the number of different points of view or evaluation criteria g_m $i=1, 2, ... , M$ considered relevant in a policy problem, where the action ***a*** is evaluated to be better than action ***b*** (both belonging to the set A) according to the *m-th* point of view if $g_m(a)>g_m(b)$. In this way a decision problem may be represented in a tabular or matrix form (as the one presented in Table 1) (Beinat and Nijkamp, 1998; Munda, 1995). The main advantage of multi-criteria models is that they make it possible to consider a large number of data, relations and objectives which are generally present in a specific real-world policy problem, so that the problem at hand can be studied in a multidimensional fashion. On the other side, an action ***a*** may be better than an action ***b*** according to one criterion and worse according to another, thus in general, there is no solution optimising all criteria simultaneously and therefore *compromise solution* have to be found.

In synthesis, the information contained in the impact matrix useful for solving the so-called multi-criterion problem is:

- *Intensity of preference* (when quantitative criterion scores are present).
- *Number* of criteria in favour of a given alternative.
- *Weight* attached to each single criterion.
- *Relationship* of each single alternative with all the other alternatives.

Combinations of this information generate different aggregation conventions, i.e. manipulation rules of the available information to arrive at a preference structure. The aggregation of several criteria implies taking a position on the fundamental issue of compensability. *Compensability* refers to the existence of trade-offs, i.e. the possibility of offsetting a disadvantage on some criteria by a sufficiently large advantage on another criterion, whereas smaller advantages would not do the same. Thus a preference relation is non-compensatory if no trade-off occurs and is compensatory otherwise. The use of weights with intensity of preference originates compensatory multi-criteria methods and gives the meaning of trade-offs to the weights. On the contrary, the use of weights with ordinal criterion scores originates non-compensatory aggregation procedures and gives the weights the meaning of importance coefficients (Roberts, 1979). We will see in Section 3 two different mathematical procedures implementing these concepts.

To develop a multi-criterion framework for dealing with sustainability indicators, there is a need to define some basic concepts[1]:

Dimension: is the highest hierarchical level of analysis and indicates the scope of objectives, individual indicators and variables. For example, a sustainability composite indicator can include economic, social and environmental dimensions.

Objective: an objective indicates the direction of change desired. For example, within the economic dimension GDP has to be maximised; within the social dimension social exclusion has to be minimised; within the environmental dimension CO_2 emissions have to be minimised.

Individual indicator: it is the basis for evaluation in relation to a given objective (any objective may imply a number of different individual indicators). It is a *function* that associates each single country (region or city) with a variable indicating its desirability according to expected consequences related to the same objective. For example, GDP, saving rate and inflation rate inside the objective "growth maximisation".

Variable: is a constructed measure stemming from a process that represents, at a given point in space and time, a shared perception of a real-world state of affairs consistent with a given individual indicator. To give an example, in comparing two countries, inside the economic dimension, one objective can be "maximisation of economic growth"; the individual indicator might be R&D performance, the indicator score or variable can be "number of patents per million of inhabitants". Another example: an objective connected with the social dimension

[1] Some of these definitions were inspired by discussions with M. Giampietro and M. Nardo.

can be “maximisation of the residential attractiveness”. A possible individual indicator is then “residential density”. The variable providing the individual indicator score might be the ratio persons per hectare.
A composite indicator or synthetic index is an aggregate of all dimensions, objectives, individual indicators and variables used. This implies that what formally defines a composite indicator is the *set of properties underlying its aggregation convention*.

This paper is divided as follows: Section 2 critically surveys the main attempts of “measuring sustainability” by means of aggregate indexes. Section 3 discusses by using an illustrative example and some formal arguments the use of a multi-criterion framework for ranking countries, regions or cities. Section 4 deals with the issue of sustainability indicators benchmarking. The policy relevance of this exercise is stressed. Finally some conclusions are offered.

2. Measuring Sustainability: the Issue of Sustainability Assessment Indexes

The purpose of "green accounting" is to provide information on the sustainability of the economy but there is no settled doctrine on how to combine different and sometimes contradictory indicators and indexes in a way immediately useful for policy (in the sense that GDP or other macroeconomic statistics have been useful for policy) (Funtowicz et al., 1999, 2002). The expression *"Taking nature into account"* (much used both in the UN system and in the European Union) hides the tension between money valuation, and appraisal through physical indicators and indexes (which themselves might show contradictory trends). So far, the elementary question of whether the European economy is moving towards sustainability or away from sustainability cannot be answered with consensus on the indicators and the integrative framework to be used (see e.g. Allen et al., 2002; Barbier and Markandya, 1990; Chichilnisky, 1996; Faucheux and O’Connor, 1998; Horwarth and Norgaard, 1990, 1992; Munda, 1997a; Musu and Siniscalco, 1996; Pearce et al., 1996).

A point of scientific controversy present in the contemporary debate is on the use of monetary or physical indexes. Examples of monetary indexes are Daly and Cobb (1989) ISEW (Index of Sustainable Economic Welfare), Pearce and Atkinson (1993) Weak Sustainability Index, the so-called El Serafy approach (Yusuf et al., 1989). Examples of physical indexes are HANPP (Human Appropriation of Net Primary Production (Vitousek et al., 1986), the Ecological Footprint (Wackernagel and Rees, 1995), MIPS (Material Input Per unit of Service) (Schmidt-Bleek, 1994).

Although these approaches may look different, in reality, they all have some common characteristics:

1. The subcomponents needed for the building the aggregate index are *ad hoc*. No clear justification is given why e.g. diet enters in the computation of the ecological footprint and the generation of waste does not.
2. All the indexes are based on the assumptions that a common measurement rod needs to be established for aggregation purposes (money, energy, space, and so on). This creates the need of making very strong assumptions on conversion coefficients to be used and on *compensability* allowed (i.e. till which point better economical performances may cause environmental destruction or social exclusion?). The mathematical aggregation convention behind an index thus needs an explicit and well thought formulation.

