Computational Statistics and Data Analysis 53 (2009) 3516–3531

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A neuro-computational intelligence analysis of the ecological footprint of nations

Mohamed M. Mostafa *, Rajan Nataraajan

Auburn University, College of Business, 415 West Magnolia Avenue, Auburn, AL 36849, USA

ARTICLE INFO

Article history:
Received 19 November 2007
Received in revised form 3 March 2009
Accepted 4 March 2009
Available online 13 March 2009

ABSTRACT

The per capita ecological footprint (EF) is one of the most-widely recognized measures of environmental sustainability. It seeks to quantify the Earth's biological capacity required to support human activity. This study uses three neuro-computational methodologies: multi-layer perceptron neural network (MLP), probabilistic neural network (PNN) and generalized regression neural network (GRNN) to predict and classify the EF of 140 nations. Accuracy indices are used to assess the prediction and classification accuracy of the three methodologies. The study shows that neuro-computational models outperform traditional statistical techniques such as regression analysis and discriminant analysis in predicting and classifying per capita EF due to their robustness and flexibility of modeling algorithms.

1. Introduction

The EF approach was developed by Wackernagel and Rees (1996). It is measured as the total area of productive land and water required to continuously produce all resources consumed and to assimilate all wastes generated by a defined population in a specific location. Since its development, the EF has become "the most widely-used measure of environmental sustainability" (Binningsbo et al., 2007, p. 337). The usefulness of the EF is that it aggregates typically complex resource use patterns into a single number (Constanza, 2000). According to Dauvergne (2005), the EF is an innovative way to compare the ecological impact of nations across the globe. The validity of the per capita EF is also empirically grounded as it was found to be significantly correlated with key environmental impacts, such as national emissions of ozone depleting substances (Prescott-Allen, 2001) and nuclear power generation (WRI, 2000)

The EF includes six different resources: crop land and pasture land for production of food and goods, built-up land to support infrastructure, forest for the production of wood products, fish for food production, and carbon assimilating capacity for carbon dioxide emissions from fossil fuels. Both land and bio-capacity are measured in global hectares (gha). A global hectare represents a hectare of land with world average bio-productivity. It is estimated that the EF of the global population is at least 30% larger than the Earth's bio-capacity (McDonald and Patterson, 2004). In 2003, the global per capita EF was 2.2 gha, while the global bio-capacity was 1.8 gha/cap. The EF of nations ranged from 0.5 gha/cap in Bangladesh to around 10 gha/cap in the U.S. (White, 2007). These figures can be used as a benchmark for assessing the sustainability of nations. Nations with EF at or below 1.8 gha/cap have a global impact that could be replicated by other nations without threatening long-term sustainability.

Although the EF approach has been applied at various levels, including global (e.g. Rice (2007)), national (e.g. Van Vuuren and Smeets (2000)), municipal/institutional (e.g. Barrett and Scott (2003)), and individual levels (e.g. Crompton et al. (2002)), no previous studies have attempted to use neuro-computational techniques to predict and classify the EF of nations. In this

* Corresponding author.
E-mail address: mmm0010@auburn.edu (M.M. Mostafa).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2009.03.003

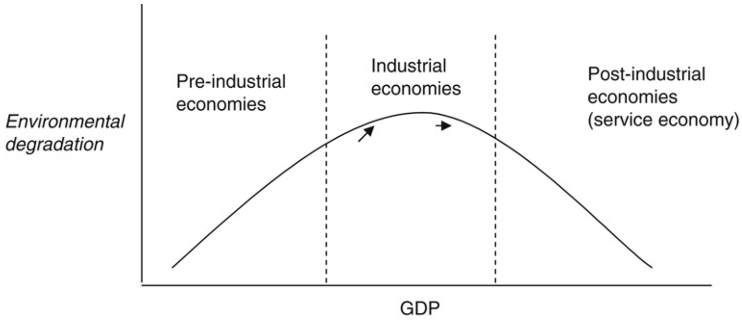

Fig. 1. The environmental Kuznets curve.

research we aim to fill this research gap by predicting and classifying the EF of 140 nations through the use of intelligent modeling techniques. More specifically, the purpose of this research is twofold:

- To determine the major factors that affect the EF of nations; and
- To benchmark the performance of neuro-computational models against traditional statistical techniques.

This paper is organized as follows. The next section summarizes the EF literature and develops the research hypotheses. The methodology used to conduct the analysis follows. The subsequent section presents empirical results of the analysis. Next, the paper sets out some implications of the analysis. The final section of the paper deals with the research limitations and explores avenues for future research.

2. Literature review and hypotheses development

Drawing on research from North America, Australasia, and Europe, there is a wealth of evidence to suggest that a wide variety of factors influence EF. These can be characterized as affluence as measured by gross domestic product (GDP), world system position (WSP), export dependence as measured by the percentage of exports to total GDP, service intensity, domestic income inequality, urbanization, and human capital. Previous research found GDP to be a robust predictor of per capita EF. For example, York et al. (2003) found that population and affluence account for 95% of the variance of total national EF. Jorgenson (2005) found GDP is positively correlated with per capita EF. Venetoulis (2001) examined carbon points for cities in Los Angeles County, California. The results of this study indicates a positive relationship between FP and per capita income. At the individual level data, Ryu and Brody (2004) found that high household incomes have significantly larger EF. Some authors postulate that EF initially rises with GDP growth, and then falls as per capita income continues to rise (e.g. Grossman and Krueger (1996)). This relationship is known in the literature as the Environmental Kuznets Curve (EKC) which describes the commonly observed inverted-U-shape relationship between environmental degradation and per-capita income. Fig. 1 illustrates this relationship. The empirical robustness of this relationship remains debatable (e.g. Dasgupta et al. (2002)). For example, Managi (2006) argues that the evidence of the inverted U-shape relationship applies only to a subset of environmental measures.

Per capita EF has been found to be largely a function of a country's position in the core/periphery hierarchy of the world-system. For example, Jorgenson (2003) found the position at the core of the world economy is causally linked with the highest per capita EF, followed by the semi-periphery and periphery. Hornborg (1998) found that consumers in core regions have a relative economic advantage compared to non-core countries, which enables them to consume natural and produced commodities at higher levels. Kentor (1998) also found that military size as well as continual research and development in core countries elevate both total GDP and consumption levels, which in turn affect EF. Several researchers refer to this body of literature as eco-structuralism (e.g. Grant et al. (2002)). They argue that WSP provides a needed counterbalance to conventional explanations of environmental degradation such as lifestyles and individual consumption habits.

Previous research found a negative relationship between export dependence and per capita EF. For example, Jorgenson (2005) found that exports intensity as measured by exports as percentage of total GDP is inversely related to the size of a nation's EF. This empirical finding is in line with what is referred to in the field of ecology as the "Netherlands fallacy", which refer to the error in assuming that the environmental impacts of wealthy nations will be contained within their national borders (Ehrlich and Holdren, 1971). The uneven ecological exchange theory (e.g. Bunker (1984) and Salzman (2000)) argues that service-based economies experience higher levels of growth in resource consumption which adversely affects EF. Jorgenson and Burns (2007) found that nations with a greater intensity in the services sector experience higher increases in per capita FP. Some evidence suggests that domestic income inequality is negatively related to the relative size of a nation's footprint (Jorgenson and Burns, 2007). However, Jorgenson (2004) found that the effect of domestic income inequality is positive in the core and negative in all other zones of the world economy stratification. Several authors found a positive relationship between urbanization as measured by the percentage of total population living in urban areas and EF levels (e.g. Jorgenson (2004) and York et al. (2003)). Jorgenson (2003) found urbanization to have the second largest positive impact on EF after WSP. However, some cross-national studies found that urbanization reduces domestic levels of deforestation (Ehrhardt-Martinez et al., 2002). Finally, Jorgenson (2003) found a positive relationship between literacy rate and EF levels. He argues that higher levels of literacy correspond with higher incomes, which allow for greater material consumption. The discussion presented above suggests the following hypotheses:

H1: Gross domestic product is positively related to growth in per capita EF.
H2: The relationship between economic growth and EF follows an inverted U-shaped EKC.
H3: World system position is positively related to growth in per capita EF.
H4: Export dependence is inversely related to growth in per capita EF.
H5: Services intensity is positively related to growth in per capita EF.
H6: Domestic income inequality is negatively related to growth in per capita EF.
H7: Urbanization level is positively related to growth in per capita EF.
H8: Literacy rate is positively related to growth in per capita EF.

