Air Qual Atmos Health (2017) 10:15–23
DOI 10.1007/s11869-016-0403-6

Prediction of nitrogen oxides emissions at the national level based on optimized artificial neural network model

Lidija J. Stamenković[1] · Davor Z. Antanasijević[2] · Mirjana Đ. Ristić[1] · Aleksandra A. Perić-Grujić[1] · Viktor V. Pocajt[1]

Received: 19 January 2016 / Accepted: 5 April 2016 / Published online: 14 April 2016

Abstract Nitrogen oxides (NO_x) emissions into the atmosphere have multiple negative effects on the environment and effects directly and indirectly on human health. This paper describes the development of a model for NO_x emission prediction at the national level based on artificial neural networks (ANNs) and on widely available sustainability, industrial, and economical parameters as input variables. In this study, 11 sustainability, industrial, and economical parameters were chosen as potential input variables. The ANN models were trained, validated, and tested with available data for 17 European countries, USA, China, Japan, Russia, and India for the years 2001 to 2008. The ANN modeling was performed using general regression neural network (GRNN), and correlation and variance inflation factor (VIF) analysis were applied to reduce the number of input variables. The best results were obtained using the selection of inputs based on the correlation between input variables, which provided a more accurate prediction than the GRNN model created with all initial selected input variables. Sensitivity analysis showed that the input variables with the largest influences on the GRNN model results were (in descending order) electricity production from oil sources, agricultural land, fossil fuel energy consumption, number of vehicles, gross domestic product, energy use, and electricity production from coal sources.

✉ Lidija J. Stamenković
lidija.stm@gmail.com

1 Faculty of Technology and Metallurgy, University of Belgrade, Karnegijeva 4, Belgrade 11120, Serbia

2 Innovation Center of the Faculty of Technology and Metallurgy, Karnegijeva 4, Belgrade 11120, Serbia

Keywords NO_x · Emission prediction · ANN · Correlation analysis · Variance inflation factor

Introduction

In the recent decades, air pollution has become a major concern, with its transboundary nature that turned it into a global problem that needs to be confronted by all countries (Alyuz and Alp 2014), with many organizations worldwide established to help combating air pollution problems and reducing pollutant emissions (Reşitoğlu et al. 2014). Two very important air pollutants, in terms of environmental protection, are nitrogen oxide and nitrogen dioxide, collectively denoted as nitrogen oxides (NO_x). Besides their direct effects, NO_x are the precursors of ozone and fine particulate matter, which cause various health problems, including respiratory diseases and cardiovascular mortality (Pope et al. 2002; Jerrett et al. 2009; Guttikunda and Kopakka 2013; Jhun et al. 2014; Tong et al. 2015). Nitrogen oxides are emitted into the atmosphere from thousands of anthropogenic emission sources and also from a significant number of natural sources. Fossil fuel power plants, industrial production, and motor vehicles are the major sources of NO_x emissions in China, USA, and Asia, and road transport is the largest contributor in the EU countries (Klimont et al. 2001; Kim et al. 2006; Vestreng et al. 2008).

In the recent years, NO_x emission data are mainly obtained from emission inventory models, developed on national, regional, and global scale. For example, the Emission Database for Global Atmospheric Research (EDGAR), the Greenhouse and Air Pollution Interactions and Synergies (GAINS), and USEPA National Emission Inventory (NEI) models were developed to provide emissions of multiple air pollutants including NO_x (Olivier et al. 1994; Van Amstel et al. 1999; Schopp et al. 1999; Streets and Waldhoff 2000; EPA 2005; Höglund-Isaksson 2012). In these models, emissions were calculated using emission factors and activity data, which had a large number of input parameters (Zhou et al. 2014). The obtained estimates are however considered to be uncertain owing to limited availability and/or inaccurate information on specific sources, statistics on source data and process technology, locations of specific sources, and emission factors for sources (Ghude et al. 2013). This imposes significant challenges for accurate estimate of NO_x emissions, especially for developing countries.

One of the possible alternatives for modeling NO_x emissions on the national level is artificial neural networks (ANNs), which have been successfully utilized in diversified applications in various fields (May et al. 2011). ANNs are inspired by biological systems and their ability to learn and generalize stems from a group of units called neurons that are connected to each other and organized in layers (Zhang et al. 1998). Broadly spoken, they are trained using datasets consisting of input and output variables until they learn the patterns presented to them (Ata 2015). After an ANN has been trained, it is able to perform prediction or classification on the entirely new datasets (Kalogirou 2001). ANNs were successfully applied for the prediction of NO_x emissions from different emission sources such as power plants, coal-fired boilers, and natural gas engine (Tronci et al. 2002; Kesgin 2004; Zhou et al. 2004; Zheng et al. 2008; Smrekar et al. 2013) as well as for emission prediction on the national level for some other pollutants (Sözen et al. 2009; Antanasijević et al. 2014).

