J. R. Statist. Soc. A (2005)
168, *Part* 2, *pp.* 307–323

Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators

M. Saisana, A. Saltelli and S. Tarantola

European Commission, Ispra, Italy

[Received November 2002. Final revision July 2004]

Summary. Composite indicators are increasingly used for bench-marking countries' performances. Yet doubts are often raised about the robustness of the resulting countries' rankings and about the significance of the associated policy message. We propose the use of uncertainty analysis and sensitivity analysis to gain useful insights during the process of building composite indicators, including a contribution to the indicators' definition of quality and an assessment of the reliability of countries' rankings. We discuss to what extent the use of uncertainty and sensitivity analysis may increase transparency or make policy inference more defensible by applying the methodology to a known composite indicator: the United Nations's technology achievement index.

Keywords: Composite indicators; Robustness assessment; Sensitivity analysis; Uncertainty

1. Introduction

Composite indicators (often called indices) are increasingly used by statistical offices and national or international organizations to convey information on the status of countries in fields such as the environment, economy, society or technological development; a review is given in Saisana and Tarantola (2002). Composite indicators are calculated by combining well-chosen subindicators into a single index, on the basis of an underlying model of the policy domain that one wishes to measure. This is most often achieved by a weighted combination of normalized subindicators' values.

The main pros and cons of using composite indicators have been debated within the services of the European Commission. The discussion is summarized by Saisana and Tarantola (2002), from which we quote:

'**Pros**
—Composite indicators can be used to **summarise complex or multi-dimensional issues**, in view of supporting decision-makers.
—Composite indicators provide the **big picture**. They can be easier to interpret than trying to find a trend in many separate indicators. They facilitate the task of ranking countries on complex issues.
—Composite indicators can help **attracting public interest** by providing a summary figure with which to compare the performance across countries and their progress over time.
—Composite indicators could help to **reduce the size** of a list of indicators **or to include more information** within the existing size limit.

Address for correspondence: M. Saisana, Joint Research Centre, European Commission, Enrico Fermi 1, 21020 Ispra, TP 361, Italy.
E-mail: michaela.saisana@jrc.it

 0964–1998/05/168307

'**Cons**

—Composite indicators **may send misleading, non-robust policy messages** if they are poorly constructed or misinterpreted. Sensitivity analysis can be used to test composite indicators for robustness.

—The simple "big picture" results which composite indicators show may invite politicians to draw **simplistic policy conclusions**. Composite indicators should be used in combination with the sub-indicators to draw sophisticated policy conclusions.

—The construction of composite indicators involves stages where **judgement** has to be made: the selection of sub-indicators, choice of model, weighting indicators and treatment of missing values etc. These judgements should be transparent and based on sound statistical principles.

—There could be **more scope for Member States** about composite indicators than on individual indicators. The selection of sub-indicators and weights could be the target of political challenge.

—The composite indicators increase the **quantity of data** needed because data are required for all the sub-indicators and for a statistically significant analysis.'

We shall tackle in the present paper many of the points that are raised by these conclusions. In practice, it is difficult to imagine that the debate on the use of composite indicators will ever be settled. Just to give an example that is linked to our experience, official statisticians may tend to resent composite indicators, whereby a large amount of work in data collection and editing is 'wasted' or 'hidden' behind a single number of dubious significance. However, the temptation of stakeholders and practitioners to summarize complex and sometime elusive processes (e.g. sustainability or a single-market policy) into a single figure to bench-mark country performance for policy consumption seems likewise irresistible.

General procedures to assess uncertainty in building composite indicators are described in this paper. In particular, we limit ourselves to three types of uncertainties:

(a) alternative normalization methods for the values of the subindicators,
(b) alternative weighting approaches and finally
(c) uncertainty in the weights of the subindicators.

Two combined tools are suggested: uncertainty analysis (UA) and sensitivity analysis (SA). UA focuses on how uncertainty in the input factors propagates through the structure of the composite indicator and affects the values of the composite indicator. SA studies how much each individual source of uncertainty contributes to the output variance.

In the field of building composite indicators, UA is more often adopted than SA (Jamison and Sandbu, 2001; Freudenberg, 2003) and the two types of analysis are almost always treated separately. A synergistic use of UA and SA is proposed and presented in Section 2, considerably extending earlier attempts in this direction (Tarantola *et al.*, 2000). The test case, which is described in Section 3, is the technology achievement index (TAI) that was developed by the United Nations (2001) to help policy makers to define technology strategies. Two normalization methods and two participatory approaches for assigning weights to the subindicators (budget allocation (BA) and analytic hierarchy processes (AHPs)) have been applied in the present work. Section 4 discusses the results and aims to answer two key questions on the quality of the composite indicator.

(a) Does the use of one normalization method and one set of weights in the development of the composite indicator (e.g. the original TAI) provide a biased picture of the countries' performance?
(b) To what extent do the uncertain input factors (normalization methods, weighting schemes and weights) affect the countries' ranks with respect to the original TAI?

Section 5 summarizes our conclusions on the role that the combination of UA and SA can play as a quality assurance tool during the development of a composite indicator for policy making.

2. Methodological aspects in building composite indicators

Several methods for calculating composite indicators from a set of subindicators were described in Saisana and Tarantola (2002). The methods that are most frequently met in the literature are based on the rescaled values (equation (1a)) or on the standardized values (equation (1b)). A composite indicator Y_c for a given country c is most often a simple linear weighted function of a total of Q normalized subindicators $I_{q,c}$ with weights w_q,

$$Y_c = \sum_{q=1}^{Q} I_{q,c} w_q, \qquad \text{where} \begin{cases} I_{q,c} = \dfrac{x_{q,c} - \min(x_q)}{\text{range}(x_q)}, & \text{(1a)} \\[2ex] I_{q,c} = \dfrac{x_{q,c} - \text{mean}(x_q)}{\text{std}(x_q)}. & \text{(1b)} \end{cases}$$

Here $x_{q,c}$ represents the raw value of the subindicator x_q for country c.

The difference in the values of the composite indicator between two countries A and B will be an output of interest studied in our UA–SA.

$$D_{\text{AB}} = \sum_{q=1}^{Q} (I_{q,\text{A}} - I_{q,\text{B}}) w_q. \tag{2}$$

Additionally, the average shift in countries' ranks will be explored. This statistic captures in a single number the relative shift in the position of the entire system of countries. It can be quantified as the average of the absolute differences in countries' ranks with respect to a reference ranking over the M countries:

$$\bar{R}_{\text{S}} = \frac{1}{M} \sum_{c=1}^{M} |\text{rank}_{\text{ref}}(Y_c) - \text{rank}(Y_c)|. \tag{3}$$

The investigation of the Y_c, D_{AB} and $\bar{R}_{\text{S}}$ will be the scope of the UA and SA, targeting the questions that were raised in Section 1 on the quality of the composite indicator.

