A Two-Phase Model for the Evaluation of Urbanization Impacts on Carbon Dioxide Emissions from Transport in the European Union

Nikola PETROVIĆ, Nebojša BOJOVIĆ, Dragan MARINKOVIĆ*, Vesna JOVANOVIĆ, Saša MILANOVIĆ

Abstract: Urbanization has contributed to an increase in the transport activities, and thus to a greater degradation of air quality. The urban indicators in the European Union (EU) countries have different negative effects on air quality. That is why the EU has been making special efforts. Although in recent years special attention has been paid to the analysis of the Environmental Kuznets Curve (EKC), the main problem related to this analysis is the choice of air quality indicators. In order to overcome this problem, a two-phase model was developed. In the first phase, extreme values of the impact of (urban) inputs on the output (carbon dioxide (CO2) emissions) variable of the EU countries from 2000 to 2014 are determined, using the methodology of artificial neural networks (ANN), in order to select an input variable. In the second phase, starting from the selected variable, a cluster analysis of the EU countries is applied in order to examine the legality of the EKC. Based on the obtained results, it is shown that the connection between the methodology of ANN, cluster analysis and the laws of the EKC can be used to examine the impact of widely available urban indicators on various air quality indicators.

Keywords: artificial neural networks; CO2; environmental kuznets curve; urbanization

1 INTRODUCTION

Urbanization leads to a relative concentration of residents and transport activities in urban areas. With the increase in population density and the increase in urban population, the need for greater mobility is strong. The concentration of jobs and higher education institutions in cities attracts more people of working age. This concentration of people and activities in cities often creates a high level of local pollution with an impact on the air. However, the biggest challenges facing European cities are no longer local but global. Addressing climate change requires that all cities reduce CO2 emissions.

Urban development with the characteristics of low population density, spreading, dependent environmental and social impacts is often called the uncontrolled urban expansion or the so-called Sprawl index. Uncontrolled urban sprawl tends to increase car use as well as negative impacts on biodiversity. The Sprawl index is an indicator that can be considered a more relevant parameter of the impact of urbanization on the transition to sustainable urban mobility [1]. The problem with the use of this indicator is the limited statistical data (especially for the period from 2000 to 2014 and for the EU member states analysed in this paper).

During the process of urbanization in the period between 1961 and 1991, the share of inhabitants in cities on the territory of today's EU increased in relation to the total number of inhabitants from 65% to 71%. However, between 1991 and 2011, the observed share increased by only one percent, i.e. to 72%. The share of the population in European cities with at least 50000 inhabitants is low: 39% compared to 52% globally. All cities in the following EU member states have seen an increase in population (Cyprus, Denmark, Finland, Luxembourg and Sweden), while all cities in Estonia and Latvia have seen a decline in population over the last twenty years [2].

In the literature, special attention is paid to urbanization and its impact on air pollution. Population growth is one of the significant factors of air pollution in developed and developing countries [3-5]. Using a dynamic approach, Mamun et al. [6] examined the relationship between CO2 emissions and population growth between 1980 and 2009. Their results indicated that an increase in population affects the increase in CO2 emissions. Zhang et al. [7] came to the conclusion that increasing the share of urban in the total population, as well as increasing the number of inhabitants affects the increase in CO2 emissions.

Al Mulali et al. pointed out that there is interdependence between urbanization, energy consumption and CO2 emissions [8]. The same conclusion was reached by Martinez-Zarzoso and Maruotti [9] who pointed out that in developed countries urbanization has a negative impact on CO2 emissions, while in underdeveloped countries, i.e. in countries with low GDP per capita, urbanization has a positive effect on CO2 emissions.

Xu et al. have pointed to the existence of an inverse relationship between urbanization and carbon emissions [10], Katircioglu et al. tested the interdependence of urbanization and CO2 emissions in Turkey. The results indicated that there is no inverse U curve between the observed variables [11].


However, it is necessary to point out that in the literature, special attention is not paid to examining the validity of the EKC, which starts from urbanization and CO2 emitted by transport.

In the introductory discussion, a review of the relevant literature is given, and then the next chapter presents the Environmental Kuznets Curve, the problem of omitting variables and the principles of the ELM method. The third chapter describes a two-phase model for evaluating the impact of urbanization on air pollution in the EU. In the first
phase of the model, the impact of urban parameters on CO₂ emissions emitted by transport is determined using the ELM method, while in the second phase of the model, a cluster analysis of the EU countries is performed and the dependence of isolated factors and CO₂ emissions from transport is determined. The presentation and discussion of the obtained results is given in the fourth chapter. Finally, the concluding remarks, in the form of advantages and disadvantages of the application of the implemented two-phase model, are presented.

