Songklanakarin J. Sci. Technol.
44 (5), 1256–1263, Sep. – Oct. 2022

http://www.sjst.psu.ac.th

Original Article

Prediction of PM10 pollution using principal component regression and hybrid artificial neural network model

Sateesh N Hosamane*

Department of Chemical Engineering, KLE Dr. M. S. Sheshgiri College of Engineering and Technology, Udyambag, Belgaum, 590008 India

Received: 11 September 2021; Revised: 13 July 2022; Accepted: 28 August 2022

Abstract

Air pollution, especially particulate matter (PM) pollution, has a significant impact on India. PM pollution is due to roadside dust, fossil fuel use, vehicular population, and industrial emissions. PM10 forecasting model development is essential because it permits the experts and the citizens to take appropriate actions to restrict their exposure and execute protective measures to improve air quality. This study aimed to develop a specialized computational intelligence methodology that uses principal component (PC) based artificial neural networks (ANN). The model was used to forecast PM10 in ambient air using meteorological data. This application is demonstrated for monitoring data from the urban area of Belagavi city of Karnataka state, India. Principal component analysis (PCA) was applied to understand the interactions between PM10 concentration and meteorological data. The analysis found that the PCANN model is better than the principal component regression (PCR) model, based on using various performance indexes (MAE, MSE, MAPE, RMSE, R, and R2). The PM10 predictive model performance was satisfactory, with a MAPE of 0.069. The overall predictive capability of PM10 was 89.59% in terms of R.

Keywords: air pollution, PM10, principal component, neural network, prediction

1. Introduction

Air pollution is a universal public health and environmental hazard across the globe (European Environment Agency ([EEA], 2013; Kolehmainen, Martikainen, & Ruuskanen, 2001). Urban air pollution is a complex mixture of toxic components, which may induce acute and chronic responses in sensitive groups. According to the World Health Organization, the top 20 polluted cities globally are situated in Asia. It is projected that people living in these unhygienic cities will lose about three years in longevity (World Health Organization [WHO], 2016). The leading impact on community health is from the quality of air that we breathe, which depends predominantly on particulate matter (PM). PM having a particle diameter of less the 10 micrometers (μm), is referred to as inhalable particles (PM10), and the concentration level of PM10 in the atmosphere is the primary measure of air quality. Many cities in India report very high annual average concentrations of suspended PM. They are the most significant environmental worries due to the harmful effects on living organisms and on the Earth's atmosphere (de Mattos Neto, Cavalcanti, Madeiro, & Ferreira, 2015). The level of injury depends upon the exposure time and the type and concentration of particles in the air (EEA, 2013). The correlation between PM10 and health effects is well recognized. Therefore, this is the most critical type of air pollutant that adversely influences human health (Hosamane, Prashanth, & Virupakshi, 2020). The extent of damage generally depends upon the period of exposure and concentration levels of particles in the air (Dadvand *et al*., 2013; Feng & Yang, 2012). Several epidemiological studies have associated the concentration of these pollutants with cardiovascular and respiratory diseases (Duan, Hao, & Yang, 2020). Early prediction of air pollutants contributing to PM10 has gained considerable concurrent interest due to its impact on the environment and the humans (Yadav & Nath, 2019). The concentration of PM10 over a specific locality is strongly influenced by emission sources, atmospheric conditions, and chemical transformations. The interactions between pollutant concentration and meteorology