3. The policy objective is often not clear. Inter-country or inter-city comparisons are a different policy objective than managing a particular country or city sustainability. In this latter case, a benchmarking exercise becomes essential. Aggregate indexes are somewhat confusing, if one wishes to derive policy suggestions. For example, by looking at ISEW, we could know that indeed a country has a worst sustainability performance than the one pictured by standard GDP, but so what? ISEW being so aggregated does not supply any clear information of the cause of this bad performance and thus is useless for policy-making (while conventional GDP is at least giving clear information on the economic performance). The same applies to the ecological footprint, which sometimes can even give misleading policy suggestions (giving that diet is used, a more energy intensive agriculture might reduce the ecological footprint of e.g. a city, but in reality its environmental performance would be much worst!) or to the weak sustainability index (which is nothing but the classical golden rule of growth theory, where environmental physical destruction is never considered – above all if it is externalised outside the national borders).
4. All these approaches belong to the more general family of composite indicators (OECD, 2003; Saisana and Tarantola, 2003), and as a consequence, the assumptions used for their construction are common to them all.

This paper builds on the following main assumption: *when dealing with sustainability indicators and indexes neither an economic reductionism nor an ecological one is possible.* Since in general, economic sustainability has an ecological cost and ecological sustainability has an economic cost, an integrative framework able to tackle conflicts such as multi-criteria evaluation is needed for sustainability "measurement".
At this stage two questions arise:

1. When a ranking of different countries, cities or regions is provided, from where this ranking is coming from?
2. Is a ranking of any utility for policy-making?

These questions will be discussed in the next two Sections.

3. Warning! Not Always Rankings Have to Be Trusted …

Let's take into consideration an illustrative example regarding 4 cities, 2 belonging to highly industrialized Countries (Amsterdam and New York) and 2 belonging to transitional economies (Budapest and Moscow). The indicators used are typical of the literature on urban sustainability (see e.g. Barbiroli, 1993 or the Urban Indicator Programme). The profiles (i.e. the score of each city according to each indicator) of these 4 cities are the ones described in Table 1.

Matrix type Impact Case Study

Criteria \ Alternatives	Budapest	Moscow	Amsterdam	New York
Houses owned (%)	50.5	40.2	2.2	10.3
Residential density (pers. /hectare)	123.3	225.2	152.1	72
Use of private car (%)	31.1	10	60	32.5
Mean travel time to work (minutes)	40	62	22	36.5
Solid waste generated per capita (t./year)	0.2	0.29	0.4	0.61
City product per person (US$/year)	4750	5100	28251	30952
Income disparity (Q5/Q!)	9.19	7.61	5.25	14.81
Households below poverty line (%)	36.6	15	20.5	16.3
Crime rate per 1000 (theft)	39.4	4.3	144.05	56.7

Table 1. Impact Matrix for the 4 Chosen Cities According to the Selected Indicators

A standard approach is to rank these cities by constructing a composite indicator. A typical composite indicator, *I*, is built as follows (OECD, 2003, p. 5):

$$I = \sum_{i=1}^{N} w_i x_i \tag{1}$$

where x_i is a normalised variable and w_i a weight attached to x_i, with $\sum_{i=1}^{N} w_i = 1$ and $0 \le w_i \le 1$, $i = 1,2,\dots, N$.

It is clear that from a mathematical point of view a composite indicator entails a weighted linear aggregation rule applied to a set of variables. The main technical steps needed for its construction are two:

1. *Standardisation of the variables to allow comparison without scale effect,*
2. *Weighted summation of these variables.*

The standardisation step is a very delicate one. Main sources of uncertainty and imprecise assessment here are:

- *Normalisation technique* used for the different measurement units dealt with.
- *Scale adjustment* used, for example population or GDP of each country considered.
- *Common measurement unit* used (money, energy, space and so on).

Several techniques can be used to standardise variables (OECD, 2003; Saisana and Tarantola, 2002). However, although each normalisation technique entails different absolute values, the ranking provided remains constant. In our example, the *"distance from the best and worst performers"* technique is applied, where positioning is in relation to the global maximum and minimum and the index takes values between 0 (laggard) and 100 (leader):

$$100\left(\frac{\text{actual value - minimum value}}{\text{maximum value - minimum value}}\right) \tag{2}$$

By applying equation (2) to the values contained in Table 1, the results presented in Table 2 are obtained.

100	78.674	0	16.770
33.485	100	52.28	0
42.2	0	100	45
45	100	0	36.25
0	21.95	48.78	100
0	1.335	89.691	100
41.213	24.686	0	100
100	0	25.462	6.018
25.116	0	100	37.495

Table 2. Normalised Impact Matrix

The indicators "houses owned" and "city product per person" have to be maximised. All the others have to be minimised. To apply equation (1) it is thus necessary to transform the indicator scores of these indicators by using the simple equation (100 – normalised indicator score). By applying this transformation to the values contained in Table 2, the results presented in Table 3 are obtained.

100	78.674	0	16.770
66.515	0	47.72	100
57.8	100	0	55
55	0	100	63.75
100	78.05	51.22	0
0	1.335	89.691	100
58.787	75.314	100	0
0	100	74.538	93.982
74.884	100	0	62.505

Table 3. Normalised Impact Matrix Accounting for Minimisation Objectives

By applying equation (1) to the values contained in Table 3, the following results are obtained:
Budapest = 512.986
Moscow = 533.373
Amsterdam = 463.169
New York = 492.052

Thus the final ranking presents Amsterdam in the bottom position (worst than all the other cities considered), Moscow is in the top position, Budapest ranks second and New York ranks third.

At this point a question needs to be answered:
From where are these (somewhat surprising) results coming from and what they mean?

