EUROPEAN COMMISSION
DG Joint Research Centre

Institute for the Protection and Security of the Citizen
Econometrics and Statistical Support to Antifraud Unit
I-21020 Ispra (VA) Italy

Robustness Analysis of the

2005 Environmental Sustainability Index

Prepared by

Michaela Saisana and Michela Nardo
(Applied Statistics Group, DG JRC)

&

Tanja Srebotnjak
(Yale Center for Environmental Law and Policy, Yale University)

January 2005

2005 **EUR 21807 EN**

The Report is available online at http://farmweb.jrc.cec.eu.int/ci/bibliography.htm

EUR 21807 EN

Table of Contents

Foreword

The concept of the environment and its protection gradually evolved from the Stockholm UN Conference on the Human Environment in 1972 through the UN Conference on Environment and Development in Rio de Janeiro in 1992 to the World Summit on Sustainable Development in Johannesburg in 2002. According to the understanding that emerged from the Rio Summit and that was reaffirmed at the Johannesburg Summit, the environment is seen as one of the three pillars of sustainable development. This broader view does not restrain the distinct role of the environment but points out its essential interconnection with the other two pillars, namely the economic and social pillars. Therefore, it is appropriate to use the term "environmental sustainability" that stresses both the specificity of the environment and its fundamental anchoring within the sustainability framework.

A measure of environmental sustainability: The "Environmental Sustainability Index" (ESI) for 2005 is developed by a team of researchers from Yale and Columbia Universities, in co-operation with the World Economic Forum and methodological support from the Applied Statistics Group of the Joint Research Centre. The ESI benchmarks the ability of 146 nations to protect the environment over the next decades, by integrating 76 data sets into 21 indicators of environmental sustainability. The data set used to construct the ESI covers a wide range of aspects of sustainability ranging from variables measuring the physical state and stress of the environmental systems to the more general social and institutional capacity to respond to environmental challenges. More than simply offering a static picture of environmental performance, the ESI offers a fertile ground for discussions on the multidimensional nature of sustainability, and on the multiple challenges arising from development and industrialization. Those challenges do not only relate to natural resource depletion, pollution, and ecosystem destruction, but also to economic development status, short-term thinking and lack of investment in capacity and infrastructure committed to pollution control and ecosystem protection.

The 2005 ESI, which was officially released at the World Economic Forum Annual Meeting in Davos in January 2005, was integrated for the first time with the outcomes from a sensitivity analysis carried out at the Joint Research Centre. Each composite indicator is based on assumptions on the way to select and combine the underlying variables. The findings of the sensitivity analysis, which systematically assesses the behaviour of the ESI under these assumptions, support the ESI as a robust composite indicator. Most country ranks do not change dramatically when tested against various combinations of aggregation, weighting and editing procedures. This robustness analysis, not performed on the previous versions of the index, confirms a reasonable degree of robustness of the ESI and provides a solid basis for discussion during the debates that usually follow the issue of the index.

It is the purpose of this document to present thoroughly the methodological steps and the results of the sensitivity analysis on the 2005 ESI.

Important note

Excerpts from the material presented here are presented in the 2005 Environmental Sustainability Index Report that is available on line www.yale.edu/esi:

Esty, Daniel C., Marc A. Levy, Tanja Srebotnjak, and Alexander de Sherbinin (2005). *2005 Environmental Sustainability Index: Benchmarking National Environmental Stewardship.* New Haven, Conn.: Yale Center for Environmental Law & Policy.

1. Introduction to Robustness Analysis for Composite Indicators

In the process of modelling a natural system or process, the observer extracts a closed, salient set of features from an open system, makes assumptions about the system's mechanisms and processes, and encodes this in a formal system (for example a computer programme) in the hope that its predictions (the output from the computer programme) will allow some useful inference on the system under study. Several practitioners have noted that the encoding process is fraught with uncertainties of different order, in the inclusion – exclusion of features, in the hypothesised mechanisms, in the value of the system parameters, in the observations, so that modelling should naturally include a careful mapping of all these uncertainties/assumptions onto the space of the model output/inferences. A combined use of Uncertainty and Sensitivity analysis can assist in evaluating whether the space of the inference is still narrow enough to be meaningful or if it is too wide to draw any useful conclusions about the system of interest. The latter outcome then requires a revision of the model or additional collection of data.

There is hence a clear incentive to supplement the mapping "*assumption* → *inference*" just mentioned with an assessment of the relative importance of different uncertainty sources in determining the uncertainty in the inference. This assessment, known as sensitivity analysis, allows, for instance, to determine which aspect of the system needs a better characterisation (e.g. the uncertainty in the initial conditions is too large and more data on present day species density are needed).

There is no reason to limit our argument to natural systems. Man-made systems (e.g. a chemical factory) or social ones (a school, a political party, a country) can all be modelled for different purposes (is the factory safe, the school efficient, the country's economy in recession or expansion) subject to the same constraints as to the tradeoffs between modelling uncertainty accurately and obtaining useful inference. Statistical indicators are themselves models, in that a considerable amount of inference and encoding participates to their construction. Composite indicators are *a fortiori* models, in that beside the issues of inclusion, exclusion, data errors and so on, they include a weighting step which might be considered, depending on the circumstances, as highly subjective.

In the specific case of composite indicators, the soundness, i.e., accuracy and precision, of resulting measures depends on a number of factors, including:

i. The model chosen for estimating the measurement error in the data, which is based on available information on variance estimation.
ii. The mechanism for including or excluding variables in the index.
iii. The transformation and/or trimming of variables during the construction process of the index.
iv. The type of normalisation scheme, such as re-scaling or standardisation, applied to remove scale effects from the variables.
v. The amount of missing data and the choice of imputation algorithm used to replace missing observations.
vi. The choice of the weights, e.g., equal weights or weights derived from factor analysis or expert opinion models.
vii. The level of aggregation, e.g., at the indicator or at the sub-indices level.
viii. The choice of aggregation system, e.g., additive, multiplicative, or multi-criteria analysis.

All these assumptions can heavily influence countries rank in a composite indicator and should be taken into account before attempting any interpretation of the results.

2. Robustness Analysis of the 2005 ESI

The robustness analysis of the ESI is essentially based on simulations that are carried on the set of equations that constitute the *index model*. Among the chief questions in assessing the robustness of the ESI ranking is how sensitive it is to changes in its structure and aggregation.

While uncertainty arises from all of the items listed above only some are significant and can be measured. The measurement error is unknown for virtually all variables, and the inclusion criteria, transformations and winsorisation, and normalisation to z-scores were found to negligibly change the country ranks. They are thus excluded from the results presented in this report.