2.1. *Uncertainty analysis*

In the first stage, the uncertain input factors in the estimation of the outputs Y_c, D_{AB} and $\bar{R}_{\text{S}}$ must be acknowledged. In general, the uncertainties in the development of a composite indicator will arise from some or all of the steps in the construction line:

(a) selection of subindicators,
(b) data selection,
(c) data editing,
(d) data normalization,
(e) weighting scheme,
(f) weights' values and
(g) composite indicator formula.

The most debated problem in building composite indicators is the difficulty in assessing properly the plurality of perspectives about the relative importance of the subindicators. Experience shows that disputes over the appropriate method of establishing weights cannot be easily resolved. Cox *et al.* (1992) summarized the difficulties that are commonly encountered when proposing weights to combine indicators into a single measure, and they concluded that many published weighting schemes are either based on too complex multivariate methods or have little meaning to society. For these reasons, participatory approaches, such as BA or AHPs, are often preferred, as they allow for an expression of the relative importance of the subindicators from the societal viewpoint. In BA experts are invited to distribute a budget of points over a number of subindicators, paying more for those indicators whose importance they want to emphasize (Moldan *et al.*, 1997). The AHP is a widely used technique for multiattribute decision-making (Saaty, 1980, 1987). The AHP is based on ordinal pairwise comparisons of subindicators. For a given objective, the comparisons are made per pairs of subindicators, and the strength of preference is expressed on a semantic scale of 1 (equality) to 9 (i.e. a subindicator can be voted to be nine times more important than the subindicator with which it is being compared). The relative weights of the subindicators are calculated by using an eigenvector technique, which as described in Saaty (1980) serves to resolve inconsistencies (e.g. a better than b better than c better than a loops). Note that the AHP approach would also allow the analyst to 'grade' the expert by measuring his or her inconsistencies.

In this work, we focus on three points of the (a)–(g) chain of composite indicator building, which can introduce uncertainty in the output variables Y_c, D_{AB} and $\bar{R}_{\mathrm{S}}$: the type of normalization (be it with rescaled (equation (1a)) or standardized values (equation (1b))), the weighting scheme (be it BA or AHP) and the subindicators' weights. We anticipate here that we shall translate all these uncertainties into a set of scalar input factors, to be sampled from their distributions. As a result, all outputs Y_c, D_{AB} and $\bar{R}_{\mathrm{S}}$ are non-linear functions of the uncertain input factors, and the estimation of the probability distribution functions (PDFs) of Y_c, D_{AB} and $\bar{R}_{\mathrm{S}}$ is the purpose of the UA. The UA procedure is essentially based on simulations that are carried out on each of equations (1)–(3), termed henceforth the *model*. Various methods are available for evaluating output uncertainty.

In the following, the Monte Carlo approach is presented, which is based on performing multiple evaluations of the model with k randomly selected model input factors. The procedure involves six steps.

(a) Assign a PDF to each input factor X_i. The first input factor, X_1, is the trigger to select the type of normalization method; the second input factor, X_2, is the trigger to select the weighting scheme. Factors X_3–X_k are random numbers that are used to select the $Q\,(=k-2)$ uncertain weights.
(b) Generate randomly N combinations of independent input factors $\mathbf{X}^l$, with $l = 1, \ldots, N$ (a set $\mathbf{X}^l = X_1^l, X_2^l, \ldots, X_k^l$ of input factors is called a sample).
(c) For each sample l, select a normalization method and weighting scheme based on X_1, $X_2, \ldots$.
(d) For each sample l, use factors X_3–X_k to select the weights.
(e) Evaluate the model, i.e. by computing the output value Y^l, where Y^l is either Y_c, the value of the composite indicator for each country, or D_{AB}, the difference between two countries, or $\bar{R}_{\mathrm{S}}$, the averaged shift in countries' ranks.
(f) Close the loop over l, and analyse the resulting output vector $\mathbf{Y}^l$, with $l = 1, \ldots, N$.

The generation of samples can be performed by using various procedures, such as simple random sampling, stratified sampling, quasi-random sampling or others (Saltelli, Chan and Scott, 2000). The sequence of $\mathbf{Y}^l$ allows the empirical PDF of the output Y to be built. The characteristics of the PDF, such as the variance and higher order moments, can be estimated with an arbitrary level of precision that is related to the size of the simulation N.

2.2. Sensitivity analysis using variance-based techniques

A necessary step when designing an SA is to identify a few summary variables that describe concisely, yet exhaustively, the message that is provided by a model (Saltelli, Tarantola and

Campolongo, 2000). In the present application for instance, although an uncertainty analysis of Y_c is necessary, SA is only applied to the two summary model outputs, D_{AB} and $\bar{R}_{\text{S}}$, as they are relevant to the quality assessment of a composite indicator.

For non-linear models, such as our composite indicator happens to be when normalization methods, weighting schemes and weights are all sampled, variance-based techniques for SA are the most appropriate (Saltelli, Tarantola and Campolongo, 2000). The importance of a given input factor X_i can be measured via the so-called *sensitivity index*, which is defined as the fractional contribution to the model output variance due to the uncertainty in X_i. For k independent input factors, the sensitivity indices can be computed by using the following decomposition formula for the total output variance $V(Y)$ of the output Y:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \ldots + V_{12\ldots k} \tag{4}$$

where

$$V_i = V_{X_i}\{E_{\mathbf{X}_{-i}}(Y|X_i)\}, \tag{5}$$

$$V_{ij} = V_{X_i X_j}\{E_{\mathbf{X}_{-ij}}(Y|X_i, X_j)\} - V_{X_i}\{E_{\mathbf{X}_{-i}}(Y|X_i)\} - V_{X_j}\{E_{\mathbf{X}_{-j}}(Y|X_j)\} \tag{6}$$

and so on. In computing $V_{X_i}\{E_{\mathbf{X}_{-i}}(Y|X_i)\}$, the expectation $E_{\mathbf{X}_{-i}}$ would call for an integral over $\mathbf{X}_{-i}$, i.e. over all factors except X_i, including the marginal distributions for these factors, whereas the variance V_{X_i} would imply a further integral over X_i and its marginal distribution. A first measure of the fraction of the unconditional output variance $V(Y)$ that is accounted for by the uncertainty in X_i is the first-order sensitivity index for the factor X_i defined as

$$S_i = V_i / V. \tag{7}$$

Terms above the first term in equation (4) are known as interactions. A model without interactions among its input factors is said to be additive. In this case, $\Sigma_{i=1}^k S_i = 1$ and the first-order conditional variances of equation (5) are all that we need to know to decompose the model output variance. For a non-additive model, higher order sensitivity indices, which are responsible for interaction effects among sets of input factors, must be computed. However, higher order sensitivity indices are usually not estimated, as in a model with k factors the total number of indices (including the S_is) that should be estimated is as high as $2^k - 1$. For this reason, a more compact sensitivity measure is used. This is the total effect sensitivity index, which concentrates in one single term all the interactions involving a given factor X_i. To exemplify, for a model of $k = 3$ independent factors, the three total sensitivity indices would be

$$S_{T1} = \frac{V(Y) - V_{X_2 X_3}\{E_{X_1}(Y|X_2, X_3)\}}{V(Y)} = S_1 + S_{12} + S_{13} + S_{123}. \tag{8}$$