In contrast to emission inventory models, the novelty of the ANN approach is that it does not require the availability of a large number of technology and process-specific data for emission factors and activity levels. Models may be created using a significantly lower number of widely available sustainability, economical, and industrial parameters as inputs (Antanasijević et al. 2014; Alyuz and Alp 2014; Shi et al. 2014). In order to obtain an ANN model with good prediction capability, the inputs are often selected by applying correlation analysis or similar statistical technique (Galelli et al. 2014).

This paper presents the development of a general regression neural network (GRNN) model for the prediction of national emissions of NO_x for the USA, China, Japan, Russia, India, and 17 European countries, using sustainability, economical, and industrial indicators as inputs. The application of correlation analysis and variance inflation factor (VIF) analysis for the selection of appropriate input variables will be described, together with the comparison of results of the GRNN models created with different combinations of inputs.

Materials and methods

Input and output data

Generally speaking, at the global level, fuel combustion in power plants, industry, road transport, domestic sectors and agricultural activities are the largest contributors to NO_x emissions. In other words, the economic and industrial development leads to greater production of air pollutants, including NO_x emissions, and this relationship has already proved in several studies that implied creating ANN models to predict annual emissions (Antanasijević et al. 2014; Stamenković et al. 2015). Following this approach, 11 indicators related to NO_x emissions were selected as potential model inputs (Table 1). The dataset for selected input variables used in this study was obtained from three statistical databases: World Bank (The World Bank 2015), OICA-Production Statistics (OICA 2015), and the international statistics database NationMaster (NationMaster 2015). The list of chosen input variables statistics for 17 European countries, USA, China, Japan, Russia, and India and their descriptive statistics are presented in Table 1.

Table 1 The descriptive statistics of the initial selected input variables

Input variable	Unit	Minimum	Maximum	Mean	Standard deviation
Electricity production from coal sources (EPCS)	kWh pc	147.570	7288.78	1987.83	1657.65
Electricity production from oil sources (EPOS)	kWh pc	7.920	1535.48	234.20	281.17
Road sector diesel fuel consumption (RDFC)	toe pc	0.010	0.69	0.30	0.17
Road sector energy consumption (RSEC)	toe pc	0.027	1.80	0.59	0.36
Gross domestic product (GDP)	€ pc	371.722	40,400.00	20,978.89	12,432.13
Energy use (EU)	toe pc	0.438	782.83	9.21	59.60
Fossil fuel energy consumption (FFEC)	%	0.325	0.96	0.77	0.15
Electricity production from renewable sources (EPRS)	kWh pc	23.330	9366.16	1417.97	1947.67
Number of vehicles (NV)	–	10,510	12,279,582	2,263,830	3,289,624
Electric power consumption (EPC)	kWh pc	399.571	17,212.95	6780.86	3931.43
Agricultural land (AL)	%	0.073	0.74	0.44	0.19

pc per capita, *toe* tons of oil equivalent

Table 2 The descriptive statistics of NO_x emissions at the national level for the study period 2001–2008

	NO_x emissions (kg pc)			
	Minimum	Maximum	Mean	Standard deviation
Country				
Belgium	24.30	31.92	29.22	2.47
Czech Republic	25.98	32.68	29.62	2.64
Germany	21.07	25.37	23.14	1.42
Spain	41.13	47.84	46.38	2.24
France	21.31	28.26	25.51	2.34
Italy	21.06	26.97	24.27	2.08
Hungary	15.94	20.77	18.17	2.05
Netherlands	25.35	31.24	28.50	2.01
Austria	25.78	29.78	28.25	1.34
Poland	21.16	22.69	22.12	0.51
Portugal	24.33	29.15	27.40	1.57
Romania	12.86	14.87	13.95	0.73
Slovenia	24.20	29.76	26.18	1.66
Slovakia	17.41	20.04	18.42	0.82
Finland	37.07	50.57	44.30	4.79
Sweden	31.83	36.46	34.85	1.57
UK	28.21	34.54	31.64	1.83
USA	46.81	62.36	55.55	5.02
China	9.09	15.66	12.12	2.33
India	5.04	6.07	5.49	0.35
Japan	18.74	21.55	20.50	0.94
Russia	29.81	43.17	34.74	4.98
Dataset				
Training data	5.04	62.36	27.90	11.85
Test data	5.88	51.97	25.45	9.97

In order to achieve better performance of the ANN model and also to allow the comparison of countries of different sizes, the selected input and output parameters have been normalized per capita. The ANN models were trained using the data for 17 EU countries, China, USA, Japan, India, and Russia for the period 2001–2006, while the data for the period 2007–2008 was used for testing the models. The number of observations in the training dataset was 132, while the number of observations in the test dataset was 44. In addition, 20 % of randomly selected training datasets were used for model validation. The descriptive statistics of NO_x emissions at the national level obtained from the EDGAR project (JRC-IES 2015) for each country included in this study and for different datasets are presented in Table 2.

Input selection techniques

The selection of input variables from available potential input data is a fundamental and yet crucial consideration in identifying the optimal functional form of ANN models. During the selection process, it is important to select the inputs that are the most significant and have the lowest overall cross-correlation, which means the lowest redundancy (May et al. 2011). In this study, two model-free statistical techniques were utilized for the selection of input variables: correlation analysis and VIF statistic.

