Environmental Health Engineering and Management Journal 2015, 2(3), 117–122

http ehemj.com

Environmental Health Engineering and Management Journal

Original Article

Open Access Publish Free

Using ANN and EPR models to predict carbon monoxide concentrations in urban area of Tabriz

Mohammad Shakerkhatibi[1*], Nahideh Mohammadi[2], Khaled Zoroufchi Benis[3], Alireza Behrooz Sarand[4], Esmaeil Fatehifar[5], Ahmad Asl Hashemi[6]

[1]Assistant Professor, Department of Environmental Health Engineering, School of Health, Tabriz University of Medical Sciences, Tabriz, Iran

[2]MSc Student, Student Research Committee, Tabriz University of Medical Sciences, Tabriz, Iran

[3]Assistant Professor, Environmental Engineering Research Center, Faculty of Chemical Engineering, Sahand University of Technology, Tabriz, Iran

[4]Assistant Professor, Department of Chemical Engineering, Urmia University of Technology, Urmia, Iran

[5]Professor, Environmental Engineering Research Center, Faculty of Chemical Engineering, Sahand University of Technology, Tabriz, Iran

[6]Lecturer, Department of Environmental Health Engineering, School of Health, Tabriz University of Medical Sciences, Tabriz, Iran

Abstract

Background: Forecasting of air pollutants has become a popular topic of environmental research today. For this purpose, the artificial neural network (AAN) technique is widely used as a reliable method for forecasting air pollutants in urban areas. On the other hand, the evolutionary polynomial regression (EPR) model has recently been used as a forecasting tool in some environmental issues. In this research, we compared the ability of these models to forecast carbon monoxide (CO) concentrations in the urban area of Tabriz city.

Methods: The dataset of CO concentrations measured at the fixed stations operated by the East Azerbaijan Environmental Office along with meteorological data obtained from the East Azerbaijan Meteorological Bureau from March 2007 to March 2013, were used as input for the ANN and EPR models.

Results: Based on the results, the performance of ANN is more reliable in comparison with EPR. Using the ANN model, the correlation coefficient values at all monitoring stations were calculated above 0.85. Conversely, the R^2 values for these stations were obtained <0.41 using the EPR model.

Conclusion: The EPR model could not overcome the nonlinearities of input data. However, the ANN model displayed more accurate results compared to the EPR. Hence, the ANN models are robust tools for predicting air pollutant concentrations.

Keywords: Forecasting, ANN, EPR, Carbon monoxide, Modeling

Citation: Shakerkhatibi M, Mohammadi N, Zoroufchi Benis K, Behrooz Sarand A, Fatehifar E, Asl Hashemi A. Using ANN and EPR models to predict carbon monoxide concentrations in urban area of Tabriz. Environmental Health Engineering and Management Journal 2015; 2(3): 117–122.

Article History:
Received: 1 June 2015
Accepted: 12 August 2015
ePublished: 4 September 2015

***Correspondence to:**
Mohammad Shakerkhatibi
Email: shakerkhatibim@tbzmed.ac.ir

Introduction

Air pollution is still a serious environmental challenge in the world and is a notable environmental threat to human health (1). The problem is much more severe in urban areas of developing countries, where it affects quality of life and public health. It is usually caused by industrial activities, energy production by power plants, residential heating, fuel burning vehicles and natural disasters (1-4). Carbon monoxide (CO) is one of the most common air pollutants produced by incomplete combustion of hydrocarbons (1,5,6). Epidemiologic studies have reported short-term associations of ambient CO with daily mortality and morbidity from cardiovascular diseases (5).

For plans on how to scale down the levels of air pollutants and prevent their adverse effects, the uninterrupted assessment of air quality is required. For this purpose, measurement of air pollutants in certain points is possible by the use of fixed monitoring stations. So, in a homogeneous environment, measurements from these stations may indicate the levels of air pollutants at each station. However in most cases, dispersion of the actual pollutants remain unknown due to heavy influence by the prevailing dispersion conditions, distribution of emission sources and topography of the region (7).

