Environ Sci Pollut Res (2016) 23:10753–10762
DOI 10.1007/s11356-016-6279-z

RESEARCH ARTICLE

Estimation of NMVOC emissions using artificial neural networks and economical and sustainability indicators as inputs

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

Received: 6 December 2015 / Accepted: 11 February 2016 / Published online: 18 February 2016

Abstract This paper describes the development of an artificial neural network (ANN) model based on economical and sustainability indicators for the prediction of annual non-methane volatile organic compounds (NMVOCs) emissions in China for the period 2005–2011 and its comparison with inventory emission factor models. The NMVOCs emissions in China were estimated using ANN model which was created using available data for nine European countries, which NMVOC emission per capita approximately correspond to the Chinese emissions, for the period 2004–2012. The forward input selection strategy was used to compare the significance of particular inputs for the prediction of NMVOC emissions in the nine selected EU countries and China. The final ANN model was trained using only five input variables, and it has demonstrated similar accuracy in predicting NMVOC emissions for the selected EU countries that were used for the development of the model and then for China for which the input dataset was previously unknown to the ANN model. The obtained mean absolute percentage error (MAPE) values were 8 % for EU countries and 5 % for China. Also, the temporal trend of NMVOC emissions predicted in this study is generally consistent with the trend obtained using inventory emission models. The proposed ANN approach can represent a viable alternative for the prediction of NMVOC emissions at the national level, in particular for developing countries which are usually lacking emission data.

Keywords ANN · China · Emissions · NMVOC · Modeling

Introduction

Non-methane volatile organic compounds (NMVOCs) is a group of several hundred organic compounds with diversified chemical structures but with a similar behavior in the atmosphere, and importantly, they have a direct influence on the formation of photochemical smog, secondary organic aerosols, and organic acids in the air, and they are severely harmful to human health (Bo et al. 2008). The sources of NMVOC emissions into atmosphere are diverse and include natural, such as vegetation and agricultural respiration, as well as anthropogenic sources, such as industry, transport, and agriculture (Parra et al. 2006). The total NMVOC emissions at the national level and ratio among categories of sources vary between countries, depending on their level of development and economic structure (Niedojadlo et al. 2007).

Driven by rapid economic development and intensive energy use, emissions of atmospheric pollutants have caused serious environmental effects in China (Zhao et al. 2013). Globally, NMVOC emissions are mainly spread over industrialized countries, among which China is the largest emitter of NMVOCs in the world, with the major NMVOC emission sources being residential fuel combustion, solvent use, the petrochemical industry, on-road transportation, and fuel evaporation (Li et al. 2014). The emissions of NMVOCs in China are important because they are large in absolute terms and also a key measure for helping to understand the formation of ozone in East Asia (Klimont et al. 2002). The surface ozone and

Responsible editor: Marcus Schulz

✉ Davor Z. Antanasijević
dantanasijevic@tmf.bg.ac.rs

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

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

organic aerosol concentrations have increased considerably in urban and rural regions of China, and several studies have identified NMVOCs as the most important precursors for the formation of these pollutants (Guo et al. 2014; Qiu et al. 2014).

The main sources of national air pollution emissions data, including NMVOCs, are emission inventories, where estimates for a source of emission are performed by multiplying the emission factor and activity rate. The accuracy of an emissions inventory developed using the emission factor method depends largely on the accuracy and representativeness of the emission factors. Most of the previous work on the modeling of NMVOC emissions in China used emission factors either from the Environmental Protection Agency (US EPA) or from the European Environment Agency (EEA), generally without modifying the emission factors to fit the Chinese sources (Qiu et al. 2014). Consequently, the use of those emission factors resulted in high uncertainty in the existing emission inventories (Qiu et al. 2014). Fu et al. (2007) estimated, based on GOME satellite observations, that NMVOC emissions from China are 25 % higher than those reported by Streets et al. (2003).

Zhao et al. (2013) have developed the latest Chinese NMVOC emission inventory, with the data being derived from official Chinese Statistics, using an “emission factor method.” In this case, the emissions from each sector in each province were calculated from the activity data, technology-based emission factors, and penetration of emission control technologies. Their work has demonstrated that the emission of NMVOCs in China between 2005 and 2010 increased by 21.0 %.

In the last decade, artificial neural networks (ANNs) have arisen as an alternative approach for estimating emissions of various pollutants (Lim et al. 2007; Antanasijević et al. 2013, 2014). ANNs are considered as standard nonlinear estimators, and their predictive and generalization abilities have been demonstrated through their successful applications in a variety of fields (Gardner and Dorling 1998; Fernando et al. 2012; Singh et al. 2012; Hájek and Olej 2012; Russo et al. 2013). The main difference between the emission factor method and the ANN approach is that the ANN approach requires a significantly smaller number of input parameters and that they are widely available, even for countries where emission-related data is scarce.