3. Methodology

3.1. Multi-layer perceptron

MLP consists of sensory units that make up the input layer, one or more hidden layers of processing units (perceptrons), and one output layer of processing units (perceptrons). The MLP performs a functional mapping from the input space to the output space. An MLP with a single hidden layer having H hidden units and a single output, y, implements mappings of the form

$$y = F\left(W_0 + \sum_{h=1}^{H} W_h Z_h\right) \tag{1}$$

$$Z_h = F\left(\beta_{oh} + \sum_{j=1}^{n} \beta_{jh} X_j\right) \tag{2}$$

where Z_h is the output of the hth hidden unit, W_h is the weight between the hth hidden and the output unit, and W_0 is the output bias. There are n sensory inputs, X_j. The jth input is weighted by an amount β_j in the hth hidden unit. The output of an MLP is compared to a target output and an error is calculated. This error is back-propagated to the neural network and used to adjust the weights. This process aims at minimizing the mean square error between the network's prediction output and the target output.

One of the first successful applications of MLP is reported by Lapedes and Farber (1988). Using two deterministic chaotic time series generated by the logistic map and the Glass–Mackey equation, they designed an MLP that can accurately mimic and predict such dynamic nonlinear systems. There is an extensive literature on the financial applications of MLP (e.g. Kumar and Bhattacharya (2006) and Harvey et al. (2000)). Another major application of MLP is in electric load consumption (e.g. Darbellay and Slama (2000) and McMenamin and Monforte (1998)). Many other problems have been solved by MLP. A short list includes air pollution forecasting (e.g. Shin et al. (2005) and Videnova et al. (2006)), maritime traffic forecasting (Mostafa, 2004), airline passenger traffic forecasting (Nam and Yi, 1997), railway traffic forecasting (Zhuo et al., 2007), commodity prices (Kohzadi et al., 1996), ozone level (Ruiz-Suarez et al., 1995), student grade point averages (Gorr et al., 1994), forecasting macroeconomic data (Aminian et al., 2006), advertising (Poh et al., 1998), and market trends (Aiken and Bsat, 1999).

The MLP is the most frequently used neural network technique in pattern recognition (Bishop, 1999) and classification problems (Sharda, 1994). However, numerous researchers document the disadvantages of the MLP approach. For example, Calderon and Cheh (2002) argue that the standard MLP network is subject to problems of local minima. Swicegood and Clark (2001) claim that there is no formal method of deriving a MLP network configuration for a given classification task. Thus, there is no direct method of finding the ultimate structure for modeling process. Consequently, the refining process can be lengthy, accomplished by iterative testing of various architectural parameters and keeping only the most successful structures. Wang (1995) argues that standard MLP provides unpredictable solutions in terms of classifying statistical data.

3.2. Probabilistic neural network

PNN is used for classification problems where the objective is to assign cases to one of a number of discrete classes (Hunter, 2000). Theoretically, the PNN can classify an out-of-sample data with the maximum probability of success when enough training data is given (Enke and Thawornwog, 2005). PNNs feature a feed-forward architecture and supervised training algorithm similar to back propagation. The training pattern is presented to the input layer. The main role of the input layer is to map all the external signals into hidden layers by a scaling function through which each input neuron normalizes the range of external signals into a specific range that the neuron network can process. The neurons in the hidden layer aim to add flexibility to the performance of the PNN in order to record the knowledge of classification extracted from the training pattern. There must be, at least, as any neurons in the hidden layer as the number of training patterns (Tam et al., 2005). The summation layer consists of one neuron for each data class and sums the outputs from all hidden neurons of each respective data class. The output layer has one neuron for each possible category. The network produces activation, a value between zero and one in the output layer corresponding to the probability density function estimated from that category. The output with the highest value represents the most probable category.

The PNN was originally introduced to the neural network literature by Specht (1990). PNNs require no assumptions about distributions of random variables used to classify; they even can handle multi-modal distributions. They train quickly and as well as, or better than MLP networks. They have the ability to provide mathematically sound confidence levels and are relatively insensitive to outliers (Singer and Bliss, 2003). While the MLP network requires a validation data set (i.e., wasted cases) to search for over-fitting, PNNs use all available data in model building. The PNN is based on Bayes' classification method shown in Eq. (3), where h_i and h_j are the prior probabilities, c_i and c_j are the costs of misclassification, ad $f_i(x)$ and $f_j(x)$ are the true probability density functions:

$$h_i c_i f_i(x) > h_j c_j f_j(x). \tag{3}$$

In his seminal work, Specht (1990) presented the probability density estimates of a two class problem as

$$\hat{f}_A(\mathbf{x}) = 1/T_A \sum_{n=1}^{TA} \Phi_n(\|x - c_{An}\|) \tag{4}$$

$$\hat{f}_B(\mathbf{x}) = 1/T_B \sum_{n=1}^{TB} \Phi_n(\|x - c_{Bn}\|) \tag{5}$$

where $\mathbf{x}$ is the random vector of input data, Φ_n is a kernel density function, T_A and T_B are the number of samples for classes A and B, respectively, and c_{An} and c_{Bn} are the training data for classes A and B, respectively.

Following the EF analysis used by Andersson and Lindroth (2001), EFs of nations are grouped into four categories. EF interval set of bio-capacity-EF $>$ 1 is referred to as 'very strong ecological sustainability'. This set represents countries with an ecological surplus that exceeds the net-exports of biomass and sink-capacity. The ecological capital of these countries is intact or improving. The 0 $<$ bio-capacity-EF $\leq$ 1 is referred to as 'strong ecological sustainability'. This set includes countries with an ecological surplus smaller than their net-exports of bio-mass and sink-capacity. Consumption by the population in these countries requires less bio-capacity than what is available on its own territory. The $-1 <$ bio-capacity-EF $\leq$ 0 is referred to as 'weak ecological sustainability'. This set represents countries with an ecological deficit. The ecological capital of these countries is eroding already due to local overuse of available bio-capacity. Finally, the bio-capacity-EF < -1 is referred to as 'ecological unsustainability'. This set includes countries with an ecological deficit larger than their net-imports of bio-mass and sink-capacity. The ecological capital of these countries is decreasing at a very high rate. Thus, for the classification of EF, there are 4 different output classes and the conditional density estimates for each of these 4 classes can be written as

$$\hat{f}_x(x \mid C = i) = 1/T_i \sum_{n=1}^{Ti} \Phi_n(\|x - c_{in}\|) \quad i = 1, \ldots, 4, \tag{6}$$

where C is the class number. The selection process can be justified using Bayes Theorem. According to this theorem, the probability that the feature, x, belong to class i is

$$P(C = i \mid x) = f_x(x \mid C = i)P(C = i)/f_x(x) \quad i = 1, \ldots, 4, \tag{7}$$

where $f_x(\mathbf{x})$ is the density of $\mathbf{x}$. This can be written as

$$P(C = I \mid x) = \alpha_i f_x(x \mid C = i) \quad i = 1, \ldots, 4, \tag{8}$$

$$\text{where } \alpha_i = P(C = i)/f_x(x) \quad i = 1, \ldots, 4. \tag{9}$$

Since $f_x(x)$, the density of $\mathbf{x}$, is the same for all i and the probability that the occurrence of each EF class is the same [i.e., $P(C = 1) = P(C = 2) = \cdots = P(C = 4)$], the α_is are going to be the same for all these classes.

The PNN has been extensively used in various pattern classification tasks in the literature due to ease of training and sound statistical foundation in Bayesian estimation theory. For example, Yang and Marjorie (1999) utilized a PNN to predict the financial crisis in oil industry companies in the USA. Jin and Srinivasan (2001) proposed a new technique for freeway incident detection using PNN. Hajmeer and Basheer (2002) used PNN to study the classification of bacterial growth. Chen et al. (2003) applied PNN to stock index forecasting. Huang (2004) applied PNN to predict the class of leukaemia and colon cancer. Gerbec et al. (2005) used PNN to classify consumers' electricity load profiles. Xue et al. (2005) classified 102 active compounds from diverse medicinal plants with anticancer activity. Jin and Englande (2006) used PNN to classify whether conditions in a lake mean it is safe to swim or not. Wilson (2006) successfully tested the PNN on 209 seizures obtained from an epilepsy-monitoring unit. Laskari et al. (2006) evaluated the performance of PNN on approximation problems related to cryptography.

3.3. Generalized regression neural network

GRNN was devised by Specht (1991), casting a statistical method of function approximation into a neural network form. The GRNN, like the MLP, is able to approximate any functional relationship between inputs and outputs (Wasserman, 1993). Structurally, the GRNN resembles the MLP. However, unlike the MLP, the GRNN does not require an estimate of the number of hidden units to be made before training can take place. Furthermore, the GRNN differs from the classical MLP in that every weight is replaced by a distribution of weight which minimizes the chance of ending up in local minima. Therefore, no test and verification sets are required, and in principle all available data can be used for the training of the network (Parojcic et al., 2007). However other researchers have used holdout samples to prevent the risk of over-fitting (e.g. Ben-Nakhi and Mahmoud (2004), Raitsos et al. (2008)).

