2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

An Investigation of the Environmental Sustainability Index in Terms of its Prediction and Clustering Capabilities

Tatiana Tambouratzis

Abstract—**During the years 1999-2005, the environmental sustainability index (ESI) constituted a predominant tool for evaluating, ranking, and grouping countries in terms of their current and future potential to protect the environment. In this piece of research, an investigation of the calculation/prediction, ranking, and clustering capabilities of the ESI 2005 is performed using traditional as well as computational intelligence tools, the latter including supervised general regression artificial neural networks, probabilistic artificial neural networks, unsupervised self-organizing maps, and fuzzy clustering. The results of the investigation shed some light on the derivation of the ESI, but further research is required for elucidating – and, thus, being able to replicate - the ESI values and clusters.**

I. INTRODUCTION

THE compilation of parameters describing "how the environment is fairing" at the regional or country level [1], has proliferated in the last 15 years. For ease of expressing and handling the resulting data, the collected parameters are usually combined and, subsequently, expressed as a single cumulative index. Such a numerical expression of environmental sustainability promotes the evaluation, ranking, and grouping of countries (as well as territories) in terms of environmental protection and related issues.

Depending on the collected parameters, as well as on the interest/focus of the collecting body, a variety of indices have appeared in the relative literature, some of the most well-known being the environmental vulnerability index (EVI) [2], the environmental sustainability index (ESI) [3], the environmental performance index (EPI) [4], the dashboard of sustainability (DS) [5-6], the wellbeing index (WI) [7], the national footprints account (ecological footprint and biocapacity) [8], the living planet index (LPI) [9], the human development index (HDI) [10] etc.

In this piece of research, the ESI 2005 [3,11] is investigated in terms of its prediction, ranking, and clustering capabilities. The test-bed for this investigation is the total of 146 countries and territories (shown in Table I) used for constructing the ESI 2005. Taking into account the information contained in [11], the following question is put forward: can the ESI value/cluster of any one or more of the 146 countries of Table I be accurately predicted (i.e. in agreement with [11]) directly from the ESI values/clusters of the other countries of the Table? Further to the straightforward purpose of the validation of the ESI 2005, such knowledge is especially constructive as it establishes whether it is also possible to:

(a) calculate the ESI 2005 of a country of interest that is not included in Table I (provided, of course, that the necessary data of the country are available) in a consistent manner relative to the ESI values of the 146

TABLE I
THE COUNTRIES USED FOR THE EVALUATION OF THE ESI 2005 [11]

Albania	Dominican Republic	Latvia	Russia
Algeria	Ecuador	Lebanon	Rwanda
Angola	Egypt	Liberia	Saudi Arabia
Argentina	El Salvador	Libya	Senegal
Armenia	Estonia	Lithuania	Serbia & Montenegro
Australia	Ethiopia	Macedonia	Sierra Leone
Austria	Finland	Madagascar	Slovakia
Azerbaijan	France	Malawi	Slovenia
Bangladesh	Gabon	Malaysia	South Africa
Belarus	Gambia	Mali	South Korea
Belgium	Georgia	Mauritania	Spain
Benin	Germany	Mexico	Sri Lanka
Bhutan	Ghana	Moldova	Sudan
Bolivia	Greece	Mongolia	Sweden
Bosnia & Herzegovina	Guatemala	Morocco	Switzerland
Botswana	Guinea	Mozambique	Syria
Brazil	Guinea-Bissau	Myanmar	Taiwan
Bulgaria	Guyana	Namibia	Tajikistan
Burkina Faso	Haiti	Nepal	Tanzania
Burundi	Honduras	Netherlands	Thailand
Cambodia	Hungary	New Zealand	Turkey
Cameroon	Iceland	Nicaragua	Turkmenistan
Canada	India	Niger	Uganda
Central African Republic	Indonesia	Nigeria	Ukraine
Chad	Iran	North Korea	United Arab Emirates.
Chile	Iraq	Norway	United Kingdom
China	Ireland	Oman	United States
Colombia	Israel	Pakistan	Uruguay
Congo	Italy	Panama	Uzbekistan
Costa Rica	Jamaica	Papua New Guinea	Venezuela
Côte d'Ivoire	Japan	Paraguay	Viet Nam
Croatia	Jordan	Peru	Yemen
Cuba	Kazakhstan	Philippines	Zambia
Czech Rep.	Kenya	Poland	Zimbabwe
Democratic Republic of Congo	Kuwait	Portugal	
Denmark	Kyrgyzstan	Romania	

T. Tambouratzis is with the Department of Industrial Management & Technology, University of Piraeus, 107 Deligiorgi St, Piraeus 185 34, Attiki, Greece (phone: +30-210-4142423; fax: +30-210-4142392; e-mail: tatianatambouratzis@gmail.com).