The output of interest in the sensitivity analysis exercise is each country's rank, $Rank_c$, for $c = 1,...,146$, and the average shift in countries' rank $\overline{R}$. The latter is quantified as the average over the *146* countries of the absolute differences in countries' rank with respect to 2005 ESI:

$$\overline{R} = \frac{1}{146}\sum_{c=1}^{146}\left|Rank_{ESI\,2005,c} - Rank_c\right| \qquad (1)$$

We analysed the following questions:

1. *How do the 2005 ESI ranks compare to the most likely ranks under all scenarios?*
2. *What is the optimal scenario for each country?*
3. *Which are the most volatile countries and why?*
4. *What are the major sources of volatility in the ranking?*

As the construction of the Index is in fact a computer programme that implements factors (v) to (viii) above, the uncertainty analysis acts on a *computational model*. Various methods are available for evaluating output uncertainty (Saltelli *et al*,. 2000a).

Instead of analyzing each factor separately, the Monte Carlo approach used here considers all uncertainty sources simultaneously, thereby allows capturing potential synergistic and antagonistic effects among the uncertainty input factors. Monte Carlo analysis consists of the specification of ranges and distributions for the uncertain input factors, the generation of samples from the distributions of uncertain input factors, the evaluation of the index model given the sample input factors, and the concluding sensitivity and uncertainty analysis (Saltelli *et al.*, 2000b).

To generate a sample of uncertain input factors from their distributions (discrete in case of triggers, or continuous in case of imputed data), they are converted into a set of scalar input factors. In accordance with uncertainty factors (v) to (viii), the following uncertainty inputs are considered:

1. Imputation: The mean of 30 Marcov-Chain Monte Carlo (MCMC) simulations for each missing datum is used in the calculation of the 2005 ESI. The sensitivity analysis considers the variance associated with the 30 imputed values per missing datum to construct a distribution centered around their mean to study the effect of imputation variability on the 2005 ESI ranking.
2. Weights: Aside from the ESI's equal weights, an expert survey was used to derive opinion-based weights. The expert-weights are obtained from interviews of 17 experts of environmental sustainability and averaging their answers.
3. Aggregation level: The impact on the ESI ranking resulting from aggregation at the level of the five ESI components instead of the 21 indicators is analysed. Indicators within component are considered equally important.
4. Aggregation method: The ESI's linear aggregation scheme is compared with a multiplicative and a non-compensatory multi-criteria approach to investigate the compensability issue among indicators.

Sampling the combined space of the four input factors, some *N=10.000* combinations of the four independent input factors $\mathbf{X}^l$, $l = 1,2,...N$ are obtained using quasi-random Sobol LP_t sequences. These sequences have no randomness but offer the advantage of being uniformly distributed across the sampling space. The result is an improved convergence rate compared to a standard random Monte Carlo algorithm in the context of numerical integration (see Bratley and Fox, 1988 for a good summary description). The construction of quasi-random sequences is described in Sobol' (Sobol', 1967) and Saltelli (Saltelli *at el*., 2000a, pp. 190).

For each trial sample $\mathbf{X}^l$ the ESI is computed, generating values for the scalar output variable of interest Y^l, where Y^l is either $Rank_c$, the rank assigned by the composite indicator to each country, or $\overline{R}$, the averaged shift in countries' rank. Each output vector $\boldsymbol{Y}^l$ is linked to the corresponding generating input vector $\mathbf{X}^l$.

The sequence of $\boldsymbol{Y}^l$ allows the empirical probability distribution function (pdf) of the output $\boldsymbol{Y}$ to be built. This distribution represents the uncertainty of the output due to the uncertainty in the input. The characteristics of this pdf, such as the variance and higher order moments, can be estimated with an arbitrary level of precision that depends on the size of the simulation N.

When several layers of uncertainty are simultaneously activated, composite indicators turn out to be non-linear, possibly non-additive models (Saisana *et al.* 2005). In particular, non-additivity corresponds to the existence of interactions between the uncertain input factors, which is exemplified in the sensitivity and uncertainty analysis of the Technology Achievement Index (TAI) by Saisana et al. (Saisana *et al.*, 2005). As argued by practitioners (Saltelli *et al.*, 2000b, EPA 2004), robust, "model-free" techniques for sensitivity analysis should be used for non-linear models. Variance-based techniques have been shown to provide yield useful results in for sensitivity analysis. The discussion of their methodological formulation to compute sensitivity measures that account for the interaction between the input factors goes beyond the scope of this report and the reader is referred to (Saltelli *et al.* 2000a). Here we only display those additional properties of model-free variance-based techniques that are convenient for the present analysis:

- they allow an exploration of the whole range of variation of the input factors, instead of just sampling factors over a limited number of values, as done e.g. in fractional factorial design (Box *et al.* 1978);

- they are quantitative, and can distinguish main effects (first order) from interaction effects (second and higher order);
- they are easy to interpret and to explain;
- they allow for a sensitivity analysis whereby uncertain input factors are treated in groups instead of individually.

3. Results and discussion

The results presented and discussed herein refer to the four key questions posed in section 2.

How do the 2005 ESI ranks compare to the most likely ranks under all scenarios?

The uncertainty analysis results of the 146 countries ranks are given in Figure 1. Countries are ordered by their original 2005 ESI rank. For ease of reading, the countries in Figure 1 are split into three groups according to original 2005 ESI rank, i.e. beginning with Finland (original ESI rank =1) to Cameroon (rank = 50) in the top graph, Ecuador (rank = 51) to Kenya (rank =100) in the center graph, and India (rank = 101) to North Korea (rank =146) in the bottom graph.

The width of the 5^{th} – 95^{th} percentile bounds and the generally small deviation of the median rank (black hyphen) from the original ESI rank (grey hyphen) demonstrate only small differences between the Monte Carlo-based and the original ESI. For about 90 countries the difference between the original 2005 ESI rank and the median rank when considering different approaches/assumptions is less than 10 positions. This outcome implies an acceptable robustness of ESI, as most of the countries' most likely (median) rank is very close to the actual 2005 ESI rank. The dominant source for the observed deviations is the combined effect of multiple imputations and aggregation level. For the countries in the top group this average difference is 7 positions, which increases to 12 positions for the center group and 11 for the bottom group. Somewhat surprisingly given the uneven distribution of data availability, both OECD and non-OECD countries have an average shift in rank of almost 9 positions. Thus, the amount of missing data appears not to be a notice worthy source of rank variation.

The greatest differences between the 2005 ESI rank and the median rank of the simulations are observed for Mali, Nicaragua, Mongolia, Guinea-Bissau and Syria. The first four countries are favoured in their 2005 ESI rank by almost 35 positions when compared to their median rank, while the opposite is valid for Syria. Guatemala and El Salvador are the only two countries within the set of 146 countries whose 2005 ESI rank falls outside the estimated rank range (5^{th} -95^{th} percentiles). For these countries the ESI rank represents an "outlier scenario": all uncertainties considered they would have ranked higher.