Analogously:

$$\begin{aligned} S_{T2} &= S_2 + S_{12} + S_{23} + S_{123}, \\ S_{T3} &= S_3 + S_{13} + S_{23} + S_{123}. \end{aligned} \tag{9}$$

The conditional variance in equation (8) can be written generally as $V_{\mathbf{X}_{-i}}\{E_{X_i}(Y|\mathbf{X}_{-i})\}$ (Homma and Saltelli, 1996). It expresses the total contribution to the variance of Y due to non-X_i, i.e. to the $k-1$ remaining factors, so that $V(Y) - V_{\mathbf{X}_{-i}}\{E_{X_i}(Y|\mathbf{X}_{-i})\}$ includes all terms, i.e. a first-order term as well as interactions in equation (4), that involve factor X_i. In general $\Sigma_{i=1}^k S_{Ti} \geqslant 1$, with equality if there are no interactions. For a given factor X_i a notable difference between S_{Ti} and S_i flags an important role of interactions for that factor in Y. Highlighting interactions between input factors helps us to improve our understanding of the model structure. Estimators for both S_i and S_{Ti} are provided by a variety of methods reviewed in Chan *et al.* (2000).

Here the method of Sobol' (1993), in its improved version due to Saltelli (2002), is used. The method of Sobol' uses quasi-random sampling of the input factors. The pair (S_i, S_{Ti}) give a fairly good description of the model sensitivities at a reasonable cost, which for the improved Sobol' method is of $2n(k+1)$ model evaluations, where n represents the sample size that is required to approximate the multidimensional integration implicit in equation (5) to a plain sum. n can vary in the 100–1000 range.

The extended variance-based techniques that have been described so far display several attractive features, such as

(a) model independence (which is appropriate for non-linear and non-additive models),
(b) exploration of the whole range of variation in the input factors, instead of just sampling factors over a limited number of values, as done for example in fractional factorial designs (Box *et al.*, 1978), and
(c) consideration of interaction effects.

When the uncertain input factors X_i are dependent, the output variance cannot be decomposed as in equation (4). The S_i- and S_{Ti}-indices are still valid sensitivity measures for X_i, though their interpretation changes as, for example, S_i carries over also the effects of other factors that can be positively or negatively correlated to X_i (see Saltelli and Tarantola (2002)). Estimation procedures are offered in Saltelli *et al.* (2004).

The extended variance-based methods, including the improved version of Sobol', for both dependent and independent input factors, are implemented in the freely distributed software SIMLAB (Saltelli *et al.*, 2004).

3. Case-study: technology achievement index

The TAI is a composite indicator that was developed by the United Nations and was described in detail in United Nations (2001). The report argues that development strategies need to be redefined in the network age and calls on policy makers to take a new look at the current technological achievements. The TAI composite indicator should help a country to assess its position relative to others, possibly to bench-mark policies. Although acknowledging that many elements make up a country's technological achievement, the index suggests that an overall assessment is more easily made on the basis of a single composite measure. Like other composite indices in the *Human Development Report* series, such as the human development index, the TAI is suggested for summary purposes, to be followed by individual analysis of the underlying indicators. The design of the index reflects two particular concerns. The first is to compare policy effectiveness across all countries, regardless of the level of technological development. The second is to identify, and to send messages to, developing countries. To accomplish this, the index must be able to discriminate between countries at the lower end of the range. The TAI focuses on four dimensions of technological capacity.

(a) Creation of technology: two subindicators are used to capture the level of innovation in a society, the number of patents granted per capita (to reflect the current level of invention activities) and the receipts of royalty and licence fees from abroad per capita (to reflect the stock of successful innovations of the past that are still useful and hence have market value).
(b) Diffusion of recent innovations: all countries must adopt innovations to benefit from the opportunities of the network age. This diffusion is measured by two subindicators: diffusion of the Internet (which is indispensable to participation) and by exports of high and medium technology products as a share of all exports.

(c) Diffusion of old innovations: participation in the network age requires diffusion of many old innovations, as technological advance is a cumulative process, and widespread diffusion of older innovations is necessary for the adoption of later innovations. Two subindicators are included here, telephones and electricity, which are especially important because they are needed to use newer technologies and are also pervasive inputs to a multitude of human activities. Both indicators are expressed as logarithms, as they are important at the earlier stages of technological advance but not at the most advanced stages. Expressing the measure in logarithms ensures that, as the level increases, it contributes less to the index.

(d) Human skills: a critical mass of skills is indispensable to technological dynamism. The foundations of such ability are basic education to develop cognitive skills and skills in science and mathematics. Two subindicators are used to reflect the human skills that are needed to create and absorb innovations: mean years of schooling and gross enrolment ratio of tertiary students enrolled in science, mathematics and engineering.

The data that are used to construct the TAI are from international series that are most widely used in analyses of trends in technology and so are considered the most reliable of the available sets (United Nations (2001), page 46). For each of the eight subindicators (two subindicators per dimension) the observed minimum and maximum values (among all available countries) are chosen as 'goalposts' (Table 1) and performance in each subindicator is expressed as a value between 0 and 1 by applying equation (1a). The original TAI is calculated for 72 countries, for which data are available and of acceptable quality, as the simple average of the eight subindicators, therefore considering that all subindicators are equally important in the development of the composite indicator. The statistics for the raw data of the subindicators for the 72 countries are given in Table 1. The data for patents, royalties and Internet hosts exhibit the highest coefficient of variation (the ratio of the standard deviation to the mean), underlining the high variability in these three subindicators. The distributions of data for patents and electricity are the most skewed (positively).

