العنوان:	Prediction of Hourly Indoor Carbon Monoxide Concentrations in Semiarid Regions using Regression and Feed Forward Back Propagation as a Hybrid Model
المصدر:	مجلة كلية فلسطين التقنية للأبحاث والدراسات
الناشر:	كلية فلسطين التقنية - دير البلح
المؤلف الرئيسي:	Elbayoumi, Maher
مؤلفين آخرين:	Harb, Suheir Elbayoumi(Co-Auth)
المجلد/العدد:	ع5
محكمة:	نعم
التاريخ الميلادي:	2018
الشهر:	سبتمبر
الصفحات:	717 - 739
رقم MD:	1109636
نوع المحتوى:	بحوث ومقالات
اللغة:	English
قواعد المعلومات:	HumanIndex, EcoLink, EduSearch
مواضيع:	الهندسة المعمارية، التصميم الداخلي، جودة الهواء، ثاني أكسيد الكربون، المدارس
رابط:	http://search.mandumah.com/Record/1109636

Prediction of hourly indoor Carbon Monoxide concentrations in semiarid regions using Regression and feedforward backpropagation as a hybrid model

Maher Elbayoumi

Energy and Sustainable Environment Center, School of Engineering, Israa University

Suheir Elbayoumi Harb

Palestine Technical College, Dier Elbalah

Date of submission:27/03/2018 Date of acceptance:30/04/2018

Abstract

Accurate site-specific forecasting of indoor hourly carbon monoxide (CO) concentrations in school microenvironments is a key issue in air quality research nowadays due to its impact on children's health. This paper investigated the improvement prediction of multiple linear regression (MLR) and feed forward back propagation (FFBP) by combining them with principal component analysis (PCA) for predicting indoor CO concentration in Gaza Strip, Palestine. Measurements were carried in 12 schools from October 2012 to May 2013 (one academic year). The results suggested that the selected models are effective forecasting tools and hence can be applicable for short-term forecasting of indoor CO level. The predicted indoor CO concentration values agree strongly well with the measured data with high coefficients of determination (R^2) 0.869, 0.870, 0.885 and 0.915 for MLR, PCA-MLR, FFBP and PCA-FFBP, respectively. Overall, results showed that PCA models combined with MLR and PCA with FFBP improved MLR and FFBP models of predicting indoor CO concentration, with reduced errors by as much as 7.14% (PCA-MLR) and 56.6% (PCA-FFBP). Moreover, PCA improved the accuracy of the FFBP model by as much as by 3.3%.

Keywords: *Natural Ventilation; Children; Indoor Air Quality; Feed forward back propagation; Principal component analysis.*

1. Introduction

Carbon Monoxide (CO) is colorless and odorless pollutant that arises from both natural and anthropogenic sources. CO is one of the most characteristic traffic pollutants in urban areas and produce as a primary pollutant during the incomplete combustion of fossil fuels and biomass in fumes produced by portable generators, stoves, and gas ranges (USEPA, 2013). CO exhibits toxicity characteristics due to its higher affinity with hemoglobin and as a result reduces the delivering oxygen to the body's tissues. Sustained exposure to CO has long been associated with effects on increasing adverse cardiovascular outcomes, asthma symptoms, hospital admission rates, and heart rate among children. (Slaughter *et al.*, 2003; Liao *et al.*, 2004; Cakmak *et al.*, 2006; ATSDR, 2012). There are also adverse impacts for exposure to low concentrations of CO for a long period among children on learning ability, manual dexterity, attention level, headaches, dizziness, nausea (feeling sick) and tiredness (HPA, 2009; USEPA, 2013)

(Raub and Benignus, 2002; Goniewicz *et al.*, 2009; HPA, 2009).

Several studies have investigated diurnal and seasonal CO concentration in different type's buildings. However; these studies have mainly focused on the monitoring, but not on the prediction of indoor air quality (IAQ) inside buildings (Chaloulakou *et al.*, 2003; Currie *et al.*, 2009; Elbayoumi *et al.*, 2014a). In addition to that, indoor CO concentrations heavily depend outdoor CO concentrations, indoor sources, infiltration, ventilation, and air mixing between and within rooms and on local conditions, such as weather changes,such as differences in temperature, humidity, pressure, atmospheric stability, and wind speed. Therefore; direct and long-term measurements of CO concentrations are not practical. In the absence of effective and efficient means to directly measure indoor CO from school buildings, development of mathematical prediction models might be a good alternative to provide reasonably accurate estimates.
Multiple linear regression (MLR) is one of the most popular methodology to express response of a dependent variable of several independent variables (predictor). Several studies used MLR to correlate the outdoor CO, particulate matter (PM), ozone, NO_2 and meteorological variables with indoor concentration of such pollutant (Chaloulakou *et al.*, 2001; Adar *et al.*, 2008; Braniš and Šafránek, 2011; Elbayoumi *et al.*, 2014b). In spite of its success, MLR presents problems in identifying the most important contributors when multicollinearity, or high correlation between the independent variables in regression equation are present (Abdul-Wahab *et al.*, 2005). The most common methods for removing such multicollinearity are principal component analysis (PCA) which has been proven to be effective tools to study the relationship between voluminous data such as air pollution and meteorological records (Yeniay and Goktas, 2002; Poupard *et al.*, 2005; Ul-Saufie *et al.*, 2010). PCA is used to reduce the number of predictive variables and transform into new variables that are mutually orthogonal, or uncorrelated, as well as to determine dominant multivariate relationships (Abdul-Wahab *et al.*, 2005). However, one of the main drawback is that MLR and principal component regression (PCR) cannot adequately model the non-linear relationships (Al-Alawi *et al.*, 2008). Thus, there are now several non-linear multivariate statistical methods that are able to approximate any non-linear relationship such as artificial neural networks (ANN).

Artificial neural networks (ANN) are one of the favoured techniques in predicting a complex system. Several studies demonstrated that the performance of ANNs were generally superior in comparison to traditional statistical methods and deterministic modeling systems because of their computational efficiency, generalization ability, and their limited need of prior knowledge about the modeling process structure (Nejadkoorki and Baroutian, 2011; Lal and Tripathy, 2012; Elangasinghe *et al.*, 2014). In past decade, there have been an increasing amount of applications of ANN models in the field of air pollution forecasting (Grivas and Chaloulakou, 2006; Sousa *et al.*, 2007). In fact, ANNs have been used to predict atmospheric concentrations of several outdoor pollutants such as NO (Gardner and Dorling, 1999), ozone (Wang *et al.*, 2003), benzene (Viotti *et al.*, 2002), SO_2 (Sofuoglu *et al.*, 2006), and particulate matter (Ul-Saufie *et al.*, 2013). Despite these great features and the successful performance of ANN models in the field of air pollution forecasting, ANNs performance suffer from inconsistencies drawback which due to the large number of factors including network structure, training methods, and sample data (Al-Alawi *et al.*, 2008). It is clear that no single method is best in every situation. Hence, by using hybrid models, PCR and combination of PCA-FFBP prediction techniques, the prediction accuracy is higher than individual forecasts.

In the literature, little attention has been paid to forecasting indoor air quality within buildings. Thus, the overarching goal of this project is to present the results of the application of multivariate regression analysis (MLR and PCA-MLR (PCR)) and feedforward backpropagation (FFBP) and PCA-FFBP in predicting indoor CO concentration as the function of meteorological parameters and other pollution concentration from natural ventilated school buildings.