2 METHODS

2.1 Environmental Kuznets Curve

Since the 1990s, the Kuznets curve has taken on a new shape. Instead of the inverse U relationship between economic growth and economic inequality (the original Kuznets curve), the relationship between economic growth and environmental degradation is now considered and known as the Environmental Kuznets Curve. However, the results of the research depend on the choice of environmental quality indicators. The literature related to EKC validation testing includes a variety of both dependent and independent variables. Dependent variables are environmental degradation indicators or environmental quality indicators (such as CO₂, SO₂, sulphur, arsenic, lead emissions, deforestation, water pollution, etc.), while the following are considered independent variables: GDP per capita, income inequality, free trade, quality of institutions, environmental regulations and corruption [19].

Based on the EKC-based research, several shortcomings can be singled out, highlighting those related to econometric issues, ranging from stationary variables to the problem of omitting variables, as well as an inadequate or non-existent theoretical and conceptual framework that does not include feedback, i.e. environmental degradation to production [20].

2.2 The Problem of Omitting Variables

One of the main shortcomings of EKC-based studies is the problem of omitted variable bias [21]. The development of simplified models that include several variables, as well as the omission of important variables, result in processes that lead to wrong conclusions and wrong predictions. Several studies have addressed the issue of omitting variables including macroeconomic variables such as price, population, income distribution, education, technology, and societal development indicators.

In order to overcome the problem related to the omission of variables, the selection of variables is first performed, i.e. the selection of the most important factor influencing air quality in order to examine the validity of the EKC. For this reason, the selection of variables can be performed using the Extreme Learning Machine (ELM) method within ANN.  

2.3 ELM Method

Today, due to advances in technology, data is being generated at an incredible pace, leading to data sets of enormous dimensions. That is why it is important to have efficient computational methods and algorithms that can deal with such large data sets, so that they can be analyzed within a reasonable time. One of the disadvantages of ANN is learning time. A large number of algorithms are used for learning neural networks, such as the Back Propagation (BP) method, the Support Vector Machine (SVM) method, the Hidden Markov Model (HMM) method, etc. [22-26]. Traditional algorithms can sometimes take up to several days to train a neural network. A special approach has become popular with ANN in recent years, the Extreme Learning Method that uses randomization in its hidden layer and with the help of which it is possible to effectively train a network. The new ELM learning algorithm is used for Single-hidden Layer Feedforward Neural Networks (SLFN) [27, 28]. With traditional learning methods, all parameters of one-way neural networks must be adjusted, while with the ELM method it is not necessary to adjust the input weight coefficients and influences of the first hidden layer [29, 30]. Through various simulation tests in research [31], it was shown that the ELM method, in addition to accelerating the learning process, also has excellent generalization performance. Unlike traditional learning algorithms used for gradient-based unidirectional neural networks, the ELM method has the following significant features:
- Extremely high speed of learning when training SLFN.
- Ability to achieve not only the smallest training error but also the smallest weight norm, i.e. obtaining better generalization performance for SLFN networks.
- Possibility to use non-differential activation functions for SLFN network training.
- Ease and convenience of use without possible problems in terms of low learning speed, local minima, too much network training, etc.

Root means square error - RMSE is most often used as a statistical indicator of the performance of the ELM method. The precision of the model is higher with the lowest possible RMSE value.

2.4 The Problem of Examining the Validity of the Environmental Kuznets Curve

When testing the validity of the EKC, researchers most often use different types of unit root tests. However, the main shortcomings of the unit root test relate to the fact that it does not take into account: 1) the heterogeneity of states in terms of observed variables; 2) the possibility of the presence of structural breakpoints that occur in the series [32]. Researchers solve the problem of respecting heterogeneity by applying Fully Modified Ordinary Least Squares – a modified method of least squares. However, by applying this method, results are obtained for each observed country, and not for several countries where the same or similar economic laws are manifested. It is necessary to point out the fact that the results related to the EKC have indicated that its validity depends on the choice of country or region. For this reason, after the first phase (the ELM method), the second phase or cluster analysis is applied in the paper to indicate which countries are valid for the EKC and which are not.

