

Article

# Application and comparison of MLR, ANN and CART models for predicting PM<sub>10</sub> concentration level of Guwahati city (India)

Abhishek Dutta<sup>1</sup>, Wanida Jinsart<sup>1\*</sup>

<sup>1</sup> Department of Environmental Science, Faculty of Science, Chulalongkorn University, 254 Phayathai Road, Pathumwan, Bangkok 10330, Thailand; duttabob@gmail.com(A.D.); jwanida2013@gmail.com (W.J.)

**Abstract:** Indian cities are increasingly becoming susceptible to PM<sub>10</sub> induced health effects which have become a matter of concern for the policymakers of the country. Air pollution is engulfing the comparatively smaller cities as the rapid pace of urbanization, and economic development seems not to lose steam ever. A review of air pollution of 28 cities of India, which includes tier-I, II, and III cities of India, found to have grossly violated both WHO and NAAQS standards in respect of acceptable daily average PM<sub>10</sub> concentrations by a wide margin. Predicting the city level PM<sub>10</sub> concentrations in advance and accordingly initiate prior actions is an acceptable solution to save the city dwellers from PM<sub>10</sub> induced health hazards. Predictive ability of three models, linear MLR, nonlinear MLP (ANN), and nonlinear CART, for one day ahead PM<sub>10</sub> concentration forecasting of tier-II Guwahati city, were tested with 2016-2018 daily average observed climate data, PM<sub>10</sub>, and gaseous pollutants. The results show that the non-linear algorithm MLP with feedforward backpropagation network topologies of ANN class, giving the best prediction value when compared with linear MLR and nonlinear CART model. ANN (MLP) approach, therefore, may be useful to effectively derive a predictive understanding of one day ahead PM<sub>10</sub> concentration level and thus provide a tool to the policymakers for improving decision-making associated with air pollution and public health.

**Keywords:** Particulate Matter; Prediction; Model comparison; Artificial neural network; Multi-variate linear regression; Small city

---

## 1. Introduction

Over the past years, airborne particulate matter (PM) concentrations in Indian cities have been rising and have become a matter of concern for the policymakers in India. The Central Pollution Control Board (CPCB) of India now included forty-one comparatively smaller tier-II cities in the list of severely air-polluted cities [1]. The World Health Organization (WHO) has been very critical of increasing air pollution in Indian cities, in general, and its adverse health effects on urban populations [2]. The total pollution mortality counts for India are the highest in the world [3]. Apte et al. [4] estimated that the level of air pollution in India would require to be declined by 20% by 2030 to maintain per-person mortality at the 2010 level. Several studies in the Indian context and elsewhere established the relationship between high PM<sub>10</sub> (particulate matter having an effective aerodynamic diameter smaller than 10  $\mu\text{m}$ ) levels with an increase in hospital admissions for lung and heart diseases [5,6,7,8]. The sources of air pollution for the Indian cities are varied. They include vehicle exhaust, small scale as well as heavy industries, power generation, brick manufacturing kilns, re-suspended dust on the roads due to vehicle movement, construction-related activities, open waste burning, combustion of various fuels, and in-situ power generation using diesel generator sets [9].

---

However, transport remains the primary emission source more or less common for big as well as smaller cities [10, 11]. For example, in the case of Bengaluru, vehicle exhaust and on-road dust resuspension together account for about 56% and 70% of total  $PM_{2.5}$  (particulate matter with diameters  $2.5 \mu m$  and smaller) and  $PM_{10}$  emissions respectively [12]. Different meteorological parameters like atmospheric pressure, wind velocity, rainfall, and relative humidity also play an important role in the  $PM_{10}$  accumulation, and hence they can govern the variability of atmospheric  $PM_{10}$  level [13, 14, 15].