Let's start with the first question. The results obtained depend on:

1. *quality of the information available* (in our case for example the data concerning Amsterdam on the use of private cars and on criminality are suspiciously high, while criminality in Moscow or residential density in New York are suspiciously low),
2. *indicators chosen* (i.e. which representation of reality we are using, e.g. whose interests we are taken into account),
3. *Direction of each indicator* (i.e. the bigger the better or vice versa, e.g. in our example, it has been used the principle that house owners should be maximized, but this could be quite disputable and culturally dependent),
4. *relative importance of these indicators* (in our case all the indicators are considered having the same importance i.e. no weighting coefficient is used),
5. *ranking method used* (in this case the linear aggregation rule).

All these uncertainties have to be taken into account when we state that a given city is "better" than another one. Points from 1 to 4 clearly concern the way a given assessment exercise is structured; thus it seems clear why in multi-criteria evaluation it is claimed that what is really important is the *"decision process"* and not the final solution, since this solution has a value only as a construction of the decision process and it is not an ultimate Truth (in Herbert Simon words, we could say that we should move from "substantive to procedural rationality").
Point 5 is more technical in nature, since it concerns the aggregation procedure used.

Munda and Nardo (2003) analyse the assumptions underlying the linear aggregation rule, used in composite indicators, and prove the following main conclusions:

1. Weights in linear aggregation rules have always the meaning of trade-off ratio. In all constructions of a composite indicator, weights are used as importance coefficients; as a consequence, a theoretical inconsistency exists.
2. The assumption of preference independence is essential for the existence of a linear aggregation rule. Unfortunately, this assumption has very strong consequences which often are not desirable in sustainability indicators. The use of a linear aggregation procedure implies that among the different ecosystem aspects there are not phenomena of synergy or conflict. This appears to be quite an unrealistic assumption (Funtowicz et al., 1990).
3. In linear aggregation rules, compensability among the different individual indicators is always assumed; this implies complete substitutability among the various components considered. For example, in a sustainability index, economic growth can always substitute any environmental destruction or inside e.g., the environmental dimension, clean air can compensate for a loss of potable water. From a descriptive point of view, such a complete compensability is often not desirable.

A simple ranking algorithm, more consistent than the linear aggregation rule in the framework of sustainability indicators, can be the following (for more details and formal proofs see Munda and Nardo, 2003).

Given a set of individual indicators $G=\{g_m\}$, $m=1,2,..., M$, and a finite set $A=\{a_n\}$, $n=1, 2,..., N$ of countries (cities or regions), let's assume that the evaluation of each country a_n with respect to an individual indicator g_m (i.e. the indicator score or variable) is based on an *interval or ratio* scale of measurement. For simplicity of exposition, let's assume that a higher value of an individual indicator is preferred to a lower one (the higher, the better), that is:

$$\begin{cases} a_j P a_k \Leftrightarrow g_m(a_j) > g_m(a_k) \\ a_j I a_k \Leftrightarrow g_m(a_j) = g_m(a_k) \end{cases} \tag{3}$$

Where, P and I indicate a preference and an indifference relation respectively, both fulfilling the transitive property.

Let's also assume the existence of a set of individual indicator weights $W=\{w_m\}$, $m=1,2,...,M$, with $\sum_{m=1}^{M} w_m = 1$, derived as importance coefficients. The mathematical problem to be dealt with is then how to use this available information to rank in a complete pre-order (i.e. without any incomparability relation) all the countries from the best to the worst one.

The mathematical aggregation convention can be divided into two main steps:

1. Pair-wise comparison of countries according to the whole set of individual indicators used.
2. Ranking of countries in a complete pre-order.

A $N \times N$ matrix, E, called *outranking matrix* (Arrow and Raynaud, 1986, Roy, 1996) can be built. Any generic element of E: e_{jk}, $j \neq k$ is the result of the pair-wise comparison, according to all the M individual indicators, between countries j and k. Such a global pair-wise comparison is obtained by means of equation (4).

$$e_{jk} = \sum_{m=1}^{M} \left(w_m(P_{jk}) + \frac{1}{2} w_m(I_{jk}) \right) \tag{4}$$

where $w_m(P_{jk})$ and $w_m(I_{jk})$ are the weights of individual indicators presenting a preference and an indifference relation respectively. It clearly holds

$$e_{jk} + e_{kj} = 1. \tag{5}$$

The maximum likelihood ranking of countries is the ranking supported by the maximum number of individual indicators for each pair-wise comparison, summed over all pairs of countries considered. More formally, all the $N(N-1)$ pair-wise comparisons compose the outranking matrix E. Call R the set of all $N!$ possible complete rankings of alternatives, $R=\{r_s\}$, $s=1,2,..., N!$. For each r_s, compute the corresponding score ϕ_s as the summation of e_{jk} over all the $\binom{N}{2}$ pairs j,k of alternatives, i.e.

$$\phi_s = \sum e_{jk} \quad . \tag{6}$$

$$\text{where } j \neq k,\ s = 1,2,...N! \ \text{ and } \ e_{jk} \in r_s$$

The final ranking (r_*) is the one which maximises equation (7), which is:

$$r_* \Leftrightarrow j_* = \max \sum e_{jk} \qquad where\ e_{jk} \in R. \tag{7}$$

Let's apply this algorithm to the impact matrix showed in Table 1.
The outranking matrix *E* is the one showed in Table 4.

