JRC Scientific and Technical Reports

Sensitivity Analysis of the 2008 Environmental Performance Index

Michaela Saisana and Andrea Saltelli

EUR 23485 EN 2008

JRC
EUROPEAN COMMISSION

ipsc
Institute for the Protection and Security of the Citizen

The Institute for the Protection and Security of the Citizen provides research-based, systems-oriented support to EU policies so as to protect the citizen against economic and technological risk. The Institute maintains and develops its expertise and networks in information, communication, space and engineering technologies in support of its mission. The strong cross-fertilisation between its nuclear and non-nuclear activities strengthens the expertise it can bring to the benefit of customers in both domains.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen
Centre for Research on Lifelong Learning (CRELL)

Contact information
Address: Michaela Saisana, JRC, TP361, via E. Fermi 2749, 21027 (VA), Italy
E-mail: michaela.saisana@jrc.it
Tel.: +39-0332-786572
Fax: +39-0332-785733

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

Composite Indicators website: http://composite-indicators.jrc.ec.europa.eu/

JRC 47026

EUR 23485 EN
ISBN 978-92-79-09702-7
ISSN 1018-5593
DOI 10.2788/91982

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Sensitivity Analysis of the 2008 Environmental Performance Index

Michaela Saisana and Andrea Saltelli

Executive Summary

An assessment of the robustness of the 2008EPI results requires the evaluation of uncertainties underlying the index and the sensitivity of the country scores and rankings to the methodological choices made during the development of the Index. To test this robustness, the EPI team has continued its partnership with the Joint Research Centre (JRC) of the European Commission in Ispra, Italy.

This JRC report shows that the 2008 EPI has an architecture that highlights the complexity of translating environmental stewardship into straightforward, clear-cut policy recipes. The trade-offs within the index dimensions are a reminder of the danger of compensability between dimensions while identifying the areas where more work is needed to achieve a coherent framework in particular in terms of the relative importance of the indicators that compose the EPI framework.

The 2008 Environmental Performance Index (EPI) is developed for 149 countries and is based on 25 indicators in six policy categories: Environmental Health, Air Pollution, Water, Biodiversity and Habitat, Productive Natural Resources, Climate Change. The EPI aims to bring a data-driven, fact-based and empirical approach to environmental protection and global sustainability.

The validity of the EPI scoring and respective ranking is assessed by evaluating how sensitive the country ranks are to the assumptions made on the index structure and the aggregation of the 25 underlying indicators. The assumptions tested are:

- measurement error of the raw data,
- choice of capping at selected targets for the 25 indicators,
- choice to correct for skewed distributions in the indicators values,
- weights of the indicators and/or of the subcomponents of the index, and finally
- aggregation function at the policy level (six policy categories).

The main conclusions are summarized below.

How do the EPI ranks compare to the ranks under alternative methodological approaches?

A total of 40,000 simulations were run in order to cover the space of uncertainties present in the 2008 EPI. We discuss ranks and not scores because non-parametric statistics are more appropriate in our case given the non-normal character of the data and the scores. In the relevant literature, the median rank is proposed as a summary measure of a rank distribution. The median rank of all combinations of assumptions indicates that for 1 out of 2 countries in the EPI, the difference between the EPI rank and the median rank is less than 15 positions (out of 149 countries). This modest sensitivity of the EPI ranking to the five assumptions (eventual measurement error in the raw data, correction of skewed data distribution, use of target values, weighting of the indicators, and aggregation function) implies a reasonably high degree of robustness of the index for those countries. For the remaining half of the countries, the EPI performance is highly sensitive to the methodological choices in the index, and should thus be considered as merely indicative. A discussion on the top performing countries is in place. The top ten performing countries in the EPI include Switzerland, Sweden, Norway, Finland, Costa Rica, Austria, New Zealand, Latria, Colombia and France. However, the simulations indicate that most of those countries should be positioned much lower. Switzerland, for example has a probability of only 31% to be ranked in the top ten countries, whilst even lower is the probability for Austria, Latvia and France. In our simulations, New Zealand scores 98% of the times in the top ten, followed by Finland, Costa Rica and Colombia. Panama, whose EPI rank is 32, should actually be considered as a top ten performing country, given that its score is among the top ten in 73% of the simulations.

Which are the most volatile countries and why?

There are several countries with a relatively high difference between their best and worst rank. A very high volatility of more than 80 positions is found for Hungary (rank: 23), Denmark (25), Albania (27), Ireland (34), Uruguay (36), Bosnia & Herzegovina (48), Belgium (57), El Salvador (65), Laos (101) and Tanzania (113). The volatility of those countries is due to the combined effect of all five assumptions, although the most influential assumptions are the use of a geometric versus a arithmetic average aggregation function at the policy level and the use of equal weighting or Factor Analysis weighting at the indicators level.

What if measurement error is incorporated?

A normally distributed random error term was added to the raw data with a mean zero and a standard deviation equal to the observed standard deviation for each indicator. Among the countries that are most affected by this assumption is Luxembourg (rank: 31), whose rank would drop by 53 positions. On the other extreme, the Philippines (rank: 61) would improve its rank and be placed in the 10^{th} position. Overall, the introduction of measurement error in the raw data has a median impact of 9 ranks and a 90^{th} percentile impact of 29 ranks. In other words, this assumptions leaves 1 out of 2 countries almost unaffected (less than 9 rank change), but 1 out of 10 countries would shift more than 29 ranks.

What if skewed distributions are not winsorized?

Winsorization was not found to have a significant impact on the EPI ranking. Most notably, Luxembourg (rank: 31) would deteriorate by 53 positions. On the other extreme, the Philippines (rank: 61) would improve and be placed in the 10^{th} position. Overall, the introduction of measurement error in the raw data has a median impact of 9 ranks and a 90^{th} percentile impact of 29 ranks. In other words, this assumptions leaves 1 out of 2 countries almost unaffected (less than 9 ranks change), but 1 out of 10 countries would shift more than 29 positions.

What if capping at target values for the indicators is not undertaken?

Luxembourg (rank: 31) and Laos (rank: 101) would see the greatest shift in their ranks (a decline of 12 and 15 positions respectively). In the best case, El Salvador (rank: 65) will improve by 9 positions. Overall, for 1 out of 2 countries, the impact of this assumption is only 3 positions, while 1 out of 10 countries shift by more than 7 positions, but not more than 15. Thus, the impact of capping at the indicators' performance targets exerts only a small impact on the EPI ranking.

What is the impact of alternative weighting schemes?

Four alterative weighting schemes, all with their implications and advantages, are deemed as the most representative in the literature of composite indicators and worth being tested in our current analysis.

- current weighting vs. FA-derived weights at the indicator level;
- current weighting vs. equal weighting at the indicator level;
- current weighting vs. equal weighting at the subcategory level;
- current weighting vs. equal weighting at the policy level;

The simulation study showed that all of these scenarios have significant influence on the EPI ranking. The scenarios with the biggest impact are: equal weighting at the policy level, equal weighting at the indicator level, and Factor Analysis derived weights at the indicator level. In any of these three cases, 1 out of 2 countries shifts less than 15 positions with respect to the original EPI ranking, whilst 1 out of 10 countries shifts more than 50 positions.

What if the aggregation function is geometric instead of arithmetic?

When a partially compensatory aggregation is performed at the policy level using the geometric mean function instead of the arithmetic mean, the impact on the EPI ranking is moderate. Sri Lanka, Peru and Egypt improve their ranks by 18 positions or more, whilst the greatest decline is observed for Uruguay (down more than 51 positions). Overall, for 1 out of 2 countries, the impact of this assumption is merely 5 positions, while 1 out of 10 countries shift by more than 18 positions (up to 51 positions).

Table of Contents

List of Tables

List of Figures

1. Introduction

The analysis presented in this report aims at validating and critically assessing the methodological approach undertaken by the EPI team at Yale and Columbia University. Although this analysis was undertaken in the past versions of the Index, the new data and framework used in 2008 necessitates such type of analysis, so as to ensure that the methodology remains appropriate. At the same time, it aims at identifying those countries with and without very robust EPI ranks. For the first group, policy signals derived from the EPI can be taken with the confidence that changes in the EPI methodology would have a negligible effect on the country's measured performance, while for the latter a more cautious approach is advised vis-à-vis translating the EPI rank into policy actions.

A clear understanding of the EPI methodology is crucial to the success of the robustness assessment of the index. In a first step, we thus considered if it is possible to reproduce the EPI results given the data and information provided to the public? The answer is "*Yes*". The EPI website provides enough information to a statistically literate public in order to replicate the EPI methodology and results.

