JRC Scientific and Technical Reports

Uncertainty and Sensitivity Analysis of the 2010 Environmental Performance Index

Michaela Saisana and Andrea Saltelli

EUR 24269 EN 2010

JRC
EUROPEAN COMMISSION

Institute for the Protection and Security of the Citizen

The mission of the Institute for the Protection and Security of the Citizen (IPSC) is to provide research results and to support EU policy-makers in their effort towards global security and towards protection of European citizens from accidents, deliberate attacks, fraud and illegal actions against EU policies.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Address: Andrea Saltelli, JRC, TP361, via E. Fermi 2749, 21027 (VA), Italy
E-mail: andrea.saltelli@jrc.ec.europa.eu
Tel.: +39-0332-789686
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 56990

EUR 24269 EN
ISBN 978-92-79-15071-5
ISSN 1018-5593
DOI 10.2788/67623

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Uncertainty and Sensitivity Analysis of the 2010 Environmental Performance Index

Michaela Saisana and Andrea Saltelli

Executive Summary

An assessment of the robustness of the 2010 Environmental Performance Index (EPI) ranks requires the evaluation of uncertainties underlying the index and the sensitivity of the country rankings to the methodological choices made during the development of the Index. To test this robustness, the Yale and Columbia University have continued their partnership with the Joint Research Centre (JRC) of the European Commission in Ispra, Italy.

This JRC report shows that the 2010 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 2010 EPI is developed for 163 countries and is based on twenty five indicators grouped in ten policy categories: Environmental burden of disease, Air pollution (effects on humans), Water (effects on humans), Air Pollution (effects on ecosystem), Water (effects on ecosystem), Biodiversity & Habitat, Forestry, Fisheries, Agriculture and Climate Change.

The EPI 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,
- EPI structure – grouping at policy categories,
- weights assigned to the indicators and/or to the policy categories,
- aggregation function at the policy or at the objectives level, and
- number of indicators or policy categories.

The main conclusions are summarized below.

2010 EPI ranks with uncertainty considerations

Iceland	1	Syrian Arab Republic	56	Tajikistan	111
Switzerland	2	Estonia	57	Mozambique	112
Costa Rica	3	Sri Lanka	58	Kuwait	113
Sweden	4	Georgia	59	Solomon Islands	114
Norway	5	Paraguay	60	South Africa	115
Mauritius	6	USA	61	Gambia	116
France	7	Brazil	62	Libyan Arab Jam.	117
Austria	8	Poland	63	Honduras	118
Cuba	9	Venezuela	64	Uganda	119
Colombia	10	Bulgaria	65	Madagascar	120
Malta	11	Israel	66	China	121
Finland	12	Thailand	67	Qatar	122
Slovakia	13	Egypt	68	India	123
UK & N. Ireland	14	Russian Federation	69	Yemen	124
New Zealand	15	Argentina	70	Pakistan	125
Chile	16	Greece	71	Tanzania (Un.R.)	126
Germany	17	Brunei Darussalam	72	Zimbabwe	127
Italy	18	f.Y.R.O.M	73	Burkina Faso	128
Portugal	19	Tunisia	74	Sudan	129
Japan	20	Djibouti	75	Zambia	130
Latvia	21	Armenia	76	Oman	131
Czech Republic	22	Turkey	77	Guinea-Bissau	132
Albania	23	Iran (Islamic Rep.)	78	Cameroon	133
Panama	24	Kyrgyzstan	79	Indonesia	134
Spain	25	Lao P. Dem. Rep.	80	Rwanda	135
Belize	26	Namibia	81	Guinea	136
Antigua-Barbuda	27	Guyana	82	Bolivia	137
Singapore	28	Uruguay	83	Papua New Guinea	138
Serbia-Montenegro	29	Azerbaijan	84	Bangladesh	139
Ecuador	30	Viet Nam	85	Burundi	140
Peru	31	Moldova Rep.	86	Ethiopia	141
Denmark	32	Ukraine	87	Mongolia	142
Hungary	33	Belgium	88	Senegal	143
El Salvador	34	Jamaica	89	Uzbekistan	144
Croatia	35	Lebanon	90	Bahrain	145
Dominican Rep.	36	Sao Tome - Principe	91	Equatorial Guinea	146
Lithuania	37	Kazakhstan	92	Korea D. P. Rep.	147
Nepal	38	Nicaragua	93	Cambodia	148
Suriname	39	Rep. Korea	94	Botswana	149
Bhutan	40	Gabon	95	Iraq	150
Luxembourg	41	Cyprus	96	Chad	151
Algeria	42	Jordan	97	U. Arab Emirates	152
Mexico	43	Bosnia-Herzegovina	98	Nigeria	153
Ireland	44	Saudi Arabia	99	Benin	154
Romania	45	Eritrea	100	Haiti	155
Canada	46	Swaziland	101	Mali	156
Netherlands	47	Côte d'Ivoire	102	Turkmenistan	157
Maldives	48	Trinidad and Tobago	103	Niger	158
Fiji	49	Guatemala	104	Togo	159
Philippines	50	Congo	105	Angola	160
Australia	51	Dem. Rep. Congo	106	Mauritania	161
Morocco	52	Malawi	107	Cent. African Rep.	162
Belarus	53	Kenya	108	Sierra Leone	163
Malaysia	54	Ghana	109		
Slovenia	55	Myanmar	110		

Legend

Countries whose EPI rank is very sensitive to the methodological assumptions (EPI rank to be treated with caution)
Countries whose EPI rank is sensitive to the methodological assumptions within acceptable limits (EPI rank reliable)
Countries whose EPI rank is very robust to the methodological assumptions (EPI rank highly reliable)

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

A total of 300 simulations were run in order to cover the space of uncertainties present in the 2010 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 simulated median rank (and its confidence interval) is proposed as a summary measure of a rank distribution. The results show that for the majority of the countries (103 of the 163), the 2010 EPI rank lies within the confidence interval for the median rank and additionally this confidence interval is narrow enough (less than 20 positions) to allow for reliable inference on those ranks, e.g. identify where environmental policies work well or where remedial action is needed.

However, the EPI ranks for the remaining 60 countries (e.g. Brunei, Cyprus, Japan, Luxembourg, Malta, Peru, Spain, UK, USA) depend strongly on the original methodological assumptions made in developing the Index and any inference on those countries rank should be formulated with great caution.

The top ten performing countries in the EPI include Iceland, Switzerland, Costa Rica, Sweden, Norway, Mauritius, France, Austria, Cuba and Colombia. However, the simulations indicate that some of those countries should be positioned much lower. Iceland, for example has a 2010 EPI rank: 1, but has a simulated median rank: 7 and a confidence interval [2, 8]. The simulations suggest that it is Switzerland and Costa Rica the two countries that excel in the 2010 EPI. Colombia and Cuba are expected to be ranked much lower (between rank 11 and 22).

What is the impact of measurement error in EPI?

A normally distributed random error term was added to the raw data with a mean zero and a standard deviation equal to one fifth of the observed standard deviation for each indicator. Overall, the introduction of measurement error in the raw data has a moderate impact on very few countries (the ten most affected countries shift roughly 10 positions), while the ranks of the majority of the countries do not change (Spearman correlation with EPI ranking is 0.997).

What is the impact of alternative weighting schemes or no structure in EPI?

Three 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: (a) current weighting vs. FA-derived weights at the indicator level; (b) current weighting vs. equal weighting at the indicator level; and (c) current weighting vs.

equal weighting at the policy level. The simulations showed that all of these scenarios have significant influence on the EPI ranking. The scenarios with the biggest impact are: equal weighting at the indicator level, followed by Factor Analysis derived weights at the indicator level, and by equal weighting at the policy level. In any of these three cases, 1 out of 2 countries shifts less than 16 positions with respect to the original EPI ranking, whilst 1 out of 10 countries shifts more than 41 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. Azerbaijan, Bolivia, Botswana, China, Egypt, Honduras, Indonesia, Cambodia, Namibia, Nicaragua, Korea and Turkmenistan improve their ranks by 20 positions or more, whilst the greatest decline is observed for Australia, Congo, Cyprus, Djibouti, Ireland, Kuwait, Luxembourg, Maldives, and Sao Tome and Principe (down more than 25 positions). Overall, for 1 out of 2 countries, the impact of this assumption is nine positions, while 1 out of 10 countries shift by more than 22 positions (maximum decline for Maldives of 64 positions).