Nowadays, modeling tools are widely used in many scientific fields, especially in environmental sciences such as air pollution (1). In such cases, various statistical and computational methods can be used to predict the concentrations of air pollutants using a series of reliable data captured in advance. In recent years, the artificial neural network (ANN) model is regarded as a cost effective method to obtain reliable prediction values of air pollutants (8). It is a computational model, which replicates the simple function of a biological network and is used to solve complex nonlinear functions. The neurons are located in the network layers of the model. The layers are defined as the input, the output and the hidden layers. There are many different types of neural networks. The most common structure of the neural network is the "feed forward" where the data flow from input to output units is strictly feed forward. The typical schematic of this structure is shown in Figure 1. ANNs are able to find and identify complex patterns in datasets which may not be well described by a simple mathematical formula or a set of known processes (9).

It has been used to predict hourly air pollutant concentrations in China (10) and has also been applied for forecasting daily PM_{10} and SO_2 concentrations in Taiyuan (11). In a study conducted to predict air pollution index (API), the ANN showed the smallest error in comparison to the autoregressive integrated moving average (ARIMA) and fuzzy time series (FTS) (12). In another study for forecasting PM_{10} concentrations in Turkey, the ANN presented more accurate results when compared with multi-linear regression (MLR) (13). Recently, similar studies have been conducted in Tehran (14), Düzce Province in Turkey (15), in Lisbon (16), in Algeciras (17) and in Spain (18), based on ANN models for forecasting air pollution.

On the other hand, the evolutionary polynomial regression (EPR) as a hybrid method in combination with the genetic algorithm (GA) and nonlinear regression has recently been used in some environmental studies for predicting the shear strength of municipal solid waste landfill (20), prediction of total sediment load of Malaysian rivers and some other environmental issues (21-24).

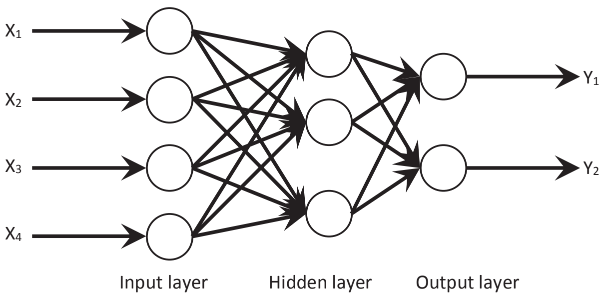

Figure 1. Typical schematic of back propagation feed forward neural network.

This research compared the ability of ANN and EPR models to forecast CO concentrations in the urban area of Tabriz. Considering the high fluctuations in the meteorological parameters of the study area (as an example: there is a temperature difference of about 50°C between the hot days of summer and the cold days of winter), an individual ANN and EPR model was constructed for each month. Having in mind that similar frameworks were used in the modeling procedures, the constructed model for March, 2013 has been explained in this study.

Methods

Study area

Tabriz, located in the northwest of Iran, the capital city of East Azerbaijan lies at 46.13 east and 38.8 north with an altitude of 1351 m above the sea level (Figure 2). According to the census conducted in 2011, it has a population exceeding 1.5 million. It has a semi-arid climate with annual precipitation around 380 mm and the mean, maximum and minimum temperatures are 13, 38 and -15°C, respectively. Table 1 presents the statistical analysis of meteorological variables.