Table 1. Statistical properties of the eight subindicators that compose the TAI

Subindicator	*Goalposts for calculating the TAI*		*Statistics across 72 countries*		
	Observed minimum value	*Observed maximum value*	*Mean*	*Standard deviation*	*Skewness*
Patents granted to residents (per million people)	0	994	109	199	3.4
Royalties and licence fees received (US $ per 1000 people)	0.0	272.6	29.3	46.1	1.8
Internet hosts (per 1000 people)	0.0	232.4	32.3	51.1	2.0
High and medium technology *exports* (% of total goods exports)	0.0	80.8	30.1	23.0	0.4
Telephones, main line and cellular (per 1000 people)	1	901	443	402	0.6
Electricity consumption (kW h per capita)	22	6969	3496	4274	2.5
Mean years of *schooling* (age 15 years and above)	0.8	12.0	7.2	2.5	−0.1
Gross tertiary science *enrolment* ratio (%)	0.1	27.4	8.1	6.2	1.0

Correlation analysis reveals that the eight subindicators have an average bivariate correlation of 0.55 and that six pairs of indicators have a correlation coefficient that is higher than 0.70. The covariance of subindicators is further investigated via factor analysis. To account for at least 90% of the variance in the data set of all the subindicators, five principal components are needed. This result indicates that the phenomenon that is described by the set of the eight subindicators is quite multidimensional. A higher correlation between the subindicators would have resulted in fewer components.

Depending on a school of thought, one may see a high correlation between subindicators as something to correct for, e.g. by making the weight for a given subindicator inversely proportional to the arithmetic mean of the coefficients of determination for each bivariate correlation that includes the given subindicator (Saisana and Tarantola, 2002). An example of this approach is the index of relative intensity of regional problems in the European Union (Commission of the European Communities, 1984). Practitioners of multicriteria decision analysis would instead tend to consider the existence of correlations as a feature of the problem, not to be corrected for, as correlated subindicators may indeed reflect non-compensational different features of the problem.

In the present study, we deviate from the original deterministic formulation of the index, in that we allow both the normalization and the weighting procedures to vary, and we sample the weights rather than keep them equal and fixed as in the original TAI. Henceforth, we indicate the extended composite indicator as Monte Carlo TAI (MCTAI). The weights for the eight subindicators have been derived from two pilot surveys that were carried out across 20 informed interviewees at the authors' institute, using the two participatory approaches BA and AHP that were described in Section 2. For the present exercise, the interviewees had no bearing on the issue being measured. In reality, interviewees should be part of a community of stakeholders with a legitimate stake. To give an example, in the internal market composite indicator, the experts were members of a committee representing the European Union member countries whose performance was being measured (Tarantola *et al.*, 2002).

Fig. 1 presents the eight scatterplots of the weights for each subindicator provided by the 20 interviewees using the two alternative weighting schemes. In other words, each point in a scatterplot represents the weight that was given to the subindicator by one interviewee when requested in a BA or an AHP approach. The deviation of the weights from the 45°-line of perfect agreement between the two weighting schemes is an interesting feature of this analysis, revealing the human tendency to reply differently to different formulations of the same question. Both weighting approaches have advantages and limitations. The weights that are provided by BA are less spread than for an AHP for each subindicator and the variance of the weights across the eight indicators is smaller for BA than for AHP. However, the AHP is based on pairwise comparisons, where perception is sufficiently high to make a distinction between subindicators. In BA all the subindicators are compared at a glance, and this might lead to circular thinking across subindicators, creating difficulties in assigning weights, particularly when the number of subindicators is high.

The 10 uncertain input factors in our analysis are described in Table 2 with their associated PDF. The input factors are sampled in a quasi-random sampling scheme (Sobol', 1967) using a base sample of size $n = 512$ (which is required for the computation of the integral; see the discussion in Section 2.2) and the composite indicator values per country are calculated by performing $2n(k+1) = 11264$ simulations. The MCTAI for the 72 countries is calculated as follows: the data for each subindicator are first normalized according to the trigger X_1 that is sampled from a uniform distribution [0, 1], where for $0 \leqslant X_1 \leqslant 0.5$ equation 1(a) is used and for $0.5 < X_1 \leqslant 1$ equation 1(b) is used. The trigger X_2, with the same type of PDF as X_1, guides the

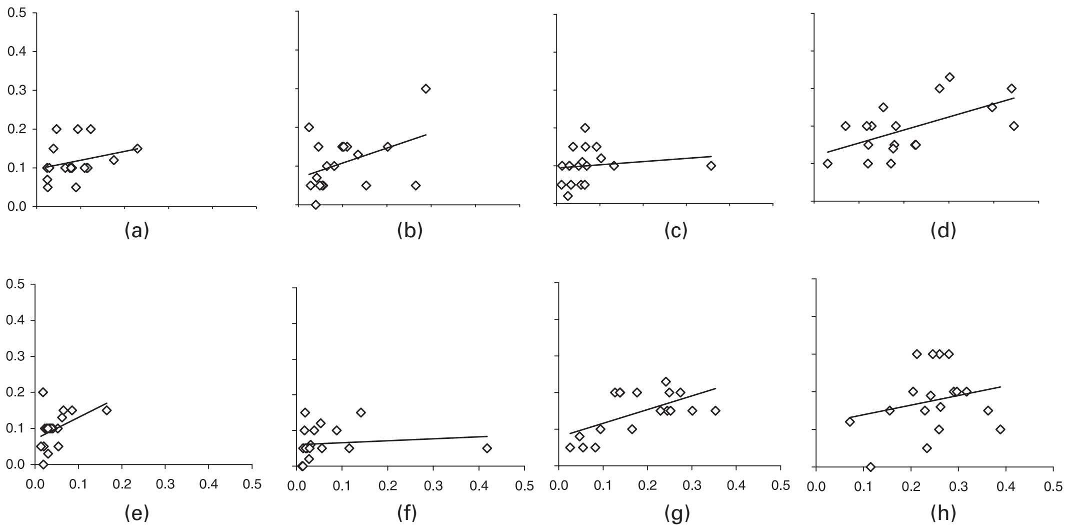

Fig. 1. Scatterplots of weights (range between 0.0 and 0.5) for the eight subindicators of MCTAI (the weights have been derived from pilot surveys of 20 informed interviewees using BA (vertical axes) and AHPs (horizontal axes); ———, best fit linear regression lines): (a) patents; (b) royalties; (c) Internet; (d) exports; (e) telephones; (f) electricity; (g) schooling; (h) enrolment