The impact of the Borda-adjusted aggregation instead is more pronounced; under this assumption half of the countries shift less than fourteen positions but the most affected countries shift between 30 and 35 positions. Overall, the Spearman correlation coefficient between the 2010 EPI ranking and this scenario is 0.90.

What are the policy implications of these findings?

The overall performance of the 163 countries studied is in general satisfactory in six of the ten policy categories. However, the remaining policy categories related to Air pollution (effects on ecosystem), Climate Change, Biodiversity & Habitat and DALY represent the main challenges for the majority of the countries: half of the countries hardly manage to achieve 50 to 60 points.

Strong determinants of good environmental performance are, among others, (1) Environmental burden of disease (DALY); (2) Indoor air pollution; (3) Outdoor air pollution; (4) Access to water; and (5) Access to sanitation. Less influential but still significant on determining the 2010 EPI ranking are: the Water quality index, the Growing stock change, Forest cover change, Agricultural subsidies and Pesticide regulation.

Other important environmental aspects, such as Non-methane volatile organic compound emissions, Critical habitat protection, Greenhouse gas emissions, and Industrial greenhouse gas emissions intensity, although they were included in the conceptual framework, they do not bear any statistically significant association to the EPI ranks. These results do not imply that keeping greenhouse gas emissions at low levels and Critical habitat protection at high levels should not be among the policy objectives of the governments world wide. They simply point to the fact that even if governments made an effort to improve these aspects, the effort would not be captured by the EPI.

In order for a country to be ranked in the top fifty in the EPI ranking must put simultaneously invest in both Objectives of the EPI within a coherent environmental performance strategy, while emphasizing reduction of the existing gaps in areas where performance is lagging. However, this does not seem to be easy given the understandable, though problematic, trade-off between Environmental Health and Ecosystem Vitality (-0.32 Spearman rank correlation). Hence, the EPI framework suggests that it is not easy to translate environmental sustainability-oriented performance into practice.

What recommendations for future versions of EPI?

The statistical analysis of the quality of the EPI shows that, although the theoretical framework and the indicators for the EPI were carefully chosen by experts, the issue of weighting is crucial to obtain a robust performance index. The current weighting and normalization schemes result 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 main assumptions tested in the uncertainty and sensitivity analysis, the country ranks are relatively reliable for 109 countries, while any conclusion on the ranks for the remaining countries should be made with great caution. An equal weighting approach or factor analysis-derived weights at the indicator level, as opposed to the current weighting scheme greatly influences the ranks. Thus, the choice of the weights must be evaluated according to the EPI's analytical rationale, policy relevance, and implied value judgments.

If the objective of EPI is to promote action on all policies categories more work would be needed to ensure that all policy fields have an impact on the aggregated EPI or, alternatively, policy categories should be given more emphasis than the aggregated measure.

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 2010 necessitates such type of analysis, so as to ensure that the methodology remains appropriate. At the same time, our study aims at identifying those countries for which the EPI ranking is robust as well as those for which it is not. . 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. For the latter a more cautious approach is advised before translating the EPI rank into policy actions.

Transparency to stakeholders is considered to be essential ingredient of well built composite indicators (OECD, 2008). A clear understanding of the EPI methodology is also necessary with a view to perform the robustness assessment of the index. Thus our first test has been: is it 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. The EPI is clear about its normative assumptions, and does not fall under the critiques of normative ambiguity at times addressed to composite indicators (see Stiglitz report, p. 65).

Indisputably, the construction of the EPI demands a sensitive balance between simplifying an environmental system and still providing sufficient detail to detect characteristic elements within it. 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. 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.

2. How does the EPI associate to its underlying components?

A simple rank correlation analysis between the 2010 EPI and the two Objectives (Table 1) reveals that the EPI is strongly correlated with the Environmental Health (environmental stress to human health) with $r_s = .77$, but it has a very low correlation with the Ecosystem Vitality ($r_S = .27$). As expected, the correlations between the 2010 EPI and the policy categories follow along the same lines (Table 2). In fact, the EPI has high correlations with the three policy categories under the Environmental Health Objective ($r_S \geq 0.70$) and only moderate to low correlation with the remaining seven policy categories under the Ecosystem Vitality Objective ($r_S < 0.5$). Practically random (non-significant at the 95% level) are the correlations between the EPI and four of the policy categories, namely to Air pollution (ecosystem), Biodiversity & Habitat, Fisheries and Climate Change.

Relationships among the policy categories themselves vary, but they are in general high among the policies within the Environmental Health and low among the policies within the Ecosystem Vitality. These results were in part expected. On one hand, the Environmental Health Objective is composed of DALY, Air Pollution (effects on humans) and Water Pollution (effects on humans). However, the DALY is calculated as an un-weighted sum of DALY data for three sources of environmental health risk −diarrhea, indoor air, and outdoor air. Thus, the three policy categories within the Environmental Health Objective provide, to a great extent, overlapping information. On the other hand, the Ecosystem Vitality is composed of policy categories that represent totally different aspects of the environmental impact on the ecosystem; this is desirable from an index development perspective since representing different dimensions is a key quality feature of a composite indicator. Yet, the negative association between several of the policy categories leads to a conclusion that there may be trade-offs between them. This creates an additional difficulty in EPI that combines different dimensions with the implicit assumption that strong performance on all policy categories should be pursued simultaneously.

A step to partially overcome these difficulties would be standardization at the level of the policy categories or – at least – at the level of the objectives. Standardizing a variable implies subtracting its mean and dividing by its standard deviation, thus rendering the variable roughly distributed as a standard normal (OECD, 2008). If standardization had been applied at the Objective level, then Ecosystem Vitality and Environmental Health would have roughly the same impact on the final EPI ranking. This possibility may be considered perhaps at a next version of the index.

Staying instead with the present EPI architecture, a recommendation that stems from the correlation analysis is that the added-value of EPI lies not in the overall country ranking but in the ten policy categories and the two objectives (Humans and Ecosystem).

One should thus try to identify linkages and trade-offs between them, instead of aggregating all into a single score.

Table 1. Spearman rank correlation coefficients between EPI and its Objectives

	Environmental Health	Ecosystem Vitality
EPI	0.77	0.27
Environmental Health		-0.32

Table 2. Spearman rank correlation coefficients between EPI and its ten policy categories

	Environmental burden of disease	Air pollution (effects on humans)	Water (effects on humans)	Air Pollution (effects on ecosystem)	Water (effects on ecosystem)	Biodiversity & Habitat	Forestry	Fisheries	Agriculture	Climate Change
EPI	.69	.75	.69	-.12*	.42	.15	.48	.17	.38	.00*
Air pollution (effects on humans)	.67									
Water (effects on humans)	.90	.72								
Air Pollution (effects on ecosystem)	-.40	-.12*	-.35							
Water (effects on ecosystem)	.21	.25	.22	.00*						
Biodiversity & Habitat	-.08*	.05*	-.04*	-.06*	.23					
Forestry	.57	.57	.59	-.19	.02*	-.21				
Fisheries	-.01*	.12*	-.09*	.03*	.10*	.13*	-.08*			
Agriculture	.21	.10*	.21	-.10*	.28	.15	-.04*	.10*		
Climate Change	-.53	-.38	-.54	.30	.00*	-.01*	-.30	.10*	-.02*	

*Coefficient not significant at 5% level.

Further study of the association between the EPI and the 25 underlying indicators reveals that the primary drivers of the EPI ranking are just five indicators: DALY, Indoor air pollution, Outdoor air pollution, Access to water and Access to sanitation (Table 3). Less influential but still significant on determining the 2010 EPI ranking are: the Water quality index, the Growing stock change and Forest cover change and the Agricultural subsidies and Pesticide regulation. The three indicators related to Climate Change, although being weighted comparatively strongly, do not exert much influence on the 2010 EPI results.

