

Ecological Footprint Model Using the Support Vector Machine Technique

Haibo Ma^{1,2*}, Wenjuan Chang², Guangbai Cui²

¹ College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang, China, ² College of Hydrology and Water Resources, Hohai University, Nanjing, China

Abstract

The per capita ecological footprint (EF) is one of the most widely recognized measures of environmental sustainability. It aims to quantify the Earth's biological resources required to support human activity. In this paper, we summarize relevant previous literature, and present five factors that influence per capita EF. These factors are: National gross domestic product (GDP), urbanization (independent of economic development), distribution of income (measured by the Gini coefficient), export dependence (measured by the percentage of exports to total GDP), and service intensity (measured by the percentage of service to total GDP). A new ecological footprint model based on a support vector machine (SVM), which is a machine-learning method based on the structural risk minimization principle from statistical learning theory was conducted to calculate the per capita EF of 24 nations using data from 123 nations. The calculation accuracy was measured by average absolute error and average relative error. They were 0.004883 and 0.351078% respectively. Our results demonstrate that the EF model based on SVM has good calculation performance.

Citation: Ma H, Chang W, Cui G (2012) Ecological Footprint Model Using the Support Vector Machine Technique. PLoS ONE 7(1): e30396. doi:10.1371/journal.pone.0030396

Editor: Attila Szolnoki, Hungarian Academy of Sciences, Hungary

Received: May 27, 2011; **Accepted:** December 15, 2011; **Published:** January 23, 2012

Copyright: © 2012 Ma et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no funding or support to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mahaibo29@163.com

Introduction

The ecological footprint (EF) approach was developed by Wackernagel and Rees [1]. It is calculated as the total area of bio-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 [2]. The EF approach provides a comprehensive unit of measurement that allows for comparisons of various types of consumption-based impacts [3]. Therefore, since its development the EF approach has become the most widely-used measure of environmental sustainability [4]. The EF approach aggregates typically complex resource use patterns into a single number [5]. The validity of the per capita EF, which traces the average amount of resources a person in a given country consumes, and the amount of waste they generate is confirmed by its significant correlation with important environmental impacts, for example, national emissions of ozone depleting substances, and nuclear power generation [6].

There are six resources considered by the EF: crop and pasture lands for production of goods and food, built land for construction, forest for the production of wood products, fossil energy for carbon dioxide emissions from fuels, and fish for food production. All of these are measured in global hectares (ha). A global hectare represents a hectare of land with global average bio-productivity. Social scientists and policymakers can compare the per capita EF of various nations to the per capita ecological capacity that exists on earth. For example, in 1996 the per capita EF ranged from 0.35 hectares to more than 16 hectares, and the majority of the estimated per capita EFs were higher than the Earth's bio-capacity per capita [7]. According to McDonald and Patterson [8], the

global EF is at least 30% larger than the Earth's bio-capacity, illustrating the severity of resource overuse. EF figures can also be used as benchmarks for assessing sustainability at a national level, for example, nations with an EF at or below 1.8 hectares per capita have a global impact that could be replicated by other nations without threatening long-term sustainability [2].

Although the EF model has been used at various levels, including global [9], municipal [10], national [9], city [11] and individual [12], no previous studies have attempted to apply a support vector machine (SVM) to predict national EF. In this paper, we seek to fill this research gap by calculating the EF of 23 nations through the use of SVM techniques. The countries analyzed in this study are listed in the Appendix S1. More specially, the purpose of this research is twofold:

First, to determine the major factors influencing national EFs, and second, to build a SVM model based on these identified factors to calculate EF.

Materials and Methods

Materials

Drawing on previous research, we found a wealth of evidence suggesting that a variety of factors influence EF. Cross-sectional analyses consistently show that national per capita ecological footprints are largely a function of gross domestic product (GDP) [13,14,15]. A negative relationship between per capita EF and export dependence (measured as the proportion of total GDP generated by exports) has also been identified [15]. According to Jorgenson and Burns [16], nations with a greater intensity in the services sector experience higher increases in per capita EF.