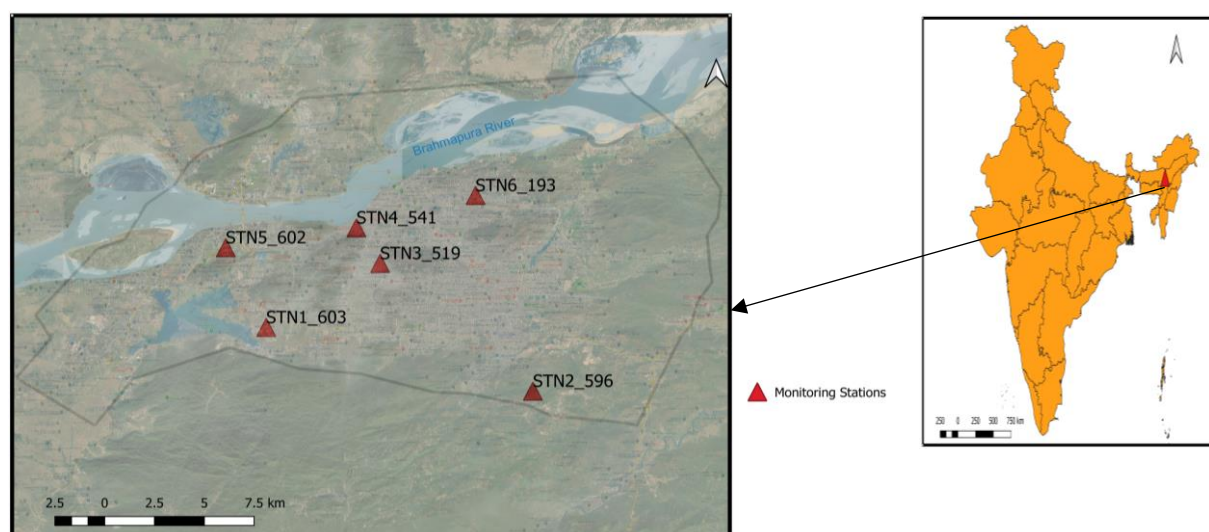
One option to the Indian policymakers to mitigate critical PM concentrations in the cities, vis a vis health effects, therefore, may be to correctly predict the concentrations at least one to two days in advance and accordingly initiate prior actions such as regulation of traffic in a planned way. However, predicting the air quality is not so straight forward job because of the complex interactions of different nonlinear parameters [16]. Shahraiyni and Sodoudi [17] surveyed 36 research studies, executed in different cities of the world, in the quest of achieving prediction accuracy in forecasting  $PM_{10}$ . In these studies, 50% of researchers employed a multi-layer perceptron (MLP) with Feedforward Backpropagation Network (FFBN) topologies, a class of Artificial Neural Network (ANN) model. Around 28% (10 studies) depended on the widely used Multiple Linear Regression (MLR) technique for  $PM_{10}$  forecasting of urban areas, in particular. Three studies (about 8%) used the Radial Basis Function (RBF) network of ANN class in forecasting city level  $PM_{10}$ . The other five studies (14%) depended on different other techniques like PNN (Pruned Neural Networks), LL (Lazy Learning), MLP and MLR combo, Elman class of Recurrent Neural Networks (RNN), and PCRA (Principal Component Regression Analysis). ANN technique appears to be providing useful results to deal with nonlinear independent variables involved in environmental pollution prediction. Hence more practitioners are resorting to ANN modeling type of data-driven approaches as alternatives to traditional deterministic or nonlinear models [18, 19]. Pollution researchers of China and elsewhere have used ANN techniques extensively to forecast airborne PM concentrations in the past. The use of MLR with stepwise inclusion of input variable has been the most commonly used tool for temporal prediction of  $PM_{2.5}$  and  $PM_{10}$  in different urban areas of India. MLR has its limitation in terms of linear representation of non-linear systems. However, researchers have, in a limited way only, showed a preference for different data-driven predictive techniques for PM forecasting in the Indian context and comparatively judge their performances (Table 1). Goyal et al. [20] forecasted respirable suspended particulate matter (RSPM) of Delhi and Hong Kong with three models (MLR, Autoregressive Integrated Moving Average or ARIMA, and a combo of MLR and ARIMA) and reported better result for the combo of MLR and ARIMA. Nagendra and Khare [21] used the ANN technique quite successfully while predicting nitrogen dioxide ( $NO_2$ ) dispersion phenomena in Delhi. Tikhe Shruti et al. [22] compared the prediction performance of air pollutants by using ANN and Genetic Programming (GP) for highly polluted Pune city of India and found GP algorithms to be superior predictor than ANN. Chelani et al. [23] used the MLP class of ANN to successfully predict the concentration of  $PM_{10}$  and toxic metals for Jaipur city of India. Kumari et al. [24] used ANN modeling for the prediction of sulphur dioxide ( $SO_2$ ) and  $PM_{10}$  concentrations in the ambient air of a south Indian small city of Vijayawada. Kaur and Mandal [25] forecasted  $PM_{2.5}$  concentration for hotspots of Delhi by employing various ANN structures FFBN, RNN, Elman, and Nonlinear Autoregressive Network with Exogenous Input (NARX).