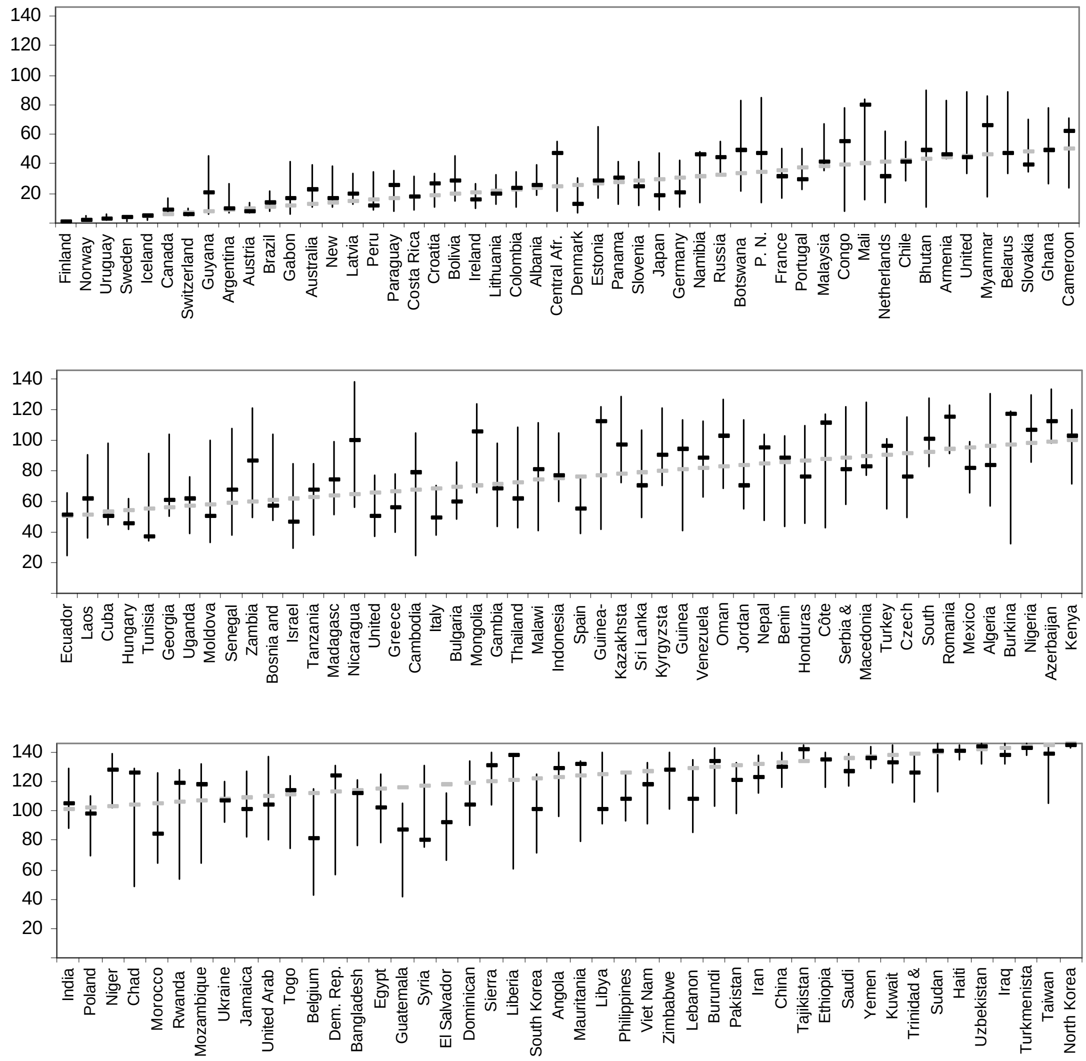

Figure 1. *Uncertainty analysis results showing the median (black mark) countries' rank and the 5th and 95th percentiles (bounds) of a country's rank distribution. Countries are ordered according to their original 2005 ESI rank (light grey mark).*

Table 1. *2005 ESI Ranking and Optimal Rank for each country under all tested combinations of uncertainty inputs*

Country	ESI Rank	Best Rank	Country	ESI Rank	Best Rank	Country	ESI Rank	Best Rank
Finland	1	1	Ecuador	51	34	India	101	88
Norway	2	2	Laos	52	46	Poland	102	77
Uruguay	3	3	Cuba	53	45	Niger	103	117
Sweden	4	2	Hungary	54	42	Chad	104	64
Iceland	5	3	Tunisia	55	34	Morocco	105	65
Canada	6	7	Georgia	56	58	Rwanda	106	84
Switzerland	7	6	Uganda	57	43	Mozambique	107	86
Guyana	8	9	Moldova	58	33	Ukraine	108	92
Argentina	9	9	Senegal	59	59	Jamaica	109	86
Austria	10	7	Zambia	60	66	United Arab Em.	110	93
Brazil	11	11	Bosnia and Herze.	61	48	Togo	111	88
Gabon	12	9	Israel	62	30	Belgium	112	43
Australia	13	18	Tanzania	63	47	Dem. Rep. Congo	113	76
New Zealand	14	12	Madagascar	64	65	Bangladesh	114	91
Latvia	15	13	Nicaragua	65	56	Egypt	115	87
Peru	16	10	United Kingdom	66	38	Guatemala	116	55
Paraguay	17	13	Greece	67	44	Syria	117	75
Costa Rica	18	12	Cambodia	68	35	El Salvador	118	75
Croatia	19	16	Italy	69	40	Dominican Rep.	119	90
Bolivia	20	20	Bulgaria	70	55	Sierra Leone	120	118
Ireland	21	11	Mongolia	71	75	Liberia	121	98
Lithuania	22	16	Gambia	72	56	South Korea	122	72
Colombia	23	16	Thailand	73.0	56	Angola	123	118
Albania	24	21	Malawi	74	56	Mauritania	124	99
Central Afr. Rep.	25	13	Indonesia	75	70	Libya	125	91
Denmark	26	8	Spain	76	44	Philippines	126	100
Estonia	27	17	Guinea-Bissau	77	71	Viet Nam	127	106
Panama	28	19	Kazakhstan	78	73	Zimbabwe	128	105
Slovenia	29	19	Sri Lanka	79	58	Lebanon	129	85
Japan	30	9	Kyrgyzstan	80	81	Burundi	130	122
Germany	31	13	Guinea	81	60	Pakistan	131	110
Namibia	32	19	Venezuela	82	64	Iran	132	119
Russia	33	37	Oman	83	72	China	133	121
Botswana	34	31	Jordan	84	55	Tajikistan	134	137
P. N. Guinea	35	27	Nepal	85	59	Ethiopia	135	125
France	36	22	Benin	86	55	Saudi Arabia	136	127
Portugal	37	23	Honduras	87	59	Yemen	137	134
Malaysia	38	36	Côte d'Ivoire	88	55	Kuwait	138	120
Congo	39	14	Serbia & Montenegro	89	75	Trinidad & Tobago	139	115
Mali	40	25	Macedonia	90	81	Sudan	140	133
Netherlands	41	14	Turkey	91	66	Haiti	141	138
Chile	42	34	Czech Rep.	92	63	Uzbekistan	142	137
Bhutan	43	31	South Africa	93	90	Iraq	143	132
Armenia	44	43	Romania	94	98	Turkmenistan	144	141
United States	45	34	Mexico	95	73	Taiwan	145	124
Myanmar	46	28	Algeria	96	57	North Korea	146	144
Belarus	47	34	Burkina Faso	97	45			
Slovakia	48	35	Nigeria	98	92			
Ghana	49	35	Azerbaijan	99	110			
Cameroon	50	32	Kenya	100	87			

