2005 Environmental Sustainability Index

Benchmarking National Environmental Stewardship

Appendix A
Methodology

This page is intentionally blank.

Table of Contents

List of Tables

List of Figures

Methodology

Considerable conceptual and analytical processing precedes the calculation of the ESI scores and rankings. The purpose of this Appendix is to provide detailed descriptions of the statistical techniques and methods used to calculate the ESI[1]. Appendices C and D provide the data underlying the ESI. We offer this detail in support of the belief that transparency is an essential foundation for good analysis and policymaking.

The issues addressed here mirror those commonly encountered in the computation of composite indices: variable selection, missing data treatment, aggregation and weighting methodologies, as well as performance testing (OECD 2003).

In addition, the Appendix describes in greater depth the methods used in the statistical analyses that support the policy conclusions presented in the report. While the core text focuses on the key messages emerging from the analyses, this section includes the results of the ESI analyses and the relationships of the index to other key socio-economic and environmental benchmarks. The statistical procedures applied in the preparation of the 2005 ESI report include cluster analysis, principal component analysis as well as stepwise and multiple regression models.

The Appendix is organized into four sections. The first section provides step-by-step explanations of the construction of the 2005 ESI. It is divided into sub-sections, which describe:

1. The selection criteria for the countries included in the ESI.
2. The standardization of the variables for cross-country comparisons.
3. The transformation of the variables for the imputation and aggregation procedures.
4. The multiple imputations algorithm used to substitute missing data.
5. The winsorization of the data.
6. The aggregation of the data to indicator scores and the final ESI score.

The next section discusses the important issues of data quality and coverage and how we have managed them in the 2005 ESI. We include the "country data review," which was carried out to crosscheck our data and to increase temporal and spatial coverage. In addition to identifying the best available data for the 2005 ESI, we also explain the logic and motivation for assessing the quality of all datasets used and provide detailed information on their sources.

The uncertainty and sensitivity analysis carried out in collaboration with the Joint Research Centre of the European Commission is presented in the third section. In a significant move towards greater transparency, we evaluate the major sources of uncertainty in the ESI, including missing data treatment, aggregation, and weighting. Each source of potential uncertainty is tested individually as well as jointly to estimate the impacts on the country rankings. The results are used to emphasize key limitations in the accuracy of the ESI scores, to address methodological criticism levied at previous ESI releases, as well as to strengthen the scientific basis for the policy conclusions presented in the report.

Finally, in the fourth section, we offer more detailed descriptions and results of the statistical analyses that form the backbone of our policy conclusions. The statistical tools used include principal component analysis, stepwise regression, and cluster analysis.

Principal component analysis is used to investigate the number of distinct dimensions that exist within the ESI indicator matrix and to show the influence of the indicators along these dimensions. It is furthermore used to determine a set of weights for the 21 indicators based on their statistical importance. These statistical weights are then compared with the equal weights used in the 2005 ESI.

In addition to identifying the most important indicators along the direction of the principal components, a stepwise regression analysis is conducted to determine the relative importance of the 76 variables with respect to the ESI score.

Because the ESI is a benchmarking tool for comparing national environmental stewardship, we emphasize the need to identify country peer-groups and "best practices" within those groups. We have conducted extensive cluster analyses, which identify seven relatively homogeneous country groupings with respect to the ESI indicators. This analysis brings to light several interesting patterns that cannot be attributed solely to the level of economic development. The characteristics with distinct patterns across the seven clusters, include population density, country size, and governance.

Calculating the ESI

1. Country Selection Criteria

A total of 146 countries met our inclusion criteria for the 2005 ESI. The decision to include a country in the index is based on country size, variable coverage, and indicator coverage as follows:

1. ***Country Size:*** Small countries are excluded. Countries with a total 2003 population under 100,000 or with land area under 5,000 square kilometers are excluded from the ESI because the nature of the interactions between elements of environmental sustainability are fundamentally different compared to larger countries. In particular, very small countries with large enough economies to be included in international data compilations resemble cities more than countries. They lack any sizable hinterland and have evolved to rely almost entirely on outsiders for provision of critical natural resources. Such profound differences make it difficult to justify including them in the same framework as other countries. However, separate ESI scores and component values for five small states are provided in Appendix E.
2. ***Variable coverage:*** While we seek to include as many countries as possible, the large number of missing observations makes it difficult to accurately and appropriately rank a country. We exclude countries that have observations for fewer than 45 of the 76 requisite data points for the ESI.
3. ***Indicator coverage:*** Some countries that survive the first two screens do not have even coverage across all 21 ESI indicators. We require that all countries in the ESI have observed variables for each of the ESI indicators, with two exceptions. Air Quality and Water Quality have relatively low country coverage across their constituent variables, but these indicators are judged too important to be eliminated. Because they are such vital issues, we want to retain the information we can for countries that report air and water quality, and we choose not to exclude the many countries that fail to report such data. If a country was missing *all* variables in *any* one of the remaining 19 indicators, it was removed.

2. Variable Standardization for Cross-Country Comparisons

To calculate the ESI scores for each country and to facilitate the aggregation of variables into indicators, the raw data need to be transformed to comparable scales. Some of the ESI variables already are denominated to make such cross-country comparison possible. Where this is not the case, we identify an appropriate denominator such as GDP, agricultural GDP, the total value of imports of goods and services, total population, the world average price of gasoline, city population, population aged 0-14 years, total land area, populated land area, as well as known amphibian, breeding bird, and mammal species.

3. Variable Transformation

After making the variables fit for cross-country comparisons, the next step is to prepare them for the imputation and aggregation processes. The procedure spelled out below explains the data transformations undertaken prior to and after the imputations, as well as the impacts they may have on the Environmental Sustainability Index scores.

First, we test all variables for normality of distribution. In many cases, the observations exhibit substantial skewness (see formula below). Most variables also exhibit patterns of heteroskedasticity, which means that the variance of the observations increases with the magnitude of the data. Both interfere with the imputation model's assumption of multivariate normality.

$$S_{x_j} = \frac{1}{\sigma_{x_j}^3} \frac{\sum_{j=1}^{p} (x_j - \mu_j)^3}{p}$$

A perfectly normally distributed variable is symmetric around its mean and hence has a skewness of zero. Skewed and/or heteroskedastic variables can be transformed to improve these properties but this may also change their distributions in ways that may affect the interpretation of the ESI scores. The logarithmic function, for example, is commonly used to reduce the influence of a few very large values by moving them closer to the mean. Similarly, it shifts very small values closer to the center of the distribution. Although the transformation may help approximating the normal distribution more closely, it will cause countries with exceptional values on a particular issue to no longer be such distinct outliers.

In addition to improving the imputation model, we also argue in favor of transformations as a means of reducing the impact of outliers on the ESI. In our experience, extremely small or large values have a relatively high probability of being measurement errors. A more normal, symmetric distribution implies that the majority of observations fall within two standard deviations of the mean (for a normal distribution, two standard deviations include 95% of the data) and extreme values occur with small probability.

However, in order to strike a balance between improving the distributional characteristics of the data and minimizing the impacts of the transformations on the ESI scores and ranks, we apply a 2-step procedure that recognizes the importance of normality for the imputations but its less significant value for the aggregation:

1. Prior to the generation of multiple imputations we transform all variables that have a skewness value larger than two using the base-10 logarithm or power transformations. In most cases the distributional effects of the transformations are beneficial.

2. After the imputations, we transform the variables back to their original scale with the exception of those variables with extreme skewness values of at least four (see Table A.1). In doing so, we ensure that only variables with extreme values outside four standard deviations are corrected for symmetry.

Table A.1: 2005 Environmental Sustainability Index – Variable Transformations after Imputations

Variable	Variable Code	Transformation	Constant*
Urban population weighted SO_2 concentration	SO2	Logarithm	0
Threatened mammal species as percentage of known mammal species in each country	PRTMAM	Logarithm	0
Freshwater availability per capita	WATAVL	Power ¼	1
Internal groundwater availability per capita	GRDAVL	Power ¼	0
Anthropogenic NO_x emissions per populated land area	NOXKM	Square root	0
Anthropogenic SO_2 emissions per populated land area	SO2KM	Logarithm	0
Anthropogenic VOC emissions per populated land area	VOCKM	Logarithm	0
Coal consumption per populated land area	COALKM	Square root	0
Vehicles in use per populated land area	CARSKM	Logarithm	0
Generation of hazardous waste	HAZWST	Power ¼	0
Industrial organic water pollutant (BOD) emissions per available freshwater	BODWAT	Square root	496
Fertilizer consumption per hectare of arable land	FERTHA	Square root	0
Pesticide consumption per hectare of arable land	PESTHA	Logarithm	0
Percentage of total forest area that is certified for sustainable management	FORCERT	Square root	0
Child death rate from respiratory diseases	DISRES	Square root	0
Average number of deaths per million inhabitants from floods, tropical cyclones, and droughts	DISCAS	Square root	0
IUCN member organizations per million population	IUCN	Square root	0
Local Agenda 21 initiatives per million people	AGENDA21	Logarithm	0
Number of ISO 14001 certified companies per billion dollars GDP (PPP)	ISO14	Square root	0
Carbon emissions per million dollars GDP	CO2GDP	Logarithm	0
Carbon emissions per capita	CO2PC	Logarithm	0

* If the observed minimum of the variable is negative, a constant is added such that the transformation of negative values can be computed. For example, if the minimum observed value is -5, a constant value of 6 is added to all observations before the logarithm or power transformation is computed.

4. Multiple Imputation of Missing Data

The question of how to treat missing or incomplete observations, which arise in virtually all types of environmental data collection, is among the most persistent and complicated problems facing policy analysts.

The degree of uncertainty due to the lack of data affects the ability to draw accurate conclusions and in many cases increases with the level of data aggregation. Insufficient data availability therefore has direct implications for effective and efficient decisionmaking.

We wish to minimize uncertainty and therefore attach substantial importance to the selection of the appropriate imputation method, i.e., the method used to fill data gaps with plausible estimates.

Two major assumptions are commonly made in the imputation literature:

1. The pattern of missing values in a multivariate vector of observations does not depend on the unobserved responses. In other words, the probability that a value is missing may be completely random (the statistical term is Missing Completely At Random or MCAR). Alternatively, it may depend on the observed values, which is called Missing At Random or MAR. The MAR assumption is more realistic for most real-life situations. If the parameters governing the missingness process are also independent of the parameters of the observed data model, the missing data mechanism is called “ignorable” and can be estimated.
2. A parameterized, functional form for the distribution of the vector observations can be formulated, and in most cases the estimates for the parameters of that form can be approximated using an iterative procedure (Johnson and Wichern 1998).

The following sections describe in detail how we selected and built the imputation model for the ESI.

Ad-hoc Methods v. More Sophisticated Approaches

The simplest ways of handling missing data are ad-hoc techniques such as *complete-case* and *available-case* methods (Little and Rubin 1987). The complete-case method uses only those observations for which all variables are observed. It is not applicable to the ESI because none of the 146 countries has observations for all 76 variables. We would hence be left with no observations in the imputation dataset.

The available-case method is based on analyzing subsets of the data for which all variables have been observed. For example, to impute missing water quality data using available cases, the imputation dataset could be limited to the water quality parameters only and all countries with one or more water quality parameters missing would be eliminated from this imputation dataset. Other variables are then imputed analogously.

It is apparent that both methods do not only lead to reduced ESI country coverage but also to potentially biased imputation results: both implicitly assume that the data are MCAR, which is highly improbable for the ESI data, because MCAR implies that all possible missingness patterns in the data matrix are equally likely.

Recognizing the complex relationships among the ESI variables we therefore opt for an imputation algorithm that broadens the base of actual experience, which allows us to involve as many countries as possible.

Table A.2 shows the top and bottom 20 countries in terms of data coverage. This list further corroborates that MCAR is not an appropriate model assumption for the ESI given the high correlation of data availability with level of income. We therefore investigated the use of a more sophisticated imputation model that does not require the data to be separated into subgroups and allows for the less restrictive MAR assumption.

The statistical foundation for dealing with ignorable MAR processes was developed in the 1970s but has been integrated only recently into standard statistical software packages. The essential idea behind MAR is that the probability that an observation is missing may not be completely random but depend on other observed variables.

More formally, if r_{ij} denotes a missingness indicator for country i and variable j, which is 1 if the country i has an observation for variable j and 0 otherwise, and if the data matrix X is partitioned into observed, X_o, and missing data, X_m, then,

$$P(r_{ij}=1 \mid X_o, X_m) = P(r_{ij}=1 \mid X_o)$$

For example, if variable X_2 is not collected anymore and is hence missing once the value for variable X_1 has reached a certain level, the probability that X_2 is missing given the value of X_1 is determined by X_1 and is a MAR process. In Table A.2 we can see a correlation between income per capita and the number of observed values. There are many other cases in which GDP per capita is a strong predictor for the values of ESI variables, and we utilized these relationships in the imputation model by including GDP per capita as an ancillary variable (see also the section dealing with deciding which variable to impute for a list of other ancillary variables).

Although the MAR assumption is more suitable for the ESI, we cannot determine if the assumption holds or if the missing data follow a non-ignorable process, i.e., a process in which the probability of X_2 missing not only depends on X_1 but also on the missing value itself.

So far, we only considered replacing a missing value with a single, plausible alternative, but imputation procedures can also generate multiple substitutes for a missing value. The key idea behind multiple imputations is to create a finite number of m completed data sets, each of which is then analyzed using

Table A.2: Countries with Highest and Lowest Data Coverage

Country	Observed	Missing	GDP / cap
Finland	75	1	$32,830
Germany	75	1	$32,800
Netherlands	75	1	$30,990
Austria	74	2	$34,240
Belgium	74	2	$31,390
France	74	2	$30,700
Ireland	74	2	$30,890
Italy	74	2	$21,480
Mexico	74	2	$3,720
Poland	74	2	$4,780
United Kingdom	74	2	$23,460
Canada	73	3	$23,840
Denmark	73	3	$39,720
South Korea	73	3	$15,290
United States	73	3	$32,510
China	72	4	$1,020
Greece	72	4	$14,760
Hungary	72	4	$5,940
Spain	72	4	$18,400
Switzerland	72	4	$45,980

Country	Observed	Missing	GDP / cap
Sudan	53	23	$350
Bosnia & Herze.	52	24	$1,720
Gabon	52	24	$4,370
Mauritania	52	24	$550
Myanmar	52	24	$1,800
Niger	52	24	$210
P. N. Guinea	52	24	$880
Yemen	52	24	$330
Dem. Rep. Congo	51	25	$90
Libya	51	25	$6,400
Sierra Leone	50	26	$170
Uzbekistan	50	26	$710
Turkmenistan	49	27	$1,050
Guyana	48	28	$940
Iraq	48	28	$1,500
Liberia	48	28	$190
North Korea	47	29	$1,300
Serbia & Montenegro	47	29	$1,900
Bhutan	45	31	$600
Guinea-Bissau	45	31	$160

Source for GDP per capita data: World Bank, World Development Indicators 2004. Data in constant 1995 US dollars.

standard statistical methods. The results of the *m* single analyses are combined to yield a final estimate of the parameter of interest. The advantage of using multiple imputations is that with repeated application of complete data analysis procedures, the uncertainty inherent in the imputation process can be captured in the variances within and between imputations.

We tested three different methods:

1. A simulation model using Markov Chain Monte Carlo (MCMC) techniques.
2. A regression-based modeling approach for missing data using observed values and existing correlations between the variables.
3. An Expectation-Maximization (EM) algorithm.

The Markov Chain Monte Carlo based imputation algorithm assumes multivariate normality of the data and generates imputations from the posterior distribution of the missing data given the observed data using a Bayesian approach. The missing data are presumed to be missing at random (MAR). Although in many cases the assumption of multivariate normality of the joint data distribution is not a realistic assumption, simulation tests have demonstrated relative robustness to deviations from this assumption (Little and Rubin 1987).

The regression imputation procedure is conceptually and computationally simple. Its underlying assumptions are that the marginal distributions of the data are normal and that linear relationships exist between the variables, which can be utilized for building linear regression models that predict the missing data. As with the MCMC model, the missing observations are assumed to be MAR.

The EM method uses an iterative process to estimate the mean vector and covariance matrix of the variables but does not generate multiple, independent draws from the data distribution. These can be obtained through the addition of a random noise, simulated from a specified distribution such as the standard normal distribution.