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

countries, and – thus – be able to

(b) rank the country of interest against the 146 countries, and

(c) group the country of interest with (i.e. assign the country to) the ESI 2005 cluster of countries having the most similar environmental and sustainability characteristics.

If these actions are possible, on the one hand valuable information can be gleaned concerning the present and future sustainability of a country of interest, while – on the other hand – the consistency of the ESI 2005 is established. The answers to the aforementioned questions have been evaluated here using 10-fold cross-validation [12].

The following text is organized as follows. Section II introduces the ESI 2005, detailing the characteristics of the collected variables, and their hierarchical organization into indicators, components, and ESI values, as these have been implemented in [11]; section III presents the prediction, classification and clustering tools employed here, subsequently describing and discussing the results of their application to the 146 countries used in [11]; section IV critically discusses the obtained results, and puts forwards some points for further investigation; finally, section V concludes the paper.

II. The ESI 2005– Inspiration And Construction

A. ESI Timeline

The ESI constitutes the result of the Environmental Performance Measurement (EPM) project, a collaboration between (a) the Yale Center for Environmental Law and Policy of Yale University, (b) the Columbia Center for International Earth Science Information Network of Columbia University, and (c) the Joint Research Centre of the European Commission, in an attempt to measure the ability of the various countries to reach and maintain environmental sustainability. The ESI was calculated, compiled, and published in 2000, 2001, 2002, and 2005, at which time it was substituted by the EPI which uses indicators that can be handled in a more straightforward manner by policy makers, environmental scientists, and the general public [4]; the EPI has been published biennially since 2006.

Due to differences in the number of countries, components, indicators, and/or variables collected each time, the ESI values compiled during the different years of publication are not fully compatible. The ESI from 2005 [11] has been employed for the present investigation as it constitutes the most complete effort (at least compared to the previous ESI versions) to evaluate, rank, and group the different countries in terms of sustainability.

B. ESI Structure and Evaluation

The ESI is constructed in a hierarchical manner. As shown in Table II, the ESI index that is used for evaluating the sustainability status of a country or territory is built up of five components (first column of Table II) which express the ability of a country to maintain – and possibly improve – its vital environmental systems, while also limiting the levels of anthropogenic stress, being resilient to environmental disturbances, responding effectively to environmental challenges, and – finally – being willing as well as able to cooperate with other (especially neighbouring) countries towards the solution of common environmental problems.

The five components are, in turn, constructed from 21 indicators (second column of Table II), with each component combining between three and six component-specific indicators; as mentioned in [11], all the indicators are equally-weighted for the formation of each component.

Each indicator is built up from a number of variables (between two and twelve) which are dedicated to accurately expressing and measuring the indicator. A total 76 variables (whose distribution among indicators is shown on the rightmost column of Table II) are employed for fully conveying the 21 indicators, with equal weights being assigned to all the variables dedicated to the same indicator (p. 13 & 66, [11]); the interested reader is referred to Table 10 (pp. 14-15 [11]) for further details on the variables.

TABLE II
THE HIERARCHICAL CONSTRUCTION OF THE ESI 2005
IN TERMS OF FIVE COMPONENTS, 21 INDICATORS AND 76 VARIABLES

Components	Indicators	Variables
Environmental Systems	Air Quality	four
	Biodiversity	five
	Land	two
	Water Quality	four
	Water Quantity	two
Reducing Environmental Stresses	Air Pollution	five
	Ecosystem Stresses	two
	Population Pressure	two
	Waste and Consumption	three
	Water Stress	four
	Managing Natural Resources	five
Reducing Human Vulnerability to Env. Stresses	Environmental Health	three
	Basic Human Sustainance	two
	Reducing Natural Disaster Vulnerability	two
Societal and Ins- titutional Capacity to Respond to En- v. Challenges	Environmental Governance	twelve
	Eco-Efficiency	two
	Private Sector Responsiveness	five
	Science and Technology	five
Global Stewardship	International Collaborative Efforts	three
	Greenhouse Gas Emissions	two
	Reducing Transboundary Env. Pressures	two

The raw data itself is transformed in a variety of ways, including scaling, extreme value removal, winsoring (trimming of the distribution tails), and other numerical as well as statistical conversions. As stated in [11], raw data processing is implemented in such a manner as to promote not only the evaluation of the 146 countries in terms of sustainability, but to further provide guidelines for their improvement in that respect: “Owing to the multi-faceted and hierarchically organized nature of the collected information, the ESI is capable not only of assessing the 146 countries in terms of sustainability, but of also offering the guidelines for further increasing sustainability, despite the incomplete, approximate, and sometimes even conflicting nature of the raw data" [11].

