Atmospheric Pollution Research 6 (2015) 99-106

Short Communication

Atmospheric Pollution Research

www.atmospolres.com

Development of artificial intelligence based NO_2 forecasting models at Taj Mahal, Agra

Dhirendra Mishra, Pramila Goyal

Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi–110016, India

ABSTRACT

The statistical regression and specific computational intelligence based models are presented in this paper for the forecasting of hourly NO_2 concentrations at a historical monument Taj Mahal, Agra. The model was developed for the purpose of public health oriented air quality forecasting. Last ten–year air pollution data analysis reveals that the concentration of air pollutants increased significantly. It is also observed that the pollution levels are always higher during the months of November at around Taj Mahal, Agra. Therefore, the hourly observed data during November were used in the development of air quality forecasting models for Agra, India. Firstly, multiple linear regression (MLR) was used for building an air quality–forecasting model to forecast the NO_2 concentrations at Agra. Further, a novel approach, based on regression models, principal component analysis (PCA) was analyzed to find the correlations of different predictor variables between meteorology and air pollutants. Then, the significant variables were taken as the input parameters to propose the reliable physical artificial neural network (ANN)–multi layer perceptron model for forecasting of air pollution in Agra. MLR and PCA–ANN models were evaluated through statistical analysis. The correlation coefficients (R) were 0.89 and 0.91 respectively, for PCA–ANN and were 0.69 and 0.89 respectively for MLR in the training and validation periods. Similarly, the values of normalized mean square error (NMSE), index of agreement (IOA) and fractional bias (FB) were in good agreement with the observed values. It was concluded that PCA–ANN model performs better and can be used for forecasting air pollution at Taj Mahal, Agra.

Keywords: *Air pollution, ANN, meteorological variables, PCA, Taj Mahal*

Corresponding Author:
Dhirendra Mishra
☎ : *+91-11-2659-6052*
🖷 : *+91-11-2659-1386*
✉ : *dhirendra.mishra@cas.iitd.ac.in*

Article History:
Received: 01 April 2014
Revised: 22 July 2014
Accepted: 23 July 2014

doi: 10.5094/APR.2015.012

1. Introduction

Agra city, where the Taj Mahal stands, has been polluted due to the increment of vehicles and industries over the past decades. There are many illegal factories, which are springing up around the Taj Mahal. The uncontrolled construction activities around the area also contribute to air pollution in the city. The river Yamuna, coming from National Capital Region (NCR) Delhi, is heavily polluted by industries impacting the magnificent statues and beautiful art of the world. When moisture gets accumulated in polluted environments, carbon, nitrogen and oxides of sulfur form weak acids (i.e., sulfuric, nitric acids etc). These acids are corrosive on textiles, paints, stone and metals. As a result many priceless works of art have been damaged in the last 20 years, irrespective of the previous 200 years due to increased these pollutants in the ambient atmosphere (Rao, 1978). Some pollutant gases like oxides of nitrogen (NO_2), carbon monoxide (CO), sulfur dioxide (SO_2) and ozone (O_3) are observed to be the most affecting pollutants for Taj Mahal. It is noticed that NO_2 emissions from anthropogenic sources have been reached up to ten times above the prescribed National Ambient Air Quality Standard (NAAQS) in the city. Combined with oxygen and moisture, NO_2 with SO_2 settles on the surface of the tomb and corrodes the marble, forming a fungus known as “marble cancer”. Also, the yellow coloring of marbles was blamed on consistently high–level air pollutants, which is generally caused by burning fossil fuels and by dust. Long–term exposures to NO_2 can also lower human resistance to respiratory infections and aggravate existing chronic respiratory diseases in the meantime. Increased levels of NO_2 can have significant impacts on people with asthma because it can cause more frequent and more intense attacks. Children with asthma and older people with heart disease are most at risk. Also, NO_2 contribute to ozone formation and can also have adverse effects on both terrestrial and aquatic ecosystems. In urban outdoor air, the presence of NO_2 is mainly due to traffic. Nitrogen oxides are emitted into the atmosphere primarily from vehicular exhaust as nitric oxide and reacts with ozone to form nitrogen dioxide (Gardner and Dorling, 1999). Mainly unvented heaters and gas stoves produce indoor NO_2 pollution. NO_2 (and other nitrogen oxides) is also a precursor for a number of harmful secondary air pollutants, including nitric acid, the nitrate in secondary inorganic aerosols, and photo oxidants (including ozone). The situation is also complicated by the fact that photochemical reactions take some time (depending on the composition of the atmosphere and meteorological parameters) and air can travel some distance before secondary pollutants are generated (WHO, 2003). Health risks from nitrogen oxides may potentially result from NO_2 itself or its reaction products including O_3 and secondary particles. These relationships are shown schematically in Figure 1. It was observed that the NO_2 level was always greater than EPA's health–based national air quality standard i.e., 0.053 ppm (annual arithmetic mean). Therefore, NO_2 was chosen for the analysis in the present study because epidemiological studies have revealed that the NO_2 is much more important for mortality from respiratory and cardiovascular causes for respiratory conditions (Burnett et al., 1997).