The relative usefulness of the three methods depends on the characteristics of the ESI data and the purpose of the analysis. Since we are interested in multiple imputations we eliminate the EM algorithm and compare the performance of the MCMC model with that of the regression model.

Comparison of Regression Imputation with MCMC Imputation

Using the ESI data, we generate imputations for both the MCMC and regression model and compare the results to see how robust the imputations and ESI scores and ranks are to the choice of imputation model. In general, we find that the differences in the results of the two methods with respect to the indicator values and ESI scores are limited, with a few exceptions. Table A.3 shows a sample of preliminary results for the ESI scores for both models using only ESI data in the first case and a set of additional socio-economic variables in the second.

Generally, we find that the inclusion of ancillary variables reduces the imputation variance of many variables that correlate with the additional data (for a list of ancillary variables refer to the sub-section Deciding Which Variables to Impute).

The ranks of the countries in the top and bottom quarter of the ESI appear to be relatively stable with only minor rank variation. Higher variation occurs in the middle 50% of the distribution. We attribute this in part to the heterogeneity of these countries with respect to environmental, institutional, and social circumstances and to the relative proximity of the ESI scores in the center of the ESI.

The deviation in means between variables imputed under the MCMC model and the regression model is higher when the fraction of missing data is large and when there are few comparable countries the imputation algorithm can build on to generate stable estimates. Variables that depend on largely unmeasured characteristics such as geography and climate are particularly affected. Such variables for which we do not have good "predictors" are used in the imputation model but are not imputed themselves (see Table A.4 for a complete list of not imputed variables.)

The relative robustness of the ESI ranks to the choice of imputation model, especially in the top and bottom quintiles, is further supported by the findings of the uncertainty and sensitivity analysis carried out with the Joint Research Centre of the European Commission, which is explained in the third section of this Appendix.

Although computationally more intensive, we use the MCMC method for the 2005 ESI because it provides the most flexible model for the ESI data and resulted in plausible imputations based on comparative tests among the three models. The exact procedure is described in the following section.

Table A.3: Impact of Imputation Model on 2005 ESI Ranks

Country	Regression		MCMC		Rank Standard Deviation	Average Rank
	No ancillary variables	With Ancillary variables	No ancillary variables	With Ancillary variables		
Finland	3	3	1	1	1.2	2.0
Sweden	1	2	4	2	1.3	2.3
Norway	2	1	2	3	0.8	2.0
Iceland	4	4	3	4	0.5	3.8
Switzerland	5	5	5	6	0.5	5.3
Canada	9	6	7	7	1.3	7.3
Austria	13	7	9	9	2.5	9.5
Australia	14	9	13	10	2.4	11.5
New Zealand	11	15	14	12	1.8	13.0
Gabon	10	17	10	18	4.4	13.8
Peru	25	18	17	20	3.6	20.0
Latvia	22	19	23	23	1.9	21.8
Colombia	60	57	22	30	19.1	42.3
Belgium	96	59	70	78	15.6	75.8
Italy	79	61	61	64	8.6	66.3
Nepal	54	63	60	58	3.8	58.8
Malawi	71	64	81	66	7.6	70.5
Chile	64	67	46	49	10.5	56.5
Myanmar	66	68	100	101	19.4	83.8
Belarus	49	69	64	76	11.5	64.5
Thailand	108	71	86	86	15.2	87.8
Chad	67	72	75	75	3.8	72.3
Ecuador	61	73	35	31	20.3	50.0
Cameroon	74	74	63	60	7.3	67.8
Madagascar	86	75	79	92	7.5	83.0
Gambia	63	76	98	97	17.0	83.5
Guinea	62	79	85	85	10.9	77.8
Russia	81	80	49	47	18.8	64.3
Côte d'Ivoire	44	81	94	98	24.6	79.3
Sri Lanka	80	82	68	83	7.0	78.3
Venezuela	123	85	76	74	22.8	89.5
Kazakhstan	105	86	91	84	9.5	91.5
Jordan	82	87	92	90	4.4	87.8
Guatemala	73	88	57	55	15.4	68.3
Benin	70	89	72	89	10.4	80.0
Senegal	83	90	88	80	4.6	85.3
Burkina Faso	41	91	93	87	24.8	78.0
Ukraine	113	92	102	105	8.7	103.0
South Korea	106	93	109	111	8.1	104.8
Iran	142	135	140	139	2.9	139.0
Syria	140	136	130	125	6.6	132.8
Libya	138	137	133	129	4.1	134.3
Uzbekistan	139	138	141	141	1.5	139.8
Nigeria	141	140	126	135	6.9	135.5
China	135	141	139	136	2.8	137.8
Kuwait	134	143	143	144	4.7	141.0
Saudi Arabia	144	144	145	146	1.0	144.8
Haiti	145	145	146	145	0.5	145.3
Yemen	143	146	144	143	1.4	144.0

Note: Results based on preliminary data, i.e., ranks do not in all cases correspond to final 2005 ESI ranking.

Markov Chain Monte Carlo Simulation

Markov Chain Monte Carlo (MCMC) simulation substitutes missing values with plausible quasi-random draws from their conditional distribution given the observed data. The MCMC approach assumes an ignorable MAR process for the missing data generating mechanism. The full data set, Y, is assumed to have a well-specified distribution, generally a multivariate normal distribution, with independent and identically distributed, or *iid*, observations. The missing values are then imputed iteratively in a Bayesian framework using a sequence of Markov Chains. Let the observed data be denoted X_o and the missing data X_m so that the full data matrix is given by $X=\{X_o, X_m\}$. The algorithm is as follows:

1. Given a prior distribution for the parameters $\boldsymbol{\theta}$ of the data model (in the case of the multivariate normal distribution the parameters are the mean and the covariance matrix) and an initial estimate of the parameters, $\boldsymbol{\theta}^{(0)}$, the missing data, X_m, are imputed through random sampling from the conditional distribution of the missing data, X_m, given the observed data, X_o, and the initial parameter estimates.
2. The completed data set is then used to update the initial parameter estimate by sampling from the joint posterior distribution of the parameters given in the completed data set. The new parameter $\boldsymbol{\theta}^{(1)}$ is then used to generate a new sample, $X_m^{(1)}$.
3. Iterating through steps 1 and 2 generates a Markov Chain of pairs of $(X_m^{(i)},\boldsymbol{\theta}^{(i)})$, which converges to the posterior conditional distribution of the missing data given the observed data. After a sufficiently long convergence time (burn-in), the first imputed data set can be drawn from the Markov Chain by sampling consecutively or every k^{th} draw ($k>0$).
4. Steps 1 to 3 are then repeated m times to generate m imputed data sets.
5. The m data sets are then analyzed individually and their results combined to a final ESI score for each country. From the m imputed data sets we can also obtain estimates of the standard errors of the missing data.

Number of Imputations

The larger the number of imputed values for each missing observation, the more that can be learned about the variation inherent in the missing observation. In the simplest case only one imputation (see single imputation methods discussed earlier) is generated. No statements can be made whether the substitute value is close to the "true" but unobserved value. The larger the number of imputations, the better our ability to estimate the variation and the more insight we have into the amount of missing information in the dataset and the band of uncertainty it creates.

Simulation studies have shown that for modest amounts of missing information (less than 30%), five to ten imputed datasets are sufficient to provide reasonable estimates of the parameters of interest.

Although we invested a great deal of effort in finding the most complete global data, the ESI still has approximately 18.6% empty cells in the data matrix. The amount of information missing may be somewhat higher depending on the importance of the variables with incomplete observations for determining a country's ESI. We therefore tested the robustness of the ESI by increasing the number of imputed datasets in our simulations from $m=10$ to $m=30$ and $m=100$.

With 30 or even 100 imputed datasets, it is possible to analyze not only the pattern of imputed values across countries for a specific variable, but also the distribution of the imputed values for a single country. We find that 30 sets of imputations provide a good compromise performance of the imputation model as well as computational efficiency.

Deciding which Variables to Impute

The ability of the imputation model to generate plausible and stable imputations depends not only on how well the data fit the model assumptions of MAR and multivariate normality but also on the inherent correlation structure.

For many aspects measured in the ESI we could identify predictor variables through correlation analysis. In addition to the existing observations for each variable, the observations of the predictors assist the model in generating more reasonable values. But we do not rely on the ESI variables alone. Previous releases of the ESI have already pointed out that certain ancillary variables such as transformations of GDP per capita, area, and population density can help to further fine-tune the predictions.

We therefore identified and include the following ancillary variables: populated land area (at least 5 persons per square kilometer), square of the base-10 logarithm GDP per capita, base-10 logarithm GDP per capita, health expenditure per capita, high technology exports as percentage of total exports, base-10 logarithm of total area, arable land as percentage of total land, base-10 logarithm of population, base-10 logarithm of population density, trade as percentage of GDP, and memberships in the Organisation for Economic Co-Operation and Development (OECD) and the Organization of the Petroleum Exporting Countries (OPEC). All data except for the populated land area dataset are from the World Bank's *World Development Indicators*.

Based on 30 fully imputed datasets, we compare the performance between imputations to check if the imputed values are stable. This is not the case for all variables. Variables that depend heavily on conditions not captured by the ESI or the ancillary variables, such as climatic, geographical, and many ecological factors, perform inadequately in the imputation model. These variables are therefore not imputed but used to assist in imputing missing values for variables that the ESI data and external data could impute in a stable manner. Table A.4 lists the variables that are not imputed.

In particular, we excluded Suspended Solids and SO_2 Exports from imputation because the results are too volatile and the fraction of missing values is large for both. We do not have sufficient confidence in being able to estimate their missing values with acceptable accuracy.

The final dataset is then obtained as the average of all values in each cell in the data matrix.

Table A.4: List of Variables not Imputed

Indicator	Variable	Code	Logic for not imputing
Biodiversity	National Biodiversity Index	NBI	Dependence on ecological and geographical factors not captured in ESI
Biodiversity	Percentage of country's territory in threatened ecoregions	ECORISK	Dependence on ecological and geographical factors not captured in ESI
Biodiversity	Threatened mammal species as percentage of known mammal species in each country	PRTMAM	Dependence on ecological and geographical factors not captured in ESI
Biodiversity	Threatened bird species as percentage of known breeding bird species in each country	PRTBRD	Dependence on ecological and geographical factors not captured in ESI
Biodiversity	Threatened amphibian species as percentage of known amphibian species in each country	PRTAMPH	Dependence on ecological and geographical factors not captured in ESI
Water Quality	Suspended solids	WQ_SS	High volatility of imputation results and dependence on factors not captured in the ESI
Water Quantity	Freshwater availability per capita	WATAVL	Dependence on ecological and geographical factors not captured in ESI
Water Quantity	Internal groundwater availability per capita	GRDAVL	Dependence on ecological and geographical factors not captured in ESI
Reducing Waste and Consumption Pressures	Generation of hazardous waste	HAZWST	Whether a country generates hazardous waste depends on factors not captured by the ESI.
Reducing Waste and Consumption Pressures	Waste recycling rates	RECYCLE	The data set is merged from two different sources, imputations would not be interpretable
Reducing Water Stress	Percentage of country under severe water stress	WATSTR	Dependence on ecological and geographical factors not captured in ESI
Natural Resource Management	Productivity overfishing	OVRFSH	Dependence on ecological and geographical factors not captured in ESI
Natural Resource Management	Salinized area due to irrigation as percentage of total arable land	IRRSAL	Dependence on ecological and geographical factors not captured in ESI
Natural Resource Management	Agricultural subsidies	AGSUB	Lack of information on external factors determining this variable
Reducing Environment-Related Natural Disaster Vulnerability	Average number of deaths per million inhabitants from floods, tropical cyclones, and droughts	DISCAS	Dependence on ecological and geographical factors not captured in ESI
Reducing Environment-Related Natural Disaster Vulnerability	Environmental Hazard Exposure Index	DISEXP	Dependence on ecological and geographical factors not captured in ESI
Environmental Governance	Local Agenda 21 initiatives per million people	AGENDA21	Lack of information on external factors determining this variable
Environmental Governance	Civil and Political Liberties	CIVLIB	Complete coverage
Environmental Governance	Percentage of variables missing from the CGSDI "Rio to Joburg Dashboard"	CSDMIS	Information which variables from the CSD CG list are missing cannot be imputed
Environmental Governance	Knowledge creation in environmental science, technology, and policy	KNWLDG	Lack of information on external factors determining this variable
Environmental Governance	Democracy measure	POLITY	Lack of information on external factors determining this variable
Eco-efficiency	Hydropower and renewable energy production as a percentage of total energy consumption	RENPC	Renewable energy sources depend on geography, climate, and other factors not captured by the ESI
Private Sector Responsiveness	Dow Jones Sustainability Group Index (DJSGI)	DJSGI	Not applicable
Private Sector Responsiveness	Average Innovest EcoValue rating of firms headquartered in a country	ECOVAL	Not applicable
Private Sector Responsiveness	Number of ISO 14001 certified companies per billion dollars GDP (PPP)	ISO14	Not applicable
Private Sector Responsiveness	Participation in the Responsible Care Program of the Chemical Manufacturer's Association	RESCARE	Not applicable
Participation in International Collaborative Efforts	Number of memberships in environmental intergovernmental organizations	EIONUM	Not applicable
Participation in International Collaborative Efforts	Participation in international environmental agreements	PARTICIP	Not applicable
Reducing Transboundary Environmental Pressures	SO_2 Exports	SO2EXP	Dependence on factors not captured in the ESI such as prevailing winds and geographical location

5. Data Winsorization

Following imputations, we "winsorize" or trim the tails of the variable distributions. Winsorization corresponds to shifting observations in the tails of the distribution to specified percentiles.

The purpose of the winsorization is to avoid having a few extreme values overly dominate the aggregation algorithm. We apply winsorization because we believe that such extreme values are more likely to reflect data quality problems in the tails of the distribution as opposed to values closer to the center of the distribution.

For each variable, the values exceeding the 97.5 percentile are lowered to the 97.5 percentile. Similarly, values smaller than the 2.5 percentile are raised to the 2.5 percentile.

Although we apply the transformation to every variable, the total number of affected values is very small. As another quality check on the imputations, we verified whether variables with imputed values have a higher degree of observations in the extreme tails. We observe a small, significant correlation between the number of winsorized values and the number of data points imputed for the 97.5% percentile, indicating that the imputation is more likely to generate large outliers than small outliers (see Table A.5).

The ESI could be criticized for using winsorization because it changes the distribution of the variable and either benefits or penalizes countries with values outside the center 95%. But our finding that winsorization affects only a very small fraction of the data and correlates with the imputations only to a small extent convinces us believe that its benefits outweigh this potential drawback. The Uncertainty and Sensitivity Analysis in Section 3 provides further support for this methodological decision.

6. Data Aggregation and Weighting

Aggregation

Composite indices are aggregations of sets of variables for the purpose of meaningfully condensing large amounts of information. Various aggregation methods exist and the choice of an appropriate method depends on the purpose of the composite indicator as well as the nature of the subject being measured.

The most common types of indices used are weighted sums and weighted geometric means of sub-components. The ESI belongs to the first group because it is the equally weighted sum of the 21 indicators:

$$I_i = \sum_{j=1}^{p} w_j \widetilde{X}_j \qquad i = 1,\ldots,n,$$

where w_j is the j^{th} weight given to $\widetilde{X}_j$, which corresponds to the z-score of the j^{th} indicator. Each indicator is itself a weighted sum of the 2 to 12 underlying variables. Within each indicator the variables are also weighted equally.

Weighted summations, in the form of averages, are not necessarily scale invariant. That means that the resulting index value, I_i, for the i^{th} object depends on the scales of the variables aggregated in the index.

Table A.5: Correlation between Number of Imputations and Number of Winsorizations.

Winsorization	Number of Imputations: Pearson	Kendall's Tau	Spearman's Rho
2.5 Percentile	0.16	0.12	0.18
97.5 Percentile	-0.25*	-0.20*	-0.24*
2.5 and 97.5 Percentile	0.06	0.03	0.04

* Correlation is significant at the 0.05 level (2-tailed).

Multiplicative expansions from one scale to another, for example, are abundant in the environmental domain. Because of this, the construction of indices based on weighted summation needs to take into account the possibility that the index values may change depending on the scale used.