*Corresponding author
Email address: satishosamane@gmail.com

is one of the best approaches available for urban planning and management to recognize these pollutants' adverse effects (Bhaskar & Mehta, 2010). Cogitate all these factors: air quality is significant for human health and the atmosphere; therefore, it is essential to extend a management and control approach that upholds the environment (Park *et al*., 2018). Two methods are generally used to predict pollutant concentrations in ambient air: deterministic modeling or stochastic approximation (Ding, Zhang, & Leung, 2016). The statistical models are an alternate choice to analyze extreme air pollution events. Several models have been developed to predict pollutant concentration using multiple linear regression (MLR) (Ng & Awang, 2018, Pires, Martins, Sousa, Alvim-Ferraz, & Pereira, 2008). The linear statistical models are susceptible to outliers and any errors corrupting the data (Cocchi Greco & Trivisano, 2007). To identify this complexity, other statistical methods, such as the principal component analysis (PCA), have proven advantageous in reducing data. The most critical variables are highlighted in the data set for further analysis (Nagendra & Khare, 2003; Thurston, Ito, & Lall, 2011; Wolff, Korsog, Kelly, & Ferman, 1985). Various researchers have proposed combinations of PCA-based regression models referred to as principal component regression (PCR), to predict pollutant concentrations, and these can overcome problems from multicollinearity among variables (Abdullah, Ismail, Fong, & Ahmed, 2016: Kumar & Goyal, 2011; Pires *et al*., 2008). Several researchers have published work on artificial neural networks (ANNs), and principal component (PC) based ANN models (Akkoyunlu, Yetilmezsoy, Erturk, & Oztemel, 2010; Baawain & Al- Series, 2014; Sousa, Martins, Alvim-Ferraz, & Pereira, 2007; Tran, Kim, Kim, Choi, & Choi, 2018). These models have established good accuracy with meteorological parameters as input variables in appropriately predicting air pollutants. Application of ANN in the field of air quality prediction has gained acceptance in the present, and is extensively accepted by the emerging nations, including India (Chellali, Abderrahim, Hamou, Nebatti, & Janovec, 2016; Chelani, Gajghate, & Hasan, 2002; Mishra, Goyal, & Upadhyay, 2015; Srimuruganandam & Nagendra, 2015). The present study focuses on building a precise approach to handling data to deliver improved modeling. The air quality of Belagavi is deteriorating, and no air quality models are well established to understand the future air pollution. The main objective of this paper is to demonstrate the application of a principal component (PC) based artificial neural network (ANN) model to forecast the daily average concentration (24-hour) of PM10 for Belagavi city. The proposed model is compared with principal component regression (PCR) models for validation results.

2. Area Description and Data Used

2.1 Study area

Belagavi is a town within Karnataka (India) at 15.87◦n 74.5◦e and has a mean height of 751 meters. The town is located in the north-western part of Karnataka and is bordered by two provinces. The city has a population of 500,000 according to the 2011 census, with a sex ratio of 988 women per 1000 men (Census, 2011). The air pollution monitoring site is located at a traffic intersection, which is packed all day due to the schools, colleges, and railway station located on the main highway leading to the state of Goa. Figure 1 shows a map and satellite image of the monitoring area.

2.2 Collection of data

The air pollution monitoring was conducted from February 2011 to January 2016 for five years. The site was equipped with Envirotech APM 460 dust sampler and the gas sampler RDS 433. Each week two to three sets samples were planned, and air pollution was monitored according to the central pollution control board (CPCB, 2009). The sampler was kept 3 to 5 m above the ground, and the samples were monitored and analysed in the laboratory. The RDS 460 air sampler was used with a flow rate of 0.8-1.25 m^3/min to remove dust from the air. The particles trapped on the filter paper are referred to as RSPM or PM10 (size <10 microns), while collected under a rotary cup are un-breathable sized particles (size >10 microns).

2.3 Data for the study

The daily averages (24 hours) of five metrological factors and three pollutant concentrations were gathered for five years. The typical annual averages of temperature (T), wind velocity (WS), wind direction (WD), precipitation (R), and humidity (H) were found to be 25.43 ºC, 2.79 m/s, 193.70 degrees, 0.397 mm, and 64.46%. The statistical summary descriptions of meteorological parameters and PM10, SO2 and NO2 (μg/m^3) concentrations are shown in Table1.

Figure 1. Location and aerial photo of monitoring site

Table 1. Basic statistics for monitored data

Variable	Mean	SD	Minimum	Maximum
T,	25.43	2.91	19.73	34.11
WS	2.79	1.10	0.18	12.11
WD	193.70	80.46	22.55	329.50
R	0.397	0.813	0.00	9.23
H	64.46	20.30	17.50	99.07
PM_{10}	88.25	33.12	11.46	224.72
SO_2	7.714	3.032	0.19	21.89
NO_2	28.231	11.29	7.17	87.57

Figures 2 shows the variation of PM10 with temperature, wind speed, wind direction, rain, and humidity. The concentration of PM10 slightly decreased with increasing wind speed, and at higher wind velocities the pollutants get dispersed. Therefore, the volume and dilution of air pollutants are controlled by wind direction and wind speed. The temperature is the most crucial metrological parameter that influences the pollutant concentrations of PM10 and SO2, and the concentrations increase with temperature. Rainfall has a significant effect on PM10. As the rainfall increases, the concentration of PM10 decreases. PM10 concentration decreases with relative humidity (Bhaskar & Mehta, 2010; Kumar & Goyal, 2011).

