Atmospheric Pollution Research 9 (2018) 388–397

HOSTED BY

Contents lists available at ScienceDirect

Atmospheric Pollution Research

journal homepage: www.elsevier.com/locate/apr

Multiple-input–multiple-output general regression neural networks model for the simultaneous estimation of traffic-related air pollutant emissions

Davor Antanasijević[a,*], Viktor Pocajt[b], Aleksandra Perić-Grujić[b], Mirjana Ristić[b]

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

A R T I C L E I N F O

Keywords:
ANN
MIMO modeling
Traffic emission
Outliers
Air pollutants

A B S T R A C T

Traffic-related air pollutant emissions have become a global environmental problem, especially in urban areas. The estimation of pollutant emissions is based on complex models that require the use of detailed travel-activity data, which is often unavailable and in particular, in developing countries. In order to overcome this issue, an alternative multiple-input–multiple-output general regression neural network model, based on basic socio-economic and transport related indicators, is proposed for the simultaneous prediction of sulphur oxides (SOx), nitrogen oxides (NOx), ammonia (NH_3), non-methane volatile organic compounds (NMVOC) and particulate matter emissions at the national level. The best model, created using only six inputs, has MAPE (mean absolute percentage error) values on testing in the range of 12–15% for all studied pollutants, except NMVOC (MAPE = 21%). The obtained predictions for SOx, NH_3 and PM10 emissions were in good agreement with the reported emissions ($R^2 \geq 0.93$), while the predictions for NOx and NMVOC are somewhat less accurate ($R^2 \approx 0.85$). It can be concluded that the presented ANN approach can offer a simple and relatively accurate alternative method for the estimation of traffic-related air pollutant emissions.

1. Introduction

In last decades, due to rapid motorization and the sheer growth of the number of vehicles, traffic-related emissions in developing countries have been growing strongly causing air quality problems, especially in urban areas (Liaquat et al., 2010). Motor vehicles emit, among other pollutants, sulphur oxides (SOx), nitrogen oxides (NOx), ammonia (NH_3), non-methane volatile organic compounds (NMVOC) and particulate matter (PM), and represent a significant source of air pollution. Traffic-related pollutants, such as NO_2 and PM, are of particular concern to health, while other gaseous pollutants contribute to global warming, atmospheric acidification and the formation of secondary pollutants (Xia and Shao, 2005; Kousoulidou et al., 2008; Amato et al., 2014). Also, traffic-related air pollutants have received more concerns in the recent years because of its adverse health effect of early life exposure during pregnancy (Deng et al., 2016; Schultz et al., 2017; Song et al., 2017).

The quantification of road transport emissions is required in order to assess population exposure and impact on air quality (HEI Panel on the Health Effects of Traffic-Related Air Pollution, 2010). The development of traffic-related emission inventories is based on complex emissions models, e.g. EPA MOVES2014 (Motor Vehicle Emission Simulator) and EEA COPERT4 (Computer Programme to calculate Emissions from Road Transport), which require the use of detailed travel-activity data, such as types of vehicles, vehicle-miles traveled, and number of trips. Furthermore, the transferability and applicability of emission factors (EFs) investigated in the laboratory to real-world traffic conditions seems to be problematic in many cases (Corsmeier et al., 2005), yielding substantial uncertainties and limitations in the resulting emissions estimates (NARSTO, 2005). Also, EF values depend on traveling behavior and driving conditions, and can vary between different locations within the same country (Berkowicz et al., 2006; Davison et al., 2015; Reyna et al., 2015). In addition, road transport EF databases, developed for the USA and EU countries, often are not transferable to other countries, and substantial work is required in order to obtain suitable EFs.

Since the available transport data for developing countries is much less representative and reliable, and the variation in vehicle technology and driving conditions is much larger, Uherek et al. (2010) have assumed up to three times higher uncertainty for each pollutant in comparison with OECD members, achieving about 30–40% for SO_2 and NOx, 60% for NMVOC and 75% for PM emissions.

Peer review under responsibility of Turkish National Committee for Air Pollution Research and Control.
* Corresponding author.
E-mail address: dantanasijevic@tmf.bg.ac.rs (D. Antanasijević).

https://doi.org/10.1016/j.apr.2017.10.011
Received 28 August 2017; Received in revised form 25 October 2017; Accepted 30 October 2017
Available online 22 November 2017
1309-1042/

In recent years, non-linear statistical modeling techniques, mainly artificial neural networks (ANNs), have arisen as an alternative approach for the estimation of total national air pollutant emissions (Antanasijević et al., 2013, 2014; Stamenković et al., 2015, 2016, 2017). Those studies demonstrated that ANN models that use widely available socioeconomic and sustainability indicators as inputs can produce emission estimates with satisfactory accuracy. Concerning that those models can be applied only for the estimation of total national emissions, this methodology has been extended for the prediction of sectoral air pollutant emissions, e.g. energy related GHG emissions (Antanasijević et al., 2015) and ammonia emission from field-applied manure (Lim et al., 2007). In the transportation sector, the use of ANNs was limited to the estimation of exhaust emissions of gasoline engine (Sayin et al., 2007) or certain types of vehicles, e.g. Mudgal et al., (2011) predicted the emissions of five pollutants from biodiesel fueled transit buses using separate ANN models. The simplicity and robustness of ANN methodology is especially useful in cases when the emission inventories methodology cannot be applied because of a lack of data.