Table 2. The 10 uncertain input factors for the analysis

Input factor	*Definition*	*PDF*	*Range*
X_1, rescaled or standardized values	Normalization method	Uniform	[0, 1], where [0, 0.5] ≡ rescaled values and (0.5, 1] ≡ standardized values
X_2, BA or AHP	Weighting scheme	Uniform	[0, 1], where [0, 0.5] ≡ BA and and (0.5, 1] ≡ AHP
X_3, W-patents	Weights list for patents	Discrete, uniform	[1, 2, ..., 20]
X_4, W-royalties	Weights list for royalties	Discrete, uniform	[1, 2, ..., 20]
X_5, W-internet	Weights list for Internet	Discrete, uniform	[1, 2, ..., 20]
X_6, W-exports	Weights list for exports	Discrete, uniform	[1, 2, ..., 20]
X_7, W-telephone	Weights list for telephone	Discrete, uniform	[1, 2, ..., 20]
X_8, W-electricity	Weights list for electricity	Discrete, uniform	[1, 2, ..., 20]
X_9, W-schooling	Weights list for schooling	Discrete, uniform	[1, 2, ..., 20]
X_{10}, W-enrolment	Weights list for enrolment	Discrete, uniform	[1, 2, ..., 20]

selection of the weighting scheme. Then the weights are sampled independently of one another as follows. Each factor X_3–X_{10} is sampled from a discrete distribution [1–20], representing the number of the interviewee, whose assigned weight is taken for the simulation. After all the weights have been assigned, they are scaled to a unit sum. The MCTAI is then calculated as the weighted average of the normalized subindicators. Linear scaling is finally applied to the composite indicator values per type of normalization method, to convert them in the range [0, 100].

Note that this assessment of uncertainties covers only points (d)–(f) in the uncertainty propagation chain (a)–(g) that was described in Section 2.1. To give an example, given the lack of information on the uncertainty in the values of the subindicators, the data are considered to be error free for this test case, and are not, therefore, among the input factors in the Monte-

Carlo-based UA. In Tarantola *et al.* (2000) the reader can find an analysis where data, weights and model uncertainties are compounded. One might likewise investigate the inclusion or exclusion of individual subindicators. This was not done explicitly here, though zero weights were occasionally assigned by the experts in the BA approach (Fig. 1).

4. Results

4.1. Uncertainty analysis on the composite indicator values

The MCTAI values for the 72 countries have been calculated for each of the 11 264 combinations of normalization method, weighting scheme and set of weights. Fig. 2 displays the median (black bar) and the corresponding fifth and 95th percentiles (bounds) of the distribution of the MCTAI per country. The crosses indicate the original value of the 2001 TAI.

From this analysis we can conclude the following.

(a) For most countries, the original TAI value is very close to the MCTAI median value of the distribution that acknowledges all three types of uncertainty. This implies that the original TAI, despite being developed by using one normalization method and one set of weights (equal weights), provides a picture of the countries' technological achievements that is not generally biased.
(b) There is, however, one country, the Netherlands, that is strongly favoured in the original TAI. Originally, the performance of the Netherlands follows those of the five best countries, including Finland, the USA, Sweden, Japan and Korea. However, when accounting for changes in the normalization method, the weighting scheme and the set of weights, its performance falls at the ninth position, behind Singapore, the UK and Australia. Singapore, in contrast, is ranked sixth in the MCTAI, instead of 10th in the original TAI.
(c) The United Nations's TAI is not intended to be a measure of which country is leading in global technology development, but to focus on how well the country as a whole is participating in creating and using technology. Take for example Finland ($\mathrm{TAI}_{\mathrm{original}} = 74.4$) and the USA ($\mathrm{TAI}_{\mathrm{original}} = 73.3$). The USA, a global technology powerhouse, has far more inventions and Internet hosts in total than Finland does, although in Finland more is being done to develop a technological skill base throughout the population. When accounting for the uncertainties in the input factors, the message is reinforced, as the distance between the two countries is more evident ($\mathrm{MCTAI}_{\mathrm{median,Finland}} = 83.1$; $\mathrm{MCTAI}_{\mathrm{median,USA}} = 78.1$).

4.2. Sensitivity analysis on selected outputs using Sobol' sensitivity measures

The previous analysis motivates a quantitative estimation of the effect of the variation in the input factors on two output variables of interest:

(a) the difference D_{AB} (equation (2)) in the composite indicator values between two countries (the Netherlands and Singapore are selected owing to the discussion above);
(b) the average shift in countries' ranks, $\bar{R}_{\mathrm{S}}$ (equation (3)), taking the original TAI as the reference.

The sensitivity measures S_i and S_{Ti}, described previously in Section 2.2, have been computed using the same set of 11 264 simulations for the two output variables of interest.

4.2.1. *Output 1: $D_{\mathrm{NL,SG}}$, difference in performance between the Netherlands and Singapore*

The histogram that is displayed in Fig. 3 represents the outcome of the UA on the differences in the composite indicator values between the Netherlands and Singapore (ranks sixth and 10th