What is the optimal scenario for each country?

Interpreting the 5th percentile of a country's rank distribution as its best rank under all assumptions made in the index model, the most pronounced improvement in performance among the top 50 countries is observed for Congo, the Netherlands, and Japan, which gain more than 20 positions in the ranking if the index were calculated according to a different structure (see Table 1). Specifically, the Netherlands and Japan greatly advance their rank if aggregation takes place at the components level and Congo's improvement is due to the alternative imputation method.

Among the center 50 countries in the 2005 ESI, the most pronounced improvement occurs for Burkina Faso and Algeria, which gain more than 40 positions in the ranking under a different structure in the composite indicator. In particular, Burkina Faso owes its improvement to the imputation, while Algeria improves its rank under aggregation at the indicators' level.

Belgium, South Korea and Guatemala also achieve improvement of more than 50 positions in the bottom 46 countries: the first two countries thanks to the aggregation at the components' level, while Guatemala thanks to imputation.

Which are the most volatile countries and why?

Let 'volatility' be defined as the difference between a country's best and worst rank, which are given by the 5th and the 95th percentiles of the rank distribution respectively. For the first 10 countries in the 2005 ESI rank, except for Guyana and Argentina, the volatility is low ranging from 2 to 4 positions, which suggests a robust performance for those countries. Guyana's high volatility (23 positions) is mainly attributed to the impact of imputation variation since 28 variables out of 76 have been imputed and its combined effect with the choice of the aggregation level. Argentina's volatility of 9 positions is entirely due to imputation, although only 5 variables are imputed.

Table 2 presents the 15 countries that are affected strongly by the choices made during the building of the ESI. These countries, with a difference in their best and worst rank (5th and 95th percentiles) of some 50 to 80 positions, are placed between the 39 (Congo) and 113 (Dem. Rep. Congo) ESI rank. Only Congo, Mali, Myanmar and Belarus are ranked among the top 50 in the ESI. The volatility of those countries can be attributed mainly to the high variance of imputation and its combined effect with the choice of aggregation level, as indicated by the Sobol' sensitivity indices (Sobol', 1993). In some simulation runs the imputed values are high, partly compensating for the low scores in other variables and improving the country's rank. In other runs, instead, the imputed value is very low, thereby reducing the country's position even further.

Table 2. *Most volatile countries in the 146 countries' set*

Country	Rank ESI	Rank range	Country	Rank ESI	Rank range
Congo	39	[14, 78]	Côte d'Ivoire	88	[55, 117]
Mali	40	[25, 84]	Czech Rep.	92	[63, 115]
Myanmar	46	[28, 80]	Algeria	96	[57, 125]
Belarus	47	[34, 87]	Burkina Faso	97	[45, 119]
Nicaragua	65	[56, 134]	Chad	104	[64, 129]
Cambodia	68	[35, 105]	Belgium	112	[43, 108]
Guinea-Bissau	77	[71, 122]	Dem. Rep. Congo	113	[76, 131]
Oman	83	[72, 122]			

What are the major sources of volatility in the ranking?

This question addresses the independent impact of each of the four assumptions during the ESI development and identifies the countries whose rank is most significantly affected by one or more of the assumptions, i.e.

- What if no imputation is carried out, instead of using the mean of imputations (as in 2005 ESI)?
- What if a set of weights provided by experts' on the topic is used for the 21 indicators, instead of the equal weighting (as in 2005 ESI)?
- What if aggregation is applied at the components' level, instead at the indicators' level (as in 2005 ESI)?
- What if a non-compensatory aggregation scheme is used, instead of the linear aggregation scheme (as in 2005 ESI)?

Imputation

It is reasonable to assume that the imputation effect should correlate positively with the fraction of missing data. This inverse relationship between data availability and increased rank volatility is, however, not straightforward. While countries with many missing values may be susceptible to increased variation due to imputation variance, the extent of the imputation effect depends in part on the imputation model used. The results suggest a mixed picture with the imputation effect varying more strongly between certain countries than according to the amount of missing data. Overall, the imputation effect on the average rank shift is one of the largest sources of uncertainty in the ESI.

Among the countries that are missing almost 1/3 of the observations only Guinea-Bissau and Myanmar experience a high impact on their rank due to imputation (Table 3). If no imputation is applied, the countries that are favoured the most are Syria, Algeria, Belgium, and the Dominican Republic, which improve by 29-37 positions. The opposite holds for Mali, Guinea-Bissau, Myanmar, and Zambia, which deteriorate between 27 to 44 positions. Overall for the set of 146 countries, the decision to impute missing values using a MCMC model has an average impact of 10 ranks and a rank-order correlation coefficient of 0.95.

***Table 3.** ESI ranking with imputed missing data versus ESI ranking without imputation: list of the countries that largely improve or worsen their rank position.*

		ESI rank with imputation	*rank without imputation*	*Change in rank*
Improvement	Syria	117	80	*-37*
	Algeria	96	64	*-32*
	Belgium	112	82	*-30*
	Dominican Rep.	119	90	*-29*
Deterioration	Mali	40	84	*+44*
	Guinea-Bissau	77	114	*+37*
	Myanmar	46	76	*+30*
	Zambia	60	87	*+27*
	Average change over 146 countries			*10*

Linear weighting versus experts' weighting

Weights can be assigned to the indicators according to their respective relevance in the context of a complex phenomenon such as environmental sustainability. In a *"budget allocation scheme"* (BA) such weights are elicited from experts, who are assumed to have an understanding of the phenomenon being modelled by the composite indicator.