Indisputably, the construction of the EPI demands a sensitive balance between simplifying an environmental system and still providing sufficient detail to detect characteristic differences (Diener and Suh, 1997). This leaves scientists and policymakers with a complex and synthetic measure that is almost impossible to verify against true conditions, particularly since environmental performance cannot be measured directly (Eyles and Furgal, 2002; von Schirnding 2002). It is therefore taken for granted that the EPI can not be verified. Yet, in order to enable informed policymaking and be useful as a policy and analytical assessment tool, the EPI needs to be assessed in regard to its validity and potential biases.

The EPI Framework

The EPI framework and methodology are described in detail in (Esty *et al.*, 2008). We provide a hand-waiving description next.

The EPI builds on measures relevant to two core objectives:

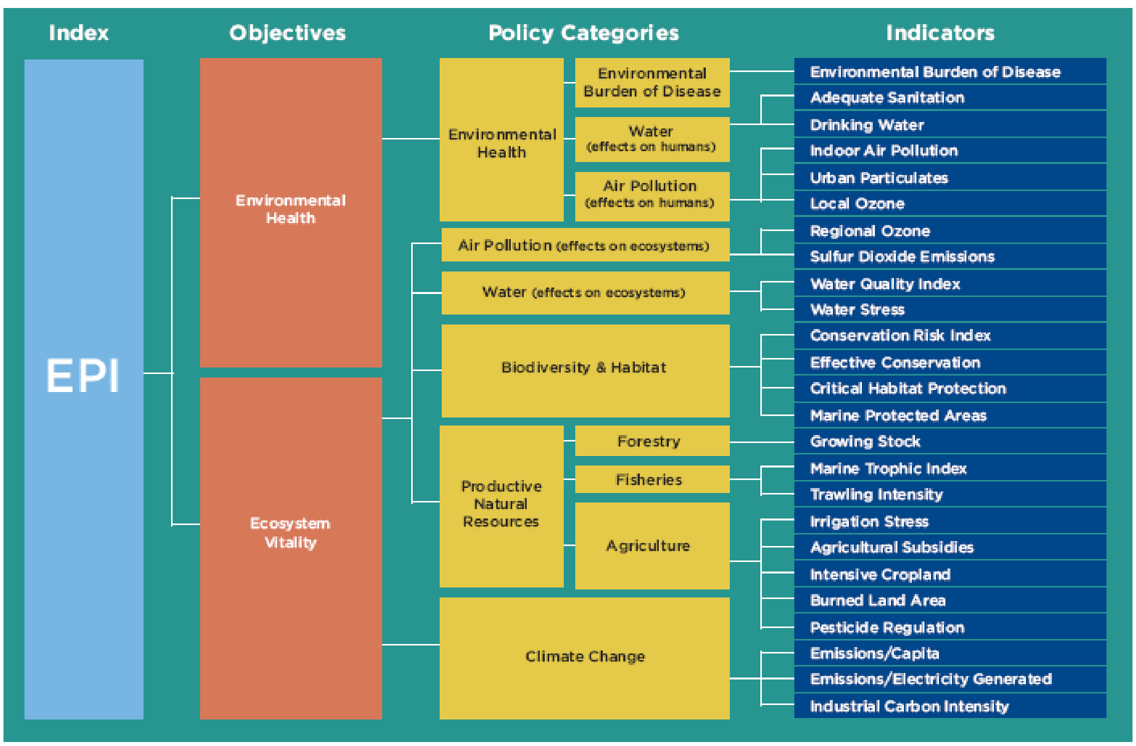
- reducing environmental stresses to human health (the Environmental Health objective); and
- protecting ecosystems and natural resources (the Ecosystem Vitality objective).

The EPI team selected the 25 indicators through: a broad-based review of the environmental science literature; in-depth consultation with a group of scientific advisors in each policy category; the evidence from the Millennium Ecosystem Assessment, the Intergovernmental Panel on Climate Change, the Global Environmental Outlook-4, and other assessments; environmental policy debates surrounding multilateral environmental agreements; and expert judgment. Each indicator builds on a foundation either in environmental health or ecological science. Some of these metrics track the underlying concept closely. Others are "proxy" variables that imperfectly reflect the theoretical focus. The 25 indicators each represent core elements of the environmental policy challenge. For each indicator, a relevant long-term public health or ecosystem sustainability goal is identified. These targets are drawn from 1) treaties or other internationally agreed upon goals; 2) standards set by international organizations; 3) leading national regulatory requirements; or the 4) prevailing scientific consensus. The indicators serve as a gauge of long-term environmental policy success. For each country and each indicator, a proximity-to-target value is calculated based on the distance from a country's current results to the policy target.

In calculating EPI scores, the EPI team averaged around isolated data gaps. But countries with more than a few missing data values (preventing any of the category scores from being calculated) were dropped from the Index. The data matrix covers 149 countries for which an EPI can be calculated across the 25 indicators. Data gaps mean that another 90 or so countries cannot be ranked in the 2008 EPI.

Using the 25 indicators, scores are calculated at three levels of aggregation (see Figure 1). First, building on two to eight underlying indicators (each representing a data set), a score is calculated for each of the six core policy categories – Environmental Health, Air Quality, Water Resources, Biodiversity and Habitat, Productive Natural Resources, and Climate Change. In some cases, subcategories are also tracked. This level of aggregation permits countries to track their relative performance within these well-established policy areas – or at the disaggregated indicator level. Second, the Environmental Health subcategories and the Ecosystem Vitality categories are weighted and then aggregated. Finally, the overall Environmental Performance Index is calculated, based on the arithmetic mean of the two broad objective scores.

Figure 1. Environmental Performance Index Framework



Targets

The EPI builds on a set of carefully chosen policy targets (see last column of Table 1). The EPI team opted to measure success against these targets, so as to provide useful information about country-specific conditions and policy results, as well as areas in need of increased attention and resources. A proximity-to-target measure helps to clarify comparative rankings, demonstrate which countries are leading or lagging in each area, and whether (as a global aggregate) the world is on a sustainable trajectory. Whenever possible the targets are based on international treaties and agreements. For issues with no international agreements, the EPI team looked next to environmental and public health standards developed by international organizations and national governments, the scientific literature, and finally, expert opinion from around the world. Only a few of the indicators have explicit consensus targets established at a global scale. This suggests that there is also a need for the international and national policy communities to be clearer about the long-term goals of environmental policies set at all levels. International agreements are often based on compromises, however, and targets derived from them do not necessarily reflect environmental performance required for full sustainability.

Table 1. Weights (in red), Sources, and Targets of EPI components

Index	Objectives	Policy Categories	Subcategories	Indicators	Indicator Code	Data Source	Target
EPI	Environmental Health 50%		Environmental burden of disease 25%		DALY	WHO	0 DALYs
			Water (effects on humans) 12.5%	Adequate sanitation 6.25%	ACSAT	WHO-UNICEF Joint Monitoring Program	100%
				Drinking water 6.25%	WATSUP	WHO-UNICEF Joint Monitoring Program	100%
			Air Pollution (effects on humans) 12.5%	Urban particulates 5%	PM10	World Bank, WHO	20 ug/m^3
				Indoor air pollution 5%	INDOOR	WHO	0%
				Health ozone 2.5%	OZONE_H	MOZART II model	0 exceedance above 85 ppb
	Ecosystem Vitality 50%	Air Pollution (effects on ecosystems) 2.5%		Ecosystem ozone 1.25%	OZONE_E	MOZART II model	0 exceedance above 3,000 AOT40. AOT40 is cumulative exceedance above 40 ppb during daylight summer hours
				Sulfur dioxide emissions 1.25%	SO2	EDGAR/Netherlands	0 tons SO_2 / populated land
		Water (effects on ecosystems) 7.5%		Water quality 3.25%	WATQI	UNEP GEMS/Water	100 score
				Water stress 3.25%	WATSTR	UNH Water Systems Analysis	0% territory under water stress
		Biodiversity & Habitat 7.5%		Conservation risk index [7.5 / (2+AZE weight + MPAEEZ weight)]%	CRI	The Nature Conservancy calculation	0.5 ratio
				Effective conservation [7.5 / (2+AZE weight + MPAEEZ weight)]%	EFFCON	The Nature Conservancy calculation	10%
				Critical habitat protection* [if no AZE sites: 0; if AZE sites: 7.5 / (2+AZE weight + MPAEEZ weight)]%	AZE	The Nature Conservancy calculation	100%
				Marine Protected Areas* [minimum of 7.5*EEZ area / land area and 7.5, divided by (2+AZE weight + MPAEEZ weight)]%	MPAEEZ	Sea Around Us Project, Fisheries Centre, UBC	10%
	Ecosystem Vitality 50%	Productive Natural Resources 7.5%	Forestry* 2.5%	Growing stock change 2.5%	FORGRO	FAO	ratio of at least 1
			Fisheries* 2.5%	Marine Trophic Index 1.25%	MTI	UBC, Sea Around Us Project	no decline
				Trawling intensity 1.25%	EEZTD	UBC, Sea Around Us Project	0%
			Agriculture* 2.5%	Irrigation Stress* 0.5%	IRRSTR	CIESIN calculation	0%
				Agricultural Subsidies 0.5%	AGSUB	World Bank, World Development Report	0
				Intensive cropland 0.5%	AGINT	CIESIN calculation	0%
				Burned Land Area 0.5%	BURNED	CIESIN calculation	0%
				Pesticide Regulation 0.5%	PEST	UNEP-Chemicals	9 banned POP chemicals and full participation in Rotterdam and Stockholm Conventions
		Climate Change 25%		Emissions per capita 8.33%	GHGCAP	IEA, CDIAC, Houghton	2.24 Mt CO_2 eq. (Estimated value associated with 50% reduction in global GHG emissions by 2050, against 1990 levels)
				Emissions per electricity generation 8.33%	CO2KWH	IEA	0 g CO_2 per kWh
				Industrial carbon intensity 8.33%	CO2IND	IEA, WDI	0.85 tons of CO2 per $1000 (USD, 2005, PPP) of industrial GDP (Estimated value associated with 50% reduction in global GHG emissions by 2050, against 1990 levels)