	Budapest	**Moscow**	**Amsterdam**	**New York**
Budapest	0	4	4	5
Moscow	5	0	5	6
Amsterdam	5	4	0	3
New York	4	3	6	0

Table 4. Outranking Matrix of the 4 Cities According to the 9 Indicators

The 24 possible rankings and the corresponding scores j_s are the following:

B	**A**	**D**	**C**	**31**	C	B	D	A	27
B	**D**	**C**	**A**	**31**	D	B	A	C	27
A	B	D	C	30	D	C	B	A	27
B	D	A	C	30	A	C	B	D	26
B	C	A	D	29	A	D	C	B	26
B	A	C	D	28	D	A	B	C	26
B	C	D	A	28	D	C	A	B	26
C	B	A	D	28	D	A	C	B	25
D	B	C	A	28	C	A	D	B	24
A	B	C	D	27	C	D	B	A	24
A	D	B	C	27	A	C	D	B	23
C	A	B	D	27	C	D	A	B	23

Where A is Budapest, B is Moscow, C is Amsterdam and D is New York.
Also in this case Moscow is clearly in the top position. New York is surely better than Amsterdam. The position of Budapest with respect to both New York and Amsterdam is not well defined.

Let's now look at an interesting feature of this procedure connected to the problem structuring step, i.e. the use of weights as importance coefficients. This concept can be synthesised as follows: *"… if we have two non-equal numbers to construct a vector in* R^2*, then it is preferable to place the greatest number in the position corresponding to the most important criterion."* (Podinovskii, 1994, p. 241).

Let's look at Table 1 again. The 9 indicators used seem reasonable; they indeed belong to three dimensions, i.e. economical, social and environmental, considered essential in any sustainability assessment. Let's then try to understand to which dimension each single indicator belongs. Roughly the following classification may be made:

Economic dimension

1. City product per person

Environmental dimension

2. Use of private car
3. Solid waste generated per capita

Social dimension

4. Houses owned
5. Residential density
6. Mean travel time to work
7. Income disparity
8. Households below poverty line
9. Crime rate

Clearly the social dimension is receiving implicitly a much bigger weight than any other dimension (considering that 6 indicators over 9 belong to this dimension). A reasonable decision might be to consider the three dimensions equally important. This would imply to give the same weight to each dimension considered and finally to split this weight among the indicators. That is, each dimension has a weight of 0.333; then the economic indicator has a weight of 0.333, the 2 environmental indicators have a weight of 0.1666 each, and each one of the 6 social indicators receives a weight equal to 0.0555. As one can see, if dimensions are considered, weighting indicators by means of importance coefficients is crucial.

Let's now see if this weighting exercise provokes any change in the final ranking. The new outranking matrix is the one presented in Table 5.

	Budapest	**Moscow**	**Amsterdam**	**New York**
Budapest	0	0.3	0.4	0.4
Moscow	0.7	0	0.5	0.6
Amsterdam	0.6	0.5	0	0.3
New York	0.6	0.4	0.7	0

Table 5. Weighted Outranking Matrix

The 24 possible rankings and the new corresponding scores j_s are the following (Where A is Budapest, B is Moscow, C is Amsterdam and D is New York):

B	**D**	**C**	**A**	**3,6**	B	C	A	D	2,9
D	B	C	A	3,5	C	B	A	D	2,9
D	C	B	A	3,5	A	B	D	C	2,9
B	D	A	C	3,5	B	A	C	D	2,8
D	B	A	C	3,4	A	D	B	C	2,8
B	A	D	C	3,3	A	D	C	B	2,8
B	C	D	A	3,2	C	D	A	B	2,7
C	B	D	A	3,2	C	A	B	D	2,6
D	C	A	B	3,2	C	A	D	B	2,5
C	D	B	A	3,1	A	B	C	D	2,5
D	A	B	C	3,1	A	C	B	D	2,5
D	A	C	B	3,1	A	C	D	B	2,4

As one can see, Moscow is still on the top position, but this time Budapest is on the bottom one. New York scores again better than Amsterdam.

Concluding, we can state that an advantage of this algorithm is to highlight the fact that rankings are not always robust and thus uncertainty sometimes exists. This uncertainty is completely ignored by the linear aggregation rule. Moreover, the use of weights as importance coefficients can change the problem modelling significantly. However one has to note that the improvement of the mathematical aggregation procedure does not change the results spectacularly. The structuring process, and in this case above all, the input information used for the indicator scores determine clearly the ranking. *Garbage in, garbage out* phenomena are almost impossible to avoid (Funtowicz and Ravetz, 1990).

4. Aiding Policy Decisions: Sustainability Benchmarking

The second question looks even nastier i.e. *is all this effort we have done of any use?* Even if we have a very reliable ranking, which is the utility of knowing that Moscow is overall better than Amsterdam or vice versa? Let's try to put some light on this issue. First of all, one should note that for the majority of indicators used in assessment exercises no clear reference point is available, for instance, when GDP is used nobody knows the ideal value of a Country GDP, thus it is quite common to compare with other Countries GDP, e.g. the USA one. Let's continue the example of our 4 cities to see how the assessment of various indicators can easily be used for policy purposes.

In order to get a set of reference values, an "ideal point" can be defined by choosing the best values reached in any single indicator. This is a well established technique in multi-criteria evaluation literature (see e.g. Yu, 1985; Zeleny, 1982) and has the advantage of indicating "real world ideal values". In our case study, the vector defining the ideal value (called "ideal city" is the one presented in Table 6.

Matrix type Impact Case Study ecolindic2.nd

Criteria \ Alternatives	Ideal City
Houses owned (%)	50.5
Residential density (pers. /hectare)	72
Use of private car (%)	10
Mean travel time to work (minutes)	22
Solid waste generated per capita (t./year)	0.2
City product per person (US$/year)	30952
Income disparity (Q5/Q!)	5.25
Households below poverty line (%)	15
Crime rate per 1000 (theft)	4.3

Table 6. Multidimensional Representation of the "Ideal City"

For mathematical procedures useful for a sustainability benchmarking exercise, in my opinion, the following properties are desirable:

1. To avoid the aggregation of all the indicators in one single aggregate function. This

approach is not desirable because it does not give useful information on the behaviour of the single indicators so that its policy usefulness is very limited.