2. Methodology

2.1 Study area

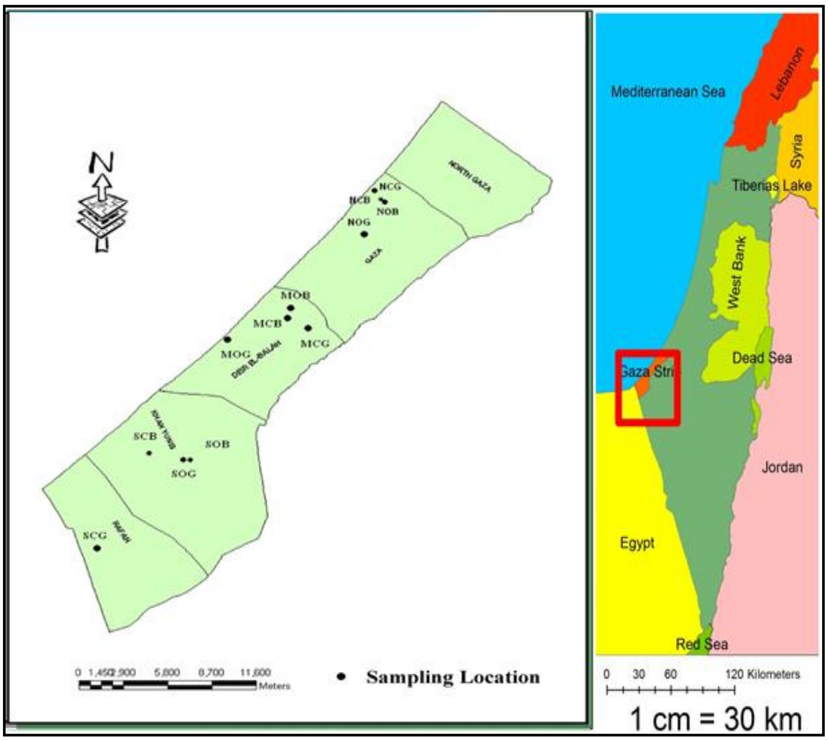
Gaza Strip (365 km^2) is located on the eastern coast of the Mediterranean sea between longitudes 34 ° 15' and 35 ° 40 ' east, and latitude 29 ° 30' and 23 ° 15' north. Climatically, the average daily temperature fluctuates from 24°C in summer season to 15 °C in winter. Meanwhile, the daily relative humidity varies between 62.5% in the daytime and 83.4% at night in the summer, and between 51.6% in the daytime and 81.3% at night in winter. The monthly average wind speed for Gaza is 3 m/s. (Koçak *et al.*, 2010; PMD, 2012). The major source of CO in Gaza Strip is the exhaust of about 60 901 motor vehicles, most of which are more than 15 years old and are out-dated (PCBS, 2012). Exhaust contains large quantities of CO, CO_2, $PM_{2.5}$, and hydrocarbons. In addition, during the frequent power outages, many people and institutions use portable electrical generators. Most of the generators involved were placed outside but were very close to the buildings to allow the generators to connect to the central electric panel (Elbayoumi *et al.*, 2014a). CO from these sources can build up in enclosed or partially enclosed spaces.

2.2. Description of Sampling Locations

The concentrations of pollutants were monitored at 12 schools located in north, middle and south of Gaza strip from October 2012 to May 2013 for one academic year. The sampling schools were purposely selected to reflect the diverse natures of human and vehicular activities. The description of sampling locations is summarized in Figure 1 and Table 1. In each selected school, three representative classrooms were selected for three sampling days. Sampling was conducted both inside and outside the selected classrooms during the studying activities.

Figure (1)

Map of Gaza strip and the monitoring schools



2.3 Pollutants measurements and instrumentation

The measurements were taken place in each site during schools hours for three consecutive days. The samplers were placed inside the classroom opposite the blackboard at least 1 m from the wall and at least 1.5 m height from the floor (Blondeau *et al.*, 2005; WHO, 2011). For outdoor sampling the samplers were placed at the front side of the building, usually near the playground area. A Kanomax IAQ Monitor was used for measuring CO and CO_2 concentrations. Meanwhile, the ventilation rate (VR) was calculated using the indoor concentration of carbon dioxide as a surrogate of the ventilation levels peroccupant (Kulshreshtha and Khare, 2011; WHO, 2011). The mass concentrations of particles ($PM_{2.5}$ and PM_{10}) have been monitored using handheld optical particle counter (HAL-HPC300). The monitor performs particulate size measurements by using laser light scattering. Air with multiple particle sizes passes through a flat laser beam produced by an ultra-low maintenance laser diode. A 3–channel pulse height analyzer for size classification detects the scattering signals.

Table (1)

Characteristic of monitoring schools

Code on the map	Number of students	Distance from main road (m)	Width of main road (m)	Description
NOB	623	43	10	Medium population density/ medium traffic volume
NCG	1183	50	10	High population density/ medium traffic volume
NCB	1066	30	20	Medium population density/high traffic volume
NOG	883	58	20	Medium population density/ medium traffic volume
MCB	733	43	10	High population density/ medium traffic volume
MCG	903	50	20	High population density/ medium traffic volume
MOG	1024	50	16	Low population density/very low traffic volume
MOB	712	65	10	High population density/ medium traffic volume
SCG	578	55	12	High population density/ medium traffic volume
SCB	729	50	10	High population density/ medium traffic volume
SOG	1132	40	20	Medium population density/ high traffic volume
SOB	1448	55	30	Medium population density/ high traffic volume

2.4 Meteorological Data

The surface wind speed (WS), ambient temperature (Temp), relative humidity(RH) and dew point temperature (TDP) in each site were simultaneously measured at the same time with pollutants measurement. A Kanomax IAQ Monitor was used for temperature, relative humidity measurements and Smart Sensor Electronic Anemometer was used for wind speed.

3. Statistical Analyses

3.1 Data Interpretation

A vital step in the development of a forecast indoor air model is the choice of input parameters (Jef *et al.*, 2005). Sensitivity analysis is a very useful method for ranking the importance of input variables by assessing their contribution (percentage) to the variability of the model output. In order to choose the most appropriate set of inputs parameters for FFBP, a number of statistical methods can be applied such as stepwise regression (SR), principal component analysis (PCA) and cluster analysis (Wilks, 2011). The importance of these methods is to reducing the number of input variables into the models, thus considerably diminishing redundant information, instabilities and over-fitting. Here, the selection of variables for each season model was made independently for each monitoring schools through a forward stepwise regression (FSR). During this procedure, which starts with the variable most correlated with the target, additional variables are added which, together with the previously selected variables, most accurately predict the target (Wilks, 2011). The procedure stops when any new variable does not significantly reduce the prediction error. Significance is measured by a partial F-test applied at 5% and by using the standardized regression determination coefficient(R^2) values (Wilks, 2011). All 15 potential predictors for indoor CO were first considered. The use of the FSR for each monitoring school in every season has reduced the complexity by retaining substantially less variables. The analysis of the data was carried out using the statistical software, SPSS (Statistical Package for Social Science, version 22) and MATLAB, version10. The data had been classified randomly into two sets using MATLAB software. Set 1 which consists of 70% of the original data, were used for model formulation and data set 2 (30%) were used for model validation.