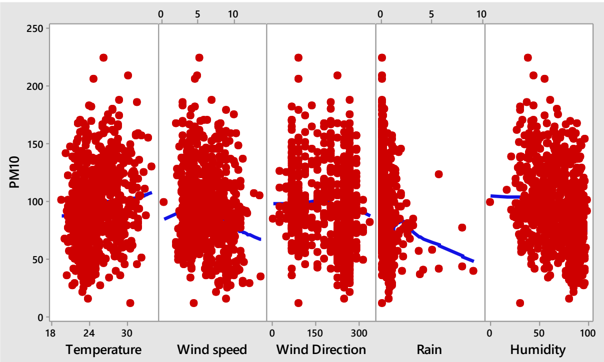

Figure 2. Effect of meteorological parameters on PM10

2.4 Computation of principal components (PC)

The principal difference between these two (PCA and MLR) techniques is in the transformed inputs reducing the dimension/complexity of the data, by creating a small number of input variables. The use of PC with regression reduces collinearity in the dataset, which could cause undesirable effects on model predictions. It also determines the independent functional variables for the prediction of pollutant concentration (de Souza, Reisen, Franco, Ispány, Bondon, & Santos, 2018). In the present investigation, the covariance matrix of the original data was assessed. The eigenvalues of the covariance matrix are found from the usual determinant condition:

$$|C - \lambda I| = 0 \tag{1}$$

where 'λ' and 'I' represent an eigenvalue and the identity matrix.

An eigenvector 'e' of the covariance matrix C for that eigenvalue satisfies:

$$Ce = \lambda e \tag{2}$$

The eigenvectors of the covariance matrix enable orthogonal rectilinear regression. The PC with the highest total variance is used for the selection. Variance described by the j^{th} PC is:

$$\textit{The variance} = (\lambda i \text{ - } \Sigma\lambda n) \tag{3}$$

The PC related to the dominant eigenvalue, the primary PC (PC1), represents a linear correlation of the variables accounting for the largest part of the total data variance. The PC2 describes maximum flexibility that PC1 can report then on. All items with eigenvalues >1 should be stored.

2.5 Principal component regressions (PCR)

PCR is a regression technique with an application of PCA. The regression model produces the response, including PCA and MLR, in determining unknown coefficients in the model. The success of PCR depends on the selection of the critical elements used for regression; PCR can predict the outcome well on model assumptions. Now this modified data is employed as input to multiple rows of recovery processes.

$$Y = \alpha_0 + \alpha_1(PC1) + \alpha_1(PC2) + ... + \alpha_n(PCn) + e \tag{4}$$

where $\alpha_0, \alpha_1, \alpha_2$ ...... α_n are coefficients of the regression model as estimated by the least-squares method, and 'e' is a random error. PCs were calculated from the data and are used as inputs to build the model with the PCR technique. The covariance matrix for a given input is decided. PCs were determined on the idea of variance defined by the eigenvalues of the covariance matrix. The principal components, whose variance of ≥ 1 or ~ 60% was explained, supported the analysis of the eight variables utilized in the prediction.

2.6 PC based ANN Model

The multilayer perceptron (MPL) with back-propagation (BP) was used for the specification with a single input, hidden, and output layers. The hidden layer is improved by performing several tests with different architectures, since this yields low values (MSE) for small intervals (convergence). Figure 3 is the architecture of the developed PCANN model.

The training data were optimized as the input set for the model/network to get good prediction results. This flexibility enables ANN to receive new features for compatible entities. Various training functions are available to perform the loop calculations. Most ANN models are characterized by functionality, activation function efficiency (selection), topology, and the learning rule. A three-layer NN with a logarithmic sigmoid function (logsig) within the hierarchy and a linear function is employed within the output layer in the present work. The magnitude of the error confirmed the performance of the developed ANN models.