Of the 25 indicators included in the 2010 EPI framework, there are twelve indicators that appear to be randomly associated with either the overall EPI and/or with the Objective they belong to (Table 3). These indicators are:

- Non-methane volatile organic compound emissions,
- Water quality Index, and Water stress Index,
- Biome protection, Marine protection, and Critical habitat protection,
- Marine trophic index,
- Agricultural water intensity, Agricultural subsidies, and Pesticide regulation,
- Greenhouse gas emissions per capita, and Industrial greenhouse gas emissions intensity.

Table 3. Spearman rank correlation coefficients between EPI and its indicators

Indicators in the EPI framework	**Correlation with EPI**	**Correlation with the Environmental Health**
Environmental burden of disease -DALY	.69	.95
Indoor air pollution	.62	.88
Outdoor air pollution	.60	.52
Access to Water	.65	.90
Access to sanitation	.66	.92
		Correlation with the Ecosystem Vitality
Sulfur dioxide emissions	-.30	.36
Nitrogen oxides emissions	-.30	.32
Non-methane volatile organic compound emissions	-.09*	.21
Ecosystem ozone	.26	-.16
Water quality Index	.46	.12*
Water stress Index	.10*	.30
Water scarcity index	.21	.35
Biome protection	.14*	.24
Marine protection	.25	.08*
Critical habitat protection	.19*	.27
Growing stock change	.54	-.22
Forest cover change	.48	-.17
Marine trophic index	.10*	-.02*
Trawling intensity	.22	.31
Agricultural water intensity	.11*	.39
Agricultural subsidies	-.45	.02*
Pesticide regulation	.52	.08*
Greenhouse gas emissions per capita	-.15*	.65
CO2 emissions per electricity generation	.37	.38
Industrial greenhouse gas emissions intensity	-.08*	.29

* coefficient not significant ($p > 0.05$).

The random association between the EPI ranks (or objectives' ranks) and these twelve indicators should not be taken to mean that these indicators do not describe important environmental issues. Instead, these random associations imply that even if some countries improve their relative position in any of those twelve indicators, this improvement will not lead to a better position either in the EPI rank and/or in the respective objectives' rank. Parsimony principles would suggest excluding the non-influential indicators from the EPI framework (Booysen, 2002; Gall, 2007). This, however, may not be advisable from a policy perspective, as excluding certain indicators will be resisted by experts due to the relevance of the indicators to the issue. As already shown above, it is difficult in an environmental study, and given the multidimensionality of the subject, to aggregate to a single measure without losing track of individually relevant dimensions.

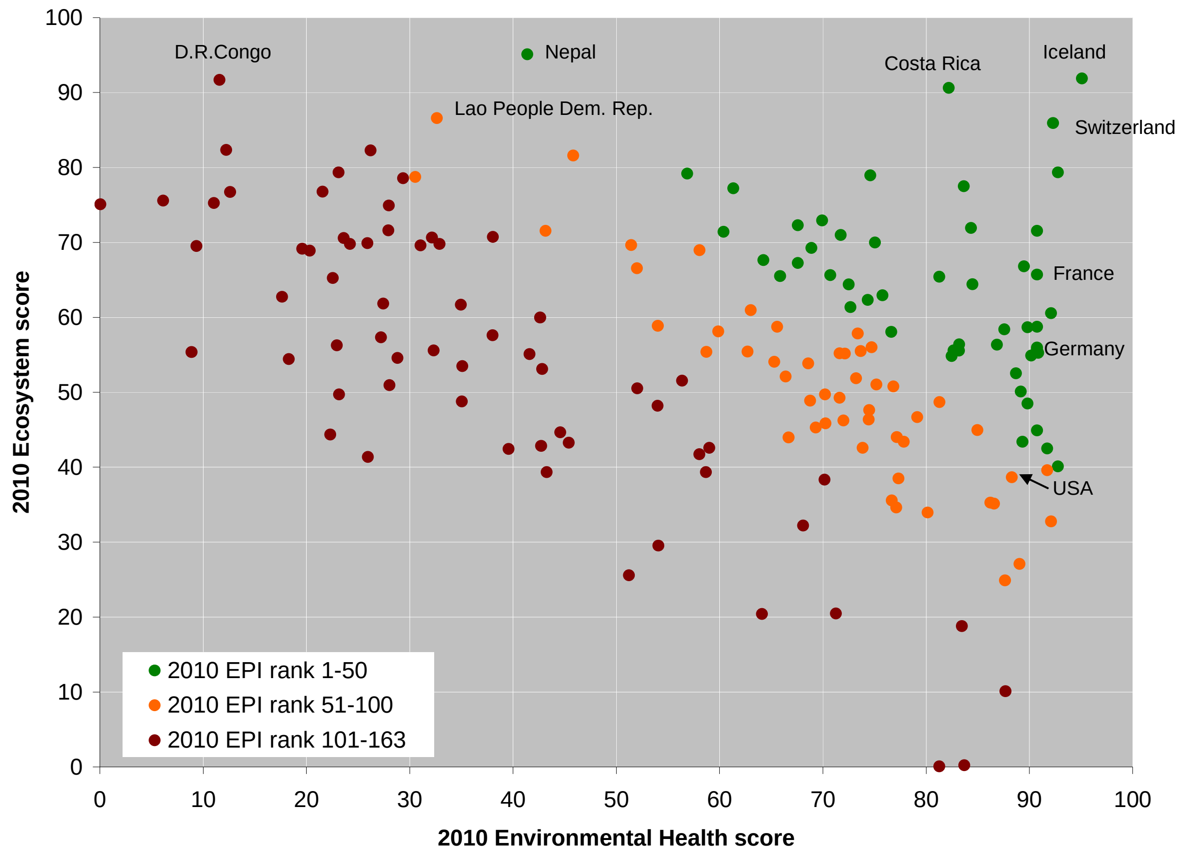
The scatter plot between the two EPI Objectives in Figure 1 shows that in order for a country to be ranked in the top fifty in the EPI ranking must put simultaneously invest in both Objectives of the EPI within a coherent environmental performance strategy, while emphasizing reduction of the existing gaps in areas where performance is lagging. However, this does not seem to be easy given the understandable − though problematic– trade-off between the two Objectives (low but significant negative association between Environmental Health and Ecosystem Vitality, $r_S = -.32$). Hence, the EPI framework suggests that it is not easy to translate environmental sustainability-oriented performance into practice.

Note that part of the problem also stems from the linear aggregation approach, which, while commonly adopted in most of the existing composite indicators, is also the one fraught with more methodological problems due to its inherent compensability and to the well known misperception of weights taken as measures of importance.

It is easy to illustrate this for the case of EPI. To a stakeholder the information that Ecosystem Vitality and Environmental Health each 'weighs' 50% of the total is automatically translated into them being equally important. As mentioned above this is not the case. Environmental Health and Ecosystem Vitality have different variances and despite the equal weights they do weight differently in EPI. This is well known to practitioners, who prefer to eschew linear aggregation in favour of e.g. partial ordering or multi-criteria (e.g. Borda- or Condorcet-based) aggregation (Munda, 2008). To make an example, when using a Condorcet-based aggregation the weights retain in full the meaning of importance. As mentioned, developers in general prefer linear aggregation for its simplicity, transparency and reproducibility. One needs software to apply non compensatory methods such as e.g. Condorcet. A possible way to alleviate this trade off between model simplicity and analytic coherence would be to ensure that – even if weights are not importance – at least they do not deviate too much from it. A way of doing this is by standardizing the variables of the policy categories or the objectives as appropriate.

Overall, correlation analysis results indicate that the 2010 EPI has an architecture that highlights the complexity of translating environmental stewardship into straightforward, clear-cut policy recipes. The trade-offs within the EPI policy categories included under the Ecosystem Objective 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 1. Scatterplot of the two EPI Objectives



3. How robust are EPI ranks to the methodological assumptions?

International statistical organizations have made progress in establishing good practices in the construction of composite indicators and ranking systems (OECD, 2008) and practitioners strongly recommend undertaking a robustness analysis before making the composite indicator public (Kennedy, 2007; Saltelli et al., 2008). We shall make use of these tools to investigate the methodological robustness of the 2010 EPI ranking.