In correlation analysis, input variables were selected on the basis of input-output correlation and on the basis of input cross-correlation. In the former case, the input variables that had the lowest correlation with the output were removed and a new model was created without these input variables. In the latter case, models with different combinations of mutually low correlated inputs ($r < 0.7$) were created and evaluated.

VIF is a method for the assessment of multicollinearity; i.e., its value corresponds to the collinearity of analyzed variables (Vu et al. 2015). The VIF is determined based on the linear relationship between the inputs (Alin 2010)

$$VIF_j = \frac{1}{\left(1 - R_j^2\right)}$$

where R_j^2 is the coefficient of determination of the regression of one input on all other inputs.

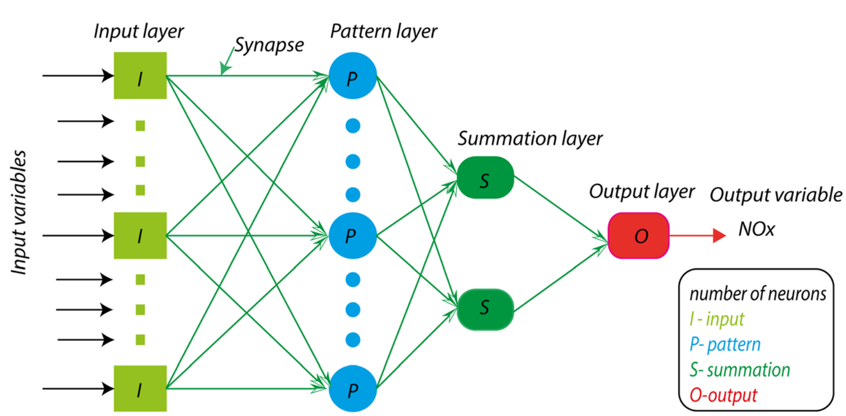

Fig. 1 The GRNN architecture

Table 3 Details of created GRNN models

Model	Input selection	Critical value[a]	Removed inputs	Architecture[b]
GRNN1	All available	–	–	11-106-2-1
GRNN2	Correlation analysis (input-output correlation)	$r < 0.1$	FFEC	10-106-2-1
GRNN3	Correlation analysis (input cross-correlation)	$r > 0.7$	a: RSEC, GDP, EPC, EPRS	a: 7-106-2-1
			b: RDFC, GDP, FFEC	b: 8-106-2-1
			c: RSEC, FFEC, EPC	c: 8-106-2-1
			d: RDFC, RSEC, EPRS, EPC	d: 7-106-2-1
GRNN4	VIF analysis	VIF > 5	a: EPC	a: 10-106-2-1
			b: EPC, GDP	b: 9-106-2-1
			c: EPC, GDP, RSEC	c: 8-106-2-1

[a] For input removal

[b] The number of neurons per layer

General regression neural networks

GRNN architecture and training

Artificial neural network is organized in layers, where each of layers consists of interconnected processing units called neurons. The way that neurons are interconnected is called the topology or architecture of ANN. In this study, for the creation of ANN models for the prediction of NO_x emissions, an architecture known as general regression neural network (GRNN) was used, since it has already demonstrated good results in environmental modeling (Singh et al. 2012; Antanasijević et al. 2015). The advantages of GRNNs are their ability to be trained quickly and to make accurate predictions on sparse datasets. The GRNN works by measuring how far a given input vector is from the vectors in the training dataset in the N-dimensional space, where N is the number of inputs (Kalogirou 2003).

It is a one-pass supervised learning network consisting of four layers, where each layer has a different role. The number of neurons in each GRNN layer is determined by the number of input/output variables and data patterns in the dataset used for training:

- The input layer: each neuron in this layer represents one input variable;
- The pattern layer: the number of neurons is equivalent to the number of training cases;
- The summation layer: it has two different neurons that compute the weighted and unweighted sum of the pattern outputs;
- The output layer: the number of neurons in output layer corresponds to the number of outputs.

A schematic representation of GRNN architecture is presented in Fig. 1.

Table 4 The results of correlation analysis

	EPCS	EPOS	RDFC	RSEC	GDP	EU	FFEC	EPRS	NV	EPC	AL
EPCS	1										
EPOS	−0.07	1									
RDFC	0.07	0.23	1								
RSEC	0.53	0.27	*0.72*	1							
GDP	0.15	0.26	*0.81*	*0.78*	1						
EU	0.26	0.05	0.03	0.27	0.07	1					
FFEC	0.32	0.20	−0.19	−0.11	−0.30	0.06	1				
EPRS	−0.18	0.03	0.31	0.32	0.45	−0.02	*−0.70*	1			
NV	0.20	0.13	0.49	0.38	0.41	0.00	−0.12	−0.06	1		
EPC	0.30	0.09	0.52	*0.73*	*0.77*	0.14	−0.55	0.69	0.25	1	
AL	0.02	−0.15	−0.06	−0.20	−0.26	0.00	0.48	−0.64	−0.14	−0.64	1
NO_x	*0.50*	*0.21*	*0.56*	*0.78*	*0.58*	*0.25*	−0.08	*0.33*	*0.21*	*0.73*	*−0.32*