This paper describes the development of an ANN model for the prediction of annual emissions of NMVOCs in China, using broadly available economical and sustainability indicators as inputs. Since the quantity of available Chinese NMVOC emissions data was not sufficient to properly train the ANN model, the NMVOC emissions model was developed using the available data for some of the largest NMVOC emitting EU countries, which NMVOC emissions per capita approximately correspond to the Chinese emission, and the developed ANN model was then applied utilizing the Chinese input data.

Materials and methods

Data collection and selection

Selection of inputs

The selection of appropriate input variables is a complex and crucial task, because the performance of the resulting model may be compromised if significant input variables are not included in the model. Conversely, if it includes too large a number of input variables, computational efficiency may be decreased, training becomes more difficult, and the model parameters are less well defined, potentially making the model validation in terms of physical plausibility, as well as knowledge extraction, problematic (Li et al. 2015). Generally, the selection of input variables can be divided into two

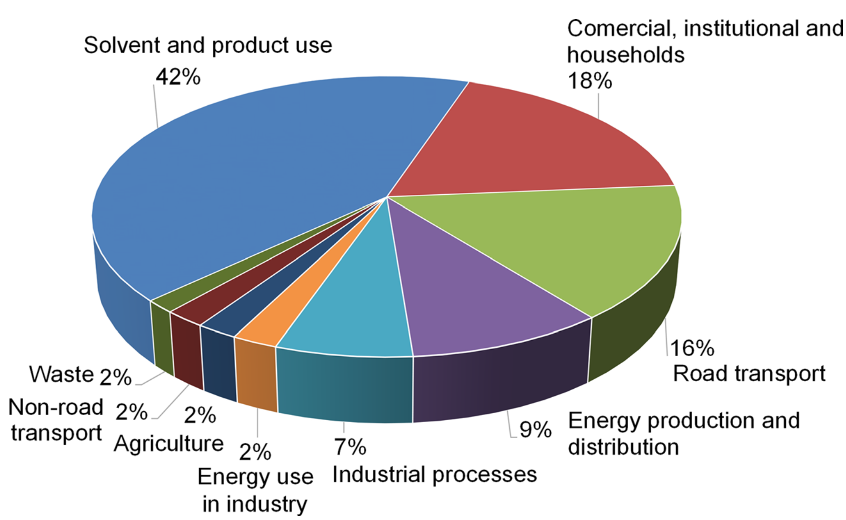


Fig. 1 Contribution of different sectors to NMVOC emission in EU countries

Table 1 Input variables and it descriptive statistics for selected EU countries

Input variable[a]	Unit after normalization	Descriptive statistics		
		Mean	Min	Max
Electricity generated from renewable sources (EGRS)	%	17.9	1.7	60.0
Electricity consumption by households (ECH)	toe pc	0.18	0.10	0.41
Agricultural land (AL)	ha pc	0.37	0.11	1.07
Chemical shipments (CS)[b]	$ pc	3514	1140	14942
Motor vehicle production (MVP)[c]	number[d]	39.74	3.33	88.70
Consumption total petroleum products (CTPP)	toe pc	1.59	1.01	2.36
Final energy consumption in transport (FECT)	toe pc	0.89	0.66	1.35
Gross domestic product (GDP)	€ pc	29627	20200	41000
Municipal waste generation (MWG)	kg pc	554	455	792
Gross inland energy consumption (GIEC)	toe pc	4.13	2.72	5.78

[a] Eurostat (2015)

[b] American Chemistry Council (ACC 2015)

[c] The International Organization of Motor Vehicle Manufacturers (2015)

[d] Per 1000 people

approaches, namely model-free and model-based. The major difference between these approaches is that the model-free approach does not depend on the model structure and training (Alves da Silva et al. 2008).

In this study, the definition of the input variables was based on forward selection strategy that selects individual variables one at a time. This selection method starts by training n single-input ANN models, where n is the number of potential variables, and selecting the input variables that maximize the model performance (May et al. 2011). The forward selection strategy proved to be suitable for the comparison of the significance of inputs in cases when the model has been created and trained on one country or defined set of countries and then needs to be applied to other countries.

Available input and output variables

The contribution of each sector in Europe's NMVOC emissions, obtained from the national emissions reported to the Convention on Long-range Transboundary Air Pollution (LRTAP Convention), according to the European Environment Agency (EEA 2014) is presented in Fig. 1.