3.2 Multiple Linear Regression (MLR)

Stepwise multiple regression was carried for CO and the result was checked for multicollinearity by examining the variance inflation factors (VIF) of the predictor variables. Durbin Watson statistic used to check if the model does not have any first order autocorrelation problem. MLR can be expressed according to the following Equation (1):

$$y = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki} + \varepsilon \qquad \text{.......................} \ (1)$$

where, b_k is the regression coefficients, x_k is the explanatory variables, i=1,2,……k and ε is stochastic error associated with the regression (Agirre-Basurko *et al.*, 2006) . The residuals (or error) were checked to evaluate if they were normally distributed with zero mean and constant variance to verify the adequacy of the statistical model (Al-Alawi *et al.*, 2008).

3.3 **Principal Component Analysis (PCA)**

PCA is a multivariate technique that is widely used in dealing with a large amount of data in monitoring studies, such as air pollution studies. In this study, this technique is applied for variables reduction and to provide the most relevant variables in CO variations (Dominick *et al.*, 2012). The PCs were extracted so that the first principal component (PC1) accounted for the largest amount of total variation in the data set, whereas the following components accounted for the remaining variations that were not considered in PC1 (Kovač-Andrić *et al.*, 2009). In general, the PCs are expressed in Equation 2 as follows:

$$PC_i = a_{1i}\, v_1 + a_{2i}\, v_2 + \cdots + a_{ni}\, v_n \qquad \ldots\ldots\ldots\ldots\ldots \quad (2)$$

where PC_i is the principal component i and a_{ni} is the loading (correlation coefficient) of the original variable V_1 (Özbay *et al.*, 2011).

Each PC represents a linear combination of data (variables) at specific coordinates at different values of chosen parameters (Elbayoumi *et al.*, 2014b). PCs are computed by calculating eigenvalues and eigenvectors. Eigenvalues will determine the eigenvectors and PCs; for each PC, only eigenvalues larger or equal to 1 are considered significant (Ul-Saufie *et al.*, 2013). Rotated PCs using varimax rotation to maximize the relationship between the PCs and original variables (Abdul-Wahab *et al.*, 2005). Dominick *et al.* (2012) reported that varimax rotation ensures that each variable maximally correlated with only one component and has minimal association with other components. The significant variables for each component are determined based on the loading factor where greater than 0.5 is considered strong, 0.4 is moderate, and 0.30 is weak.

3.4 **Feed-forward Backpropagation (FFBP)**

The multilayer perceptron (MLP) is the most common and successful neural network architecture with feed-forward backpropagation network topologies because the model displays an efficient learning environment, minimizing error between the target and obtained values, and has the common learning algorithm of a neural network (Ul-Saufie *et al.*, 2013).

As shown in Figure 2 the network of FFBP usually consists of an input layer, some hidden layers and an output layer and each layer is comprised of several operating neurons. Each neuron is connected to every neuron in nearby layers through adaptable synaptic weights which determines the strength of the relationship between two connected neurons. In each layer every neuron sums all the inputs that it receives from previous layer and formed the neuron output through predefined activation, or transfer function. Learning is defined as a network's ability to change weights by using backpropagation algorithm through two phases. In the forward phase, the training data set is propagated through the hidden layer and comes out of the neural network through the output layer. The output values are then compared to actual target output values. The error between the output layer and the actual values are calculated and propagated back towards the hidden layer (Ul-Saufie *et al.*, 2013). In the backward phase, derivatives of network error, with respect to the networks, are fed back to the network and used to adjust the weights so that errors were reduced with each iteration, resulting in improved FFBP models and the neural model gets closer and closer to producing the desired outputs.

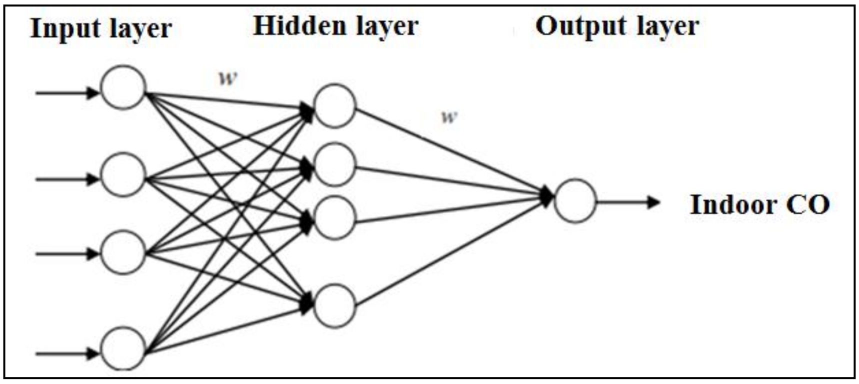
As net architecture authors used a 3-layer perceptron model. The first input layer contains the input variables which were selected by using the FSR method to reduce the complexity by retaining substantially less variables. In hidden layers architecture of the FFBP network the main problem is deciding i.e. the number of hidden layers and value of neurons in each hidden layer and activation function. Number of neurons in hidden layer has a strong influence on the output. The optimum number of neurons is very important because too few neurons will contribute to under-fitting, while too many neurons lead to over-fitting. Since a fixed scientific solution for the design of an optimal ANN model does not exist, the only method available is to try different numbers of neurons to observe how the results look. Using Equation (3), the initial number of neurons was 9 neurons and the number was increased until a relatively stable and optimal value was achieved (Yang *et al.*, 2005).

$$n_h = 2 \times n_i + 1 \qquad \text{……………..} (3)$$

Where n_i is the number of input neurons and n_h is the number of hidden neurons. The last layer is the output layer, which consists of the target of the forecasting model. Here, indoor CO was used as the output variable.

Figure (2)

Architecture of FFBP models for the prediction of indoor CO concentrations

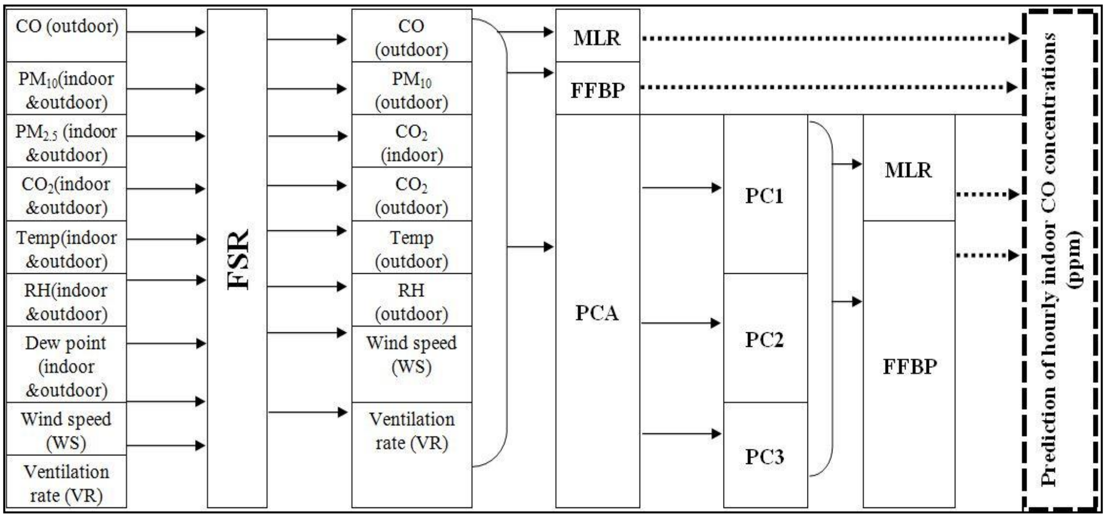


3.5 **Hybrid Models**

Hybrid models are models combine MLR and FFBP techniques with PCA. PCR is a combination of MLR and PCA. The use of PCs as input in MLR and FFBP is intended to reduce the complexity and multicollinearity problems of the models. The selected variables with high loading from PCA ensured that the majority of the original variances were included in the models, and they were ideal for use as independent variables in MLR and FFBP (Gervasi, 2008; Gvozdić *et al.*, 2011). Figure 3 shows the architecture of a PCR and PCA-FFBP model for prediction of indoor CO concentrations.

