

Development and implementation of the numerical model for predicting the values of ecological footprint, based on the Monte Carlo methodology

Radmila Jankovic

Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

rjankovic@mi.sanu.ac.rs

Ivan Mihajlovic

Technical faculty in Bor, University of Belgrade, Serbia,
imihajlovic@tfbor.bg.ac.rs

Abstract: The impact of human activities on the environment can be observed through the ecological footprint - the biologically productive area of land and water that is needed to satisfy human demands. In order to achieve and maintain sustainability, Earth's natural capital needs to be preserved. Thus, it is of high importance for scientific and general community to analyze and predict the ecological footprint in order to successfully manage natural resources and protect the environment. The aim of this paper is to develop and implement a numerical model based on Monte Carlo methodology, for predicting the values of ecological footprint (EF). The model is based on systematic analysis of six input variables: (1) Rural population, (2) Urban population, (3) GDP per capita, (4) Energy use, (5) Electric power consumption, (6) Electricity production, and one output variable which is the total ecological footprint of consumption. The dataset included data from European, North American, South American, Asian and African countries, as well from Australia. Predicted values from the model were then compared with the measured ones, in order to verify the accuracy of the model.

Keywords: ecological footprint, Monte Carlo, numerical model

1 Introduction

Sustainability has become a major concern of today's society. All countries worldwide need to deal with environmental problems [1] especially those affecting climate change. One way to assess the Earth's sustainability is by calculating the Ecological Footprint.

The Ecological Footprint can be defined as a technique that measures the limits of our planet and the extent to which humanity is exceeding those limits [2]. Basically, the ecological footprint analyses the relation between the demand and supply of environmental resources, where the EF considers the demand that humanity places on the biosphere, while biocapacity presents the amount of biologically productive area that can be used for the needs of humanity [2]. Environmental resources here include: cropland used for providing plant-based foods and fibers, grazing land for providing animal products, fishing grounds for fish products, forests for timber and forest products, uptake land for the neutralization of anthropogenic carbon dioxide emissions, and built-up land for infrastructure [2].

Presenting it simple, the ecological footprint represents the material balance of the Earth, observed as closed system. This way, a state of ecological deficit occurs when the demand placed on the biosphere exceeds the region's biocapacity. Unfortunately, at the global level, an equivalent of 1.7 Earths is used to provide the present resources that humanity needs [3].

Industrial activities are based on smart and automated industrial concepts, and in most contemporary conditions, on Industry 4.0 concept which includes cyber-physical systems, the Internet of Things, cloud computing artificial intelligence, and cognitive computing. However, even such processes require resources and especially energy. As industry becomes more environmentally friendly, consequently it also becomes more energy dependent. This way, the pollution and environmental burden is now transferred from the production to the consumption sphere, especially when it comes to energy consumption.

The aim of this research is to develop adequate prediction model for Ecological Footprint estimation, based on energy related parameters, and then validate the model by using a Monte Carlo simulation. The EF data is already available in different publicly available data sources, so the main motive here is not the calculation itself, but accessing the dependencies between the socio-demographic and energy related inputs, and the total Ecological Footprint of consumption.

The paper is organized as follows. Section 2 presents a review of the literature, while Section 3 describes data and methodology used in this research. Section 4 shows the obtained results from both models. The last section presents conclusions of this study.

2 Literature review

In reference [4], author used a Bayesian linear regression model and Markov Chain Monte Carlo for simulation to predict the ecological footprint of 140 nations. The results indicate that urbanization level along with the world system position positively affect the ecological footprint per capita, while income is negatively related. Despite these results, the author emphasized the importance of a longitudinal approach, rather than a cross-sectional one which was used in the study.

In a paper [5], the authors used the Markov chain to estimate the ecological footprint of Beijing. The results indicated that consumption patterns and environmental policies kept ecological footprint stable, while energy consumption was the biggest contributor to ecological footprint. Moreover, population growth and urbanization level impact the intensity of ecological footprint. The authors suggested the improvement of urban design, energy efficiency improvements, and changes of consumption patterns.