Italic numbers indicate the highly correlated inputs and inputs highly correlated with NO_x emission

Table 5 VIF analysis and created GRNN models

Input	VIF values
EPC	17.03
GDP	8.68
RSEC	7.82
RDFC	4.85
FFEC	4.63
EPCS	4.57
AL	4.51
EPRS	4.22
NV	2.38
EPOS	1.58
EU	1.19

During the training process, 80 % of training cases were used to adjust GRNN weights while optimal smoothing factor (SF) was determined using genetic algorithms and the remaining 20 % of training data (for details, see, e.g., Ben-Nakhi and Mahmoud (2002)). The SF is the only parameter that needs to be determined during the GRNN training, and therefore, it defines the prediction ability of the created model. As is stated by Specht (1991), larger SF provides more smooth regression surface while lower values provide a closer non-linear approximation to the output values.

GRNN model development

In this study, the NeuroShell 2 software was used for the creation ANN models (Ward Systems Group Inc. 2008). Four GRNN models were created using combinations of inputs determined by several approaches (Table 3), and all models were trained with min-max normalized values in the range of 0 to 1.

The first approach for input selection was based on correlation analysis. The GRNN2 was created on the basis of *input-output correlation*, where only the input variables correlated with the output with a coefficient of correlation higher than 0.1 were used. In the present case, as can be seen in Table 4, only one input variable (fossil fuel energy consumption) needed to be removed from the initial input vector.

The GRNN3 model was created using only the input variables with the mutual coefficient of correlation lower than 0.7. In this study, there are six pairs of highly correlated input variables [road sector energy consumption (RSEC)-road sector diesel fuel consumption (RDFC), gross domestic product (GDP)-RDFC, GDP-RSEC, electric power consumption (EPC)-RSEC, GDP-EPC, and electricity production from renewable sources (EPRS)-fossil fuel energy consumption (FFEC)]. Therefore, in order to obtain an independent set of input variables, four GRNN models with different combinations of independent input variables were created.

The second input selection approach was based on the VIF analysis. Since the critical value of VIF that indicates serious multicollinearity is problem specific, the GRNN models were created by sequentially removing inputs with the highest VIF value (Table 5).

Model performance indicators

Statistical analysis of the model performance generally compares the values predicted by the model with the observed or actual data. The performance of each model created in this study and its ability to produce accurate predictions were determined using the statistical indicators which are commonly used in the related literature (Moriasi et al. 2007): the index of model performance (dr) (Willmott et al. 2012), the Nash-Sutcliffe efficiency (NSE) (Wang et al. 2009), the correlation coefficient (*r*), the root mean square error (RMSE), the mean absolute error (MAE), the RMSE-observation standard

Table 6 Performance indicators for the created GRNN models

Model	Performance indicators						
	d_r	NSE	RMSE (kg pc)	MAE (kg pc)	RSR	FA 1.25 (%)	*r*
GRNN1	0.92	0.89	3.78	2.97	0.34	89	0.97
GRNN2	0.93	0.91	3.33	2.61	0.30	95	0.97
GRNN3a	0.92	0.86	4.10	2.86	0.38	91	0.95
GRNN3b	0.93	0.91	3.34	2.69	0.30	93	0.97
GRNN3c	0.93	0.91	3.29	2.63	0.30	98	0.98
GRNN3d	*0.94*	*0.93*	*2.80*	*2.20*	*0.26*	*95*	*0.98*
GRNN4a	0.93	0.92	3.09	2.46	0.28	93	0.97
GRNN4b	0.91	0.87	3.97	3.13	0.37	82	0.96
GRNN4c	0.93	0.87	3.87	2.64	0.36	95	0.95

Italic numbers indicate the best performance model

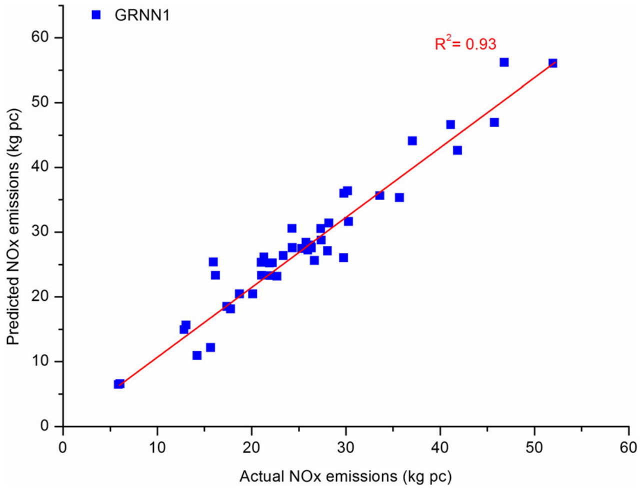

Fig. 2 Comparison of the actual data and GRNN1 model results (test dataset)

deviation ratio (RSR), and the percent of predictions within the factor of 1.25 of observed values (FA1.25).