*Averaged around if missing data or not applicable to country

Calculating the EPI

To make the 25 indicators comparable, each metric was converted to a proximity-to-target measure with a range of 0 to 100.

Initially, the distribution of each indicator was examined by the EPI team to identify whether extreme values skew the aggregations of some indicators. Extreme outliers (greater than or equal to three standard deviations from the mean) are more likely to be the result of data processing (especially for modeled data) than actual performance. Accordingly, outliers were adjusted using a recognized statistical technique called winsorization – in this case trimming at the 95th percentile of the distribution. In a small number of cases even this level of winsorization left significant outliers, and in such cases winsorization at a greater level based on a comparison of the two alternative values (see Esty *et al.*, 2008, Appendix E).

A second decision concerned the treatment of countries that exceeded the long-term performance or sustainability target. To avoid rewarding "over-performance," no indicator values above the long-term target were used. In the few cases where a country did better than the target, the value was reset so that it was equal to the target. Once those two adjustments were made, a simple arithmetic transformation was undertaken: the observed values were placed onto a 0 to 100 scale (100 ≡ target value, 0 ≡ worst observed value).

Weighting and aggregation, in particular, are two areas of inescapable methodological controversy. While the field of composite index construction has become a well-recognized subset of statistical analysis, there is no clear consensus on how best to construct composite indices. 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.

To help identify appropriate groupings and weights for each indicator, a principal component analysis (PCA) was carried out. Most categories did not have clear referents in the PCA results. Absent a PCA-derived basis for weighting the indicators, equal weights were used with some refinements determined by the EPI team with expert guidance.

The Environmental Health and Ecosystem Vitality subcategories each represent 50% of the total EPI score. This equal division of the EPI into issues related to (1) humans and (2) nature is not a matter of science but rather policy judgment. But this even weighting of the two overarching objectives of environmental policy reflects a widely-held intuition, and this choice (used in the 2006 Pilot EPI) has not been generally criticized. Indeed, for every "deep ecologist" who favors more weight being placed on Ecosystem Vitality, there is a "humans first" environmental policymaker who prefers that the tilt go the other way.

Within the Environmental Health Objective/Policy Category, the Environmental Burden of Disease (DALY) indicator is weighted 50% and accordingly contributes 25% of the overall EPI score, because it is widely regarded to be the most comprehensive and carefully-defined measure of environmental health burdens. The effects of Water and Air Pollution on human health comprise the remainder of the Environmental Health subcategory and are each allocated a quarter of the total score for Environmental Health, reflecting a widespread policy consensus. The two water-related Environmental Health indicators (Adequate Sanitation and Drinking Water) are equally weighted. In the Air Pollution subcategory, Urban Particulates and Indoor Air Pollution receive equal weights, and double the weight given to the effects of ground-level Ozone on human health. Urban particulates and indoor air pollution are widely acknowledged by the United Nations Environment Programme (UNEP), World Health Organization (WHO), and United Nations Children's Fund (UNICEF) as important indicators of the burden of air pollution on human health. There is, however, a growing literature that suggests a link between ozone exposure and human health. The EPI human exposure to ozone metric assesses person-days of exposure per year to ground-level ozone exceeding 85 parts per billion (ppb). Because this indicator is experimental, the EPI team assigned half the weight of those with known reliability.

Within the Ecosystem Vitality Objective, the Climate Change indicator carries 50% of the weight (i.e., 25% within the total EPI). This is owing to the increasing importance attached to climate change in policy discussions, and its potential to have far reaching impacts across all aspects of ecosystem vitality and natural resource management. The Air Pollution (effects on ecosystems) policy category is weighted at 5% of the Ecosystem Vitality Objective. This slightly lower weight when compared to water, biodiversity, and productive natural resources is owing to the fact that Air Pollution is already partially captured in the Environmental Health Objective. The remaining indicators: Water, Biodiversity, and Productive Natural Resources, are each evenly weighted to cover the remaining 22.5% of the Ecosystem Vitality Objective.

2. How is the EPI associated to its subcomponents and policy categories?

Following the replication process, correlation analysis is performed to examine the relationship between the EPI scores and the indicator scores, the policy scores and finally the objectives scores. Correlation analysis is a basic but widely used tool for "confirming" the mathematical design of indices. Booysen (2002) recommends that a weak correlation between an underlying indicator and an index should result in the exclusion of the respective indicator from the process. A major drawback of correlation analysis though is the fact that a strong correlation does not necessarily imply a strong influence or representation of the indicator in the overall index. In other words, any random variable could potentially show strong correlation with the index without actually being part of the index.

A simple rank correlation analysis between the EPI scores and the category scores (Table 2) reveals that the EPI has very high correlation with the Environmental Health category ($r_S = 0.90$) and the Water category ($r_S = 0.59$), and a fairly strong relationship with the Productive Natural Resources ($r_S = 0.34$) and the Climate Change ($r_S = 0.18$) categories. However, the relation of the EPI to two of the six policy categories, namely to Air pollution and Biodiversity & Habitat, appears to be random and non-significant at the 95% level. Relationships among the policy categories themselves vary, but they are in general low and in most cases random. These results indicate that the six policy categories represent totally different aspects of environmental performance – which is desirable from an index development perspective. Although it is desired not to have very high association between the main components of a composite indicator (since representing different dimensions is a key quality feature of a composite indicator), the negative association between several of the policy categories leads to a conclusion that there may be trade-offs between them, which creates an additional difficulty in an index that combines such different dimensions with the implicit assumption that strong performance on all policy categories is possible simultaneously. In this case it may be argued that there should be no single measure of environmental performance, but rather one should focus on the six policy categories and identify linkages and trade-offs between them, instead of attempting to aggregate them into a single score.

Table 2. Spearman rank correlation coefficients for the EPI, the two objectives and the six policy categories

	Policy categories						*Objectives*	
	Environmental Health	Air pollution (effects on nature)	Water (effects on nature)	Biodiversity & Habitat	Productive Natural Resources	Climate Change	Ecosystem Vitality	Environmental Health
EPI	0.90	-0.09*	0.59	-0.04*	0.34	0.18	0.29	0.90
Environmental Health		-0.18	0.42	-0.22	0.29	-0.16	-0.08*	
Air pollution (effects on nature)			-0.06*	-0.12	0.05*	0.07*		
Water (effects on nature)				-0.04*	0.18	0.26		
Biodiversity & Habitat					-0.01*	0.18		
Productive Natural Resources						-0.08*		