In [6] the environmental sustainability of China was investigated and predicted future ecological footprint using the linear autoregressive integrated moving average method (ARIMA) and the artificial neural network (ANN). Their results indicate that the ecological footprint in China will continue with the rising trend, while the overall ecological security is predicted to continue to worsen. The authors also state that a combination of ARIMA and ANN models can make the prediction results more reliable [6].

The authors in [7] predicted ecological footprint in Europe using Proportional-Odds Cumulative Logistic regression, based on innovation factors, the degree of economic freedom, and whether or not the country is a member of European Union. Their analysis indicated that ecological footprint depends on the employment in foreign controlled enterprises, eco-innovation index, and region. Moreover, it is concluded that Europe's ecological footprint from 2006 to 2014 decreased, while the biocapacity increased. It is also found that Luxembourg had the highest ecological footprint, followed by Denmark and Estonia. On the contrary, the lowest ecological footprint was found for non-EU countries characterized by a lower economic development, with the lowest values of ecological footprint in Moldova and Albania. Among the EU members, Romania, followed by Bulgaria, had the lowest ecological footprint level [7].

In paper [8], the authors analyzed the ecological footprint of Beijing using the support vector machine (SVM), where a novel model was introduced based on which the prediction of ecological footprint was made for the period from 2016 to 2020. The results indicate that SVM achieved higher prediction accuracy than BPNN. The ecological footprint of Beijing is indicated to increase by 2020 [8].

3 Data and methodology

3.1 Data

The ecological footprint data, to be used for analysis and modeling, was downloaded from the Global Footprint Network [3]. The data comprises of six land use types: (1) Cropland, (2) Grazing land, (3) Fishing grounds, (4) Forest Land, (5) Build-up land, and (6) Carbon uptake land. Ecological footprint can be viewed from the aspect of production (EF of production), and consumption (EF of consumption). The ecological footprint of production can be calculated as following [2]:

$$EF_p = \sum_i \frac{P_i}{Y_{N,i}} \cdot YF_{N,i} \cdot EQF_i = \sum_i \frac{P_i}{Y_{W,i}} \cdot EQF_i \quad (1)$$

Where P is the amount of each primary product i harvested in the nation, $Y_{N,i}$ is the annual average yield for the production of commodity i , $YF_{N,i}$ is the country specific yield factor for the production of each product i , $Y_{W,i}$ is the average world yield for commodity i , and EQF_i is the equivalence factor for the land use type producing products i [2].

The ecological footprint of consumption can be calculated as follows [2]:

$$EF_c = EF_p + EF_i - EF_e \quad (2)$$

Where EF_p is the ecological footprint of production, EF_i is the imported commodity flow, and EF_e is exported commodity flow [2].

The ecological footprint of consumption is the most commonly reported EF [2]. It is also worth noticing that the Ecological Footprint is expressed in global hectares (gha) [2].

Besides the total ecological footprint of consumption, other parameters were also collected for the analyses: Population number (urban, rural, and total), GDP per capita (in constant 2010 USD), Energy use (kg of oil equivalent per capita), Electric power consumption (kWh per capita), and Electricity production (% of total) from six different sources (coal, hydroelectric, natural gas, nuclear, oil, and renewable sources). This additional data was downloaded from World Bank reports [9]. Based on different Electricity production sources, six datasets were created, next to the original dataset which consists of 3133 datalines, and involves 72 countries from all over the world for the period of 1972-2014.

In this paper, the dataset containing the total EF of consumption as the output variable, and Population numbers, GDP per capita, Energy use, Electric power consumption, and Electricity production from coal sources, as input variables, was used for the analysis. This dataset consist of 1353 datalines, and involves only

those values for which the Electricity production from coal sources exceeds 4.99%.

The purpose of this study is to predict the behavior of the Ecological Footprint based on different parameters. Hence, the total Ecological Footprint of consumption represents an output variable, while socioeconomic and energy related parameters represent the input variables.

3.2 Methodology

Correlation analysis was firstly performed in order to access the existence and strength of the linear relationship between variables. Then, a multiple linear regression model was developed, and a model equation was formed which was later used for further analysis. Multiple linear regression analysis creates a relation between the input variables (predictors) and the dependent variable, and can be represented using the following equation [10]:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i \quad (3)$$

Where Y_i is the dependent variable, β_0 represents the intercept, β_1 is the coefficient of the independent variable, x_i is the predictor variable, and ε_i is the random error [10].