Many previous studies have shown that the Artificial Neural Network (ANN) based air pollution models perform better than other statistical techniques (Gardner and Dorling, 1999; Lu et al., 2003; Wang et al., 2003a; Kumar and Goyal, 2013) and ANN based air pollution forecasting models are being implemented in some cities (Wang et al., 2003b; Kumar and Goyal, 2011). However, very few recent studies have investigated its internal mechanism to understand the extent to which the modeled function identifies the relative contribution of these controlling emissions and meteorological parameters to the observed concentrations. Time–lagged models (models that can forecast concentrations), for example, one hour ahead, three hours ahead or 24 hours ahead, are found to give reliable predictions (Gardner and Dorling, 1999) but are not very useful when large gaps in data are present due to equipment failure. There are many analogous methods and their combinations including Multiple Linear Regression (MLR), ANN and Principal Component Analysis (PCA) were also adopted in the many previous studies over different cities to forecast the pollutants (Kumar and Goyal, 2011). In the present study MLR as well as PCA–ANN techniques has been used for NO_2 forecasting at Taj Mahal, Agra. The PCA is chosen to reduce the dimension of input variables in the ANN model with lowering the training time and preserving or even improving the ANN model's accuracy. The main aim of this study is to demonstrate a methodology for extracting knowledge of the relative contribution of different meteorological parameters on pollutant concentrations and use it in the construction of a robust PCA–ANN model (based on pattern recognition). It is useful in situations where one has to rely only on a few meteorological predictors to model pollutant concentrations and also be used to identify the key meteorological parameters that should be measured with a greater sensitivity. Therefore, it could be said that this is the first study, which uses the statistical as well as artificial intelligence methodologies to forecast air pollutants over the region of a historical monument, Taj Mahal, Agra.

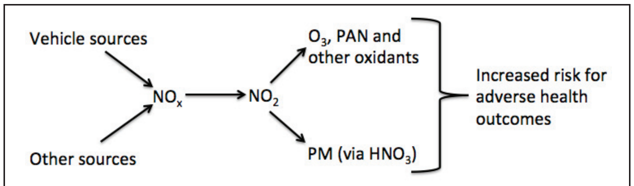

Figure 1. *Relationship of nitrogen oxide emissions with formation of NO_2 and other harmful reaction products including O_3 and PM.*

2. Materials and Methods

2.1. Study area and data sample

Agra is best known for the Taj Mahal and as an important tourist destination, transport hub and commercial center. It stretches across 26°44'N to 27°25'N and 77°26'E to 78°32'E. It is a district of Uttar Pradesh state and its borders touch Rajasthan to from west and south. Figure 2 shows the Agra city map with Central Pollution Control Board (CPCB) monitoring station, Taj Mahal with data sampling station Sanjay Place.

In the present study, the meteorological parameters i.e., Temperature (Temp), Relative Humidity (RH), Wind Speed (WS), Wind Direction Index (WDI), Vertical Wind Speed (VWS), Solar Radiation (SR) and Pressure (Pres) as well as the air pollutants CO, O_3 and SO_2 have been used to develop an ANN–MLP forecasting model for Taj Mahal city, Agra. Central Pollution Control Board (CPCB) at Sanjay Place, Agra, India sampled the relevant data. The sampling period was 18 November to 27 November 2013 at Sanjay Place, Agra. The missing values in the data have been interpolated by the neighboring values using linear interpolation. The proposed methodology is for understanding the contribution of different predictor variables using a neural network to learn as much as possible from concentrations between physical values and future pollutant concentrations, in order to achieve better accuracy. The sampled data are considered to be very close to the actual values (there are also measurement errors, but they should be significantly smaller). Due to the circular nature of wind directions, wind direction index has been used in this study (Goyal et al., 2014). According to Gardner and Dorling (1999), models including or omitting emissions resulted in extremely similar results. Therefore, the unavailable emission data were not used in the model development. The maximum observed concentrations of NO_2 during the sampling periods were sometimes more than 5–6 times by the prescribed NAAQS. Low concentrations of O_3 were caused by the depletion of ozone in the oxidation of traffic–originated nitrogen oxides. The CO concentrations were also much higher compared to mean values (4–5 times). Table 1 presents the basic statistics about the parameters studied in this study.