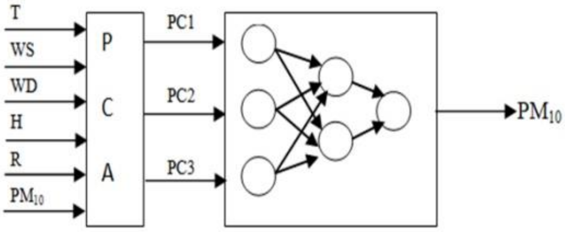

Figure 3. PC based ANN Model

2.6 Evaluation of model performance

In performance evaluation of the built model the correctness of the responses was assessed. The following performance indicators were used for the analysis. The statistical properties of each of these mathematical indices are as follows:

2.6.1 Mean absolute error (MAE)

It tests the scale of the error rate in the predictor data, irrespective of the direction. MAE is a clear translation as the absolute difference between *Yi* and *Xi*

$$MAE = \frac{\sum_{i=1}^{n}\left|Y_i - X_i\right|}{n} \tag{5}$$

2.6.2 Mean squared error (MSE)

Error implying a measure to estimate how close the register line is to the set of points. It does this by measuring the distances from points to the reference line (called "errors") and minimizing them. The squaring of error is needed to get rid of any negative values.

$$MSE = \frac{\sum_{i=1}^{n}\left(Y_i - X_i\right)^2}{n} \tag{6}$$

2.6.3 Mean absolute percentage error (MAPE)

A mathematical measure of how accurate the process is. Accuracy measures are calculated in percentage and compute the mean absolute error percent of each time in real values divided by actual values and given by:

$$MAPE = \frac{100}{n} * \sum_{i=1}^{n}\left|\frac{X_i - Y_i}{X_i}\right| \tag{7}$$

2.6.4 Root mean square error (RMSE)

RMSE is a standard deviation of residuals (error of prediction), and the degree of how far from line data is also known as the distribution of residuals. Usually, it tells us how strong the data is around the line for the fit.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left|Y_i - X_i\right|^2}{n}} \tag{8}$$

2.6.5 Correlation coefficient (R)

It measures how strong the relationship is between the two variables. R (Pearson's) is a global correlation that is used for linear regression.

$$R = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y_i}\right)\left(X_i - \overline{X_i}\right)}{\left\{\left[\sum_{i=1}^{n}\left(X_i - \overline{X_i}\right)^2\right]\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y_i}\right)^2\right]\right\}^{1/2}} \tag{9}$$

Here 'n' is a total count of data records, Y_i are the data forecasted, and X_i are the original (predictor) data. A zero error would indicate that all of the output data were identified by the model correctly. The accuracy of the model was evaluated with 90% confidence level.

3. Results and Discussion

3.1 Principal component regressions (PCR)

The covariance matrix, known as the loading matrix, describes the amount of variance for every component shown in Table 2. PCs with an eigenvalue >1 were considered to analyse 60% of the variance described in the PC matrix covariance. The total variance calculated for the first three PCs is nearly 60%, of which WS, R, H, and PM10 are listed; PC1 accounts for 31.8% of the variance in the data. T and WD are grouped into the PC2, which is explained at 14.3%, and PC3 defines SO2 and NO2 by 12.3% of the data variance.

The predictive model of PCR using three principal components was identified using MINITAB17 statistical software. The primary goal of this exercise was to extract the essential information (PCs) from the data table and to

Table 2. Principal component analysis

Component	T	WS	WD	R	H	PM_{10}	SO_2	NO_2	Eigenvalue	Cumulative
PC1	-0.331	0.441	0.322	0.38	0.531	-0.378	-0.042	-0.149	2.543	0.318
PC2	-0.69	0.255	-0.629	-0.128	-0.045	0.175	0.115	-0.041	1.148	0.461
PC3	-0.098	0.081	-0.011	0.026	0.053	0.149	-0.819	0.537	1.005	0.587
PC4	-0.001	-0.022	-0.008	0.291	0.126	0.072	0.538	0.777	0.988	0.711
PC5	0.004	0.458	0.45	-0.527	0.12	0.521	0.151	0.042	0.842	0.816
PC6	0.026	-0.079	0.017	0.631	0.047	0.716	-0.043	-0.282	0.662	0.899
PC7	-0.469	-0.719	0.271	-0.218	0.364	0.096	0.009	-0.011	0.4701	0.957
PC8	0.432	-0.018	-0.474	-0.179	0.741	0.081	-0.023	-0.049	0.3417	1

rearrange the description of the data set. MLR method gave

$$PM10 = 45.62 + 1.191 * vPC1 - 0.638 * vPC2 + 1.164 * vPC3 \tag{10}$$

The regression coefficients are statistically significant ($P<0.05$), and the residuals' distribution is normal, indicating a good fit by the model equation. The value of R for the performance of PCR was found to be 78.90%.