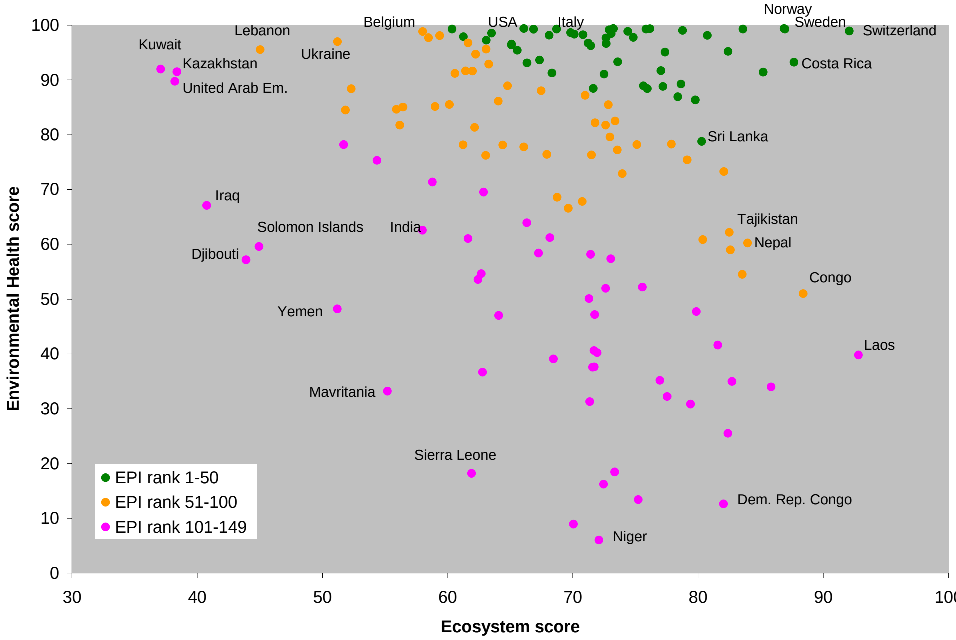
* coefficient not significant at the 95% level

Further study of the association between the EPI and the 25 underlying indicators reveals that there is a strong dominance of just a few indicators in the overall EPI. Thus, the primary drivers of the EPI ranking are four indicators: the Environmental Burden of Disease (DALY), the Adequate Sanitation (ACSAT), the Drinking Water (WATSUP) and the Indoor Air Pollution (INDOOR). Somewhat surprisingly, the three indicators related to climate change, although being weighted comparatively strongly, do not exert much influence on the EPI results. Parsimony principles would suggest excluding the non-influential indicators from the EPI framework (Gall, 2007). This, however, may not be advisable from a policy perspective, unless excluding certain indicators is supported by expert opinion on the relevance of the indicators to the issue. An eventual revision of the EPI framework may be undertaken in terms of the weighting issue.

The scatter plot between the two main Objectives of the EPI, Environmental Health and Ecosystem Vitality, in Figure 2 points to an understandable - though problematic – trade-off between these two objectives. Countries may end up choosing one or the other path in pursuing environmental performance in a somewhat mutually exclusive pattern, perhaps descriptive of different scales and time horizons. This graph, therefore, points to a major problem in translating sustainability-oriented performance into practice. At the same time, the high association between the EPI scores and the Environmental Health scores, and the random association between the EPI scores and the Ecosystem Vitality scores (see Table 2) leads to an Ecosystem's performance behaving as a noise term superimposed to Environmental Health.

The conclusions from this preliminary analysis already point to the conclusion that the 2008 EPI has an architecture that highlights the complexity of translating environmental stewardship into straightforward, clear-cut policy recipes. The trade-offs within the index dimensions are a reminder of the danger of compensability among the dimensions while identifying the areas where more work is needed to achieve a coherent framework in particular in terms of the relative importance of the indicators that compose the framework.

Figure 2. Scatterplot of the two EPI Objectives



3. Robustness of the EPI results to the methodological assumptions

There is ample evidence of the creativity in the community of composite indicators developers, which not only comes as a response to the demands of the user/stakeholder community, but it also reflects the disagreements within the research community on which indicators influence a particular phenomenon and on their relative importance (Cutter *et al.*, 2003). When building an index to capture environmental performance, it is therefore necessary to take stock of existing methodologies to avoid skewing the assessment and decision-making.

By acknowledging a variety of methodological assumptions in the development of an index that are intrinsic to policy research, one can determine whether the main results change substantially when the assumptions are varied over a reasonable range of possibilities (Saisana *et al.*, 2005; Saisana and Tarantola, 2002; Saltelli *et al.*, 2000, JRC/OECD, 2008). The advantages offered by considering different scenarios to build the EPI could be: to gauge the robustness of the EPI results, to increase its transparency, to identify the countries whose performance improves or deteriorates under certain assumptions, and to help frame the debate around the use of the EPI for policy- making. The alternative scenarios to build the EPI should, however, bear certain quality features:

1. No strong dominance of a few indicators at the expense of others in the index.
2. No deliberate bias of the index results against a few countries.
3. Simplicity and easy reproduction of the index.

In the case of the 2008 EPI, the assumptions that needed to be tested, are: (1) measurement error of the raw data, (2) choice of capping the 25 indicators at the selected targets, (3) choice to correct for skewed distributions in the indicator values, (4) weights assigned to the indicators and/or to the subcomponents of the index, and finally (5) aggregation function at the policy level. The analysis that we have undertaken maps the effects of these uncertainties and assumptions on the EPI country rankings. We also seek to use uncertainty and sensitivity analyses to assess whether useful conclusions can be drawn from the index given the construction methodology selected.

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

All of the five assumptions discussed above can heavily influence the output—and reliability—of the EPI. Using uncertainty and sensitivity analysis, we systematically evaluated the impact that the methodological and conceptual choices highlighted above have on the robustness of the EPI scoring and ranking. Our study aimed to answer four main questions.

1. What associations are there between the EPI and its indicators and/or subcomponents?
2. How do the EPI ranks compare to the ranks under combinations of alternative scenarios derived from the 5 assumptions?
3. Which countries have the most volatile ranks and why?
4. What are the major sources of variability in the EPI rankings?

The first question has already been discussed previously. Next, we will focus on the remaining three questions which call for a combined application of uncertainty and sensitivity analysis.

Our approach

We focus on testing the five central methodological issues, which are translated into 40,000 simulations of different combinations of them.

To be more specific, the <u>measurement error</u> is introduced by adding to each value in the dataset a random error with a mean equal to zero and standard deviation equal to the observed standard deviation of the corresponding indicator. Some thousands of alternative datasets that include error in some of the data values are generated. The two triggers on <u>capping at target values</u> and <u>correcting for skewed data distributions</u> are binary (yes/no). Regarding the <u>weights</u> to be attached to the indicators and/or the subcomponents, we have identified four alternatives to the current one: Factor analysis-derived weights at the indicator level; equal weighting at the indicator level; equal weighting at the subcategory level (and relative weights within each subcategory as in the EPI); equal weighting at the policy level (and relative weights within each policy category and subcategory as in the EPI). Finally, a binary trigger determines the <u>aggregation function (at the policy level)</u> to be an arithmetic or a geometric average. In the latter case, the use of a geometric aggregation would penalize countries that compensate very low performance in some policy categories with very high performance in other policy categories. Given that environmental excellence is understood to mean strong performance on the different EPI categories simultaneously, compensation at the policy level should be penalized. We undertook a saturated sampling of the space of input factors.

The combinations of these assumptions are translated into a set of N=40,000 simulations in a Monte Carlo framework. The composite index is then evaluated N times, and the EPI scores and ranks obtained are associated with the corresponding draws of input factors to appraise their influence. When several layers of uncertainty are simultaneously activated, composite indicators turn out to be non-linear, possibly non-additive models, due to interactions between the assumptions (Saisana *et al.*, 2005). As a result, all EPI scores and ranks are non-linear functions of the input factors (/assumptions) and the purpose of the uncertainty analysis is the estimation of their probability distribution functions.

As argued by practitioners (Saltelli *et al.*, 2008), robust, “model-free” techniques for sensitivity analysis should be used for non-linear models. Variance-based techniques have been shown to yield useful results for sensitivity analysis.

4. How do the EPI ranks compare to the ranks under all scenarios?

The uncertainty analysis results from the Monte Carlo simulations for the 149 countries are given in detail in Table 3. They reveal whether any deliberate bias against some countries is introduced by making certain methodological choices in building the EPI and respond to arguments made by Andrews *et al.* (2004: 1323) that many indices “rarely have adequate scientific foundations to support precise rankings: […] typical practice is to acknowledge uncertainty in the text of the report and then to present a table with unambiguous rankings”.

Countries are ordered by their original EPI score. The numbers represent the frequency of a country being among the top 10, top 10-20, and so on. Just to give an example, New Zealand is among the top 10 performing countries in 98% of the simulations. Costa Rica and Finland follow, with a frequency of 81% to be ranked among the top 10. Interestingly, Switzerland, which scores top in the original EPI, is almost as likely to be among the top 10, top10-20 or top 20-30 countries. These frequencies indicate the uncertainty about the countries scores in the EPI. In fact, approximately half of the countries in the EPI are placed correctly in the environmental performance ladder, whilst the other half of the countries can fluctuate significantly between various positions, and any conclusion on the performance of these countries should be drawn with great caution. These results depend on the theoretical framework and the set of indicators, but are independent of the methodology (methodology-free results), given that they represent a whole set of alternative scenarios. The dominant source for the observed deviations arises from the choice of the weights and its combined effect with the choice of the aggregation function at the policy level. As Table 3 demonstrates, countries with high or low performance in the EPI do not have wide variations in their ranks under alternative scenarios. The exceptions to this rule are Austria, Canada, and Iceland. In our simulations Austria ranked between the top 10 to the top 40-50. Another interesting example is Iceland (rank: 11) whose score can be anywhere within the top10-20 to top 80-90. Canada, on the other hand (rank: 12) has a 58% frequency to be ranked in the top10 and 33% to be ranked among the top10-20. This result suggests that in fact Canada outperforms Iceland on the environmental issues measured in the EPI given the current framework.

