Available online at http://ap.iauardabil.ac.ir
Islamic Azad University, Ardabil Branch
Anthropogenic Pollution, Vol 1 (1), 2017: 61-67
DOI: 10.22034/apj.2017.1.1.6167

APJ

Prediction of Ground-Level Air Pollution Using Artificial Neural Network in Tehran

[1*]Afshin Khoshand, [2]Mahshid Shahbazi Sehrani, [3]Hamidreza Kamalan, [4]Siamak Bodaghpour

1. Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran -Email: khoshand@kntu.ac.ir
2. Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran
3. Department of Civil Engineering, Pardis Branch, Islamic Azad University, Pardis, Iran
4. Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran

Received: 9 September 2017 / **Accepted**: 18 October 2017 / **Published**: 30 October 2017

ABSTRACT: Novel technologies and subsequent pollutions are serious threats to the environment and public health. The environmental pollutions, especially air pollution, are currently leading environmental concerns in developing countries, including Iran. In the present study, the air quality and meteorological data were employed to achieve potent models based on an Artificial Neural Network (ANN) for the prediction of air pollution in Tehran, Iran. The developed models manage to predict daily concentrations of various air pollutants such as O_3, PM_{10}, NO_2, CO, and $PM_{2.5}$. The required data were collected daily through the Air Quality Organization from all air quality stations of Tehran within a four-year period (from 2012 to 2015). Training the models was on the basis of Multi-Layer Perceptron (MLP) with the Back Propagation (BP) algorithm using MATLAB program. The results indicated appropriate agreement between the observed and predicted concentrations, as the values of the coefficient of multiple determinations (R^2) for all models were more than 0.83. In conclusion, the studied meteorological parameters are effective on all pollutants concentrations.
Keywords: Air Pollution, Artificial Neural Network, MATLAB, Tehran

1. INTRODUCTION

Despite technological advancements concerning the emissions of air pollutants over the last years, the air pollution remains a major concern in industrial cities, which can be harmful not only for the environment but also for human health (Schindler et al. 2017). Moreover, urban residents are influenced by the health impact of air pollution in their close neighborhood (Chay and Greenstone 2005) and exhibit eagerly willingness to live in less polluted neighborhoods (Smith and Huang 1995; Bickersta and Walker 2001; Lera-Lopez et al. 2012). The air pollutants exert a wide range of adverse effects on the biological, physical, and economic systems of which the biological impacts on human health are major concerns (Barai et. al. 2003). Atmospheric pollution can be caused by mobile, stationary and area sources. Primary (which is emitted directly from a source) and secondary (which is not directly emitted as such) pollutants can result in the ambient air quality degradation in the industrial cities and put people at risk of daily exposure to air pollution (Baawain et al. 2007). The negative effects of air pollution determine the importance of predicting air pollution at ground level as an effective alarming system that allow time to generate a specific response in severe episodes. It is not easy to predict the air quality due to the fact that incomplete or absence of reliable environmental data often comes across in the environmental research (Baawain and Al-Serihi 2014). This situation might be a result of inappropriate sampling, mistakes in measurements and obvious errors in data acquisition (Junninen et al. 2004). However, discontinuities in data represent an important barrier for proposition schemes of time series that usually need relentless data to achieve satisfactory effectiveness (Sahin et al. 2011). Therefore, general modeling approach that can deal with discontinuous and noise in data as well as capture the complex interactions within data with satisfactory efficiency is necessary to obtain reliable predictive outcomes. The ANN models seem an appropriate choice for this purpose since they have been found to be remarkably well in capturing complex interactions within the given input parameters (Baawain et al. 2007).

A review of the available literature reveals that ANNs have been applied successfully to predict the ground-level air pollution. According to de Gennaro et al. (2013), Cheng et al. (2012), Gobakisa et al. (2011), and Kurt and Oktay (2010), the neural networks are promising tools for air quality prediction when comparing with other statistical models, such as regression-based models. Moreover, the ANNs, in particular the MLP, have a reliable performance in dealing with highly nonlinear systems such as the phenomenon of interaction between air pollution and climatic condition (Abdul-Wahab and Al-Alawi 2008).