The GRNN is a method of estimating the joint probability density function (pdf) of x and y, giving only a training set. The estimated value is the most probable value of y and is defined by

$$E(y \mid \mathbf{x})y = \hat{y}(\mathbf{x}) = \int_{-\infty}^{+\infty} yf(\mathbf{x}, y)\mathrm{d}y \Big/ \int_{-\infty}^{+\infty} f(\mathbf{x}, y)\mathrm{d}y. \tag{10}$$

The density function $f(\mathbf{x}, y)$ can be estimated from the training set using Parzen's estimator (Parzen, 1962)

$$f(\mathbf{x}, y) = 1/(2\pi)^{(p+1)/2}\sigma^{(p+1)}1/n\sum_{i=1}^{n}\exp[-(x-x^i)^{\mathrm{T}}(x-x^i)/2\sigma^2]\exp[-(y-y^i)^2/2\sigma^2]. \tag{11}$$

The probability estimate $f(x, y)$ assigns a sample probability of width σ for each sample x^i and y^i, and the probability estimate is the sum of these sample probabilities (Specht, 1991). Defining the scalar function D_i^2

$$D_i^2 = (x - x_i)^{\mathrm{T}}(x - x_i) \tag{12}$$

and assessing the indicated integration yields the following:

$$Y(\mathbf{x}) = \sum_{i=1}^{n} Y^i \exp(-D_i^2/2\sigma^2) \Big/ \sum_{i=1}^{n}\exp(-D_i^2/2\sigma^2). \tag{13}$$

The resulting regression (13) is directly applicable to problems involving numerical data.

Due to its good performance in noisy environments, the GRNN has been extensively used in various prediction and forecasting tasks in the literature. For example, Gaetz et al. (1998) analyzed EEG activity of the brain using a GRNN. Chtioui et al. (1999) used a GRNN for leaf wetness prediction. Ibric et al. (2002) used a GRNN in the design of extended-release aspirin tablets. Cigizoglu (2005) employed a GRNN to forecast monthly water flow in Turkey. In this study, GRNN forecasting performance was found to be superior than the MLP and other statistical and stochastic methods. Kim and Lee (2005) used a GRNN-based genetic algorithm to predict silicon oxynitride etching. Hanna et al. (2007) developed a GRNN model to predict seismic condition in sites susceptible to liquefaction. Shie (2008) used a hybrid method integrating a GRNN and a sequential quadratic programming method to determine an optimal parameter setting of an injection-molding process.

4. Results

4.1. Preliminary data analysis

In this study per capita EF data were taken from White (2007). This data set is the latest available data on per capita EF of nations. Population and land area data were taken from the World Bank (2002). We use the natural logarithm transformation to correct for outliers. WSP data were taken from a world indicator developed by Kentor (2000). This is a multidimensional indicator that combines economic, military, and export dependence characteristics in one measure. Kentor's WSP index is used in other cross-national studies of natural resource consumption and other environmental outcomes (e.g. Jorgenson (2003)). GDP data were taken from the World Bank (2002). Consistent with other cross-national studies (e.g. Jorgenson (2003)), we use the natural logarithm transformation of the GDP data to correct for excessive skewness. Exports as percentage of total GDP data were taken from the World Bank (2002). This data set measures the export intensity and export dependence of a nation. Services as percentage of total GDP data were taken from the World Bank (2002). This data measure the extent to which a domestic economy is service based. Domestic income inequality data were taken from the World Bank (2002). These data are measured as Gini coefficients. A Gini score of zero means perfect equality while a score of one hundred suggests perfect inequality. Urbanization level data were taken from the World Bank (2002). Following Jorgenson and Burns (2007), we regress the data on per capita GDP and use the residuals as measure of urbanization independent of economic development. The residualized data are used to minimize collinearity. Finally, literacy rate data were taken from the World Bank (2002). Literacy rate measures the percent of a nations population over the age of fifteen that can read and write. Table 1 provides descriptive statistics for all variables used in the analysis while Fig. 2 provides boxplots for the major variables included in the analysis.

Though it does not prove causation, correlation can serve as predictor of causation (Sekaran, 2000). The product moment correlations between the variables are shown in Table 2. This table was constructed to get a feel for the associations among the ten variables used in the study. Most of the correlation coefficients were significant and had the expected sign.

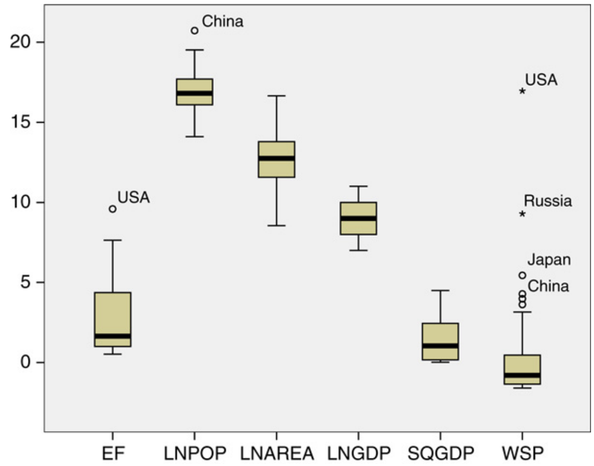

Fig. 2. Boxplots of major variables included in the analysis.

Table 1
Descriptive statistics of variables used in the analysis ($N = 140$).

Variable	Mean	SD	Minimum	Maximum
EF	2.44	2.04	0.52	9.59
WSP*	0.07	2.59	−1.59	16.96
GDP (ln)	8.67	1.19	6.00	11.00
GDP squared (ln)	1.39	1.32	0.00	5.17
PercServ (ln)	3.59	0.55	2.08	4.76
Service as % of GDP	52.63	13.11	18.74	76.88
GINI	41.06	10.52	21.80	73.90
URB (residualized)	0.00	16.69	−33.08	38.94
LIT	80.57	20.80	17.00	100.00

* $N = 87$ (due to missing data).

Table 2
Product–moment correlations matrix.

	1	2	3	4	5	6	7	8	9	10
1. POP										
2. AREA	.547**									
3. WSP	.335**	.112								
4. GDP	−.073	−.104	.543**							
5. SRGDP	.011	−.027	.376**	.009						
6. %EX	−.317**	−.174*	−.185*	.348**	−.208*					
7. %SER	.010	−.140	.353**	.495**	.013	−.067				
8. GINI	.031	.147	−.297**	−.373**	−.254**	.242**	.166			
9. URB	.024	.004	.401**	.726**	−.061	.261**	.414**	−.201*		
10. LIT	.013	.065	−.279**	−.484**	.160	−.278**	−.383**	.114	−.486**	
11. EF	−.054	−.024	.688**	.772**	.350**	.270**	.419**	−.425**	.142	.398**

** Correlation is significant at the .05 level (2-tailed).
* Correlation is significant at the .01 level (2-tailed).

4.2. Hierarchical linear regression

Hierarchical regression analysis was used to test the research hypotheses. This method is also known as incremental variance partitioning (Pedhazur, 1982). This approach allows us to focus on the variables forming the hypotheses, and at the same time sieve out the influence of the control variables that might have a moderating effect on EF. Also this method allows the researcher to control the order of the variables entered into the regression model, allowing us to assess the incremental predictive ability of any variable of interest (McQuarrie, 1988).

Prior research has demonstrated that per capita EF can be influenced by a nation's population and land area (e.g. York et al. (2003)). Therefore, these two variables were included as the first of nine blocks in hierarchical multiple regression. The two variables were included as controls to reduce the possibility of spurious relationships based on these variables. Table 3 shows a summary of results of the hierarchical regression analysis. As seen in Table 3, when the two control variables were entered into the regression equation in the first step, the coefficient of determination (R^2) was found to be 0.012 indicating that 1.2% of per capita EF is explained by these variables. By adding the GDP variable in step 2, R^2 increased to 0.650 or 65%.

Table 3
Hierarchical regression results.