Table 3. Probabilities of country ranks in the Environmental Performance Index under all tested combinations of input factors (probabilities less than 5% are not shown)

2008 EPI: Uncertainty analysis for 149 countries

	Rank 1-10	Rank 11-20	Rank 21-30	Rank 31-40	Rank 41-50	Rank 51-60	Rank 61-70	Rank 71-80	Rank 81-90	Rank 91-100	Rank 101-110	Rank 111-120	Rank 121-130	Rank 131-140	Rank 141-149
Switzerland	31	30	20	11	6										
Sweden	63	25	10												
Norway	55	31	6												
Finland	81	16													
Costa Rica	81	16													
Austria	15	19	16	21	18										
New Zealand	98														
Latvia	25	39	26	6											
Colombia	74	18	5												
France	15	26	30	14	13										
Iceland		11	15	5	14	15	9	10	13						
Canada	58	33	9												
Germany		14	40	21	20										
United Kingdom	11	44	29	11											
Slovenia		9	18	25	23	8	10								
Lithuania		16	20	14	9	8	9	6	9						
Slovakia		15	21		14	25	5	6	8						
Portugal	20	46	16	11	5										
Estonia	56	34													
Croatia		16	19	23	10	6	9	5	5	5					
Japan	6	38	35	14	5										
Ecuador	63	26		5											
Hungary				6		6	13	16	20		6	10	9		
Italy		6	28	24	16	13	5								
Denmark			8		9	6	15	13	14	8	6	11			
Malaysia	31	48	15	5											
Albania			9	11	6	13	10	16		5	6		9	5	
Russia	9	33	43	9											
Chile	16	46	25	8											
Spain		5	30	18	19	14	11								
Luxembourg		9		15	16	20	26	5	5						
Panama	73	20													
Dominican Rep.	18	54	21	6											
Ireland			5	16	13	15	13	13			9	5			
Brazil	5	20	29	24	11										
Uruguay				11	15	9	8	9	10		9			14	
Georgia				8	8	19	15	16	10	13	5				
Argentina			10	23	28	24	11								
United States	5	23	19	24	13	8									
Taiwan				20	13	19	16	10	13						
Cuba		5	24	29	19	13	5								
Poland		5	11	20	35	15									
Belarus					11	10	10	18	16	16	13				
Greece				8	18	14	19	15		5	10	6			
Venezuela	5	11	36	25	18	5									
Australia	30	30	14	10	9										
Mexico		11	15	34	28		6								
Bosnia & Herzegovina					5		10	11	24	9	6	8		14	6
Israel				5	31	19	19	13	5	6					
Sri Lanka	19	36	16	16	10										
South Korea					6		14	14	19	9	8	13	8		
Cyprus				10	9	25	14	28	6						
Thailand		8	30	35	11	11									
Jamaica				8	15	24	11	11	9	10	5				
Netherlands						9	11	14	10	21	9	11	9		
Bulgaria				5	19	25	15	8	10	6					
Belgium						13	6	11	6	6	16	10	13	9	
Mauritius					6	9	19	18	8	16	15				
Tunisia						5	10	10	10	14	19	18	9		
Peru			15	30	18	30									
Philippines		6	13	26	21	16	9		5						
Armenia							6	13	19	8	16	18	8	6	
Paraguay					11	18	20	18	9	8	5	6			
Gabon		6	35	28	16	5	6								
El Salvador			5	6		13	16	9	10	9	8	9	8	5	
Algeria			5	5	15	26	24	11	6	5					
Iran			11	23	26	18	16								
Czech Rep.						9	8	15	11	13	19	15	10		
Guatemala		10	16	23	26	14	8								
Jordan						8	14	24	20	6	14	8			
Egypt			19	21	24	13	10	6							
Turkey							18	15	18	16	9	6	6		
Honduras		9	28	20	15	13	8	5							
Fyrom						5	5	15	10	18	21	13	6	5	
Ukraine							8	15	6	23	11	10	13	10	
Viet Nam				5	10	18	29	20	8						
Nicaragua			8	21	28	14	10	14							
Saudi Arabia						11	13	23	16	10	11				
Tajikistan							9	6	14	19	18	19	11		
Azerbaijan									10	23	9	24	26	5	
Nepal					5	9	8	15	28	14	13	6			
Morocco								14	20	15	13	15	16		
Romania						14	10	33	16	11		5			
Belize			9	29	16	13		14	6						
Turkmenistan					5	6	14	8	19	25	11	9			
Ghana		11	14	16	10	10	13	9	8	5					
Moldova										8		10	13	30	31
Namibia							16	16	25	18	16				
Trinidad & Tobago				6	20	11	23	8	13	8		5			
Lebanon						5	15	13	13	5	5	6	20	8	11
Oman						10	25	18	24	5	10				
Fiji									6	15	13	19	14	20	5
Congo				9		23	18	13	13	13	9				
Kyrgyzstan						5	6	15	13	30	15	9			
Zimbabwe					21	11	15	10	15	13	6				
Kenya		18	15	11	11	11	11	6	8	6					
South Africa					14	16	20	11	16	11					
Botswana						5			13	18	18	19	16	6	
Syria										16	11	21	13	30	6
Mongolia				5		13	9	25	18	15	10				
Laos				10	8	9	10	6	10	9	19	11			
Indonesia				9	10	11	23	19	14	5	6				
Côte d'Ivoire			10	13	15	20	8	11		9	6				
Myanmar							6	5	16	24	26	15	5		
China							9	13	9	25	19	13	5		
Uzbekistan											14	16	29	29	11
Kazakhstan										5	15	16	36	24	
Guyana						6	10	19	15	20	20	5			
Papua New Guinea						6	10	14	9	11	20	20	8		
Bolivia							8	10	23	13	20	11			
Kuwait											9	5	15	28	41
United Arab Em.										10	9	36	19	11	11
Tanzania				11	13	16	9	11	11	8	6	6			
Cameroon						6	10	6	13	23	23	15			
Senegal						5	6	9	16	30	24	6			
Togo							8	6	18	18	18	18	10		
Uganda						8	5	5	6	11	20	21	11	10	
Swaziland									6		16	31	24	15	
Haiti											10	21	23	30	10
India										11	15	31	25	13	
Malawi					9	13	13	14		11	15	9	6	8	
Eritrea								6	13	16	16	25	18		
Ethiopia								6	8	9	8	9	25	26	5
Pakistan									23	9	26	18	18		
Bangladesh												9	18	24	48
Nigeria								6	5	13	15	24	23	6	5
Benin						10	11	10	14	13	9	11	13		
Central Afr. Rep.											13	14	16	38	13
Sudan												10	34	46	6
Zambia									10	10	14	9	21	21	11
Rwanda					6		11	18	11	18	5	13	6	9	
Burundi								9	8	15	9	18	29	11	
Madagascar									8	13	16	20	21	15	
Mozambique							6	6	9	11	14	18	21	9	
Iraq													11	26	60
Cambodia										8	15	11	31	28	
Solomon Islands														16	81
Guinea											13	14	23	36	6
Djibouti												8	18	35	39
Guinea-Bissau										15	14	28	19	15	5
Yemen													6	29	63
Dem. Rep. Congo												13	29	26	23
Chad												8	16	33	40
Burkina Faso											9	6	18	43	25
Mali												5	18	36	41
Mauritania												9	25	40	24
Sierra Leone													11	18	70
Angola														19	79
Niger													6	19	73

Legend:
Probability between 5 and 15%
Probability between 15 and 30%
Probability between 30 and 50%
Probability greater than 50%
Probability lower than 5% is not shown

Source: JRC calculations
Notes:
1. Numbers express probabilities for the country rank
2. Countries listed based on the 2008 EPI scores from highest to lowest (left to right)

5. Which countries have the most volatile ranks and why?

We will use the term "volatility" as a measure of the difference between a country's best and worst rank, calculated from the 5th and the 95th percentiles of the rank distribution simulations. For Finland, Costa Rica, New Zealand, Colombia and Panama, we can reasonably state that they have a top 10 performance (frequency greater than 70%) and very low volatility in their scores. Interestingly, Panama is ranked 32nd in the EPI – a rank that occurs less than 5% of the times in our simulations. Table 4 presents the 20 countries that are affected most strongly by the methodological choices made during the construction of the EPI. These countries, with a difference in their best and worst rank (5th and 95th percentiles) of at least 80 positions, are ranked between 11th (Iceland) and 131st (Rwanda). A number of those countries such as Lithuania, Hungary, Denmark, Albania, Ireland, Uruguay, and Bosnia & Herzegovina are ranked among the top 50 in the EPI. The volatility of those countries' ranks can be attributed mainly to the choice of the weighting combined with the aggregation scheme at the policy level.