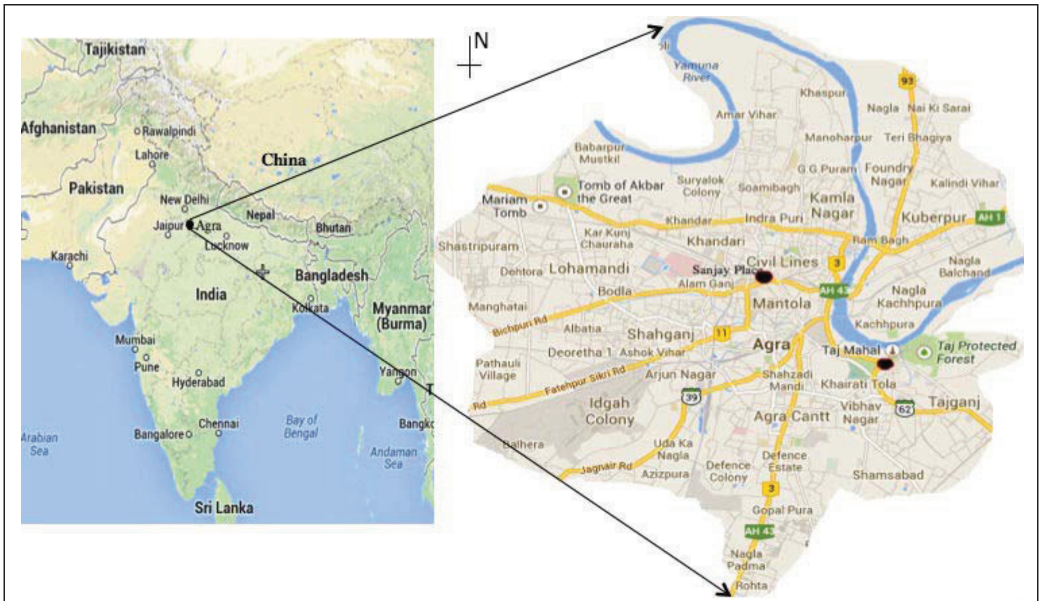

Figure 2*. Agra city map with CPCB monitoring station–Taj Mahal and data sampling station–Sanjay Place (Source: www.google.co.in/map).*

***Table 1**. Statistics of the measured values during the study period*

Pollutants/Variables	Minimum	Maximum	Mean	St. Dev.
NO_2 (µg/m^3)	4.74	392.41	51.87	14.36
SO_2 (µg/m^3)	3.19	31.46	12.60	4.07
CO (µg/m^3)	40.00	7 730.00	1 574.26	1 176.83
O_3 (µg/m^3)	2.31	104.80	9.88	7.62
Benzene (µg/m^3)	0.38	1.67	0.70	0.29
Toluene (µg/m^3)	1.17	19.11	4.27	3.61
Xylene (µg/m^3)	0.39	11.15	1.00	1.16
Temperature (°C)	11.53	28.75	20.12	4.90
Relative Humidity (%)	30.94	73.14	48.85	10.64
Wind Speed (m/s)	0.27	6.88	0.82	1.28
Vertical Wind Speed (m/s)	–0.39	–0.04	–0.20	0.04
Wind Direction Index	0.00	2.00	0.91	0.73
Solar Radiation (W/m^2)	22.53	693.01	196.80	233.54
Pressure (mm Hg)	745.17	746.64	745.83	0.43

2.2. Multiple Linear Regression (MLR) model

A forecast can be expressed as a function of a certain number of factors that determine its outcome. MLR technique includes one dependent variable to be predicted and two or more independent variables. In general, multiple linear regression can be expressed as in Equation (1):

$$Y = b_1 + b_2X_2 + \cdots + b_kX_k + e \tag{1}$$

where, Y is the dependent variable, X_2, X_3,..., X_k are the independent variables, b_1, b_2,..., b_k are linear regression parameters. In this study, NO_2 is the dependent variable and air pollutant concentrations and meteorological variables are independent variables, e is an estimated error term which is obtained from independent random sampling from the normal distribution with mean zero and constant variance. The task of regression modeling is to estimate the b_1, b_2,..., b_k which can be done using least square error technique.

2.3. Principal Component Analysis (PCA)

Among the many methods available for visualizing ambient air pollution and meteorological data, PCA has been used for this study, as it is able to capture the predominant parameters of the data. PCA is a computer intelligence method originating from multivariate statistical analysis, which allows for the selection of the major factors within a certain dataset (Jolliffe, 2002). It is capable of identifying interrelations within the dataset. So, PCA was considered as a tool capable of providing an overview of the interdependencies and variability of data and extracting information for forecasting mechanisms (Kumar and Goyal, 2013). The role of PCA is to reduce the number of predictor variables and transform them into new variables, which are called Principal Components (PC). These are obtained from the independent linear combinations of original data, which retain the maximum possible variance of the same data. The correlation matrix of the normalized input data can compute the PCs. The eigenvalues of the correlation matrix "C" are obtained from its characteristic equation as given below:

$$|C - \lambda I| = 0 \tag{2}$$

where, λ is the eigenvalue and I is the identity matrix. For each eigenvalue, λ a non–zero eigenvector e exits, which can be defined as:

$$C_e = \lambda\, e \tag{3}$$

The eigenvectors are derived from the correlation matrix. Their associated eigenvalues represent the total amount of variances, which are explained by each of the eigenvectors. The higher order principal components having minimal amounts of the total variance can be viewed as noise. The i^{th} variance of i^{th} PC is given as:

$$Variance = \frac{\lambda_i}{\sum_n \lambda_n} \tag{4}$$

The PCs are associated with the greatest eigenvalue represent the linear combination of the variables, which is accounting for the maximum total variability in the data. After getting all the PCs, the initial data set is transformed into the orthogonal set by multiplying the eigenvectors.