An artificial neural network model (ANN) was also formed in order to compare and evaluate the best prediction performance between nonlinear (ANN) and linear (MLRA) modeling. The multilayer perceptron is a type of artificial neural network that was used, which consists of input layers, hidden layers, and output layers [11]. The neurons are interconnected except the neurons that are in the same layer. The networks works in a way that the input layer receives the information, then it multiplies the received values by weights, and passes the information to the hidden layer which processes it. Lastly, the output layer makes predictions [12].

Lastly, a Monte Carlo simulation was used to validate and to derive final conclusions on the efficiency of each model for predicting the Ecological Footprint. The analyses were performed using SPSS v.24, and Distribution Analyzer software.

4 Analyses and results

First, a correlation analysis was performed for all variables, excluding Electricity production from all other sources, except from coal. The results of correlation analysis are represented in Table 1.

The correlation analysis indicates that the strongest relationship occurs between the Total Ecological Footprint (TEF) and Urban Population (UP) ($r=0.893$, $p<0.001$). Moreover, there is a weak relationship between the total ecological footprint and Energy Use (EU) ($r=0.180$, $p<0.001$), as well as between TEF and Electricity production from coal sources (EPCoal) ($r=0.263$, $p=0.001$). There is no statistically significant relationship between TEF and GDP.

Based on the correlation analysis and literature findings, this analysis involves (1) Urban population, (2) Energy use, (3) GDP, and (4) Electricity production from coal sources, as input variables, and (5) Total Ecological Footprint as an output variable. Previous research indicates that the GDP per capita is an important factor which can have a valuable impact on sustainability [13,14,15] hence, this parameter was also retained for the analysis.

Table 5 Correlation analysis results

Correlations		TEF	UP	GDP	EU	EPCoal
TEF	r	1	.893**	.013	.180**	.263**
	p		.000	.635	.000	.000
	N	1353	1353	1353	1352	1353
UP	r	.893**	1	-.168**	-.064*	.291**
	p	.000		.000	.019	.000
	N	1353	1353	1353	1352	1353
GDP	r	.013	-.168**	1	.805**	.008
	p	.635	.000		.000	.759
	N	1353	1353	1353	1352	1353
EU	r	.180**	-.064*	.805**	1	.088**
	p	.000	.019	.000		.001
	N	1352	1352	1352	1352	1352
EPCoal	r	.263**	.291**	.008	.088**	1
	p	.000	.000	.759	.001	
	N	1353	1353	1353	1352	1353

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

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

r – Pearson Correlation, p – Sig. (2-tailed), N –number of datalines

TEF - total Ecological Footprint, UP - Urban Population, RP - Rural Population, GDP - GDP per capita (constant 2010 USD), EU - Energy use (kg of oil equivalent per capita), EPC - Electric power consumption (kwh per capita), EPCoal - Electricity production from coal sources (% of total, only values above 4.999%).

4.1 MLRA

In order to predict the values of Ecological footprint, a Multiple Linear Regression Analysis was performed. The analysis included the total ecological footprint as dependent variable, and four predictors: Electricity production from coal sources, GDP per capita, Urban population, and Energy use.

The results of the analysis indicate that the Electricity production from coal sources, GDP per capita, Urban population, and Energy use explain 85.6% of the variance ($r^2=0.856$, $p<0.000$). Table 2 presents the results of the MLRA.

The regression equation is:

$$Total\ EF = -1.643 \cdot 10^8 + 6.579 \cdot UP - 3061.341 \cdot GDP + 88477.866 \cdot EU - 717260.214 \cdot EPCoal \quad (4)$$

Table 6 Results of MLRA

Coefficients ^a						
Model	Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
	B	Std. Error			Tolerance	VIF
(Constant)	-1.643E8	1,491E7	-11.015	.000		
UP	6.579	.080	82.181	.000	.879	1.137
GDP	-3061.341	682.664	-4.484	.000	.338	2.961
EU	88477.866	5135.478	17.229	.000	.343	2.915
EPCoal	-717260.214	287439.40	-2.495	.013	.902	1.109

$r=0.925$, $r^2=0.856$, $F=1999.499$, $p<0.001$.