Against the above background, in this paper, we first made a review of the literature on air pollution to assess the level of  $PM_{10}$  concentration in the Indian cities to understand the gravity of the problem across Indian cities clearly. A comprehensive and comparative understanding of the  $PM_{10}$  concentration level across the cities of the country is still lagging. Secondly, the sampling site is a representation of a tier-II city of India, where air pollution-induced health hazards are increasing at a faster rate. We have analyzed three years of  $PM_{10}$  concentration of an Indian tier-II city, Guwahati to understand its correlation with meteorological variables. Rapid urbanization and their contribution to air pollution have made smaller Indian cities like Guwahati vulnerable too. A recent study, conducted in Guwahati, computed Hazard Quotient (HQ) based on Indian National Ambient Air Quality Standards (NAAQS) and WHO city-level air quality standards and indicated quite a high

degree of health risk for the city dwellers [26]. Third, we tested the predictive ability of three models, based on linear MLR, nonlinear ANN (MLP), and nonlinear CART (Classification and Regression Trees) analyses, for one day ahead  $PM_{10}$  concentration forecasting of Guwahati city. These models were critically assessed through a comparative evaluation of performance indicators keeping in mind the end goal to choose the best-fitted model for accurate forecasting  $PM_{10}$  at the city level. Though a higher level of mortality burden associated with air pollution exposure in Indian tier-II cities has been reported, not many studies have been performed on  $PM_{10}$  forecasting to date. It will be advantageous for the policymakers to take policy decision if a snapshot of the  $PM_{10}$  concentration level is known one day ahead with precision for a city like Guwahati through the use of the optimum of the statistical approaches readily available for use. This study is novel due to the fact that the testing of the predictive ability of three models has been carried out in the backdrop of tier-II cities of India. Unlike previous modeling efforts (Table 1), this is the first instance concerning the application of CART analysis as a statistical procedure for prediction of  $PM_{10}$  in a comparative set up of Indian city. In the recent past, Gocheva-Ilieva and Stoimenova [27] employed CART in predicting  $PM_{10}$  for the Pleven city of Bulgaria and claimed of very accurate model performance. The CART technique is evolving as a new method for analysis and forecasting of  $PM_{10}$  very fast [28].

## 2. Location of the study

The model development for forecasting of  $PM_{10}$  was attempted in the north-eastern Indian tier-II city of Guwahati, which is the capital city of the state of Assam, India. The city with about one million population has 340 km<sup>2</sup> urban space contiguous with a large river (the Brahmaputra) flowing through the city. It has a scenic landscape with the Brahmaputra River flowing on one side with the foothills of the Shillong plateau on the other side. Vehicular growth (both light and heavy vehicles) in the city was notable in the past decade with about a reported sharp rise of 87%. Guwahati has a humid subtropical climate. The four major seasons of the city are winter (December to February), spring (March to May), summer (June to August), and autumn (September to November), with the differing meteorological condition. Guwahati has been recognized as one of India's most rapidly growing cities in India for the last 10-12 years. There is black carbon pollution in the city air due to the rapid urbanization and poor environmental quality control [29]. Guwahati has six ambient air monitoring stations, set up under the National Air Quality Monitoring Programme (NAMP), to measure key pollutants [30]. One of the NAMP stations can measure  $PM_{2.5}$  while the newly developed CAAQM (Continuous Ambient Air Quality Monitoring) station started functioning only during mid of 2019. The locations covered for collecting data are shown in Table 2 and Figure 1 below.