Figure (3)

Architecture of a PC-MLR and PC-FFBP model for the prediction of indoor CO concentrations.



3.6 **Performance Indicators**

The analysis of prediction performance typically involves calculation of errors between observed y and predicted y values. In this study five performance indicators were used which are normalized absolute error (NAE), root mean square error (RMSE), predication accuracy (PA), coefficient of determination (R^2) and index of agreement (IA). Normalized absolute error (NAE) and root mean square error (RMSE) were used to find the error of the model where value closer to 0 indicated a better model. Meanwhile, the other three performance indicators, i.e. index of agreement (IA), prediction accuracy (PA) and coefficient of determination (R^2) were used to check the accuracy of the model result, where a higher accuracy is given by value closer to 1. The equations used are given in Equations (4)- (8) (Karppinen *et al.*, 2000; Gervasi, 2008):

The Normalized Absolute Error (NAE).

$$NAE = \frac{\sum_{i=1}^{N}|P_i - O_i|}{\sum_{i=1}^{N} O_i} \qquad \ldots\ldots\ldots\ldots\ldots \quad (4)$$

The Root Mean Square Error (RMSE).

$$RMSE = \sqrt{\frac{1}{N-1}\sum_{i-1}^{N}(P_i - O_i)^2} \qquad \ldots\ldots\ldots\ldots..(5)$$

The Coefficient of Determination (R^2)

$$R^2 = \left(\frac{\sum_{i-1}^{N}(P_i-\bar{P})(O_i-\bar{O})}{N.S_{pred}.S_{obs}}\right)^2 \qquad \ldots\ldots\ldots\ldots..(6)$$

The Prediction Accuracy (PA)

$$PA = \frac{\sum_{i=1}^{N}(P_i-\bar{P})}{\sum_{i=1}^{N}(O_i-\bar{O})} \qquad \ldots\ldots\ldots\ldots..(7)$$

The Index of Agreement (IA)

$$IA = 1 - \left[\frac{\sum_{i=1}^{N}(P-O_i)^2}{\sum_{i=1}^{N}(|P_i-\bar{O}|+|O_i-\bar{O}|)^2}\right] \qquad \ldots\ldots\ldots\ldots..(8)$$

Where N is the number of sample, O_i is the indoor CO concentrations values and P_i is the predicted CO, P is the mean of predicted indoor CO. O is the mean of measured indoor CO, $\bar{P}$ is the average of predicted indoor CO, $\bar{O}$ is the average of measured CO. S_{pred} is a standard deviation of the predicted indoor CO. S_{obs} is a standard deviation of the observed indoor CO.

3.8 Quality Control

The instruments used to measure air pollutants and meteorological parameters were calibrated accordance with manufacturer instructions with a specific time schedule based on the type of measurements used. All gas analysers were auto-calibrated daily using a specific gas calibrator. Daily zero and span was performed for handheld optical particle counter (HAL-HPC300) by using Zero-Count Filter (Hal, 2012).

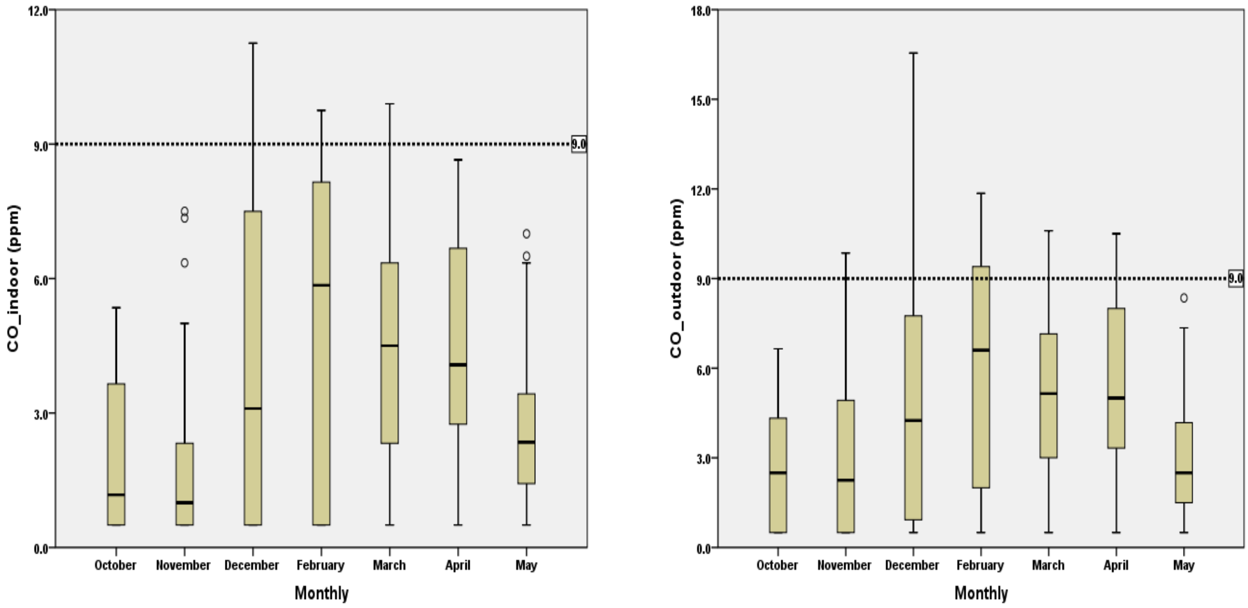
4 Result and Discussion

4.1 **Descriptive statistics**

Figure 4 shows the box plot and descriptive statistics of daily indoor and outdoor CO concentrations from 2012 to 2013. According to Ramli et al. (2010), the box plot is a simple graphical display that is ideal for making comparisons. In December 2012, February and April 2013, both indoor and outdoor CO concentrations were higher than the remaining months. The outdoor concentrations were 4.80 ppm, 6.07 ppm, and 5.35 ppm for December, February and April, respectively. The school locations displayed in Table 1 are very close to street intersections, and most of the schools located in overpopulated areas are characterized by congested traffic. Thus, frequent traffic jams resulting from poorly maintained roads, high traffic density, and very low wind speed are considered the main factors that contribute to high emission, accumulation, and low dilution of generated CO. Moreover, during December and February months where the weather is cold, winter time, the catalytic converters of vehicles take time to reach the operating temperature when the engine is cold, thereby resulting in increased CO production (Marković *et al.*, 2008). In addition, the CO production in overcrowded residential areas increases when cars move slowly near schools. Thus, the location of schools may be influences indoor CO concentrations. Most of time, maximum daily concentrations were below World Health Organization guidelines of 9 ppm except for some exceedences that were observed during November, December, February, March and April. Total number of exceedences that were recorded on outdoor concentrations is 82 exceedences and 22 exceedences on indoor concentrations.The indoor/outdoor (I/O) ratio were less than 1.00 during the monitoring period. Similar results were obtained by Chaloulakou et al. (2003), who revealed that air pollutants, such as CO, that are non-reactive and cannot be absorbed strongly on walls have an I/O ratio close to 1.0 in the absence of indoor sources. Thus, the building envelope provides little protection from outdoor CO pollution, and peaks in indoor concentrations reached the extremes of outdoor concentrations regardless of the airtightness in these buildings.