2. MATERIALS AND METHODS

2.1. Artificial Neural Networks

The neural networks are a computational technique based on a large collection of neural units modeling the routes of which the brain addresses problems with huge clusters of biological neurons connected by axons. The ANNs are computing systems, motivated by biological models, made up of a number of simple and highly interconnected processing components, which process information by its dynamic state response to external inputs. The number of layers can vary from a single layer to multiple layers in the ANNs. The neurons may be linked in many ways (Agirre-Basurko et. al. 2006; Haykin et. al. 1999), depending on the characteristics of the proposed problem. To design an ANN model, the air pollution system is considered to be a system receiving information from distinct sets of inputs, X_i (i = 1 to n), namely climatic parameters and air pollution properties, which produces a specific output (Gardner and Dorling 1998).

2.2. Area Description

Tehran as one of the industrial cities in Iran covers a total area of 730 km^2. The major pollutants emitted in atmosphere by mobile, stationary and area sources in Tehran include O_3, PM_{10}, NO_2, CO and $PM_{2.5}$.

2.3. Data Sets

The data used in this study are air quality and meteorological parameters with daily concentrations of O_3, PM_{10}, NO_2, CO and $PM_{2.5}$, as well as 3-hour records of air temperature, relative humidity, wind speed and wind direction for a period from 2012 to 2015. Iran Meteorological Organization (IMO) provided all of these meteorological data. The used ambient air quality data were collected through the Air Quality Organization from all the air quality stations in Tehran. The 3-hour meteorological data were processed into daily format and used to construct a data series in a two- dimensional table (the columns represent variables including concentrations of O_3, PM_{10}, NO_2, CO and $PM_{2.5}$, and air temperature, relative humidity, wind speed and wind direction, and the rows refer to observation date).

The data were divided into two sets, learning set for training the ANN and testing set for verifying the efficiency and correctness of the developed model. Because there is no universal rule to determine the size of subsets, the data set for this project were randomly subjected to a ratio of 3:1 between training and testing sets.

2.4. Data Preparation

Preparing data for the neural network data analysis is an important and critical step that has an immense impact on the success and performance of the ANN results (Yu et al. 2006). The data preparation in current study was included firstly inspecting the data set for missing data and data noise. The entire data set covered a period of 1 January 2012 to 30 December 2015. Some data were missed due to instrument calibrations or malfunctions. Therefore, the datasets with errors and/or incomplete information were entirely removed. Finally, extreme values were deleted using the Standardized Score Method (SSM). These gaps refilled by the linear interpolation method. Since a separate model was developed for each pollutant, the selection of input variables varied from one model to another one. The selection of predictor variables was based on a comprehensive review of the theoretically addressed chemical and weather processes that influence the formation and concentration of atmospheric pollutants, as described by Seinfeld and Pandis (1998), Wayne (1985) and USEPA (2003). Furthermore, in order to support the neural network to deal with the data, all the input data were normalized to the range of [0, 1] by linear scaling.

2.5. Model Development

To develop the prediction model for air quality, MATLAB Neural Network Toolbox was used since is a flexible and user-friendly toolbox. The Neural Network Toolbox offers a broad variety of parameters for neural network development which can be chosen.

This software is able to normalize all the data to the range of [0, 1]. The feedforward Back Propagation (BP) Multi-Layer Perceptron (MLP) neural network architecture was selected for the air quality modeling in the current research. BP is one of the most popular training procedures, which is used for non-linear models (Fig.1) (Trigo et. al. 1999).