One of the advantages of cluster analysis is reflected in the possibility of its application in various scientific
disciplines. Researchers are often faced with a large number of observed units that have different characteristics. The possibility of interpretation of the observed units is possible after the application of cluster analysis or cluster formation. Based on that, it can be pointed out that the next advantage of applying cluster analysis is the reduction of set characteristics to cluster characteristics, while the loss of information related to the entire population is minimal. The main disadvantage of this analysis is that the solutions of its application depend on the choice of variables. This shortcoming was overcome by applying the ELM method.

3 TWO-PHASE MODEL

The first step in modeling the impact of ANN and the ELM method is to define the input and output parameters (variables) of the model [33-38] in order to identify the laws and relationships between them (Fig. 1). Data collection was performed from the websites of the European Commission and the World Data Bank for the EU countries in the period from 2000 to 2014. In the process of collecting, the problem of incomplete data arose and one important advantage of ANN was used, which is the possibility of using incomplete data in its work.

Namely, in the first phase, the developed ANN/ELM model quantifies the impact of adopted input and output parameters for the EU countries [39, 40] in order to select the ones on which, in the second phase of the model [41], clustering and interdependence are examined, i.e. the existence of an inverse U relationship.

3.1 The First Phase of the Model

To determine the impact of urban parameters on CO₂ emissions stemming from transport activity, the ANN model was developed, applying the ELM method, which is a three-layer one-way neural network and consists of one input, one hidden and one output layer (Fig. 2).

The choice of this network was made on the basis of the ability to approximate any arbitrary continuous function from several real variables. Two data sets were created for the EU countries: the data on input variables and output variables in the period from 2000 to 2007 were used for training, while the data from 2008 to 2014 were used for neural network testing. For the comparison of RMSE errors to be the most adequate, it is very important that the number of training epochs for all inputs to the neural network is the same. Specifically, during the processing of the mentioned input-output data and determining the influence of input variables in the ANN model, one epoch was used for neural network training.

The content of the input layer consists of three neurons, i.e. three different neural network inputs. The output layer of the neural network consists of a single neuron, whose output covers each combination of input variables, for CO₂ emissions from transport. The sigmoid (logistic) function was used as the activation function for the output layer, while the linear (purelin) function was used for the hidden layer. The adopted learning algorithm independently adjusts the parameters of the neural network in order to find their right combination with which the neural network approximates the nonlinear function with high quality. Networks are trained for input data in such a way as to determine the specific RMSE error of each input to a specific output. The input variable with the lowest RMSE training error obtained has the greatest impact on the observed output variable and vice versa, i.e. the input variable with the highest RMSE error has the smallest impact on the adopted output variable.

In the process of training the network, the number of neurons in the hidden layer was changed, which also depended on the number of input neurons, in order to obtain the best results in the output layer. In RMSE testing, error is used to monitor the regression flow between training and test data sets. The neural network training process needs to be stopped when the RMSE testing error starts to show its sudden increase in relation to the RMSE training error, i.e. that there is a deviation between the training and testing data sets. Using the developed ANN, the modeling of the impact of urbanization on CO₂ emissions was performed on the basis of input and output data for the EU countries in the time interval from 2000 to 2014 (Fig. 2). The input parameters used in this model are percentage of urban in total population, annual urban population growth in percentages and population density shown per capita per km² of state. As the output parameter, CO₂ in tons emissions from transport was analyzed. All input and output variables are modeled within the Matlab R2015b software.

Figure 1 A Two-phase model for the evaluation of urbanization impacts on CO₂ emissions from transport in the EU [42]
### 3.2 The Second Phase of the Model

Cluster analysis is a multivariable technique of grouping similar units into different groups or clusters, i.e., dividing a set into subsets according to a certain measure of distance [43]. The characteristics of observation units, i.e., cluster characteristics, are determined on the basis of variables. Today, statistical software is used to implement cluster analysis such as Statistica and SPSS. The paper uses STATISTICA 8.0 software for cluster analysis.