2. To avoid complete compensability, i.e. the possibility that a good score on one indicator can always compensate a very bad score on another indicator. Economic development implies the creation of new assets in terms of physical, social and economic structures. At the same time, like in any process of "creative destruction", traditional physical, social and cultural assets derived from our common heritage may disappear. Complete compensability implies that an excellent performance on the economic dimension can justify any type of very bad performance on the other dimensions, which is exactly what the concept of sustainability tries to avoid.
3. To be as much transparent as possible to the general public. In sustainability management and planning distributional issues play a central role. If a given policy option is evaluated to be "good" or to be "bad", key questions are "good" or "bad" for which point of view? For whom? How long? Any policy option always implies winners and losers, thus it is important to check if a policy option looks good just because some dimensions (e.g. the environmental) or some social groups (e.g. the lower income groups) are not taken into account (Giampietro, 1994).

A first very simple procedure can be the application of a normalisation rule known as *"distance from the group leader"*, which assigns 100 to the leading country and other countries are ranked as percentage points away from the leader (OECD, 2003).

By applying this normalisation rule (taking care that when the objective is minimisation the leader is the city with the lowest indicator score) to the indicator scores of Amsterdam and New York (the leader is of course the ideal city) the results presented in Table 7 are obtained. As shown in Figures 1, 2 and 3 these results can also be presented graphically for making their interpretability easier. The numerical results are synthesised by using the so-called radar diagrams, where the ideal city reaches the score 100 on any indicator.

	Amsterdam	**New York**
Economic dimension		
City product per person	91.27	100
Environmental dimension		
Use of private car	16.6	30.7
Solid waste generated per capita	50	32.78
Social dimension		
Houses owned	4.35	20.39
Residential density	47.33	100
Mean travel time to work	100	60.27
Income disparity	100	37.47
Households below poverty line	73.17	92.02
Crime rate	2.98	7.58

Table 7. Benchmarking Exercise by Using the Distance from the Leader Method

According to these computations, Amsterdam reaches the ideal values on mean travel time to work and income disparity thus these issues can be considered solved. Households below poverty line and city product per person are not a problem. All the issues connected to the other indicators are, on the contrary, a problem for this city. Their policy relevance is high and their solution should be urgent (above all, the crime rate).

New York is doing perfectly on city product per person and residential density (where it meets perfectly the ideal values), more or less well on households below poverty line and mean travel time to work (where it is not so far from the ideal values) and more or less bad on the other values where it is definitely worse than the ideal values used, and as a consequence, in our hypothetical situation, the issues connected with these indicators should be considered important policy priorities.

In reality, one should note that these ideal values depend on the cities we are comparing. In this case, the 4 cities are so heterogeneous that probably their comparison is meaningless. Moreover again the issue of information quality applies: who does really believe that residential density is not a problem in New York?