Table 1: Common predictor variables proposed for each model

Input parameters	WD[1]	WD	WD	WD	WD
	WS[2]	WS	WS	WS	WS
	T[3]	T	T	T	T
	RH[4]	RH	RH	RH	RH
	CO NO_2	$PM_{2.5}$	CO O_3	NO_2	PM_{10}
Output parameter	O_3	PM_{10}	NO_2	CO	$PM_{2.5}$

[1] Wind Direction; [2] Wind Speed; [3] Temperature; [4] Relative Humidity

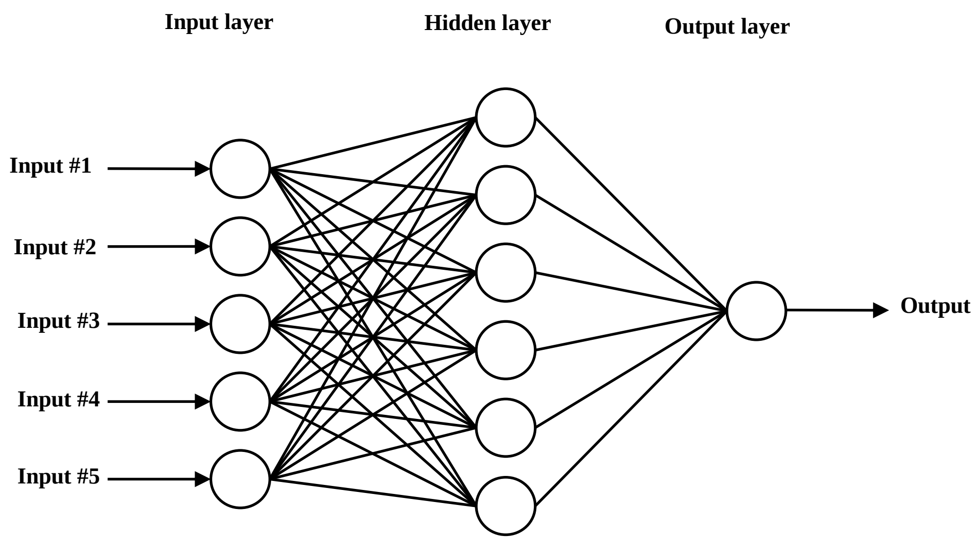

Figure 1: MLP neural network architecture for air quality prediction

The main objective of the current study was to design the ANN simulations of typical air pollutant concentrations from 2012 to 2015. Therefore, separate ANN model was developed for each pollutant of O_3, PM_{10}, NO_2, CO and $PM_{2.5}$. Table 1 shows the selected pollutants and meteorological parameters as input parameters for each model. The models predict the concentration of each pollutant on the next day using the previous conditions of all input parameters. In the present research, the MLP models were run in a supervised manner based on trial and error technique by which several alternative adjustments were incorporated to improve the model performance. These adjustments included alternative use of different number of neurons in the hidden layer, different epochs, and applying of different activation functions. The network training continued until achieving the highest correlation between the genuine and predicted output, which is expressed by coefficient of multiple determinations (R^2) and root means square error (RMSE).

3. RESULTS

The MLP-BP architectures that provided the best models are shown in Table 2. The best performing ANN models were found using training function TRAINLM and transfer function TANSIG. The results showed excellent performance for the developed networks of $PM_{2.5}$ and PM_{10} according to R^2 (0.94 and 0.92, respectively) and RMSE (0.01810 and 0.00322, respectively).

The obtained results for O_3, NO_2 and CO networks have an acceptable agreement with the experimental data (R^2 of 0.88, 0.83, and 0.83, and RMSE of 0.04697,0.0474 and 0.0417, respectively). Figures 2 to 6 provide the comparison between the observed values and predicted concentration of each pollutant.

The MLP-BP model that provided the best results to predict the O_3 consisted of one hidden layer with 50 hidden neurons. The obtained result indicates that the developed model (Fig.2) could explain 88% of the variability in O3 concentrations. The selected input variables for the model included CO, NO_2, wind speed, wind direction, temperature and humidity. The best performance for O_3 is probably due to the mechanism of O_3 production. The ambient air temperature and humidity are directly related to the solar radiation and highly important for ozone production (Wallace and Hobbs 1977; Seinfeld and Pandis 1998).