The aggregation therefore requires that the $(n \times p)$ matrix X of n countries and p variables is normalized, i.e., all variables are on the same scale, in order to avoid distortions due to variables with very large values or variances. Most economic indices are built on a monetary unit of measurement, which provides a unified framework for comparing country performance. Environmental data do not generally have a common scale and normalization is necessary to remove the scale effects of different units of measurement without changing the relative distances between observations.

The ESI preserves the relative distances between countries' values by converting all variables to z-scores, which are obtained by subtracting the mean from the observation and dividing the result by the standard deviation of the variable. For variables in which high values correspond to low levels of environmental sustainability, we reverse the order by subtracting the observation from the mean and dividing the result by the standard deviation. In other words, for variables such as "percentage of land area under protected status" we use the conventional z-score, whereas for variables such as "percentage of mammals threatened" we produce a z-score in which higher percentages of threatened mammals correspond to lower levels of environmental sustainability.

Although normalization of the variables to z-scores removes the scale effects, z-scores depend on observed data statistics. They are "relative transformations" and change every time the ESI is updated due to shifts in the distribution of the variables over time. Furthermore, if all countries improve their performance on a given variable by the same amount between two time periods, the z-scores will remain the same even though performance has improved across the board.

The relationship among the variables and their individual contribution to the ESI merits significant attention. Linear weighted summation implies that the variables are preferentially independent (Munda and Nardo 2003b). Preferential independence means that the trade-off ratio between any two variables in a set, $\wp$, of variables is independent of the values taken on by the variables in $\wp^c$ (the complement of $\wp$). Under preferential independence, the summation of variables in the ESI corresponds to their marginal contributions to environmental sustainability, and requires that there are no synergistic or antagonistic effects among the variables. This is hardly a realistic assumption for environmental data. Given, for example, the proven synergistic relationships between several SO_2 and NO_x in the formation of acid rain, we cannot realistically assume preferential independence.

Weighted geometric mean aggregation is a potential alternative. It is defined as

$$I_i = \prod_{j=1}^{p} \left(w_j X_j\right)^{1/p} \quad i = 1, \ldots, n$$

Ebert and Welsch demonstrate that in the case of strictly positive, ratio-scale noncomparable variables, including many environmental variables, the aggregation by geometric mean can provide meaningful indices, i.e., indices with unambiguous orderings (Ebert and Welsch 2004). Another, more advanced approach is the multi-criteria decision method, which does not allow poor performance on one variable to be compensated by good performance on another.

When comparing the properties of the three different aggregation methods, the trade-offs become clear. When the objective is to design the best possible index, considerations of the most advanced statistical techniques available are important. On the other hand, if transparency and easy understanding by non-experts is equally important, the logical framework of the ESI represents a useful and valid alternative.

In an important expansion of our analysis of the properties of the ESI, we include in this Appendix the uncertainty and sensitivity analysis of the ESI, carried out by the Joint Research Centre of the European Commission in Ispra, Italy. The analysis identifies and quantifies the impact of the different sources of uncertainty in the ESI data as well as the effects of the weighting and aggregation methods on the rankings of the ESI.

Weighting

In composite indices, the choice of weights can reflect the importance given to the variables comprising the index or the substitution rates between them. In other instances, the weights are used to adjust for unequal variances of the variables, and hence their unequal levels of certainty. The specification of the weights is thus an integral part of index development and below we discuss our logic and motivation for choosing equal weights for the 21 indicators in the ESI.

Different methods to determine weights have been developed. They include data-dependent statistical tools as well as judgment-based expert opinions and budget allocation schemes.

Relative weights can be derived from least squares estimation, i.e., the line fitting method that minimizes the sums of squares of the relative distances of points from their expected value. Least squares minimization is the procedure underlying the linear regression model. A frequently occurring problem in least squares is that larger values tend to be associated with larger standard errors. Large observations will therefore have a disproportional influence on the sum of squares compared to smaller values. A weighted least squares approach corrects for this effect.

Principal component analysis and factor analysis are also useful statistical tools for estimating weights. They build on the relative importance of the variables for the principal components.

Statistically determined weights have the advantage that they apply a neutral and data-reliant weighting. However, statistical weights do not always reflect the priorities of decisionmakers or the budget constraints that limit free choice among a range of policy options.

Various methods for eliciting subjective preferences have been developed using elements ranging from budget allocation techniques to correspondence analysis. Regardless of whether the weights are determined statistically or subjectively, in most cases there exists no unique set of weights.

The ESI uses equal weights at both the indicator and the variable level. Our argument for equal indicator weights is based on the premise that no objective mechanism exists to determine the relative importance of the different aspects of environmental sustainability. At the country level, the indicators would almost certainly be weighted differently, but we cannot determine a globally applicable, differential set of weights that would allow a fair comparison between countries. As unsatisfactory as the choice of equal weights may appear, it is a neutral and justifiable allocation of importance across the indicators. Moreover, the principal component analysis in section 4 demonstrates that, even if the weights are determined through statistical means, no indicator stands out as being more or less important than others.

The variables within each of the 21 indicators are equally weighted because we think that each variable contributes roughly proportionately to the indicator to which it is allocated. In cases in which a country is missing a variable (and it is not imputed), the variable is not included in the average.

We note here that an interactive form of the ESI, that allows the user to set his or her own weights and to re-calculate an ESI based on these weights, is under development and will be made available on our website.

Data Quality and Coverage

1. Variable Grading

One of the most important conclusions of the ESI is the need for better data and a policy commitment to developing the necessary analytic underpinnings for a more data-driven approach to environmental decisionmaking. To further facilitate this process, we evaluated all ESI data sets with respect to the following criteria:

Relevancy:

1. The degree to which the variable matches the issue of interest.

Accuracy:

1. The reliability of the data source.
2. Whether the variable methodology is well established and widely adopted.
3. The availability of other data for cross-checking to assess the accuracy of the variable.

Coverage in space and time:

1. The availability of the most recent data.
2. The frequency with which the variables are updated.
3. The spatial coverage of the variable.
4. Whether the time series data can be constructed.

Certain variables are based on more than one data source, in which case, each data source is rated separately. In most cases, there are no deviations between the ratings of the sources. In the few instances where they are judged differently, this has been marked.

The evaluation of the variables was conducted by team members at the Yale Center for Environmental Law and Policy and Columbia University's Center for International Earth Science Information Network and combined into a single rating. The participants of the 2005 ESI Expert Review Meeting in December were also asked to comment on the preliminary "grades" and evaluations.

The evaluation process is inescapably subjective and limited by the knowledge base of the research teams. The goal of this exercise is not to establish a definitive quality assessment for each dataset, rather it is to begin a dialogue about data quality and to encourage further investments in data collection and methodological improvements. The grading scale used for the evaluation rates each variable according to its relevancy, accuracy, and coverage in space and time using grades ranging from A (Excellent) to F (Unacceptable), or U (Unknown).

The resulting matrix of variable grades summarizing our assessment of the relevancy, accuracy, and coverage of the variables in the ESI is shown in Table A.6.

Table A.6: Quality Assessment of ESI Variables

Component	Indicator Number	Indicator	Variable Number	Variable	Variable Description	Match between variable and issue	Reliability of data source	Variable methodology	Cross-check criteria	Most recent data set	Frequency of update	Spatial coverage	Consistent time series
Environmental Systems	1	Air Quality	1	NO2	Urban population weighted NO_2 concentration	A	A	B	C	A-	A-	D	B-
			2	SO2	Urban population weighted SO_2 concentration	A	A	B	B-	A-	A-	D	B-
			3	TSP	Urban population weighted TSP concentration	A	A	B	B-	A-	A-	D	C-
			4	INDOOR	Indoor air pollution from solid fuel use	B	B-	C	D+	A	U	B	F
	2	Biodiversity	5	ECORISK	Percentage of country's territory in threatened ecoregions	B	B	C	D	A	A	A	F
			6	PRTBRD	Threatened bird species as percentage of known breeding bird species in each country	B	A-	B+	B-	A	A-	A	A-
			7	PRTMAM	Threatened mammal species as percentage of known mammal species in each country	B	A-	B+	B-	A	A-	A	A-
			8	PRTAMPH	Threatened amphibian species as percentage of known amphibian species in each country	B	A-	A-	B-	A	B	A	B
			9	NBI	National Biodiversity Index	A	A-	B	B	A	U	B	D
	3	Land	10	ANTH10	Percentage of total land area (including inland waters) having very low anthropogenic impact	A-	B	B-	B-	A-	D	A	C-
			11	ANTH40	Percentage of total land area (including inland waters) having very high anthropogenic impact	A-	B-	B-	B-	A-	D	A	C-
	4	Water Quality	12	WQ_DO	Dissolved oxygen concentration	A	B+	B-	B	A	A	D	C+
			13	WQ_EC	Electrical conductivity	A-	B+	B-	B	A	A	D	C+
			14	WQ_PH	Phosphorus concentration	A	B+	B-	B	A	A	D	C+
			15	WQ_SS	Suspended solids	A	B+	B-	B	A	A	D	C+
	5	Water Quantity	16	WATAVL	Freshwater availability per capita	A	B	A-	B	C	A	B	A-
			17	GRDAVL	Internal groundwater availability per capita	A-	B	C-	C	A	C	B	D
Reducing Environmental Stresses	6	Reducing Air Pollution	18	NOXKM	Coal consumption per populated land area	A	A-	B	B+	B	B-	C	C-
			19	SO2KM	Anthropogenic NO_x emissions per populated land area	A	A-	A-	B+	B	B-	C-	C-
			20	VOCKM	Anthropogenic SO_2 emissions per populated land area	A	A-	B	B+	B	B-	C	C-
			21	COALKM	Anthropogenic VOC emissions per populated land area	C	A-	A-	A-	A	A	A	A
			22	CARSKM	Vehicles in use per populated land area	C+	A	A	A-	A	A	A	A-
	7	Reducing Ecosystem Stress	23	FOREST	Annual average forest cover change rate from 1990 to 2000	A-	A-	B-	B	B	B-	B	C-
			24	ACEXC	Acidification exceedance from anthropogenic sulfur deposition	A	C	B	B	D	F	A	F
	8	Reducing Population Pressure	25	GR2050	Percentage change in projected population 2004-2050	A	B	A	A	A	A	A	A
			26	TFR	Total Fertility Rate	A	A-	A	A-	A	A	A	A
	9	Reducing Waste & Consumption Pressures	27	EFPC	Ecological Footprint per capita	A	B	B-	C-	B	A-	B	B
			28	RECYCLE	Waste recycling rates	B	A	C	B-	B	A-	C	D
			29	HAZWST	Generation of hazardous waste	B	A-	B	A	B	A/B	C	C
	10	Reducing Water Stress	30	BODWAT	Industrial organic water pollutant (BOD) emissions per available freshwater	A	A	A-	A	B	A	B-	A-

Component	Indicator Number	Indicator	Variable Number	Variable	Variable Description	Match between variable and issue	Reliability of data source	Variable methodology	Cross-check criteria	Most recent data set	Frequency of update	Spatial coverage	Consistent time series
Reducing Environmental Stresses	10	Reducing Water Stress	31	FERTHA	Fertilizer consumption per hectare of arable land	B-	A-	B+	B	A	A	A	A
			32	PESTHA	Pesticide consumption per hectare of arable land	B	A-	A-	B	C	A	D	A
			33	WATSTR	Percentage of country under severe water stress	A	B	C	B-	C	C	B	D
	11	Natural Resource Management	34	OVRFSH	Productivity overfishing	B	B	C+	D	C	U	B	F
			35	IRRSAL	Salinized area due to irrigation as percentage of total arable land	A	B	C	D	C	C	D	F
			36	FORCERT	Percentage of total forest area that is certified for sustainable management	B	A	B+	B+	A	A	A	B
			37	WEFSUB	World Economic Forum Survey on subsidies	C	B	B-	D	A	A	A	B-
			38	AGSUB	Agricultural subsidies	B	A&B	A&C	B-	B	B	B	D
Reducing Human Vulnerability	12	Environmental Health	39	DISINT	Death rate from intestinal infectious diseases	B	A-	C-	B-	B	A	C	B
			40	DISRES	Child death rate from respiratory diseases	B	A-	C-	B-	B	A	C	B
			41	U5MORT	Children under 5 mortality rate per 1000 live births	C+	A	A	A	A	A-	A	A-
	13	Basic Human Sustenance	42	UND_NO	Proportion of undernourished in total population	B	A	B-	B	B	A-	A-	B
			43	WATSUP	Percentage of population with access to improved drinking water source	A	B	B	B	A	A	A-	B-
	14	Reducing Environment-Related Natural	44	DISCAS	Average number of deaths per million inhabitants from floods, tropical cyclones,	B+	A	B	B+	B	A	B	A
			45	DISEXP	Natural Disaster Exposure Index	B+	A	B	B+	A	A	B	A
Social and Institutional Capacity	15	Environmental Governance	46	PRAREA	Percentage of total land area under protected status	C	B+	B	A-	A	A-	A	B-
			47	GASPR	Ratio of gasoline price to world average	B-	A	B-	A-	A	A-	B	A-
			48	CSDMIS	Percentage of variables missing from the CGSDI "Rio to Joburg Dashboard"	B-	B	A	A-	A	A	B-	B
			49	KNWLDG	Knowledge creation in environmental science, technology, and policy	C	B+	D	B	A	F	C	C
			50	IUCN	IUCN member organizations per million population	B-	A	A	A-	A	A	A	A
			51	AGENDA21	Local Agenda 21 initiatives per million people	A-	B	A-	B-	A	B	C	B
			52	GRAFT	Corruption measure	A-	B-	B	D	A	B	A	B-
			53	LAW	Rule of law	A-	B-	B	C+	A	A-	A	B-
			54	CIVLIB	Civil and Political Liberties	A-	B	B-	D	A	A	A	A
			55	WEFGOV	World Economic Forum Survey on environmental governance	B-	B	B-	D	A	A	C	B-
			56	GOVEFF	Government effectiveness	A-	B-	B	C+	A	A-	A	B-
			57	POLITY	Democracy measure	B-	B+	A-	C	A	A	B	A
	16	Eco-efficiency	58	ENEFF	Energy efficiency	A	A	A-	A-	A	A	B	A
			59	RENPC	Hydropower and renewable energy production as a percentage of total energy	A	A	A	A-	A	A	B	A
	17	Private Sector Responsiveness	60	DJSGI	Dow Jones Sustainability Group Index (DJSGI)	B	B	C	C	A	A	D	B
			61	ECOVAL	Average Innovest EcoValue rating of firms headquartered in a country	A-	A	B	A-	A	A	D-	B
			62	ISO14	Number of ISO 14001 certified companies per billion dollars GDP (PPP)	B	B-	A	A-	A	A	A	A
			63	WEFPRI	World Economic Forum Survey on private sector environmental innovation	B-	B	B-	D	A	A	C	B-
			64	RESCARE	Participation in the Responsible Care, Program of the Chemical Manufacturers Association	C	A	D	A	A	A	A	B

Component	Indicator Number	Indicator	Variable Number	Variable	Variable Description	Match between variable and issue	Reliability of data source	Variable methodology	Cross-check criteria	Most recent data set	Frequency of update	Spatial coverage	Consistent time series
Social and Institutional Capacity	18	Science and Technology	65	INNOV	Innovation Index	B-	B	C+	A	B	B	B	B
			66	DAI	Digital Access Index	A-	A&B	C	B-	A	U	A	B
			67	PECR	Female primary education completion rate	B	B	B	A-	A	A-	B	A
			68	ENROL	Gross tertiary enrollment rate	C	B+	A	A-	A	A	A-	A
			69	RESEARCH	Number of researchers per million inhabitants	B	B	B	B	A	A	D	B
Global Stewardship	19	Participation in International Collaborative Efforts	70	EIONUM	Number of memberships in environmental intergovernmental organizations	B	B	D	D	A	A-	A	B
			71	FUNDING	Contribution to international and bilateral funding of environmental projects and development aid	B	A	C-	B-	A	B-	A	B-
			72	PARTICIP	Participation in international environmental agreements	B	A	D	A-	A	A	A	B-
	20	Greenhouse Gas Emissions	73	CO2GDP	Carbon emissions per million US dollars GDP	A	B+	B+	A-	A-	A	A-	A
			74	CO2PC	Carbon emissions per capita	A	B+	B+	A-	B+	A	A-	A
	21	Reducing Transboundary Environmental Pressures	75	SO2EXP	SO_2 exports	A-	B+	B	A/D	A	A/D	D	A/D
			76	POLEXP	Import of polluting goods and raw materials as percentage of total imports of goods and services	B	A	C	A	A	A	B	A

2. Country Data Review Initiative

One of our main objectives is to advance the global availability of reliable, timely, and comparable environmental information for environmental decisionmaking.