Model	Standardized β	R^2	R^2 change	F change	Sig.	Tolerance	VIF
1 (Constant)		.012	.012	.519	.597		
LnPop	−.074					.754	1.326
LnArea	−.053					.754	1.326
2 (Constant)		.650	.638	151.166	.000		
LnPop	−.061					.754	1.326
LnArea	.062					.743	1.347
LnGDP	.808					.977	1.023
3 (Constant)		.766	.116	40.683	.000		
LnPop	−.045					.753	1.328
LnArea	.039					.740	1.351
LnGDP	.721					.919	1.088
LnSqGDP	.352					.939	1.065
4 (Constant)		.828	.062	.007	.932		
LnPop	−.185					.609	1.642
LnArea	.042					.740	1.351
LnGDP	.539					.616	1.623
LnSqGDP	.257					.827	1.209
WSP	.356					.491	2.036
5 (Constant)		.828	.000	.007	.932		
LnPop	−.186					.572	1.749
LnArea	.042					.737	1.357
LnGDP	.540					.589	1.698
LnSqGDP	.257					.822	1.217
WSP	.356					.486	2.060
PercEx	−.004					.799	1.251
6 (Constant)		.832	.003	1.631	.205		
LnPop	−.186					.572	1.749
LnArea	.055					.711	1.407
LnGDP	.497					.448	2.234
LnSqGDP	.257					.822	1.217
WSP	.350					.483	2.070
PercEx	.020					.704	1.420
PercServ	.078					.575	1.739
7 (Constant)		.846	.015	7.399	.008		
LnPop	−.231					.528	1.892
LnArea	.088					.675	1.481
LnGDP	.432					.396	2.522
LnSqGDP	.198					.685	1.460
WSP	.362					.481	2.081
PercEx	−.017					.662	1.511
PercServ	.090					.572	1.749
Gini	−.157					.590	1.696
8 (Constant)		.850	.004	2.001	.161		
LnPop	−.218					.515	1.942
LnArea	.081					.670	1.491
LnGDP	.400					.360	2.780
LnSqGDP	.231					.574	1.741
WSP	.361					.480	2.082
PercEx	−.019					.661	1.749
PercServ	.091					.572	1.749
Gini	−.171					.573	1.746
URB	.076					.667	1.500
9 (Constant)		.853	.003	1.343	.250		
LnPop	−.219					.515	1.942
LnArea	.077					.667	1.499
LnGDP	.388					.353	2.834
LnSqGDP	.228					.573	1.744
WSP	.359					.480	2.083
PercEx	−.027					.652	1.533
PercServ	.063					.487	2.054
Gini	−.177					.568	1.759
URB	.063					.636	1.573
Lit	.065					.618	1.617

This R^2 change (0.638) is significant ($p < 0.001$). This implies that the additional 63.8% of variation in per capita EF is explained by GDP.

In the third step, GDP squared was entered. This variable was entered to test the existence of an EKC. When the variable was entered the R^2 increased from 63.8% to 76.6% indicating a change of 11.6%, which is significant ($p < 0.001$). By adding

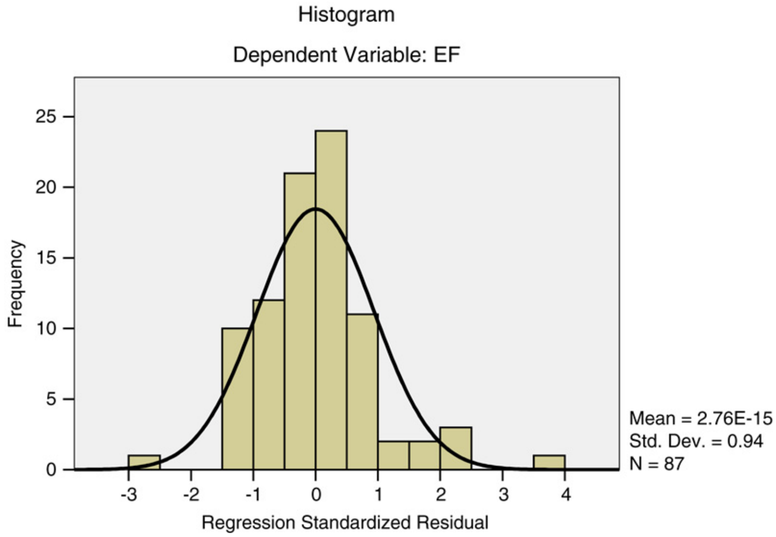

Fig. 3. Histogram of regression standardized residuals.

Table 4
Hypotheses testing results.

Hypothesis	Relationship	Result
WSP → EF	Positive	Supported
GDP → EF	Positive	Supported
SQGDP → EF	Negative	Not supported
EXP → EF	Negative	Not supported
SER → EF	Positive	Not supported
Gini → EF	Negative	Supported
URB → EF	Positive	Not supported
LIT → EF	Positive	Not supported

the WSP variable in step 4, R^2 increased to .828% or 82.8%. This change is significant ($p < 0.001$). This implies that the additional 6.2% of variance in per capita EF is explained by WSP. In the fifth step, exports as a percentage of total GDP was entered in the equation in order to gauge its impact as an independent predictor. From the regression equation (model 5), it can be seen that R^2 did not increase which indicates that the export as a percentage of total GDP variable is insignificant. Services as a percentage of total GDP was entered in the sixth step. This variable was found to be insignificant ($p = 0.205$). In the seventh step the domestic income inequality was entered into the regression equation. By entering this variable R^2 increased from 0.832 to 0.846. This change was significant at the 0.05 level. The urbanization level variable was entered in the eighth step while literacy rate was entered in the ninth step. The two variables were found insignificant at the 0.001 level. Table 3 reports also the variance inflation factor (VIF) and tolerance values for each independent variable in all models tested. The values reported show that multicollinearity is not a serious problem. The largest VIF value among all X variables is 2.780. The largest VIF value is often used as an indicator of the severity of multicollinearity. A maximum VIF value in excess of 10 is frequently taken as an indication that multicollinearity may be unduly influencing the least square estimates (Kutner et al., 2005). The shape of histogram of regression residuals seems to depart slightly from the bell-shaped curve a shown Fig. 3. This was confirmed by the normal probability plot of residuals (Fig. 4) which suggests that the underlying distributions are not normal. Additionally, the final regression model explains 85% of the variation in the per capita EF. This indicates that here still remains a 15% of the variability in the per capita EF data that are not linearly explicable. Seeking an enhanced model that can capture the nonlinear aspects of the relation under investigation is thus warranted.

We also found GDP to be positively related to per capita EF which supports hypothesis 1. Surprisingly, we found no evidence of the EKC as the coefficient for the quadratic per capita GDP is positive and significant-opposite of that necessary to generate an EKC. However, recent studies using the EF as a measure of environmental degradation have found essentially no support for the EKC (e.g. Bagliani et al. (2008)). We found that EF is likely to be based on the nation's WSP which supports hypothesis 3. The signs of both the export dependence and the service intensity hypotheses were in the expected direction, however they did no reach statistical significance. The domestic income inequality hypothesis was negatively related to per capita EF which supports hypothesis 6. The signs of both urbanization and Literacy rate were in the expected direction, however, they did not reach statistical significance. Table 4 summarizes the results of hypotheses testing.

It should be noted that we tested the research hypotheses using hierarchical linear regression and LDA as NNs are of limited ability to statistically test and interpret hypotheses concerning the roles of specific variables that are included in the

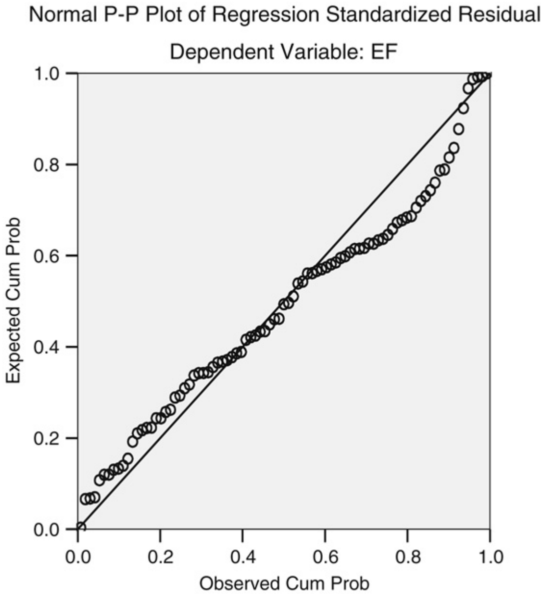

Fig. 4. Normal probability plot of regression standardized residuals.

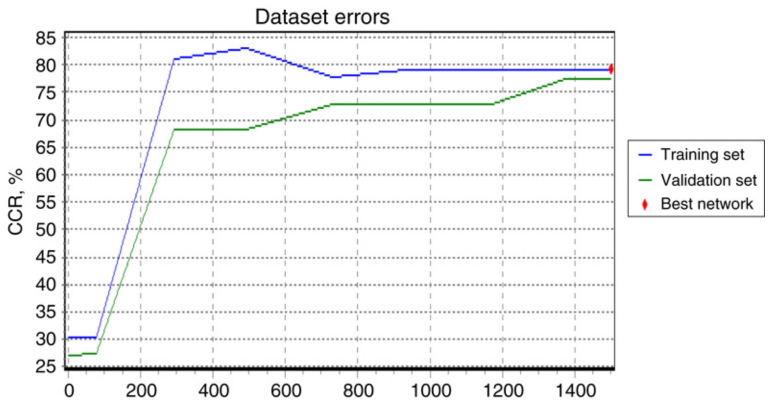

Fig. 5. Correct classification rate (CCR) for the MLP neural network

models as predictors. However, this limitation "should not be a serious drawback if one simply desires classification from the model" (Swicegood and Clark, 2001, p. 176).