Several studies have shown that sustainability and economical/industrial parameters are suitable for the creation of ANN models used for the estimation of pollutant emission into the atmosphere (Alonso et al. 2007; Sözen et al. 2009; Antanasijević et al. 2013, 2014; Khoshnevisan et al. 2013). In order to cover the activity of the key emission sources, different economical and sustainability indicators were identified as potential model inputs (Table 1). Beside the activity level indicators, the gross domestic product (GDP) is selected in order to quantify the use of emission control technologies that can significantly alter the emissions of air pollutants. In order to be included into the dataset for the creation of NMVOC model, each EU country needs to meet two criterions: to have approximately similar NMVOC emissions per capita as China and to have available data for all the inputs used within the studied period. The list of nine EU countries chosen for this research and descriptive statistics of their annual NMVOC emissions

Table 2 Descriptive statistic of NMVOC emission (kg pc) in selected EU countries in the period 2004–2012

Country	NMVOC emission		
	Mean	Min	Max
France	15.53	11.04	22.02
Germany	12.72	11.23	14.07
Italy	18.13	14.49	22.09
UK	16.51	13.30	20.81
Belgium	11.93	9.48	14.43
Ireland	11.85	9.57	14.63
The Netherlands	9.81	8.88	10.79
Spain	16.79	13.44	21.18
Sweden	21.13	19.85	22.72

Table 3 Dataset for China

Input/output[a]	Unit	Year						
		2005	2006	2007	2008	2009	2010	2011
EGRS[b]	%	18.1	17.4	17.0	19.5	19.5	21.2	19.0
ECH	toe pc	0.019	0.022	0.027	0.029	0.031	0.033	0.036
AL	ha pc	0.100	0.099	0.099	0.092	0.091	0.090	0.090
CS[c]	\$ pc	225	281	380	508	564	732	970
MVP[d]	number[e]	4.4	5.5	6.7	7.0	10.4	13.7	13.7
CTPP	toe pc	0.27	0.29	0.29	0.30	0.33	0.36	0.37
FECT	toe pc	0.08	0.10	0.11	0.12	0.12	0.14	0.15
GDP[f]	€ pc	1380	1650	2114	2722	2989	3535	4343
MWG[g]	kg pc	119	113	115	117	118	122	125
GIEC	toe pc	1.27	1.38	1.49	1.54	1.61	1.70	1.81
NMVOC[h]	kg pc	14.5	15.3	15.6	15.6	16.3	17.1	–

[a] National Bureau of statistics of the people's Republic of China (2015)

[b] US Energy Information Administration (EIA) (2015)

[c] American Chemistry Council (ACC 2015)

[d] The International Organization of Motor Vehicle Manufacturers (2015)

[e] Per 1000 people

[f] The World Bank (2015)

[g] The Organisation for Economic Co-operation and Development (OECD) (2015)

[h] Zhao et al. (Zhao et al. 2013)

(kg per capita), obtained from the EDGAR project (JRC-IES 2011), are presented in Table 2, while the descriptive statistics of inputs for selected EU countries is presented in Table 1.

To achieve better performance of the ANN model and also to allow comparison of countries of different sizes, the selected input and output parameters are normalized per capita. The NMVOC model was trained with the data for nine European countries for the period 2004–2011, while testing was performed with the data from 2012. In addition, 20 % of the data had been randomly extracted from the training dataset and used for the internal validation of models, in order to prevent the overtraining of network. After the NMVOC model had been tested with the data from the EU countries, it was applied to the Chinese data for the period 2005–2011 (Table 3).

Table 4 The R^2 value for created single-input ANN models

Input	EU countries[a] R^2	China[b] R^2
EGRS	0.79	0.58
ECH	0.69	N/A
AL	0.68	N/A
CS	0.28	N/A
MVP	0.28	0.74
CTPP	0.26	N/A
FECT	0.05	N/A
GDP	0.01	N/A
MWG	0.00	N/A
GIEC	0.00	N/A