- Dependent Variable: total EF
- Predictors: (Constant), Urban Population (UP), GDP per capita (GDP), Energy use (EU), Electricity production from coal sources (EPCoal)

Figure 1 shows the scatterplot of the regression adjusted predicted values of total Ecological Footprint, and actually measured values of total Ecological Footprint. The data fits the model well with $r^2=0.853$.

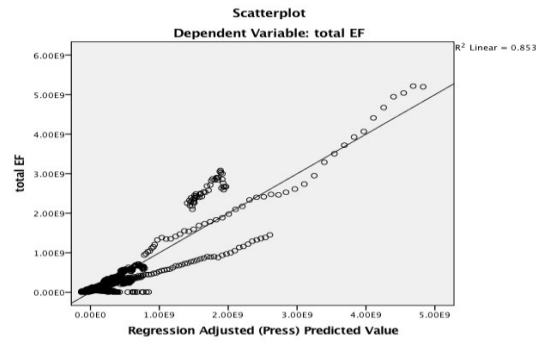


Figure 1 The scatterplot of the MLRA

4.2 ANN

An artificial neural network method was developed and compared to the MLR model [16]. The dataset was first divided into 70% training and 30% testing set, and then the Multilayer Perceptron method was applied. Sigmoid activation function was used. Figure 2 shows the layers of the network.

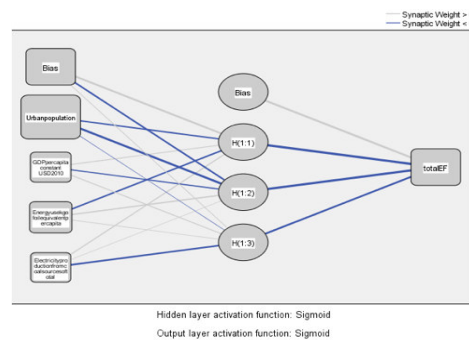


Figure 2 The architecture of the ANN

Regarding the importance of each input variable, Urban population is observed to be of highest importance, followed by the Energy use, GDP per capita, and Electricity production from coal sources (Table 3).

Table 7 Independent variable importance

	Importance	Normalized Importance
Urban population	0.821	100.0%
GDP per capita (constant USD 2010)	0.045	5.5%
Energy use (kg of oil equivalent per capita)	0.112	13.7%
Electricity production from coal sources (% of total)	0.022	2.7%

4.3 Testing the prediction potential of both models

In order to access the prediction potential of prediction models, both ANN, and MLRA models have been tested on a dataset including data only regarding China. China was selected for this experiment because of excellent model fitting. Figure 3 shows the actual and predicted values of the total Ecological Footprint. As observed, the prediction potential of the MLR model is better than the ANN model, but both models can still be effectively used for Ecological Footprint prediction.

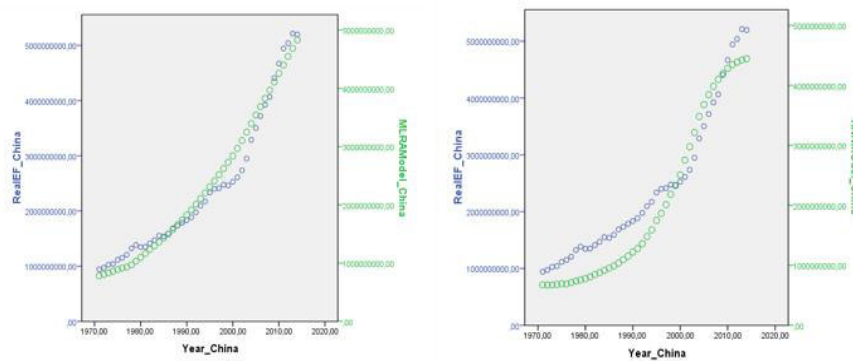


Figure 3 Prediction potential of the models for China

4.4 Monte Carlo Simulation

The process of making predictions necessarily involves a significant uncertainty which can be explained by using probability distributions. This way the variables can have different probabilities that different outcomes may occur. In doing so, Monte Carlo simulations can be performed. The first step of Monte Carlo analysis is to define the most suitable distribution of each variable. Figure 4 shows these distributions. Based on defined distributions, random values for all input variables were generated. Using the developed prediction models, values of the dependent variable Y (EF) were calculated for all randomly generated inputs.