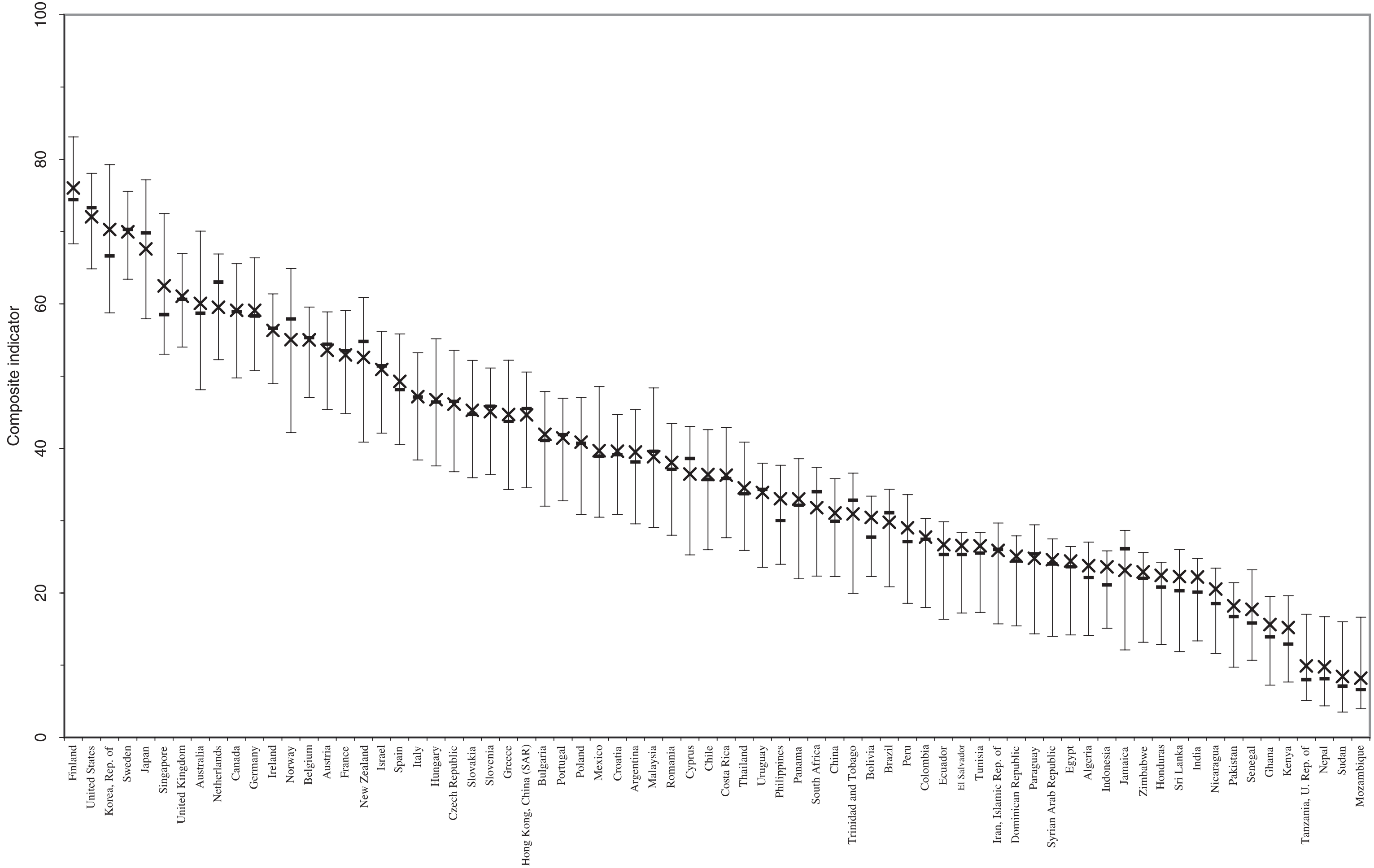

Fig. 2. Results of UA showing the original TAI for 2001 (–), and the median (×) and the corresponding 5th and 95th percentiles (bounds) of the distribution of the MCTAI for 72 countries (uncertain input factors: normalization method, weighting scheme and weights for the subindicators; the countries are ordered according to their median values)

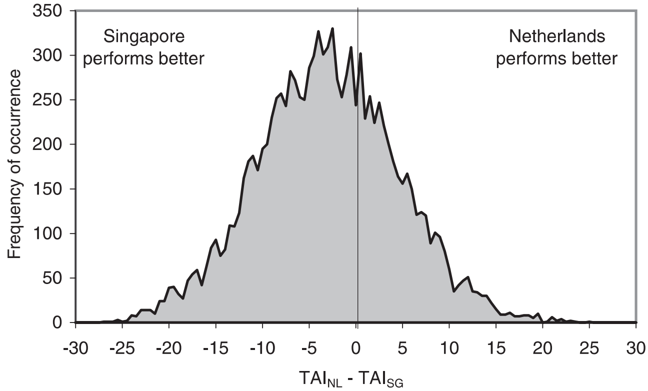

Fig. 3. Results of UA for the output variable $D_{\mathrm{NL,SG}}$: difference in the composite indicator values between the Netherlands and Singapore (uncertain input factors: normalization method, weighting scheme and weights for the subindicators)

Table 3. Sobol' sensitivity measures of the first-order and total effect for the output $D_{\mathrm{NL,SG}}$: difference in the composite indicator values between the Netherlands and Singapore†

Input factor	*First-order effect (S_i)*	*Total effect (S_{Ti})*	$S_{Ti} - S_i$
X_1	0.00	0.03	0.03
X_2	*0.14*	0.56	*0.42*
X_3	0.00	0.00	0.00
X_4	0.01	0.04	0.03
X_5	0.02	0.05	0.03
X_6	0.02	0.17	*0.15*
X_7	0.02	0.07	0.06
X_8	*0.17*	0.37	*0.20*
X_9	0.02	0.10	0.08
X_{10}	*0.12*	0.29	*0.16*
Sum	0.52	1.69	1.17

†Values in italics (greater than 0.10) indicate important input factors based on S_i or $S_{Ti} - S_i$.

in the original TAI and ninth and sixth in the MCTAI respectively). The left-hand region of Fig. 3, where Singapore performs better than the Netherlands, covers about 65% of the total area and this shows that Singapore participates in creating and using technology more than the Netherlands does, in contradiction with the original TAI.

The sensitivity measures S_i and S_{Ti} for $D_{\mathrm{NL,SG}}$ are shown in Table 3. When we use S_i for SA, we are looking for important input factors that—if fixed singly—would reduce the variance the most in the output variable. Importance in SA, though, is a relative notion and there is no established threshold: one usually looks at the S_i-values and the distances between them and considers the first few factors as important. In this work, an input factor will be considered to be important if $S_i > 0.10$ (i.e. if the input factor explains more than $1/k$ of the output variance). Thus, we see that the weights for electricity and enrolment are important variables, as well as the type of weighting approach (BA or AHP). In contrast, the selection of the normalization method does not affect the variance of $D_{\mathrm{NL,SG}}$. All the input factors, taken singly, explain 52% of the output variance. The remaining 48% is the fraction of the output variance that is not explained by the input factors taken singly, but by interactions between the factors themselves. The larger the difference $S_{Ti} - S_i$, the more that factor is involved in interactions with the other factors. We can see that the trigger for the weighting scheme (BA or AHP) has a strong interaction with other factors, mainly with the weights for electricity, enrolment and exports. When we use S_{Ti} for SA, we customarily look for factors with very small S_{Ti} , which can thus confidently be declared non-influential, and hence for example fixed in a subsequent iteration of the analysis (Saltelli and Tarantola, 2002). The high value of S_{Ti} for the weight of export tells us that it cannot be fixed in spite of its low S_i.

The information that is provided by SA, e.g. the importance of the trigger for the weighting scheme, of the weights of electricity and enrolment, can be used to investigate further the difference in the ranking of the Netherlands (sixth in the original TAI; ninth in the MCTAI).

Let us start by investigating the relative performance between the Netherlands and Singapore in the two-dimensional space of the 'trigger for weighting scheme' and the 'weight of electricity'. Fig. 4(a) shows the box plots of the values for the weight of electricity for the two weighting schemes (i.e. BA or AHP) for the cases where the performance of the Netherlands is better than, worse than or equal to that of Singapore. The total of 11264 samples has been used. The same plot has been repeated for the 'trigger for weighting scheme' and the 'weight of enrolment' and is shown in Fig. 4(b). With respect to the original TAI, MCTAI favours Singapore when high weights are assigned to the enrolment subindicator, for which Singapore is much better than the Netherlands (in $3684+3699=7383$ samples), and/or to the electricity subindicator, for which Singapore is marginally better than the Netherlands. The relative performance between the Netherlands and Singapore on the 'weight of enrolment–weight of electricity' plane, which is not shown here, confirms that Singapore is better than the Netherlands mostly when the weight for enrolment is high, whereas the weight for electricity can only make a difference when extremely high values for it are assigned in the AHP.