N/A not applicable

[a] When applied on data from 2012

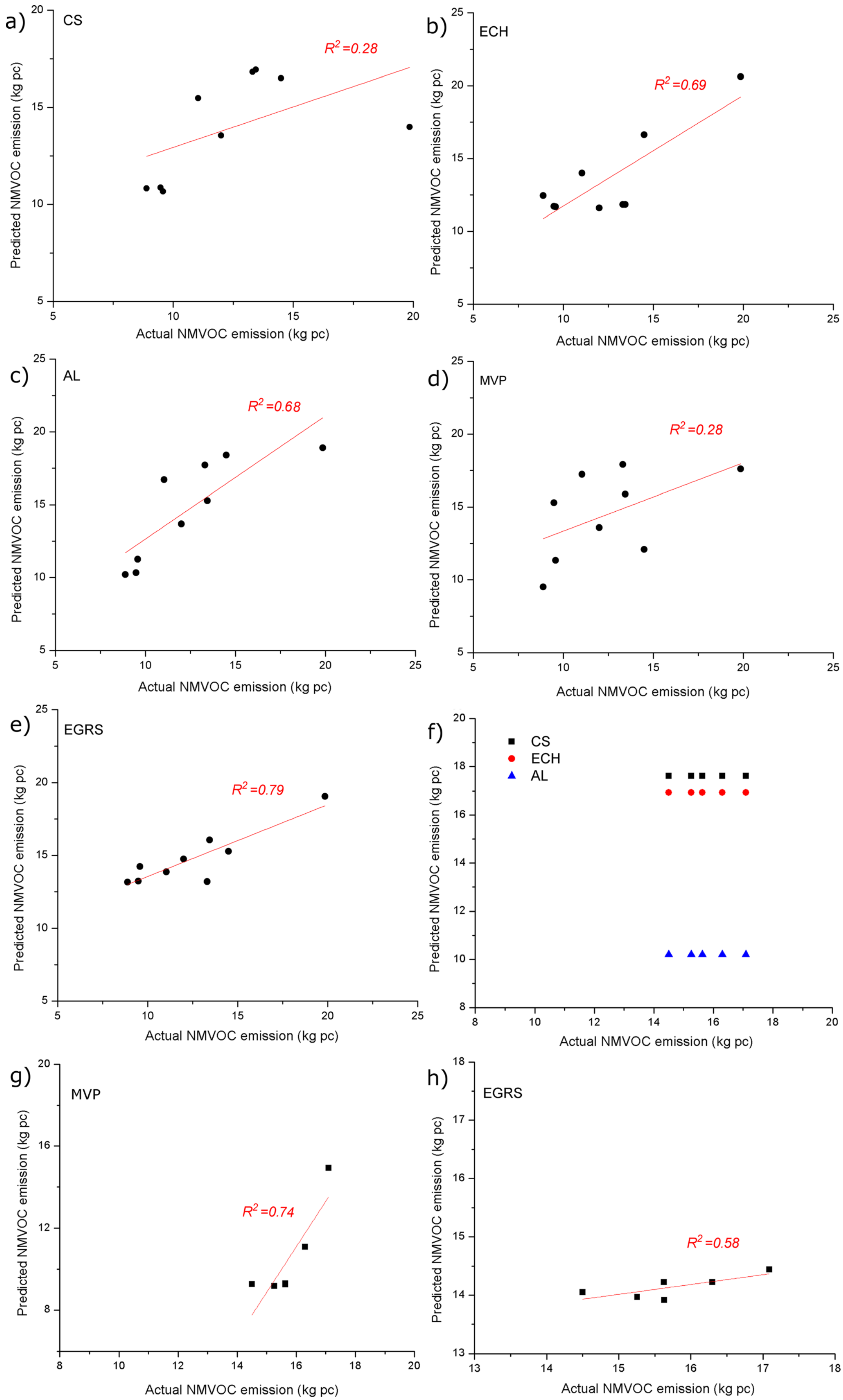
[b] When applied on data from 2005–2010

Brief description of ANNs

ANNs have been developed for modeling complex and nonlinear relations between a predicted parameter and predictor variables. Typically, an ANN architecture consists of neurons that are grouped into three layers: the input layer, the hidden layer, and the output layer (Özbay 2012). Input neurons receive the values of input parameters that are fed into the network and store the scaled input values, while the results are calculated in the hidden layer and presented in the output layer by the output neurons (Pai et al. 2013).

Neurons from the different layers are connected to each other by synapses, for which strength is quantified by a

Fig. 2 Comparison of the actual and predicted NMVOC emission using single-input ANN models: **a**–**e** testing with EU data for 2012 and **f**–**h** application on Chinese data 2005–2010


a)
CS
R^2 =0.28
Predicted NMVOC emission (kg pc)
Actual NMVOC emission (kg pc)
b)
ECH
R^2 =0.69
Predicted NMVOC emission (kg pc)
Actual NMVOC emission (kg pc)
c)
AL
R^2 =0.68
Predicted NMVOC emission (kg pc)
Actual NMVOC emission (kg pc)
d)
MVP
R^2 =0.28
Predicted NMVOC emission (kg pc)
Actual NMVOC emission (kg pc)
e)
EGRS
R^2 =0.79
Predicted NMVOC emission (kg pc)
Actual NMVOC emission (kg pc)
f)
CS
ECH
AL
Predicted NMVOC emission (kg pc)
Actual NMVOC emission (kg pc)
g)
MVP
R^2 =0.74
Predicted NMVOC emission (kg pc)
Actual NMVOC emission (kg pc)
h)
EGRS
R^2 =0.58
Predicted NMVOC emission (kg pc)
Actual NMVOC emission (kg pc)

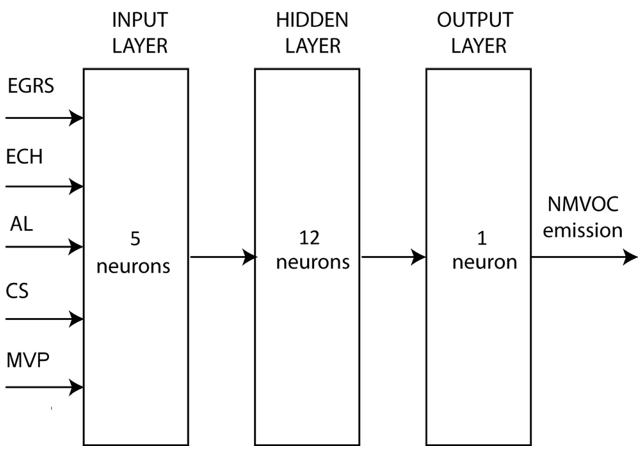


Fig. 3 The schematic representation of NMVOC ANN model

weighting factor. The weighted values are determined using a learning algorithm. Backpropagation (BP) is one of the most popular and common training algorithms used (Russo et al. 2013). The term backpropagation refers to the way the error is computed (Ul-Saufie et al. 2013):

- The training data set is propagated through the hidden layer and comes out of the neural network through the output layer.
- The output values obtained are then compared to the actual output values.
- The error between the output layer and the actual values is calculated and propagated back towards the hidden layer.
- This is followed by modifications to connection strengths based on the differences between computed and observed information signals of the outputs units.