For this purpose, we provided our updated data for the 2005 ESI to the environmental ministries and statistical offices of 152 countries, requesting that they review the data for accuracy and provide, where applicable, corrections or recent updates.[2] We also set up a website through which we were able to provide regular updates and additional information on the ongoing data review process. A total of 62 countries responded to our request. Of these, 25 countries sent us updated and additional data and 14 provided useful feedback on methodological aspects of the ESI. Thirty-nine of the countries also sent us references to reports and websites or informed us that they had no comments on the data we sent (see Table A.7 for a detailed list of responses).

We also made it clear in our data review that we support the established environmental data collection activities of international institutions, especially the United Nations system of data collections, and requested that responses also be submitted to the respective international organizations compiling the statistics.

We utilized all information from the responses that was consistent with our methodology. Through the metadata provided by countries and follow-up communication with our contacts in the countries we were able to determine the consistency of the data with those provided by international sources. Table C.1 in Appendix C – Variable Profiles provides source information, including country sources where country data were incorporated, for all variables.

Table A.7: Responses by Countries that Provided Data

Country	Data	Reports/ Websites	Commentary	Other
Albania	♦			
Argentina				♦
Australia	♦			
Austria	♦		♦	
Azerbaijan				♦
Belarus				♦
Belgium	♦		♦	
Botswana				♦
Cameroon				♦
Canada	♦		♦	
Costa Rica	♦			
Croatia		♦		
Czech Republic				♦
Denmark		♦		
El Salvador				♦
Estonia		♦		
Finland	♦		♦	
France		♦		
Germany				♦
Greece				♦
Guatemala				♦
Hong Kong				♦
Hungary				♦
Iceland				♦
India				♦
Indonesia		♦		
Ireland	♦		♦	
Israel				♦
Italy	♦			
Japan	♦		♦	
Jordan	♦			

(continued)

Country	Data	Reports/ Websites	Commentary	Other
Korea	♦		♦	
Latvia				♦
Lebanon				♦
Lithuania	♦	♦	♦	
Madagascar	♦			
Malawi				♦
Malaysia				♦
Mauritius	♦	♦		
Nepal	♦			
New Zealand		♦		
Nigeria				♦
Pakistan				♦
Philippines		♦		
Poland	♦	♦		
Portugal				♦
Romania		♦		
Singapore	♦			
Slovak Republic			♦	♦
Slovenia		♦	♦	♦
South Africa	♦		♦	
Sweden				♦
Switzerland			♦	
Taiwan	♦		♦	
Thailand		♦		
Trinidad & Tob.				♦
Turkey	♦			
Uganda	♦			
United Arab Em.	♦			
United Kingdom	♦			
United States			♦	
Zimbabwe	♦			
TOTALS	25	13	14	26

3. Search for Additional and Better Data

In our attempt to update the ESI with the most recent, comparable, and high-quality data, we searched extensively for data to both improve current proxy variables in the ESI and to fill important gaps in the range of environmental, socio-economic, and institutional topics that the ESI indicators cover.

We carefully reviewed critiques of previous ESI reports and addressed a range of peer review comments to identify issues that are not adequately addressed by the ESI. An important outcome of this review and analysis is the revision of the ESI structure. The 2005 ESI includes 14 new variables, which are allocated to an improved 21-indicator framework. Two indicators – Natural Resource Management and Reducing Environment-Related Natural Disaster Vulnerability – have been added to the 2005 ESI. The Capacity for Debate indicator used in the 2002 ESI has been folded into the Environmental Governance indicator as we became convinced that they track the same phenomenon. The description and logic for each variable is given in Table A.8 while Table A.9 explains the replacements and deletions we have made in the variable composition.

Table A.8: Variable Additions to the 2005 ESI (alphabetical order)

Variable	Variable Description	Units	Logic
AGENDA21	Local Agenda 21 initiatives per million people	Number of Local Agenda 21 initiatives per million people	Local Agenda 21 (LA21) is an international sustainability planning process that provides an opportunity for local governments to work with their communities to create a sustainable future. The number of Local Agenda 21 initiatives in a country measures the degree to which civil society is engaged in environmental governance.
AGSUB	Agricultural subsidies	Percentage of total agricultural GDP (USD) that comes from subsidies	Agricultural subsidies reduce environmental sustainability primarily by creating price distortions, promoting the production of input intensive crops, wasteful use of natural resource inputs; use of marginal and fragile lands, and rent-seeking behavior.
DAI	Digital Access Index	Score between 0 and 1 with higher scores corresponding to better access	The Internet has created a new economy and promoted an unprecedented increase in the amount of environmental information that can be accessed and disseminated worldwide. Access to the Internet thus is important for access to information, stakeholder participation, decisionmaking, and generation of innovative solutions to environmental problems.
DISCAS	Average number of deaths per million inhabitants from floods, tropical cyclones, and droughts	Average number of deaths per million inhabitants	Vulnerability to natural disasters is a function of the severity of the hazard and the resilience of the socioeconomic system to perturbations. High vulnerability, as reflected in large numbers of disaster-related casualties, affects a country's ability to achieve longer-term sustainable development by redirecting resources to disaster recovery and reducing future resiliency.
DISEXP	Environmental Hazard Exposure Index	Average number of hazards to which the population is exposed (between 0 and 4)	Vulnerability to natural disasters is a function of the severity of the hazard and the resiliency of the socioeconomic system to perturbations. High exposure to natural hazards means that resources that could be used to achieve longer-term sustainable development must either be used for preventative measures or for disaster response.
FORCERT	Percentage of total forest area that is certified for sustainable management	Percentage of total forest area that is FSC or PEFC certified	This variable measures the extent to which a country seeks sustainable forestry practices.
GOVEFF	Government effectiveness	Z-score with high values corresponding to high levels of effectiveness	Governmental Effectiveness is defined in this data set as "quality of public service provision, the quality of the bureaucracy, the competence of civil servants, the independence of the civil service from political pressures, and the credibility of the government's commitment to policies." It is relevant for environmental sustainability because basic governmental competence enhances a society's ability to monitor and respond to environmental challenges.
GRDAVL	Internal ground water availability per capita	Thousand cubic meters per capita	Surface water is an important part of the picture of a country's water resources. The more groundwater is available per capita, the higher the probability that a country can sustainably manage its groundwater resources, e.g. for agricultural production.
INDOOR	Indoor air pollution from solid fuel use	Percentage of households using solid fuels, adjusted for ventilation	The public health community has drawn attention to the deleterious effects of indoor air pollution, especially on women who cook inside using solid fuels. High exposure to the fumes from solid fuel combustion is dangerous to human health. Solid fuel use has further consequences for deforestation and soil depletion because of dung collection.
IRRSAL	Salinized area due to irrigation as percentage of total arable land	Percentage of total arable land salinized due to irrigation	Soil salinization is a form of land degradation. The transport of salts to the land's surface due to irrigation renders the land unfit for production, and is therefore unsustainable in the long-term.
LAW	Rule of law	Z-score with high values corresponding to high degrees of rule of law	The rule of law is important in terms of establishing the "rules of the game" for the private sector, and for ensuring that violations of environmental regulations are enforced.
OVRFSH	Productivity overfishing	Score between 1 and 7 with high scores corresponding to overfishing	Overfishing of a country's exclusive economic zone is unsustainable.
POLEXP	Import of polluting goods and raw materials as percentage of total imports of goods and services	Import of polluting goods and raw materials as percentage of total imports of goods and services	Countries that import a large volume of commodities that are associated with negative environmental externalities at the point of extraction or processing may not be pursuing an environmentally sustainable path because of the likelihood that their actions are contributing to damage abroad. This measure does not take into account variation in actual environmental externalities within exporting countries, nor does it factor in other relevant imports that are not classified as commodities; as such it should be considered a rough proxy.
RESEARCH	Number of researchers per million inhabitants	Number of researchers per million inhabitants	Scientific capacity is important for the development of new technologies for sustainable environmental management.

The new variables greatly strengthen the ESI's capacity to assess key aspects of environmental sustainability. The gains emerge in some cases through better measures such as the Digital Access Index, which replaces the Number of Internet Hosts per million Inhabitants, or through incorporating a policy element that was previously unaddressed, such as Agricultural Subsidies as a proxy for agricultural sustainability and Indoor Air Pollution from Solid Fuel Use as a proxy for air quality.

In some cases, the new datasets are only rough gauges of issues we wish to track, e.g., overfishing and agricultural subsidies. But they reflect our best effort to produce a useful assessment of very complex concepts and to capture critical dimensions of sustainability that are often difficult to measure.

For other ESI variables, we could not identify better measures but succeeded in improving their geographical coverage by merging several data sources. In this context, several water and air quality variables were supplemented with information from additional sources. Despite their crucial influence on public health, infrastructure, and associated economic impacts, a real shortcoming exists with respect to ambient air pollution and water pollution. If it were not for their importance, the variables allocated to these two indicators would not have met our criteria for inclusion in the Index.

Table A.9: Summary of Changes in Variable Composition

Variable Replacements	
New in 2005 ESI	**Previously in 2002 ESI**
Percentage of variables missing from the CGSDI "Rio to Joburg Dashboard"	Percentage of ESI variables in publicly available data sets
Generation of hazardous waste	Radioactive waste
Gross tertiary enrollment rate	Technology Achievement Index
Digital Access Index	Technology Achievement Index
Percentage of total forest area that is certified for sustainable management	FSC accredited forest area as percent of total forest area
Female primary education completion rate	Technology Achievement Index
Participation in international environmental agreements	Percentage of CITES reporting requirements met; Participation in Vienna Convention / Montreal Protocol; Participation in Climate Change Convention
Contribution to international and bilateral funding of environmental projects and development aid	Global Environmental Facility participation; Participation in Montreal Protocol multilateral fund
Freshwater availability per capita	Internal renewable water per capita; Per capita water inflow from other countries
New or Additional Variables or Data Sources	**Logic**
National Biodiversity Index	Improving the Biodiversity indicator
Percentage of country's territory in threatened ecoregions	Improving the Biodiversity indicator
Threatened amphibian species as percentage of known amphibian species in each country	Improving the Biodiversity indicator
Knowledge creation in environmental science, technology, and policy	Knowledge generation in environmental science and policy facilitates development of innovative environmental technologies and policies
Participation in Responsible Care Program of the Chemical Manufacturer's Association	Voluntary and self-regulatory program of the chemical industry that, albeit non-binding, demonstrates willingness of private sector to take more responsibility for environmental protection and resource management
Waste recycling rates	Waste and consumption intensities can be counter-balanced by high resource recycling rates

Table A.9 continued on next page

New or Additional Variables or Data Sources	Logic
Dissolved oxygen	Increasing geographical coverage
Electrical conductivity	Increasing geographical coverage
Phosphorus concentration	Increasing geographical coverage
Suspended solids	Increasing geographical coverage
Anthropogenic SO_2 emissions per populated land area	Increasing geographical coverage
Anthropogenic NO_x emissions per populated land area	Increasing geographical coverage
Anthropogenic VOC emissions per populated land area	Increasing geographical coverage
Agricultural subsidies	Important proxy for measuring sustainable agricultural practices
Productivity overfishing	Important proxy for measuring sustainable fisheries management
Local Agenda 21 initiatives per million people	Gauges country's capacity and ability to implement sustainable development strategies at the local level
Average number of deaths per million inhabitants from floods, tropical cyclones, and droughts	Assessing a country's vulnerability to environmental disasters
Environmental Hazard Exposure Index	Assessing a country's vulnerability to environmental disasters
Government effectiveness	Effective government is important for sustainable natural resource use and management
Internal groundwater availability per capita	Supplementing surface water availability
Indoor air pollution from solid fuel use	Indoor air quality is at least as important an environmental health factor as ambient air quality
Salinized area due to irrigation as percentage of total arable land	Proxy for sustainable agricultural practices
Rule of law	Effective law enforcement is important for sustainable natural resource use and management
Number of researchers per million inhabitants	Gauges a country's capacity to generate and adopt innovative technologies and to implement them
Variables deleted	**Logic**
World Business Council on Sustainable Development memberships	Memberships do not imply tangible actions by private sector
CFC Consumption	CFC consumption successfully regulated under Vienna Convention and Montreal Protocol (and Amendments)
Subsidies for commercial fishing sector	Important but outdated data set
Total marine fish catch	Inadequate measure of transboundary pressure
Seafood consumption per capita	Inadequate measure of transboundary pressure

One solution to the problem of insufficient national data is to use modeling data. If the phenomenon of interest is regional or global in scope, complex modeling systems built on observed input data, for example meteorological records, can achieve astonishing accuracy. The ESI used data from several widely accepted models. The variables for which we adopted model estimates are water availability and water stress (WaterGap model version 2.1e, Kassel University, Germany), excessive acidification (Stockholm Environment Institute at York), long-range air transport of sulfur dioxide (Europe's EMEP program and IIASA), anthropogenic emissions of NO_x, SO_2, and VOCs modeled by the Intergovernmental Panel on Climate Change (IPCC), and populated land area measured as the area of a country with a population density of at least 5 people per square kilometer. This data set was constructed by CIESIN as part of the Gridded Population of the World GPW version 3 program using nine geospatially referenced input data sets.

Finally, we also received custom-made data sets from two private entities that evaluate corporate sustainability: the EcoValue21 rating from Innovest and the Dow Jones Sustainability Index from the Dow Jones Sustainability Group. These data sets have real limitations as proxies for private sector contributions to environmental sustainability. Notably, they are oriented to the environmental stewardship of large companies and are thus likely to be skewed toward efforts in the developed world. We include these variables to highlight the central role of business in the quest for environmental progress in every society. However, finding better ways to gauge private sector environmental performance must be seen as a high priority.

Despite our efforts to find data or build our own measures, persistent shortcomings exist with respect to long-term local, regional, and global environmental processes such as the evolution of biological diversity in ecosystems, the flux, dispersion, and deposition of long-range air pollution, and the monitoring of global weather, hydrological, and climate processes.

Enormous scientific progress has been made in understanding the functioning of these systems. However, global data availability is lagging behind. We believe that the Environmental Sustainability Index could be improved by including data on several variables, all of which are believed to have significant impact on natural resource use, human health, and ecosystem resilience. Among these variables are emissions of Persistent Organic Pollutants (POPs) as well as emissions of mercury and lead. However, we decided not to include any information on these variables in the Index because of their lack of sufficient quality and coverage. Other measures of importance but lacking data include toxic and solid waste management, wetlands loss, nuclear reactor safety, and sustainable agricultural practices.

Uncertainty and Sensitivity Analysis of the 2005 ESI

Prepared by Michaela Saisana, Michela Nardo, and Andrea Saltelli (Applied Statistics Group), Joint Research Centre of the European Commission

Sensitivity analysis is the study of how output variation in models such as the ESI can be apportioned, qualitatively or quantitatively, to different sources of variation in the assumptions. In addition, it measures how the given composite indicator depends upon the information that composes it. Sensitivity analysis is closely related to uncertainty analysis, which aims to quantify the overall variation in the countries' ranking resulting from the uncertainties in the model input.

A combination of uncertainty and sensitivity analysis can help to gauge the robustness of the ESI, to increase its transparency, and to frame policy discussions. The validity and robustness of the ESI depends on a number of factors including:

- The model chosen for estimating the measurement error in the data, which is based on available information on variance estimation.
- The mechanism for including or excluding variables in the index.
- The transformation and/or trimming of variables during the construction process of the index.
- The type of normalization scheme, such as re-scaling or standardization, applied to remove scale effects from the variables.
- The amount of missing data and the choice of imputation algorithm, in this case Markov-Chain Monte Carlo (MCMC) simulations or the EM algorithm.
- The choice of weights, e.g., equal weights or weights derived from factor analysis and expert opinion models.
- The level of aggregation, at the indicator or the component level.
- The choice of aggregation system, e.g., additive, multiplicative, or multi-criteria analysis.