From Fig. 4 it can be observed that Urban population variable and total EF variable best fit Pearson family distributions, while GDP per capita fits the Negative Exponential distribution. The energy use variable and the electricity production from coal sources best fit the Johnson family distribution.

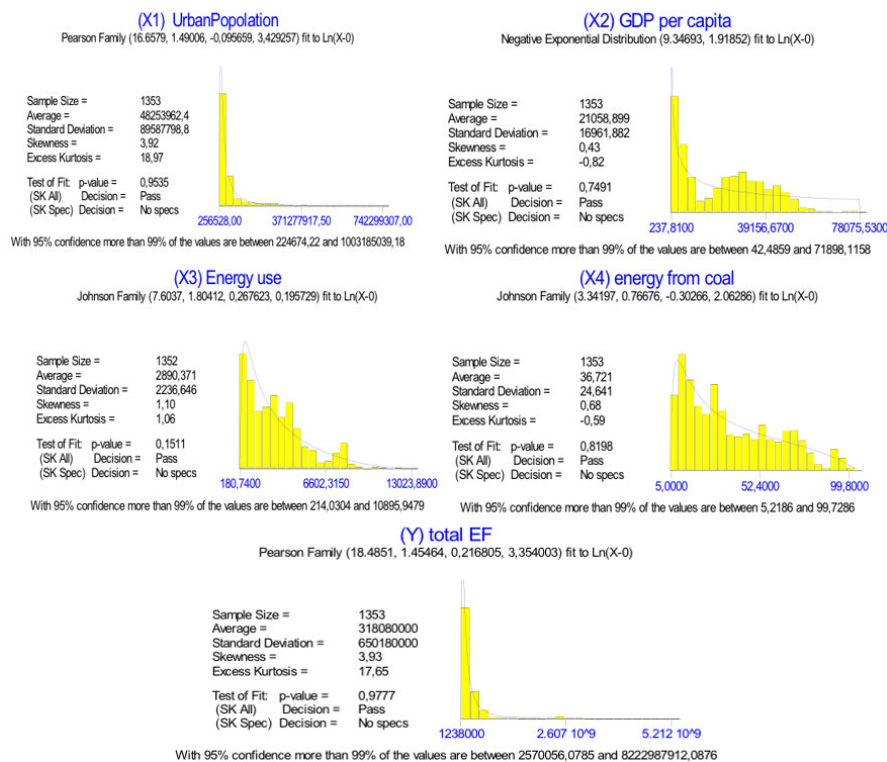


Figure 4 The distribution of each variable

After the Monte Carlo simulation, and based on the MLR equation, the predicted values of total EF were calculated, and are shown in figure 5.

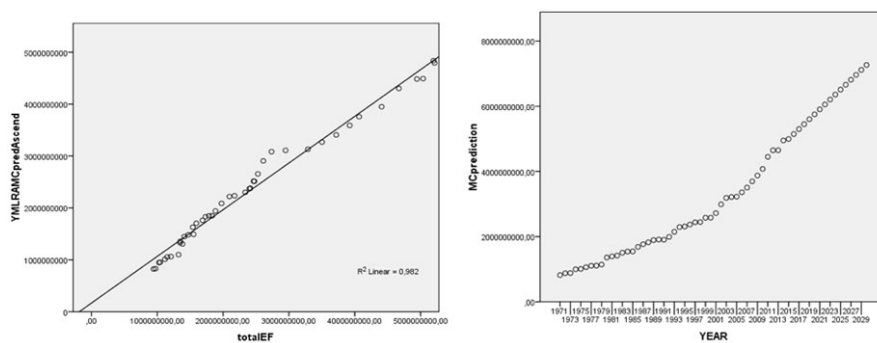


Figure 5 The predicted values of total EF of consumption per capita

Conclusions

Ecological Footprint is the measure of the human activity toward further progress, and also toward further exhausting of the biological potential for renewal and

recuperation. This fact makes EF an important part of sustainability, and hence a critical factor in improving ecological conditions worldwide.