4.3. MLP and PNN-based classification

There are many software packages available for analyzing MLP models. We chose the NeuroIntelligence package (Alyuda Research Company, 2003). This software applies artificial intelligence techniques to automatically find the efficient MLP architecture. Typically, the application of MLP requires a training data set and a testing data set (Lek and Guegan, 1999). The training data set is used to train the MLP and must have enough examples of data to be representative for the overall problem. The testing data set should be independent of the training set and is used to assess the classification accuracy of the MLP after training. Following Lim and Kirikoshi (2005), an error back-propagation algorithm with weight updates occurring after each epoch was used for MLP training. The learning rate was set at 0.1. After 1500 iterations, the correct classification rate (CCR) reached 79% as seen in Fig. 5. Table 5 reports the properties and predictive accuracy of the MLP model. As can be observed, the MLP classifier predicted a training sample with 79% accuracy and a validation sample with 78% accuracy.

The PNN network used in this study was trained with 80% of the data selected randomly. Network training is a process by which the connection weights and biases of the NN are adapted through a continuous process of simulation by the environment in which the network is embedded. The primary purpose of training is to minimize an error function by searching for a set of connections strengths and biases that causes the NN to produce outputs that are equal or close to targets. A number of training algorithms can be used. In practice, the Levenberg–Marquardt routine often finds better optima for a

Table 5
MLP properties.

Parameter	Value
Architecture	[8-7-4]
Input activation function	Logistic
Classification model	Confidence limits
Accept/reject level	0.5
Training algorithm	Quick propagation
Number of iterations	1500
Number of weights	81
Fitness	3.67
Akaike Information Criterion (AIC)	−459.57
Output error	Cross-entropy
Network error (training)	0.106
Network error (validation)	0.000
Error improvement	0.000
CCR % (training)	78.9
CCR % (validation)	77.7

Table 6
PNN properties.

Parameter	Value
Configuration	PNN category predictor
Independent category variables	(Ecological footprint)
Independent numeric variables	(WSP, GDP, SRGDP, %EX, %SER, GINI, URB, LIT)
Number of cases (training)	70
Number of trials (training)	103
% Bad predictions (training)	1.43
Mean incorrect probability % (training)	5.75
SD of incorrect probability % (training)	9.36
Number of cases (testing)	17
% Bad predictions (testing)	17.65
Mean incorrect probability % (testing)	32.05
SD of incorrect probability % (testing)	31.82

variety of problems than the other optimization techniques (Shavlick et al., 1991). After the training phase, the NN model is applied to the data set to classify each nation's EF into one of the four categories.

There are many computer software packages available for building and analyzing PNNs. Because of its extensive capabilities for building networks based on a variety of training and learning methods, NeuralTools Professional package (Palisade Corporation, 2005) was chosen in this study. This software automatically scales all input data. Scaling involves mapping each variable to a range with minimum and maximum values of 0 and 1. NeuralTools Professional software uses a non-linear scaling function known as the 'tanh', which scales inputs to a $(-1, 1)$ range. This function tends to squeeze data together at the low and high ends of the original data range. It may thus be helpful in reducing the effects of outliers (Tam et al., 2005). PNN properties are shown in Table 6.

To study the effectiveness of the MLP and PNN-based classification of EF, the results of MLP and PNN were compared with the traditional multiple discriminant analysis (MDA). MDA is frequently used supervised pattern recognition technique. A linear function of the variables is sought, which maximizes the ratio of between-class variance and minimizes the ratio of within-class variance. MDA is an extremely simple and efficient method of classification. Indeed, it cannot be outperformed if the two distributions are normal and have the same dispersion matrix (i.e., Bayes limit). Fig. 6 shows the canonical discriminant functions' EF group centroids. A common measure of predictive models is the percentage of observation correctly classified or the hit ratio. Table 7 reports the predictive accuracy of the two NN models as compared with the MDA. As can be observed in Panel (a), the MLP classifier predicted the training sample with 79% accuracy and the test sample with 77.7%. In panel (b), the PNN classifier predicted the training sample with 98.6% accuracy and the test sample with 82.4% accuracy. In Panel (c), the MDA model had an accuracy rate of 74.7%. Press's Q statistic of 114.67 is greater than the critical value of 11.34 (χ^2 value with 3 *df* is statistically significant at the 0.01 level of significance). That is the discriminant analysis classification accuracy is greater than that expected by chance.

Despite the satisfactory classification performance of the NN models in this study, such models are often criticized as black boxes that do not allow decision-makers to make inferences on how the input variables affect the models' results (e.g. Lee et al. (2006)). One way to address this issue is to conduct a variable impact analysis (VIA). In this study VIA was performed using the variable impact option in NeuralTools software. The purpose of variable impact analysis is to measure the sensitivity of net predictions to changes in independent variables. Fig. 7 shows the most important input variable for the PNN is WSP followed by GDP and domestic income inequality as measured by the Gini coefficient. The lower the percent value for a given variable, the less that variable affects the predictions. The results of the analysis can help in the selection of a new set of independent variables, one that will allow more accurate predictions. For example, a variable with a low

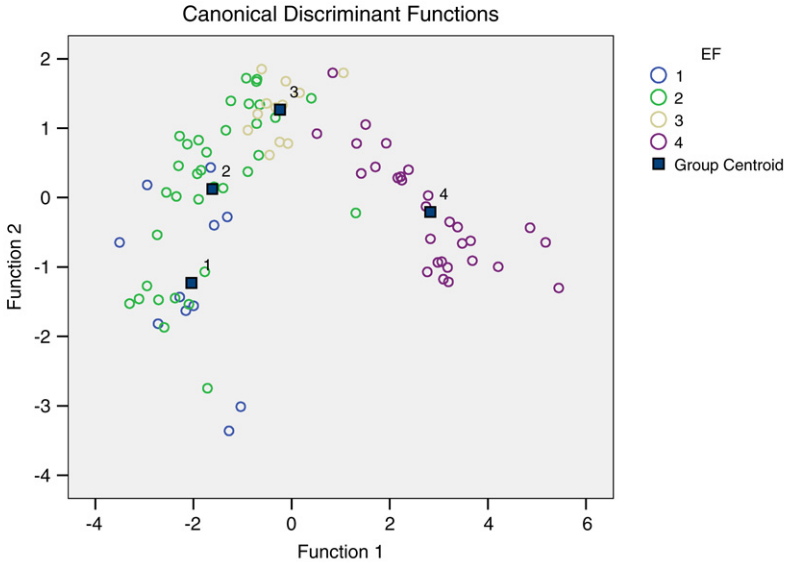

Fig. 6. Discriminant analysis group centroids.

Table 7
Predictive accuracy of classification models.

Panel (a): MLP*					
Class count (training)	1	2	3	4	Total
1	7	9	0	0	16
2	0	36	0	2	38
3	0	3	6	3	12
4	0	1	2	26	29
Class count (testing)					
1	8	1	0	0	9
2	3	18	0	2	23
3	0	2	4	0	6
4	0	1	1	5	7
Panel (b): PNN**					
Class count (training)	1	2	3	4	Total
1	10	0	0	0	10
2	1	22	0	0	23
3	0	0	10	0	10
4	0	0	0	27	27
Class count (testing)					
1	1	0	0	0	1
2	3	9	0	0	12
3	0	0	2	0	2
4	0	0	0	2	2
Panel (c): MDA***					
Class count	1	2	3	4	Total
1	8	3	0	0	11
2	9	18	7	1	35
3	0	0	12	0	12
4	0	0	2	27	29
Class per cent					
1	72.7	27.3	0.0	0.0	100.00
2	25.7	51.4	20.0	2.9	100.00
3	0.0	0.0	100	00.0	100.00
4	0.0	0.0	6.9	93.1	100.00

* 79% of original grouped cases correctly classified (training). 77.7% of original grouped cases correctly classified (testing).
** 98.6% of original grouped cases correctly classified (training). 82.4% of original grouped cases correctly classified (testing).
*** 74.7% of original grouped cases correctly classified (67.8% of cross-validated grouped cases correctly classified).

impact value can be eliminated in favor of some new variables. PNN results confirm the hierarchical regression results in this study and also confirm the other regression and structural equation modeling results of previous research. These results

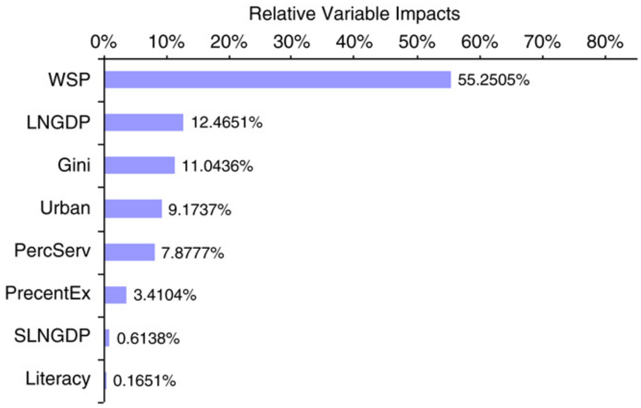

Fig. 7. Relative variable impact analysis.