All these assumptions can heavily influence the ESI country rankings and should be taken into account before attempting an interpretation of the results. The Joint Research Centre of the European Commission in Ispra, Italy, systematically evaluated the impacts that the above conceptual and methodological choices have on the robustness of the ESI ranking using uncertainty analysis and sensitivity analysis.

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 winsorization, and normalization to z-scores were found to negligibly change the country ranks. They are thus excluded from the results presented in this Section.

The output of interest in all tested scenarios of the sensitivity analysis is each country's rank. This is denoted $Rank_c$ for $c = 1,...,146$. The average shift, $\overline{R}$, in the ranks across countries, is calculated as the average of the absolute differences in countries' rank with respect to the original ESI rank:

$$\overline{R} = \frac{1}{146}\sum_{c=1}^{146}\left|Rank_{ESI2005,c} - Rank_c\right| \qquad (1)$$

We analyzed the following issues:

1. How do the ESI 2005 ranks compare to the most likely rank 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?

The sensitivity analysis procedure is a simulation-based procedure that acts on the equations that create the ESI model. Each equation corresponds to one step in the ESI construction. Although a range of methods exists for evaluating output uncertainty (Saltelli, Chan et al. 2000) we choose a Monte Carlo approach because it considers all uncertainty sources simultaneously. The simultaneity of the approach allows us to capture all possible synergistic effects among uncertain input factors, including their interactions as well as individual effects.

1. Our Approach

All uncertainties are then translated into a set of scalar input factors, which are sampled from their distributions (discrete in the case of triggers, or continuous in the case of imputed data). We specified the following inputs of uncertainty:

1. Imputation: We consider the variance associated with the *m=30* fully imputed datasets that are generated for each missing datum to construct a distribution centered around the mean. This allows us to study the effect of imputation variability on the ESI ranking.
2. Weighting schemes: We consider an expert opinion model as an alternative weighting scheme to the equal weighting approach used in the original ESI. A sample expert rating of a set of ESI indicators was obtained by averaging the opinion from 17 experts working on a broad spectrum of environmental sustainability and policy issues.
3. Aggregation level: We studied the impact of aggregation at the level of the five components compared to the 21 indicators in the original ESI.
4. Aggregation method: We compare the ESI's linear aggregation model with a non-compensatory multi-criteria model to account for the compensability issue among indicators.

By sampling the input space we obtained some *N=10,000* combinations of the 4 independent input factors $\mathbf{X}^l$, $l = 1,2,...N$, where N corresponds to the total number of simulations. For each trial sample $\mathbf{X}^l$, the ESI was computed, generating values for the scalar output variable of interest Y^l, where Y^l was either $Rank_c$, the rank assigned by the index to each country, or $\overline{R}$, the averaged shift in countries' rank. Each output vector, $\mathbf{Y}^l$, is then associated with the corresponding generating input vector $\mathbf{X}^l$.

For the choice of sampling method we consider simple random sampling, stratified sampling, quasi-random sampling and others (Saltelli, Chan et al. 2000). We use the sampling strategy based on Sobol sequences vectors (LP_τ sequences, (Sobol 1967)), which are quasi random sequences, to produce sample points that best scan the entire space of possible combinations between the input factors (Sobol 1976). Quasi-random sequences are used in place of random points to guarantee convergence of estimates. Moreover, Sobol sequences usually result in better convergence when employed in numerical integration. Bratley and Fox provide a good summary description (Bratley and Fox 1988).

The sequence of Y^l allows estimation of the empirical probability distribution function (pdf) of the output Y. The distribution reflects the uncertainty of the output due to the uncertainty in the input. Its characteristics, such as the variance and higher order moments, can be estimated with an arbitrary level of precision that only depends on the number of simulations, N.

The present analysis models several inputs of uncertainty simultaneously, which causes the index to be non-linear (Saisana, Tarantola et al. 2005). As argued by practitioners (Saltelli, Tarantola et al. 2000; EPA 2004), robust, "model-free" techniques for sensitivity analysis should be used for non-linear models.

Variance-based techniques for sensitivity analysis are model free and display the following additional properties convenient for the present analysis:

- Exploration of the whole range of variation in the input factors, instead of only sampling factors over a limited number of values, as done in fractional factorial design (Box, Hunter et al. 1978);
- Distinguish main effects (first order) and interaction effects (higher order);
- Easy interpretation and explanation;
- Simultaneous consideration of uncertainty factors;
- Justification of rigorous settings for sensitivity analysis, as is discussed later in this section.

2. Results and Discussion

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

The uncertainty analysis results of the 146 countries ranks are given in Figure A.1. Countries are ordered by their original ESI 2005 rank.

The original ESI ranks (grey mark) and the Monte Carlo based median ranks (black mark) rarely deviate: In most cases the $5^{th} - 95^{th}$ percentile bounds overlap the original 2005 ESI rank. For about 90 countries the difference between the ESI rank and the median rank when considering alternative approaches/assumptions is less than 10 positions.

This outcome reinforces the conclusion that the ESI is a fairly robust index. The main source of the variation is the combined effect of imputation and aggregation level. For countries in the first group, the average rank deviation is 7 positions, which increases to 12 positions for the second group and 11 for the third group. Surprisingly, both OECD and non-OECD countries have an average shift in rank of almost 9 positions. These findings indicate that the number of imputations for each country is less important than the imputation model itself.

Five countries have above average differences between the ESI rank and the simulation-based median rank: Mali, Nicaragua, Mongolia, Guinea-Bissau and Syria. The 2005 ESI rank for the first four countries is almost 35 positions higher when compared to their median rank, while the opposite is valid for Syria.

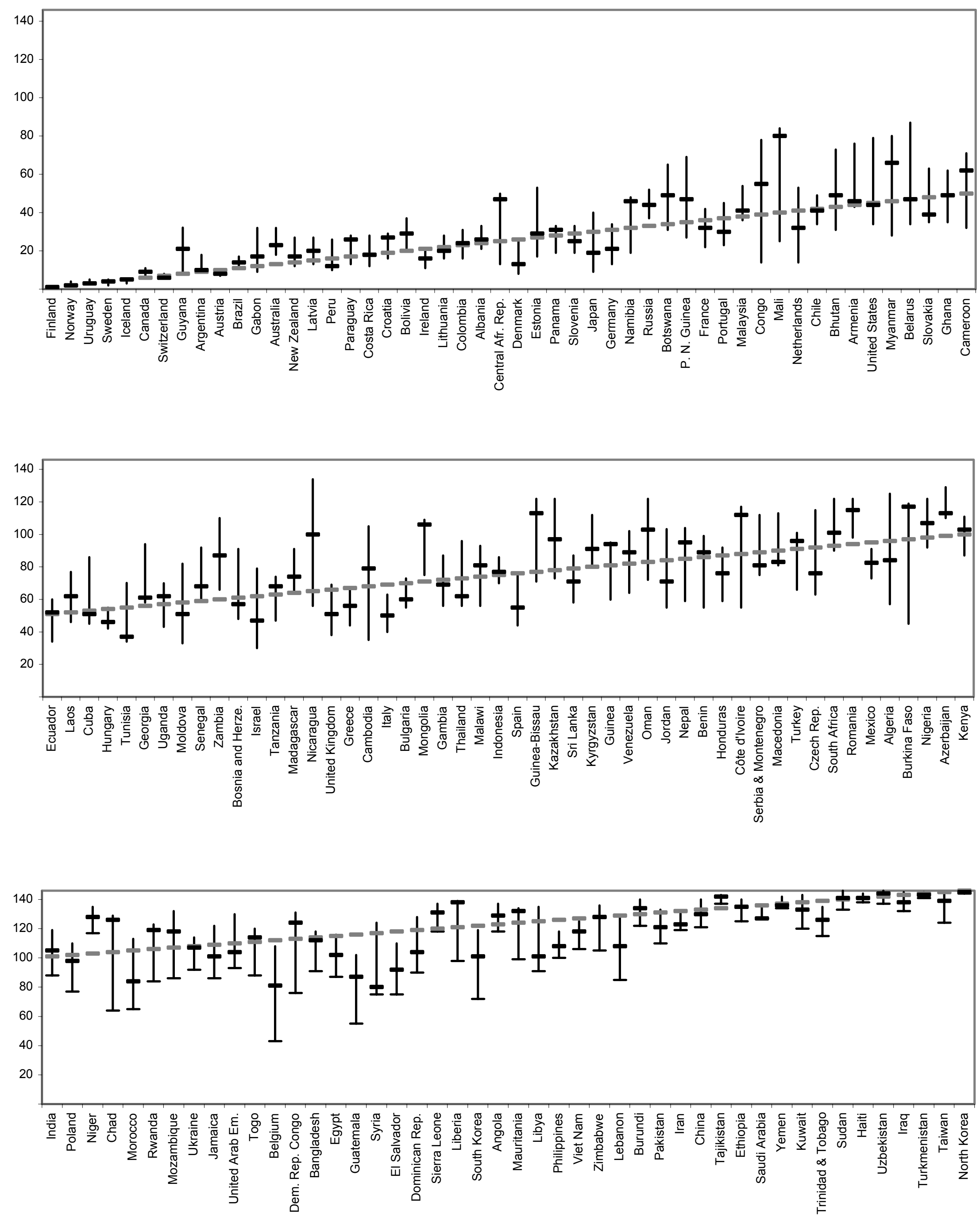

Figure A.1: 2005 ESI Rank v. Median Rank

Note: Grey marks correspond to actual ESI rank and black marks correspond to median rank. Whiskers show 5th and 95th percentiles (bounds) of rank distribution.

Table A.10: 2005 ESI Ranking and Optimal Rank for Each Country under All Combinations of Uncertainty Inputs

Country	2005 ESI Rank	Best Rank	Country	2005 ESI Rank	Best Rank	Country	2005 ESI Rank	Best Rank
Finland	1	1	Cameroon	50	32	Azerbaijan	99	110
Norway	2	2	Ecuador	51	34	Kenya	100	87
Uruguay	3	3	Laos	52	46	India	101	88
Sweden	4	2	Cuba	53	45	Poland	102	77
Iceland	5	3	Hungary	54	42	Niger	103	117
Canada	6	7	Tunisia	55	34	Chad	104	64
Switzerland	7	6	Georgia	56	58	Morocco	105	65
Guyana	8	9	Uganda	57	43	Rwanda	106	84
Argentina	9	9	Moldova	58	33	Mozambique	107	86
Austria	10	7	Senegal	59	59	Ukraine	108	92
Brazil	11	11	Zambia	60	66	Jamaica	109	86
Gabon	12	9	Bosnia & Herze.	61	48	United Arab Em.	110	93
Australia	13	18	Israel	62	30	Togo	111	88
New Zealand	14	12	Tanzania	63	47	Belgium	112	43
Latvia	15	13	Madagascar	64	65	Dem. Rep. Congo	113	76
Peru	16	10	United Kingdom	65	56	Bangladesh	114	91
Paraguay	17	13	Nicaragua	66	38	Egypt	115	87
Costa Rica	18	12	Greece	67	44	Guatemala	116	55
Croatia	19	16	Cambodia	68	35	Syria	117	75
Bolivia	20	20	Italy	69	40	El Salvador	118	75
Ireland	21	11	Bulgaria	70	55	Dominican Rep.	119	90
Lithuania	22	16	Mongolia	71	75	Sierra Leone	120	118
Colombia	23	16	Gambia	72	56	Liberia	121	98
Albania	24	21	Thailand	73	56	South Korea	122	72
Central Afr. Rep.	25	13	Malawi	74	56	Angola	123	118
Denmark	26	8	Indonesia	75	70	Mauritania	124	99
Estonia	27	17	Spain	76	44	Philippines	125	91
Panama	28	19	Guinea-Bissau	77	71	Libya	126	100
Slovenia	29	19	Kazakhstan	78	73	Viet Nam	127	106
Japan	30	9	Sri Lanka	79	58	Zimbabwe	128	105
Germany	31	13	Kyrgyzstan	80	81	Lebanon	129	85
Namibia	32	19	Guinea	81	60	Burundi	130	122
Russia	33	37	Venezuela	82	64	Pakistan	131	110
Botswana	34	31	Oman	83	72	Iran	132	119
P. N. Guinea	35	27	Jordan	84	55	China	133	121
France	36	22	Nepal	85	59	Tajikistan	134	137
Portugal	37	23	Benin	86	55	Ethiopia	135	125
Malaysia	38	36	Honduras	87	59	Saudi Arabia	136	127
Congo	39	14	Côte d'Ivoire	88	55	Yemen	137	134
Netherlands	40	25	Serbia & Montenegro	89	75	Kuwait	138	120
Mali	41	14	Macedonia	90	81	Trinidad & Tobago	139	115
Chile	42	34	Turkey	91	66	Sudan	140	133
Bhutan	43	31	Czech Rep.	92	63	Haiti	141	138
Armenia	44	43	South Africa	93	90	Uzbekistan	142	137
United States	45	34	Romania	94	98	Iraq	143	132
Myanmar	46	28	Mexico	95	73	Turkmenistan	144	141
Belarus	47	34	Algeria	96	57	Taiwan	145	124
Slovakia	48	35	Burkina Faso	97	45	North Korea	146	144
Ghana	49	35	Nigeria	98	92			

2. What is the optimal scenario for each country?

We interpret the 5th percentile of a country's rank distribution as its best rank. We note in Table A.10 that among the first 50 countries the most pronounced improvement in the performance are observed for Congo, the Netherlands, and Japan, which all gain more than 20 positions under a different scenario in the sensitivity analysis.

Among the countries ranked between 51st and 100th in the ESI, the most pronounced improvement under a different structure would have been for Burkina Faso and Algeria (gaining more than 40 positions in the ranking). In particular, Burkina Faso owes its improvement to the imputation, while Algeria improves its rank under aggregation at the indicator level.

Among the lowest ranked 46 countries, Belgium, South Korea and Guatemala display the most pronounced improvement (more than 50 positions). For Belgium and South Korea, this is due to aggregation at the components' level, while Guatemala's rank alters due to imputation.

3. Which countries have the most volatile rankings and why?

In order to provide an estimate of the magnitude of movement in ranks under the different simulation models, we define 'volatility' 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.

The volatility for the top ten countries, with the exception of Guyana and Argentina, suggests a robust performance for those countries. Guyana's high volatility of 23 positions is mainly attributed to the high variability in the imputation – 28 variables out of the total of 76 have been imputed – and its interaction with the aggregation level. Argentina's volatility of 9 positions is entirely due to imputation, although only 5 variables have been imputed.

Table A.11 presents the 15 countries that are affected the most by the construction procedure of the index. These countries ranked between 13 and 39 and experience differences in their best and worst ranks of 50 to 80 positions.

Only Congo, Mali, Myanmar and Belarus are ranked among the top 50 in the ESI. Their volatility can be attributed mainly to the interaction effect of imputation and aggregation level, as indicated by the Sobol sensitivity indices (1993). In some simulation runs the imputed values are favorable, partly compensating for the low scores in other variables and improving the country's rank. In other runs, however, the imputed value is far below average performance, which lowers the country's position.

4. What are the largest influences on the 2005 ESI?

To answer this question, we focus on the following comparisons:

- Imputation versus no imputation

Table A.11: Most Volatile Countries in the 2005 ESI

Country	Rank ESI	Range of Ranks	Country	Rank ESI	Range of Ranks
Congo	39	14 to 78	Côte d'Ivoire	88	55 to 117
Mali	41	25 to 84	Czech Rep.	92	63 to 115
Myanmar	46	28 to 80	Algeria	96	57 to 125
Belarus	47	34 to 87	Burkina Faso	97	45 to 119
Nicaragua	66	56 to 134	Chad	104	64 to 129
Cambodia	68	35 to 105	Belgium	112	43 to 108
Guinea-Bissau	77	71 to 122	Dem. Rep. Congo	113	76 to 131
Oman	83	72 to 122			

- Expert-weighting versus equal weighting of the 21 indicators
- Aggregation at the components level versus at the indicators level
- Non-compensatory aggregation scheme versus linear aggregation

Imputation

Imputation should be more influential for countries where missing data are a large problem. However, this relationship is not straightforward. Among the countries that miss almost 33% of their observations, only Guinea-Bissau and Myanmar are strongly affected by the imputations (Table A.12). Without imputation, Syria, Algeria, Belgium and the Dominican Republic improve their ranks between 29 and 37 positions. Conversely, Mali, Guinea-Bissau, Myanmar, and Zambia, decline 27 to 43 positions. Overall, the imputation has an average impact of 10 ranks and a rank-order correlation coefficient of 0.949.