When building an index to capture environmental performance along two main axis − Humans and Ecosystem− it is necessary to take stock of existing methodologies in order to avoid possible bias in the assessment and decision-making. By conducting uncertainty analysis and hence acknowledging the variety of methodological assumptions involved in the development of an index, one can determine whether the main results change substantially when the main assumptions are varied over a reasonable range of possibilities. This approach helps to avert the criticism addressed to composite measures or rankings, namely that they are presented as if they had been calculated under conditions of certainty (while this is rarely the case) and then taken at face value by end-users (Sharpe, 2004; Saisana et al., 2005; Saisana and Saltelli, 2008a). The objective of UA is not to establish the truth or to verify whether the EPI is a legitimate model to measure environmental performance world wide, but rather to test whether the ranking itself and/or its associated inferences are robust or volatile with respect to changes in the methodological assumptions within a plausible and legitimate range.

Further, the type of uncertainty analysis we will apply here allows us to propose an alternative measure for ranking countries which is dependant of the framework (selected set of indicators) but not on the methodological choices (weighting or type of aggregation). We adopt for this study a multi-modelling approach (Saisana, 2008; Saisana and Munda, 2008), whereby different combinations of aggregation and weighting are taken as different models within the same normative framework. Applying these models to the EPI indicators allows us to produce a simulated median ranking for EPI, which is dependant on the framework of the 25 EPI indicators but robust with respect to the methodological assumptions. With this new measure, we can contrast country performance with respect to the original 2010 EPI ranking.

3.1 Multi-modelling approach

In the case of the 2010 EPI, the assumptions that needed to be tested are:

- measurement error of the raw data,
- EPI structure – grouping at policy categories,
- weights assigned to the indicators and/or to the policy categories,
- aggregation function at the policy or at the objectives level, and
- number of indicators or policy categories.

(a) 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 one fifth of the observed standard deviation for each indicator. Several alternative datasets that include error in some of the data values are generated to this end.

(b and c) Assumption on the EPI structure and the weighting scheme: In the 2010 EPI an expert-based weighting scheme was used. Although this is a legitimate choice, it is not unique and it is hard to find a theoretical justification for it. To anticipate criticism, we tested three alternative and legitimate options: factor analysis derived weights[1] across all 25 indicators; equal weighting across all 25 indicators; and equal weighting across the 10 policy categories.

(d) Assumption on the aggregation function: The EPI rankings are built using a weighted arithmetic average, hence a linear aggregation rule (Eq. (1)) of the 25 indicators. Decision theory practitioners have challenged aggregations based on additive models because of inherent theoretical inconsistencies (Munda, 2008) and the fully compensatory nature of linear aggregation, in which an x% increase in one indicator can offset an y% decrease in another, where y depends from the ratio of the weights of the two variables. This is the reason why practitioners call weights in linear aggregation 'trade-off coefficients', not to be confused with measures of importance.

We would argue that at the first level of aggregation, the calculation of the 2010 EPI policy categories as a weighted arithmetic average of the indicators has the advantage of "compensating" for eventual inconsistencies in the data. At the second level of aggregation, instead, namely from the policy categories into the overall EPI, the use of a less compensatory aggregation function would be more advantageous, as it would imply that a country should place more effort in improving itself in those policy categories where it is relatively weak. To this end, we applied two alternative aggregation functions: a geometric weighted average (Eq. (2)) and a multi-criteria method[2].

In the case of the geometric averaging, we shifted slightly the policy categories scores to above 1.00 to allow for the proper use of the geometric aggregation. From the multi-criteria literature, we selected a method suggested by Brand et al. (2007) (Eq. (3)) because it can deal with a large number of countries and it can also deal with eventual ties in the policy categories scores.

---

[1] upon factor rotation and squaring of the factor loadings, as described in Nicoletti et al. (2000)

[2] Both geometric aggregation and the Borda method applied here are less compensatory than linear weighting. For details see OECD (2008).

Weighted Arithmetic Average score: $y_j = \sum_{i=1}^{n} w_i \cdot x_{ij}$ (1)

Weighted Geometric Average score: $y_j = \prod_{i=1}^{n} x_{ij}^{w_i}$ (2)

Borda adjusted score: $y_j = \sum_{i=1}^{n} (m_{ij} + \frac{k_{ij}}{2}) \cdot w_i$ (3)

y_j : composite indicator score for country j, w_i : weight attached to policy category i, x_{ij} : score for country j on policy category i, m_{ij} : number of countries that have weaker performance than country j relative to policy category i; k_{ij} : number of countries with equivalent performance to country j relative to policy category i.

(e) Assumption on the number of indicators and policy categories: We have either kept all 25 indicators or in some cases excluded one at a time. We have done the same for the ten policy categories, that is either kept all ten policy categories or in some cases excluded one at a time.[3] This statistical procedure is a tool to test the robustness of inference and should not be seen as a disturbance of the framework. In fact it makes it possible to assess the impact of assigning a zero weight to an indicator or to a policy category, combined with the other assumptions on the weighting method and aggregation rule. Eliminating an indicator or a policy category from the framework can also be seen as "tuning" the ranking in favour of countries which have a comparative disadvantage on that aspect (Grupp and Mogee, 2004)[4].

The analysis of capping the raw data at target values and of correcting for skewed data distributions (winsorization) were not included in this year's assessment of the EPI because they were found to be of almost no importance in the 2008 EPI (Saisana and Saltelli, 2008b).

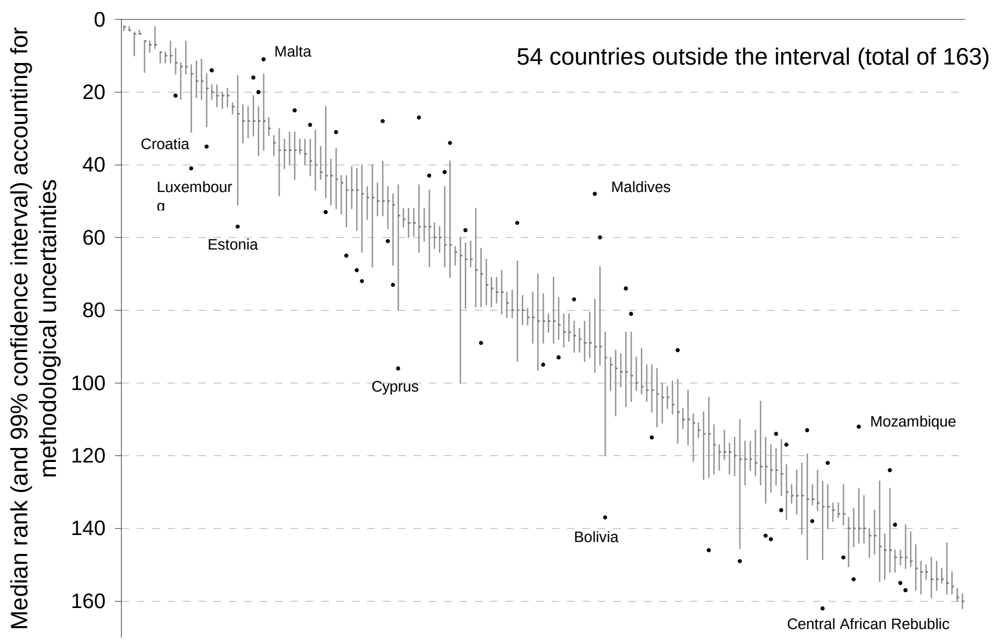
[3] Note that when an indicator is excluded from the framework, all policy categories are kept. Also when one policy category is excluded, all the indicators for the remaining nine categories are included.
[4] Note that large variations in the median rank of countries are not due to the elimination of one indicator (or policy category) at a time. In fact, the Spearman rank correlation coefficient between the 2010 EPI ranking and the median of the 25 rankings produced by eliminating one indicator (while keeping fixed the weighting scheme and aggregation method) from the respective framework is greater than 0.998. The same comment holds for the elimination of one policy category at a time. Instead, this exercise allows us to get less volatile estimates of the median rank. To be more specific, had one estimated the bootstrapped confidence interval for the median rank by using only those scenarios that employ the full framework, there would have been roughly 30% more countries with confidence intervals greater than 20 positions compared to those reported above for the 300 scenarios.