Further details on the BP algorithm and its variants can be found elsewhere (Bishop 2006; Bolanča et al. 2008).

Among several other popular types of ANNs, the three-layer feed forward neural networks trained with a backpropagation algorithm (BPNN) are most widely used, because of their ability to generalize well in diversified research areas (Kalogirou 2003).

In this study, the input values were scaled in the range [−1,1] using the linear function, while the sigmoid (logistic) function was used as a neuron activation function. The determination of weighting factors was performed using the conjugate gradient BP algorithm (Bolanča et al. 2008) with the learning rate and momentum set to 0.1, while the initial weights were set to 0.3. The number of neurons in the input (N_i) and output (N_o) layer is equal to the number of inputs and outputs that are used in the BPNN model, respectively. In order to avoid time-consuming trial and error procedure, the number of hidden neurons (N_h) is determined using following rule-of-thumb method:

$$N_h = \frac{(N_i + N_o)}{2} + \sqrt{n_{tr}} \tag{1}$$

where n_{tr} is the number of training samples.

Performance metrics

The performance of models created in this study was evaluated using the following statistical criteria:

Nash-Sutcliffe efficiency (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(C_{o_i} - C_{P_i}\right)^2}{\sum_{i=1}^{n} \left(C_{o_i} - \overline{C}_o\right)^2} \tag{2}$$