The multi-input-single-output (MISO) and multiple-input–multiple-output (MIMO) ANN architectures are two basic structures that have been frequently reported in the literature. Concerning that MIMO ANN architecture simplifies the ANN-based model development, it has been used for the prediction of various scientific and engineering fields, e.g. energy related analysis (Gareta et al., 2006; Kumara et al., 2013; An et al., 2013), flood forecasting (Chang et al., 2007), meteorological parameters prediction (Raza and Jothiprakash, 2014), physical and chemical properties prediction (Ghaedi, 2015), etc.

The aim of this work is to present the development of a MIMO ANN model for simultaneous prediction of five traffic-related pollutants (SOx, NOx, NH_3, NMVOC and PM10). All studied pollutants came from the same source (traffic), thus following the parsimony principle a desired level of accuracy can be achieved using a single (MIMO) model with as few inputs as possible.

2. Materials and methods

In this section, a description of data sources for traffic emission is given, along with its statistics and correlation analysis. The same information is then presented for the selected inputs, together with a brief overview of GRNN architecture used for the creation of MIMO model.

2.1. Traffic emission data

The road transport emission data for SOx, NOx, NH_3, NMVOC and particulates with a diameter < 10 μm (PM10) were obtained from the European Environment Agency (EEA) through the Eurostat Air Pollution Database (Eurostat, 2015). This Air Pollution Database was created using annual reports under the Convention on Long-range Transboundary Air Pollution (LRTAP Convention). The emission data covers 26 European countries with a combined population of more than 500 million people and a period spanning nine years (2005–2013). Traffic emissions are expressed in kg per capita in order to allow comparison of countries of different sizes. Available data is split into two datasets (with the ratio of cca. 4:1): the training set, which comprised the data from 2005 to 2011 and was used for the development of the model, and the test set that includes the remaining data (22%) from 2012 to 2013. In the studied European countries (see Table S1 in Supplementary material) significantly lower emissions per capita (< 1 kg pc) have been reported for SOx and NH_3, while other pollutants, e.g. NOx, are emitted up to 18 kg pc. Descriptive statistics (mean, max, min and range) for each pollutant and each dataset are presented in Table 1. A notable decrease of emissions of SOx, NOx and NMVOC between the periods of the training and test datasets can be observed.

The correlation analysis results (Table 2) show that the mutual correlation between the outputs is low ($r \le 0.55$), hence the simultaneous prediction of uncorrelated outputs presented in this case further increases the complexity of ANN learning.

Table 1
Descriptive statistics for pollutants and datasets.

Output variables	Statistics	Training data (n = 182)	Test data (n = 52)	Reduction [%]
SOx [kg pc]	Mean	0.052	0.012	77
	St. dev.	0.133	0.008	
	Range	0.004–1.004	0.004–0.051	95
NOx [kg pc]	Mean	8.732	6.818	22
	St. dev.	2.299	1.728	
	Range	3.945–17.895	3.443–11.733	41
NH_3 [kg pc]	Mean	0.207	0.149	28
	St. dev.	0.127	0.079	
	Range	0.019–0.600	0.019–0.292	53
NMVOC [kg pc]	Mean	2.850	1.770	38
	St. dev.	1.201	0.759	
	Range	0.751–6.697	0.532–3.803	45
PM10 [kg pc]	Mean	0.665	0.535	20
	St. dev.	0.338	0.305	
	Range	0.213–2.084	0.220–1.959	7

Table 2
Correlation analysis of output data.

	SOx	NOx	NH_3	NMVOC	PM10
SOx	1				
NOx	−0.02	1			
NH_3	−0.06	0.48	1		
NMVOC	0.18	0.30	0.41	1	
PM10	0.01	0.44	0.54	0.55	1

2.2. Input data

Inputs were selected with a general aim to obtain a simple, robust and widely applicable model. Therefore, only basic socioeconomic and transport related indicators were utilized for the development of the model. Motorization rate (Mrate), Age of passengers cars (APC) and Final energy consumption in transport (FECtr) were selected because they provide information related to the number, age and use of vehicles. Gross domestic product (GDP) was also used because of the relationship between environmental quality and economic development (Kuznets curve) (Kuznets, 1955). Considering that traffic-related emissions depend on driving conditions, the urban population indicator (UP), which refers to people living in urban areas, was taken into account as well. Since several studies have reported that residential density has an effect on transport emissions (Brownstone and Golob, 2009; Hong and Shen, 2013), the population density (PD) was also used as model input. Finally, in order to cover the average distance of a round work trip, job density (JD) was calculated by multiplying PD with the employment rate. Apart from APC, which is calculated as presented in our previous study (Antanasijević et al., 2014), and UP that is acquired from World Bank, all other inputs were obtained from Eurostat. Descriptive statistics of selected inputs are presented in Table 3.

As in the case of the model outputs, the correlation analysis was used in order to quantify mutual correlation among the selected inputs. In contrast to the correlation of outputs, where high mutual correlation is useful, correlated input data often introduces confusion during the ANN learning process (Abdul-Wahab et al., 2005).

As can be seen in Table 4, besides the JD and PD which are perfectly correlated, the average mutual correlation of other inputs is low ($r_{mean} = 0.39$ and $r < 0.80$). In order to determine the particular significance of JD and PD on the performance of the model, two separate GRNN models with six input parameters were created.