Results and discussion

Input selection and model results

The values of performance indicators for all created GRNN models are presented in Table 6.

The values of performance indicators (dr = 0.92, FA1.25 = 89 %) obtained by the GRNN1 model can be regarded as good. A comparison of actual and predicted values of NO_x emissions for the test dataset created is shown in Fig. 2.

Since the GRNN1 model, which is created with all initial selected inputs, provided good agreement with the actual NO_x emission, it can be concluded that the initial input selection was adequate.

It can be observed from Table 6 that the eight GRNN models created using different input selection approaches demonstrated significantly better results compared with GRNN1 model which used all initial input variables. The GRNN2 model based on input-output correlation showed satisfactory results, but not the best (dr = 0.93 and FA1.25 = 95 %); also, in this case, the number of input variables was the highest compared with the other GRNN models. From the models based on VIF analysis, GRNN4a demonstrated superior performance (dr = 0.93 and FA1.25 = 93 %) in comparison with GRNN4b and GRNN4c. It can be concluded that further reduction of input variables after eliminating EPC led to the slightly degraded results, the reason being that from both GRNN4b and GRNN4c models, GDP was removed as an input variable. In fact, from the results in Table 6, it can be seen that all models created without GDP as an input variable have slightly lower performance than the other models, which probably occurred due to the existence of three highly correlated pairs between GDP and other inputs (Table 4). The GRNN3d model demonstrated the best performance indicator values from all the models created both on the basis of VIF and input-output correlation analysis and also significantly better performance than the GRNN1 model. As it can be seen, the GRNN3d model was created using the smallest number of input variables, that is 7, which represents a reduction of 37 % compared to the GRNN1 model. The plots of actual versus predicted NO_x emissions as well as the FA1.25 plot for the GRNN3d model are presented in Fig. 3. It can be clearly seen that the GRNN3d model produces a very good agreement between the observed and modeled NO_x emissions (R^2 = 0.95) (Fig. 3a) and have reduced deviation in comparison with the GRNN1 model (Table 6).

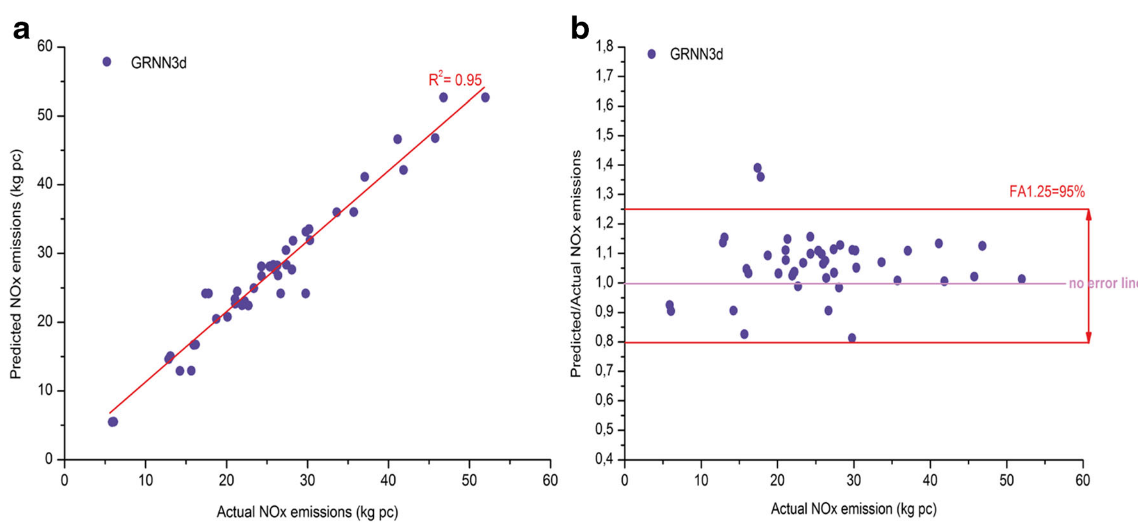

Fig. 3 The results of GRNN3d model for the test dataset: **a** comparison of actual and predicted NO_x emissions and **b** FA1.25 plot

As can be observed in Fig. 4a, predictions for Slovakia can be regarded as extreme outliers (rel. err. >35 %), while the maximum relative errors for 2007 and 2008 predictions were made for Romania and Slovenia, respectably. In general, Eastern European countries tend to have emission inventories that exhibit higher uncertainty, which is most probably the reason for higher deviations between the predicted and actual values. Also, 1-year-ahead predictions prove to be more accurate that 2-year-ahead predictions: for 2007 data, 75 % of predictions have the error of up to 7 %, while for 2008 for the same fraction, the error is 15 % (Fig. 4a). A similar trend can be observed in the case of five non-European countries (Fig. 4b), with error ranges that are higher for China and Russia.