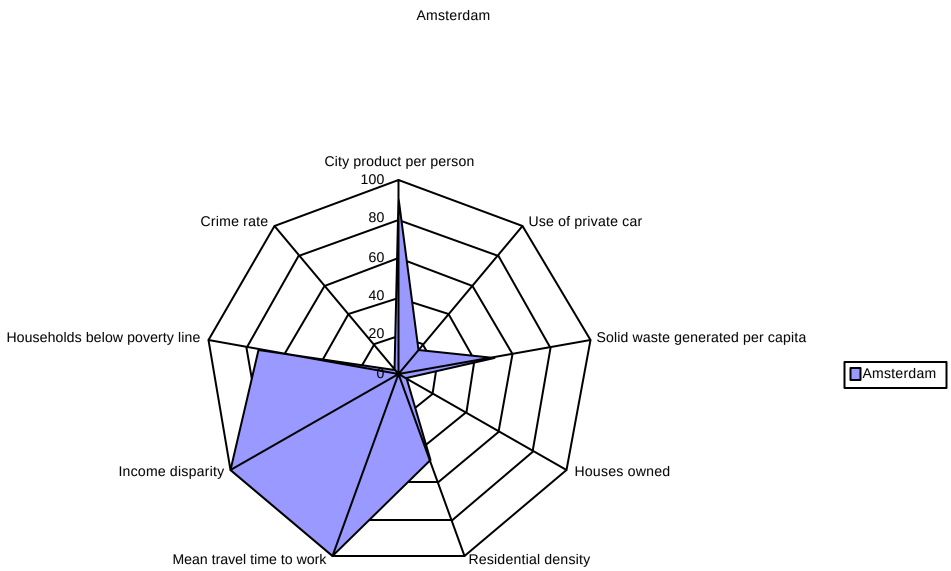

Figure 1. Radar Diagram for Amsterdam Sustainability Benchmarking

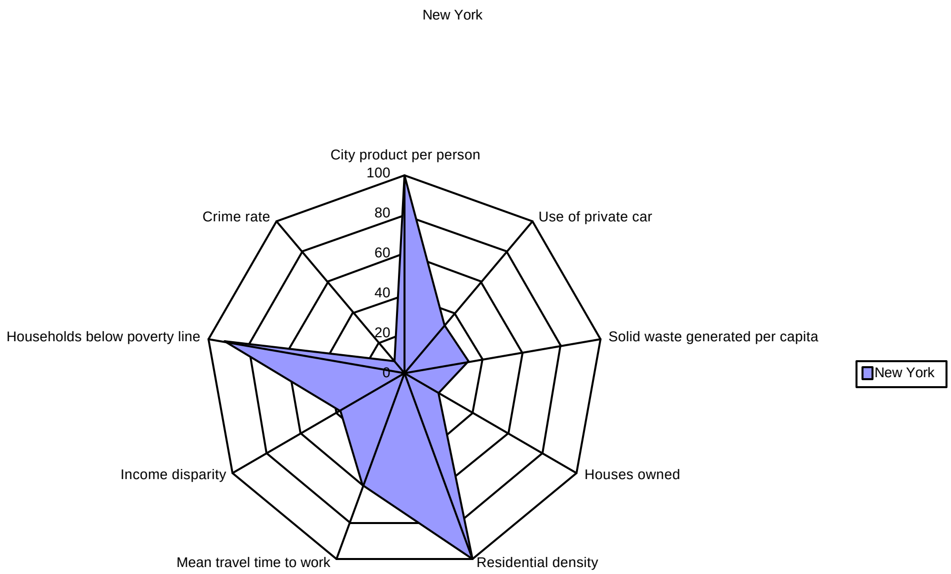

Figure 2. Radar Diagram for New York Sustainability Benchmarking

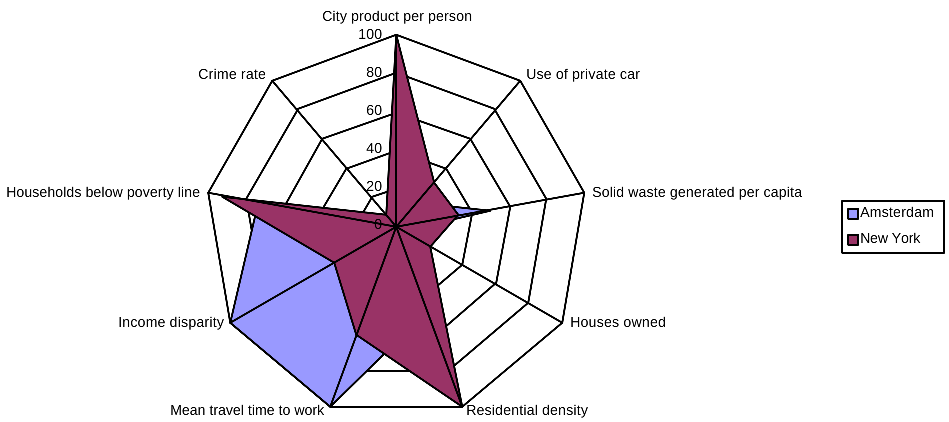

Figure 3. Radar Diagram Comparing Amsterdam and New York

Of course the distance from the group leader method is a very simple approach. A more sophisticated approach for benchmarking can be one based on the notion of a fuzzy preference relation (Munda, 1995; Munda, forthcoming). When dealing with sustainability indicators an essential step is the definition of sustainability ideal values and as a consequence the concept of distance from these sustainability ideal values has to be dealt with. This is a typical example of a fuzzy concept (Zadeh, 1965), i.e. what exactly means close or distant?

In this framework, an interesting concept is the one of a *fuzzy preference relation* (Kacprzyk and Roubens, 1988). If A is assumed to be a finite set of N alternatives, a *fuzzy preference relation* is an element of the $N \times N$ matrix $R = (r_{jk})$, i.e.

$$R_{jk} = \mu_R(a_j, a_k) \text{ with } j,k = 1, 2, \ldots, N \quad \text{and} \quad 0 \le r_{jk} \le 1. \tag{8}$$

$r_{jk} = 1$ indicates the maximum credibility degree of preference of a_j over a_k; each value of r_{jk} in the open interval $(0.5, 1)$ indicates a definite preference of a_j to a_k (a higher value means a stronger credibility); $r_{jk} = 0.5$ indicates the indifference between a_j and a_k. This definition implies that fuzzy preference relations can be used as mathematical models of intensity of preference. This is crucial in our case, since the concept of distance is based on the idea of intensity (i.e. how much distant?).

These concepts can be adapted in the scope of sustainability indicators benchmarking as follows.

Let's consider a given city, region or country X we would like to bring into play for the sustainability assessment exercise and a set of sustainability indicators G. Let's assume that the measurement of the indicator scores is based on an *interval or ratio* scale, for simplicity of exposition, here the assumption is made that a higher value of an indicator is preferred to a lower one (the higher, the better). For each indicator g_m it is established a sustainability ideal value S_m to be met. Let p_{1m} and p_{2m} be constant preference thresholds and q_{1m} and q_{2m} constant indifference thresholds for the indicator g_m.

Then the credibility degree μ of the fuzzy relations close, very close, distant and very distant between X and the various S_m can be computed by using equations showed in (9).

$$\begin{cases} \mu_{(\text{verydistant})} = \left[1 + p_{1m}\left(S_m - g_m(X)\right)^{-2}\right]^{-2} \\ \mu_{(\text{distant})} = \left[1 + p_{2m}\left(S_m - g_m(X)\right)^{-2}\right]^{-1} \\ \mu_{(\text{veryequal})} = e^{-q_{1m}\left(S_m - g_m(X)\right)^2} \\ \mu_{(\text{equal})} = e^{-q_{2m}\left|S_m - g_m(X)\right|} \end{cases} \tag{9}$$

However one has to admit that often more mathematical sophistication does not change real-world policy suggestions significantly. When time constraints exist or the distance measurements are clear cut values, a method such as the distance from the group leader can be used with no particular analytical cost. On the contrary issues such as quality of the information used and the method adopted for deriving ideal values (e.g. comparing homogeneous cities or using European averages for the benchmark of a city in Europe (Giampietro and Mauymi, 2000)) are always influencing results heavily.

5. Conclusion

There is no doubt that there is a lot of complexity and fuzziness inherent in the sustainability concept. A possible reduction of this complexity, a pre-condition for management and planning actions, introduces the problem of the descriptors used: indicators and indices. Often, some indicators improve while others deteriorate. This is the classical conflictual situation dealt with in multi-criteria decision theory; in particular non-compensatory methods are quite relevant, since compensability implies substitutability between different types of capital. The defence of the use of a multi-criterion framework for the assessment (by assessment here is meant the ranking of Countries, cities or regions and their benchmarking) of sustainability is the main argument of this paper.