4.2.2. Output 2: $\bar{R}_{\mathrm{S}}$, average shift in countries' ranks with respect to the original technology achievement index

The Sobol' measures of first-order (S_i) and total effect (S_{Ti}) for the second output are shown in Table 4. The trigger for the weighting scheme and the weight of exports are the most important factors, explaining 35% of the output variance, whereas all the input factors taken singly explain 44% of the output variance. The high influence of the weight of exports on the variance of $\bar{R}_{\mathrm{S}}$ is due to Australia, Singapore, Norway, New Zealand, Greece, Mexico, Chile and Malaysia, for which the subindicator's values for exports are lower than the 20th or higher than the 92nd percentiles. The remaining 56% of the output variance that is not explained by the input factors taken singly is due to the interactions between the factors themselves. We can see that the trigger for the weighting scheme has a strong interaction with the weight of exports as for $D_{\mathrm{NL,SG}}$. Note that the weight of exports contributes to the output variance more through interactions than singly. Interestingly, the normalization method has no influence on the variation of the output.

This type of analysis has quantified the influence of the input factors on the average shift in countries' ranks with respect to the original TAI and has indicated that it is the weighting scheme interacting mainly with the weight of exports that can influence the ranking of the countries in the MCTAI with respect to the original TAI.

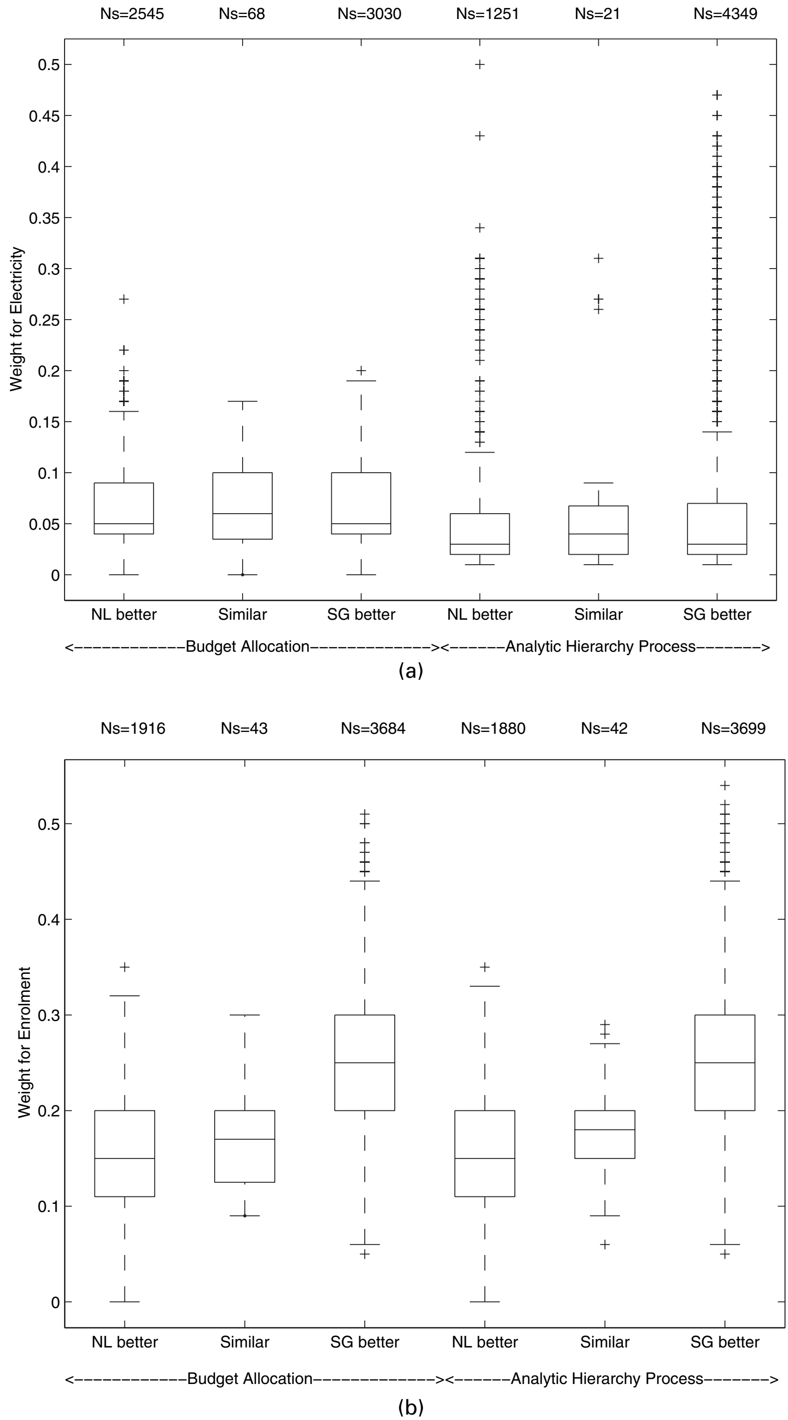

Fig. 4. Box plots of 11 264 values for (a) the weight for electricity and (b) the weight for enrolment with respect to the two weighting schemes (BA and the AHP) and the relative performance between the Netherlands (NL) and Singapore (SG) (the numbers of samples, Ns, in each group are given on top of each plot): +, outlier

Table 4. Sobol' sensitivity measures of the first-order and total effect for the output 'average shift in countries' ranks with respect to the original TAI'†

Input factor	*First-order index* (S_i)	*Total effect index* (S_{Ti})	$S_{Ti} - S_i$
X_1	0.00	0.06	0.06
X_2	*0.21*	0.56	*0.36*
X_3	0.00	0.00	0.00
X_4	0.02	0.02	0.01
X_5	0.01	0.03	0.02
X_6	*0.14*	0.44	*0.30*
X_7	0.03	0.05	0.02
X_8	0.04	0.13	0.10
X_9	0.00	0.09	0.09
X_{10}	0.00	0.14	*0.14*
Sum	0.44	1.53	1.09

†Values in italics (greater than 0.10) indicate important input factors based on S_i or $S_{Ti} - S_i$.

4.3. Worthiness of the composite indicator

Now that—having propagated uncertainties—the value of the composite indicator is no longer a simple number, but a distribution of values, the composite indicator might be seen to lose relevance if a high fraction of countries were to overlap with one another. A first remark is that some of the overlap in Fig. 2 is only exaggerated; once the correlation between the MCTAI for pairs of countries is explicitly acknowledged, the overlap becomes much smaller than is implied by the bars in Fig. 2. On a more general level, there is a trade-off between the level of uncertainty that is included in the composite indicator and its worthiness, which is herein considered as the capacity of the same index to discriminate effectively between countries. If the purpose of an index in a given policy context were to be to embarrass, e.g. to spur lazy countries into action, then again including layers of uncertainty may be seen as counter-productive. However, a careful analysis of uncertainties seems to us to make the comparison more robust. The index is no longer a magic number corresponding to a crisp normalization, weighting scheme and weight assignment, but reflects uncertainty and ambiguity in a more transparent and defensible fashion.