ANN and EPR models

The hourly CO concentrations from March 2007 to March 2013 were obtained from five fixed monitoring stations designated as S1, S2, S3, S4, and S5 (Figure 3), operated by the East Azerbaijan Environmental Office and were used in the modeling procedures. Selection of model input variables with the most significant impact on model performance, is an important step in the ANN model development process. The next step is to build an input-output database required for ANN training. Since the main objective of constructing the ANN model is to forecast pollutant concentrations, the meteorological parameters which are predictable by conventional weather forecasting were used as model inputs (4). The selected meteorological variables included air temperature, relative humidity and wind speed and direction, were obtained from the East Azerbaijan Meteorological Bureau. In addition to

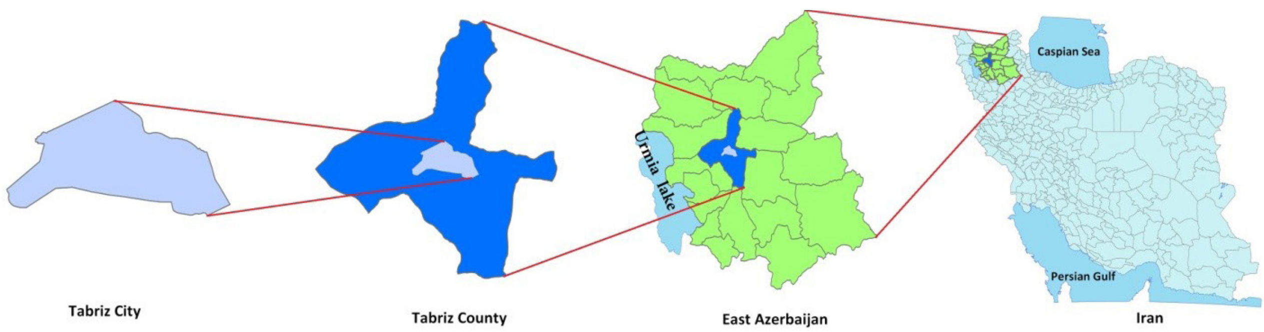

Figure 2. The location of Tabriz in the northwest of Iran.

Table 1. Statistics of meteorological variables

Parameter	Average	Median	SD	Min.	Max.
Temperature (°C)	4.7	4.4	5.7	-12.4	23
Wind speed (m/s)	3.5	3	2.5	0	14
Wind direct	156	120	121.7	0	360
Relative humidity (%)	58.4	58	19.84	18	99

Abbreviations: SD, standard deviation; Min, Minimum; Max, Maximum.

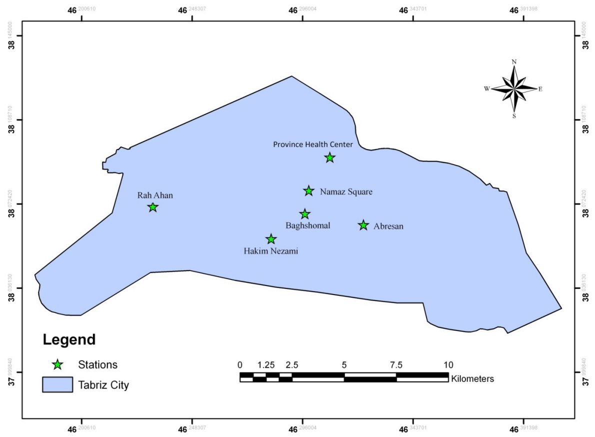

Figure 3. Locations of fixed air pollution monitoring stations in Tabriz.

meteorological variables, the hour of the day and day of the week were chosen as model inputs to cover sub-daily and sub-weekly variations in traffic pattern. For example, traffic patterns during weekends are significantly different from those on weekdays in the middle of the week, and weekdays right before and after the weekends.

A feed forward multi-layer perceptron network was used to forecast CO concentrations at the monitoring sites. The settings of ANN used in this study are presented in Table 2. The CO concentrations dataset were divided into 3 subsets, namely training (70%-80% of all), validation (10%-15% of all) and a test set (10%-15% of all) as an input for the ANN. The training set is a set of samples used to adjust or train weights in the neural network to produce the desired outcome. The validation set was used to further refine the neural network construction and find the best network configuration. The testing set was used to determine the performance of the fully trained neural network by computation of an error metric i.e., mean squared error (MSE) or maximum absolute error (MAE). The EPR model was carried out according to the procedure proposed by Giustolisi and Savic (25). To avoid network overflow, as a result of very large or very small values of input variables, all meteorological and pollutant values as model input had to be normalized (26), before entering the model. To this end, all pollutant and meteorological data were normalized ahead of prediction using Equation 1 in the range of -0.9-0.9. After prediction with Equation 2, the output data were transformed back to the real values.