Linear Weighting v. Budget Allocation (BA)

The ESI uses equal weights to calculate the country scores from the 21 indicators. As alternative weighting schemes we test a *"budget allocation scheme,"* in which the weights are obtained from experts with a demonstrated understanding of environmental sustainability.

For the ESI composite indicator, the 21 experts present at the December 2004 ESI Expert Review Workshop were each given a "budget" of 100 points and asked to allocate them to the 21 indicators according to their personal judgment of the relative importance of the indicators.

Four of those experts assigned zero priority points to a significant number of indicators and were therefore eliminated from the sample. The sets of weights obtained by the 17 remaining experts together with the overall average are listed in Table A.13.

The average expert weighting is slightly different from the equal weighting used in the ESI: the indicators within the Systems and Stresses components were weighted somewhat higher than the indicators within the Human Vulnerability, Social and Institutional Capacity, and Global Stewardship. Nevertheless, the variance of experts' opinions is rather large, varying from 40-80% of the mean weight. This explains the difference between the ESI ranking and the one provided by Budget Allocation. Overall, the weighting has an average impact of 5 ranks in the simulations and a rank-order correlation coefficient of 0.989 (Table A.14).

Table A.12: Most Improvement with Imputation v. No Imputation.

	Imputation	ESI Rank with Imputation	Rank without Imputation	Change in Rank
Improvement	Syria	117	80	*-37*
	Algeria	96	64	*-32*
	Belgium	112	82	*-30*
	Dominican Republic	119	90	*-29*
Deterioration	Mali	41	84	*+43*
	Guinea-Bissau	77	114	*+37*
	Myanmar	46	76	*+30*
	Zambia	60	87	*+27*
Average change over 146 countries:				*10*

Table A.13: Expert Group Weights for 2005 ESI Indicators

	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	0.1
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	0.1
Reducing Air Pollution	0.06	0.05	0.05	0.02	0.05	0.07	0.08	0.04	0.03	0.1
Reducing Ecosystem Stresses	0.06	0.05	0.06	0.02	0.05	0.05	0.06	0.06	0.03	0.02
Reducing Population Growth	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	0.1
Natural Resource Management	0.07	0.09	0.06	0.02	0.04	0.07	0	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
Reducing Environment-Related Natural Disaster Vulnerability	0.05	0	0.05	0.04	0.06	0.07	0	0.04	0	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	0.05	0.05	0.06	0.03	0.11	0.05
Participation in International Collaborative Efforts	0.04	0.02	0.03	0	0.04	0.04	0.04	0.02	0.03	0.02
Greenhouse Gas Emissions	0.04	0.02	0.03	0.1	0.06	0.09	0.07	0.05	0.03	0.1
Reducing Transboundary Environmental Pressures	0.06	0.05	0.03	0.06	0.04	0	0.06	0.05	0.03	0.02

	Experts								
	11	12	13	14	15	16	17	Average	Equal weighting
Air Quality	0.05	0.1	0.06	0.06	0.07	0.05	0.1	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	0.1	0.02	0.04	0.07	0.05	0.1	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	0.1	0.05	0.05	0.06	0.05	0.05	0.06	0.05
Reducing Ecosystem Stresses	0.05	0.1	0.08	0.05	0.04	0.05	0.07	0.05	0.05
Reducing Population Growth	0.05	0.01	0.06	0.05	0.05	0.02	0.01	0.04	0.05
Reducing Waste & Consumption Pressures	0.05	0.1	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	0.02	0	0	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
Reducing Environment-Related Natural Disaster Vulnerability	0.03	0	0.05	0	0	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	0.1	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 A.14: Most Improvement/Deterioration for Equal Weighting (EW) v. Budget Allocation (BA).

	Weighting	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 Emirates	110	127	*+ 17*
	South Africa	93	109	*+ 16*
	Italy	69	82	*+13*
	Nicaragua	66	78	*+ 12*
Average change over 146 countries:				*5*

Because the experts weighting assigns larger weights to indicators within the Systems and Stresses Components of ESI compared to the remaining indicators, it has a positive impact on the rank of countries such as Sri Lanka and Niger, but a negative effect on others such as the Chile, South Africa or Italy.

Aggregation at the Components Level v. Aggregation at the Indicators Level

In order to further assess the robustness of the ESI, we analyze the possibility of equally weighting the five components Environmental Systems, Reducing Environmental Stresses, Human Vulnerability, Social and Institutional Capacity, and Global Stewardship, instead of the 21 indicators.

Figure A.2 compares the ranking obtained from equally weighting the 21 indicators with those obtained by equally weighting the 5 components (indicators within component receive equal weight). We find that by changing the aggregation level, the average shift of the top 40 and the bottom 30 countries of the ESI 2005 is 7 positions and the shift of the remaining countries averages 11 positions. As expected, mid-level performers display higher variability than the top and bottom of the ranking.

Weighting the five components instead of the indicators affects only 38 countries by more than 10 positions. The average impact is 8 ranks and the rank-order correlation coefficient remains very high at 0.964.

If component weighting were used in the ESI, Belgium and South Korea would improve their ranks by almost 40 positions (Table A.15). On the contrary, countries such as Congo or Nicaragua would see their ranks decline by some 30 positions.

The reason for these substantial shifts for some countries is due to their relatively good performance in the systems and stresses components, which are more heavily weighted when the aggregation is takes place at the indicators level.

Linear Aggregation v. Non-Compensatory Multi-Criteria

The literature on index development offers a suite of aggregation techniques, including additive methods. However, additive aggregations imply certain properties and requirements for the indicators and the associated weights, which are often not desirable and at times difficult to verify. Other, less widespread, aggregation methods include multiplicative (geometric) and non-linear aggregations such as multi-criteria analysis.

Several authors (Debreu 1960; Keeney and Raiffa 1976; Krantz, Luce et al. 1971) note that an additive aggregation function for a given set of indicators exists only if these indicators are mutually preferentially independent. Preferential independence is a very strong condition because 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 the estimation of the marginal contribution of each indicator to the index. This marginal contribution can then be added together to yield a total value.

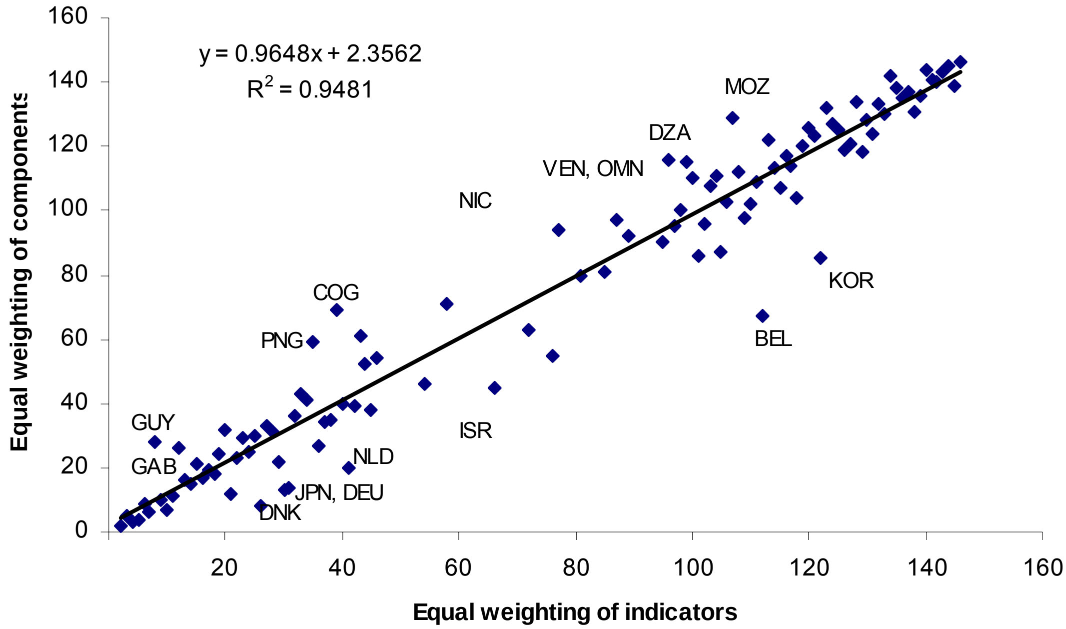

Figure A.2: Equal Weighting of the 21 Indicators v. Equal Weighting of the 5 Components.

Table A.15: Most Improvement/Deterioration in Ranks of Equal Weighting of Indicators (EWI) v. Equal Weighting of Components (EWC).

	Weighting	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	40	20	*– 20*
Deterioration	Congo	39	69	*+ 30*
	Nicaragua	66	93	*+ 27*
	P. N. Guinea	35	59	*+ 24*
	Venezuela	82	106	*+ 24*
	Oman	83	105	*+ 22*
Average change over 146 countries:				*8*

However, it is unrealistic to assume that no synergies exist among the indicators of the ESI (Funtowicz, G. et al. 1990). The combined impact of the acidifying substances SO_2, NO_X, NH_3 and O_3 on plant growth, for example, is substantially more severe that the (linear) addition of the impacts of each of these substances alone would be (Dietz and Straaten 1992).

Furthermore, linear aggregation entails full compensability: a poor performance in some indicators can be compensated by a good performance in others. Yet not everybody would trade an increase in the 'Participation in International Collaborative Efforts' indicator with a decrease in the 'Biodiversity' indicator. Taken to its extreme, full compensability implies that weights become substitution rates (e.g., how much 'Biodiversity' can be traded against 'Participation'), and do not indicate the importance of the indicator with which they are associated.

This means that a potential inconsistency exists between the way the weights are used and their theoretical meaning. For the weights to be interpreted as "importance coefficients" (e.g. place the greatest weight on the most important "dimension"), non-compensatory aggregation procedures should be used to construct composite indices (Podinovskii 1994). This can be done using a non-compensatory multi-criteria approach

A Non-Compensatory Multi-Criteria Approach (NCMC)

A non-compensatory multi-criteria approach (NCMC) is based on mathematical aggregation conventions that can be divided into two main steps, the pair-wise comparison of countries according to the whole set of indicators used and the ranking of countries in a complete pre-order.

The result of the first step is an ($M \times M$) matrix where M corresponds to the number of countries, commonly termed outranking matrix (Arrow and Raynaud 1986; Roy 1996). The information in the outranking matrix is used in the second step taking into consideration the intensity of preference (i.e., the difference in rank between countries for a given indicator); the number of indicators in favor of a given country; the weight attached to each indicator; and the relationship of each country with respect to all the other countries.

There are several ranking procedures for this second step (Young 1988). One possible algorithm is derived from the Condorcet-Kemeny-Young-Levenglick (CKYL) ranking procedure (Munda and Nardo 2003a). According to CKYL, the ranking of countries with the highest likelihood is the one 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 of overcoming 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, the approach does not require any transformation of the raw data, such as truncation, logarithmic transformation or normalization to assure the comparability of indicators.

Figure A.3 compares the results of the non-compensatory multi-criteria method with the ranking of the original ESI. In both cases we weight all 21 indicators equally. It is apparent that the aggregation method primarily affects the mid-range countries and, to a lesser extent, the laggards. Overall, the aggregation scheme has an average impact of 8 ranks and a rank-order correlation coefficient of 0.962, very similar to the impact of weighting the components instead of the indicators. In particular, while the top 50 countries move an average of only 5 positions, the next 50 countries' volatility averages 12 positions, and the lowest 46 countries shift ranks on average by 8 positions.

Both aggregation schemes, therefore, produce comparable rankings (the R^2 is 0.92). Using the NCMC, only 43 out of 146 countries display a change of more than 10 positions and none of these countries is in the top 30.

When compensability among indicators is not allowed, countries performing poorly on a number of indicators decline in rank while countries with moderate values tend to improve their situation. Table A.16 shows the countries displaying the largest variation in their ranks.

3. Conclusions

We can assess the validity of the ESI rankings by evaluating how sensitive they are to the assumptions that have been made in the structure and aggregation of the indicators. Uncertainty and sensitivity analysis allows us to assess the impact of four main methodological sources of uncertainty: variability in the imputation of missing data, equal versus

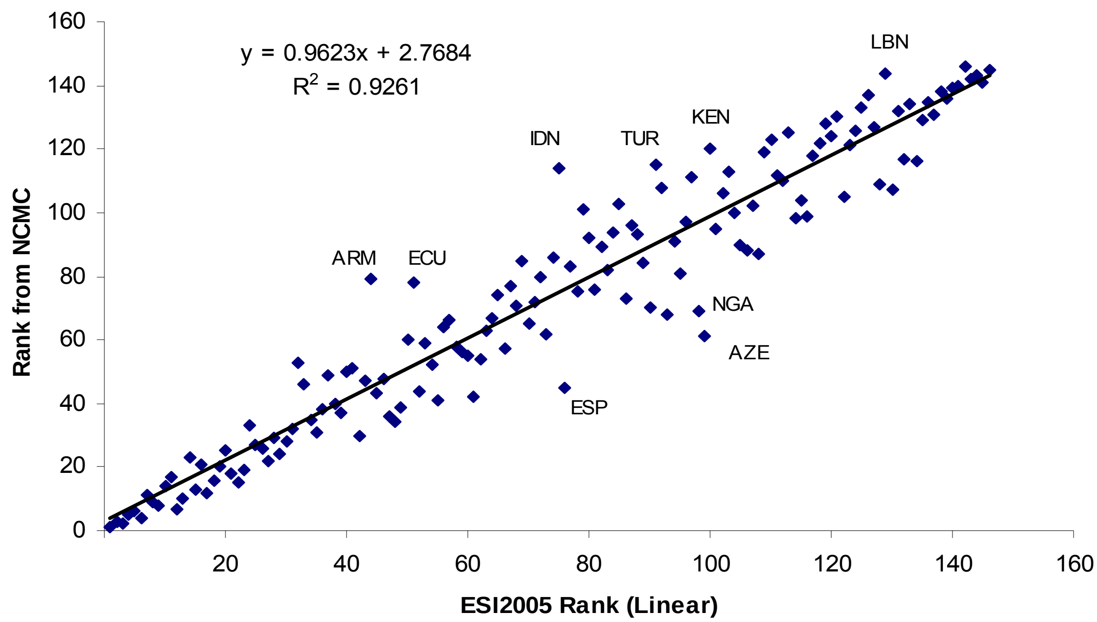

Figure A.3: Linear Aggregation of Indicators v. Non-Compensatory Multi-Criteria (NCMC) Aggregation of Indicators

Table A.16: Most Improvement/Deterioration in Ranks of Linear Aggregation (LIN) v. Non-Compensatory Multi-Criteria (NCMC) Aggregation.

	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*

experts opinion weighting of indicators, aggregation at indicators versus at components level, and linear verses non-compensatory aggregation scheme. The main findings can be summarized as follows:

Which countries have the most volatile ranks 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).

Would the ESI be more stable if no imputation had been carried out? Imputation should be more influential for countries where missing data are a large problem. However, this relation is not straightforward. Among the countries that are missing almost 33% of the observations, only Guinea-Bissau and Myanmar are highly impacted by imputation. If no imputation were carried out, Syria, Algeria, Belgium and Dominican Republic could improve their ranks by 9 to 37 positions. Conversely, Mali, Guinea-Bissau, Myanmar and Zambia would decline in the ranking by 27 to 43 positions. Overall, imputation changes a country's rank by 10 positions on average.

What if a "non-compensatory" aggregation scheme had been used, instead of the linear aggregation scheme? Aggregation schemes matter mainly for the mid-performing countries. When the assumption of compensability among indicators is removed, countries having very poor performance in some indicators, such as Indonesia or Armenia, decline in rank, whereas countries with fewer extreme values, such as Azerbaijan or Spain, improve their position. Overall, the aggregation scheme methodology has an average impact of 8 ranks.

What if aggregation had been applied at the component level instead of at the indicator 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 set of expert-derived weights had been used for the 21 indicators instead of the equal weighting? 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.

Statistical Analyses of the ESI for Policy Conclusions

1. Principal Component Analysis

Principal component analysis is a statistical method for identifying the key drivers or dimensions in a multivariate model. It is a useful tool to investigate the relationships between the 21 indicators in the ESI. This section describes in greater detail the steps and statistical assumptions underlying the method, followed by the results of applying principal component analysis to the ESI.