2.4. Artificial neural network – multi layer perceptron model

The forecasting of air quality can be considered as a non–linear regression problem between predictand (here, hourly concentration) and predictors (such as meteorological and air quality variables). ANN models are capable of approximating any smooth differentiable function and used for modeling complex non–linear processes. ANN models have been utilized for several tasks within the air quality domains, such as forecasting, function approximation and pattern classification (Gardner and Dorling, 1999). Multi Layer perceptrons have been applied successfully to solve some difficult and diverse problems, by training them in a supervised manner with a highly popular algorithm. In the current study, PCA–ANN models were applied in order to forecast, hourly mean concentrations of NO_2 in the Agra city. Like other studies, it was observed that the pollutants are present continuously throughout the day, but hourly variations of concentrations are more effective in developing statistical models. Therefore, hourly data was chosen for training and validation.

Technically, the process of ANN is briefly explained as follows. The neuron is the basic information processing unit of an ANN. It consists of a set of links, describing the neuron inputs, with weights W_1, W_2, ..., W_m and an adder function (linear combiner) for computing the weighted sum of the inputs i.e. $u = \sum_{j=1}^{m} W_j\, X_j$. Finally, activation function φ for limiting the amplitude of the neuron output i.e. $y = \varphi(u + b)$, where b denotes bias. All the processes are depicted in Figure 3.

The neural networks basically involve the collection of neurons that are configured in two or more layers. The combination of neurons into multilayer structures gives the power of pattern recognition and prediction. The multi layer feed–forward network comprises of an input layer, hidden layer, and output layer. Specifically, the input layer is a layer that is directly connected to outside information. All data in the input layer will be feed–forwarded to the hidden layer as the next layer. Meanwhile, the hidden layer functions as feature detectors of input signals and releases them to the output layer. Finally, the output layer is considered as a collector of the features detected and as a producer of the response. In the networks, the output from output layer is the function of the linear combination of hidden unit's activation; whereas the hidden unit's activation function is in the form of a non–linear function of the weighted sum of inputs. The building and training of the network are then carried out using Matlab 7.12 (Licensed, IIT Delhi). Under the assumption that concentration pattern does not change significantly from one day to the next, the proposed model can be used to forecast concentrations for the consecutive hour by providing values of new predictor variables.

3. Results and Discussion

3.1. Common characteristics of observed concentrations

In this study, the observed concentrations of air pollutants and meteorological variables for Taj Mahal, Agra were investigated, with the aim to develop an air pollution forecasting model. The variation in annual average concentrations of various air pollutants (NO_2, RSPM, SPM, and SO_2) at Taj Mahal, Agra, over the period of 2004 to 2013 (ten years) is shown in Figure 4. It was observed that the concentration of NO_2 increased significantly during the last ten years, while the concentrations of RSPM, SPM and SO_2 did not. Further, a comparison of the monthly average values of various criteria pollutants (NO_2, RSPM, SPM and SO_2) at Taj Mahal, Agra, over the period of 2011 to 2013 are given in Figure 5. The monthly variations of air pollutant concentrations were almost same. The minimum values were observed during July and August (monsoon season) and maximums during December and January (winter season). For NO_2 (Figure 5a), the standard deviation is found to be the maximum as 8.32 μg/m^3 in 2013, followed by 6.67 and 4.61 μg/m^3 respectively, during 2012 and 2011. The maximum standard deviations are observed for RSPM and SPM in 2012 (Figures 5b–5c) and SO_2 shows little variations. It is also well understood that the precipitation also affects the pollutant concentrations, as the concentrations are observed lower on rainy months compared to non–rainy months. This may be due to the washing out of pollutants by rain, resulting in a reduction in their concentrations.

It can be concluded from the above analysis that there are increments of NO_2 concentrations during November and December (winter season) and exceeded the prescribed NAAQS (approximately 30 μg/m^3) many times during these months. Also, the tourist count was observed maximum during these months. Thus the study of NO_2 forecasting model is important in Agra city.