$$X_{norm}=\frac{X-X_{min}}{X_{max}-X_{min}}\times(r_{max}-r_{min})+r_{min} \quad (1)$$

Table 2. The settings of the ANN

Train parameter	Value
Train function	Train lm
Divide function	Divide rand
Train ratio	0.7-0.8
Valid ratio	0.1-0.15
Test ratio	0.1-0.15
Performance function	MAE
Validation checks	300
Number of epochs	10 000
Number of layers	3-5
Number of neurons	8-25

Abbreviation: ANN, artificial neural network.

$$X=\left(\frac{(X_{norm}-r_{min})\times(X_{max}-X_{min})}{r_{max}}\right)+r_{min} \quad (2)$$

Where X_{norm} is the normalized value, X is the original value, X_{min} and X_{max} are the minimum and maximum values of X, and r_{min} and r_{max} are the values of -0.9 and 0.9, respectively.

In order to evaluate the accuracy of the constructed models to predict the CO levels, four statistical parameters were calculated. These parameters are: MSE, root mean squared error (RMSE), MAE and correlation coefficient (R^2). These parameters were calculated using Equations 3 to 6 as follows:

$$MSE=\frac{1}{N}\sum(Pi-Oi)^2 \quad (3)$$

$$RMSE=\left[\frac{1}{N}\sum_{i=1}^{N}(Pi-Oi)]^{1/2}\right] \quad (4)$$

$$MAE=\frac{1}{N}\sum||(Pi-Oi)| \quad (5)$$

$$R^2=\frac{\sum_{i-1}^{n}\left[P_i-\bar{O}_i\right]^2}{\sum_{i-1}^{n}\left[P_i-\bar{O}\right]^2} \quad (6)$$

Results

Table 3 presents the statistical analysis of measured CO concentrations, at 5 monitoring stations. The 8-hour average concentrations of 1.4, 2.6, 3.9, 2.8, and 3.7 ppm were measured for CO concentration at the monitoring stations of S1, S2, S3, S4 and S5, respectively.

The ANN performance and its comparison with the EPR results and real time measured values in stations S1-S5 are shown in Figures 4 to 8, respectively. The results revealed good agreement between measured and ANN predicted CO concentrations at all stations; however, the measured data were not associated with EPR outputs.

Totally, the performance of ANN and EPR models in predicting CO concentrations is shown in Figure 9. As illustrated, the levels of CO predicted using EPR, were less accurate (R^2 = 0.41) compared to the results with the ANN

Table 3. Statistical analysis of CO concentrations (ppm)

Station	Average	Median	SD	Min.	Max.
S1	1.4	1.1	0.6	0.1	14.3
S2	2.6	2.5	1.6	0.0	12.2
S3	3.9	2.9	1.9	0.0	24.1
S4	2.8	2.5	1.9	0.0	21.9
S5	3.7	3.5	2.2	0.0	19.3

Abbreviations: CO, carbon monoxide; SD, standard deviation; Min, Minimum; Max, Maximum.

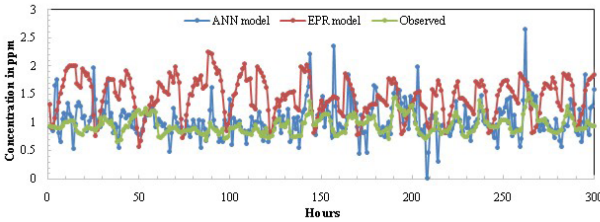

Figure 4. Hourly mean predicted and measured concentrations for CO at station S1.