3.2 PCA based artificial neural network (ANN)

The hybrid ANN model was used to forecast PM10 using PCs as input variables. The model was trained and validated for PM10 concentration as output with metrological variables as inputs, using logarithmic sigmoid (logsig) activation functions. The PCs were used as input data to the developed model; the first three PCs were selected in place of the original eight monitored parameters.

3.3 PC based ANN model performance

Various measures were used to evaluate the effectiveness of the hidden layer, activation function, and rate of learning by the selected algorithm. Existing research uses a neural network of 5 - 25 neurons as a hidden range for the optimization process. Ten neurons made excellent forecasts using this method. Networks with two or more hidden layers do not correlate well with the estimated contaminant levels (Chellali *et al*., 2016). The network was trained with defined data set, and the model was evaluated. Table 3 reviews the assessment of the model, done in the training and the testing data, with varied numbers of neurons in the hidden layer. As soon as a positive ANN construct is established, the apparent PM10 concentrations are compared with the predicted estimates. Model accuracy was evaluated using R, achieving 0.886, 0.8959, and 0.8884 for training, testing, and validation of the data set. MAPE values were 0.072, 0.076, 0.094, and 0.080 in experiments changing the neurons in the hidden layer to 10, 15, 20, and 25, respectively. Ten neurons achieved acceptable MAPE with 0.072 for training and 0.075 for testing data. The performance of the structures was measured on all three levels; the model's performance is revealed in Table 3. The optimized performance of estimated PM10 by the hybrid model for the prediction is shown as training, testing and validation of ANN in Figure 4. The network composition for the prediction of PM10 and the efficiency of the proposed PCANN technique is shown in Table 4. The following observations can be made for the optimized network design. With an increased neuron count, the network produced several outputs with different MSE for the training set. Further increasing the neurons (>10) caused infeasible results and MSEs started increasing during validation. With the choice of 10 neurons, the training was very fast in 10 to 12 Epochs. Having less neurons will cause underfitting, and conversely, an increase in neurons will cause overfitting as well as prolonged training. Table 4 shows that the MAE grew from 4.882 to 6.248 as the neuron count was increased from ten to twenty. Decent results were found between predicted and observed PM10 and are credited due to the robust effect of the particular entries. It was observed that the model did not reasonably predict the minimum and maximum concentrations of PM10 due to unexpected changes in the local environment (rain during summer and sudden increase in the use of diesel generators) near the monitoring site.

The results are similar to prior research by various investigators. The ANN model has been applied to predict PM10 concentrations in Hammastation in Algiers using climate change. Model estimates were satisfactory, with R values ranging from 0.77 to 0.92 (Chellali *et al*., 2016). The ANN model was demonstrated for the city of Varanasi, India. The models used to predict daily PM10 concentrations had a predictive accuracy of 9.86% in terms of MAPE (Yadav & Nath, 2019). The novel method for estimating air pollution concentrations of PM10 using multi-linear regression (MLR) and ANN was proposed in the Eastern region of peninsular Malaysia. R^2, RMSE and MSE were evaluated for the models

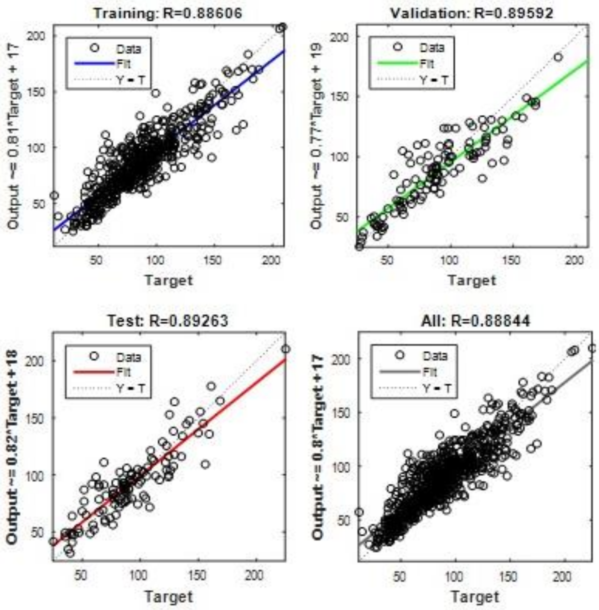

Figure 4. Training, testing and validation of PM_{10} ($\mu g/m^3$)