It is stated that the 21 indicators constitute the basis for expressing environmental sustainability, and thus for evaluating the ESI values of the 146 countries (p. 64 & 66, [11]); as already mentioned, equal weights have been employed to the 21 indicators for producing the 146 ESI values (Figure 1 and text in p. 13, [11]). Uncertainty and sensitivity analyses have verified the general stability of the ESI values, as well as of the ranking of the 146 countries, even in the case where expert-derived – rather than equal – weights have been assigned to the 21 indicators for evaluating the ESI (p. 38, [11]). As also suggested in the same document (p. 24, [11]), the overall performance of a country can be best understood by looking not only at the ESI value or ranking, but by also investigating the 21 indicators. In a consistent fashion, is also the indicators that have also been employed for clustering the ESI values and countries (p. 94, and pp. 97-98, [11]).

C. ESI Merit/Significance

The ESI value of a country or territory is used for benchmarking national environmental stewardship. In other words, not only is a country with a high ESI value considered as having achieved a high level of environmental sustainability, which it is likely to maintain in the future, but the ESI value also quantifies the likelihood that a country shall – at least – preserve its environmental resources, and avoid environmental deterioration over a period of several decades.

Further to providing a measure of environmental sustainability of the 146 countries of Table I, the ESI values have further been used for:

- Comparing each country with the other countries in terms of ESI values.
- Ranking the 146 countries in terms of environmental sustainability (Table 11, page 22, [11]).
- Grouping the 146 countries into seven sets of countries via statistical cluster analysis. As mentioned in [11], seven clusters have been selected as optimal since "We can see these clusters as having observable similarities and thus representing a useful point of departure for policy comparisons" (p. 29, [11]). As further detailed, "...the clustering is optimal for the countries of highest and lowest ESI values, but not clear for the middle ones". By having "countries in the same cluster characterized by similar system scores, stress scores, vulnerability, capacity, and stewardship" (Table 14, page 30, [11]), but by never having a country that is superior to all the other countries of the same cluster in terms of component and indicator values, sustainability is further promoted as follows: each country is provided with guidelines and examples (derived from the other countries in the same cluster that are superior in terms of one or more components, indicators, or variables) for improving the indicator as well as component values, and thus increasing its environmental sustainability.

The interested reader can also find a comparison of the ESI 2005 and other sustainability indicators in [11] (Appendix F, pp. 383-390). Finally, all of the techniques employed in the following sections have been implemented using MATLAB 2009a [13].

III. The Prediction and Clustering Potential of ESI 2005

A. ESI Prediction

As already mentioned in Section I, it is of special interest to establish the means of deriving the ESI values of the 146 countries employed for creating the ESI 2005.

The methodology employed in [11] for producing the ESI values of the 146 countries constitutes a first-degree polynomial with equal-valued coefficients for the 21 polynomial terms (i.e. the terms corresponding to the 21 indicators), excluding the constant term: "We settled on uniform weighting of the 21 indicators because simple aggregation is transparent and easy to understand. Moreover, when we asked leading experts from the governmental, business, and non-governmental sectors to rank the indicators, none stood out as being of substantially higher or lower importance than the others. Similarly, when we tried to use statistical methods (including principal component analysis) to identify appropriate weights, nearly equal values were suggested across all 21 indicators". It is also mentioned in [11] that the same ESI values are produced even when the restriction of equal coefficients is relaxed.

In an effort to duplicate the polynomials as well as the methodology followed in [11], the relationship between ESI 2005 indicators and ESI 2005 values has been reconstructed using polynomial approximation [14]. In this case, all 146 countries have been used for creating the two polynomials, i.e. 10-fold cross-validation has not been implemented. Two polynomials have been created, the first polynomial simply aiming at maximal prediction accuracy, the second polynomial further enforcing the use of equal coefficients for the 21 polynomial terms; the results of the two polynomials are shown in Table III.