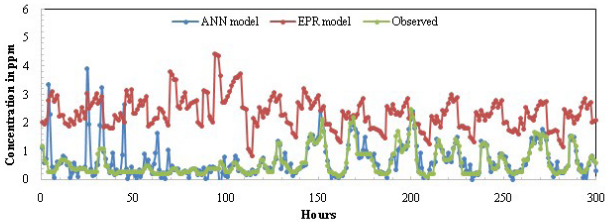

Figure 5. Hourly mean predicted and measured concentrations for CO at station S2.

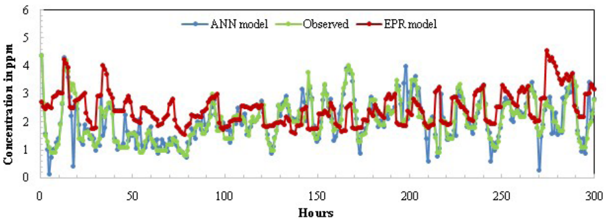

Figure 6. Hourly mean predicted and measured concentrations for CO concentration at station S3.

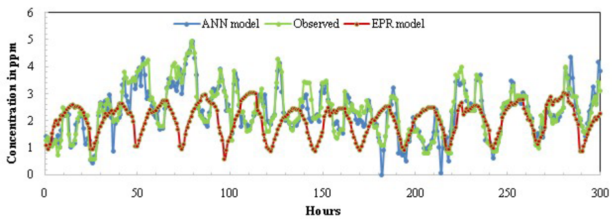

Figure 7. Hourly mean predicted and measured concentrations for CO concentration at station S4.

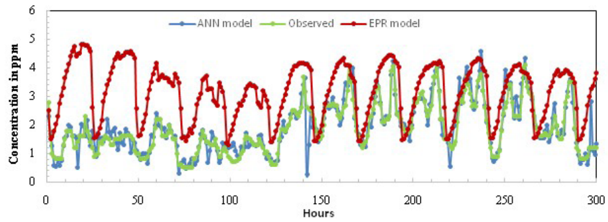

Figure 8. Hourly mean predicted and measured concentrations for CO concentration at station S5.

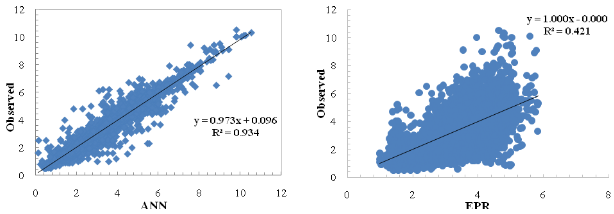

Figure 9. Predicted with ANN and EPR models vs. observed concentrations for CO concentration.

(R^2 = 0.93), which indicate that the EPR model could not overcome the nonlinearities of the variables such as meteorological parameters. The performance of constructed models were assessed by calculating MAE, MSE, RMSE and R^2 for all studied monitoring stations (Table 4).

Discussion

A comparison of CO concentrations with the National Ambient Air Quality Standards (35 ppm) revealed that the CO concentration was at an acceptable level at all monitoring stations, during the study period. As previously mentioned, a separate model was built for each month of year, resulting in 60 ANN/EPR models for 5 monitoring stations at 12 months of the year. The important benefit of constructing a monthly ANN/EPR model is that it reduces the modeling errors caused by the inherent variability of the meteorological parameters. The input parameters were selected based on several previous studies which have frequently used these parameters for training ANN to predict hourly air pollutant concentrations (11,13,15,18,27,28). Also, examination of various configurations of input variables and evaluation of model performance was carried out to select the most effective input variables. The results of investigations which are consistent with those of Cai et al (10) and Arhami et al (4) showed that adding or replacing other meteorological variables did not translate to better predictions. The sensitivity analyses were done by omitting input variables one at a time and constructing a new model using the remaining input variables. The results showed that the maximum performance of the model was reached when all six parameters were used as inputs and all parameters had significant influence on the performance of the ANN model. As can be seen in Table 4, at all stations the values of calculated MAE, MSE and RMSE were lesser for ANN predicted values. In addition, at all stations the values of correlation coeficients were much more for ANN than EPR. These results reveal the fact that the ANN model is a successful tool for forecasting of not only the CO concentration, but also the concentration of other air pollutants. For primary pollutants such as CO, which have inert behavior in the