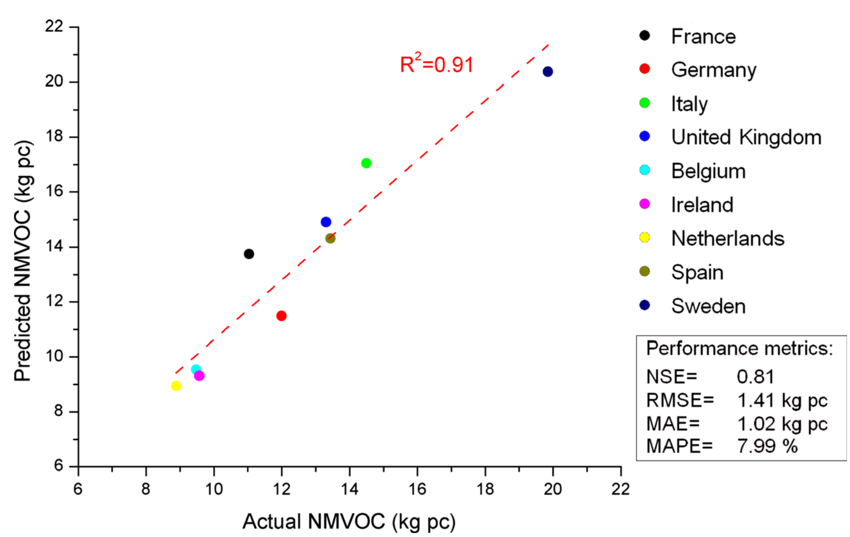


Fig. 4 Comparison of the actual and predicted NMVOC emission for EU countries

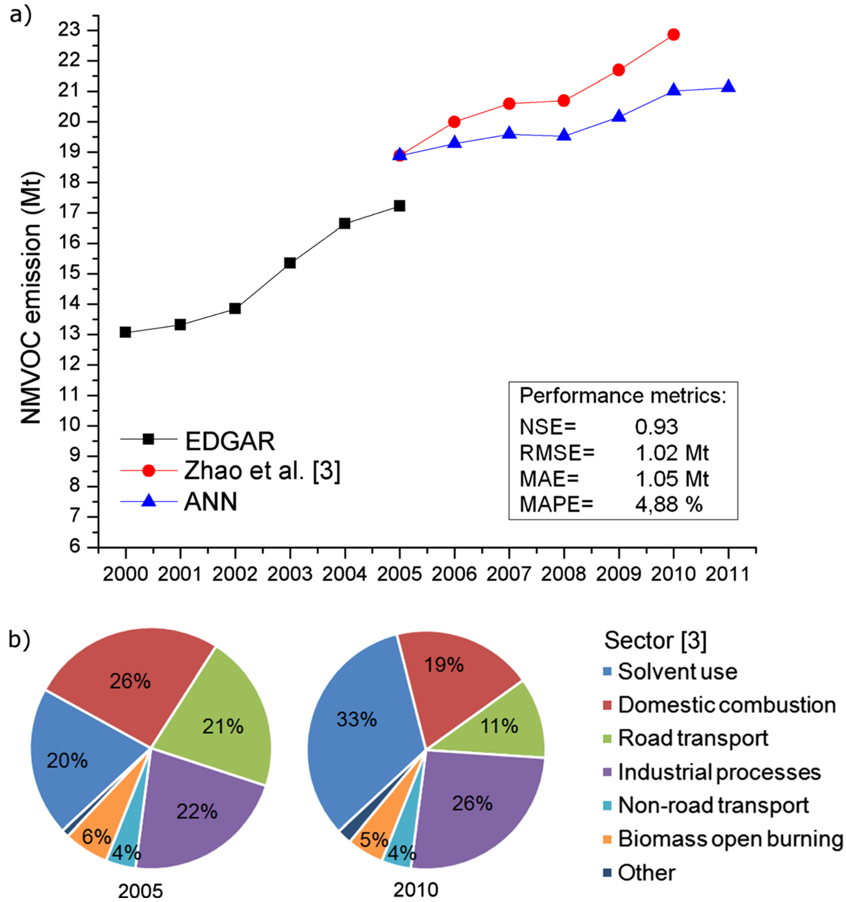


Fig. 5 Comparison of **a** the actual and predicted NMVOC emission for China and **b** contribution of different sectors to NMVOC emission in China Zhao et al. (2013)

The root mean squared error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(C_{O_i} - C_{P_i}\right)^2} \tag{3}$$

The mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|C_{O_i} - C_{P_i}\right| \tag{4}$$

Mean absolute percentage error (MAPE)

$$MAPE = 100\frac{1}{k}\sum\frac{\left|C_o - C_p\right|}{C_o} \tag{5}$$

Coefficient of determination (R^2)

$$R^2 = \left[\sum\left(C_p - \overline{C}_p\right)\left(C_o - \overline{C}_o\right)\right] / \sum\left(C_o - \overline{C}_o\right)^2 \sum\left(C_p - \overline{C}_p\right)^2 \tag{6}$$

where C_p and C_o are the predicted emissions and those obtained from EDGAR, respectively.

NSE is a normalized statistic that determines the relative magnitude of the residual variance ("noise") compared to the measured data variance ("information") (Nash and Sutcliffe 1970). NSE ranges between $-\infty$ and 1, where values between 0 and 1 represent acceptable model performance. RMSE provides a global idea of the difference between the observed and modeled values. MAE quantifies residual errors (Salazar-Ruiz et al. 2008). MAPE shows the mean value of relative error which was created by the model and expresses errors as a percentage of the observed data. The coefficient R^2, which ranges from 0 to 1, is defined as the proportion of the variance in the dependent variable that is predictable from the independent variable.

Results and discussion

First, single-input ANN models were created, in order to select the most significant inputs. The obtained R^2 values for the created single-input ANN models in their testing with EU data and their application on the Chinese data are presented in Table 4, while the actual vs. predicted plots for the most significant inputs are presented in Fig. 2.

As it can be seen in Table 4, in the case of the EU countries, all potential inputs, except final energy consumption in transport (FECT), gross domestic product (GDP), municipal waste generation (MWG), and gross inland energy consumption (GIEC), can be consider as significant for the prediction of NMVOC emissions. The abovementioned inputs are less significant, probably due to the similar development level and environmental standards that are implemented in the selected EU countries.

In contrast, the application of those models on the Chinese data showed that only two single-input ANN models, namely EGRS and MVP (Fig. 2g, h), are sufficiently sensitive to demonstrate significant differences in results as a consequence of input variations for different years.

Consequently, the final multi-input ANN model was created using those two inputs (EGRS and MVP) and including in addition ECH, AL, and CS, in order to enhance the performance of the model for the EU countries and thus implicitly for China. A schematic representation of the three-layer ANN model with five inputs used for the prediction of NMVOC emissions is presented in Fig. 3.

The comparison of actual and predicted NMVOC emissions for EU countries using the test dataset for 2012 is presented in Fig. 4. The results obtained can be regarded as very good, with the R^2 value of 0.91 and MAPE of 8 %.

The prediction of NMVOC emissions in China using the created ANN model for the period 2005–2011, along with the NMVOC emissions for period 2000–2005 obtained from EDGAR (JRC-IES 2011) and the NMVOC emission estimates for period 2005–2010 by Zhao et al. (2013), is presented in Fig. 5. A higher error (MAPE of 8 %) obtained on the testing with EU countries in comparison with the error obtained for Chinese data (MAPE of 4.88 %) is probably related with the higher heterogeneity of EU dataset.