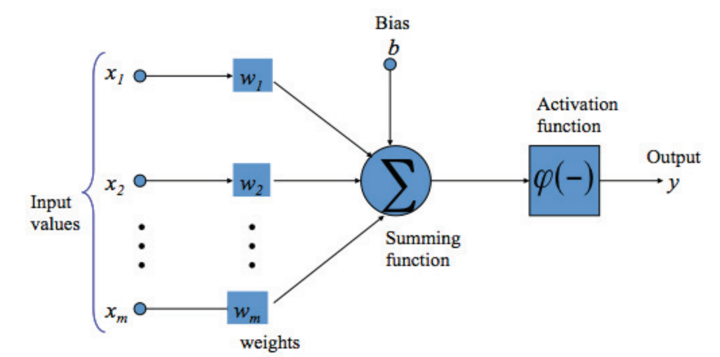

Figure 3. *The architecture of an artificial neural network.*

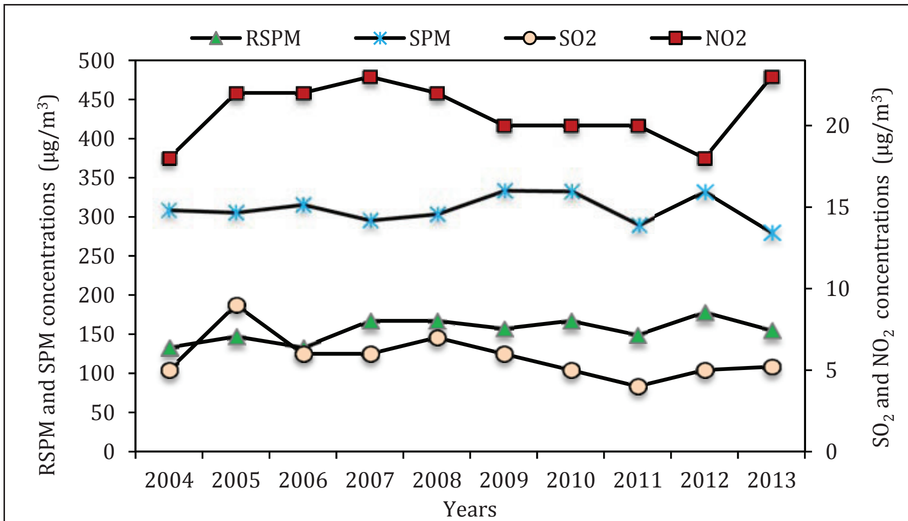

Figure 4. *Variation of the annual average concentrations (μg/m^3) of criteria pollutants.*

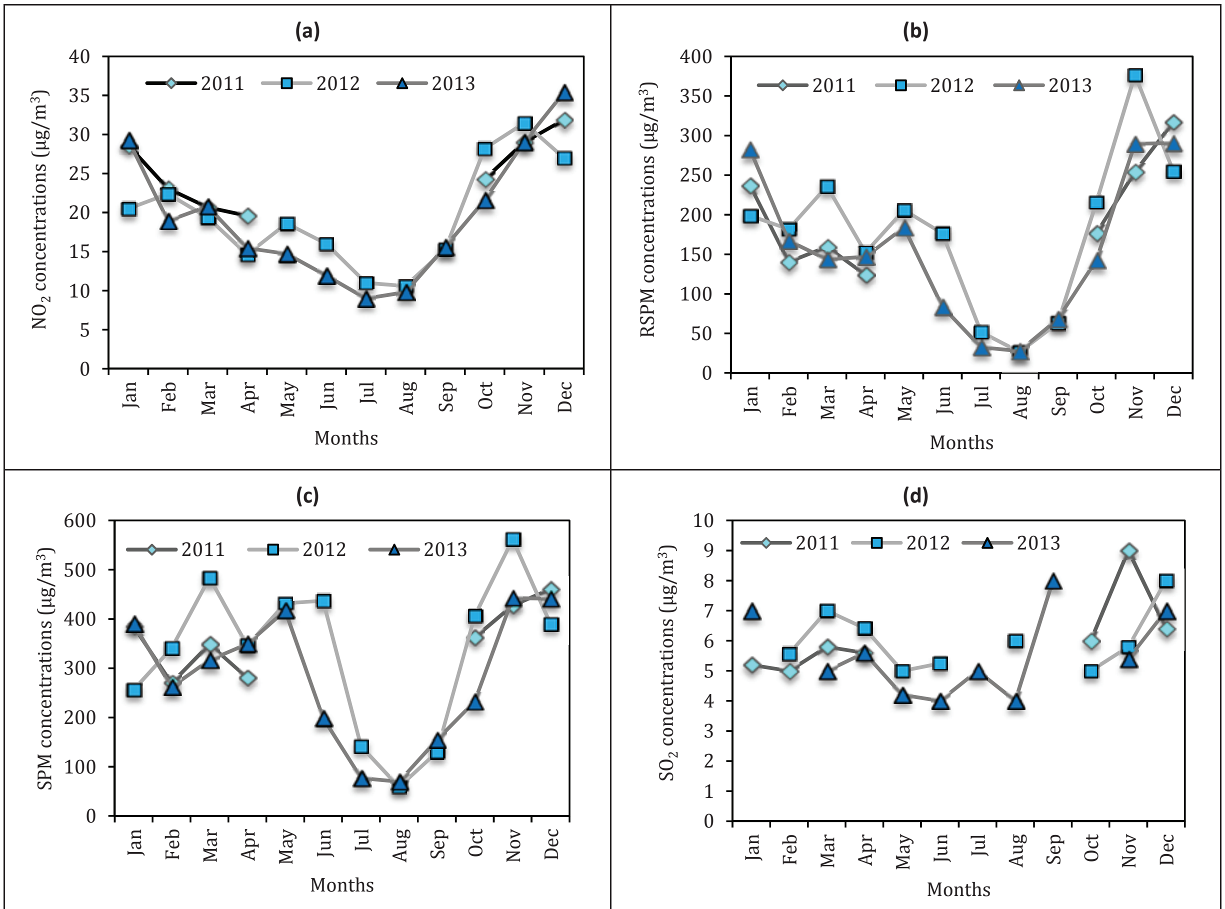

Figure 5. *Variation in the monthly average concentrations (µg/m³) of **(a)** NO_2, **(b)** RSPM, **(c)** SPM and **(d)** SO_2.*