TABLE III
POLYNOMIAL APPROXIMATION ACCURACY OF THE ESI 2005

errors	mean	min	max	std
optimal polynomial	0.19	7×10^{-4}	0.92	0.16
polynomial with equal coefficients	0.22	20×10^{-4}	1.25	0.19

Prediction accuracy has been found most satisfactory, with the use of equal coefficients for the 21 polynomial terms being found slightly impaired when compared to the polynomial approximation with no constraints applied to the coefficient values.

However, and despite the seemingly insignificant errors shown in Table III, when ranking the 146 countries using (a) the polynomial approximation of the ESI values with no constraints on the coefficient values, and (b) the polynomial approximation of the ESI values with equal coefficient values, 44 and 41, respectively,[1] changes in country rank

[1] i.e. 30 and 28% of the 146 countries.

(position) have been observed when compared with the ranks assigned to the 146 counties using the ESI 2005 values given in [11].

In an attempt to clarify this issue, the four-layer general regression artificial neural network (GRNN) [15] has also been implemented for predicting the ESI 2005 values of the 146 countries. The GRNN implements a free-form (non-parametric) regression, and is composed of four layers, namely: the input layer with as many nodes as there are input dimensions; the second layer which encodes one training pattern per node; the third layer which employs two nodes for evaluating the similarity between the input pattern and each node of the second layer; the fourth layer which comprises as many nodes as there are output dimensions, combines the responses of the nodes of the third layer, and produces the final prediction. A single presentation of the training patterns is sufficient for setting up the GRNN, while the single trainable parameter, the spread (σ), is used for optimizing the degree of interpolation between the training patterns for the evaluation of the output: an appropriate value of σ allows the regression surface to approximate the Bayes optimal.

For ESI value prediction, the GRNN has been employed in three distinct ways, namely:

(a) directly from the 21 indicators to the ESI values (single GRNN), thus following the manner of deriving the ESI described in [11],

(b) directly from the five components to the ESI values (single GRNN), and

(c) indirectly, i.e. by predicting the five components from the 21 indicators using the GRNN developed in (a), and subsequently predicting the ESI from the five predicted components using the GRNN developed in (b).

Additionally, two kinds of data scaling have been implemented, namely:

(i) scaling every input dimension, as well as every output dimension, independently of each other in the range [0.1 0.9] (scaled data), and

(ii) uniform scaling all the input dimensions simultaneously, and all the output dimensions simultaneously (uniscaled data).

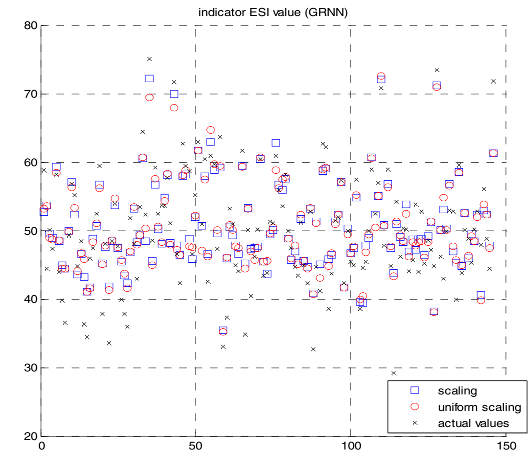
Distinct parameter values have been found optimal for the different means of ESI 2005 prediction as well as for the two kinds of scaling, resulting in

- six GRNNs using the totality of countries for training and testing, and
- six sets of 10 GRNNs, trained and tested using 10-fold cross-validation (one GRNN for each fold).

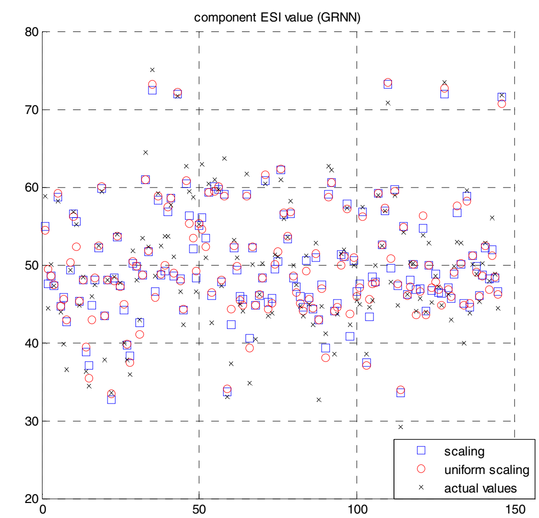
In all cases, the optimal value of σ has been determined within the interval [0.01 0.99] with steps of 0.01.