Sensitivity analysis

In order to determine the influence of each input variable on the GRNN3d model, sensitivity analysis was performed.

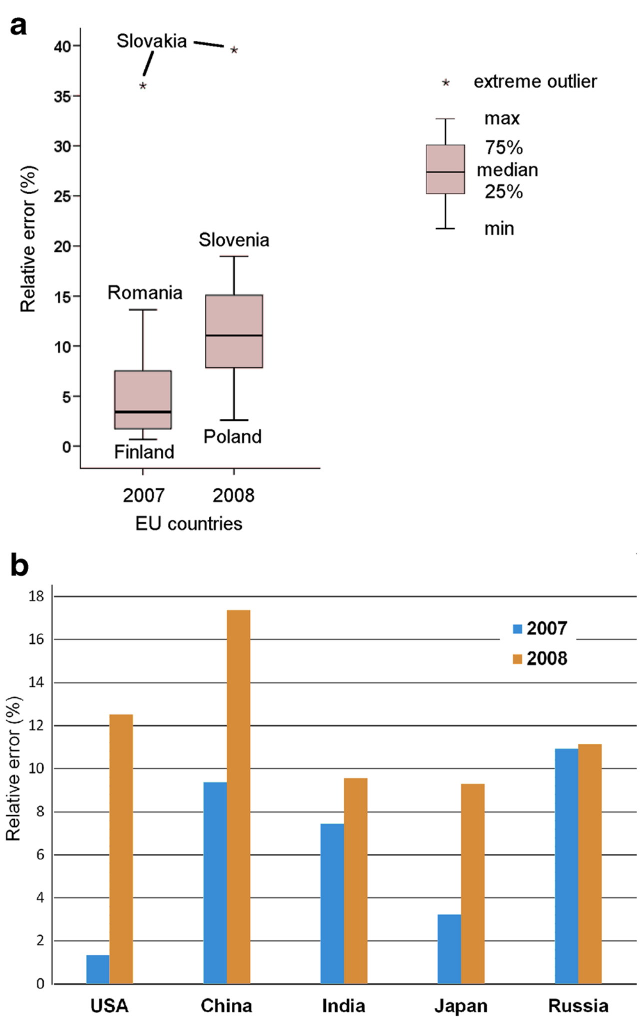

Fig. 4 Relative errors for **a** European and **b** non-European countries

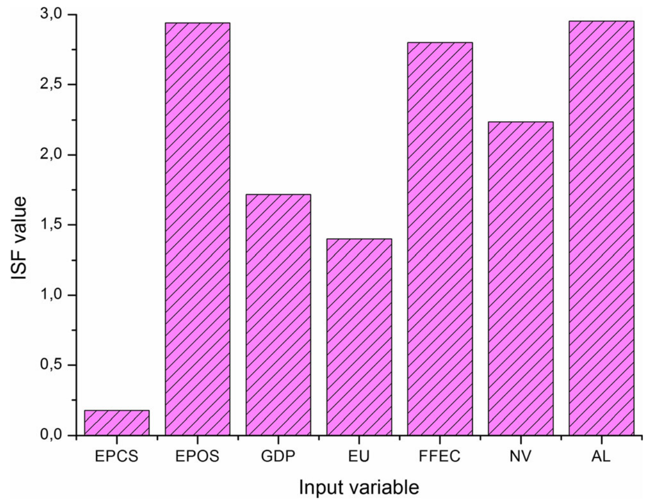

Fig. 5 Individual smoothing factors for input variables used in the created GRNN3d model

During the training process of GRNN network, besides using a genetic algorithm for the determination of the overall smoothing factor, individual smoothing factors (ISF) for every input variable can also be calculated. Individual smoothing factors indicate the significance of each input variable for the GRNN model (Kalogirou 2003). Consequently, sensitivity analysis in this study was performed on the basis of ISF values (Fig. 5). According to ISF values, the most significant input variables are the electricity production from oil sources (EPOS) followed by agricultural land (AL), FFEC, and number of vehicles (NV). Gross domestic product (GDP) and energy use (EU) have lower but still significant influence on the model output. Therefore, these ISF values confirmed the fact that the energy and transport sectors have the most significant impact on NO_x emission.

Conclusions

In this paper, the artificial neural network (ANN) approach was used for the prediction of nitrogen oxides (NO_x) emissions at the national level for 17 European countries (EU Members), USA, China, India, Japan, and Russia. The ANN model was created using general regression neural network (GRNN) architecture and sustainability, economical, industrial, and agricultural indicators as input variables. In order to optimize the GRNN model, two input selection techniques (correlation and VIF analyses) were applied.

From the eight created GRNN models, the best performance was achieved after the reduction of input variables using the input cross-correlation procedure (dr = 0.94, NSE = 0.93, RMSE = 2.80 kg pc, MAE = 2.20 kg pc, FA1.25 = 95 %), which was a significant improvement in comparison with the non-optimized GRNN model, created with all available input variables (dr = 0.92, NSE = 0.89, RMSE = 3.78 kg pc, MAE = 2.97 kg pc, FA1.25 = 89 %).