One should note that the construction of any assessment exercise depends on very strong assumptions about (1) the *purpose* of this construction, e.g. to evaluate the sustainability of a given city, (2) the *scale* of analysis, e.g. a block inside a city, the administrative unit constituting a Commune or the whole metropolitan area and (3) the set of dimensions (economic, social, environmental etc.), objectives and indicators used for the evaluation process. A reductionist approach can be defined as the use of just *one measurable indicator* (e.g. the monetary city product per person), *one dimension* (e.g. economic), *one scale of analysis* (e.g. the Commune), *one objective* (e.g. the maximisation of economic efficiency) and *one time horizon*. All these issues have to be taken into account when we state that a given city is "better" than another one.

A sustainability policy exercise implies difficult decisions such as the choice of indicators, their policy prioritization and the choice of ideal values; such an exercise is not a technical issue only, it is mainly a socio-political issue. Behind a list of indicators and a list of targets there would always be a history of scientific research and political controversy. When science is used in policy, the appropriate management of quality has to be enriched to include this multiplicity of participants and perspectives. The criteria of quality in this new context will presuppose ethical principles. But in this case, the principles will be explicit and will become part of the dialogue (Funtowicz and Ravetz, 1994).

In the context of sustainability assessment the concept of Social Multi-Criteria Evaluation (SMCE) (Munda, 2003) is then very relevant. SMCE principles can be synthesised as follows (see Figure 4):

- The use of a multi-criteria framework is a very efficient tool to implement a *multi/inter-disciplinary* approach.
- Science for policy implies a *responsibility* of the scientists towards the whole society and not just towards a mythical decision-maker.
- *Public participation* is a necessary component but not a sufficient one. Participation techniques are a tool for improving the knowledge of the problem at hand and not for receiving inputs to be used uncritically in the evaluation process. Social participation does not imply lack of responsibility.

- *Ethical judgments* are unavoidable components of the evaluation exercise. These judgments influence heavily the results. As a consequence, *transparency* on the assumptions used is essential.
- In this framework, mathematical aggregation conventions play a significant role, i.e. to assure that the rankings obtained are *consistent* with the information and the assumptions used.

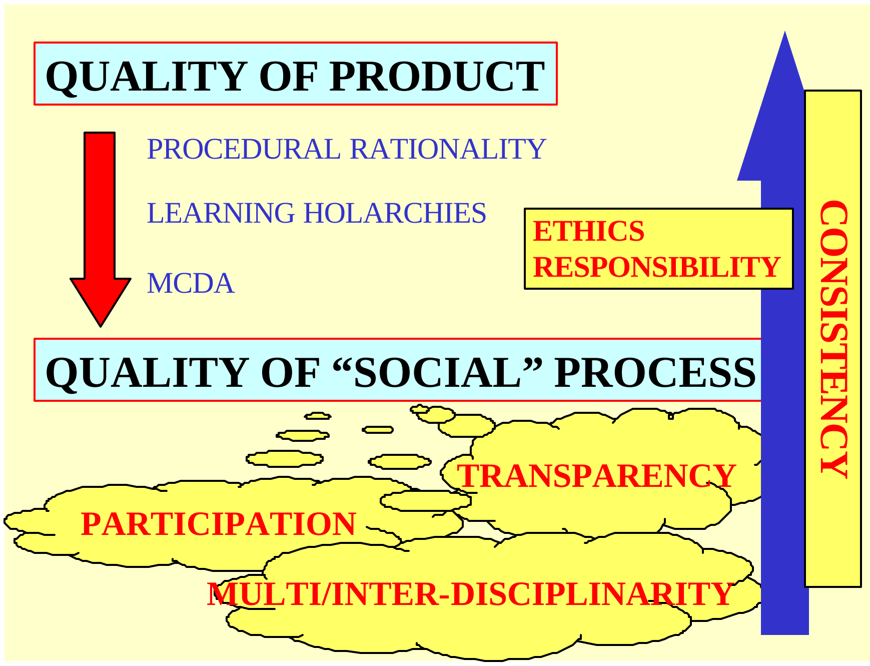

Figure 4. Synthesis of a Social Multi-Criteria Evaluation Process

ACKNOWLEDGEMENTS

Comments by the anonymous referees are gratefully acknowledged. This research has been partly financed by the Autonomous Government of Catalonia (Generalitat de Catalunya) under the contract *“Emmarcar i coordinar l’actuació en la investigació d’una proposta metodològica per a la construcció d’índex de sostenibilitat per a Catalunya”*.

REFERENCES

Allen T.F.H., Tainter J.A., Hoekstra T.W. (2002), *Supply-side sustainability*, Columbia University Press, New York.

Arrow K.J., Raynaud H. (1986) - *Social choice and multicriterion decision making*, M.I.T. Press, Cambridge.

Barbier E.B., Markandya A, (1990) -The conditions for achieving environmentally sustainable growth, *European Economic Review*, 34, pp. 659-669.

Barbiroli G. (1993) – Direct and indirect indicators of sustainable development, Proceedings International Conference on Environmental Pollution, Barcelona, pp. 489-496.

Beinat E. Nijkamp P. (eds.) (1998) - *Multicriteria evaluation in land-use management: methodologies and case studies*, Kluwer, Dordrecht.

Chichilnisky G. (1996) – An axiomatic approach to sustainable development, *Social Choice and Welfare*, 13(2), pp. 219-248.

Daly H.E., Cobb J.J. (1989), *For the common good: redirecting the economy toward community, the environment and a sustainable future*, Beacon Press, Boston.

Faucheux S. and O'Connor M. (eds.) (1998) - *Valuation for sustainable development: Methods and policy indicators*, Edward Elgar.