Some evidence suggests that domestic income inequality is negatively related to the relative size of a nation's per capita EF [16]. Jorgenson (2003) found that urbanization has a positive impact on EF [13]. From the above, it can be seen that the factors that influence EF can be characterized as affluence (as measured by GDP), export dependence, service intensity, domestic income inequality, and urbanization.

Methodology

The SVM is a machine-learning method based on the structural risk minimization principle from statistical learning theory. It maps input data x into a higher-dimensional feature space Q by non-linear mapping to yield and solve a linear regression problem in this feature space [17]. The regression approximation addresses the problem of estimating a function based on a given set $\{(x_i, y_i)\}_i^n$, where x_i denotes the input vector, y_i denotes the output value, and n denotes the total number of data patterns. In SVM, the regression function is given as the following:

$$g(x) = w * \varphi(x) + d \quad (1)$$

Where d is a scalar threshold, w is the weight vector, and $\varphi(x)$ is the high-dimensional feature space that is nonlinearly mapped from the input space x .

Support vector regression (SVR) performs linear regression in the high-dimensional feature space by ϵ -insensitive loss. At the same time, to prevent over-fitting and thereby improving the generalization capability, the following regularized functional involving summation of the empirical risk and a complexity term $\|w\|^2/2$, is minimized. The coefficients w and d can be estimated by minimizing the regularized risk function:

$$\text{Min } \|w\|^2/2$$

$$\text{s.t. } \begin{cases} y_i - \varphi(\omega, x_i) - d \leq \lambda \\ \varphi(\omega, x_i) + d - y_i \leq \lambda \end{cases} \quad (2)$$

The regression problem is transformed into the following constrained formation:

$$\begin{aligned} \text{Min } & \frac{1}{2} \|w\|^2 + \rho \sum_{i=1}^n (\delta_i + \delta_i^*) \\ \text{s.t. } & \begin{cases} y_i - \langle w, x_i \rangle - d \leq \lambda + \delta_i & \delta_i \geq 0 \\ \langle w, x_i \rangle + d - y_i \leq \lambda + \delta_i^* & \delta_i^* \geq 0 \end{cases} \end{aligned} \quad (3)$$

Where the constant ρ stands for the penalty degree of the sample with error exceeding λ . Two positive slack variables δ and δ^* represent the distance from actual values to the corresponding boundary values of $\lambda - tube$.

A dual problem can then be derived by using the optimization method to maximize the function:

$$\begin{aligned} \text{Max } & \sum_{i=1}^n y_i (v_i - v_i^*) - \lambda \sum_{i=1}^n (v_i + v_i^*) - \\ & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (v_i - v_i^*) (v_j - v_j^*) K(x_i, x_j) \\ \text{s.t. } & \sum_{i=1}^n (v_i - v_i^*) = 0 \quad \text{and} \quad 0 \leq v_i, v_i^* \leq \rho \end{aligned} \quad (4)$$

Where v_i and v_i^* are the Lagrange multiplier.

The SVM for function fitting obtained by using the above mentioned maximization function is then given by the following function:

$$g(x) = \sum_{i=1}^n (v_i - v_i^*) K(x_i, x) + d \quad (5)$$

In Equation 5, sample points that appear with non-zero coefficients are the so-called support vectors. The kernel function $K(x_i, x_j) = \varphi(x_i) \varphi(x_j)$ satisfies Mercer's conditions and performs the non-linear mapping.

Results and Discussion

Preliminary data analysis

In this study, per capita EF was taken from White [18,19], the latest data on national level per capita EF, and GDP data were taken from the World Bank [20]. To correct for excessive skewness, we use the natural logarithm transformation of GDP data. Export data as a percentage of total GDP were taken from the World Bank [20] and used as a measure of export intensity and export dependence. Service data as a percentage of total GDP were taken from the World Bank [20] and used as an indicator of service intensity. Domestic income inequality data were taken from the World Bank [20] and are presented as GINI coefficients, which measure the distribution of income within countries. A GINI index score of zero suggests perfect equality, while an index score of 100 suggests perfect inequality. Urbanization data were taken from the World Bank [20], and are measured as the percentage of the total population living in cities, which represents a country's relative level of urbanization. Following Jorgenson and Burns [16], we regress these data on per capita GDP and use the residuals as measures of urbanization to minimize collinearity.