It can be seen that the NMVOC estimates for 2005 obtained by EDGAR and by Zhao et al. (Zhao et al. 2013) differ by approximately 10 %. The differences between the two emission inventories can be traced down to the use of different emission factors or the use of national statistics that differed from the internationally available equivalent data (Van Amstel et al. 1999). Conversely, the predictions of NMVOC emissions obtained using the ANN model notably corresponds to the estimates made by Zhao et al. (2013) with a MAPE of 4.88 %. The differences between these two emission estimates are caused by a slower emission increase in 2006 and 2007 predicted by the ANN model: Zhao et al. (2013) estimated an increase in NMVOC emissions of 1.1 and 0.6 Mt in 2006 and 2007, respectively, while the ANN model predicts an increase of only 0.4 and 0.3 Mt during these 2 years. The stagnation of NMVOC emissions in 2008 is predicted by both models, as well as a progressive increase during the years 2009 and 2010. The stagnation of NMVOC emissions during 2008 is probably caused by an increase of 2.5 % in the production of electricity from renewable sources in China in comparison with 2007 (Table 3). Despite the discrepancy in increasing rates for some years, the temporal trend is generally consistent between Zhao et al. (2013) and this study. The stagnation of NMVOC emissions similar to that in 2008 is predicted for 2011 by the ANN model.

In the studied period, the "solvent use" sector became the dominant source of NMVOC emissions in China as well, with an increase of its share from 20 to 33 % (Fig. 5) (Zhao et al. 2013). Since the uncertainties of emissions data related to solvent use are typically very high, ranging up to ±30 % (Theloke and Friedrich 2004) with its share increasing, it can be assumed that the uncertainty of NMVOC emission estimates in later years has also increased, which led to differences in the predictions of NMVOC emissions between the ANN model and Zhao et al. (2013) (Fig. 5).

Conclusions

The main objective of this study was the development of an ANN model based on economical and sustainability indicators for the prediction of annual NMVOC emissions in China for the period 2005–2011 and its subsequent benchmarking with inventory emission factor models.

Available data for nine European countries (France, Germany, Italy, UK, Belgium, Ireland, The Netherlands, Spain, and Sweden) for the period 2004–2012 was used for training and testing the ANN model. The forward input selection strategy was used in order to simultaneously compare the significance of particular inputs for the prediction of NMVOC emissions in the EU countries and China, since the Chinese data was not used for the training of the ANN model.

It can be concluded that the applied use of this input selection approach was successful, resulting in an ANN model that demonstrated similar accuracy in predicting NMVOC emissions for the selected EU countries that were used for the development of the model and then for China for which the input dataset was previously unknown to the ANN model, e.g., the MAPE values were 8 % for EU countries and 5 % for China. The temporal trend of NMVOC emissions predicted in this study is generally consistent with the trend obtained using corresponding inventory emission factor models.

The presented ANN obtained these results, however, by using only five input variables instead of a large number of country and emission source-specific data, such as hundreds of emission factors and activity rates. Therefore, the proposed ANN approach can represent a viable alternative for the prediction of NMVOC emissions at the national level, in particular for developing countries that are also the largest emitters of pollutants into atmosphere, such as China, India, and Russia, and which usually lack emission inventory data.

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

Compliance with ethical standards

Funding This study was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Project No. 172007).

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

Alonso JM, Alvarruiz F, Desantes JM, Hernández L, Hernández V, Moltó G (2007) Combining neural networks and genetic algorithms to predict and reduce diesel engine emissions. IEEE T Evolut Comput 11:46–55

Alves da Silva AP, Ferreira VH, Velasquez RMG (2008) Input space to neural network based load forecasters. Int J Forecast 24:616–629

American Chemistry Council (2015) Chemical Shipments http://www.americanchemistry.com/. Accessed 21 April 2015

Antanasijević DZ, Pocajt VV, Povrenović DS et al (2013) PM 10 emission forecasting using artificial neural networks and genetic algorithm input variable optimization. Sci Total Environ 443:511–519

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

Bishop CM (2006) Pattern Recognition and Machine Learning. Springer, Singapore

Bo Y, Cai H, Xie SD (2008) Spatial and temporal variation of historical anthropogenic NMVOCs emission inventories in China. Atmos Chem Phys 8:7297–7316

Bolanča T, Cerjan-Stefanović Š, Ukić Š et al (2008) Application of different training methodologies for the development of a back propagation artificial neural network retention model in ion chromatography. J Chemom 22:106–113

European Environment Agency (EEA) (2014) Non-methane volatile organic compounds (NMVOC) emissions. http://www.eea.europa.eu/data-and-maps/indicators/eea-32-non-methane-volatile-1/assessment-4. Accessed 11 April 2015

Eurostat (2015) Database. http://ec.europa.eu/eurostat/data/database. Accessed 11 April 2015

Fernando HJS, Mammarella MC, Grandoni G et al (2012) Forecasting PM10 in metropolitan areas: efficacy of neural networks. Environ Pollut 163:62–67

Fu T-M, Jacob DJ, Palmer PI et al (2007) Space-based formaldehyde measurements as constraints on volatile organic compound emissions in east and south Asia and implications for ozone. J Geophys Res 112:1–15

Gardner M, Dorling S (1998) Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ 32:2627–2636

Guo S, Hu M, Zamora ML et al (2014) Elucidating severe urban haze formation in China. Proc Natl Acad Sci U S A 111:17373–17378