When training and testing on the 146 countries, prediction accuracy reached – as expected – 100% for a wide range of values of σ for all cases (a) to (c), and for both kinds of data scaling. However, the optimal values of σ during 10-fold cross-validation were found to be quite specific (as shown in Table IV), again showing perfect recall of the training sets, but producing a significant drop in prediction accuracy for the 10 test sets; for case (c), the optimal values derived from (a) and (b) have been employed directly for training and testing.

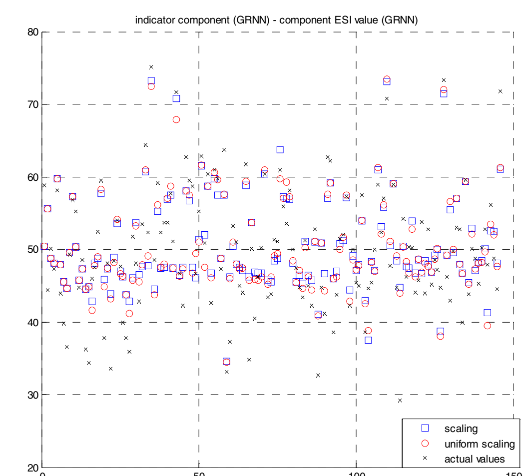
The results of GRNN prediction accuracy that appear in



(a)



(b)



(c)

Fig. 1. ESI value prediction via GRNNs (a) directly from the 21 indicators, (b) directly from the five components; (c) indirectly.

Table IV are restricted to the test sets of the 10 folds. No superiority of either form of scaling is apparent, although an advantage of using small σ values (slightly smaller for uniscaled rather than scaled data) is observed. Additionally, indirect prediction of the ESI value is found impaired relative to the two ways of direct ESI prediction, which is far from surprising given the accumulation of prediction errors that carry on from the first to the second GRNN.

TABLE IV
GRNN TEST RESULTS FOR PREDICTING THE ESI VALUE VIA 10-FOLD CROSS-VALIDATION

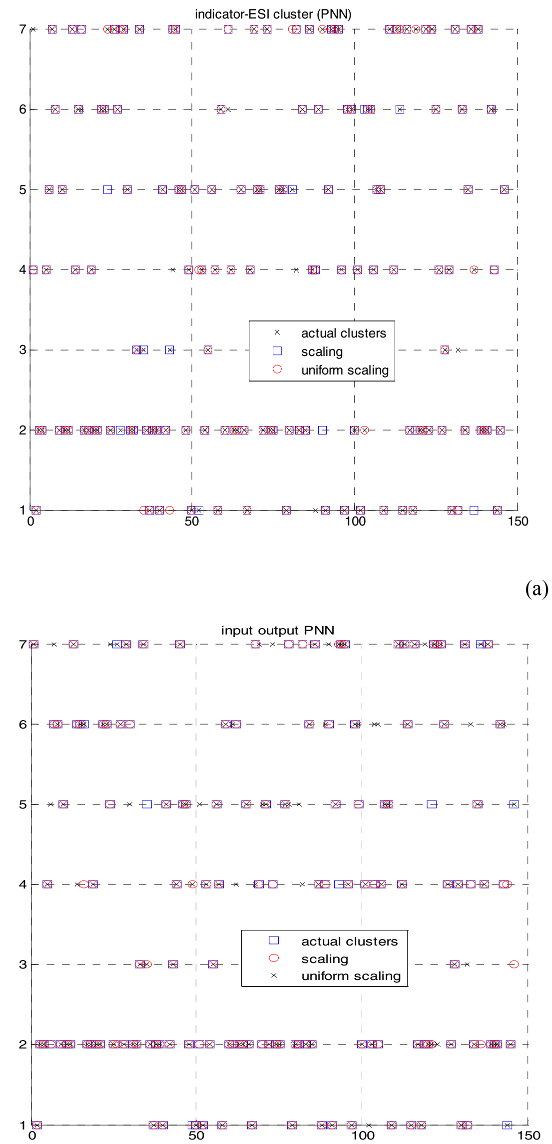
errors	mean	min	max	std
from 21 indicators, scaled (σ=0.2)	3.16	0.03	14.56	2.62
from 21 indicators, uniscaled (σ=0.12)	3.16	0.02	14.08	2.66
from five components, scaled (σ=0.1)	1.82	0.01	10.30	1.75
from five components, uniscaled (σ=0.07)	1.90	0.02	10.29	1.80
indirectly, scaled (σ=0.21)	8.73	0.13	29.11	6.36
indirectly, uniscaled (σ=0.12)	8.82	0.02	28.40	6.41