Cluster analysis does not have a mechanism that allows the distinction between relevant and irrelevant variables on the basis of which relatively homogeneous clusters are formed. Therefore, the choice of the included variables on the basis of which the cluster analysis is performed must be subject to conceptual considerations. The results of cluster analysis depend on both the representativeness of the sample and the multicollinearity of the variables on the basis of which the cluster analysis is performed [44]. Researchers are often faced with a large number of observed units that have different characteristics. The possibility of interpretation of the observed units is possible after the application of cluster analysis or cluster formation. Based on that, it can be pointed out that the next advantage of applying cluster analysis is the reduction of set characteristics to cluster characteristics, while the loss of information related to the entire population is minimal. The main disadvantage of this analysis is that the solutions of its application depend on the choice of the cluster analysis method. Clustering is a process that allows a basic set to be divided into subsets or clusters. A center or center of gravity is formed for each cluster, and distance or distances are used as a measure of similarity in this case between the EU countries. In other words, the existence of an inverse relationship between the observed input and output parameters of the EU countries is examined.

### 4 RESULTS AND DISCUSSION

#### 4.1 Results of Application of the First Phase of the Model

RMSE errors after training and testing of ANN as the impact of urbanization on air pollution caused by transport activities are given in Tab. 1. In the created ANN model all input and output variables are modeled and the obtained value of the least RMSE training error is 52,393 while the highest is 42,646.

<table>
<thead>
<tr>
<th>Table 1 Impacts of urban indicators on air pollution</th>
<th>CO2 RMSE training</th>
<th>CO2 RMSE testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban population</td>
<td>43,5233</td>
<td>51,8307</td>
</tr>
<tr>
<td>Urban population growth</td>
<td>52,3932</td>
<td>52,6333</td>
</tr>
<tr>
<td>Population density</td>
<td>42,6460</td>
<td>50,7458</td>
</tr>
</tbody>
</table>

The obtained results indicate the fact that the population density input parameter has the smallest RMSE training error. This fact unequivocally leads to the conclusion that the population density parameter has the greatest impact on CO2 emitted by transport. In doing so, the urban population growth input parameter has the largest RMSE training error. Bearing in mind that the largest deviation is the actual and projected value, it can be unequivocally concluded that the urban population grows.

After quantification of the impact, ANN methodology, and the observed parameters on air pollution caused by transport activities, the clustering of the EU countries was performed and, based on the identified major impacts, the validity of the existence of the EKC was examined.

#### 4.2 Examination of the Environmental Kuznets Curve

After quantification and determining the greatest impact of the observed parameters of the EU countries, a cluster analysis of the EU countries was performed and the legality of the EKC was examined (which was proven), starting from CO2 emissions from transport activities, GDP per capita and population density (Tab. 2). The first cluster includes the countries that record significantly less GDP per capita but also CO2 emitted by transport compared to the countries belonging to the second cluster. However, the countries belonging to the first cluster record the highest population density, i.e., higher population density compared to the countries belonging to the second cluster.
Countries belonging to the third cluster record the lowest CO₂ emitted by transport, but also the lowest population density in relation to the observed clusters. Fig. 3a and Fig. 3b show that the countries belonging to clusters I and II, from the observed period, went through a phase characterized by an increase in CO₂ emissions due to increasing population density and entered a phase characterized by a decrease in CO₂ emitted by transport due to increased density population. At the beginning of the observed period, the countries belonging to the third cluster (Fig. 3c) were at the end of the phase characterized by the increase of CO₂ emitted by transport due to the increase in population density.

It is necessary to point out the fact that the observed population density at which the reduction of CO₂ emitted by transport occurs is not the same for the observed clusters. For the first cluster it is a population density of about 280 inhabitants/km², for the second cluster a population density of about 180 inhabitants/km², and for the third cluster it is a population density of about 82 inhabitants/km².

5 CONCLUSIONS

Transport represents one of the largest emitters of harmful substances, such as carbon dioxide, which affect air quality especially in urban areas. Urbanization parameters such as the share of urban in the total population, the growth of urban population on an annual level and population density in the European Union (EU) have different negative effects on air quality. A two-phase model was developed to analyze the impact of urbanization indicators on carbon dioxide emissions emitted by transport in the observed time period from 2000 to 2014 for the EU states.

In the first phase of the model, ANN analysis was conducted, and based on the obtained results it can be concluded that population density has the greatest impact, while urban population growth has the least impact on CO₂. After the cluster analysis was performed according to the CO₂ emissions from transport, GDP per capita and population density, the validity of the EKC was examined, which was also proved by the existence of an inverse U relation. The first cluster includes the countries that record less GDP per capita and CO₂ emitted by transport compared to the countries that belong to the second cluster. However, the countries belonging to the first cluster record the highest population density, i.e. higher population density compared to the countries belonging to the second cluster. The countries belonging to the third cluster record the lowest CO₂ emitted by transport, but also the lowest population density in relation to the observed clusters.