Figure 4

Box plots and descriptive statistics for monthly indoor and outdoor CO concentrations



Monthly	Pollutant	Minimum	Maximum	Mean	Std. Deviation	I/O ratio
October	CO (indoor)	0.50	5.35	1.95	1.68	0.75
	CO(outdoor)	0.50	6.65	2.61	2.06	
November	CO (indoor)	0.50	7.50	1.81	1.78	0.76
	CO(outdoor)	0.50	9.85	2.69	2.30	
December	CO (indoor)	0.50	11.25	4.00	3.43	0.83
	CO(outdoor)	0.50	16.55	4.80	3.67	
February	CO (indoor)	0.50	9.75	5.08	3.34	0.84
	CO(outdoor)	0.50	11.85	6.07	3.74	
March	CO (indoor)	0.50	9.90	4.27	2.72	0.84
	CO(outdoor)	0.50	10.60	5.08	3.00	
April	CO (indoor)	0.50	8.65	4.45	2.41	0.83
	CO(outdoor)	0.50	10.50	5.35	2.87	
May	CO (indoor)	0.50	7.00	2.62	1.84	0.87
	CO(outdoor)	0.50	8.35	3.02	2.17	

4.2 Bivariate correlation analysis

Several studies confirmed that the IAQ is dependent on outdoor concentrations and local conditions, such as weather changes and seasonal variations (Roberts, 2004; Kam et al., 2011). Therefore, bivariate correlation was used to identify the factors that may influence indoor CO concentrations as presented in Table 2. A moderate relationship exists between indoor and outdoor CO concentrations ($r = 0.47$). The value of the correlation coefficient (r) between the indoor and outdoor data can be used as indicator of the degree to which CO measured indoors is attributed to the infiltration from outdoors. Morawska et al. (2001) and Chaloulakou et al. (2003) showed that the indoor peak concentrations of CO are slightly dampened and lag behind outdoor peaks, thus suggesting that indoor CO concentrations are not immediately affected by outdoor concentration changes due to changes in air exchange air. Moreover, a positive correlation exists between indoor and outdoor CO concentration and PM_{10}, $PM_{2.5}$ and CO_2 due to the same emission source. The PM - CO correlations observed in this study are consistent with the finding Dominick et al., (2012) study). The indoor CO concentration was found to be negatively correlated with indoor temperature and relative humidity. In naturally ventilated buildings, high building ventilation rate promptly brings indoor humidity to the same level encountered outside. Thus, a negative correlation between humidity and CO infiltration and/or build-up inside the building is expected. Furthermore, a negative correlation exists between indoor and outdoor CO concentration and both ventilation rate and wind speed because low wind speeds favour the accumulation of pollutants (low wind speeds are also related to stable atmospheric conditions).

Table (2)

Correlation coefficients between indoor and outdoor CO **and meteorological Parameters**

	CO (indoor)	CO (outdoor)
CO (indoor)	1	0.47*
CO (outdoor)	0.47*	1
$PM_{2.5}$(indoor)	0.38*	0.23*
$PM_{2.5}$ (outdoor)	0.34*	0.19*
PM_{10} (indoor)	0.15*	0.13*
PM_{10} (outdoor)	0.29*	0.37*
CO_2 (indoor)	0.26*	0.31*
CO_2 (outdoor)	0.42*	0.40*
Temp (indoor)	-0.41*	-0.39*
Temp (outdoor)	-.42*	-.34*
RH (indoor)	-0.08*	-.02
RH (outdoor)	-0.06	-0.01
VR	-0.21*	-0.16*
WS	-0.10*	-0.09*

**Correlation is significant at the 0.01 level (2-tailed).*

4.3 Principal component analysis

Sensitivity analysis, using FSR technique, was undertaken so as to examine the level of importance for 15 variables i.e. outdoor CO, indoor and outdoor PM_{10}, indoor and outdoor $PM_{2.5}$, indoor and outdoor CO_2, temperature, relative humidity, dew point temperature, ventilation rate ,and wind speed. The standardised regression determination coefficient (R^2) values were used to estimate the relationship between indoor CO and these variables. The results show that 8 variables were identified before the extraction using PCA as shown in Table 3. After the extraction was applied, three factors were considered as the principal component based on eigenvalues of more than 1. PCA procedures was followed by a varimax rotation to maximize the loading of a predictor variable and the higher loading variable with absolute values greater than 40% were selected for the principal component interpretation **(**Abdul-Wahab *et al.*, 2005**)**.The factor loading values after rotation are very important in interpretation of PCA results. The factor loadings correlate the factors and the variables, the higher the factor loading, the more variable contributes to the variation of the PC (Ul-Saufie *et al.*, 2013). The eigenvalues for all linear components after rotation are shown in Table 3. The cumulative variance of the principal components is 74.06%. The first PC explains 39. 98% of the total variation in the data set, which indicates a heavy load on relative humidity, temperature dew point, ventilation rate, wind speed and outdoor CO. The second PC, which accounts for approximately 20.96% of the total variation, loads heavily on indoor and outdoor PM_{10} and $PM_{2.5}$. Among the principal components, the third account for approximately 13.12% and load heavily on outdoor and indoor temperature.

Table (3)

Rotated principal component loadings matrix

	Component		
	1	2	3
CO(outdoor)	-0.60		
RH (indoor)	-0.86		
RH(outdoor)	-0.80		
VR	0.83		
WS	0.82		
CO_2(indoor)	0.62		
TDP(indoor)	0.90		
TDP(outdoor)	0.90		
$PM_{2.5}$(indoor)		0.77	
PM_{10}(indoor)		0.77	
PM_{10}(outdoor)		0.75	
$PM_{2.5}$(outdoor)		0.73	
Temp(indoor)			0.93
Temp(outdoor)			0.93
Eigenvalue	6.00	3.14	1.97
% of Variance	39.98	20.96	13.12
Cumulative %	39.98	60.94	74.06

4.4 **Multiple linear regression (MLR)**

The stepwise MLR models of indoor CO prediction using the original parameters and PCs as the inputs were conducted with regression assumptions approximately satisfied (Table 4). The four goodness of fit measures showed that the residual distributions were approximately normal, with zero means and no detectable serial and the correlation coefficients of the regressions were all highly statistically significant ($P< 0.01$) (Abdul-Wahab *et al.*, 2005). The result showed that the developed models did not encounter multicollinearity problems as the VIF was less than 3.0. In addition, the tolerance values for the variables in MLR model are higher than 0.3. In accordance with the findings of Field et al. (2009), the tolerance value must be smaller than 0.1 to indicate a multicollinearity problem. However, DW may indicate slightly positive autocorrelation problems in the model because these values ranged from 1.966 to 1.961. The developed MLR and PCR models were also assessed using the coefficient of determination (R^2), which was used as an indicator of the ability of the selected variables to explain the variations in indoor CO concentration (Abdul-Wahab et al., 2005). As presented in Table 4 when the four best variables are fitted to the CO data, the value of the R^2 is approximately (0.72). Thus, approximately 72% of the variation in the indoor CO concentrations can be explained by the four variables, as listed in the table. Meanwhile, the usage of PCs as the inputs in MLR could improve the efficiency of the model to explain the variations in CO concentrations. During these time periods, R^2 values for PCR was approximately (0.73), as reported in Table 4. Therefore, approximately 73% of the variation in the indoor CO concentrations can be explained by the three independent components.