Table 4. Most volatile countries in the EPI

Country	EPI Rank	Range of Simulation Ranks	Country	EPI Rank	Range of Simulation Ranks
Iceland	11	[14,95]	El Salvador	65	[31,129]
Lithuania	16	[16,98]	Ghana	86	[12,93]
Hungary	23	[33,129]	Lebanon	89	[62,143]
Denmark	25	[25,131]	Kenya	96	[13,98]
Albania	27	[25,132]	Laos	101	[29,116]
Ireland	34	[24,114]	Côte d'Ivoire	103	[21,103]
Uruguay	36	[31,139]	Tanzania	113	[23,113]
Bosnia & Herzegovina	48	[48,141]	Uganda	117	[55,134]
South Korea	51	[42,125]	Malawi	121	[48,132]
Belgium	57	[42,137]	Benin	127	[51,130]
			Rwanda	131	[45,131]

6. What are the sources of major impact on the variability of the EPI ranking?

We now focus on assessing the impact of each of the five assumptions individually, which amounts to a total of eight different scenarios. We undertake the following comparisons:

Measurement error

- current case without measurement error in the data vs. measurement error in the data;

Winsorisation

- current winsorisation approach vs. no winorisation;

Target values

- current target values v. no target values;

Weighting

- current weighting vs. FA-derived weights at the indicator level;
- current weighting vs. equal weighting at the indicator level;
- current weighting vs. equal weighting at the subcategory level;
- current weighting vs. equal weighting at the policy level;

Aggregation

- current arithmetic aggregation vs. geometric aggregation at the policy level.

Measurement error

It is reasonable to assume that the raw data are not flawless and that despite efforts to guarantee the most reliable sources for them, errors may still be present. To account for this, we have added a normally distributed random error term to the raw data with a mean zero and a standard deviation equal to the observed one for each indicator. Table 5 presents the countries that are mostly affected by this assumption. Most notably, Luxembourg (rank: 31) would deteriorate its rank by 53 positions. On the other extreme, the Philippines (rank: 61) would improve its rank and be placed in the 10^{th} position. Overall, the introduction of measurement error in the raw data has a median impact of 9 ranks and a 90^{th} percentile impact of 29 positions. In other words, this assumptions leaves 1 out of 2 countries almost unaffected (less than 9 positions change), but 1 out of 10 countries would shift more than 29 positions.

Table 5. Countries most affected by measurement error compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Colombia	9	42	-33	Costa Rica
Iceland	11	47	-36	Dominican Rep.
Estonia	19	60	-41	Norway
Luxembourg	31	84	-53	Finland
Dominican Rep.	33	2	31	Canada
Cuba	41	74	-33	
Poland	42	83	-41	**Bottom five countries**
South Korea	51	18	33	Cambodia
Peru	60	27	33	Mauritania
Philippines	61	10	51	Angola

Iran	67	32	35	Burkina Faso
Honduras	73	38	35	Sierra Leone
Nepal	81	115	-34	
Fiji	94	54	40	Median change: 9 ranks
South Africa	97	57	40	90th percentile change: 29 ranks

Winsorization

Winsorization is also expected to have an impact on the rankings, particularly for those countries that present a few extreme values. Table 6 presents the countries that are mostly affected by the choice of not winsorizing, as opposed to the current one. In the best case, South Africa (rank: 97) improves its position by 16, whilst in the worst case, Botswana (rank: 98) declines by 21 ranks. For 1 out of 2 countries, the impact of this assumption is only 5 positions, while 1 out of 10 countries shift by more than 11 positions, but not more than 21.

Table 6. Countries most affected by not winsorizing skewed distributions compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Hungary	23	39	-16	Sweden
Luxembourg	31	48	-17	Norway
Georgia	37	50	-13	Switzerland
Belarus	43	56	-13	New Zealand
Bosnia & Herzegovina	48	61	-13	Costa Rica
Tajikistan	79	95	-16	
Azerbaijan	80	96	-16	**Bottom five countries**
Lebanon	89	75	14	Mali
Fiji	94	107	-13	Chad
South Africa	97	81	16	Sierra Leone
Botswana	98	119	-21	Niger
Indonesia	102	87	15	Angola
Côte d'Ivoire	103	91	12	
Uzbekistan	106	125	-19	Median change: 5 ranks
Tanzania	113	99	14	90th percentile change: 11 ranks

Targets

Allowing for “extra credit” when exceeding the indicator targets is also expected to have an impact on the results. Table 7 presents the countries that are mostly affected by this assumption. Luxembourg (rank: 31) and Laos (rank: 101) would see the greatest shift in their ranks (a decline of 12 and 15 positions respectively). In the best case, El Salvador (rank: 65) will improve by 9 positions. Overall, for 1 out of 2 countries, the impact of this assumption is only 3 positions, while 1 out of 10 countries shift by more than 7 positions, but not more than 15. The two assumptions on the use of target values and on the winsorization are thus by far the least influential methodological decision in the EPI, a result that we will confirm below.

Table 7. Countries most affected by not capping the indicators at the performance target compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Slovakia	17	28	-11	Norway
Hungary	23	33	-10	Sweden
Luxembourg	31	43	-12	Switzerland
Bosnia & Herzegovina	48	57	-9	Costa Rica
Sri Lanka	50	40	10	New Zealand
Jamaica	53	61	-8	
Philippines	61	53	8	**Bottom five countries**
El Salvador	65	56	9	Mali
Saudi Arabia	78	86	-8	Burkina Faso
Azerbaijan	80	89	-9	Sierra Leone
Trinidad & Tobago	91	83	8	Angola
Lebanon	89	81	8	Niger
Laos	101	116	-15	
Cameroon	114	105	9	Median change: 3 ranks
Central Afr. Rep.	128	136	-8	90^{th} percentile change: 7 ranks

Alternative weighting schemes

Four alterative weighting schemes, all with their implications and advantages, are deemed as the most representative in the literature of composite indicators and worth being tested in our current analysis.

- current weighting vs. FA-derived weights at the indicator level;
- current weighting vs. equal weighting at the indicator level;
- current weighting vs. equal weighting at the subcategory level;
- current weighting vs. equal weighting at the policy level;

Using FA-derived weights at the indicator level significantly affects the country rankings. Half of the countries shift fewer than 16 positions but 15 countries shift more than 47 positions. Table 8 shows the countries that experience the biggest shift in their rank due to this assumption.

Table 8. Countries most affected by the FA weights compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Lithuania	16	63	-47	Switzerland
Hungary	23	75	-52	Finland
Denmark	25	79	-54	New Zealand
Albania	27	93	-66	Estonia
Georgia	37	87	-50	Austria
Bosnia & Herzegovina	48	99	-51	
South Korea	51	105	-54	**Bottom five countries**
Egypt	71	23	48	Angola
Saudi Arabia	78	17	61	Yemen
Belize	84	21	63	Bangladesh
Moldova	87	134	-47	Solomon Islands
Trinidad & Tobago	91	40	51	Sierra Leone
Zimbabwe	95	48	47	
Kenya	96	45	51	Median change: 16 ranks
Mongolia	100	33	67	90^{th} percentile change: 47 ranks

Equal weighting at the indicator level would increase the weight of the indicators in the Air Pollution (effects on nature) subcategory, the Water (effects on nature), the Biodiversity and Habitat category, and the Productive Natural Resources category. A total of seventeen indicators will increase their weight, as opposed to the current weighting scheme. The remaining eight indicators will reduce their weight, in particular, the DALY indicator and the three indicators related to Climate Change. The countries whose EPI ranks are most affected by this change are shown in Table 9. The countries that improve their ranks the most are Laos, Kenya, Mongolia and Malawi (by more than 60 positions upwards). On the other hand, Denmark and South Korea decline more than 70 positions. Overall, for 1 out of 2 countries, the impact of this assumption is 15 positions, while 1 out of 10 countries shift by more than 48 positions (up to 72 positions).