Table 1 provides descriptive statistics for all variables used in the analysis. The product moment correlations between variables are shown in the Table 2. Although correlations do not prove causation, they can be used to generate hypotheses; therefore Table 2 is presented to highlight the correlations among the five variables used for analysis. It indicates that most of the correlations were significant and in the expected direction.

SVM analysis

We used data of 123 countries (shown in Appendix S1) to establish and test the SVM-based model. We used data of 99

Table 1. Descriptive statistics of all the variables used in this study (N = 123).

Variable	Mean	SD	Minimum	Maximum
EF	2.436	1.9517	0.52	9.59
GDP(ln)	7.618	1.6658	4.449	10.805
Service (%)	54.066	13.003	20.542	78.53
Export (%)	36.311	18.280	7.272	83.83
Gini	38.931	10.040	19.5	62.9
urban	0.441	1.053	0.001	7.296

Note: EF = ecological footprint; GDP(ln) = gross domestic product (the natural logarithm transformation); service (%) = service as the percentage of GDP; export (%) = export as the percentage of GDP. Gini = national income disequilibrium; urban = urbanization level (residualized). doi:10.1371/journal.pone.0030396.t001

Table 2. Product moment correlations matrix.

	1	2	3	4	5	6
1.GDP(ln)	1					
2.urban(residualized)	0.423**	1				
3.Service (%)	0.729**	0.257**	1			
4.GINI	-0.394**	-0.255**	-0.241**	1		
5.Export(%)	0.284**	0.304	0	-0.202*	1	
7.EF	0.860**	0.347**	0.613**	-0.432**	0.218*	1

**correlation is significant at the level of 0.01 (2-tailed).

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

doi:10.1371/journal.pone.0030396.t002

countries (80% of the data) to establish the model, and 24 countries (20% of the data) to test the accuracy of the model. For the 24 countries used to test model accuracy, the calculations were conducted in alphabetical order; therefore the results are presented alphabetically. When the correct model was established, there was no need for further “training” or “test” data. The model only required five variables to calculate a nation’s EF. In addition, the model was designed to achieve short calculation times. Therefore, compared to traditional EF techniques, the SVM technique was very easy to apply.

According to the method of Liu, Zhuang, and Liu [17], we used the particle swarm optimization technique to choose the optimal parameters for the SVM model. The optimal parameters are as follows: $\rho = 1000$, $\delta = 513$, $\lambda = 0.001$. The EF model was then determined by these three parameters and the data of 99 countries. Following this, we used the model to calculate the EF

of the other 24 countries. Model accuracy was measured by absolute and relative error. The calculation performance is displayed in Figure 1. The calculation results are presented in Table 3. Figure 1 and Table 3 show that the EF model based on SVM can calculate EF perfectly. The average absolute error is 0.004883, and the average relative error is only 0.351078%. Therefore, we were successful in establishing an EF model, and we can use it to calculate the EF of any nation using only five nation-specific variables.