For the ESI composite indicator, each of the 21 experts interviewed[1] was given a "budget" of 100 points, and asked to allot them to the 21 indicators according to the importance they attach to the indicator. Four of those experts gave zero priority points to a significant number of indicators, which corresponds to not to include the indicator and its underlying variables into the ESI. Since the elimination of a substantial number of indicators was not envisioned for a comprehensive, broad-based ESI, these four samples were eliminated from the sample.

The sets of weights obtained by the experts, together with their average are listed in Table 4.

Overall, the average expert-weighting supplies nearly equal weights. with the indicators in the Environmental Systems and Stresses components judged slightly more important and the Vulnerability, Institutional Capacity, and Global Stewardship indicators somewhat less important. Nevertheless the variance of the opinion survey is rather large and ranges from 40-80% of the mean weight. This explains the difference between the 2005 ESI ranking and the one provided by the BA, when the mean of all experts' is considered. Overall for the set of 146 countries, the assumption on the weighting set has an average impact of 5 ranks (Table 5) and a rank-order correlation coefficient of 0.99.

The experts' weighting, by assigning larger weights to indicators within the System and Stress Components of ESI and less to the remaining indicators, has a positive impact on the rank of countries such as Sri Lanka and Niger, whilst a negative effect on others such as the Chile, South Africa or Italy.

[1] The interview was carried out during the "Expert Review meeting of the 2005 Environmental Sustainability Index" held on 9-10 December 2004 at Yale University

Table 4. *Experts' weighting for each indicator of the ESI*

	experts									
	1	2	3	4	5	6	7	8	9	10
Air Quality	0.03	0.05	0.09	0.14	0.04	0.02	0.03	0.05	0.03	0.02
Biodiversity	0.05	0.09	0.07	0.14	0.05	0.05	0.02	0.05	0.03	0.10
Land	0.05	0.09	0.06	0.14	0.05	0.02	0.04	0.06	0.11	0.05
Water Quality	0.05	0.05	0.09	0.14	0.06	0.02	0.03	0.05	0.03	0.02
Water Quantity	0.05	0.02	0.05	0.02	0.04	0.07	0.04	0.06	0.03	0.10
Reducing Air Pollution	0.06	0.05	0.05	0.02	0.05	0.07	0.08	0.04	0.03	0.10
Reducing Ecosystem Stress	0.06	0.05	0.06	0.02	0.05	0.05	0.06	0.06	0.03	0.02
Reducing Population Pressure	0.04	0.05	0.07	0.02	0.06	0.05	0.08	0.06	0.03	0.02
Reducing Waste & Consumption Pressures	0.06	0.05	0.05	0.02	0.06	0.05	0.08	0.05	0.03	0.05
Reducing Water Stress	0.06	0.05	0.04	0.02	0.06	0.07	0.05	0.05	0.03	0.10
Natural Resource Management	0.07	0.09	0.06	0.02	0.04	0.07	0.00	0.06	0.05	0.05
Environmental Health	0.05	0.09	0.04	0.02	0.06	0.05	0.05	0.06	0.03	0.05
Basic Human Sustenance	0.05	0.05	0.04	0.02	0.05	0.05	0.05	0.06	0.11	0.05
Exposure to Natural Disasters	0.05	0.00	0.05	0.04	0.06	0.07	0.00	0.04	0.00	0.02
Environmental Governance	0.03	0.03	0.03	0.02	0.04	0.05	0.04	0.01	0.14	0.03
Eco-efficiency	0.04	0.02	0.03	0.02	0.04	0.05	0.02	0.05	0.11	0.02
Private Sector Responsiveness	0.03	0.05	0.03	0.02	0.05	0.05	0.06	0.05	0.03	0.05
Science and Technology	0.03	0.05	0.05	0.00	0.05	0.05	0.06	0.03	0.11	0.05
Participation in International Collaborative Efforts	0.04	0.02	0.03	0.00	0.04	0.04	0.04	0.02	0.03	0.02
Greenhouse Gas Emissions	0.04	0.02	0.03	0.10	0.06	0.09	0.07	0.05	0.03	0.10
Reducing Transboundary Environmental Pressures	0.06	0.05	0.03	0.06	0.04	0.00	0.06	0.05	0.03	0.02

	expert								
	11	12	13	14	15	16	17	**average**	**Equal weighting**
Air Quality	0.05	0.10	0.06	0.06	0.07	0.05	0.10	0.06	0.05
Biodiversity	0.05	0.05	0.06	0.05	0.06	0.05	0.02	0.06	0.05
Land	0.05	0.05	0.02	0.04	0.04	0.05	0.03	0.06	0.05
Water Quality	0.05	0.10	0.02	0.04	0.07	0.05	0.10	0.06	0.05
Water Quantity	0.05	0.05	0.02	0.06	0.03	0.05	0.04	0.05	0.05
Reducing Air Pollution	0.05	0.10	0.05	0.05	0.06	0.05	0.05	0.06	0.05
Reducing Ecosystem Stress	0.05	0.10	0.08	0.05	0.04	0.05	0.07	0.05	0.05
Reducing Population Pressure	0.05	0.01	0.06	0.05	0.05	0.02	0.01	0.04	0.05
Reducing Waste & Consumption Pressures	0.05	0.10	0.08	0.05	0.05	0.05	0.03	0.05	0.05
Reducing Water Stress	0.05	0.03	0.02	0.05	0.06	0.05	0.03	0.05	0.05
Natural Resource Management	0.05	0.00	0.02	0.00	0.00	0.05	0.03	0.04	0.05
Environmental Health	0.03	0.05	0.05	0.06	0.06	0.08	0.02	0.05	0.05
Basic Human Sustenance	0.03	0.05	0.02	0.04	0.05	0.05	0.05	0.05	0.05
Exposure to Natural Disasters	0.03	0.00	0.05	0.00	0.00	0.04	0.04	0.03	0.05
Environmental Governance	0.06	0.02	0.01	0.05	0.04	0.06	0.12	0.05	0.05
Eco-efficiency	0.05	0.05	0.08	0.05	0.05	0.05	0.02	0.04	0.05
Private Sector Responsiveness	0.05	0.01	0.06	0.04	0.06	0.05	0.05	0.04	0.05
Science and Technology	0.07	0.05	0.02	0.06	0.04	0.05	0.03	0.05	0.05
Participation in International Collaborative Efforts	0.05	0.01	0.04	0.05	0.05	0.04	0.07	0.03	0.05
Greenhouse Gas Emissions	0.05	0.10	0.08	0.06	0.05	0.04	0.05	0.06	0.05
Reducing Transboundary Environmental Pressures	0.05	0.01	0.08	0.04	0.04	0.04	0.05	0.04	0.05

Table 5. *ESI ranking with equal weighting (EW) versus ESI ranking with BA: list of the countries that largely improve or worsen their rank position.*

		ESI rank with EW	*rank with BA*	*Change in rank*
Improvement	Sri Lanka	79	61	*-18*
	Niger	103	86	*-17*
	Dem. Rep. Congo	113	98	*-15*
	El Salvador	118	103	*-15*
	Hungary	54	40	*-14*
Deterioration	Chile	42	59	*+17*
	United Arab Em.	110	127	*+17*
	South Africa	93	109	*+16*
	Nicaragua	65	78	*+13*
	Italy	69	82	*+13*
	Average change over 146 countries			*5*

Aggregation at the components' level versus aggregation at the indicators' level

Equally weighting the five components (Stress, System, Human Vulnerability, Institutional Capacity and Global Stewardship) instead of the 21 indicators offers another possible index model for the ESI.