3.2. Modeling: selection of input variables and evaluation

Table 2 illustrates how NO_2 concentrations are related to meteorological variables and other air pollutants. The meteorological variables are temperature, relative humidity, wind speed, wind direction index, vertical wind speed, solar radiation and pressure. The air pollutants are CO, NO, O_3, SO_2, Benzene (Ben), Toluene (Tol) and Xylene (Xyl). At high values of solar radiation (during daytime), NO_2 concentrations are linearly related to both temperature (positively) and relative humidity (negatively).

Table 2 shows that NO_2 concentrations are negatively correlated with solar radiation, wind speed, temperature, pressure, ozone and positively correlated with carbon monoxide, benzene, toluene, sulfur dioxide, relative humidity etc. It reflects that the correlations between most of the variables are not poor as their values are more than 0.056, which represents that the correlation coefficient is significant. Since the critical value of the correlation coefficient of the 99th percentile confidence interval using student's *t*-test for the study period is found to be 0.056, which is statistically significant. Since NO_2 and CO is the precursor gases of surface ozone and the increase of ozone in winter months is attributed to the increase in NO_2 concentrations. Again, the SO_2 conversion reaction increases in the presence of O_3 and relative humidity. Whenever, the atmospheric relative humidity tends to be greater than 80%, the pollutants absorb moisture and starts growing rapidly in size, i.e., the resultant increased scattering may lead the increase in NO_2 concentrations (Goyal et al., 2014). The decrease in pressure enhances low pressure as well as wind speed, which may increase the concentration of NO_2. Also, it is a well–known fact that the high wind speed causes low concentration of pollutants. It has also been found that NO_2 has positive correlation (directly proportional) with many variables such as benzene, toluene and xylene. The reduction in solar radiation results in low temperature and low boundary layer height, which leads to high concentration of pollutants.

The NO_2 forecasting conditions in terms of independent variables have been studied during November in Agra. The hourly meteorological and air quality data available for the year 2013 were used for training and validation of the forecasting ability of the MLR model. The hourly data for eight days (192 hours data) was selected for training and next two days (48 hours) data for validation of the proposed model. A MLR model shown in Equation (5) was derived with the help of hourly time series data to explain the variation in NO_2 concentrations due to meteorological variables and air pollutants:

$$\begin{aligned} NO_2 = {} & 0.16\,CO - 0.425\,O_3 - 0.018\,SR + 7.147\,Temp \\ & -76.536\,Pres - 1.362\,Tol - 1.945\,Xyl \\ & +8.665\,Ben + 56\,981.193 \end{aligned} \tag{5}$$

where the aberrations show their usual meaning.

PCs were computed for the available air quality and meteorological variables. The correlation matrix of normalized input data was obtained and the PCs were determined on the basis of the variance explained by the eigenvalues of the correlation matrix. The PCs and the cumulative amount of variance are presented in Table 3 with its eigenvalues and amount of variance. However, the PCs, whose cumulative amounts of variance are approximately 80%, are used in the model and remaining components were excluded (Table 3). Therefore, only 5 new variables (PCs) were used instead of the original 14 variables. The coefficients of each input variable corresponding to all five PCs were calculated. Concerning the PCA–ANN model specifications and the way that their results are evaluated, the training process of the PCA–ANN models was based on the back–propagation algorithm. The PCA–ANN models have been trained and validated separately with the hourly data of air pollutants and meteorological variables. The hyperbolic tangent function has been used as the activation function. The performance of MLR and PCA–ANN model for NO_2 for 48 hours is shown in Figure 6.

The performances of the models were also assessed on the basis of statistical measures. The evaluations of the final forecasting models are based on statistical indices, such as correlation coefficient (*R*), Index of Agreement (IOA), Normalized Mean Square Error (NMSE) and Fractional Bias (FB). More details about the statistical indices are included in the Appendix. The statistical performances computed between observed and predicted NO_2 concentrations from MLR and PCA–ANN models are given in Table 4.