Further investigation in terms of actual ESI values and predicted values by the GRNNs, for cases (a) and (b) only, and restricted to the results of 10-fold cross-validation, shows that prediction is consistently more accurate when the five components are employed than when the 21 indicators are used, a finding that holds true under all the error measures shown in Table IV. Another interesting point is that the prediction of low ESI values is significantly impaired relative to that of large and medium ESI values. This is clearly shown in Figure 1, where it can also seen that – in these cases – the GRNN employing the components as inputs is clearly superior to the one using the indicators. It should be mentioned at this point that the observed inaccuracy cannot be attributed to extrapolation errors, since tests removing the extreme patterns did not significantly improve overall prediction accuracy in any of the three cases.

The aforementioned findings are contrary to what is supported in [11], namely that the ESI values of the 146 countries are derived from the 21 indicators, rather than the five components: "We consider the 21 indicators to be the fundamental building blocks of environmental sustainability – and it is these 21 indicators that are aggregated to create the ESI" [11]. From the results, it can be derived that using the components rather than the indicators allows the more accurate prediction/evaluation of the ESI values, especially for countries demonstrating low levels of sustainability. Still, the inaccuracy of ESI value prediction, especially when contrasted with the perfect recall of the training patterns, points towards some kind of transformation of the data (especially for deriving the indicators) that occlude the relationship between GRNN inputs (indicators) and outputs (ESI values).

B. ESI Classification - Cluster Prediction

It is also of interest to determine whether the cluster to which a given country belongs can be accurately predicted according to the ESI-derived clusters that appear in [11].



(b)

Fig. 2. ESI classification into the seven ESI clusters via PRNNs (a) directly from the 21 indicators, (b) directly from the five components.

The four-layer PNN [16] has been employed to this end, an artificial neural network architecture that is very similar to the GRNN in terms of structure as well as training, and – in fact – constitutes the classification counterpart of the GRNN. For uniformity and ease of comparison, the same folds have been employed for performing 10-fold cross-validation. Only clustering directly from the 21 indicators and from the five components has been performed, i.e. indirect clustering using two PNNs has not been investigated further. Both scaling and uniform scaling of the indicators and components have been tested.

Prediction of the ESI clusters to which the 146 countries has been found quite satisfactory – though by no means as precise as would be expected – , with 102 countries being clustered correctly by all four PNNs; Figure 2 demonstrates the ability of the PNN to correctly assign the majority of countries to their ESI cluster for both kinds of scaling. Furthermore, as shown in Table V, the values of the σ parameter remain quite low, with those used for uniformly scaled inputs being smaller than those used for scaled inputs, and scaling independently per input dimension found superior to uniform scaling.

TABLE V
PNN TEST RESULTS FOR PREDICTING THE ESI CLUSTER

Accuracy (%)	Directly from 21 indicators		Directly from five components	
	scaled	uniscaled	scaled	uniscaled
σ	0.25	0.19	0.15	0.11
correct	91.10	79.45	90.41	78.77

Different misclassifications are observed by the two kinds of scaling, differences that are – in fact – more accentuated than the differences observed by the use of indicators or of components for the prediction of the ESI cluster. Overall, scaling independently per dimension is preferable to uniform scaling.

The assumption that the indicators are employed for clustering appears valid; however the use of components might still constitute an acceptable implementation given its reduced computational complexity. As shown in Table VI, no direct effect of the size of the cluster – as proposed in [11] - on the number of misclassifications is apparent.

TABLE VI
PNN MISCLASSIFICATIONS PER CLUSTER

Cluster (number of countries)	Directly from 21 indicators		Directly from five components	
	scaled	uniscaled	scaled	uniscaled
1 (17)	4	2	3	4
2 (41)	1	2	1	1
3 (8)	1	2	3	1
4 (18)	3	7	2	6
5 (19)	0	5	1	5
6 (19)	3	5	4	7
7(24)	4	6	3	5

It appears that countries from clusters 3, 4, 6, and 7 are misclassified (i.e. assigned to a different cluster than the one stated in [11]) more often than countries belonging to other clusters. This is in partial accordance with the findings of Section *II.C* and validity of the different clusters (mentioned at the end of Section *II.C*), as clusters 6 and 7 contain the countries of the lowest ESI values.