The combinations of these assumptions are translated into a set of roughly $N \approx 300$ 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 assumptions to appraise their influence.

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

The uncertainty analysis results from the Monte Carlo simulations for the 163 countries are given in detail in Figure 2. The graph presents the 'median' performance across all 300 models as a summary measure of the plurality of stakeholders' views on how to combine the information in order to assess environmental performance. The 99% confidence interval for each country and the countries whose original 2010 EPI rank does not fall within this interval are also displayed. Confidence intervals were estimated using bootstrap (1000 samples taken with replacement, see Efron, 1979).

Figure 2. Simulated median and its 99% confidence interval for the EPI ranks



Note: The dots relate a country's 2010 EPI rank to the median rank calculated over the set of plausible scenarios (roughly 300 models) generated in our uncertainty analysis to account for measurement error in the raw data, structure, weights, aggregation function, indicators/policy categories. Ranks that fall outside the interval are marked in black.

While for the majority of the countries the EPI rank lies within the confidence interval estimated in our simulations, 54 countries appear to be slightly misplaced. For example, Japan, Malta and Peru have been favoured by the choices made in the 2010 EPI, while Brunei, Cyprus and Luxembourg were placed in a worse position than our simulations would suggest. Needless to say that these shifts were non-intentional, but they were inherent in the methodological choices in the EPI construction, while uncertainty analysis brings them into light. Any message conveyed by the 2010 EPI for those 54 countries should, therefore, be formulated with great caution and considered only as contingent on the original methodological assumptions made in developing the Index (see Table 4).

Table 4. Countries whose EPI rank lies outside the simulated confidence interval

"favored" by the 2010 EPI (alphabetical order)	EPI rank	Simulated conf. int.	"disfavored" by the 2010 EPI (alphabetical order)	EPI rank	Simulated conf. int.
Algeria	42	[46, 68]	Belarus	53	[24, 49]
Antigua and Barbuda	27	[46, 64]	Benin	154	[135, 145]
Bangladesh	139	[146, 152]	Bolivia	137	[86, 120]
Chile	16	[21, 32]	Botswana	149	[110, 146]
El Salvador	34	[39, 71]	Brunei Darussalam	72	[40, 64]
Japan	20	[24, 38]	Bulgaria	65	[43, 57]
Kuwait	113	[120, 149]	Cambodia	148	[128, 139]
Libyan Arab Jamahiriya	117	[123, 138]	Central African Rep.	162	[127, 149]
Maldives	48	[77, 97]	Croatia	35	[15, 30]
Malta	11	[15, 36]	Cyprus	96	[46, 80]
Mexico	43	[47, 68]	Equatorial Guinea	146	[104, 126]
Mozambique	112	[129, 144]	Estonia	57	[16, 51]
Namibia	81	[86, 105]	Gabon	95	[76, 89]
Paraguay	60	[68, 95]	Haiti	155	[146, 150]
Peru	31	[36, 52]	Jamaica	89	[63, 79]
Qatar	122	[128, 140]	Latvia	21	[8, 15]
Sao Tome and Principe	91	[99, 117]	Luxembourg	41	[13, 31]
Serbia and Montenegro	29	[33, 43]	Mongolia	142	[115, 133]
Singapore	28	[39, 54]	Nicaragua	93	[77, 88]
Solomon Islands	114	[118, 128]	Papua New Guinea	138	[128, 134]
Spain	25	[31, 44]	Russian Federation	69	[41, 58]
Sri Lanka	58	[62, 80]	Rwanda	135	[116, 131]
Syrian Arab Republic	56	[67, 94]	Senegal	143	[117, 130]
Tunisia	74	[86, 107]	South Africa	115	[95, 108]
Turkey	77	[83, 92]	F.Y.R.O.M	73	[48, 68]
UK & N. Ireland	14	[18, 22]	Turkmenistan	157	[139, 151]
Yemen	124	[129, 152]	USA	61	[46, 58]

The widest confidence intervals for the median rank are estimated for twenty four countries (>20 positions) which are shown in Table 5. A very high volatility, between 32 and 40 positions is found for El Salvador (rank: 34), Estonia (57), Cyprus (96), Trinidad and Tobago (103), Bolivia (137) and Botswana (149). The volatility of those countries is due to the combined effect of all five assumptions, although the most influential assumptions are the use of equal weighting or Factor Analysis weighting at the indicators level and the use of geometric versus a arithmetic average aggregation function at the policy level. Most of these countries were also found above to be misplaced in the EPI ranking.

Despite these concerns, for the majority of the countries, namely for 103 of the 163 countries, the 2010 EPI rank lies within the confidence interval for the median rank and additionally this confidence interval is narrow enough (less than 20 positions) to allow for reliable inference on those ranks. Hence, for those countries the EPI rank can be used as an indication of where environmental policies work well and where remedial action is needed.

Table 5. Most volatile countries in the EPI

Country (alphabetical order)	EPI rank	Simulated conf. int.
Algeria	42	[46, 68]
Belarus	53	[24, 49]
Bolivia	137	[86, 120]
Botswana	149	[110, 146]
Brunei Darussalam	72	[40, 64]
Burkina Faso	128	[105, 128]
Central African Rep.	162	[127, 149]
Cyprus	96	[46, 80]
Egypt	68	[70, 97]
El Salvador	34	[39, 71]
Equatorial Guinea	146	[104, 126]
Estonia	57	[16, 51]
Kuwait	113	[120, 149]
Lao People's Dem. Rep.	80	[52, 79]
Maldives	48	[77, 97]
Malta	11	[15, 36]
Mexico	43	[47, 68]
Nepal	38	[40, 68]
Pakistan	125	[127, 155]
Paraguay	60	[68, 95]
Syrian Arab Rep.	56	[67, 94]
Trinidad and Tobago	103	[60, 100]
Tunisia	74	[86, 107]
Yemen	124	[129, 152]

A discussion on the top performing countries is in place. The top ten performing countries in the EPI include Iceland, Switzerland, Costa Rica, Sweden, Norway, Mauritius, France, Austria, Cuba and Colombia. Most of these countries were also among to the top ten performing countries also in 2008 EPI (namely Switzerland, Sweden, Norway, Costa Rica, Austria and France). However, the simulations indicate that some of those countries should be positioned much lower. Iceland, for example has a 2010 EPI rank: 1, but has a simulated median rank: 7 and a confidence interval [2, 8]. The simulations suggest that it is Switzerland and Costa Rica the two countries that excel in the 2010 EPI. Colombia and Cuba are expected to be ranked much lower (between rank 11 and 22).

Table 6 presents the 2010 EPI ranks under these uncertainty considerations and could be used as a guide on the interpretation of the 2010 EPI results.

These simulations have helped us to estimate country ranks that depend on the 25 indicators of environmental performance, as these were selected by the EPI team and the invited experts, but are independent of the methodological choices made during the EPI development.