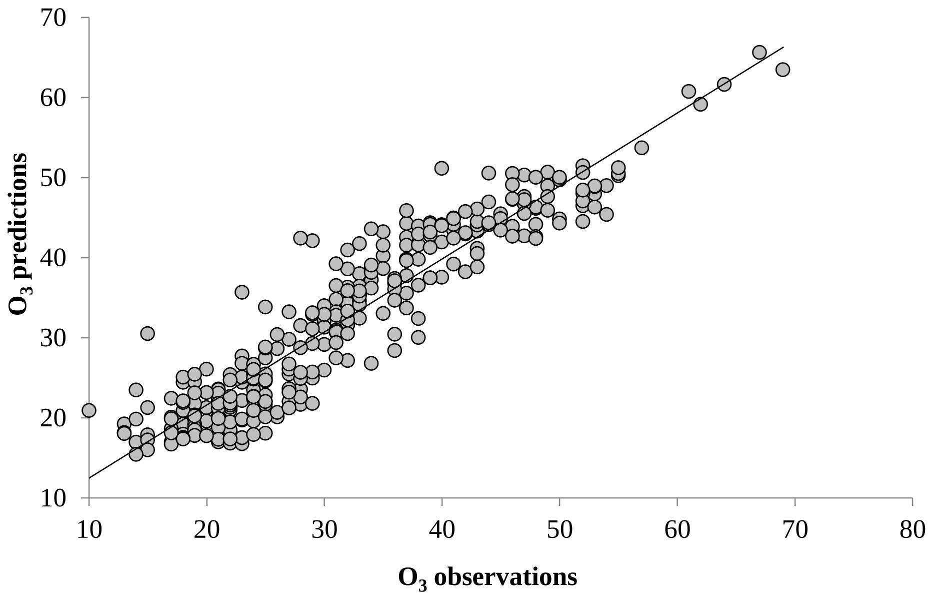

Figure 2: ANN predicted versus observation values for O_3

The ANN model used to predict PM_{10} concentrations included one hidden layer with 65 hidden neurons. Fig.3 shows the obtained R^2 (0.88) for this model for testing data of PM_{10}, indicating the importance of the selected input variables including $PM_{2.5}$, wind speed, wind direction, temperature and humidity in the prediction of PM_{10} concentration.

The best MLP-BP model used to predict NO_2 concentrations consisted of one hidden layer with 70 hidden neurons. The performance of ANN for NO_2 prediction was acceptable with R^2 value of 0.87. The high performance between measured and ANN predicted values of NO_2 has been presented in Fig.4.

The best MLP-BP architecture for CO prediction included one hidden layer with 68 hidden neurons. The ANN model performed very well in predicting the CO concentrations as a function of NO_2 and meteorological conditions including wind direction, wind speed, air temperature and humidity. The R^2 was 0.83, which reflects the influence of meteorological conditions in regulating the variability in CO concentrations (Fig.5).

The appropriate MLP-BP architecture for $PM_{2.5}$ prediction included one hidden layer with 60 hidden neurons. Figure 6 illustrates that the obtained R^2 vale is 0.9 for testing data of $PM_{2.5}$. It emphasizes the importance of input variables, which includes PM_{10}, wind speed, wind direction, temperature and humidity in prediction of $PM_{2.5}$ concentrations. It can be concluded that meteorological parameters are effective on all of studied pollutants concentrations.