As a final observation, the – at best – 10% misclassification rate when grouping the 146 countries in their cluster according to [11] is still quite large, again hinting towards a different treatment of the data during clustering, or the use of somehow transformed data submitted to clustering.

C. ESI Clustering of Countries

Given that the classification of the countries to their preferred ESI cluster exceeds 90% for the scaled ESI 2005 data, both when using the 21 indicators and the five components, it becomes necessary to go one step back and also determine how consistent the original clustering of the 146 countries into seven clusters (as proposed in [11]) is, and – thus – be able to gleam valuable information concerning how the seven ESI 2005 clusters of the 146 countries have been derived.

To this end, the 146 countries have been clustered not only in terms of their ESI values, but also in terms of their indicators and components, as a means of establishing the actual criterion of grouping countries together.

Three distinct methodologies, namely the self-organizing map (SOM) [17], fuzzy c-means clustering (FCM) [18], and k-means clustering (KMC) [19], have been used to this end, each of them independently clustering the indicators, the components, and the ESI values so as to implement their methodology-derived interpretation of seven-group clustering.

The SOM operates by mapping the similarities of high-dimensional patterns into one- or two-dimensional arrangements of points (maps) in an unsupervised manner, while still preserving the topological properties of the original pattern space. The 1-D SOMs employed here have seven nodes, i.e. as many nodes as there are clusters. Although running along a single dimension, they have been found adequately flexible to cover the problem space not only of one, but also of 21 and 5 dimensions, respectively[2].

No such restrictions apply for FCM and KMC. The FCM assigns each pattern to each available cluster with a degree of belonging (rather than a binary "belong" to a single cluster and "do-not-belong" to the other clusters), thus accommodating patterns that do not belong entirely to a single cluster, but share similarities with more clusters. Finally, KMC creates – and iteratively adjusts – a Voronoi tessellation of the pattern space according to the number of desired clusters, such that each pattern belongs to the cluster of its nearest centroid.

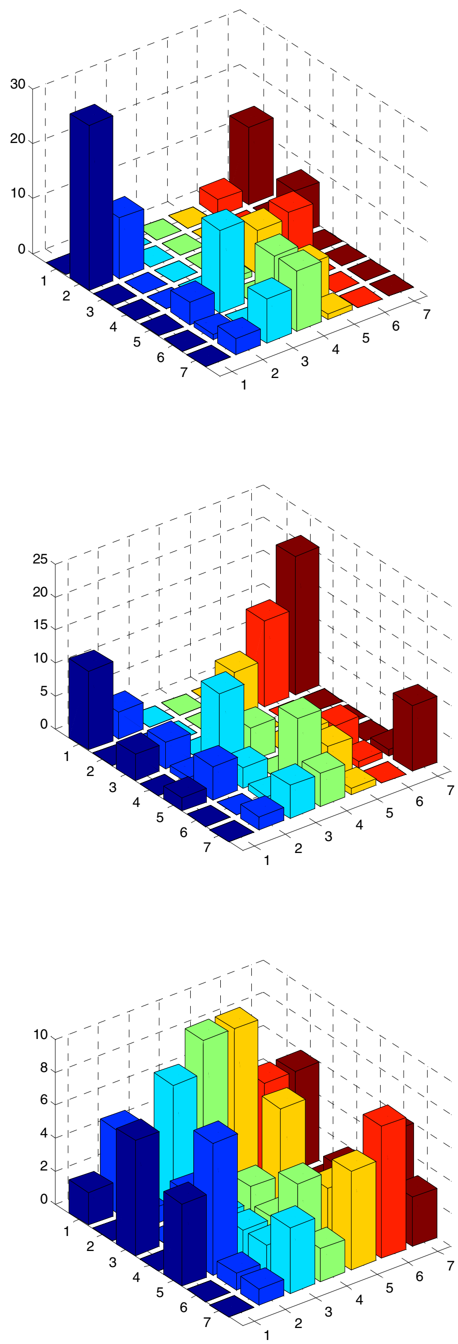
TABLE VII
TOTAL NUMBER OF MATCHES BETWEEN CLUSTERS

Methodology	from 21 indicators	from five components	from ESI values
SOM	75	65	42
FCM	68	70	52
KMC	88	70	46

Following training, the seven (i.e. as many as there are ESI clusters) pairwise intersections between the countries assigned to each SOM/FCM/KMC-derived cluster and the countries actually belonging to each ESI cluster (such that each combination of derived and ESI cluster is used exactly once), have been selected and used to represent the compatibility between clusters. Table VII illustrates the number of matches between the clusters created by each

[2] A 2x4 SOM was also investigated, but found to perform slightly worse than the SOM described here.

technique and the clustering put forward in [11], expressed independently for indicators, components and ESI values.