Table 9. Countries most affected by using equal weights at the indicator level compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Hungary	23	80	-57	Switzerland
Denmark	25	97	-72	Finland
South Korea	51	122	-71	New Zealand
Belgium	57	115	-58	Estonia
Tunisia	59	117	-58	Colombia
Ukraine	75	124	-49	
Belize	84	35	49	**Bottom five countries**
Moldova	87	139	-52	Yemen
Congo	92	39	53	Angola
Kenya	96	29	67	Iraq
Mongolia	100	33	67	Bangladesh
Laos	101	17	84	Solomon Islands
Côte d'Ivoire	103	49	54	
Malawi	121	55	66	Median change: 15 ranks
Rwanda	131	77	54	90^{th} percentile change: 48 ranks

We next tested the impact of an equal weighting at the subcategory level, whilst the relative weights for the indicators within each subcategory remain as in the EPI. This is expected to have a less pronounced impact on the EPI ranks because this assumption assigns greater weight to the six of the ten subcategories and reduces the weight of the other four and in particular the weight of the climate change and of the environmental burden of disease (DALY). As a consequence, the countries whose EPI ranks are most affected by this change are given in Table 10. The countries that improve their ranks the most are Trinidad & Tobago and Laos (improvement of more than 38 positions). On the other hand, Denmark and Taiwan decline more than 50 positions. Overall, for 1 out of 2 countries, the impact of this assumption is 9 positions, while 1 out of 10 countries shift by more than 26 positions (up to 51 positions).

Table 10. Countries most affected by equal weighting at the subcategory level compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Denmark	25	76	-51	Switzerland
Argentina	38	65	-27	Finland
Taiwan	40	90	-50	New Zealand
Australia	46	18	28	Sweden
South Korea	51	100	-49	Colombia
Netherlands	54	86	-32	
Belgium	57	101	-44	**Bottom five countries**
Mauritius	58	29	29	Dem. Rep. Congo
Tunisia	59	92	-33	Niger
Gabon	64	37	27	Bangladesh
Belize	84	49	35	Angola
Trinidad & Tobago	91	50	41	Mauritania
Fiji	94	66	28	
Mongolia	100	72	28	Median change: 9 ranks
Laos	101	63	38	90^{th} percentile change: 26 ranks

We conclude the assessment of the impact of different weighting methods by evaluating the impact of equal weighting at the policy level. The relative weights within the policy categories and within the subcategories remain the same as in the EPI. A weight of 1/6 (=16.7%) is thus assigned to each policy category, thus reducing significantly the previously assigned weight of 50% to the environmental health and the weight of 25% assigned original to climate change. All policy categories now have a weight to 16.7%. The countries whose EPI ranks are most affected by this change are given in Table 11. The countries with the most notable improvement in their ranks are Laos and Kenya (improvement of more than 78 positions). On the other hand, Belgium and South Korea decline more than 75 positions. Overall, for 1 out of 2 countries, the impact of this assumption is 18 positions, while 1 out of 10 countries shift by more than 486 positions (up to 91 positions).

Table 11. Countries most affected by equal weighting at the policy category level compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Denmark	25	77	-52	Switzerland
United States	39	87	-48	Finland
Taiwan	40	101	-61	Sweden
South Korea	51	126	-75	Norway
Netherlands	54	122	-68	New Zealand
Belgium	57	138	-81	
Tunisia	59	111	-52	**Bottom five countries**
Armenia	62	110	-48	Solomon Islands
Ukraine	75	123	-48	Djibouti
Belize	84	30	54	Yemen
Lebanon	89	137	-48	Iraq
Congo	92	23	69	Kuwait
Kenya	96	18	78	
Mongolia	100	35	65	Median change: 18 ranks
Laos	101	10	91	90^{th} percentile change: 48 ranks

Aggregation scheme at the policy level

We assume that full compensability is allowed among the indicators within each policy category but not desirable across the policy categories, consistently with the current theories that environmental aspects should be partially compensatory. Table 12 presents those countries for which the most notable shift in the country rank occurs when a partially compensatory aggregation is performed at the policy level, i.e., a geometric mean function instead of an arithmetic mean function. Sri Lanka, Peru and Egypt improve their ranks by 18 positions or more, whilst the most decline is observed for Uruguay (down more than 51 positions). Overall, for 1 out of 2 countries, the impact of this assumption is merely 5 positions, while 1 out of 10 countries shift by more than 18 positions (up to 51 positions).

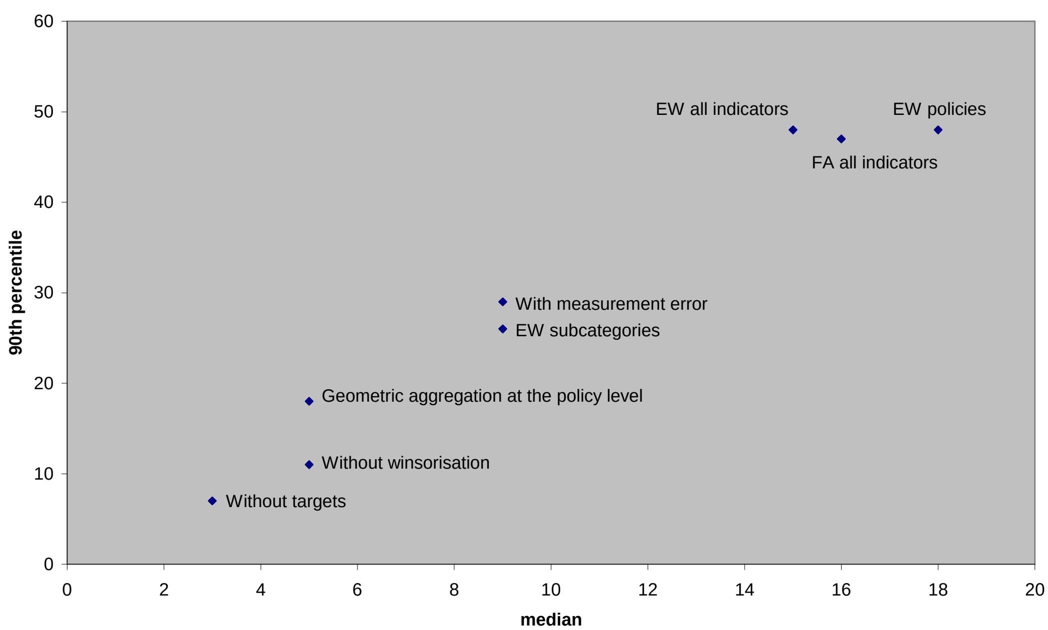
Table 12. Countries most affected by geometric aggregation at the policy level compared to the original EPI.

	EPI rank	Rank	Difference	Top five countries
Hungary	23	45	-22	Switzerland
Albania	27	62	-35	Norway
Ireland	34	58	-24	Sweden
Uruguay	36	87	-51	Finland
Greece	44	66	-22	Costa Rica
Bosnia & Herzegovina	48	94	-46	
Sri Lanka	50	31	19	**Bottom five countries**
Peru	60	42	18	Dem. Rep. Congo
El Salvador	65	83	-18	Mali
Egypt	71	51	20	Sierra Leone
Turkey	72	91	-19	Angola
Ukraine	75	96	-21	Niger
Moldova	87	113	-26	
Lebanon	89	119	-30	Median change: 5 ranks
Kazakhstan	107	126	-19	90^{th} percentile change: 18 ranks

As expected and confirmed in all cases discussed above, middle-of-the-road performers display higher variability than the top and bottom countries. Yet, these results have shown that it is not possible to know a priori which of the middle-of-the-road performers are heavily affected by the methodological choices and which countries are less sensitive.

Summing up, when only one assumption is changed at a time, the most significant impact to the EPI ranking is attributable to the weighting method, in particular when choosing equal weights at the policy level (and original weights within each policy) compared to the original EPI, equally weighting all indicators, or using factor analysis derived weights at the indicators level (Figure 3). In any of these three cases, 1 out of 2 countries ("median" to be read in the horizontal axis of Figure 3) shifts less than 15 positions with respect to the original EPI ranking, whilst 1 out of 10 countries ("90^{th} percentile" to be read on the vertical axis of Figure 3) shifts more than 50 positions. The addition of measurement error and the impact of an equal weighting at the subcategories also have significant impact on the EPI ranking (1 out of 2 countries shifts less than 9 positions, but 1 out of 10 countries shift close to 30 positions or more). The least influential input factor is the decision on whether to cap performance at the indicator targets and winsorisation. In fact, 1 out of 2 countries shift less than five positions in the overall ranking and 1 out of 10 countries shift more than 10 positions, but not more than 21 positions.

Figure 3. Sensitivity analysis: impact of one-at-a-time changes in the five tested assumptions on the EPI ranking.