Figure 2 compares the ranking obtained from both approaches and it emerges that, by changing the aggregation level, the average shift of the top 40 and the bottom 30 countries of the 2005 ESI is 7 positions whereas the shift of the remaining countries is 11 positions on average. As expected middle-of-the-road performers display higher variability than the top and bottom countries.

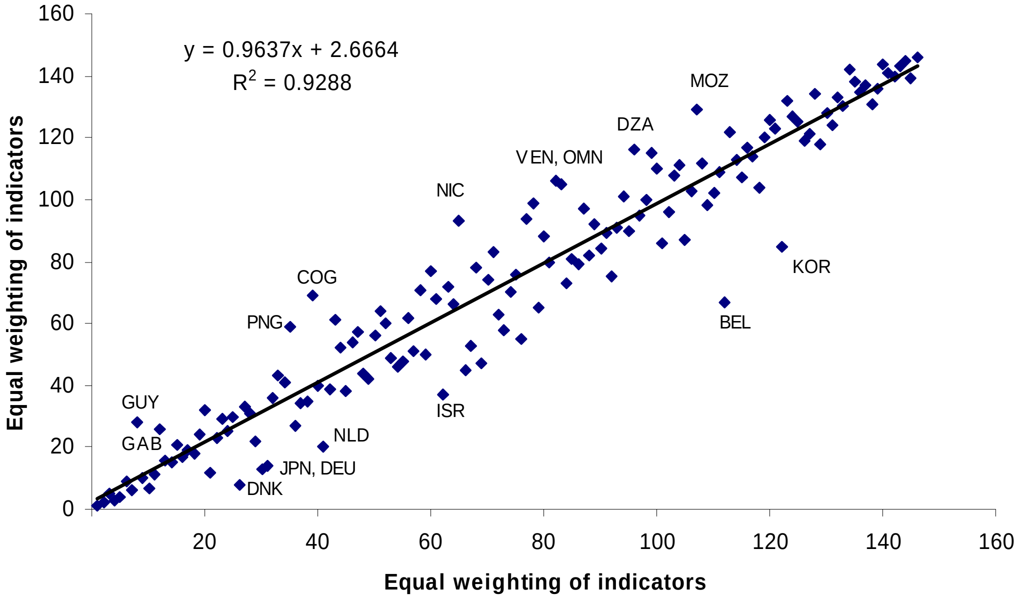

Figure 2. *ESI ranking: comparison of equal weighting of the 21 indicators versus equal weighting of the 5 components.*

Table 6. *ESI ranking with equal weighting (EWI) of all 21 indicators versus ESI ranking with equal weighting of the 5 components (EWC): list of the countries that largely improve or worsen their rank position.*

		ESI rank with EWI	*rank with EWC*	*Change in Rank*
Improvement	Belgium	112	67	*-45*
	South Korea	122	85	*-37*
	Israel	62	37	*-25*
	Italy	69	47	*-22*
	Netherlands	41	20	*-21*
Deterioration	Congo	39	69	*+30*
	Nicaragua	65	93	*+28*
	P. N. Guinea	35	59	*+24*
	Venezuela	82	106	*+24*
	Oman	83	105	*+22*
Average change over 146 countries				*8*

Weighting components instead of indicators affects only 38 of the 146 countries seriously, i.e. by more than 10 positions. Ten of these countries are among the first 50 positions of the 2005 ESI. The average impact is 8 ranks and the rank-order correlation coefficient is high at 0.96. Belgium and South Korea improve their rank by almost 40 positions if weighting is done at the components level (Table 6). On the contrary, countries such as Congo or Nicaragua drop some 30 positions. The reason has to be sought in the relatively good performance of these latter countries in stress and system components that are more heavily weighted when the aggregation is at the indicators' level. The aggregation at the components level, in fact, favors components made by a small number of indicators which receive a score higher than the sum of its equally weighted component indicators.

Linear aggregation versus non-compensatory multi-criteria

The literature of composite indicators offers several examples of aggregation techniques. Additive techniques are the most frequent ones. However, additive aggregations imply requirements and properties, both of indicators and of the associated weights, which are often not desirable and at times difficult to meet or cumbersome to verify. Other, less widespread, aggregation methods are proposed in the literature of composite indicators, such as multiplicative (or geometric) aggregations or non-linear aggregations (e.g. multi-criteria analysis).

Several authors (Debreu, 1960; Keeney and Raiffa, 1976; Krantz et al., 1971) note that an additive aggregation function for a given set of indicators exists if and only if these indicators are mutually preferentially independent. Preferential independence is a very strong condition since it implies that the trade-off ratio between two indicators is independent of the values of the remaining indicators (Ting, 1971). In practice, this means that an additive aggregation function allows for an assessment of the marginal contribution of each variable separately. This marginal contribution can then be added together to yield a total value. If, for example, environmental dimensions are involved, the use of a linear aggregation procedure implies that among the different aspects of an ecosystem there are no phenomena of synergy or conflict. This appears to be quite an unrealistic assumption (Funtowicz et al., 1990). For example, laboratory experiments made clear that the combined impact of the acidifying substances SO_2, NO_X, NH_3 and O_3 on plant growth is substantially more severe than the (linear) addition of the impacts of each of these substances alone (Dietz and van der Straaten, 1992).

Furthermore linear aggregation entails full compensability: a poor performance in some indicators can be compensated for by a good performance in others. Yet, not everybody would trade an increase in ‘Participation in International Collaborative Efforts’ with a decrease in the ‘Biodiversity’ or an increase in ‘Greenhouse Gas Emissions’. A perverse implication of full compensability is that the weights necessarily have the meaning of substitution rates (e.g. how much ‘Biodiversity’ can be traded against ‘Participation’), and do not indicate the importance of the indicator associated.