Hájek P, Olej V (2012) Ozone prediction on the basis of neural networks, support vector regression and methods with uncertainty. Ecol Inf 12: 31–42

Institute for Environment and Sustainability (JRC-IES) (2011) Global Emissions EDGAR v4.2. http://edgar.jrc.ec.europa.eu/overview.php?v=42. Accessed 22 April 2015

Kalogirou SA (2003) Artificial intelligence for the modeling and control of combustion processes: a review. Prog Energy Combust Sci 29: 515–566

Khoshnevisan B, Rafiee S, Omid M, Yousefi M, Movahedi M (2013) Modeling of energy consumption and GHG (greenhouse gas) emissions in wheat production in Esfahan province of Iran using artificial neural networks. Energy 52:333–338

Klimont Z, Streets DG, Gupta S et al (2002) Anthropogenic emissions of non-methane volatile organic compounds in China. Atmos Environ 36:1309–1322

Li M, Zhang Q, Streets DG et al (2014) Mapping Asian anthropogenic emissions of non-methane volatile organic compounds to multiple chemical mechanisms. Atmos Chem Phys 14:5617–5638

Li X, Maier HR, Zecchin AC (2015) Improved PMI-based input variable selection approach for artificial neural network and other data driven environmental and water resource models. Environ Model Softw 65: 15–29

Lim Y, Moon Y-S, Kim T-W (2007) Artificial neural network approach for prediction of ammonia emission from field-applied manure and relative significance assessment of ammonia emission factors. Eur J Agron 26:425–434

May R, Dandy G, Maier H (2011) Review of input variable selection methods for artificial neural networks. In: Suzuki K (ed) Artificial neural networks—methodological advances and biomedical applications. InTech, pp 19–44

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models: part 1. A Discus Princ J Hydrol 10:282–290

National Bureau of statistics of the people's Republic of China (2015) http://www.stats.gov.cn/

Niedojadlo A, Becker KH, Kurtenbach R, Wiesen P (2007) The contribution of traffic and solvent use to the total NMVOC emission in a German city derived from measurements and CMB modelling. Atmos Environ 41:7108–7126

Özbay B (2012) Modeling the effects of meteorological factors on SO2 and PM10 concentrations with statistical approaches. Clean Soil, Air, Water 40:571–577

Pai TY, Hanaki K, Su HC, Yu LF (2013) A 24-h forecast of oxidant concentration in Tokyo using neural network and fuzzy learning approach. Clean Soil, Air, Water 41:729–736

Parra R, Jiménez P, Baldasano JM (2006) Development of the high spatial resolution EMICAT2000 emission model for air pollutants from the north-eastern Iberian Peninsula (Catalonia, Spain). Environ Pollut 140:200–219

Qiu K, Yang L, Lin J et al (2014) Historical industrial emissions of non-methane volatile organic compounds in China for the period of 1980–2010. Atmos Environ 86:102–112

Russo A, Raischel F, Lind PG (2013) Air quality prediction using optimal neural networks with stochastic variables. Atmos Environ 79:822–830

Salazar-Ruiz E, Ordieres JB, Vergara EP, Capuz-Rizo SF (2008) Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence-based models in Mexicali, Baja California (Mexico) and Calexico, California (US). Environ Model Softw 23:1056–1069

Singh KP, Gupta S, Kumar A, Shukla SP (2012) Linear and nonlinear modeling approaches for urban air quality prediction. Sci Total Environ 426:244–255

Sözen A, Gülseven Z, Arcaklioğlu E (2009) Estimation of GHG emissions in Turkey using energy and economic indicators. Energ Source Part A 31:1141–1159

Streets DG, Bond TC, Carmichael GR et al (2003) An inventory of gaseous and primary aerosol emissions in Asia in the year 2000. J Geophys Res 108:8809

The International Organization of Motor Vehicle Manufacturers (2015) Production statistics. http://www.oica.net/category/production-statistics/. Accessed 11 April 2015

The Organisation for Economic Co-operation and Development (OECD) (2015) Municipal Waste. Treatment, Generation and, http://stats.oecd.org/Index.aspx?DataSetCode=WASTE. Accessed 21 April 2015

The World Bank (2015) Data. http://data.worldbank.org/country. Accessed 21 April 2015

Theloke J, Friedrich R (2004) NMVOC emmissions from solvent use in Germany. In: Friedrich R, Reis S (eds) Emissions of air pollutants: measurements, calculations and uncertainties. Springer-Verlag Berlin, Heidelberg, Berlin, Heidelberg, pp 63–79

Ul-Saufie AZ, Yahaya AS, Ramli NA et al (2013) Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). Atmos Environ 77:621–630

US Energy Information Administration (EIA) (2015) Electricity generated from renewable sources http://www.eia.gov/. Accessed 23 April 2015

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 Policy 2:275–293

Zhao B, Wang S, Dong X et al (2013) Environmental effects of the recent emission changes in China: implications for particulate matter pollution and soil acidification. Environ Res Lett 8:024031