Principal component analysis is designed to summarize a p-dimensional dataset into a smaller number, q, of dimensions while preserving the variation in the data to the maximum extent possible. The objective to maximize the amount of variance explained is equivalent to losing as little of the information in the data as possible. The q new dimensions are constructed such that:

1. They are linear combinations of the original variables.
2. They are independent of each other.
3. Each dimension captures a successively smaller amount of the total variation in the data.

The number of linear combinations of variables can theoretically range from none to all p variables but the goal is to find the q $(0<q<p)$ of linear combinations of the p variables that "best" summarize the information in the data.

While principal component analysis provides considerable flexibility in determining q, the objective is to capture those features in the data that help better understand an issue of interest or to discover interesting new patterns among the relationships between variables.

The p original variables are combined into q linear combinations, which form the new principal components of the system. A standardized linear combination Z_1 of a data vector, $X_1=(X_{11}, X_{12}, \ldots, X_{1p})$, of length p is defined as:

$$Z_1 = w_1{}^t X_1 \text{, where } \sum_{i=1}^{p} w_i{}^2 = 1$$

Principal component analysis chooses the weights by determining the linear combination of all p variables in the transformed data set that maximizes the variance of the data. That is, the vector w of weights is calculated such that the squared difference of the new variable values and their respective means is maximized in relation to the total variance of the untransformed data.

The results for w_1 determine the first principal component. The second principal component with weights w_2 is then obtained analogously by maximizing the variance orthogonal to the direction of the first component. The third principal component with weights w_3 maximizes the residual variance in the direction orthogonal to the first and second components, and so forth.

The orthogonality of the principal components means that they are statistically independent. For example, if all water indicators of the ESI formed one principal component and all air emission indicators formed another, then any changes in either set of indicators would have no impact on the other.

The consecutive process of maximizing residual variance implies that at every step less variance is remaining. Once it falls below a specified threshold, the procedure is halted and no more additional principal components are calculated. Several criteria exist to determine the threshold value. One method considers the eigenvalues of the data matrix. The eigenvalue, λ, is the value that solves the equation

$$X_{corr} a = \lambda a \text{,}$$

where X_{corr} is the $(p \times p)$ correlation matrix calculated from the data for n countries and p variables and a is a vector in $\Re^p \neq 0$.

The eigenvalues, $\lambda_1, \ldots, \lambda_p$ decrease in magnitude: $\lambda_1 > \lambda_2 > \ldots \lambda_p$. The first λ_j that is less then 1 corresponds to the j^{th} principal component that explains less variance than is contained in the original, untransformed data. Values $\lambda<1$ therefore indicate that there is no gain to be expected from adding the principal component to the set of selected components. The first *(i-1)* components are sufficient to summarize the data.

Another rule of thumb for determining the number of principal components is to plot the eigenvalues in decreasing order and to connect the values in the plots by straight lines. The resulting plot is called a scree plot and usually has the form of an "elbow", starting from larger eigenvalues and dropping quickly to a lower value after which the decrease is more gradual until all *p* principal components are added to the system. The point where the transition from strong decrease in λ_i to λ_{i+1} to a more gradual decline occurs is often chosen for *q*. This "elbow" criterion generally tends to yield fewer components than the $\lambda<1$ criterion.

A third approach using the Longman-Allen values builds on the fact that in a random multivariate normal distribution, all eigenvalues should be of approximately the same size. A random *p* dimensional normal data set is generated and the eigenvalues calculated. They are added to the scree plot. All eigenvalues of the original data matrix *X* that lie above the Longman-Allen values signify principal components that represent non-random information in the data and should therefore be retained.

In the analysis of the ESI indicator data, we use eigenvalues and the scree plot to specify the number of principal components for the 21 indicators. The resulting factor loadings of the indicators on each principal component indicate their importance, i.e., the higher the loading of an indicator, the more useful it is for explaining variation in the direction of the principal component. Variables with similarly large loadings on the same principal component can be interpreted as being related along the direction of this component. The interpretation for the ESI is that these variables measure latent concepts such as air or water quality.

As noted earlier, the loadings from principal component analysis can also be treated as inherent weights of the variables or indicators for the aggregation process. As statistically derived weights they can be compared with:

1. The equal weights chosen for the ESI at both the variable and the indicator level.
2. The preferences a panel of experts would give to the 21 indicators of the ESI.

The uncertainty and sensitivity analysis in this Appendix analyzes the differences in these approaches with respect to the resulting ESI values and ranks.

Results of the Principal Component Analysis

Our results indicate the existence of six principal components for the 21 indicators, which explain more than 76% of the variation in the data. Although the number of components selected depends to a certain extent on the decision criteria chosen to determine the cut-off point for adding more components, the scree-plot, $\lambda>1$, and explained variance criteria all support the choice of six principal components (see Table A.17 and Figure A.3 for a summary of the results).

After deciding to keep six principal components in the model, we need to repeat the model to re-allocate the indicator loadings on the selected components. For better interpretability of the results, we choose a Varimax rotation, which rotates the principal components in six-dimensional space in such a way that maximizes each indicator's loadings on only one of the six directions. After 36 iterations the rotation algorithm has converged and the rotated component matrix is shown in Table A.18.

Table A.17: Determining the Number of Principal Components – Cumulative Variance Explained.

Principal Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	7.57	36.07	36.07
2	2.96	14.07	50.14
3	2.22	10.55	60.69
4	1.20	5.70	66.39
5	1.11	5.30	71.69
6	1.02	4.84	76.53
7	0.67	3.21	79.75
8	0.65	3.08	82.82
9	0.57	2.72	85.54
10	0.53	2.53	88.07
11	0.47	2.24	90.31
12	0.37	1.75	92.06
13	0.32	1.54	93.59
14	0.26	1.25	94.84
15	0.21	0.99	95.83
16	0.20	0.96	96.79
17	0.19	0.92	97.70
18	0.16	0.75	98.45
19	0.14	0.64	99.09
20	0.10	0.49	99.58
21	0.09	0.42	100.00

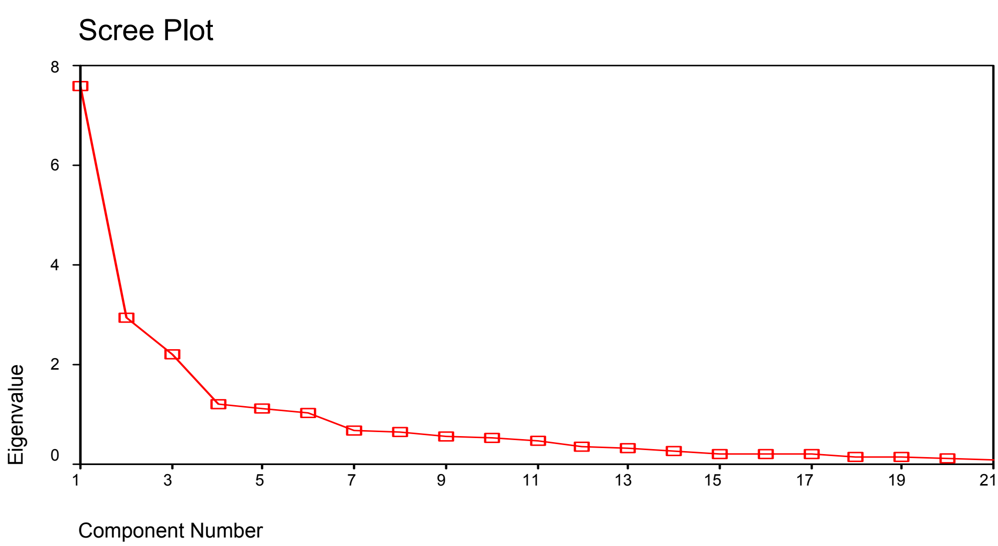

Figure A.4: Scree plot of Eigenvalues v. Principal Components

Table A.18: Rotated Component Loading Matrix

Indicator	Principal Component 1	2	3	4	5	6	Weights (scaled to 1)
Air Quality	0.17	-0.81	0.06	-0.1	0.27	0.19	0.05
Biodiversity	-0.20	0.32	0.15	0.04	0.59	-0.40	0.04
Land	-0.41	0.27	0.41	-0.5	0.22	-0.30	0.05
Water Quality	0.41	-0.08	0.71	-0	0.16	0.06	0.04
Water Quantity	-0.08	0.17	0.84	-0.1	0.01	-0.10	0.05
Reducing Air Pollution	-0.67	0.48	0.25	-0.1	0.12	0.11	0.05
Reducing Ecosystem Stresses	-0.14	-0.18	0.02	-0.8	0.00	0.07	0.05
Reducing Population Growth	0.54	-0.65	0.06	0.03	-0.20	-0.10	0.05
Reducing Waste and Consumption Pressures	-0.32	0.37	-0.14	0.51	-0.10	0.18	0.03
Reducing Water Stress	-0.55	0.38	0.47	0.16	0.26	0.10	0.05
Natural Resource Management	-0.72	-0.07	0.30	0.3	-0.10	-0.20	0.05
Environmental Health	0.70	-0.43	0.17	0.12	-0.30	-0.10	0.05
Basic Human Sustenance	0.68	-0.53	0.00	-0.1	-0.10	-0.20	0.05
Reducing Environment-Related Natural Disaster Vulnerability	0.07	-0.32	0.08	-0.1	0.81	0.14	0.05
Environmental Governance	0.86	-0.11	0.23	0.2	0.03	0.00	0.05
Eco-Efficiency	0.08	0.77	0.39	0.18	-0.10	0.15	0.05
Private Sector Environmental Responsiveness	0.89	-0.10	0.01	0.07	0.09	0.00	0.05
Science & Technology	0.79	-0.49	0.10	-0.1	-0.10	-0.10	0.06
Participation in Global Collaborative Efforts	0.76	0.34	0.04	0.05	0.00	-0.10	0.05
Greenhouse Gas Emissions	-0.07	0.80	0.20	0.22	0.04	0.33	0.05
Reducing Transboundary Air Pollution	-0.17	0.21	0.01	0.02	0.01	0.83	0.05

Rotation method: Varimax with Kaiser Normalization >=0.75* >=0.5* >=0.25*
* Absolute value

From Table A.17 we already expected that most indicators would load highly on the first, second, and third principal component because they have the highest eigenvalues. Since the eigenvalues are calculated using the correlation matrix of the input data, they represent the variance explained by each principal component.

The factor loadings matrix highlights which indicators load together on the same component as well as which indicators do not load strongly on any of the six components.

The results demonstrate several important characteristics of the ESI: Firstly, the ESI is a multidimensional index and environmental sustainability is a multidimensional concept. Although the number of principal components is smaller than the number of ESI indicators, six components are required to capture at least 75% of the variation in the data. The rotated principal components also load strongly on distinct sets of indicators, which corroborates our assumption that if the ESI were based on a small number of indicators such as the Human Development Index (HDI) produced by the United Nations Development Program (UNDP), it would not fully describe all dimensions of environmental sustainability.

Secondly, the analysis of the component loadings matrix in Table A.18 above suggests that some indicators relate more closely to each other than others. These sets of indicators have high loadings on the same principal component and in the same direction along the component.

Thirdly, since no indicator has low loadings on all six principal components, we can conclude that none of them is redundant in the calculation of the ESI.

Principal component 1 is determined predominantly by indicators belonging to the Social and Institutional Capacity component: Environmental Governance, Private Sector Environmental Responsiveness, Science & Technology, and Participation in Global Collaborative Efforts are the most influential indicators of this principal component. They are among the most influential indicators in the dataset, a result that confirms the findings of the correlation analysis, which also demonstrates that Environmental Governance and Participation in International Collaborative Efforts correlate most significantly with the overall ESI.

Aside from governance, principal component 1 is dominated by the indicators Natural Resource Management, Reducing Air Pollution, and Reducing Water Stress.

Other interesting findings exist for components 2 through 6. While the second component correlates strongly with several ESI indicators; most prominently with Air Quality, Eco-efficiency, and Greenhouse Gas Emissions; the remaining four components are determined by only one or two indicators. Given that all axes are orthogonal to each other, this means that the indicators loading on principal components 3 to 6 measure distinctly different aspects of environmental sustainability than are captured by components 1 and 2.

Component 3, for example, correlates most significantly with the quantity and quality of country's water resources as measures by Water Quality and Water Quantity.

Component 5, on the other hand, highlights clearly the importance of the new Vulnerability to Environment-Related Natural Disasters indicator. This indicator assesses a country's vulnerability to natural disasters that have a strong climate and weather component such as droughts, floods, and tropical cyclones. High losses of human and economic capital due to natural disasters reflect not only a natural geography-related susceptibility of the country to adverse catastrophic natural events but also a lack of capacity to prepare and deal with such events. The sixth dimension is dominated by the environmental impact countries have on other countries, as measured through the Reducing Transboundary Environmental Pressures indicator. It thereby emphasizes the importance of an indicator that is difficult to measure but vitally important to the overall sustainability picture.

The second important application of principal component analysis to the ESI consists of its ability to determine the statistical weights of the indicators. We calculate the weights of the 21 indicators as follows. Using the Varimax rotated component loading matrix, the six factor loadings of each indicator were squared to avoid negative weights and added together, thereby reflecting the total squared loadings across the six principal components. The sum of squared loadings for the 21 indicators was then re-scaled so that the final weights add up to 1. If an indicator has comparatively strong capacity to explain the variation in the data, it would be expected to receive a relatively high weight, and vice versa.

The weights estimated through principal component analysis for the 21 indicators are nearly identical, representing approximately 1/21. This finding lends further support to the choice of equal weights on the indicator level for calculating the ESI and supports the finding of the uncertainty and sensitivity analysis that budget allocation and the multi-criteria decision model do not substantially affect the ESI ranks. It should be noted, however, that weights estimated through principal component analysis reflect the average weight of each indicator, not the set of weights any particular country might apply in efforts to prioritize environmental policy.

2. Stepwise Linear Regression Analysis

Stepwise linear regression is an iterative regression method that determines the most influential variables among a set of variables. The three standard types of performing stepwise linear regression are *forward*, *backward*, and *exhaustive*. Although each method is built on the same objective of identifying the most powerful predictors in a regression model, the methods can lead to different answers.

Forward stepwise regression starts with a "zero-model" and adds one variable at a time. The variables with the highest R^2 are retained in the model and the search starts again for the next most powerful predictor, and so forth until all variables have been added. Cut-off values can be set to exclude those variables that do not add to the explanatory power of the model and to terminate the process once a desirable R^2 has been reached.

Backward stepwise regression is similar to the forward method but starts with the full model, i.e., all variables in the regression model. It then removes one variable at a time and excludes the variable that causes the smallest decrease in R^2. It then starts again removing one variable at a time, excluding the next worst predictor, and so forth until no more variables are left. Cut-off values can be set so as to avoid discarding useful variables and to prevent the model R^2 falling below a desirable level.

Exhaustive stepwise regression is a combination of the two methods above in that it adds and removes variables to find the best combination of predictors. This method is computationally much more intensive, especially as the number of variables in the data set increases, but has the advantage of performing the most extensive search for the best predictors.

We apply an exhaustive stepwise regression model to determine which of the 76 ESI variables are the most useful predictors of the ESI.

For the variable model, we set the entry level of significance to 0.05, i.e., for a variable to be included in the model, it must explain at least 5% of the ESI's variance. The level of significance to remain in the model is set to 0.10 or 10% of the variance in the ESI. After 45 iterations of the procedure no more change in the model composition occurs. Based on the adjusted R^2 value statistic we select a model with 12 variables, which cumulatively explain approximately 89% of the variation in the ESI.

The selected variables and summary statistics summary are shown in Tables A.19 and A.20. Overall, Air Quality, Imports of Polluting Goods, Water Quality and Quantity, Environmental Governance, Fertility Rates, High Anthropogenic Land Conversion, and Deaths from Natural Disasters are the most important predictors for the ESI. The results thereby confirm the studies that have focused on "governance" as a critical driver of policy success (Esty and Porter 2001) but also suggest that environmental quality and stresses have important implications for the ESI scores.

It should be noted, however, that due to differential weighting of variables in the global Index, the importance of the variables as determined by statistical analysis is somewhat confounded with the magnitude of the implicit weights for each variable. Implicit weights for individual variables range from 1/42 for variables in indicators with only two variables (such as Water Quantity and Eco-Efficiency) to 1/252 for the 12 variables in the Environmental Governance indicator.