In this study, the applicability of the developed models for predicting the indoor CO concentration variations was assessed using hourly average monitoring records from validation data set. The performance levels of the validation MLR and PCR models indicated that relatively strong relationships were obtained between the observed and predicted values and R^2 for these models was almost 0.87. Thus, R^2 can explain approximately 87% of the variation in the indoor CO by using both models. In addition to that, the values of RMSE ranged from 0.027 % for PCR model to 0.029% for MLR model.

By comparing the performance of the two models, even the R^2 values for the both models were comparable PCR model produced the lowest RMSE comparing with MLR.

Table (4)

Summary models for indoor CO concentration predictions based on original parameters and PCA as inputs.

Method	Models	R^2	Range of VIF	Durbin-Watson
MLR	$CO=0.07+1.07CO(out)+0.07CO_2(out)+0.06PM_{10}(out)-0.10Temp(out)$	0.72	1.07-2.32	1.96
PCR	$CO = 0.08+0.88PC1 +0.19PC2-0.03PC3$	0.73	1.00	1.96

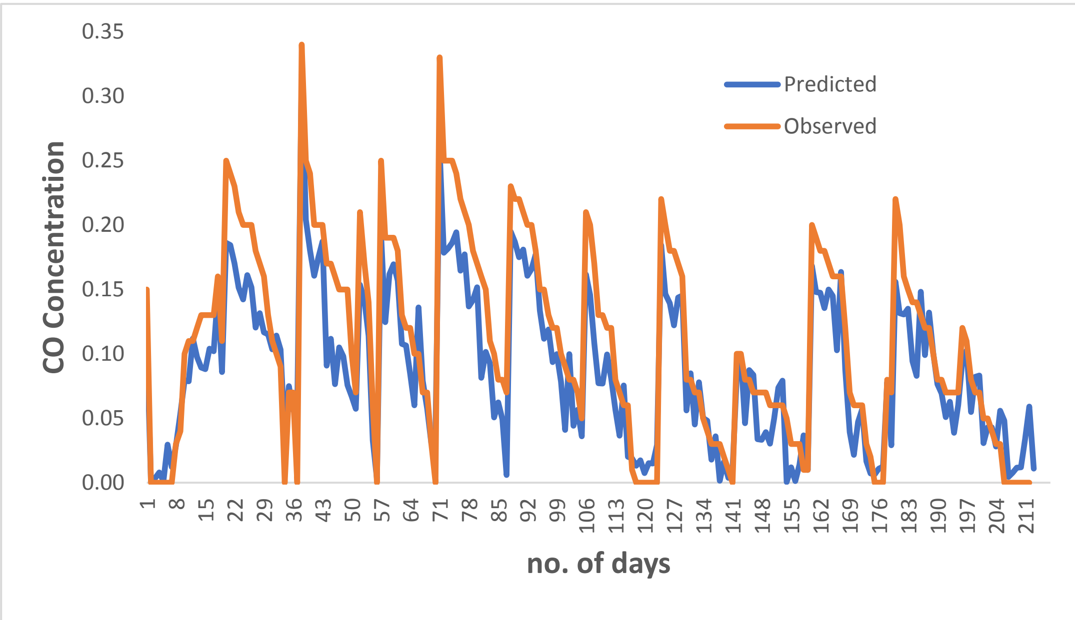
4.5 **Feedforward backpropagation (FFBP)**

The optimum trained FFBP and PCA- FFBP structures were selected according to the minimum RMSE, NAE, and maximum PA, R^2 and IA of the test sets. The best number of neurons in the hidden layer was 9 for actual parameters as inputs and 4 for PCA as inputs. The optimal activation functions are obtained using sigmoid transfer function. The best transfer function is purelin and purelin for original input and PCA inputs. The coefficient of determination, R^2, results show that independent variables are fitted to indoor CO data and can explain approximately between 88%- 91% of the variation in the indoor CO by using FFBP and PCA-FFBP, respectively. In addition to that, the values of RMSE ranged from 0.019 % for PCA-FFBP model to 0.034% for FFBP model. By comparing the performance of the two models, PCA- FFBP model produced the lowest RMSE value and highest R^2 values. Figure 5 shows the daily cycles of measured indoor concentrations, together with the calculated indoor concentrations by using FFBP model. The indoor concentrations predicted by the model follow the changes of the measured indoor concentrations, but the response of the model is delayed and cannot follow exactly the sharp outdoor concentration changes. For example, the model did not detect several days of abnormal air quality such as days no.21, 37, 53, 57, 70, 89, 109, and 125.

a seasonal variation in Gaza, such as the temperature, affects other variables, such as the fluctuations of relative humidity, the changes of the number of vehicles due to the number of the passengers, the seasonal fluctuations of the outdoor air quality due to the thermodynamic changes of the chemical reactions in an engine of the vehicles, such as the partial burning of nitrogen and carbon components due to the incomplete combustion of the gasoline, and so on. This results in the seasonal variations of air pollutants and their sources due to the fluctuations of the vehicles' emission components.

Figure 5

Measured indoor and calculated model indoor hourly CO concentrations



4.6 **Comparisons**

Five performance indicators were used to evaluate and compare between the four models (MLR, PCR, FFBP and PCA-FFBP) used to predict indoor CO as shown in Table 5. For a good model NAE and RMSE value should approach zero, while PA, R^2 and IA should be closer to one. The results suggest that the four models are effective forecasting tools and hence can be applicable for short-term forecasting of indoor CO level. The performance of FFBP outperformed MLR prediction accuracy (R^2) as calculated by the percentage differences as much as 1.8%. However, the prediction error (RMSE) for FFBP was larger than MLR as much as 15.8%.

PCA didn't improve the accuracy measures of MLR model comparing with PCR model where the results of PA, R^2 and IA were almost comparable with MLR but PCA reduced the prediction error in PCR comparing with MLR as calculated by the percentage differences as much as 13.0% and 7.14% for NAE and RMSE, respectively. Furthermore, PCA improves the accuracy of FFBP model where PCA-FFBP model results showed the highest accuracy measures i.e. 0.978 (IA), 0.961 (PA), and 0.915 (R^2) and the lowest error indicators i.e. 0.144 (NAE) and 0.019 (RMSE) comparing with the other three models. In addition, the performance of PCA-FFBP outperformed FFBP prediction accuracy (R^2) as calculated by the percentage differences as much as 3.3% and with errors (RMSE) reduced as much as 56.6% for prediction of indoor CO. Thus, the results obtained from the PCA-FFBP models were encouraging compared to the constructed multiple linear regression model.