(a)

(b)

(c)

Fig. 3. Examples of cluster agreement between SOM-based clustering of the five components (a), FCM of the 21 indicators (b), and KMC of the ESI values (c), and the seven ESI-derived clusters. The optimal agreement is the one that contains the most countries when selecting one bar from each row and each column (total of seven bars selected); ideally, this corresponds to a bar chart seven high bars that are located at distinct rows and columns (i.e. no two high bars share their row or column).

The results are far from satisfactory, especially since at best around half of the 146 countries appear to be clustered in accordance with the clustering put forward in [11]. In terms of combinations of data and clustering techniques, the best results are obtained when using the indicators as the basis of clustering for the SOM and KMC, and when using the five components as the basis of clustering for the FCM; the worst results are always observed when using the ESI values. It should further be mentioned that clustering using the five components produces the most consistent (though not necessarily best) results, a finding that may be of interest in terms of clustering.

Figure 3 shows examples of the pairwise intersections between (a) a SOM-derived clustering of the components and the ESI clusters, (b) a FCM-derived clustering of the indicators and the ESI clusters, and (c) a KMC-derived clustering of the ESI values and the ESI clusters. It is noted that the optimal results would require a single raised bar per row as well as per column.

IV. Conclusion

An initial investigation of the hierarchical construction of the ESI 2005 via indicators and components has been performed using the ESI values and ESI clusters described in [11]. Some points of discussion – and areas of further research – remain concerning the derivation of the ESI values as well as clusters, which – at present – remains elusive. It appears that some kind of transformation of the data (especially concerning the indicators) does not permit the precise replication of the relationship described in [11] between the ESI 2005 indicators/components and the ESI values/rankings/clusters.

References

[1] A. de Sherbinin, A. Reuben, M.A. Levy, and L. Johnson, *Indicators in Practice: How Environmental Indicators are Being Used in Policy and Management Contexts*, http://ciesin.columbia.edu/binaries/web/global/news/2013/indicatorsinpractice.pdf

[2] www.sopac.org/index.php/environmental-vulnerability-index.

[3] D.C. Esty, M. Levy, T. Srebotnjak, and A. de Sherbinin *2005 Environmental Sustainability Index: Benchmarking National Environmental Stewardship*, New Haven: Yale Center for Environmental Law & Policy, 2005

[4] *Measuring Progress: a practical guide from the developers of the Environmental Performance Index (EPI)*, www.epi.yale.edu/dowloads

[5] http://www.iisd/dashboard.asp

[6] http://esl.jrc.ec.europa.eu/dashbrds.htm

[7] www.gallup.com/poll/123215/gallup-helathways-index.aspx

[8] http://www.footprintnetwork.org/en/index.php/GFN/footprint-for-nations/

[9] http://wwf.panda.org/about_our_earth_/all_publications/living_planet_report/

[10] http://hdr.undp.org/en/statistics/hdi

[11] http://www.yale.edu/esi/ESI2005_Main_report.pdf

[12] P. A. Devijver and J. Kittler, *Pattern Recognition: a Statistical Approach*, Prentice-Hall, London, 1982

[13] The MathWorks 2009. R2009a MatLab & Simulink.

[14] C.R. Rao, H. Toutenburg, A. Fieger, C. Heumann, T. Nittner and S. Scheid, *"Linear Models: Least Squares and Alternatives (2nd Edition)"*, Springer-Verlag Series in Statistics, New York, U.S.A., 1999

[15] D.F. Specht, "*A General Regression Neural Network*". IEEE Transactions on Neural Networks, vol. 2, pp. 568-576, 1991

[16] D.F. Specht, "*Probabilistic Neural Networks and the Polynomial Adaline as Complementary Techniques to Classification*". IEEE Transactions on Neural Networks, vol. 1, pp. 111-121, 1990

[17] T. Kohonen, *"Self-Organized Formation of Topologically Correct Feature Maps"*. Biological Cybernetics, vol. 43, no. 1, pp. 59-69, 1982

[18] J.C. Bezdec, *"Pattern Recognition with Fuzzy Objective Function Algorithms"*, Plenum Press, New York, 1981

[19] J.B. MaxQueen, *"Some Methods for Classification and Analysis of Multivariate Observations",* Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281-197, University of California Press, 1967