Table A.19: Summary of Stepwise Regression Variable Selection (Transformed variables)

Model		Unstandardized Coefficients			
Variable	Variable Description	beta	Std. Error	t	p-value
(Constant)	Intercept	49.88	0.23	216.61	<0.0001
DISRES	Child death rate from respiratory infections	2.17	0.35	6.29	<0.0001
WATAVL	Water availability per capita	3.23	0.28	11.70	<0.0001
WEFGOV	World Economic Forum Survey on environmental governance	4.37	0.40	11.00	<0.0001
COALKM	Coal consumption per populated land area	1.91	0.34	5.69	<0.0001
FERTHA	Fertilizer consumption per hectare of arable land	1.67	0.33	5.03	<0.0001
POLEXP	Import of polluting goods and raw materials as percentage of total imports of goods and services	1.50	0.26	5.73	<0.0001
WQ_DO	Dissolved oxygen concentration	1.48	0.33	4.51	<0.0001
TFR	Total Fertility Rate	2.51	0.37	6.75	<0.0001
ANTH40	Percentage of total land area (including inland waters) having very high anthropogenic impact	1.93	0.35	5.59	<0.0001
GASPR	Ratio of gasoline price to world average	1.32	0.32	4.19	<0.0001
SO2KM	Anthropogenic SO_2 emissions per populated land area	1.23	0.35	3.55	<0.0001
DISCAS	Average number of deaths per million inhabitants from floods, tropical cyclones, and droughts	0.81	0.26	3.14	<0.001

Table A.20: Stepwise Regression Model Summaries for 1 to 12 Variables.

Model	R^2	Adjusted R^2	Std. Error of the Estimate	Change Statistics				
				R^2 Change	F Change	df1	df2	Sig. F Change
1	0.35	0.35	6.84	0.35	78.72	1	144	0
2	0.55	0.55	5.7	0.2	64.03	1	143	0
3	0.68	0.67	4.85	0.13	55.63	1	142	0
4	0.74	0.73	4.39	0.06	32.14	1	141	0
5	0.77	0.76	4.12	0.03	20.3	1	140	0
6	0.8	0.79	3.86	0.03	20.47	1	139	0
7	0.83	0.82	3.56	0.03	25.13	1	138	0
8	0.85	0.84	3.36	0.02	18.5	1	137	0
9	0.87	0.86	3.15	0.02	19.85	1	136	0
10	0.89	0.88	2.96	0.02	19.26	1	135	0
11	0.89	0.89	2.87	0.01	8.88	1	134	0
12	**0.9**	**0.89**	**2.78**	**0.01**	**9.87**	**1**	**133**	**0**

3. Cluster Analysis

Cluster analysis is a statistical technique used to separate a large group of objects into sub-groups with similar characteristics. We use this technique to identify groupings of relevant peer countries.

Within each peer group, countries have a better basis for benchmarking their environmental performance because the group members are the most homogeneous with respect to their ESI indicators and the differences across the groups are maximized.

Using the ESI indicators to determine peer groups of countries for finding common benchmarks for performance evaluation is of enormous value. Cluster analysis helps to advance this process by grouping beyond level of development alone. In doing so, it enables countries to identify others who are similarly situated – thus providing a good place to start in the search for best practices.

We tested hierarchical agglomerative and divisive clustering methods as well as different distance metrics but found that Ward's method of agglomerative clustering consistently produced the best results.

A feature of agglomerative clustering is that it starts with as many individual clusters as there are countries. It then successively combines countries that are most similar to each other with respect to a quantitative similarity measure until all countries are joined in a single cluster. The similarity measure decreases during this process, while the within-cluster dissimilarity increases as more and more countries are added. The trade-off lies therefore in choosing a similarity measure, or "pruning value", that yields both a relatively small number of clusters and a high level of similarity. We determine that 7 clusters yield a reasonable division between the countries.

Another clustering method, we use the *k* means algorithm developed by Hartigan and Wong (Hartigan and Wong 1979) to determine cluster membership of the countries. *K* means is a non-hierarchical method that requires that the number of clusters, *k*, be specified upfront (hence the preliminary use of Ward's method). It then iteratively finds the disjoint partition of the objects into *k* homogeneous groups such that the sum of squares within the clusters is minimized. The algorithm converges in fewer than 1000 iterations. The cluster membership is shown in Table A.21. Table A.22 provides additional cluster information.

Table A.21: Cluster Membership for k Means Clustering

Low system and stress scores; low vulnerability and high capacity; moderate stewardship	Moderate system and stress scores; high vulnerability and low capacity; above average stewardship	Above average system score; low vulnerability; high capacity; moderate stresses and stewardship	Moderate system, stresses, and capacity scores; low vulnerability and stewardship	Above average system score, moderate stresses, vulnerability, capacity, and stewardship	Moderate system, stresses, and vulnerability scores; low capacity and stewardship	Low system score; moderate stresses, vulnerability, capacity, and stewardship
Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Austria	Angola	Australia	Bosnia & Herze.	Argentina	Algeria	Albania
Belgium	Benin	Canada	Bulgaria	Bolivia	Armenia	Bangladesh
Denmark	Bhutan	Finland	Croatia	Botswana	Azerbaijan	China
France	Burkina Faso	Iceland	Czech Rep.	Brazil	Belarus	Cuba
Germany	Burundi	New Zealand	Estonia	Chile	Iraq	Dominican Rep.
Ireland	Cambodia	Norway	Greece	Colombia	Kazakhstan	Egypt
Israel	Cameroon	Sweden	Hungary	Costa Rica	Kuwait	El Salvador
Italy	Central Afr. Rep.	United States	Jamaica	Ecuador	Kyrgyzstan	Georgia
Japan	Chad		Latvia	Gabon	Libya	India
Netherlands	Congo		Lebanon	Guatemala	Moldova	Indonesia
Portugal	Côte d'Ivoire		Lithuania	Guyana	Mongolia	Iran
Slovenia	Dem. Rep. Congo		Macedonia	Honduras	North Korea	Jordan
South Korea	Ethiopia		Poland	Namibia	Oman	Malaysia
Spain	Gambia		Romania	Nicaragua	Russia	Mexico
Switzerland	Ghana		Serbia & Montenegro	Panama	Saudi Arabia	Morocco
Taiwan	Guinea		Slovakia	Paraguay	Turkmenistan	Pakistan
United Kingdom	Guinea-Bissau		Trinidad & Tobago	Peru	Ukraine	Philippines
	Haiti		Turkey	Uruguay	United Arab Em.	South Africa
	Kenya			Venezuela	Uzbekistan	Sri Lanka
	Laos					Syria
	Liberia					Thailand
	Madagascar					Tunisia
	Malawi					Viet Nam
	Mali					Zimbabwe
	Mauritania					
	Mozambique					
	Myanmar					
	Nepal					
	Niger					
	Nigeria					
	P. N. Guinea					
	Rwanda					
	Senegal					
	Sierra Leone					
	Sudan					
	Tajikistan					
	Tanzania					
	Togo					
	Uganda					
	Yemen					
	Zambia					

Table A.22: Additional Characteristics of Clusters

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
	Number of countries	17	41	8	18	19	19	24
	Average ESI scores	52.9	47.1	66.3	49.6	57.1	44.0	46.2
Average values of ESI Component Values	Environmental Systems	39.1	50.8	75.6	43.4	66.9	51.5	37.4
	Reducing Environmental Stresses	33.9	54.7	44.0	50.9	55.7	52.6	50.9
	Reducing Human Vulnerability	71.3	26.6	78.0	72.2	51.0	54.2	49.4
	Social and Institutional Capacity	77.7	36.1	83.5	52.3	52.1	29.6	44.4
	Global Stewardship	57.5	63.6	49.4	31.4	54.5	26.8	52.2
Average values of other character-istics	GDP/capita	$27,480	$420	$29,860	$4,390	$2,980	$3,810	$1,730
	Population (million)	33.6	19.0	46.1	11.8	21.2	20.7	149
	Total Area (thousand square kilometers)	171	539	3,466	123	1,026	1,561	1,010
	Population Density (per square kilometer)	238	70.3	13.5	122	32.1	56.0	174
	Environmental Governance Indicator (z-score)	1.0	-0.5	1.0	0.2	0.1	-0.6	-0.2

The cluster analysis reveals clear linkages between group membership and the average performance along the five ESI components. It also suggests the existence of relationships between cluster membership and additional characteristics such as average income per capita, population density, and area size.

The geographic pattern of the clusters is striking, especially since no geographical data was used in the analysis. We interpret this feature as a result of the many similarities of countries in close geographical proximity in regard to environmental conditions and pressures, economic and trade linkages, as well as with respect to social and cultural communalities.

Cluster 1 and 3 represent the developed countries with 24 of 29 OECD countries present (Luxembourg is too small to be included in the ESI). Interestingly, Taiwan is a member of cluster 1, which is characterized by high population density and industrialization combined with high social and institutional capacity. With the exception of South Korea and Taiwan, these countries share high to moderately high ESI scores. Although Taiwan is likely to be seen as an outlier in the group, its cluster membership suggests that its indicator values are more similar to this group of countries than to any of the remaining six clusters.

The differentiation between cluster 1 and 3 appears to follow characteristics captured in the distribution of ESI scores between developed and developing countries and further fine-grains the results of the analysis into the relationships between economic development and environmental sustainability. Despite comparable per capita incomes and good environmental governance, the average ESI scores for cluster 1 and 3 are markedly different (excluding the low scores of South Korea and Taiwan from cluster 1 only lifts the average ESI score by 2 points). The most prominent difference exists in the Environmental Stress component. Clearly, developed countries with large land area, low population densities – by far the lowest of all 7 clusters – and a rich natural resource base enjoy a comparative advantage because the absorptive capacity of their environments is bigger than that of smaller sized, high population density, developed countries. Although we try to correct the variables underlying the indicators for the most prevalent distortions due to size, the cluster results indicate that large area size is advantageous for environmental sustainability.

Cluster 2 is composed of the least developed countries that are characterized by weak governance and high human vulnerability. Another group of developing countries is formed by cluster 7. Cluster 2 and 7 differ in their average population size as well as their Environmental Systems and Human Vulnerability components scores. Cluster 7 includes four of the most populous countries in the world: China, India, Indonesia, and Bangladesh. Only the large geographic area of several countries in this cluster reduces their average population densities to more moderate values. This cluster's average ESI scores are only slightly higher than those of Cluster 6, which includes many of the lowest ranked countries in the ESI.

Cluster 4 includes many Eastern European countries with moderate incomes but relatively high environmental stresses, which might be a legacy of their former economic systems as well as their high average population density.

Cluster 5 comprises most of the Latin American countries, and has the second highest average ESI score and population density after Cluster 3. The good performance of the countries in this cluster has already become apparent in the high rankings of Uruguay, Guyana, Argentina, Brazil, Peru, Paraguay, and Costa Rica among the top 20 ESI countries.

Cluster 6 by contrast, has the lowest average ESI scores and is characterized by very low average scores for Social and Institutional Capacity and Global Stewardship. The countries of the Middle East and Central Asia dominate this cluster. The group is characterized by moderate environmental systems and stresses scores as well as an average human vulnerability to environmental shocks. The Social and Institutional Capacity and Global Stewardship components are the lowest across the seven clusters.

Appendix A References

Arrow, Kenneth J. and Hervé Raynaud (1986). *Social Choice and Multicriterion Decision Making*. Cambridge: M.I.T. Press.

Box, George, William Hunter, et al. (1978). *Statistics for Experimenters*. New York: John Wiley and Sons.

Bratley, Paul and Bennett L. Fox (1988). "Algorithm 659 Implementing Sobel's Quasirandom Sequence Generator." *ACM Trans.* Math. Software 14: 88-100.

Debreu, Gerard (1960). "Topological Methods in Cardinal Utility Theory". *Mathematical Methods in Social Sciences* (K. J. Arrow, S. Karlin and P. Suppes, ed.). Stanford: Stanford University Press.

Dietz, Frank J. and Jan van der Straaten (1992). "Rethinking Environmental Economics: Missing Links Between Economic Theory and Environmental Policy." *Journal of Economic Issues.* XXVI (1): 27-51.

Ebert, Udo and Heinz Welsch (2004). "Meaningful Environmental Indices: A Social Choice Approach." *Journal of Environmental Economics and Management.* 47: 270-283.

Environmental Protection Agency (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).

Esty, Daniel C. and Michael E. Porter (2001). "Ranking National Environmental Regulation and Performance: A Leading Indicator of Future Competitiveness?" *The Global Competitiveness Report 2001* (M. E. Porter, J. Sachs and et. al., ed.). New York: Oxford University Press.

Funtowicz, Silvio O., Giuseppe Munda, et al. (1990). "The Aggregation of Environmental Data Using Multicriteria Methods." *Environmetrics.* Vol. 1(4): 353-360.

Hartigan, John and M.A. Wong (1979). "A k-means Clustering Algorithm." *Journal of Applied Statistics.* 28: 100-108.

Johnson, Richard A. and Dean W. Wichern (1998). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.

Keeney, Ralph and Howard Raiffa (1976). *Decision with Multiple Objectives: Preferences and Value Trade-offs*. New York: Wiley.

Krantz, David H., R.Duncan Luce, et al. (1971). *Foundations of Measurement, Vol. 1, Additive and Polynomial Representations*. New York: Academic Press.

Little, Roderick J.A. and Donald B. Rubin (1987). *Statistical Analysis with Missing Data*. New York: John Wiley.

Munda, Giuseppe and Michela Nardo (2003a). *Methodological Modelling of Composite Indicators for Ranking Countries*, Joint Research Center, Ispra (www.lamsade.dauphine.fr/mcda/biblio/Year/2003.complete.html).

Munda, Giuseppe and Michela Nardo (2003b) *Mathematical Modelling of Composite Indicators for Ranking Countries*, Proceedings of the First OECD/JRC Workshop on Composite Indicators of Country Performance, JRC, Ispra.

Organisation for Economic Co-operation and Development (OECD) (2003). *Composite Indicators of Country Performance: A Critical Assessment*, DST/IND (2003):5. Paris.

Podinovskii, Vladislav V. (1994). "Criteria Importance Theory." *Mathematical Social Science.* 27: 237-252.

Roy, Bernard (1996). *Multicriteria Methodology for Decision Analysis*. Dordrecht: Kluwer.

Saisana, Michaela, Stefano Tarantola, et al. (2005). "Uncertainty and Sensitvity Techniques as Tools for the Analysis and Validation of Composite Indicators." *Journal of the Royal Statistical Society A.* 168(2): 1-17.

Saltelli, Andrea, Karen Chan, et al. (2000). *Sensitivity Analysis*. Probability and Statistics Series. New York: John Wiley & Sons.

Saltelli, Andrea, Stefano Tarantola, et al. (2000). "Sensitivity Analysis as an Ingredient of Modelling." *Statistical Science.* 15: 377-395.

Sobol, Ilya M. (1967). "On the Distibution of Points in a Cube and the Approximate Evaluation of Integrals." *USSR Journal of Computational Mathematics and Mathematical Physics.* 7: 86-112.

Sobol, Ilya M. (1976). "Uniformly Distributed Sequences with an Addition Uniform Property." *USSR Journal of Computational Mathematics and Mathematical Physics.* 16: 236-242.

Sobol, Ilya M. (1993). "Sensitvity Analysis for Non-linear Mathematical Models." *Mathematical Modelling & Computational Experiment.* 1: 407-414.

Ting, H.M. (1971). *Aggregation of Attributes for Multiattributed Utility Assessment*. Technical Report n. 66, Operations Research Center. Cambridge: MIT.

Young, Hobart P. (1988). "Condorcet's Theory of Voting." *American Political Science Review.* 82 (4): 1231-1244.

Endnotes

[1] For more information on the statistical analyses included in the 2005 ESI report, please contact the Project Director, Tanja Srebotnjak, at Tanja.Srebotnjak@Yale.edu.

[2] To identify contact addresses for environment ministries and national statistical offices we used several sources, including the database on statistical offices of the United Nations Statistics Division (UNSD) and the list of environmental ministries of the United Nations Environment Programme (UNEP). We were unable to find contact details for a small number of environmental ministries and some request were returned as undeliverable.

This page is intentionally blank.