Table 6. 2010 EPI ranks with uncertainty considerations

Country	EPI rank	Median rank	[99% conf. int.]	Country	EPI rank	Median rank	[99% conf. int.]	Country	EPI rank	Median rank	[99% conf. int.]
Iceland	**1**	7	[2, 8]	Syrian Arab Rep.	56	80	[67, 94]	**Tajikistan**	**111**	111	[109, 122]
Switzerland	**2**	2	[2, 3]	Estonia	57	26	[16, 51]	Mozambique	112	140	[129, 144]
Costa Rica	**3**	3	[2, 3]	Sri Lanka	58	66	[62, 80]	Kuwait	113	132	[120, 149]
Sweden	**4**	4	[3, 4]	**Georgia**	**59**	60	[56, 63]	Solomon Islands	114	124	[118, 128]
Norway	**5**	10	[6, 12]	Paraguay	60	90	[68, 95]	South Africa	115	102	[95, 108]
Mauritius	**6**	13	[6, 15]	USA	61	50	[46, 58]	**Gambia**	**116**	119	[117, 124]
France	**7**	10	[9, 12]	**Brazil**	**62**	56	[52, 60]	Libyan Ar. Jam.	117	130	[123, 138]
Austria	**8**	7	[6, 9]	**Poland**	**63**	60	[57, 64]	**Honduras**	**118**	119	[113, 121]
Cuba	**9**	17	[11, 22]	**Venezuela**	**64**	64	[63, 68]	**Uganda**	**119**	119	[117, 125]
Colombia	**10**	17	[12, 22]	Bulgaria	65	47	[43, 57]	**Madagascar**	**120**	120	[115, 126]
Malta	11	28	[15, 36]	**Israel**	**66**	75	[69, 81]	**China**	**121**	122	[113, 126]
Finland	**12**	9	[9, 12]	**Thailand**	**67**	66	[61, 68]	Qatar	122	134	[128, 140]
Slovakia	**13**	4	[4, 10]	Egypt	68	83	[70, 97]	**India**	**123**	131	[124, 136]
UK & N. Ireland	14	20	[18, 22]	Russian Federation	69	47	[41, 58]	Yemen	124	146	[129, 152]
New Zealand	**15**	6	[6, 15]	**Argentina**	**70**	74	[71, 79]	Pakistan	125	145	[127, 155]
Chile	16	28	[21, 32]	**Greece**	**71**	73	[66, 79]	**Tanzania**	**126**	117	[112, 125]
Germany	**17**	21	[18, 24]	Brunei Darussalam	72	48	[40, 64]	**Zimbabwe**	**127**	121	[116, 124]
Italy	**18**	21	[20, 25]	f.Y.R.O.M	73	51	[48, 68]	Burkina Faso	128	123	[105, 128]
Portugal	**19**	21	[19, 24]	Tunisia	74	97	[86, 107]	**Sudan**	**129**	140	[131, 144]
Japan	20	28	[24, 38]	**Djibouti**	**75**	78	[75, 82]	**Zambia**	**130**	121	[116, 128]
Latvia	21	12	[8, 15]	**Armenia**	**76**	75	[71, 78]	**Oman**	**131**	131	[122, 142]
Czech Republic	**22**	13	[12, 22]	Turkey	77	87	[83, 92]	**Guinea-Bissau**	**132**	133	[124, 135]
Albania	**23**	28	[24, 34]	**Iran (Islam. Rep.)**	**78**	86	[81, 90]	**Cameroon**	**133**	131	[128, 133]
Panama	**24**	24	[23, 26]	**Kyrgyzstan**	**79**	83	[81, 86]	**Indonesia**	**134**	135	[133, 138]
Spain	25	36	[31, 44]	Lao P. Dem. Rep.	80	69	[52, 79]	Rwanda	135	125	[116, 131]
Belize	**26**	28	[24, 33]	Namibia	81	98	[86, 105]	**Guinea**	**136**	136	[133, 137]
Antigua-Barbuda	27	57	[46, 64]	**Guyana**	**82**	80	[75, 82]	Bolivia	137	93	[86, 120]
Singapore	28	50	[39, 54]	**Uruguay**	**83**	83	[71, 89]	Papua N.Guinea	138	132	[128, 134]
Serbia-Montenegro	29	39	[33, 43]	**Azerbaijan**	**84**	82	[80, 84]	Bangladesh	139	148	[146, 152]
Ecuador	**30**	30	[27, 32]	**Viet Nam**	**85**	88	[85, 92]	**Burundi**	**140**	140	[137, 151]
Peru	31	44	[36, 52]	**Moldova Rep.**	**86**	86	[81, 89]	**Ethiopia**	**141**	142	[140, 148]
Denmark	**32**	36	[33, 41]	**Ukraine**	**87**	80	[76, 84]	Mongolia	142	123	[115, 133]
Hungary	**33**	34	[32, 38]	**Belgium**	**88**	96	[91, 109]	Senegal	143	124	[117, 130]
El Salvador	34	62	[39, 71]	Jamaica	89	70	[63, 79]	**Uzbekistan**	**144**	146	[142, 154]
Croatia	35	19	[15, 30]	**Lebanon**	**90**	89	[81, 93]	**Bahrain**	**145**	154	[148, 159]
Dominican Rep.	**36**	36	[31, 40]	S. Tome- Principe	91	108	[99, 117]	Eq. Guinea	146	114	[104, 126]
Lithuania	**37**	36	[33, 37]	**Kazakhstan**	**92**	82	[75, 89]	**Korea D.P.Rep.**	**147**	142	[135, 147]
Nepal	38	49	[40, 68]	Nicaragua	93	84	[77, 88]	Cambodia	148	136	[128, 139]
Suriname	**39**	37	[33, 40]	**Rep. Korea**	**94**	95	[93, 102]	Botswana	149	121	[110, 146]
Bhutan	**40**	40	[31, 47]	Gabon	95	83	[76, 89]	**Iraq**	**150**	154	[149, 157]
Luxembourg	41	15	[13, 31]	Cyprus	96	54	[46, 80]	**Chad**	**151**	149	[141, 151]
Algeria	42	62	[46, 68]	**Jordan**	**97**	89	[83, 94]	**U. Ar. Emirates**	**152**	152	[149, 158]
Mexico	43	57	[47, 68]	**Bosnia-Herzeg.**	**98**	103	[96, 112]	**Nigeria**	**153**	152	[150, 154]
Ireland	**44**	45	[43, 54]	**Saudi Arabia**	**99**	97	[92, 101]	Benin	154	140	[135, 145]
Romania	**45**	42	[35, 44]	**Eritrea**	**100**	100	[93, 102]	Haiti	155	148	[146, 150]
Canada	**46**	50	[45, 54]	**Swaziland**	**101**	110	[102, 117]	**Mali**	**156**	154	[151, 155]
Netherlands	**47**	49	[46, 55]	**Côte d'Ivoire**	**102**	102	[95, 105]	Turkmenistan	157	148	[139, 151]
Maldives	48	90	[77, 97]	Trinidad&Tobago	103	65	[60, 100]	**Niger**	**158**	155	[144, 158]
Fiji	**49**	47	[41, 52]	**Guatemala**	**104**	104	[101, 107]	**Togo**	**159**	156	[152, 158]
Philippines	**50**	56	[52, 60]	**Congo**	**105**	101	[91, 103]	**Angola**	**160**	151	[145, 157]
Australia	**51**	36	[30, 49]	**Dem. Rep. Congo**	**106**	114	[108, 127]	**Mauritania**	**161**	159	[157, 160]
Morocco	**52**	57	[52, 61]	**Malawi**	**107**	110	[107, 113]	Cent. African R.	162	134	[127, 149]
Belarus	53	43	[24, 49]	**Kenya**	**108**	106	[100, 109]	**Sierra Leone**	**163**	160	[158, 162]
Malaysia	**54**	43	[39, 51]	**Ghana**	**109**	104	[100, 111]				
Slovenia	**55**	55	[52, 57]	**Myanmar**	**110**	113	[111, 115]				

Legend

Countries whose EPI rank is very sensitive to the methodological assumptions (EPI rank to be treated with caution)
Countries whose EPI rank is sensitive to the methodological assumptions within acceptable limits (EPI rank reliable)
Countries whose EPI rank is very robust to the methodological assumptions (EPI rank highly reliable)

3.3 Which assumptions have the highest impact on the EPI ranking?

Complementary to the uncertainty analysis, a sensitivity analysis makes it possible to assess the impact of a modeling scenario on the 2010 EPI ranking. To this end, we calculate for each country the absolute rank shift between the EPI rank and the rank provided by a scenario and then summarize these shifts over all 163 countries by using the 50th percentile, the 90th percentile and the Spearman rank correlation coefficient, which serve as our sensitivity measures. Table 7 provides the sensitivity analysis results for selected scenarios that are based on the entire set of 25 indicators.

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 one fifth of the observed standard deviation for each indicator. Overall, the introduction of measurement error in the raw data has a moderate impact on very few countries (the ten most affected countries shift roughly 10 positions), while the ranks of the majority of the countries do not change (Spearman correlation with EPI ranking is 0.997).