The implication of the ambiguity in the use of additive aggregation is the existence of a theoretical inconsistency in the way weights are actually used and their real theoretical meaning. For the weights to be interpreted as “importance coefficients” (in jargon symmetrical importance of variables, e.g. place the greatest weight beside the most important “dimension”) non-compensatory aggregation procedures must be used to construct composite indicators (Podinovskii, 1994). This can be done using a non-compensatory multi-criteria approach.

A Non-Compensatory Multi-Criteria Approach (NCMC) is based on mathematical aggregation conventions that can be divided into two main steps:

- Pairwise comparison of countries according to the whole set of indicators used.
- Ranking of countries in a complete pre-order.

The result of the first step is an $M \times M$ matrix (M = number of countries), commonly termed *outranking matrix* (Arrow and Raynaud, 1986; Roy, 1996). The way the information contained in the outranking matrix is used in the second step takes into consideration the following points:

- Intensity of preference (i.e. how much one country is better than another country for a given indicator);
- Number of indicators in favour of a given country;
- Weight attached to each indicator;
- Relationship of each country with respect to all the other countries.

There are several ranking procedures for this second step (see Young, 1988). One possible algorithm is derived from the Condorcet-Kemeny-Young-Levenglick (CKYL) ranking procedure (Munda and Nardo, 2003). We offer here a 'hand waiving' description of the NCMC algorithm. Imagine we have three countries, A, B and C and we aim at ranking their overall performance according to N indicators. We build to this effect an 'outranking matrix' whose entries eij tells us how much country 'i' does better than country 'j'. eij is in fact the sum of all weights of all indicators for which country 'i' does better than country 'j'. eji will likewise be the sum of all weights for which the reverse is true. If the two countries do equally well on one variable, its weight is split between eij and eji. As a result eij + eji =1 if weights have been scaled to unity. We now write down all permutations of county order (ABC,ACB,BAC,BCA,CAB,CBA) and compute for each of them the ordered sum of the scores, e.g. for ABC we compute Y=eAB+eAC+eBC. We do this for all permutations and take as the multicriteria country ranking the one with the highest total score Y. Note that this ordering is only based on the weights, and on the sign of the difference between countries values for a given indicator, the magnitude of the difference being ignored. With this approach no compensation occurs, to exemplify, a country that does marginally better on many indicators comes out better than a country that does a lot better on a few ones because it cannot compensate deficiencies in some dimensions with outstanding performances in others. According to CKYL the ranking of countries with the highest likelihood is the ranking supported by the maximum number of indicators for each pair-wise comparison, summed over all pairs of countries considered. The multi-criteria method has the advantage to overcome some of the problems inherent in additive or multiplicative aggregations: preference dependence between indicators and the meaning of trade-offs given to the weights. Furthermore, both qualitative and quantitative information can be treated simultaneously. In addition, it does not need any manipulation of the raw data (e.g. winsorization, data transformation) or normalization to assure the comparability of indicators.

Figure 3 compares the ranking obtained with the non-compensatory multicriteria method to that of the 2005 ESI. In both cases all 21 indicators are weighted equally. The figure illuminates that the aggregation method used affects principally the middle-of-the-road and, to a lesser extent, the bottom-ranked countries. Overall, for the 146 countries included in the ESI, the assumption on the aggregation scheme has an average impact of 8 ranks and a rank-order correlation coefficient of 0.96, very similar to the impact of aggregating at the components' instead of indicators' level. In particular, while the top 50 countries only move on average by 5 positions, the following 50 countries move on average by 12 positions and the latter 46 countries by 8 positions.

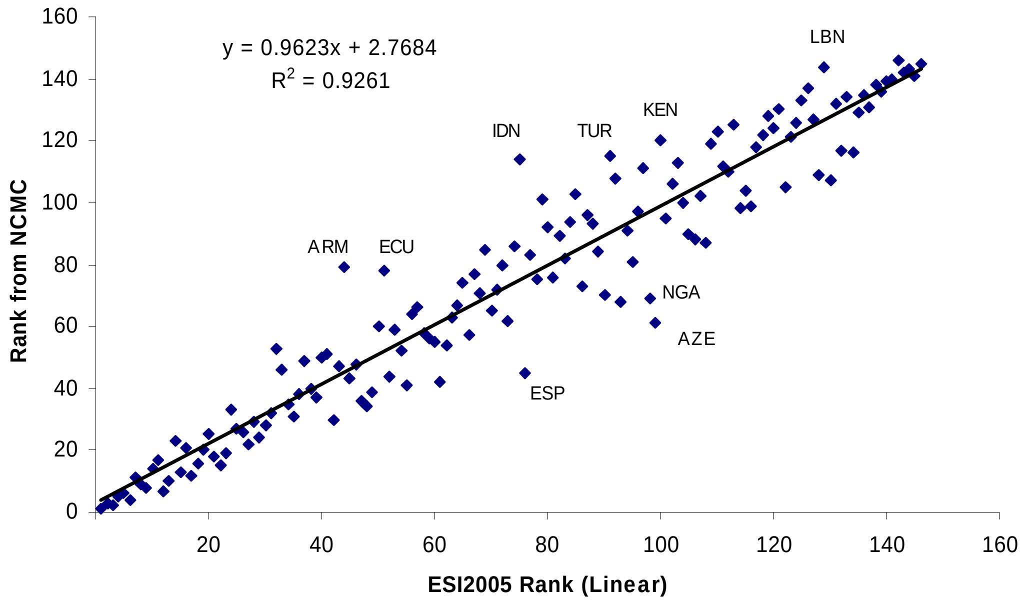

Figure 3. *ESI ranking: comparison of two alternative aggregation methods. Linear aggregation of the 21 indicators (2005 ESI) versus non-compensatory multi-criteria (NCMA) aggregation of the 21 indicators.*

Both aggregation schemes, therefore, seem to produce comparable rankings (the R^2 is 0.92). Using the NCMC, in fact, only 43 out of 146 countries display a change in rank higher than 10 positions (none before the 30th ESI rank). When compensability among indicators is not allowed, countries performing very poorly on some indicators, such as Indonesia or Armenia see their rank lowered with respect to the linear aggregation, whereas countries that have less extreme values improve their situation, such as Azerbaijan or Spain. Table 7 shows the countries displaying the largest variation in their ranks.