Table 3. Performance evaluation of PM_{10} PCANN model

Nodes	10	15	20	25	Nodes	10	15	20	25
Indicator	Training results				Indicator	Testing results			
MAE	4.882	5.363	6.248	5.615	MAE	5.459	5.614	5.606	5.442
MSE	39.51	48.74	63.44	53.3	MSE	50.38	53.74	53.88	50.83
RMSE	6.286	6.981	7.965	7.301	RMSE	7.098	7.331	7.340	7.129
MAPE	0.072	0.076	0.094	0.08	MAPE	0.075	0.076	0.076	0.074

to compare the techniques. The results show that the ANN model outperformed the MLR by 0.25 to 12.25 in terms of RMSE (Yusof, Azid, Sani, Samsudin, Amin, Rani, & Jamalani, 2019). Therefore, a PCANN method with transformed neurons with appropriate activation function was adequate for predicting PM10 concentrations in Belagavi.

3.4 Comparison of PCANN and PCR

The performance indicators for PCR and PCANN are shown in Table 5, where the neural networks gave better predictions as shown in Figure 5. The PCANN technique is one of the best for predicting PM10. The overall performance of the developed model using an ANN was adequate or reasonable and could be considered for operational use. The general prophetic competence of the PM10 model was about 90% in terms of R.

4. Conclusions

The main objective of this study was to improve the performance in predicting PM10 concentrations by applying PCs as input variables to an ANN. The method appears to be one of the best, after optimizing the size of hidden layer in the ANN architecture. The use of PCA in ANN model construction proved effective as it addresses collinearity problems in MLR. It also enabled the reduction of the number of translator predictions. The hybrid PCANN model performed well and can predict PM10 concentrations across the year. The PM10 prediction results for PCANN models are much more precise than the results of PCR. It is concluded that ANN models can perform better when PCA is applied for data reduction. The method used here can be fully applied to further environmental and engineering-related problems dealing with nonlinear relationships. Finally, concerned authorities can use air quality predictions to disseminate relevant information to the general public, to protect their health and take critical preventative measures during extreme air pollution and winter.

Acknowledgements

The authors is grateful to the Principal and Management of KLE Dr. M. S. Sheshgiri college of Engineering and Technology, Belagavi, for providing the facilities to carry out this work.

References

Abdullah, S., Ismail, M., Fong, S. Y., & Ahmed, N. (2016). Evaluation for long term PM 10 Concentration Forecasting using Multi Linear Regression (MLR) and Principal Component Regression (PCR) Models. *EnvironmentAsia*, *9*(2). Retrieved from https://www.thaiscience.info/Journals/Article/ENVA/10982852.pdf

Akkoyunlu, A., Yetilmezsoy, K., Erturk, F., & Oztemel, E. (2010). A neural network-based approach for the prediction of urban SO2 concentrations in the Istanbul metropolitan area. *International Journal of Environment and Pollution*, *40*(4), 301-321. Retrieved from https://doi.org/10.1504/IJEP.2010.031752

Baawain, M. S., & Al-Serihi, A. S. (2014). Systematic approach for the prediction of ground-level air pollution (around an industrial port) using an artificial neural network. *Aerosol and Air Quality Research*, *14*(1), 124-134. Retrieved from https://aaqr.org/articles/aaqr-13-06-oa-0191.pdf

Table 5. Performance of PCANN and PCR model

Indicator	PCANN	Indicator	PCR
MAD	5.107	MAD	6.750
MSE	44.872	MSE	102.120
RMSE	6.633	RMSE	10.105
MAPE	0.069	MAPE	0.088

Table 4. Optimized PCANN model performance for PM_{10}

Input data for PM_{10}	R^2				
	Network*	Training	Testing	Validation	Function
3PCs (8 input parameters)	3-10-1-1	88.60%	89.59%	88.84%	logsig
RMSE	3-10-1-1	6.286	7.098	8.061	logsig