What is the impact of alternative weighting schemes or no structure in EPI?

Three 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 policy level.

The simulations showed that all of these scenarios have significant influence on the EPI ranking. The scenarios with the biggest impact are: equal weighting at the indicator level, followed by Factor Analysis derived weights at the indicator level, and by equal weighting at the policy level. In any of these three cases, 1 out of 2 countries shifts less than 16 positions with respect to the original EPI ranking, whilst 1 out of 10 countries shifts more than 41 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. Azerbaijan, Bolivia, Botswana, China, Egypt, Honduras, Indonesia, Cambodia, Namibia, Nicaragua, Korea and Turkmenistan improve their ranks by 20 positions or more, whilst the greatest decline is observed for Australia, Congo, Cyprus, Djibouti, Ireland, Kuwait, Luxembourg, Maldives, and Sao Tome and Principe (down more than 25 positions). Overall, for 1 out of 2 countries, the impact of this assumption is nine positions, while 1 out of 10 countries shift by more than 22 positions (maximum decline for Maldives of 64 positions).

The impact of the Borda-adjusted aggregation instead is more pronounced; under this assumption half of the countries shift less than fourteen positions but the most affected countries shift between 30 and 35 positions. Overall, the Spearman correlation coefficient between the 2010 EPI ranking and this scenario is 0.90.

Table 7. Impact of the methodological assumptions on the EPI ranking

Scenario	**50th prctile**	**90th prctile**	**Spearman rank corr. with EPI**
Measurement error in the raw data	3	7	0.99
Geometric aggregation of the policy categories	9	22	0.95
Equal weights for the ten policy categories	11	30	0.92
Equal weights for the ten policy categories and Borda-adjusted aggregation	12	33	0.91
Factor Analysis-weights for the 25 indicators	12	36	0.90
Equal weights for the ten policy categories and geometric aggregation	12	36	0.89
Borda-adjusted aggregation for the ten policy categories	14	35	0.90
Equal weights for the 25 indicators	16	41	0.86

Note: The 50th and 90th percentiles are calculated over the absolute rank shift between the EPI rank and the rank provided by a given scenario (over all 163 countries).

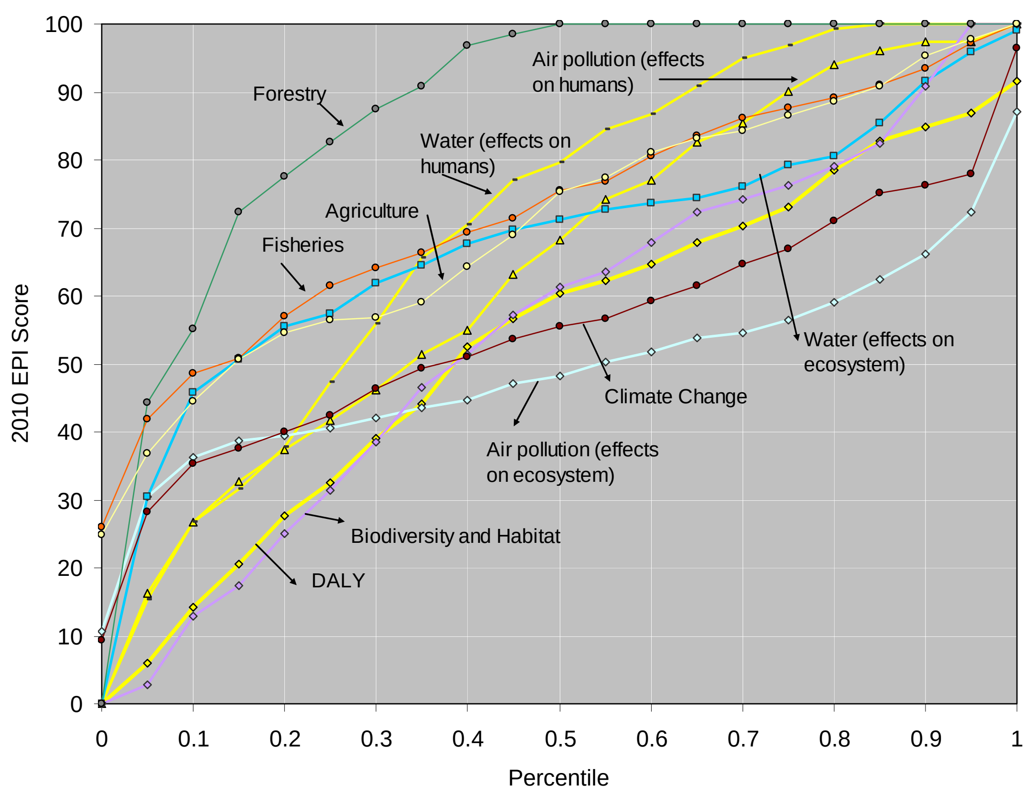
Although the different scenarios produce relatively different rankings compared to the EPI ranking, the Spearman rank correlation between the 2010 EPI and the median of all 300 scenarios considered is 0.96, which shows a high degree of confidence in the overall EPI classification. However, certain countries are more sensitive than others in the methodological choices and hence their ranks need to be treated with caution when such ranks are used to formulate policy statements.

4. What are the policy implications of these findings?

While the 2010 EPI ranks are reliable for the majority of the countries analyzed (for 103 out of 163), for the remaining countries the EPI ranks need not be taken at face value as they are particularly sensitive to the methodological assumptions in the Index development. However, the overall 2010 EPI results provide a reliable picture of the situation at global level (high degree of correlation between the simulated median ranking and the EPI ranking). Hence, while a country will score higher than some and lower than others, the added value of the EPI should not be seen as identifying winners and losers. Instead, the EPI can be used to generate a discussion about what policies contribute to good environmental performance and also provide insight into the nature of environmental policy challenges at the global scale.

Along these lines, Figure 3 shows that at a global scale, the best overall environmental performance is found in the Forestry policy category, in which half of the countries score 100 points and 80% of the countries obtain scores greater than 78 points. Also satisfactory is overall country performance on Air pollution (effects on humans), Water (effects on humans), Agriculture and Fisheries. There is one policy category for which most countries' performance is particularly worrying: Air pollution (effects on ecosystem). Half of the countries do not score more than 50 points and not a single country achieves a 100 score. Also worrying is overall country performance for the Climate Change, Biodiversity & Habitat and DALY. These four policy categories need remedial action and pose the highest environmental challenges at the global scale.

Figure 3. 2006 Index and pillar scores (and ranks)



The high degree of confidence in the overall EPI results suggests that robust conclusions can be drawn by studying the associations between the EPI scores and variables of interest such GDP per capita, the Human Development Index or other. However, we remind the reader that caution is needed when taking the 2010 EPI ranks at face value, at least for sixty of the countries included in the EPI.

5. Conclusions

The 2010 Environmental Performance Index, developed by the Yale and Columbia University distils key aspects of environmental performance in ten policy categories: Environmental burden of disease, Air pollution (effects on humans), Water (effects on humans), Air Pollution (effects on ecosystem), Water (effects on ecosystem), Biodiversity & Habitat, Forestry, Fisheries, Agriculture and Climate Change. These dimensions of environmental performance include a total of 25 indicators. As always when combining statistical indicators to capture a complex dimension, the EPI contains normative as well as analytic ingredients, in a mixture of that serves both analysis and advocacy addressed to 163 countries.

We subjected the 2010 EPI to thorough validity testing. We conducted an uncertainty analysis to assess the impact on the EPI ranking of simultaneous variations in the methodological assumptions related to the measurement error in the raw data, the structure of the indicators and the weights attached to them, the aggregation function at the policy level and the number of indicators (or policy categories) included in the framework. The effect proved to be acceptable for 109 countries (out of 164), but important for the remaining countries (e.g., Latvia, Luxembourg, Croatia, Spain, USA, Mexico, Mongolia, Qatar, and Haiti). Any Index-driven narrative on those countries should be considered only as contingent on the original methodological assumptions made in developing the Index.