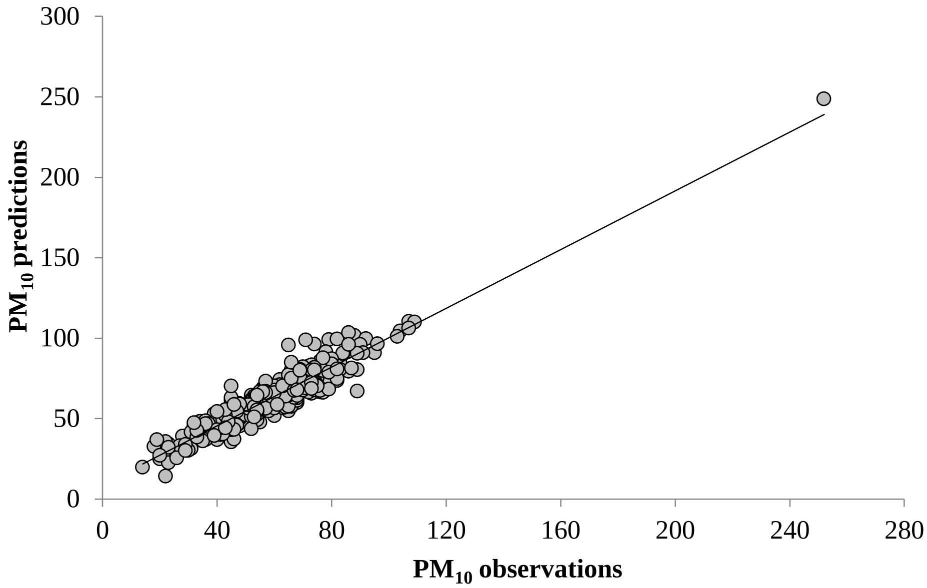

Figure 3: ANN predicted versus observation values for PM_{10}

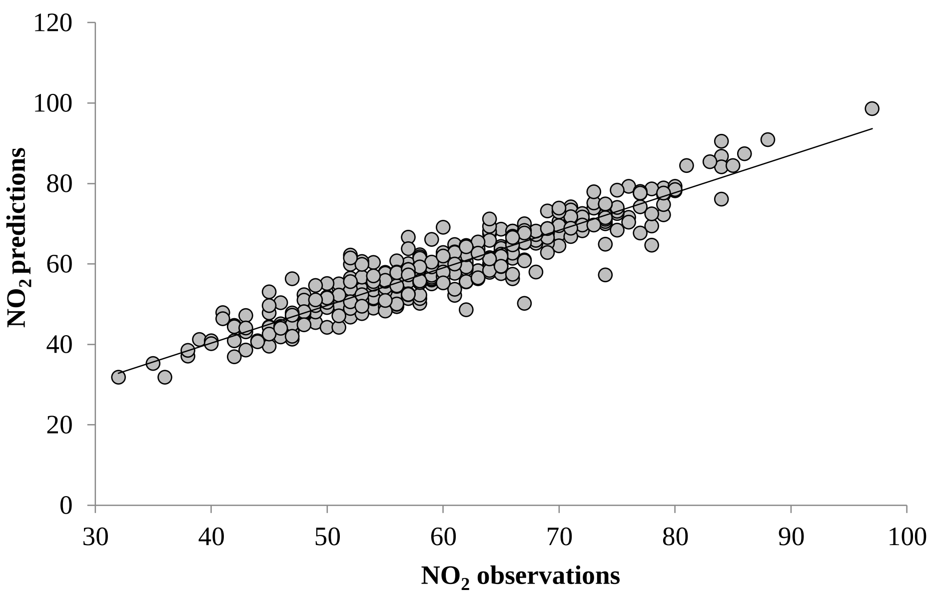

Figure 4: ANN predicted versus observation values for NO_2

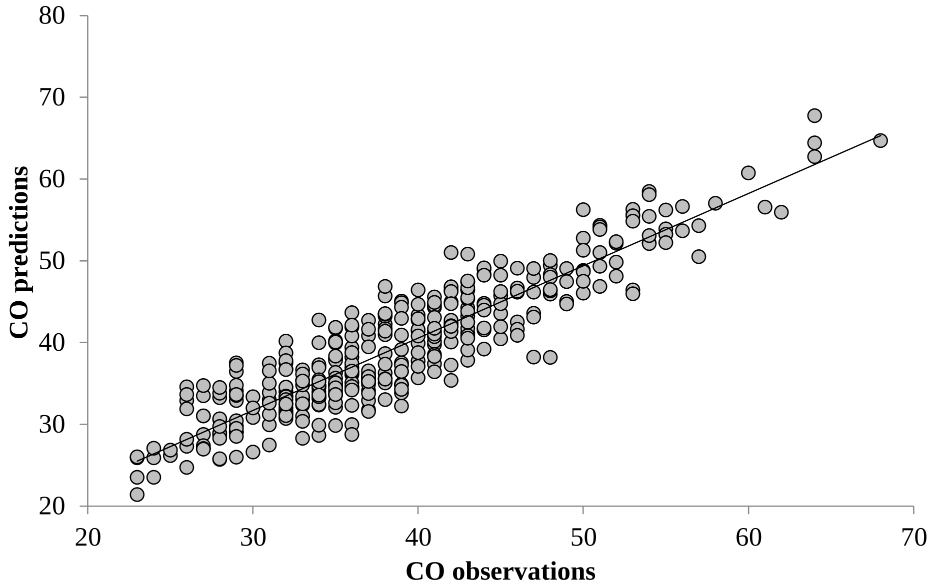

Figure 5: ANN predicted versus observation values for CO

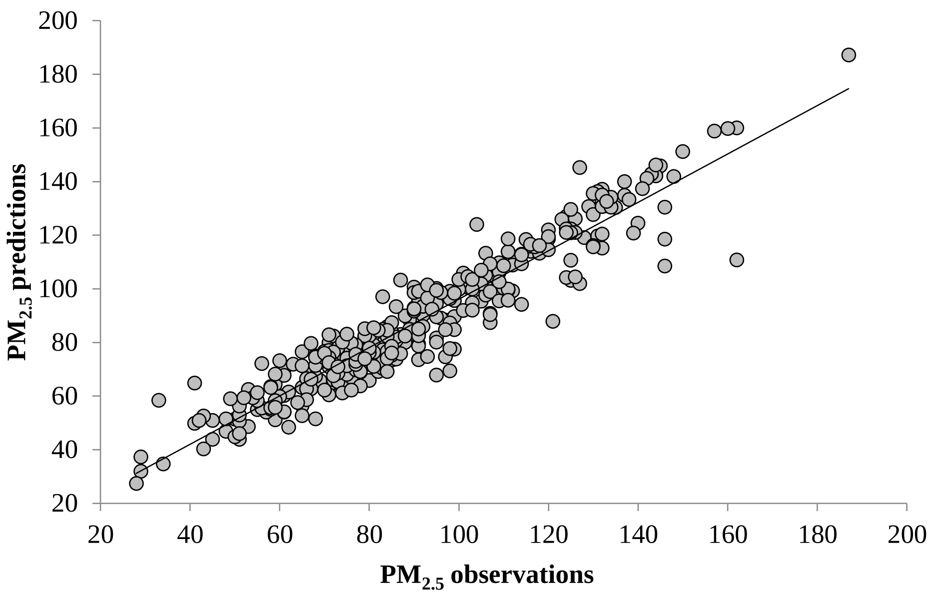

Figure 6: ANN predicted versus observation values for $PM_{2.5}$

Table 2: Best architectures for MLP neural network models

Model	Hidden Neurons No.	Training Function	Transfer Function	Training R	Testing R	R^2	RMSE
O_3	50	TRAINLM	TANSIG	0.9085	0.88532	0.8768	0.04697
PM_{10}	65			0.97533	0.92486	0.8833	0.00322
NO_2	70			0.90162	0.83322	0.8695	0.0474
CO	68			0.86686	0.83727	0.834	0.0417
$PM_{2.5}$	60			0.96453	0.94189	0.9004	0.0181

4. CONCLUSION

The present work has studied the use of ANN to predict the models for the ground-level air pollution in Tehran. The main objective was to predict the concentration of pollutants, including O_3, PM_{10}, NO_2, CO and $PM_{2.5}$ in the atmosphere according to their relationship with the air quality data of the previous day and meteorological conditions. Therefore, each pollutant was separately modeled. Each ANN model was trained using previous day conditions in order to predict the next day concentrations. The ANN models were trained by historical daily time series of air quality measurement as well as meteorological measurements. The models were developed using feedforward MLP technique based on the BP algorithm. This study has found that generally air pollution concentrations using Tehran data set have been well predicted. The use of ANN models as the coefficient of multiple determinations (R^2) was found to exceed 0.83. According to R^2 value obtained from Figures 2 to 6, the best models were provided for $PM_{2.5}$, PM_{10}, O_3, NO_2 and CO, respectively. It is concluded that meteorological parameters are effective on the concentrations of all these pollutants. These results provide a system of air pollution forecasting in Tehran, which can be used by environmentalists to create an early alert of air quality for the public and take the necessary and suitable actions.