*Network (a-b-c-d) - number of (inputs-hidden nodes-hidden layer-output layer)

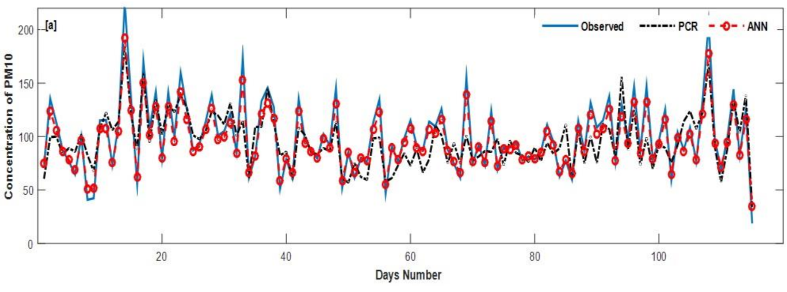

Figure 5. Performance comparison of PCR and PC based ANN model

Bhaskar, B. V., & Mehta, V. M. (2010). Atmospheric particulate pollutants and their relationship with meteorology in Ahmedabad. *Aerosol and Air Quality Research*, *10*(4), 301-315. Retrieved from https://aaqr.org/articles/aaqr-09-10-oa-0069.pdf

Census (2011). Retrieved from http://www.census2011.co.in/census/city/430- belgaum. html.

Chelani, A. B., Gajghate, D. G., & Hasan, M. Z. (2002). Prediction of ambient PM10 and toxic metals using artificial neural networks. *Journal of the Air and Waste Management Association*, *52*(7), 805-810. Retrieved from https://www.tandfonline.com/doi/pdf/10.1080 /10473289.2002.10470827

Chellali, M. R., Abderrahim, H., Hamou, A., Nebatti, A., & Janovec, J. (2016). Artificial neural network models for prediction of daily fine particulate matter concentrations in Algiers. *Environmental Science and Pollution Research*, *23*(14),14008-14017. Retrieved from https://doi.org/10.1007/s11356-016-6565-9

Cocchi, D., Greco, F., & Trivisano, C. (2007). Hierarchical space-time modelling of PM10 pollution. *Atmospheric Environment*, *41*(3), 532-542. Retrieved from https://doi.org/10.1016/j.atmosenv.2006.08.032

CPCB (2009), Guidelines for ambient air quality monitoring guidelines for ambient air quality monitoring. Retrieved from http://cpcb.nic.in › openpdffile

Dadvand, P., Parker, J., Bell, M. L., Bonzini, M., Brauer, M., Darrow, L. A., & Woodruff, T. J. (2013). Maternal exposure to particulate air pollution and term birth weight: a multi-country evaluation of effect and heterogeneity. *Environmental Health Perspectives*, *121*(3), 267-373. Retrieved from https://ehp.niehs.nih.gov/doi/full/10.1289/ehp.1205575

Ding, W., Zhang, J., & Leung, Y. (2016). Prediction of air pollutant concentration based on sparse response back-propagation training feedforward neural networks. *Environmental Science and Pollution Research*, *23*(19), 19481-19494. Retrieved from https://doi.org/10.1007/s11356-016-7149-4

de Mattos Neto, P. S., Cavalcanti, G. D., Madeiro, F., & Ferreira, T. A. (2015). An approach to improve the performance of PM forecasters. *PloS one*, *10*(9), e0138507. Retrieved from https://journals.plos.org/plosone/article?id=10.1371 /journa.pone.0138507

de Souza, J. B., Reisen, V. A., Franco, G. C., Ispány, M., Bondon, P., & Santos, J. M. (2018). Generalized additive models with principal component analysis: an application to time series of respiratory disease and air pollution data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *67*(2), 453-480. Retrieved from https://doi.org/10.1016/j.scitotenv.2018.04.273

Duan, R. R., Hao, K., & Yang, T. (2020). Air pollution and chronic obstructive pulmonary disease. *Chronic Diseases and Translational Medicine*, *6*(04), 260-269. Retrieved from https://doi.org/10.1016/j.cdtm.2020.05.004

European Environment Agency (2013). Air quality in Europe-2013, Report No 9/2013. Retrieved from https://www.eea.europa.eu › publications › download