Table (5)

Performance indicator for ANN models vs.

MLR models for indoor CO

Model	NAE	RMSE	R^2	PA	IA
MLR	0.213	0.029	0.869	0.937	0.965
PCR	0.187	0.027	0.870	0.937	0.961
FFBP	0.260	0.034	0.885	0.945	0.936
PCA- FFBP	0.144	0.019	0.915	0.961	0.978

5 Conclusion

Previously, MLR, PCR, and ANN (FFBP) and PCA-FFBP neural network methods were used effectively to study air pollution and meteorological records. In this study the capability of these techniques to predict indoor CO concentrations in natural ventilated schools located in Gaza Strip, Palestine was employed. To raise the efficiency of FFBP neural network prediction, FSR method was used to select the key input variables for the optimal structure of the models. Overall, it was found that the four models i.e. MLR, PCR, FFBP and PCA-FFBP are effective forecasting tools and hence can be applicable for short-term forecasting of indoor CO level. However, the model's performance results varied that the PCA-FFBP model gave the highest quality of prediction with lowest error 0.144 (NAE) and 0.019 (RMSE) and with greatest accuracy 0.978 (IA), 0.961 (PA), and 0.915 (R^2) compared with the remaining models. It can be recommended that PCA-FFBP neural network could be a new promising methodology instead of a MLR and FFBP to predict IAQ in naturally ventilated buildings. It is evident that further study should be implemented in other schools, considering different localities and different climate for better protection of human health and life quality.

3 Acknowledgement

Special thanks are expressed to the schools involved in this study. The study could not have been completed without the cooperation of the students and the support of the schools principals and the teachers who were responsible for the day to day operation of the indoor and outdoor sampling.

References

1. Abdul-Wahab, S.A., Bakheit, C.S. and Al-Alawi S.M. "Principal Component and Multiple Regression Analysis in Modelling of Ground-Level Ozone and Factors Affecting Its Concentrations." Environmental Modelling and Software,vol. 20, no. 10, pp. 1263-71,2005.
2. Adar, S.D, Mark D., Sullivan J. R., Compher, M., Szpiro, A., and Liu, L. "Predicting Airborne Particle Levels Aboard Washington State School Buses." Atmospheric Environment,vol. 42, no. 33, pp. 7590-99,2008.
3. Agirre-Basurko, E, Ibarra-Berastegi, G., and Madariaga I. "Regression and Multilayer Perceptron-Based Models to Forecast Hourly O3 and No2 Levels in the Bilbao Area." Environmental Modelling and Software,vol. 21, no. 4, pp. 430-46, 2006.
4. Al-Alawi, S. M., Abdul-Wahab, S. A. and Bakheit C. S. "Combining Principal Component Regression and Artificial Neural Networks for More Accurate Predictions of Ground-Level Ozone." Environmental Modelling & Software, vol. 23, no. 4, pp. 396-403, 2008..
5. ATSDR. Toxicological Profile for CO. 1st ed. Atlanta: Agency for Toxic Substances and Disease Registry: U.S. Department of Health and Human Services, 2012.
6. Blondeau, P., Iordache, V., Poupard, O., Genin, D., and Allard, F. "Relationship between Outdoor and Indoor Air Quality in Eight French Schools." Indoor Air, vol. 15, no. 1, pp. 2-12 , 2005.
7. Braniš, M., and Šafránek. J. "Characterization of Coarse Particulate Matter in

School Gyms." Environmental Research, vol. 111, no. 4, pp. 485-91, 2011.
Cakmak, S., Dales, R. E., and Judek S. "Do Gender, Education, and Income .8
Modify the Effect of Air Pollution Gases on Cardiac Disease?". Journal of
Occupational and Environmental Medicine,vol. 48, no. 1,pp. 89-94, 2006.
Chaloulakou, A., Mavroidis, I., and Duci, A."Indoor and Outdoor Carbon .9
Monoxide Concentration Relationships at Different Microenvironments in
the Athens Area." Chemosphere,vol. 52, no. 6, pp. 1007-19, 2003.
Chaloulakou, A., Mavroidis, I., Spyrellis, N. "A Comparison of Indoor and .10
Outdoor Carbon Monoxide Concentration Levels at a Public School in
Athens. Application of a Statistical Prediction Tool." In 7th International
Conference on Environmental Science and Technology. Ermoupolis, Syros .11
island, Greece, 2001.
Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., and Rivkin S. G. .12
"Does Pollution Increase School Absences?". The Review of Economics and
Statistics.vol. 91, no. 4, pp. 682-94, 2009.
Dominick, D., Juahir, H., Latif, M.T., Zain, S.M., and Aris, A.Z. "Spatial .13
Assessment of Air Quality Patterns in Malaysia Using Multivariate
Analysis." Atmospheric Environment,vol. 60, no. 1, pp. 172-81, 2012.
Elangasinghe, M.A., Dirks, K.N., Singhal, N., Costello, S.B., Longley, I., .14
and Salmond, J.A. "A Simple Semi-Empirical Technique for Apportioning
the Impact of Roadways on Air Quality in an Urban Neighbourhood."
Atmospheric Environment,vol. 83, pp. 99-108, 2014.
Elbayoumi, M, Ramli, NA., Yusof, NF., and Al Madhoun, W. "The Effect of .15
Seasonal Variation on Indoor and Outdoor Carbon Monoxide Concentrations
in Eastern Mediterranean Climate ". Atmospheric Pollution Research , vol. 5,
no. 2, pp. 315-24 , 2014.
Elbayoumi, M, Ramli, NA., Yusof, NF., Yahaya, AS., Al Madhoun, W. and .16
Ul-Saufie, AZ. "Multivariate Methods for Indoor Pm10 and Pm2.5
Modelling in Naturally Ventilated Schools Buildings." Atmospheric
Environment,vol. 94, pp. 11-21, 2014.
Gardner, MW., and Dorling, SR. "Neural Network Modelling and Prediction .17
of Hourly No and No2 Concentrations in Urban Air in London."
Atmospheric Environment,vol. 33, no. 5, pp. 709-19, 1999.
Gervasi, O. Computational Sicience and Its Applications-Iccsa, edited by O. .18
Gervasi. Vol. 1st, Italy: Springer, 2008.
Goniewicz, M., Czogała, J., Kośmider, L., Koszowski, B., Zielińska-Danch, .19
W., and Sobczak, A. "Exposure to Carbon Monoxide from Second-Hand
Tobacco Smoke in Polish Pubs." Central European Journal of Public
Health,vol. 17, no. 4, pp. 220-9, 2009.
Grivas, G., and Chaloulakou, A. "Artificial Neural Network Models for .20
Prediction of Pm10 Hourly Concentrations, in the Greater Area of Athens,
Greece." Atmospheric Environment, vol. 40, no. 7 ,pp. 1216-29, 2006.