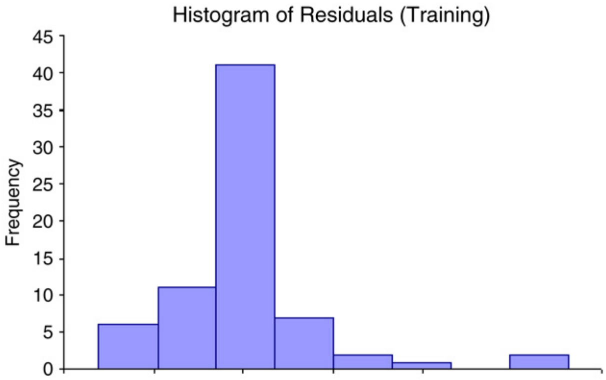

Fig. 8. GRNN histogram of residuals (training).

show that WSP and GDP are the most important factors in determining the per capita EF of a nation (e.g. Jorgenson (2003) and Jorgenson and Burns (2007)).

4.4. GRNN-based prediction

NeuralTools Professional package (Palisade Corporation, 2005) was chosen to build the GRNN in this study. This software uses the conjugate gradient descent optimization technique to train the GRNN. The error measure used during the training to evaluate different smoothing factors is the mean square error. GRNN properties used in this study are shown in Table 8. Fig. 8 shows the histogram of GRNN residuals. This Figure shows that no pattern is detected in the GRNN residuals, which suggests that the errors are approximately normally distributed with a mean of zero. The GRNN was then used to predict EF. These predicted values were then plotted against the observed EF values as shown in Fig. 9. Here the straight line indicates the ideal trend where the predicted and observed EF values will be the same at different levels. The trend line of predicted EF values very closely follows the ideal pattern. The results were satisfactory, as indicated by the relatively high R^2 value of 94%. The fact that the R^2 showed such an increase in the GRNN model indicates the data are best represented in a model able to detect nonlinear relationships. An improved performance from the use of NN versus regression was also found in the work by Larrain (2007) who used both methods to predict treasury bills and inventories. Similar results were reported in Hurrion and Birgil (1999) and in Radhakrishnan and Nandan (2005).

5. Implications, limitations and future research

Our results confirm the theoretical work by Hecht-Nielson (1989) who has shown that NNs can learn input–output relationships to the point of making perfect forecasts with the data on which the network is trained. However, perfect forecasts with the training data do not guarantee optimal forecasts with the testing data due to differences in the two data sets. The good performance of the NN models in predicting and classifying per capita EF can be traced to it's inherent non-linearity. This makes an NN ideal for dealing with non-linear relations that may exist in the data. Thus, neuro-computational

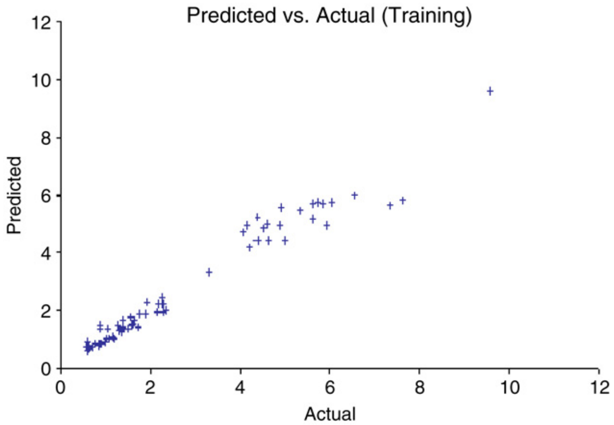

Fig. 9. GRNN predicted vs. actual values (training).

Table 8
GRNN properties.

Parameter	Value
Configuration	GRNN numeric predictor
Independent category variables	(Ecological footprint)
Independent numeric variables	(WSP, GDP, SRGDP, %EX, %SER, GINI, URB, LIT)
Number of cases (training)	70
Number of trials (training)	99
% Bad predictions (training)	7.15
Root mean square error (training)	0.42
Mean absolute error (training)	0.24
SD of absolute error (training)	0.35
Number of cases (testing)	17
% Bad predictions (testing)	23.53
Root mean square error (testing)	0.69
Mean absolute error (testing)	0.47
SD of absolute error (testing)	0.51

models are needed to better understand the inner dynamics of nations' EF. Our results are also in line with the findings of other researchers who have investigated the performance of NN compared to other traditional statistical techniques, such as regression analysis, discriminant analysis, and logistic regression analysis. For example, in a study of clinical diagnosis of cancers, Shan et al. (2002) found a hit ratio of 85% for the PNN model compared to 80% for the MDA model. In a study of credit-scoring models used in commercial and consumer lending decisions, Bensic et al. (2005) compared the performance of logistic regression, neural networks and decision trees. The PNN model produced the highest hit rate and the lowest type I error. Similar findings have been reported in a study examining the performance of NN in predicting bankruptcy (Anandarajan et al., 2001) and diagnosis of acute appendicitis (Sakai et al., 2007).

Despite the significant contributions of this study, it suffers from a number of limitations. First, this study has used a cross-sectional rather than a longitudinal approach. This implies that much more emphasis has been placed on observing per capita EF of nations than in observing changes in global EF. There would seem to be therefore a need for much more longitudinal research to focus on observing changes in EF behavior over time. Second, despite the satisfactory performance of the NN models in this study, future research might improve the performance of the NN models used in this study by integrating fuzzy discriminant analysis and genetic algorithms (GA) with NN models. Mirmirani and Li (2004) pointed out that traditional algorithms search for optimal weight vectors for a neural network with a given architecture, while GA can yield an efficient exploration of the search space when the modeler has little *apriori* knowledge of the structure of problem domains. Finally, future research might use other NN architectures such as self-organizing maps (SOMs) to classify the EF of nations. Due to the unsupervised character of their learning algorithm and the excellent visualization ability, SOMs have been recently used in myriad classification tasks. Examples include classifying cognitive performance in schizophrenic patients and healthy individuals (Silver and Shmoish, 2008), mutual funds classification (Moreno et al., 2006), crude oil classification (Fonseca et al., 2006), and classifying magnetic resonance brain images (Chaplot et al., 2006).

Appendix

See Table A.1.

Table A.1
Countries included in the analysis.

Country	EF	Country	EF	Country	EF
Albania	1.43	Ghana	0.96	Nigeria	1.17
Algeria	1.58	Greece	5	Norway	5.85
Angola	1.01	Guatemala	1.29	Pakistan	0.6
Argentina	2.26	Guinea	0.94	Panama	1.89
Armenia	1.1	Guinea-Bissau	0.66	Paraguay	1.63
Australia	6.56	Haiti	0.56	Peru	0.87
Austria	4.94	Honduras	1.27	Philippines	1.05
Azerbaigan	1.75	Hungary	3.5	Poland	3.29
Bangladesh	0.52	India	0.75	Portugal	4.19
Belarus	3.32	Indonesia	1.06	Romania	2.35
Belgium	5.62	Iran	2.38	Russia	4.41
Benin	0.82	Ireland	4.95	Rwanda	0.66
Bolivia	1.34	Israel	4.62	Saudi Arabia	4.64
Bosnia Herzegovina	2.33	Italy	4.15	Senegal	1.16
Botswana	1.58	Jamaica	1.75	Serbia and Montenegro	2.28
Brazil	2.15	Japan	4.35	Sierra Leone	0.72
Bulgaria	3.11	Jordan	1.76	Slovakia	3.23
Burkina Faso	0.99	Kazakhstan	3.97	Slovenia	3.42
Burundi	0.67	Kenya	0.81	South Africa	2.29
Cambodia	0.71	Korea DR	1.44	Spain	5.36
Cameroon	0.83	Korea Republic	4.05	Sri Lanka	1
Canada	7.61	Kuwait	7.35	Sudan	1
Central African Rep	0.87	Kyrgyzstan	1.25	Swaziland	1.15
Chad	1.03	Laos	0.89	Sweden	6.07
Chile	2.33	Latvia	2.59	Switzerland	5.15
China	1.64	Lebanon	2.91	Syria	1.73
Colombia	1.28	Libya	3.42	Tajikistan	0.64
Congo	0.62	Lithuania	4.44	Tanzania	0.7
Congo Dem Rep	0.58	Macedonia	2.32	Thailand	1.38
Costa Rica	1.98	Madagascar	0.71	Togo	0.87
Cote d'Ivoire	0.75	Malawi	0.56	Trinidad and Tobago	3.13
Croatia	2.94	Malaysia	2.25	Tunisia	1.54
Cuba	1.54	Mali	0.84	Turkey	2.06
Czech Republic	4.91	Mauritania	1.26	Turkmenistan	3.47
Denmark	5.75	Mauritius	1.87	Uganda	1.08
Dominican Republic	1.61	Mexico	2.56	Ukraine	3.19
Ecuador	1.49	Moldova Republic	1.27	United Arab Emirates	7.87
Egypt	1.35	Mongolia	3.09	United Kingdom	5.59
El Salvador	1.37	Morocco	0.88	United States	9.59
Eritrea	0.71	Mozambique	0.63	Uruguay	1.92
Estonia	6.47	Myanmar	0.91	Uzbekistan	1.83
Ethiopia	0.82	Namibia	1.14	Venezuela	2.18
Finland	7.64	Nepal	0.69	Vietnam	0.88
France	5.63	Netherlands	4.39	Yemen	0.85
Gabon	1.38	New Zealand	5.94	Zambia	0.63
Gambia	1.37	Nicaragua	1.18	Zimbabwe	0.85
Germany	4.55	Niger	1.11		

Source: White (2007).