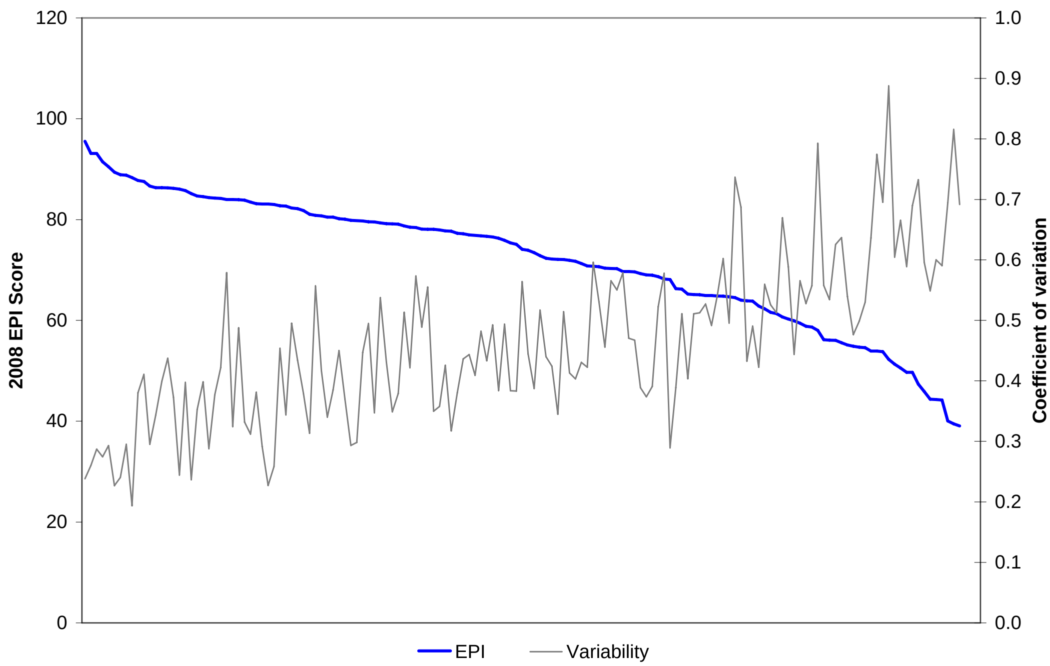
Note: median versus 90^{th} percentile of the absolute differences in the rank score between a given scenario and the EPI. EW stands for equal weighting.

When all sources of uncertainty are allowed to vary simultaneously their combined effect becomes even more pronounced. The use of geometric aggregation combined with equal weighting at the policy level, with/without targets, without winsorization, and without measurement error affects half of the countries by more than 39 positions, of which 1 out of 10 is affected by a median shift of 69 positions.

7. EPI and Variability

Countries that are situated in the top or mid-way in the EPI ranking tend to score uniformly high on the various indicators. In other words, these countries display a relatively low variability, which equals the coefficient of variation across the 25 indicators values for a given country. Figure 4 shows that the variability increases further down the EPI ranking. This scissors pattern is evident, and pronounced. The correlation coefficient between the EPI and the coefficient of variation series is equal to $r = -0.78$, indicating *a fairly high degree of reverse association between the EPI scores and the variability in the underlying indicators*. For comparison purposes, in the case of the Trade and Development Index (UNCTAD, 2005) that is based on eleven components and developed for 110 countries, the correlation coefficient between the index scores and the coefficients of variation series was much higher and equal to $r = -0.93$.

Figure 4. The scissor diagram of EPI and variability



An implication of this finding is that while changes in the EPI scores over time could be regarded as a quantitative indication of trends in environmental performance, those with respect to the variability of the ranks could be seen as qualitative changes. Reducing even further the variability in the indicators should be among the objectives of environmental policies and strategies. To be successful, a country must put simultaneously invest in multiple goals within a coherent environmental performance strategy, while emphasizing reduction of the existing gaps in areas where performance is lagging. By demonstrating significant inter-country differences in the values of the coefficient of variation, the scissors diagram points to the importance of country-specific approaches to environmental strategies. At the same time, though, it is unlikely that these variations will be reduced without coherent environmental policies and decision-making.

8. Concluding remarks

The methodological approach used to construct the 2008 EPI was studied in this report. The "statistical" filters of index quality show that, although the theoretical framework and the indicators were carefully chosen by experts, the issue of weighting is crucial to obtain a robust performance index. The current weighting scheme results in an EPI that is dominated by very few indicators while having an almost random association with several other underlying indicators. With respect to the five input factors tested in the sensitivity and uncertainty analysis, the country rankings are relatively reliable for approximately half of the countries, while any conclusion on the ranking for the other half of the countries should be made with great caution. An equal weighting approach at the indicator level, or at the policy level, as opposed to the current weighting scheme greatly influences the ranks. Thus, the choice of the weights must be evaluated according to its analytical rationale, policy relevance, and implied value judgments. The real value of the EPI lies not in the overall ranking of the countries, but rather in the solid framework and construction of the indicators. It is from this perspective that further revision of the index should be considered if the goal is to arrive at a single number that provides meaningful input to policy-making.

9. References

Andrews C. J., Hassenzahl D.M., Johnson B.B, 2004, Accommodating uncertainty in comparative risk, Risk Analysis 24(5):1323-1335.

Booysen F., 2002, An overview and evaluation of composite indices of development, Social Indicators Research 59 (2):115-151.

Cutter S. L., Boruff B.J., Shirley, W.L, 2003, Social vulnerability to environmental hazards. Social Science Quarterly 84(1):242-261.

Diener Ed, Suh E., 1997, Measuring quality of life: Economic, social, and subjective indicators. Social Indicators Research 40(1-2):189-216.

Esty D. C., Levy M.A., Kim C.H., de Sherbinin A., Srebotnjak T., Mara V., 2008, 2008 Environmental Performance Index. New Haven: Yale Center for Environmental Law and Policy.

Eyles J., Furgal C., 2002, Indicators in environmental health: identifying and selecting common sets. Canadian Journal of Public Health 93(S1):S62-67.

Gall M, 2007, Indices of social vulnerability to natural hazards: A comparative evaluation, PhD dissertation, Department of Geography, University of South Carolina.

JRC/OECD, 2008, Handbook on Constructing Composite Indicators. Methodology and user Guide, OECD Publishing, ISBN 978-92-64-04345-9.

Saisana M., Tarantola S., 2002, State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development, EUR Report 20408 EN, European Commission, JRC, IPSC, Ispra, Italy, pp. 72.

Saisana M., Tarantola S., Saltelli A., 2005, Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators, Journal of the Royal Statistical Society A, 168(2):307-323.

Saltelli A., Chan K., Scott M., 2000, Sensitivity Analysis. John Wiley & Sons Ltd.

Saltelli A., Ratto M., Andres T., Campolongo F., Cariboni J., Gatelli D., Saisana M., Tarantola S., 2008, Global sensitivity analysis. The Primer, John Wiley & Sons, England.

UNCTD, 2005, Trade and Development Index, Developing countries in international trade, United Nations Conference on Trade and Development.

von Schirnding Y., 2002, Health in sustainable development planning: the role of indicators, WHO/HDE/HID/02.11. Geneva: World Health Organization (WHO).

European Commission

EUR 23485 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen
Title: Sensitivity Analysis of the 2008 Environmental Performance Index
Author(s): Michaela Saisana and Andrea Saltelli
Luxembourg: Office for Official Publications of the European Communities
2008 – 24 pp. – 21 x 29.70 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-09702-7
DOI 10.2788/91982

Abstract

An assessment of the robustness of the 2008EPI results requires the evaluation of uncertainties underlying the index and the sensitivity of the country scores and rankings to the methodological choices made during the development of the Index. To test this robustness, the EPI team has continued its partnership with the Joint Research Centre (JRC) of the European Commission in Ispra, Italy.

This JRC report shows that the 2008 EPI has an architecture that highlights the complexity of translating environmental stewardship into straightforward, clear-cut policy recipes. The trade-offs within the index dimensions are a reminder of the danger of compensability between dimensions while identifying the areas where more work is needed to achieve a coherent framework in particular in terms of the relative importance of the indicators that compose the EPI framework.

The 2008 Environmental Performance Index (EPI) is developed for 149 countries and is based on 25 indicators in six policy categories: Environmental Health, Air Pollution, Water, Biodiversity and Habitat, Productive Natural Resources, Climate Change. The EPI aims to bring a data-driven, fact-based and empirical approach to environmental protection and global sustainability.

The validity of the EPI scoring and respective ranking is assessed by evaluating how sensitive the country ranks are to the assumptions made on the index structure and the aggregation of the 25 underlying indicators. The assumptions tested are:

- measurement error of the raw data,
- choice of capping at selected targets for the 25 indicators,
- choice to correct for skewed distributions in the indicators values,
- weights of the indicators and/or of the subcomponents of the index, and finally
- aggregation function at the policy level (six policy categories).

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

LB-NA-23485-EN-C

JRC
EUROPEAN COMMISSION

Publications Office
Publications.eu.int

ISBN 978-92-79-09702-7