In this research, EF was not calculated, but rather its dependence on important inputs was assessed. Two models were created, MLRA and ANN, and further used for making predictions of the total Ecological footprint of consumption, based on the GDP per capita, Energy use, Urban population, and Electricity production from coal sources. Monte Carlo Simulation was used for the validation of both models.

The numerical models developed in this research are proved to be efficient for predicting the Ecological Footprint, based on observed importance of potential inputs. The Multiple Linear Regression model performed slightly better than the ANN, with more accurate predictions.

Acknowledgement

This work was supported by the Mathematical Institute of the Serbian Academy of Sciences and Arts (Project III44006).

References

- [1] Charfeddine, L., Al-Malk, A.Y. and Al Korbi, K. Is it possible to improve environmental quality without reducing economic growth: Evidence from the Qatar economy. *Renewable and Sustainable Energy Reviews*, 82, 2018, pp.25-39.
- [2] Borucke, M., Moore, D., Cranston, G., Gracey, K., Iha, K., Larson, J., Lazarus, E., Morales, J.C., Wackernagel, M. and Galli, A., Accounting for demand and supply of the biosphere's regenerative capacity: The National Footprint Accounts' underlying methodology and framework. *Ecological indicators*, 24, 2013, pp.518-533.
- [3] Global Footprint Network. Ecological footprint. Footprint Network web site: <http://www.footprintnetwork.org>, 2018.
- [4] Mostafa, M. M., A Bayesian approach to analyzing the ecological footprint of 140 nations. *Ecological Indicators*, 10(4), 2010, pp. 808-817.
- [5] Lu, Y., & Chen, B., Urban ecological footprint prediction based on the Markov chain. *Journal of Cleaner Production*, 163, 2017, pp. 146-153.
- [6] Wang, Z., Yang, L., Yin, J., & Zhang, B., Assessment and prediction of environmental sustainability in China based on a modified ecological footprint model. *Resources, Conservation and Recycling*, 132, 2018, pp. 301-313.
- [7] Ghita, S., Saseanu, A., Gogonea, R. M., & Huidumac-Petrescu, C. E., Perspectives of Ecological Footprint in European Context under the Impact of Information Society and Sustainable Development. *Sustainability*, 10(9), 2018, pp. 3224.

- [8] Liu, L., & Lei, Y., An accurate ecological footprint analysis and prediction for Beijing based on SVM model. *Ecological Informatics*, 44, 2018, pp. 33-42.
- [9] The World Bank, World Bank Open Data, 2018, Retrieved from <https://data.worldbank.org>
- [10] Eberly, L.E. "Multiple linear regression." In *Topics in Biostatistics*, Humana Press, 2007, pp. 165-187.
- [11] Haykin, S.S. *Neural networks and learning machines*. Vol. 3. Upper Saddle River, NJ: Pearson, 2009.
- [12] Zacharis, N. Z., Predicting student academic performance in blended learning using Artificial Neural Networks. *International Journal of Artificial Intelligence and Applications*, 7 (5), 2016, pp. 17-29.
- [13] Rainham, D. & Mcdowell, I., The Sustainability of Population Health. *Popul. Environ.* 26. 2005, pp. 303-324. 10.1007/s11111-005-3344-9.
- [14] Uddin, G. A., Alam, K., & Gow, J., Does ecological footprint impede economic growth? An empirical analysis based on the environmental Kuznets curve hypothesis. *Australian Economic Papers*, 55(3), 2016, pp.301-316.
- [15] Feng, Z., and Wu. J.: Relationship between the ecological footprint and the economic growth in China, in *International Conference on Remote Sensing, Environment and Transportation Engineering*, Nanjing, China, 2011, pp. 6205-6211.
- [16] Kalina, J.: High-dimensional data in economics and their (robust) analysis. *Serbian Journal of Management*, 12(1), 2017, pp. 157-169.