References

Aiken, M., Bsat, M., 1999. Forecasting market trends with neural networks. Information Systems Management 16, 42–49.

Alyuda Research Company. 2003. NeuroIntelligence User Manual (Version 2.1).

Aminian, F., Suarez, E., Aminian, M., Walz, D., 2006. Forecasting economic data with neural networks. Computational Economics 28, 71–88.

Anandarajan, M., Lee, P., Anandarajan, A., 2001. Bankruptcy prediction of financially stressed firms: An examination of the predictive accuracy of artificial neural networks.. International Journal of Intelligent Systems in Accounting, Finance & Management 10, 69–81.

Andersson, J., Lindroth, M., 2001. Ecologically unsustainable trade. Ecological Economics 37, 113–122.

Bagliani, M., Bravo, G., Dalmazzone, S., 2008. A consumption-based approach to environmental Kuznets curves using the ecological footprint indicator. Ecological Economics 65, 650–661.

Barrett, J., Scott, A., 2003. The application of the ecological footprint: A case of passenger transport in Merseyside. Local Environment 8, 167–183.

Ben-Nakhi, A., Mahmoud, M., 2004. Cooling load prediction for buildings using general regression neural networks. Energy Conversion and Management 45, 2127–2141.

Bensic, M., Sarlija, N., Zekic-Susac, M., 2005. Modeling small-business credit scoring by using logistic regression, neural networks and decision trees. Intelligent Systems in Accounting, Finance and Management 13, 133–150.

Binningsbo, H., de Soya, I., Gleditsch, N., 2007. Green giant or straw man? Environmental pressure and civil conflict, 1961–99. Population and Environment 28, 337–353.

Bishop, C., 1999. Neural Networks for Pattern Recognition. Oxford University Press, New York.

Bunker, S., 1984. Modes of extraction, unequal exchange, and the progressive underdevelopment of an extreme periphery: The Brazilian Amazon. American Journal of Sociology 89, 1017–1064.

Calderon, T., Cheh, J., 2002. A roadmap for future neural networks research in auditing and risk assessment. International Journal of Accounting Information Systems 3, 203–236.

Chaplot, S., Patnaik, L., Jagannathan, N., 2006. Classification of magnetic resonance brain images using wavelets as input to support vector machines and neural network. Biomedical Signal Processing and Control 1, 86–92.
Chen, A., Leung, M., Daouk, H., 2003. Application of neural networks to an emerging market: Forecasting and trading the Taiwan stock index. Computers & Operations Research 30, 901–923.
Chtioui, Y., Panigrahi, S., Francl, L., 1999. A generalized regression neural network and its application for leaf wetness prediction to forecast plant disease. Chemometrics and Intelligent Laboratory Systems 48, 47–58.
Cigizoglu, H., 2005. Generalized regression neural network in monthly flow forecasting. Civil Engineering and Environmental Systems 22, 71–84.
Constanza, R., 2000. The dynamics of the ecological footprint concept. Ecological Economics 32, 341–345.
Crompton, S., Roy, R., Caird, S., 2002. Household ecological footprinting for active distance learning and challenge of personal lifestyles. International Journal of Sustainability in Higher Education 3, 313–323.
Darbellay, G., Slama, M., 2000. Forecasting the short-term demand for electricity: Do neural networks stand a better chance? International Journal of Forecasting 16, 71–83.
Dasgupta, S., Laplante, B., Wang, H., Wheeler, D., 2002. Confronting the environmental Kuznets curve. Journal of Economic Perspectives 16, 147–168.
Dauvergne, P., 2005. Research in global environmental politics: History and trends. In: Dauvergne, P. (Ed.), Handbook of Global Environmental Politics. Edward Elgar, Cheltenham, pp. 8–32.
Ehrhardt-Martinez, K., Crenshaw, E., Jenkins, C., 2002. Deforestation and the environmental Kuznets curve: A cross-national investigation of intervening mechanisms. Social Science Quarterly 83, 226–243.
Ehrlich, P., Holdren, J., 1971. Impacts of population growth. Science 171, 1212–1217.
Enke, D., Thawornwog, S., 2005. The use of data mining and neural networks for forecasting stock market returns. Expert Systems with Applications 29, 927–940.
Fonseca, A., Biscaya, J., de Sousa, J., Lobo, A., 2006. Geographical classification of crude oils by Kohonen self-organizing maps. Analytica Chimica Acta 556, 374–382.
Gaetz, M., Weinberg, H., Rzempoluck, E., Jantzen, K., 1998. Neural network classification and correlation analysis of EEG and MEG activity accompanying spontaneous reversals of the Necker cube. Cognitive Brain Research 6, 335–346.
Gerbec, D., Gasperic, S., Smon, I., Gubina, F., 2005. Allocation of the load profiles to consumers using probabilistic neural networks. IEEE Transactions on Power Systems 20, 548–555.
Gorr, W., Nagin, D., Szczypula, J., 1994. Comparative study of artificial neural network and statistical models for predicting student point averages. International Journal of Forecasting 10, 17–34.
Grant, S., Jones, A., Bergesen, A., 2002. Organizational size and pollution: The case of the U.S chemical industry. American Sociological Review 67, 389–407.
Grossman, G., Krueger, B., 1996. The inverted-U: What does it mean? Environment and Development Economics 1, 119–122.
Hajmeer, M., Basheer, I., 2002. A probabilistic neural approach for modeling and classification of bacterial growth/no-growth data. Journal of Microbiological Methods 51, 217–226.
Hanna, A., Ural, D., Saygili, G., 2007. Evaluation of liquefaction potential of soil deposits using artificial neural networks. Engineering Computations 24, 5–16.
Harvey, C., Travers, K., Costa, M., 2000. Forecasting emerging market returns using neural networks. Emerging Markets Quarterly 4, 43–55.
Hecht-Nielson, R., 1989. Theory of the back-propagation neural network. In: International Joint Conference on Neural Networks. Washington, DC, 593–605.
Hornborg, A., 1998. Towards an ecological theory of unequal exchange: Articulating world system theory and ecological economics. Ecological Economics 25, 127–136.
Huang, C., 2004. Class prediction of cancer using probabilistic neural networks and relative correlation metric. Applied Artificial Intelligence 18, 117–128.
Hunter, A., 2000. Feature selection using probabilistic neural networks. Neural Computing & Applications 9, 124–132.
Hurrion, R., Birgil, S., 1999. Comparison of factorial and random experimental design methods for the development of regression and neural network simulation models. Journal of the Operational Research Society 50, 1018–1033.
Ibric, S., Jovanovic, M., Djuric, Z., Parojcic, J., Solomun, L., 2002. The application of generalized regression neural network in the modeling and optimization of aspirin extended release tablets with Eudragit RS PO as matrix substance. Journal of Controlled Release 82, 213–222.
Jin, G., Englande, A., 2006. Prediction of swimmability in a brackish water body. Management of Environmental Quality 17, 197–208.
Jin, X., Srinivasan, D., 2001. Classification of freeway traffic patterns for incident detection using constructive probabilistic neural networks. IEEE Transactions on Neural Networks 12, 1173–1187.
Jorgenson, A., 2003. Consumption and environmental degradation: A cross-national analysis of the ecological footprint. Social Problems 50, 374–394.
Jorgenson, A., 2004. Uneven processes and environmental degradation in the world economy. Human Ecology Review 11, 103–113.
Jorgenson, A., 2005. Unpacking international power and the ecological footprints of nations: A quantitative cross-national study. Sociological Perspectives 48, 383–402.
Jorgenson, A., Burns, T., 2007. The political-economic causes of change in the ecological footprints of nations, 1991–2001: A quantitative investigation. Social Science Research 36, 834–853.
Kentor, J., 1998. The long term effects of foreign investment dependence on economic growth, 1940–1990. American Journal of Sociology 103, 1024–1046.
Kentor, J., 2000. Capital and Coercion: The Economic and Military Processes That Have Shaped the World Economy 1800–1990. Garland, New York.
Kim, B., Lee, B., 2005. Prediction of silicon oxynitride plasma etching using a generalized regression neural network. Journal of Applied Physics 98, 1–6.
Kohzadi, N., Boyd, M., Kemlanshahi, B., Kaastra, I., 1996. A Comparison of artificial neural network and time series models for forecasting commodity prices. Neurocomputing 10, 169–181.
Kumar, K., Bhattacharya, S., 2006. Artificial neural network vs. linear discriminant analysis in credit ratings forecast. Review of Accounting and Finance 5, 216–227.
Kutner, M., Nachtsheim, C., Netter, J., Li, W., 2005. Applied Linear Statistical Models, 5th ed. McGraw-Hill, Boston.
Lapedes, A., Farber, R., 1988. How neural nets work? In: Anderson, D. (Ed.), Neural Information Processing Systems. American Institute of Physics, New York, pp. 442–456.
Larrain, M., 2007. The PMI, the T-bill and inventories: A comparative analysis of neural network and regression forecast. Journal of Supply Chain Management 43, 39–51.
Laskari, E., Meletiou, G., Tasoulis, D., Vrahatis, M., 2006. Studying the performance of artificial neural networks on problems related to cryptography. Nonlinear Analysis: Real World Applications 7, 937–942.
Lee, T., Chiu, C., Chou, Y., Lu, C., 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. Computational Statistics & Data analysis 50, 1113–1130.
Lek, S., Guegan, J., 1999. Artificial neural networks as a tool in ecological modeling: An introduction. Ecological Modeling 120, 65–73.
Lim, C., Kirikoshi, T., 2005. Predicting the effects of physician-directed promotion on prescription yield and sales uptake using neural networks. Journal of Targeting, Measurement and Analysis for Marketing 13, 158–167.
Managi, S., 2006. Pollution, natural resource and economic growth: An econometric analysis. International Journal of Global Environmental Issues 6, 73–88.
McDonald, G., Patterson, M., 2004. Ecological footprints and interdependencies of New Zealand regions. Ecological Economics 50, 49–67.
McMenamin, J., Monforte, F., 1998. Short term energy forecasting with neural networks. Energy Journal 19, 43–52.
Mirmirani, S., Li, H., 2004. Gold price, neural networks and genetic algorithm. Computational Economics 23, 193–200.
Moreno, D., Marco, P., Olmeda, I., 2006. Self-organizing maps could improve the classification of Spanish mutual funds. European Journal of Operational Research 147, 1039–1054.
Mostafa, M., 2004. Forecasting the Suez Canal traffic: A neural network analysis. Maritime Policy and Management 31, 139–156.