Gvozdić, V., Kovač-Andrić, E., and Brana, J. "Influence of Meteorological Factors NO2, SO2, CO and PM10 on the Concentration of O3 in the Urban Atmosphere of Eastern Croatia." Environmental Modeling & Assessment, vol. 16, no. 5, pp. 491-501, 2011. .21

Hal. "Hal-Hpc300 User Manual." http://www.haltechnologies.com/Docs/HALHPC300%20User%20Manual.pdf. 2013 .22

HPA. "Carbon Monoxide:General Information." Public Health England www.hpa.org.uk/webc/hpawebfile/hpaweb_c/1202487036268. 2013 .23

Jef, H., Clemensa, M., Gerwinb, D., Fransb, F., and Olivierc, B. "A Neural Network Forecast for Daily Average PM10 Concentrations in Belgium." Atmospheric Environment,vol. 39, no. 18, pp. 3279-89, 2005. .24

Kam, W, Cheung, K., Daher, N., Sioutas, C. "Particulate matter (PM) concentrations in underground and ground-level rail systems of the Los Angeles metro." Atmospheric Environment, vol.45, pp. 1506-1516, 2011 .25

Karppinen, A, Kukkonen, J., Elolähde, T., Konttinen, M., Koskentalo, T., and Rantakrans, E. "A Modelling System for Predicting Urban Air Pollution: Model Description and Applications in the Helsinki Metropolitan Area." Atmospheric Environment, vol. 34, no. 22, pp. 3723-33, 2000. .26

Koçak, M., Kubilay, N., Tugrul, S., and Mihalopoulos, N. "Atmospheric Nutrient Inputs to the Northern Levantine Basin from a Long-Term Observation: Sources and Comparison with Riverine Inputs." Biogeosciences,vol. 7, no. 12, pp. 4037-50, 2010. .27

Kovač-Andrić, E., Brana, J., and Gvozdić, V. "Impact of Meteorological Factors on Ozone Concentrations Modelled by Time Series Analysis and Multivariate Statistical Methods." Ecological Informatics,vol. 4, no. 2, pp. 117-22, 2009. .28

Kulshreshtha, P., and Khare, M."Indoor Exploratory Analysis of Gaseous Pollutants and Respirable Particulate Matter at Residential Homes of Delhi, India." Atmospheric Pollution Research, vol. 2, no. 3, pp. 337-50, 2011. .29

Lal, B., and Tripathy, S. "Prediction of Dust Concentration in Open Cast Coal Mine Using Artificial Neural Network." Atmospheric Pollution Research,vol. 3, no. 2, pp. 211-218, 2012. .30

Liao, D., Yinkang D., Whitsel, E., Zheng, Z., Heiss, G., Chinchilli, V., and Lin, H. "Association of Higher Levels of Ambient Criteria Pollutants with Impaired Cardiac Autonomic Control: A Population-Based Study." American Journal of Epidemiology,vol. 159, no. 8, pp. 768-77, 2004. .31

Marković, D., Marković, D., Jovanović, A., Lazić, L., and Mijić, Z. "Determination of O3, NO2, SO2, CO and PM10 Measured in Belgrade Urban Area." Environmental Monitoring and Assessment, vol. 145, no. 1-3, pp. 349-59, 2008. .32

33. Nejadkoorki, F., and Baroutian, S. "Forecasting Extreme PM1Concentrations Using Artificial Neural Networks." International Journal of Environmental Research,vol. 6, no. 1, pp. 277-84, 2011.
34. Özbay, B., Keskin, G.A., Doğruparmak, Ş.Ç., and Ayberk, S. "Multivariate Methods for Ground-Level Ozone Modeling." Atmospheric Research,vol. 102, no. 1, pp. 57-65, 2011.
35. PCBS. "Statistical Yearbook of Palestine Report." Ramalla- Palestine: Palestinian Central Bureau of Statistics (PCBS), 2012.
36. PMD. "Climate Bulletine." The Palestinian Meteorological Department (PMD), http://www.pmd.ps/ar/mna5palestine.php.2013
37. Poupard, O., Blondeau, P., Iordache, V., and Allard, F. "Statistical Analysis of Parameters Influencing the Relationship between Outdoor and Indoor Air Quality in Schools." Atmospheric Environment,vol. 39, no. 11, pp. 2071-80, 2005.
38. Ramli, N.A., Ghazali, N.A., Yahaya, A.S. "Diurnal fluctuations of ozone concentrations and its precursors and prediction of ozone using multiple linearregressions". Malaysian Journal of Environmental Management,vol. 11, no. 2, pp.57-69, 2010.
39. Raub, J., and Benignus, V. "Carbon Monoxide and the Nervous System." Neuroscience & Biobehavioral Reviews,vol. 26, no. 8, pp. 925-40, 2002.
40. Roberts, S. " Interactions between particulate air pollution and temperature in air pollution mortality time series studies."Environmental Research,vol. 96, no. 3, pp. 328-337, 2004.
41. Slaughter, J.C., Lumley T., Sheppard, L., Koenig, J. Q., and Shapiro, G. "Effects of Ambient Air Pollution on Symptom Severity and Medication Use in Children with Asthma." Annals of Allergy, Asthma and Immunology, vol. 91, no. 4, pp.346-53,2003.
42. Sofuoglu, S., Sofuoglu, A., Birgili, S., and Tayfur G. "Forecasting Ambient Air So2 Concentrations Using Artificial Neural Networks." Energy Sources, Part B,vol. 1, no. 2, pp. 127-36, 2006.
43. Sousa, S., Martins, F., Alvim-Ferraz, M., and Pereira M. "Multiple Linear Regression and Artificial Neural Networks Based on Principal Components to Predict Ozone Concentrations." Environmental Modelling & Software, vol. 22, no. 1, pp.97-103, 2007.
44. Ul-Saufie, A., Yahaya, A., Ramli, N., Awang, N., and Hamid H. "Future Daily Pm_{10} Concentrations Prediction by Combining Regression Models and Feedforward Backpropagation Models with Principle Component Analysis (PCA)." Atmospheric Environment,vol. 77, pp. 621-30, 2013.
45. Ul-Saufie, A., Yahaya, A., Ramli, N. "Improving Multiple Linear Regression Model Using Principal Component Analysis for Predicting Pm_{10} Concentration in Seberang Prai, Pulau Pinang." International Journal of Environmental Sciences,vol. 2, no. 2, pp. 403-08, 2010.

46. USEPA. "An Introduction to Indoor Air Quality (IAQ), Carbon Monoxid (CO)."UnitedStates Environmental Protection Agency, http://www.epa.gov/iaq/co.html.2013
47. Viotti, P., Liuti, G., and Genova P. "Atmospheric Urban Pollution: Applications of an Artificial Neural Network (ANN) to the City of Perugia." Ecological Modelling,vol. 148, no. 1, pp. 27-46, 2002.
48. Wang, W., Lu, W., Wang, X., and YT A. "Prediction of Maximum Daily Ozone Level Using Combined Neural Network and Statistical Characteristics." Environment International,vol. 29, no. 5, pp. 555-62, 2003.
49. WHO. Methods for monitoring indoor air quality in schools, World Health Organization (WHO) Bonn. Germany, 2011.
50. Wilks, D. Statistical Methods in the Atmospheric Sciences. Vol. 100: Academic press, 2011.
51. Yang, J., Rivard H., and Zmeureanu R. "On-Line Building Energy Prediction Using Adaptive Artificial Neural Networks." Energy and Buildings,vol. 37, no. 12, pp. 1250-59 , 2005.
52. Yeniay, O., and Goktas A. "A Comparison of Partial Least Squares Regression with Other Prediction Methods." Hacettepe Journal of Mathematics and Statistics, vol. 31, no. 99, pp. 99-111, 2002.