Funtowicz S.O., Ravetz J.R. (1990) - *Uncertainty and quality in science for policy,* Kluwer Academic Publishers, Dordrecht.

Funtowicz S.O., Munda G., Paruccini M. (1990) - The aggregation of environmental data using multicriteria methods, *Environmetrics*, Vol. 1(4), pp. 353-368.

Funtowicz S.O., Ravetz J. R. (1994) - The worth of a songbird: ecological economics as a post-normal science, *Ecological Economics*, 10, pp. 197-207.

Funtowicz S., Martinez-Alier J., Munda G. and Ravetz J. (1999) - Information tools for environmental policy under conditions of complexity, *European Environmental Agency, Experts' Corner, Environmental Issues Series, No. 9.*

Funtowicz S.O., Martinez-Alier J., Munda G., Ravetz J. (2002) – Multicriteria-based environmental policy in H. Abaza and A. Baranzini (eds.) – *Implementing sustainable development,* UNEP/Edward Elgar, Cheltenham, pp. 53-77.

Giampietro, M. (1994) - Using hierarchy theory to explore the concept of sustainable development. *Futures* 26 (6): 616-625.

Giampietro M., Mayumi K. (2000) – Multiple-scale integrated assessment of societal metabolism: introducing the approach, *Population and Environment,* Vol. 22, No. 2, pp. 109-154.

Giampietro M., Mayumi K. (2000) – Multiple-scale integrated assessment of societal metabolism: integrating biophysical and economic representations across scales, *Population and Environment,* Vol. 22, No. 2, pp. 155-210.

Horwarth R., **Norgaard R.B.** (1990) – Intergenerational resource rights, efficiency and social optimality, *Land Economics*, 66, pp. 1-11.

Horwarth R., **Norgaard R.B.** (1992) - Environmental valuation under sustainable development, *American Economic Review Papers and Proceedings*, 80, pp. 473-477.

Kacprzyk J., Roubens M. (eds.) (1988) – *Non-conventional preference relations in decision making,* Springer-Verlag, Heidelberg.

Munda G., Nijkamp P. and Rietveld P. (1995) - Qualitative multicriteria methods for fuzzy evaluation problems: an illustration of economic-ecological evaluation, *European Journal of Operational Research,* 82, pp.79-97.

Munda G. (1995) - *Multicriteria evaluation in a fuzzy environment,* Series: Contributions to economics, Physica-Verlag, Heidelberg.

Munda G. (1997a) – Environmental economics, ecological economics and the concept of sustainable development, *Environmental Values,* vol. 6, No. 2, pp. 213-233.

Munda G. (1997b) - Multicriteria evaluation as a multidimensional approach to welfare measurement, in J. van den Bergh and J. van der Straaten (eds.), *Economy and ecosystems in change: analytical and historical approaches*, Edward Elgar, Cheltenham, pp. 96 - 115.

Munda G. (2003) – "Social multi-criteria evaluation (SMCE)": methodological foundations and operational consequences, *European Journal of Operational Research.*

Munda G. and Nardo M. (2003) – Mathematical modelling of composite indicators for ranking countries, *Proceedings of the First OECD/JRC Workshop on Composite Indicators of Country Performance,* JRC, Ispra.

Munda G. (forthcoming) – A NAIADE based approach for sustainability benchmarking, *International Journal of Environment Technology and Management.*

Musu I., Siniscalco D. (1996) – *National accounts and the environment,* Kluwer Academic Publishers, Dordrecht.

OECD (2003) – *Composite indicators of country performance: a critical assessment*, DST/IND(2003)5, Paris.

Pearce D.W., Atkinson G.D. (1993) - Capital theory and the measurement of sustainable development: an indicator of "weak" sustainability, *Ecological Economics*, vol. 8, pp. 103-108.

Pearce D, Hamilton G., Atkinson G.D. (1996) – Measuring sustainable development: progress on indicators, *Environment and Development Economics,* 1, pp. 85-101.

Podinovskii V.V. (1994) -Criteria importance theory, *Mathematical Social Sciences,* 27, pp. 237 - 252.

Roberts F. S. (1979) - *Measurement theory with applications to decision making, utility and the social sciences*, Addison-Wesley, London.

Roy B. (1996) - *Multicriteria methodology for decision analysis*, Kluwer, Dordrecht.

Saisana M. and Tarantola S. (2002) – *State-of-the-art report on current methodologies and practives for composite indicator development*, EUR 20408 EN Report, European Commission, JRC, Ispra, Italy.

Simon H.A. (1976) - From Substantive to precedural rationality. In: J.S. Latsis (Editor) *Methods and appraisal in economics*, Cambridge: Cambridge University Press

Schmidt-Bleek, F. (1994) - *Wieviel umwelt braucht der mensch? MIPS Das Mass für ökologisches wirtschaften*, Birkh user, Berlin.

Urban Indicators Programme, http://www.urbanobservatory.org/indicators/database

Vitousek P., Ehrlich P., Ehrlich A. and Matson P. (1986) - Human appropriation of the products of photosynthesis, *Bioscience*, 34(6): 368 373.

Wackernagel, M. and W. E. Rees (1995) - *Our ecological footprint: Reducing human impact on the earth,* Gabriola Island, BC and Philadelphia, PA: New Society Publishers.

Yu P. L. (1985) - *Multi criteria decision making: concepts, techniques and extensions,* Plenum Press, New York.

Yusuf, J.A., S. El Serafy and E. Lutz (1989) - *Environmental accounting for sustainable development,* A UNEP World Bank Symposium, Washington D.C.

Zadeh L.A. (1965) - Fuzzy sets, *Information and Control,* 8, pp. 338-353.

Zeleny M. (1982) - *Multiple criteria decision making*, McGraw Hill, New York.