REFERENCES

Abdul-Wahab SA, Al-Alawi SM (2008) Prediction of Sulfur Dioxide (SO2) Concentration levels from the Mina Al-Fahal refinery in Oman using artificial neural networks. Am J Environ Sci 4: 473–481.

Agirre-Basurko E, Ibarra-Berastegi G, Madariaga I (2006) Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area. Environ Model Software 21: 430–446.

Baawain MS, El-Din MG and Smith DW (2007) Artificial neural networks modeling of ozone bubble columns: mass transfer coefficient, gas hold-up, and bubble size. Ozone Sci Eng 29: 343–352.

Baawain MS, Al-Serihi AS (2014) Systematic approach for the prediction of ground-level air pollution (around an industrial port) using an artificial neural network. Aerosol Air Qual Res 124–134.

Barai SV, Dikshit AK and Sharma S (2003) Neural network models for air quality prediction: a comparative study. Soft Comput Ind Appl 39:209–305.

Bickersta K, Walker G (2001) Public understandings of air pollution: The localization of environmental risk. Global Environ Change 11: 133–145.

Chay KY, Greenstone M (2005) Does air quality matter? evidence from the housing market. J Polit Econ 113 (227): 376–424.

Cheng S, Li L, Chen D, Li J (2012) A neural network based ensemble approach for improving the accuracy of meteorological fields used for regional air quality modeling. J Environ Manage 112: 404–414.

De Gennaro G, Trizio L, Di Gilio A, Pey J, Pérez N, Cusack M, Alastuey A, Querol X (2013) Neural network model for the prediction of pm10 daily concentrations in two sites in the western Mediterranean. Sci Total Environ 463–464: 875–883.

Gardner MW, Dorling SR (1998) Artificial neural networks (the multilayer perceptron). Atmos Environ 32: 2627–2636.

Gobakisa K, Kolokotsab D, Synnefac A, Saliari M, Giannopoulouc K and Santamourisc M (2011) Development of a model for urban heat island prediction using neural network techniques. J Sustainable Cities Soc. 1: 104–115.

Haykin S (1999) Neural Networks: A Comprehensive Foundation. MacMillan, New York.

Junninen H, Niska H, Tuppurainen K, Ruuskanen J, Kolehmainen M (2004) Methods for imputation of missing values in air quality data sets. Atmos Sci 38: 2895–2907.

Kurt A, Oktay AB (2010) Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. Expert Syst Appl 37: 7986–7992.

Lera-L_opez F, Faulin J, S_anchez M (2012) Determinants of the willingness-to-pay for reducing the environmental impacts of road transportation. Transp Res Part D. J Transp Environ 17: 215–220.

Şahin UA, Bayat C and Uçan ON (2011) Application of cellular neural network (CNN) to the prediction of missing air pollutant data. Atmos Res 101: 314–326.

Schindler M, Caruso G, Picard P (2017) Equilibrium and first-best city with endogenous exposure to local air pollution from traffic. Reg Sci Urban Econ 62: 12–23.

Seinfeld JH, Pandis SN (1998) Atmospheric chemistry and physics: from air pollution to climate change, John Wiley & Sons, Inc. New jersey.

Smith VK, Huang JC (1995) Can Markets Value Air Quality? A Meta- Analysis of Hedonic Property Value Models. J Polit Econ 103(1): 209–227.

Trigo R, Palutikof J (1999) Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach. Clim Res 13: 45–59.

USEPA (United States Environmental Protection Agency) (2003) Guidelines for developing an air quality (Ozone and PM2.5) forecasting program. EPA, North Carolina.

Wallace JM, Hobbs PV (1977) Atmospheric Science: An introductory survey. Academic Press, New York.

Wayne RP (1985) Chemistry of atmosphere: an introduction to the chemistry of the atmosphere of earth, the planets, and their satellites. Clarendon Press, Oxford.

Yu L, Wang S, Lai KK (2006) An Integrated Data Preparation Scheme for Neural Network Data Analysis. IEEE Trans Knowl Data Eng 18: 217–230.