Hosamane, S. N., Prashanth, K. S., & Virupakshi, A. S. (2020). Assessment and prediction of PM10 concentration using ARIMA. *Journal of Physics: Conference Series*, *1706*(1), 012132. Retrieved from https://iopscience.iop.org/article/10.1088/1742-6596/1706/1/012132/pdf

Kolehmainen, M., Martikainen, H., & Ruuskanen, J. (2001). Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment*, *35*(5), 815-825. Retrieved from https://doi.org/10.1016/S1352-2310(00)00385-X

Kumar, A., & Goyal, P. (2011). Forecasting of air quality in Delhi using principal component regression technique. *Atmospheric Pollution Research*, *2*(4), 436-444. Retrieved from https://www.sciencedirect.com/science/article/pii/S 1309104215304700

Mishra, D., Goyal, P., & Upadhyay, A. (2015). Artificial intelligence based approach to forecast PM2. 5 during haze episodes: A case study of Delhi, India. *Atmospheric Environment*, *102*, 239-248. Retrieved from https://doi.org/10.1016/j.atmosenv.2014.11.050

Nagendra, S. S., & Khare, M. (2003). Principal component analysis of urban traffic characteristics and meteorological data. *Transportation Research Part D: Transport and Environment*, *8*(4), 285-297. Retrieved from https://doi.org/10.1016/S1361-9209(03)00006-3

Ng, K. Y., & Awang, N. (2018). Multiple linear regression and regression with time series error models in forecasting PM10 concentrations in Peninsular Malaysia. *Environmental Monitoring and Assessment*, *190*(2), 1-11. Retrieved from https://doi.org/10.1007/s10661-017-6419-z

Park, S., Kim, M., Kim, M., Namgung, H. G., Kim, K. T., Cho, K. H., & Kwon, S. B. (2018). Predicting PM10 concentration in Seoul metropolitan subway stations using artificial neural network (ANN). *Journal of Hazardous Materials*, *341*, 75-82. Retrieved from https://doi.org/10.1016/j.jhazmat.2017.07.050

Pires, J. C. M., Martins, F. G., Sousa, S. I. V., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2008). Selection and validation of parameters in multiple linear and principal component regressions. *Environmental Modelling and Software*, *23*(1), 50-55. Retrieved from https://doi.org/10.1016/j.envsoft.2007.04.012

Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling and Software*, *22*(1), 97-103. Retrieved from https://doi.org/10.1016/j.envsoft. 2005. 12.002

Srimuruganandam, B., & Nagendra, S. S. (2015). ANN-based PM prediction model for assessing the temporal variability of PM10, PM2. 5 and PM1 concentrations at an urban roadway. *International Journal of Environmental Engineering*, *7*(1), 60-89. Retrieved from https://doi.org/10.1504/IJEE.2015.069266

Tran, H., Kim, J., Kim, D., Choi, M., & Choi, M. (2018). Impact of air pollution on cause-specific mortality in Korea: Results from Bayesian Model Averaging and Principle Component Regression approaches. *Science of The Total Environment*, *636*, 1020-1031. Retrieved from https://doi.org/10.1016/jscitotenv.2018.04.273

WHO (2016), Urban Ambient Air Pollution database – Update (2016), Retrieved from http://www.who.int/airpollution/data/AAP_database_summary_results_2016_v02.pdf

Wolff, G. T., Korsog, P. E., Kelly, N. A., & Ferman, M. A. (1985). Relationships between fine particulate species, gaseous pollutants and meteorological parameters in Detroit. *Atmospheric Environment (1967)*, *19*(8), 1341-1349. Retrieved from https://doi.org/10.1016/0004-6981(85)90264-1

Yadav, V., & Nath, S. (2019). Novel hybrid model for daily prediction of PM 10 using principal component analysis and artificial neural network. *International Journal of Environmental Science and Technology*, *16*(6), 2839-2848. Retrieved from https://doi.org/10.1007/s13762-018-1999-x

Yusof, K. K., Azid, A., Sani, M. S. A., Samsudin, M. S., Amin, S. N. S. M., Rani, N., & Jamalani, M. A. (2019). The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models over particulate matter (PM10) variability during haze and non-haze episodes: A decade case study. Retrieved from https://doi.org/10.11113/mjfas.v15n2.1004