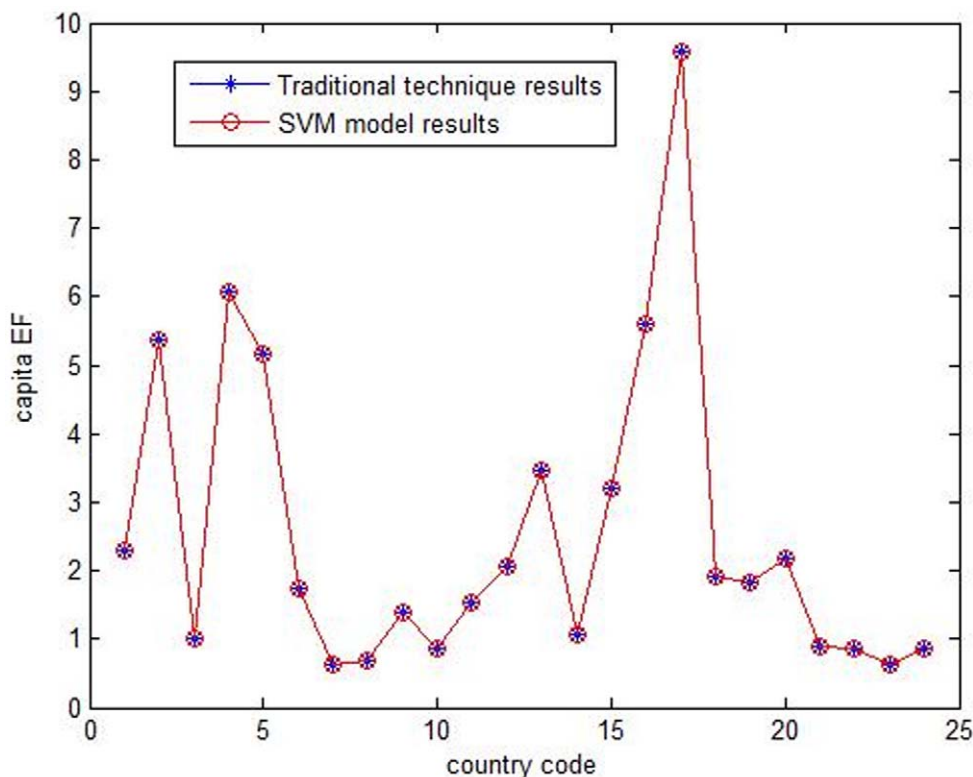
According to Table 2, we can see that the product moment correlation between GDP (ln) and EF was 0.860. We constructed a least-squares regression model and obtained the following equation:

$$EF = 1.008 * GDP(\ln) - 5.242 \quad (6)$$

The average absolute error from least-squares regression is 0.7620, and the average relative error is 44.66%. These are bigger than the errors derived from the EF model using the SVM technique with five variables. So, we can assume that the additional four variables are useful for the calculation of EF.

Implications, limitations, and future research

Our results demonstrate that national level per capita EF is influenced by the nation’s GDP, urbanization level, distribution of income (measured with the Gini coefficient), export dependency (as a percentage of total GDP), and service intensity (as a percentage of total GDP). Using these five variables, we established an SVM model to calculate EF. Compared with the traditional technique, the SVM model required less variables, and had a quicker calculation time. Therefore, the SVM technique is very easy to apply.

**Figure 1.** SVM calculation performance of per capita EF.

doi:10.1371/journal.pone.0030396.g001

Table 3. Calculation results by SVM.

Country name	Traditional technique results	SVM model results	Absolute error	Relative error(%)
South Africa	2.29	2.2946	0.0046	0.200873
Spain	5.36	5.3551	0.0049	0.09142
Sri lank	1	1.0046	0.0046	0.46
Sweden	6.07	6.0664	0.0036	0.05931
Switzerland	5.15	5.1442	0.0058	0.11262
Syria	1.73	1.7364	0.0064	0.369942
Tajikistan	0.64	0.6442	0.0042	0.65625
Tanzania	0.7	0.7069	0.0069	0.985714
Thailand	1.38	1.3866	0.0066	0.478261
Togo	0.87	0.8742	0.0042	0.482759
Tunisa	1.54	1.5469	0.0069	0.448052
Turkey	2.06	2.0629	0.0029	0.140777
Turkmenistan	3.47	3.4666	0.0034	0.09798
Uganda	1.08	1.0853	0.0053	0.490741
Ukraine	3.19	3.1847	0.0053	0.16614
United kingdom	5.59	5.5842	0.0058	0.10376
United states	9.59	9.5843	0.0057	0.05944
Uruguay	1.92	1.9231	0.0031	0.161458
Uzbekistan	1.83	1.8346	0.0046	0.251366
Venezuela	2.18	2.1832	0.0032	0.146789
Vietnam	0.88	0.8846	0.0046	0.522727
Yemen	0.85	0.8542	0.0042	0.494118
Zambia	0.63	0.6354	0.0054	0.857143
zimbabwe	0.85	0.8550	0.005	0.588235
average	2.535417	2.537425	0.004883	0.351078

doi:10.1371/journal.pone.0030396.t003

Despite the significant contributions of this study, it is subject to a number of limitations. First, this study used a cross-sectional rather than a longitudinal method. Much more emphasis was placed on observing national-level EFs than on observing changes in global EF. Much more emphasis should be placed on longitudinal research to focus on observing changes in EF behavior over time. Second, we only considered five factors that influenced per capita EF. In the future, we will explore other factors influencing per capita EF.