The trained PCA–ANN model for NO_2 concentrations shows the value of *R* as 0.89 in training and 0.90 in validation while MLR model shows 0.84 and 0.69 respectively. The values of IOA between observed and predicted NO_2 concentrations were >0.98 for both models in the training and validation dataset, which is close to its ideal value 1.0. The values of NMSE were found as 0.024 and 0.017 for MLR as well as 0.016 and 0.017 for PCA–ANN in the training and validation datasets respectively. The values of fractional bias (FB) for both models were close to 0.0 for training and validation, respectively. Thus, both models showed acceptable results in training as well as in the validation and the values of statistical measures for PCA–ANN model were close to their corresponding ideal values. Thus, it can be concluded that PCA–ANN model is performing better than MLR model. It was also observed that the high concentration values are not predictable by the proposed model, which may be due to the local anthropogenic emission activities. Overall, the performance of the PCA–ANN models is observed satisfactory and it can be considered for operational use.

Table 2. *Correlation matrix of air pollution concentrations with meteorological variables in Agra, India*

	NO_2	SO_2	CO	O_3	Tmp	RH	WS	WDI	Ben	Tol	Xyl	VWS	SR	Pres
NO_2	1.00													
SO_2	0.37	1.00												
CO	0.60	0.79	1.00											
O_3	–0.12	0.10	0.01	1.00										
Tmp	–0.14	–0.39	–0.26	0.19	1.00									
RH	0.21	0.52	0.41	–0.14	–0.86	1.00								
WS	–0.25	–0.24	–0.28	0.13	0.43	–0.45	1.00							
WDI	–0.09	–0.04	–0.04	0.04	0.02	0.05	–0.03	1.00						
Ben	0.53	0.63	0.76	–0.08	–0.43	0.41	–0.34	–0.14	1.00					
Tol	0.49	0.73	0.89	–0.13	–0.29	0.37	–0.26	–0.07	0.79	1.00				
Xyl	0.23	0.58	0.53	0.05	–0.31	0.45	–0.19	–0.14	0.59	0.47	1.00			
VWS	–0.03	0.02	0.00	–0.01	–0.15	0.12	–0.10	–0.04	0.03	0.01	–0.01	1.00		
SR	–0.46	–0.47	–0.42	0.19	0.70	–0.62	0.54	0.00	–0.47	–0.40	–0.28	–0.11	1.00	
Pres	–0.17	–0.41	–0.28	0.20	1.00	–0.86	0.44	0.02	–0.45	–0.31	–0.33	–0.15	0.71	1.00

Table 3. *Eigenvalues and variances of the computed PC's*

Principal Component	PC1	PC2	PC3	PC4	PC5
Eigenvalues	5.803	2.148	1.151	1.056	0.955
% of Variance	41.449	15.343	8.220	7.542	6.822
Cumulative variance (%)	41.449	56.793	65.013	72.555	79.377
NO_2	0.542	0.362	–0.463	0.087	0.016
SO_2	0.777	0.333	0.272	0.050	0.025
CO	0.783	0.529	–0.009	0.090	–0.031
O_3	–0.153	0.264	0.746	0.073	0.052
TEMP	–0.746	0.600	–0.114	0.050	0.037
RH	0.785	–0.435	0.197	0.043	–0.025
WS	–0.536	0.283	0.213	–0.151	0.124
WDI	–0.075	–0.154	0.145	0.875	0.099
BEN	0.811	0.342	–0.053	–0.081	–0.051
TOL	0.763	0.473	–0.097	0.040	–0.039
XYL	0.618	0.270	0.347	–0.170	–0.031
VWS	0.102	–0.242	0.106	–0.439	–0.437
SR	–0.764	0.311	0.181	–0.083	0.039
PRES	–0.766	0.578	–0.101	0.044	0.038

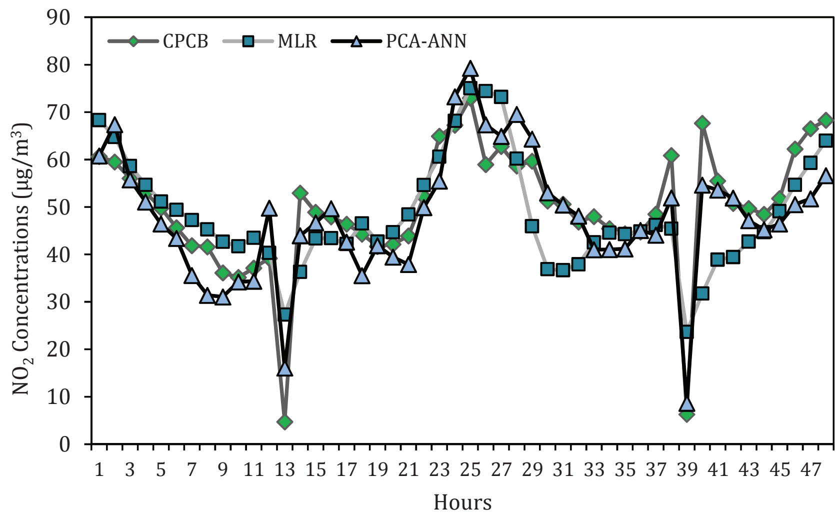

Figure 6. *Comparison of the observed (CPCB) and modeled (MLR and PCA-ANN) hourly NO_2 concentrations.*