The disadvantage of the proposed model is related to the fact that within the impact of urbanization on CO₂ emissions emitted by transport, the following indicators were considered: urban population, urban population growth and population density, i.e. the impact of quantitative population changes on air quality (participation of the able-bodied in the urban population and the educational structure of the population) on air quality in the EU. The directions for future research will be focused on: Research into the impact of qualitative changes in the population on the transport of passengers and goods, i.e. on the use of passenger and freight transport modes; Analysis of the impact of qualitative changes in population on air quality; Application of ANN models developed to simulate precise emission or pollutant concentration scenarios; Examination of the validity of the EKC, which refers to the interdependence between different modes of passenger (freight) transport and CO₂.

Table 2 Cluster analysis based on CO₂ emitted by transport, GDP per capita and population density

<table>
<thead>
<tr>
<th>States</th>
<th>Distance</th>
<th>States</th>
<th>Distance</th>
<th>States</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyprus</td>
<td>1556,741</td>
<td>Austria</td>
<td>3389,32</td>
<td>Bulgaria</td>
<td>3276,628</td>
</tr>
<tr>
<td>Greece</td>
<td>1239,614</td>
<td>Belgium</td>
<td>4564,49</td>
<td>Croatia</td>
<td>618,369</td>
</tr>
<tr>
<td>Italy</td>
<td>5023,793</td>
<td>Denmark</td>
<td>2557,55</td>
<td>Czech Republic</td>
<td>2593,132</td>
</tr>
<tr>
<td>Malta</td>
<td>3419,438</td>
<td>Finland</td>
<td>3085,88</td>
<td>Estonia</td>
<td>1509,123</td>
</tr>
<tr>
<td>Portugal</td>
<td>2528,331</td>
<td>France</td>
<td>6164,05</td>
<td>Hungary</td>
<td>744,984</td>
</tr>
<tr>
<td>Slovenia</td>
<td>2163,844</td>
<td>Germany</td>
<td>5401,62</td>
<td>Latvia</td>
<td>539,205</td>
</tr>
<tr>
<td>Spain</td>
<td>1913,216</td>
<td>Ireland</td>
<td>2356,62</td>
<td>Lithuania</td>
<td>509,205</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>25560,21</td>
<td>Poland</td>
<td>541,866</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>1646,09</td>
<td>Romania</td>
<td>2517,628</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>1448,06</td>
<td>Slovakia</td>
<td>1735,004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>5243,12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 Movement of average values of CO₂ emitted by transport and population densities in a) I, b) II and c) III cluster, respectively
The EKC has been improved with the developed two-phase model, starting from its shortcomings which are pointed out in the literature. The ANN model developed within the first phase of the two-phase model was used to quantify the impact of urbanization and transport modes at the local level, ie. the level of the urban environment. The ANN model developed within the first phase of the two-phase model can be applied to other urban areas as well. A two-phase model represents a universal platform for managing different urban indicators and their mutual influences on air pollution.

Acknowledgements

This research was financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-68/2022-14/200109).

6 REFERENCES


Contact information:
Nikola PETROVIĆ, Assistant Professor
University of Niš, Faculty of Mechanical Engineering, Aleksandra Medvedeva 14, 18106, Niš, Serbia
E-mail: petrovic.nikola@masfak.ni.ac.rs

Nebojša BOJOVIĆ, Full Professor
University of Belgrade, Faculty of Transport and Traffic Engineering Voyoode Stepe 305, 11010, Belgrade, Serbia
E-mail: nb.bojovic@af.bg.ac.rs

Dragan MARINKOVIĆ, Full Professor (Corresponding author)
University of Niš, Faculty of Mechanical Engineering in Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia
Lecturer and Researcher,
Department of Structural Analysis, Berlin Institute of Technology, Strasse des 17. Juni 135, 10623 Berlin, Germany
E-mail: dragan.marinkovic@tu-berlin.de

Vesna JOVANOVIĆ, Assistant Professor
University of Niš, Faculty of Mechanical Engineering, Aleksandra Medvedeva 14, 18106, Niš, Serbia
E-mail: vesna.jovanovic@masfak.ni.ac.rs

Saša MILANOVIĆ, Associate Professor
University of Niš, Faculty of Mechanical Engineering, Aleksandra Medvedeva 14, 18106, Niš, Serbia
E-mail: sasa.milanovic@masfak.ni.ac.rs