Table 4. Results of ANN and EPR models for predicting CO concentration

Monitoring station	Model	MAE (ppm)	MSE (ppm)	RMSE (ppm)	R^2 (%)
S1	ANN	0.22	0.14	0.38	0.85
	EPR	0.69	0.82	0.90	0.17
S2	ANN	0.19	0.18	0.46	0.93
	EPR	1.28	2.19	1.48	0.17
S3	ANN	0.25	0.20	0.45	0.91
	EPR	1.09	1.18	1.09	0.26
S4	ANN	0.19	0.10	0.33	0.90
	EPR	0.74	0.94	1.04	0.28
S5	ANN	0.25	0.18	0.42	0.93
	EPR	0.97	1.60	1.26	0.41

Abbreviations: CO, carbon monoxide; ANN, artificial neural network; EPR, evolutionary polynomial regression; MSE, mean squared error; MAE, maximum absolute error; RMSE, root mean squared error.

atmosphere, the input varibles considered in this study are enough to reach an acceptable performance of the model. However, it is important to take into account the other input varibles i.e., solar radiation and photochemical data, when using ANN models to forecast secondary pollutants such as O_3 (4,19,29,30).

Conclusion

This research studied the potential of applying ANN and EPR models for ambient CO prediction. The ANN model can be used as a reliable model for the prediction of hourly CO concentrations in urban areas. However, in the case of the EPR model, the results were not satisfactory at all monitoring stations. The presented ANN model can produce reliable simulations of not only CO, but also the concentration of other air pollutant levels. Although the presented model is only valid for Tabriz monitoring stations, but the ANN-based approach can be applied to other urban areas as well.

Acknowledgments

The authors would like to give special thanks to the East Azerbaijan Environmental Office and Tabriz Meteorological Organization for providing air pollutants and meteorological data, used in this study.

Ethical issues

We certify that all data collected during the current study is presented in this manuscript; no data from the study has been or will be published separately.

Competing interests

The authors declare they have no competing interests.

Authors' contributions

All authors were involved in study design, data collection, and article approval.

References

1. Najafpoor AA, Hosseinzadeh A, Allahyari S, Javid AB, Esmaily H. Modeling of CO and NO_X produced by vehicles in Mashhad, 2012. Environ Health Eng Manag J 2014; 1(1): 45-9.
2. Kurt A, Oktay AB. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. Expert Syst Appl 2010; 37(12): 7986-92.
3. Lu WZ, Wang WJ. Potential assessment of the support vector machine method in forecasting ambient air pollutant trends. Chemosphere 2005; 59(5): 693-701.
4. Arhami M, Kamali N, Rajabi M. Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. Environ Sci Pollut Res 2013; 20(7): 4777-89.
5. Chen R, Pan G, Zhang Y, Xu Q, Zeng G, Xu X, et al. Ambient carbon monoxide and daily mortality in three Chinese cities: the China Air Pollution and Health Effects Study (CAPES). Sci Total Environ 2011; 409(23): 4923-28.
6. Levy RJ. Carbon monoxide pollution and neurodevelopment: a public health concern. Neurotoxicol Teratol 2015; 49: 31-40.
7. Zoroufchi Benis K, Fatehifar E. Optimal design of air quality monitoring network around an oil refinery plant: a holistic approach. International J Environ Sci Technol 2015; 12(4): 1331-42.
8. Azid A, Juahir H, Toriman M, Kamarudin M, Saudi A, Hasnam C, et al. Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: a case study in Malaysia. Water Air Soil Pollut 2014; 225(8): 1-14.
9. Zare Abyaneh H. Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. J Environ Health Sci Eng 2014; 12: 40.
10. Cai M, Yin Y, Xie M. Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. Transportation Research Part D: Transport and Environment 2009; 14(1): 32-41.
11. Wang P, Liu Y, Qin Z, Zhang GA. Novel hybrid forecasting model for PM_{10} and SO_2 daily concentrations. Sci Total Environ 2015; 505: 1202-12.
12. Rahman N, Lee M, Suhartono Latif M. Artificial neural networks and fuzzy time series forecasting: an application to air quality. Qual Quant 2014.
13. Ozdemir U, Taner S. Impacts of meteorological factors on PM_{10}: Artificial Neural Networks (ANN) and Multiple Linear Regression (MLR) approaches. Environmental Forensics 2014; 15(4): 329-36.
14. Nejadkoorki F, Baroutian S. Forecasting extreme