5. Conclusions and future work

This work has shown how to use global uncertainty and SA and other statistical techniques for the quality assessment of composite indicators, in general, and demonstrated their use on the United Nations's TAI. Three types of uncertainties have been acknowledged:

(a) alternative normalization methods for the values of the subindicators,
(b) alternative weighting approaches (such as BA and the AHP) and finally
(c) uncertainty in the weights of the subindicators.

The analysis was useful in showing that

(i) the original version of the composite indicator provided a picture of the countries' relative performances that is similar to the indicator where the three types of uncertainties are accounted for,

(ii) there are cases where countries' rankings change significantly between the original and extended TAI, i.e. the Netherlands and Singapore in the present exercise, and in identifying the reason and the regions in the space of the weights that favour one country with respect to another,

(iii) the message of what the index aims at measuring (i.e. a country's participation in creating and using technology and not which country is leading in global technology development) can be better communicated to the public when the uncertainties in the input factors are taken into consideration and

(iv) the weighting approach (through interaction with a few weights for the subindicators) affects the countries' ranks (which is useful to focus efforts on reducing the uncertainty bounds for the MCTAI), whereas the normalization method has no influence on the rankings.

The iterative use of UA and SA during the preparation of composite indicators could therefore contribute to the well structuring of composite indicators, could provide information on whether the countries' rankings measure anything meaningful and could reduce the possibility that composite indicators may send misleading or non-robust policy messages.

The verification that is offered in the present work is nevertheless partial. We have not considered as uncertain the values for the subindicators, because no estimates of the measurement errors are available for the raw data. Furthermore we have implicitly assumed that all the plurality of the debate (i.e. the sources of uncertainty) is captured by the variability in the weights, be it with the BA or AHP alternatives and the normalization method. Although the latter assumption is not far fetched, and one sees in practice examples where this happens, e.g. the European Union internal market score-board and European Union Internet business readiness index (for a review, see Saisana and Tarantola (2002)), we might also have situations where the very concept of a composite indicator is rejected by some of the stakeholders, or where the model underlying the weighting is called into question. To give an example, some investigators (Munda and Nardo, 2003) have argued that, even if weights are customarily assigned as a measure of relative importance when using linear aggregation, they have in fact a meaning of a substitution rate, whereby for example an equal weight for two indicators would mean that we are willing to trade one unit down in one indicator for one unit up in another. Even if we have not propagated these other categories of uncertainty (data uncertainty and underlying model uncertainty) in our example, it should be clear to the reader that this can be done in principle without difficulty, following a similar approach to that presented herein.

References

Box, G., Hunter, W. and Hunter, J. (1978) *Statistics for Experimenters*. New York: Wiley.

Chan, K., Tarantola, S., Saltelli, A. and Sobol', I. M. (2000) Variance based methods. In *Sensitivity Analysis* (eds A. Saltelli, K. Chan and M. Scott), pp. 167–197. New York: Wiley.

Commission of the European Communities (1984) The regions of Europe: second periodic report on the social and economic situation of the regions of the Community, together with a statement of the regional policy committee, OPOCE. *Report*. Commision of the European Communities, Luxembourg.

Cox, D. R., Fitzpatrick, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J. and Jones, D. R. (1992) Quality-of-life assessment: can we keep it simple (with discussion)? *J. R. Statist. Soc.* A, **155**, 353–393.

Freudenberg, M. (2003) Composite indicators of country performance: a critical assessment. *Report DSTI/IND(2003)5*. Organisation for Economic Co-operation and Development, Paris.

Homma, T. and Saltelli, A. (1996) Importance measures in global sensitivity analysis of model output. *Reliab. Engng Syst. Sfty*, **52**, 1–17.

Jamison, D. and Sandbu, M. (2001) WHO ranking of health system performance. *Science*, **293**, 1595–1596.

Moldan, B., Billharz, S. and Matravers, R. (1997) *Sustainability Indicators: Report of the Project on Indicators of Sustainable Development*. Chichester: Wiley.
Munda, G. and Nardo, M. (2003) On the methodological foundations of composite indicators used for ranking countries. *Organisation for Economic Co-operation and Development–Joint Research Centre Wrkshp Composite Indicators of Country Performance, Ispra, May 12th*. (Available from `http://webfarm.jrc.cec.eu.int/uasa/evt-OECD-JRC.asp`.)
Saaty, R. W. (1987) The analytic hierarchy process—what it is and how it is used. *Math. Modllng*, **9**, 161–176.
Saaty, T. L. (1980) *The Analytic Hierarchy Process*. New York: McGraw-Hill.
Saisana, M. and Tarantola, S. (2002) State-of-the-art report on current methodologies and practices for composite indicator development. *Report EUR 20408 EN*. European Commission–Joint Research Centre, Ispra.
Saltelli, A. (2002) Making best use of model valuations to compute sensitivity indices. *Comput. Phys. Communs*, **145**, 280–297.
Saltelli, A., Chan, K. and Scott, M. (2000) *Sensitivity Analysis*. New York: Wiley.
Saltelli, A. and Tarantola, S. (2002) On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *J. Am. Statist. Ass.*, **97**, 702–709.
Saltelli, A., Tarantola, S. and Campolongo, F. (2000) Sensitivity analysis as an ingredient of modelling. *Statist. Sci.*, **15**, 377–395.
Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. (2004) *Sensitivity Analysis in Practice, a Guide to Assessing Scientific Models*. New York: Wiley.
Sobol', I. M. (1967) On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math. Phys.*, **7**, 86–112.
Sobol', I. M. (1993) Sensitivity analysis for non-linear mathematical models. *Math. Modllng Comput. Exp.*, **1**, 407–414.
Tarantola, S., Jesinghaus, J. and Puolamaa, M. (2000) Global sensitivity analysis: a quality assurance tool in environmental policy modelling. In *Sensitivity Analysis* (eds A. Saltelli, K. Chan and M. Scott), pp. 385–397. New York: Wiley.
Tarantola, S., Saisana, M., Saltelli, A., Schmiedel, F. and Leapman, N. (2002) Statistical techniques and participatory approaches for the composition of the European Internal Market Index 1992-2001. *Report EUR 20547 EN*. European Commission–Joint Research Centre, Ispra.
United Nations (2001) *Human Development Report*. Oxford: Oxford University Press.