Overall, the 2010 EPI gives a fair representation of the ensemble of models considered: the Spearman correlation between the 2010 Index ranking and the simulated median ranking is 0.99, whilst with the most extreme scenario (equal weights for all 25 indicators) is 0.86. These results suggest that the overall 2010 EPI results provide a reliable picture of the situation at global level and can be used to generate a discussion about what policies contribute to good environmental performance, to study the association between environmental performance and GDP, for example, and to provide insight into the nature of environmental policy challenges at the global scale. However, the country ranks, while reliable for the majority of the countries, for the remaining countries the EPI ranks need not be taken at face value as they are particularly sensitive to the methodological assumptions in the Index development.

Important findings from the analysis of the EPI results suggest that:

- The overall performance of the 163 countries is in general satisfactory in six of the ten policy categories. However, the remaining policy categories related to Air pollution (effects on ecosystem), Climate Change, Biodiversity & Habitat and DALY represent the main challenges for the majority of the countries: half of the countries hardly manage to achieve 50 to 60 points.
- Strong determinants of good environmental performance are, among others, (1) Environmental burden of disease (DALY); (2) Indoor air pollution; (3) Outdoor air pollution; (4) Access to water; and (5) Access to sanitation. Less influential but still significant on determining the 2010 EPI ranking are: the Water quality index, the Growing stock change, Forest cover change, Agricultural subsidies and Pesticide regulation.
- Other important environmental aspects, such as Non-methane volatile organic compound emissions, Critical habitat protection, Greenhouse gas emissions, and Industrial greenhouse gas emissions intensity, although they were included in the conceptual framework, they do not bear any statistically significant association to the EPI ranks. These results do not imply that keeping greenhouse gas emissions at low levels, and Critical habitat protection at high levels, should not be among the policy objectives of governments world wide. They simply point to the fact that even if governments made an effort to improve these aspects, the effort would not be captured by the EPI. The same comment holds for other indicators, such as Water stress Index, Biome protection, Marine protection, Marine trophic index, Agricultural water intensity.
- In order for a country to be ranked in the top fifty in the EPI, it must invest simultaneously in both Objectives of the EPI within a coherent environmental performance strategy, while emphasizing reduction of the existing gaps in areas where performance is lagging. However, this does not seem to be easy given the understandable − though problematic– trade-off between Environmental Health and Ecosystem Vitality. Hence, the EPI framework suggests that it is not easy to translate environmental sustainability-oriented performance into practice.

From the point of view of implications, the assessment carried out on the EPI does not represent merely a methodological or technical appendix. Composite measures are often attached to regulatory mechanisms whereby governments or organizations are rewarded or penalised according to the results of such measurements. The use and publication of composite measures can generate both positive and negative behavioural responses and if significant policy and practice decisions rest on the results, it is important to have a clear understanding of the potential risks involved in constructing a composite and arriving at a ranking or benchmarking.

The statistical analysis of the quality of the EPI shows that, although the theoretical framework and the indicators for the EPI were carefully chosen by experts, the issue of weighting is crucial to obtain a robust performance index. The current weighting and normalization schemes result 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 main assumptions tested in the uncertainty and sensitivity analysis, the country ranks are relatively reliable for 109 countries, while any conclusion on the ranks for the remaining countries should be made with great caution. An equal weighting approach or factor analysis-derived weights at the indicator level, as opposed to the current weighting scheme greatly influences the ranks. Thus, the choice of the weights must be evaluated according to the EPI's analytical rationale, policy relevance, and implied value judgments.

While an index such as EPI is intrinsically hard to compile, given the multidimensionality of the concept being measured, some improvement to the aggregation and normalization procedures are perhaps still possible and should be considered in the next version of the index. An effort should be made so that the weights of the policy categories and objectives do not deviate excessively from a measure of the relative importance of each on the final EPI rank.

References

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

Brand, D.A., Saisana, M., Rynn, L.A., Pennoni, F., Lowenfels, A. B., 2007. Comparative analysis of alcohol control policies in 30 countries. PLoS Medicine, 4(4), 752-759.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7(1), 1–26.

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

Grupp, H., Mogee, M.E., 2004. Indicators for national science and technology policy: how robust are composite indicators? Research Policy 33, 1373-1384.

Kennedy P., 2007, A Guide to Econometrics, Fifth ed. Blackwell.

Munda, G., 2008. Social Multi-criteria Evaluation for a Sustainable Economy, Springer, Berlin.

Nature News, 2007. Academics strike back at spurious rankings, Nature, 447, 31 May 2007, 514-515.

Nicoletti, G., Scarpetta, S., Boylaud, O., 2000. Summary indicators of product market regulation with an extension to employment protection legislation, OECD, Economics department working papers No. 226, ECO/WKP(99)18.

OECD, 2008, Handbook on Constructing Composite Indicators. Methodology and user Guide, OECD Publishing, Paris.

Saisana M., 2008, The 2007 Composite Learning Index: Robustness Issues and Critical Assessment, Report 23274, European Commission, JRC-IPSC, Italy.

Saisana M., Munda G., 2008, Knowledge Economy: measures and drivers, Report 23486, European Commission, JRC-IPSC.

Saisana M., Saltelli A., 2008a, Expert Panel Opinion and Global Sensitivity Analysis for Composite Indicators, Chapter 11 in Computational Methods in Transport: Verification and Validation, Vol. 62, ISSN 1439-7358, Ed. Frank Graziani, Springer Berlin Heidelberg, pp.251-275.

Saisana M., Saltelli, A., 2008b, Sensitivity Analysis for the 2008 Environmental Performance Index, Report 23485, European Commission, JRC-IPSC.

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

Sharpe A. (2004), Literature Review of Frameworks for Macro-indicators, Centre for the Study of Living Standards, Ottawa, CAN.

Stiglitz, J.E., Sen, A., Fitoussi JP, 2009, Report by the Commission on the Measurement of Economic Performance and Social Progress, www.stiglitz-sen-fitoussi.fr.

European Commission

EUR 24269 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen
Title: Uncertainty and Sensitivity Analysis of the 2010 Environmental Performance Index
Author(s): Michaela Saisana and Andrea Saltelli
Luxembourg: Office for Official Publications of the European Communities
2010 – 27 pp. – 21 x 29.70 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-15071-5
DOI 10.2788/67623

Abstract

An assessment of the robustness of the 2010 Environmental Performance Index (EPI) ranks requires the evaluation of uncertainties underlying the index and the sensitivity of the country rankings to the methodological choices made during the development of the Index. To test this robustness, the Yale and Columbia University have continued their partnership with the Joint Research Centre (JRC) of the European Commission in Ispra, Italy.

This JRC report shows that although the theoretical framework and the indicators for the EPI were carefully chosen by experts, the issue of weighting is crucial to obtain a robust performance index. The current weighting and normalization schemes result 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 main assumptions tested in the uncertainty and sensitivity analysis, the country ranks are relatively reliable for 109 countries, while any conclusion on the ranks for the remaining countries should be made with great caution. An equal weighting approach or factor analysis-derived weights at the indicator level, as opposed to the current weighting scheme greatly influences the ranks. Thus, the choice of the weights must be evaluated according to the EPI's analytical rationale, policy relevance, and implied value judgments. If the objective of EPI is to promote action on all policies categories more work would be needed to ensure that all policy fields have an impact on the aggregated EPI or, alternatively, policy categories should be given more emphasis than the aggregated measure.

The 2010 EPI is developed for 163 countries and is based on twenty five indicators grouped in ten policy categories: Environmental burden of disease, Air pollution (effects on humans), Water (effects on humans), Air Pollution (effects on ecosystem), Water (effects on ecosystem), Biodiversity & Habitat, Forestry, Fisheries, Agriculture and Climate Change.

The EPI 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 by the JRC-IPSC are:

- measurement error of the raw data,
- EPI structure – grouping at policy categories,
- weights assigned to the indicators and/or to the policy categories,
- aggregation function at the policy or at the objectives level, and
- number of indicators or 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-24269-EN-C

JRC
EUROPEAN COMMISSION

ISBN 978-92-79-15071-5

9 789279 150715

Publications Office
Publications.eu.int