PM_{10} concentrations using artificial neural networks. Int J Environ Res 2012; 6(1): 277-84.

15. Taspinar F. Improving artificial neural network model predictions of daily average PM_{10} concentrations by applying PCA and implementing seasonal models. J Air Waste Manag Assoc 2015; 65(7): 800-9.
16. Russo A, Lind PG, Raischel F, Trigo R, Mendes M. Neural network forecast of daily pollution concentration using optimal meteorological data at synoptic and local scales. Atmos Pollut Res 2015; 6(3): 540-9.
17. Munoz E, Martin ML, Turias IJ, Jimenez-Come MJ, Trujillo FJ. Prediction of PM_{10} and SO_2 exceedances to control air pollution in the Bay of Algeciras, Spain. Stoch Environ Res Risk Assess 2014; 28(6): 1409-20.
18. Sharma N, Chaudhry K, Rao CC. Vehicular pollution modeling using artificial neural network technique: a review. J Sci Ind Res (India) 2005; 64(9): 637-47.
19. Viotti P, Liuti G, Di Genova P. Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia. Ecol Modell 2002; 148(1): 27-46.
20. Keramati M, Reshad SK, Asgarpour S, Tutunchian MA. Predicting shear strength of municipal waste material by evolutionary polynomial regression (EPR). Electronic Journal of Geotechnical Engineering 2014; 19: 53-62.
21. Abdul Ghani NA, Shahin MA, Nikraz HR. Use of evolutionary polynomial regression (EPR) for prediction of total sediment load of Malaysian rivers. International Journal of Engineering 2012; 6(5): 265-77.
22. Baawain MS, Al-Serihi AS. Systematic approach for the prediction of ground-level air pollution (around an industrial port) using an artificial neural network. Aerosol Air Qual Res 2014; 14(1): 124-34.
23. Laucelli D, Giustolisi O. Scour depth modelling by a multi-objective evolutionary paradigm. Environ Model Softw 2011; 26(4): 498-509.
24. Giustolisi O, Doglioni A, Savic DA, Webb BW. A multi-model approach to analysis of environmental phenomena. Environ Model Softw 2007; 22(5): 674-82.
25. Giustolisi O, Savic DA. A symbolic data-driven technique based on evolutionary polynomial regression. Journal of Hydroinformatics 2006; 8(3): 207-22.
26. Gardner MW, Dorling SR. Neural network modelling and prediction of hourly NOx and NO_2 concentrations in urban air in London. Atmos Environ 1999; 33(5): 709-19.
27. Zhang H, Liu Y, Shi R, Yao Q. Evaluation of PM_{10} forecasting based on the artificial neural network model and intake fraction in an urban area: a case study in Taiyuan city, China. J Air Waste Manag Assoc 2013; 63(7): 755-63.
28. Russo A, Raischel F, Lind PG. Air quality prediction using optimal neural networks with stochastic variables. Atmos Environ 2013; 79: 822-30.
29. Biancofiore F, Verdecchia M, Di Carlo P, Tomassetti B, Aruffo E, Busilacchio M, et al. Analysis of surface ozone using a recurrent neural network. Sci Total Environ 2015; 514: 379-87.
30. Luna AS, Paredes ML, De Oliveira GC, Correa SM. Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. Atmos Environ 2014; 98: 98-104.