Considering the demonstrated prediction capability of the developed ANN model and the small number and availability of input parameters, it can be concluded that the model can be successfully applied for the prediction of NO_x emissions at the national level and can potentially present an alternative approach for the prediction of NO_x emissions.

Acknowledgments The authors are grateful to the Ministry of Education, Science and Technological Development of the Republic of Serbia, Project No. 172007, for the financial support.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Statement of human rights and statement on the welfare of animals This article does not contain any studies with human participants or animals performed by any of the authors.

References

Alin A (2010) Multicollinearity. Wiley Interdiscip Rev Comput Stat 2: 370–374. doi:10.1002/wics.84

Alyuz U, Alp K (2014) Emission inventory of primary air pollutants in 2010 from industrial processes in Turkey. Sci Total Environ 488–489:369–381. doi:10.1016/j.scitotenv.2014.01.123

Antanasijević DZ, Ristić MĐ, Perić-Grujić AA, Pocajt VV (2014) Forecasting GHG emissions using an optimized artificial neural network model based on correlation and principal component analysis. Int J Greenh Gas Control 20:244–253. doi:10.1016/j.ijggc.2013.11.011

Antanasijević D, Pocajt V, Ristić M, Perić-Grujić A (2015) Modeling of energy consumption and related GHG (greenhouse gas) intensity and emissions in Europe using general regression neural networks. Energy 84:816–824. doi:10.1016/j.energy.2015.03.060

Ata R (2015) Artificial neural networks applications in wind energy systems: a review. Renew Sust Energ Rev 49:534–562. doi:10.1016/j.rser.2015.04.166

Ben-Nakhi AE, Mahmoud MA (2002) Energy conservation in buildings through efficient A/C control using neural networks. Appl Energy 73:5–23. doi:10.1016/S0306-2619(02)00027-2

EPA (2005) 2002 National Emissions Inventory booklet.

Galelli S, Humphrey GB, Maier HR et al (2014) An evaluation framework for input variable selection algorithms for environmental data-driven models. Environ Model Softw 62:33–51. doi:10.1016/j.envsoft.2014.08.015

Ghude SD, Pfister GG, Jena C et al (2013) Satellite constraints of nitrogen oxide (NOx) emissions from India based on OMI observations and WRF-Chem simulations. Geophys Res Lett 40:423–428. doi:10.1029/2012GL053926

Guttikunda SK, Kopakka RV (2013) Source emissions and health impacts of urban air pollution in Hyderabad, India. Air Qual Atmos Health 7:195–207. doi:10.1007/s11869-013-0221-z

Höglund-Isaksson L (2012) Global anthropogenic methane emissions 2005-2030: technical mitigation potentials and costs. Atmos Chem Phys 12:9079–9096. doi:10.5194/acp-12-9079-2012

Jerrett M, Burnett RT, Pope CA et al (2009) Long-term ozone exposure and mortality. N Engl J Med 360:1085–95. doi:10.1056/NEJMoa0803894

Jhun I, Coull BA, Zanobetti A, Koutrakis P (2014) The impact of nitrogen oxides concentration decreases on ozone trends in the USA. Air Qual Atmos Health 8:283–292. doi:10.1007/s11869-014-0279-2

JRC-IES (2015) Institute for Environment and Sustainability. In: Global Emissions EDGAR v4.2. http://edgar.jrc.ec.europa.eu/overview.php?v=42. Accessed 19 Oct 2015

Kalogirou SA (2001) Artificial neural networks in renewable energy systems applications: a review. Renew Sustain Energy Rev 5:373–401. doi:10.1016/S1364-0321(01)00006-5

Kalogirou SA (2003) Artificial intelligence for the modeling and control of combustion processes: a review. Prog Energy Combust Sci 29: 515–566. doi:10.1016/S0360-1285(03)00058-3

Kesgin U (2004) Genetic algorithm and artificial neural network for engine optimisation of efficiency and NOx emission. Fuel 83:885–895. doi:10.1016/j.fuel.2003.10.025

Kim SW, Heckel A, McKeen SA et al (2006) Satellite-observed U.S. power plant NOx emission reductions and their impact on air quality. Geophys Res Lett 33:L22812. doi:10.1029/2006GL027749

Klimont Z, Cofala J, Schöpp W et al (2001) Projections of SO2, NOx, NH3 and VOC emissions in East Asia up to 2030. Water Air Soil Pollut 130:193–198. doi:10.1023/A:1013886429786

May R, Dandy G, Maier H (2011) Artificial neural networks—methodological advances and biomedical applications. InTech, Croatia

Moriasi DN, Arnold JG, Van Liew MW et al (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 50:885–900, doi: 10.13031/2013.23153

NationMaster (2015) NationMaster: stats by category. http://www.nationmaster.com/statistics. Accessed 7 Nov 2015