As a new approach to measuring sustainability, EF analysis has been more successful than others. Inevitably, the approach is not without its flaws [21,22]. However, its theory and application will be improved with continued study and with refinements the methodology used by organizations responsible for environmental reporting and management.

References

- Wackernagel M, Rees W (1996) Our Ecological Footprint Reducing Human Impact on the Earth New Society Publishers.
- Mostafa MM (2010) A Bayesian approach to analyzing the ecological footprint of 140 nations. *Ecological Indicators* 10: 808–817.
- Wackernagel M, Silverstein J (2000) Big things first: focusing on the scale imperative with the ecological footprint. *Ecological Economics* 32: 391–394.
- Binningsbo H, de Soya I, Gleditsch N (2007) Green giant or straw man? Environmental pressure and civil conflict. *Population and Environment* 28: 337–353.
- Costanza R (2000) The dynamics of the ecological footprint concept. *Ecological Economics* 32: 341–345.
- WRI(World Resources Institute) (2000) World Resources 2000–2001: People and Ecosystems: The Fraying Web of Life. New York.
- Jorgenson AK (2006) Global warming and the neglected greenhouse gas: a cross-national study of the social causes of methane emissions intensity, 1995. *Social Forces* 84: 1777–1796.
- McDonald GW, Patterson MG (2004) Ecological footprints and interdependencies of New Zealand regions. *Ecological Economics* 50: 49–67.
- Rice J (2007) Ecological unequal exchange: international trade and uneven utilization of environmental space in the world system. *Social Forces* 85: 1369–1392.
- Barrett J, Scott A (2003) “The application of the ecological footprint”: a case of passenger transport in Merseyside. *Local Environment* 8: 167–183.
- Li XM, Xiao RB, Yuan SH, Chen JA, Zhou JX (2010) Urban total ecological footprint forecasting by using radial basis function neural network: A case study of Wuhan city, China. *Ecological Indicators* 10: 241–248.

Supporting Information

Appendix S1 Countries analyzed in the study.
(DOC)

Acknowledgments

We appreciate anonymous reviewers for useful suggestions and precious insight.

Author Contributions

Conceived and designed the experiments: HM. Performed the experiments: WC. Analyzed the data: GC. Contributed reagents/materials/analysis tools: HM. Wrote the paper: HM. Design the ecological footprint model: HM.

12. 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.
13. Jorgenson A (2003) Consumption and environmental degradation: a cross-national analysis of the ecological footprint. *Social Problems* 50: 374–394.
14. Jorgenson A (2004) Uneven processes and environmental degradation in the world economy. *Human Ecology Review* 11: 103–113.
15. Jorgenson A (2005) Unpacking international power and the ecological footprints of nations: a quantitative cross-national study. *Sociological Perspectives* 48: 383–402.
16. 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.
17. Liu LX, Zhuang YQ, Liu XY (2011) Tax forecasting theory and model based on SVM optimized by PSO. *Expert Systems with Applications* 38: 116–120.
18. White TJ (2007) Sharing resources: The global distribution of the Ecological Footprint. *Ecological Economics* 64: 402–410.
19. WWF International (2006) *Living Planet Report*, Avenue du Mont-Blanc CH-1196 Gland, Switzerland.
20. World Bank (2002) *World Development Indicators*. Washington, DC.
21. Phil Mcmanus, Graham Haughton (2006) Planning with Ecological Footprints: a sympathetic critique of theory and practice 18: 113–127.
22. DietzSimon, NeumayerEric (2007) Weak and strong sustainability in the SEEA: Concepts and measurement 61: 617–626.