McQuarrie, E., 1988. An alternative to purchase intentions: The role of prior behavior in consumer expenditure on computers. Journal of the Market Research Society 37, 407–437.
Nam, K., Yi, J., 1997. Predicting airline passenger volume. Journal of Business Forecasting Methods & Systems 16, 14–17.
Palisade Corporation, 2005. NeuralTools Professional User Guide (Version 1.0). Palisade Corporation, Ithaca, New York.
Parojcic, J., Ibric, S., Djuric, Z., Jovanovic, M., Corrigan, O., 2007. An investigation into the usefulness of generalized regression neural network analysis in the development of level A in vitro-in vivo correlation. European Journal of Pharmaceutical Sciences 30, 264–272.
Parzen, E., 1962. On the estimation of a probability density function and mode. Annals of Mathematical Statistics 33, 1065–1076.
Pedhazur, E., 1982. Multiple Regression in Behavioral Research: Explanation and Prediction. Holt Rinehart and Winston, New York.
Poh, H., Yao, J., Jasic, T., 1998. Neural networks for the analysis and forecasting of advertising impact. International Journal of Intelligent Systems in Accounting, Finance & Management 7, 253–268.
Prescott-Allen, R., 2001. The Wellbeing of Nations: A Country-by-Country Index of Quality of Life and the Environment. Island, Washington, DC.
Radhakrishnan, T., Nandan, U., 2005. Milling force prediction using regression and neural networks. Journal of Intelligent Manufacturing 16, 93–102.
Raitsos, D., Lavender, S., Maravelias, C., Haralabous, J., Richardson, A., Reid, P., 2008. Identifying four phytoplankton functional types from space: An ecological approach. Limnology and Oceanography 53, 605–613.
Rice, J., 2007. Ecological unequal exchange: International trade and uneven utilization of environmental space in the world system. Social Forces 85, 1369–1392.
Ruiz-Suarez, J., Mayora-Ibarra, O., Torres-Jimenez, J., Ruiz-Suarez, L., 1995. Short-term ozone forecasting by artificial neural network. Advances in Engineering Software 23, 143–149.
Ryu, H., Brody, S., 2004. Examining socio-economic/demographic, environmental value, land-use, and spatial impacts on the ecological footprint for Dallas, Texas. Paper Presented at the 45th Annual Conference of the Association of Collegiate Schools of Planning. Portland, Oregon.
Sakai, S., Kobayashi, K., Toyabe, S., Mandai, N., Kanda, T., Akazawa, K., 2007. Comparison of the levels of accuracy of an artificial neural network model and a logistic regression model for the diagnosis of acute appendicitis. Journal of Medical Systems 31, 357–364.
Salzman, J., 2000. Environmental protection beyond the smokestack: Addressing the impact of the service economy. Corporate Environmental Strategy 7, 20–37.
Sekaran, U., 2000. Research Methods for Business, 3rd ed. Wiley, New York.
Shan, Y., Zhao, R., Xu, G., Liebich, Zhang, Y., 2002. Application of probabilistic neural network in the clinical diagnosis of cancers based on clinical chemistry data. Analytica Chimica Acta 471, 77–86.
Sharda, R., 1994. Neural networks for the MS/OR analyst: An application bibliography. Interfaces 24, 116–130.
Shavlick, J., Mooney, R., Towell, G., 1991. Symbolic and neural learning: An experimental comparison. Machine Learning 6, 111–143.
Shie, J., 2008. Optimization of injection-molding for mechanical process of polypropylene components via a generalized regression neural network. Polymers for Advanced Technologies 19, 73–83.
Shin, U., Ucan, O., Bayat, C., Oztorun, N., 2005. Modeling of SO_2 distribution in Istanbul using artificial neural networks. Environmental Modeling and Assessment 10, 135–142.
Silver, H., Shmoish, M., 2008. Analysis of cognitive performance in schizophrenia patients and healthy individuals with unsupervised clustering models. Psychiatry Research 159, 167–179.
Singer, D., Bliss, J., 2003. Use of a probabilistic neural network to reduce costs of selecting construction rock. Natural Resources Research 12, 135–140.
Specht, D., 1991. A general regression neural network. IEEE Transactions on Neural Networks 2, 568–576.
Specht, D., 1990. Probabilistic neural networks. Neural Networks 3, 109–118.
Swicegood, P., Clark, J., 2001. Off-site monitoring systems for prediction bank underperformance: A comparison of neural networks, discriminant analysis, and professional human judgment. International Journal of Intelligent Systems in Accounting, Finance & Management 10, 169–186.
Tam, C., Tong, T., Lau, T., Chan, K., 2005. Selection of vertical framework system by probabilistic neural network models. Construction Management and Economics 23, 245–254.
Van Vuuren, D., Smeets, E., 2000. Ecological footprints of Benin, Bhutan, Costa Rica and the Netherlands. Ecological Economics 34, 115–130.
Venetoulis, J., 2001. Consuming the Earth: Money, values, land use, and ecological footprints in Los Angeles. Ph.D. Dissertation, Claremont Graduate University, Claremont, CA.
Videnova, I., Nedialkova, D., Dimitrova, M., Popova, S., 2006. Neural networks for air pollution forecasting. Applied Artificial Intelligence 20, 493–506.
Wackernagel, M., Rees, W., 1996. Our Ecological Footprint: Reducing Human Impact on the Earth. New Society Publishers, Gabriola Island, Canada.
Wang, S., 1995. The unpredictability of standard back propagation neural networks in classification applications. Management Science 41, 555–559.
Wasserman, P., 1993. Advanced Methods in Neural Computing. Van Nostrand-Reinhold, New York.
White, T., 2007. Sharing resources: The global distribution of the ecological footprint. Ecological Economics 64, 402–410.
Wilson, S., 2006. Algorithm architectures for patient dependent seizure detection. Clinical Neurophysiology 117, 1204–1216.
World Bank, 2002. World Development Indicators. World Bank, Washington DC.
WRI, 2000. (World Resources Institute). World Resources 2000–2001: People and Ecosystems: The Fraying Web of Life. Oxford University Press, New York.
Xue, C., Zhang, X., Liu, M., Hu, Z., Fan, B., 2005. Study of probabilistic neural networks to classify the active compounds in medicinal plants. Journal of Pharmaceutical and Biomedical Analysis 38, 497–507.
Yang, Z., Marjorie, B., 1999. Probabilistic neural networks in bankruptcy prediction. Journal of Business Research 44, 67–74.
York, R., Rosa, E., Dietz, T., 2003. Footprints upon the Earth: The environmental consequences of modernity. American Sociological Review 68, 279–300.
Zhuo, W., Li-Min, J., Yong, Q., Yan-hui, W., 2007. Railway passenger traffic volume prediction based on neural network. Applied Artificial Intelligence 21, 1–10.