Table 4. *Statistical performance measures for ANN–MLP model*

Statistical Measures	Ideal Value	MLR		PCA–ANN	
		Training	Validation	Training	Validation
R	1	0.84	0.69	0.89	0.91
IOA	1	0.99	0.98	0.99	0.98
NMSE	0	0.024	0.017	0.016	0.017
FB	0	0.024	0.003	0.001	–0.021

4. Conclusions

In this study regression as well as neural network with multi–layer perceptron were used for designing the air pollution–forecasting models at historical monument, Taj Mahal, Agra. It gives the better air pollutant-forecasting approaches model based on ANN techniques for modeling continuously hourly time series data with PCA approaches performed better than the MLR techniques at Taj Mahal, Agra. A unique approach, based on general linear models, PCA, was employed in selecting the significant input variables. The model is built on measured meteorological variables and concentrations of the air pollutant concentrations. A good agreement was observed between the predicted and observed pollutant concentrations at Taj Mahal, Agra. It was observed that the both the proposed models couldn't forecast well during high concentration pollutant periods due to the input and architecture of the model itself. Further, the temporal variables, which imply variation of anthropogenic activities, such as traffic, industries, etc. are the most important for forecasting the concentrations of air pollutants in Agra. Therefore, more research work is needed for Taj Mahal, Agra to show the significant variations of the values of high concentrations as well as the temporal variables. Overall, increasing air pollution in Agra started affecting the world heritage monument, Taj Mahal. Therefore, to control the vehicular pollution, green fuels like CNG should be introduced in Agra as soon as possible.

Acknowledgment

The authors would like to thank Central Pollution Control Board, New Delhi for providing the sampled and observation data at different locations for this study in Agra. The authors also would like to thanks the anonymous reviewers for their valuable comments and suggestions.

Appendix

The measures of model performance are the correlation coefficient (*R*), index of agreement (IOA), normalized mean square error (NMSE) and fractional bias (FB). These measures were calculated as follows:

$$R = \frac{\overline{(C_O - \overline{C_O})(C_P - \overline{C_P})}}{\sigma_{C_P}\sigma_{C_O}} \tag{6}$$

$$IOA = 1 - \frac{\sum_{i=1}^{n}(C_{P_i} - C_{O_i})^2}{\sum_{i=1}^{n}\left(\left|C_{P_i} - \overline{C_O}\right| + \left|C_{O_i} - \overline{C_O}\right|\right)^2} \tag{7}$$

$$NMSE = \frac{\overline{(C_O - C_P)^2}}{\overline{C_O}\ \overline{C_P}} \tag{8}$$

$$FB = \frac{(\overline{C_P} - \overline{C_O})}{0.5(\overline{C_P} + \overline{C_O})} \tag{9}$$

where, C_P is the model prediction, C_O is the observations, $\overline{C}$ is the average over the dataset, and σ_C is the standard deviation over the data set.

References

Burnett, R.T., Dales, R.E., Brook, J.R., Raizenne, M.E., Krewski, D., 1997. Association between ambient carbon monoxide levels and hospitalizations for congestive heart failure in the elderly in 10 Canadian cities. *Epidemiology* 8, 162–167.

Gardner, M.W., Dorling, S.R., 1999. Neural network modelling and prediction of hourly NO_x and NO_2 concentrations in urban air in London. *Atmospheric Environment* 33, 709–719.

Goyal, P., Kumar, A., Mishra, D., 2014. The impact of air pollutants and meteorological variables on visibility in Delhi. *Environmental Modeling & Assessment* 19, 127–138.

Jolliffe, I. T., 2002. *Principal Component Analysis*, Springer, New York.

Kumar, A., Goyal, P., 2013. Forecasting of air quality index in Delhi using neural network based on principal component analysis. *Pure and Applied Geophysics* 170, 711–722.

Kumar, A., Goyal, P., 2011. Forecasting of air quality in Delhi using principal component regression technique. *Atmospheric Pollution Research* 2, 436–444.

Lu, W.Z., Wang, W.J., Wang, X.K., Xu, Z.B., Leung, A.Y.T., 2003. Using improved neural network model to analyze RSP, NO_x and NO_2 levels in urban air in Mong Kok, Hong Kong. *Environmental Monitoring and Assessment* 87, 235–254.

Rao, S.T., 1978. Pollution of Taj Mahal and its Environs due to the proposed Mathura Oil Refinery. *International Symposium on Environmental Agents and Their Biological Effects*, Osmania University, Hyderabad.

Wang, W.J., Xu, Z.B., Lu, J.W., 2003a. Three improved neural network models for air quality forecasting. *Engineering Computations* 20, 192–210.

Wang, W.J., Lu, W.Z., Wang, X.K., Leung, A.Y.T., 2003b. Prediction of maximum daily ozone level using combined neural network and statistical characteristics. *Environment International* 29, 555–562.

WHO (World Health Organization), 2003. Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide. Report on a WHO Working Group, Bonn, Germany, http://www.euro.who.int/__data/assets/pdf_file/0005/112199/E79097.pdf, accessed in April 2014.