Table 7. *ESI ranking with linear aggregation (LIN) versus ranking with non-compensatory multi-criteria (NCMC): list of the countries that improve or lower their rank position most noticeably.*

	Aggregation	*ESI rank with LIN*	*rank with NCMC*	*Change in Rank*
Improvement	Azerbaijan	99	61	*-38*
	Spain	76	45	*-31*
	Nigeria	98	69	*-29*
	South Africa	93	68	*-25*
	Burundi	130	107	*-23*
Deterioration	Indonesia	75	114	*+39*
	Armenia	44	79	*+35*
	Ecuador	51	78	*+27*
	Turkey	91	115	*+24*
	Sri Lanka	79	101	*+22*
	Average change over 146 countries			*8*

4. Conclusions

The validity of the ESI ranking is assessed by evaluating how sensitive it is to the assumptions that have been made about its structure and aggregation of the 21 indicators, mainly related to: variability in the imputation of missing data, equal versus experts opinion weighting, aggregation at indicators versus at components level, and linear versus non-compensatory aggregation scheme. The main findings are summarized as follows:

- How do the 2005 ESI ranks compare to the most likely ranks under all scenarios? The most likely rank of a country considering all combinations of assumptions in the sensitivity analysis rarely deviates substantially from its 2005 ESI rank indicating the robustness of the 2005 ESI ranking with respect to this source of uncertainty. For about 90 out of 146 countries the difference between the 2005 ESI rank and the most likely (median) rank is less than 10 positions.

- Which are the most volatile countries and why? The top ten ranking countries in the ESI all have modest volatility (2 to 4 positions) with the exceptions of Guyana (23 positions) and Argentina (9 positions). This small degree of sensitivity implies a robust evaluation of performance for those countries. Guyana's high volatility is mainly attributed to imputation (28 variables out of 76 have been imputed) and its combined effect with the choice of the aggregation level. Argentina's volatility is entirely due to imputation, although only 5 variables have been imputed. The countries that present the highest volatility (between 50 and 80 positions), are found between rank 39 (Congo) and rank 113 (Dem. Rep. Congo).

- What if no imputation is carried out, instead of using the imputed mean? Imputation plays an important role in the robustness of the 2005 ESI ranking. For the set of 146 countries, the assumption on imputation has an average impact of 10 ranks. However, among the countries that are missing almost 30% of the observations, only Guinea-Bissau and Myanmar experience notable rank changes due to imputation.

- What if a set of weights provided by experts' on the field had been used, instead of the equal weighting for all 21 indicators? An alternate weighting obtained by surveying the experts at the December 2004 ESI Review Meeting assigns slightly higher values to indicators within the Systems and Stresses Components of ESI and less to the remaining indicators. Using these weights has a pronounced positive effect on the rank of a few countries such as Sri Lanka and Niger, but a negative effect on others such as Chile, South Africa, or Italy. Overall, the analysis shows only a small sensitivity to the weighting assumption with an average impact of 5 ranks.

- What if aggregation is applied at the components' level instead of at the indicators' level? Weighting the five components equally has little effect on most countries, with a few significant exceptions. Belgium and South Korea would rise by almost 40 positions in the ranking if aggregation were done at the component level rather than the indicator level. Conversely, Congo and Nicaragua would fall by 30 positions. The reason for this effect lies in the fact that aggregation at the component level gives added weight to components with fewer indicators, such as Human Vulnerability and Global Stewardship. Overall, the level at which aggregation to the ESI takes place has an average impact of 8 ranks, similar to the impact of the aggregation scheme.

- What if a non-compensatory aggregation scheme is used, instead of the linear aggregation scheme? Aggregation scheme matters mainly for the average performers. When compensability among indicators is not allowed, countries having very poor performance in some indicators, such as Indonesia or Armenia lose ground with respect to the linear yardstick, whereas countries that have less extreme values improve their situation, such as Azerbaijan or Spain. Overall for the set of 146 countries, the assumption on the aggregation scheme has an average impact of 8 ranks.

References

1. Arrow K.J., and Raynaud H. (1986) *Social choice and multicriterion decision making*, M.I.T. Press, Cambridge.
2. Box, G., Hunter, W. and Hunter, J. (1978) Statistics for experimenters, New York: John Wiley and Sons.
3. Bratley P. and Fox B.L. (1988) ALGORITHM 659 Implementing Sobol's quasirandom sequence generator. ACM Trans. Math. Software 14, 88-100.
4. Debreu G. (1960) Topological methods in cardinal utility theory, in Arrow K.J., Karlin S. and Suppes P. (eds.) *Mathematical methods in social sciences*, Stanford University Press, Stanford.
5. Dietz F.J., van der Straaten J. (1992) Rethinking environmental economics: missing links between economic theory and environmental policy, *Journal of Economic Issues*, Vol. XXVI No. 1, pp. 27-51.
6. EPA (2004) Council for Regulatory Environmental Modeling, Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models", http://www.epa.gov/osp/crem/library/CREM%20Guidance%20Draft%2012_03.pdf.
7. Funtowicz S.O., Munda G., Paruccini M. (1990) The aggregation of environmental data using multicriteria methods, *Environmetrics*, Vol. 1(4), pp. 353-36.
8. Keeney R., Raiffa H. (1976) *Decision with multiple objectives: preferences and value trade-offs*, Wiley, New York.
9. Krantz D.H., Luce R.D., Suppes P. and Tversky A. (1971) *Foundations of measurement, vol. 1, Additive and polynomial representations*, Academic Press, New York.
10. 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.
11. Podinovskii V.V. (1994) Criteria importance theory, *Mathematical Social Sciences*, 27, pp. 237 - 252.
12. Roy B. (1996) *Multicriteria methodology for decision analysis*, Kluwer, Dordrecht.
13. Saisana M., Tarantola S., Saltelli A. (2005) Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators, *Journal of the Royal Statistical Society A*, 168(2), 1-17.
14. Saltelli, A., Chan, K. and Scott, M. (2000a) *Sensitivity analysis*, Probability and Statistics series, New York: John Wiley & Sons.
15. Saltelli, A., Tarantola, S. and Campolongo, F. (2000b) Sensitivity analysis as an ingredient of modelling. *Statistical Science*, 15, 377-395.
16. Sobol' I.M. (1967) On the distribution of points in a cube and the approximate evaluation of integrals. USSR Computat. Maths. Math. Phys. 7, 86-112.
17. Sobol' I.M. (1976) Uniformly distributed sequences with an addition uniform property. USSR Computat. Maths. Math. Phys. 16, 236-242.

18. Sobol', I. M. (1993) Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling & Computational Experiment* 1, 407-414.

19. Ting H.M. (1971) *Aggregation of attributes for multiattributed utility assessment*, Technical report n. 66, Operations Research Center, MIT Cambridge Mass.

20. Young H.P. (1988) Condorcet's theory of voting, *American Political Science Review*, Vol. 82, No. 4, pp. 1231-1244.

Mission of the JRC

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

EUROPEAN COMMISSION
DG Joint Research Centre