OICA (2015) The International Organization of Motor Vehicle Manufacturers. In: Production Statistics | OICA. http://www.oica.net/category/production-statistics/. Accessed 20 Oct 2015

Olivier JG, Bouwman AF, van der Maas CW, Berdowski JJ (1994) Emission database for global atmospheric research (Edgar). Environ Monit Assess 31:93–106. doi:10.1007/BF00547184

Pope CA, Burnett RT, Thun MJ et al (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287:1132–1141. doi:10.1016/j.scitotenv.2011.03.017

Reşitoğlu İA, Altinişik K, Keskin A (2014) The pollutant emissions from diesel-engine vehicles and exhaust after treatment systems. Clean Techn Environ Policy 17:15–27. doi:10.1007/s10098-014-0793-9

Schopp W, Amann M, Cofala J et al (1999) Integrated assessment of European air pollution emission control strategies. Environ Model Softw 14:1–9

Shi Y, Xia Y, Lu B et al (2014) Emission inventory and trends of NOx for China, 2000–2020. J Zhejiang Univ Sci A 15:454–464. doi:10.1631/jzus.A1300379

Singh KP, Gupta S, Kumar A, Shukla SP (2012) Linear and nonlinear modeling approaches for urban air quality prediction. Sci Total Environ 426:244–55. doi:10.1016/j.scitotenv.2012.03.076

Smrekar J, Potočnik P, Senegačnik A (2013) Multi-step-ahead prediction of NOx emissions for a coal-based boiler. Appl Energy 106:89–99. doi:10.1016/j.apenergy.2012.10.056

Sözen A, Gülseven Z, Arcaklioğlu E (2009) Estimation of GHG emissions in turkey using energy and economic indicators. Energy Sources Part A: Recovery, Utilization and Environmental Effects 31:1141–1159. doi:10.1080/15567030802089086

Specht DF (1991) A general regression neural network. IEEE transactions on neural networks/a publication of the IEEE Neural Networks Council 2:568–76. doi:10.1109/72.97934

Stamenković LJ, Antanasijević DZ, Ristić MĐ et al (2015) Modeling of methane emissions using artificial neural network approach. J Serbian Chem Soc 80:421–433. doi:10.2298/JSC020414110S

Streets DG, Waldhoff ST (2000) Present and future emissions of air pollutants in China: SO2, NOx, and CO. Atmos Environ 34:363–374. doi:10.1016/S1352-2310(99)00167-3

The World Bank (2015) World Development Indicators | Data. http://data.worldbank.org/data-catalog/world-development-indicators. Accessed 7 Nov 2015

Tong DQ, Lamsal L, Pan L et al (2015) Long-term NOx trends over large cities in the United States during the great recession: comparison of satellite retrievals, ground observations, and emission inventories. Atmos Environ 107:70–84. doi:10.1016/j.atmosenv.2015.01.035

Tronci S, Baratti R, Servida A (2002) Monitoring pollutant emissions in a power plant through neural network. Neurocomputing 43:3–15. doi:10.1016/S0925-2312(01)00617-8

Van Amstel A, Olivier J, Janssen L (1999) Analysis of differences between national inventories and an Emissions Database for Global Atmospheric Research (EDGAR). Environ Sci Pol 2:275–293. doi:10.1016/S1462-9011(99)00019-2

Vestreng V, Ntziachristos L, Semb A et al (2008) Evolution of NOx emissions in Europe with focus on road transport control measures. Atmos Chem Phys 9:1503–1520. doi:10.5194/acp-9-1503-2009

Vu DH, Muttaqi KM, Agalgaonkar AP (2015) A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. Appl Energy 140:385–394. doi:10.1016/j.apenergy.2014.12.011

Wang WC, Chau KW, Cheng CT, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J Hydrol 374:294–306. doi:10.1016/j.jhydrol.2009.06.019

Ward Systems Group Inc (2008) Neuroshell 2. Ward Systems Group Inc, Frederick

Willmott CJ, Robeson SM, Matsuura K (2012) A refined index of model performance. Int J Climatol 32:2088–2094. doi:10.1002/joc.2419

Zhang GP, Patuwo EB, Michael YH (1998) Forecasting with artificial neural networks: the state of the art. Int J Forecast 14:35–62. doi:10.1016/S0169-2070(97)00044-7

Zheng LZL, Yu SYS, Yu MYM (2008) Monitoring NOx emissions from coal fired boilers using generalized regression neural network. 2nd IEEE International Conference on Bioinformatics and Biomedical Engineering 1916–1919. doi: 10.1109/ICBBE.2008.808

Zhou H, Cen K, Fan J (2004) Modeling and optimization of the NOx emission characteristics of a tangentially fired boiler with artificial neural networks. Energy 29:167–183. doi:10.1016/j.energy.2003.08.004

Zhou Y, Cheng S, Chen D et al (2014) A new statistical approach for establishing high-resolution emission inventory of primary gaseous air pollutants. Atmos Environ 94:392–401. doi:10.1016/j.atmosenv.2014.05.047