**Figure 1.** Study location and monitoring stations.

**Table 1.** Different data-driven predictive techniques used for PM forecasting in the Indian context

Location (Type)	Method*	Predictor variables**	Target**	Remarks	Ref.
Delhi (Tier-I city)	Wavelet and RNN	CO, NO <sub>2</sub> , NO, O <sub>3</sub> , SO <sub>2</sub> , PM <sub>2.5</sub>	CO, O <sub>3</sub> , NO <sub>2</sub> , NO, SO <sub>2</sub> and PM <sub>2.5</sub>	Prediction performance was reasonably good.	[31]
Lucknow (Tier-II city)	PLSR, MPR, and ANN	T, RH, WS, SPM, NO <sub>2</sub> , SO <sub>2</sub>	RSPM, SO <sub>2</sub> and NO <sub>2</sub>	MPR and ANN performed better	[32]
Lucknow (Tier-II city)	SDT, DTF, and DTB vs. SVM	Air quality & meteorological parameter	AQI and Combined AQI	DTF and DTB outperformed the SVM	[33]
Mumbai (Tier-I city)	ANN and Kriging combination	T,RH,WS,WD, AP, NO <sub>x</sub> , SO <sub>x</sub> ,RSPM	NO <sub>x</sub> , SO <sub>x</sub> , and RSPM	ANN and Kriging performed satisfactorily	[34]
Delhi (Tier-I city)	NF compared with MLR, and ANN.	CO, O <sub>3</sub> , NO <sub>2</sub> , SO <sub>2</sub> , PM <sub>2.5</sub> , AP, T, WS, WD, RH, V, DP	PM <sub>2.5</sub>	NF model is better than ANN and MLR models.	[35]
Rourkela (Tier -II city)	WMLP, WRNN, MLP (FFBN), and RNN	T, RH, BLH, SP, WD , WS	PM <sub>2.5</sub> , PM <sub>10</sub>	WMLPNN model performed better	[36]
Varanasi (Tier-II city)	PCA, ANN (MLP) and MLR	PM <sub>2.5</sub> , NO, Benzene & VWS for PCA, ANN. SR, WS & AP for MLR	PM <sub>10</sub>	Hybrid PCA-ANN model gives a better prediction	[37]
Delhi (Tier-I city)	SVM and ANN	PM <sub>2.5</sub> , SO <sub>2</sub> , CO, NO, NO <sub>x</sub> , C <sub>7</sub> H <sub>8</sub> , NO <sub>2</sub> , VWS, WS, WD, T, RH, SR	PM <sub>2.5</sub>	ANN exhibited a better result	[38]
Delhi (Tier-I city)	ANN	Meteorological variables	PM <sub>10</sub> ,PM <sub>2.5</sub> , NO <sub>2</sub> , and O <sub>3</sub>	O <sub>3</sub> predictions are better than PM.	[39]
Chennai (Tier- I city)	ARMA/ARIMA modelling	PM <sub>10</sub> , SO <sub>2</sub> , NO <sub>2</sub>	PM <sub>10</sub> , SO <sub>2</sub> , and NO <sub>2</sub>	Forecasting efficiency good.	[40]
Delhi (Tier-I city)	Four types of ANN (FFBP, RNN, Elman, and NARX)	PM <sub>2.5</sub> , T, WS, WD, RH, SR.	PM <sub>2.5</sub>	NARX model outperforms others	[25]
Delhi (Tier-I city)	ANFIS, WANFIS, WANFIS-GA, WANFIS-PSO	PM <sub>2.5</sub>	PM <sub>2.5</sub>	WANFIS-PSO performed better	[41]
Guwahati (Tier-II city)	MLR, ANN (MLP) and CART	PM <sub>10</sub> , RF, T, RH, WS, NO <sub>2</sub> , SO <sub>2</sub> , CO	PM <sub>10</sub>	ANN(MLP) performs better	This study

\* Abbreviations of methods: PLSR: Partial least squares regression; MPR: Multivariate Polynomial Regression; SDT: Single Decision Tree; DTF: Decision Tree Forest; DTB: Decision Tree Boost; SVM: Support Vector Machine; NF: Artificial Intelligence-based Neuro-Fuzzy techniques; WMLP: Wavelet-based MLP; WRNN: Wavelet-based RNN; ARMA: Autoregressive Moving Average; NARX : Non-linear Autoregressive Network with Exogenous Input; ANFIS: Adaptive-Neuro Fuzzy Inference System; WANFIS: Wavelet ANFIS; WANFIS-GA: WANFIS Genetic Algorithm; WANFIS-PSO:

WANFIS Particle Swarm Optimization. \*\*Abbreviation of predictor variables and targets: NO :Nitrogen oxide; O<sub>3</sub>: Ozone; T: Temperature; RH: Relative Humidity; WS: Wind Speed; SPM: Suspended Particulate Matter; RSPM: Respirable Suspended Particulate Matter; WD: Wind Direction; AP: Atmospheric Pressure; NO<sub>x</sub>: Nitrogen Oxides; Sulphur oxides ; SO<sub>x</sub>; V: Visibility; DP: Dew Point; BLH: Boundary Layer Height; SP: Surface Pressure; SR: Solar Radiation; VWS: Vertical Wind Speed; AQI: Air Quality Index

**Table 2.** Monitoring stations and their UTM coordinates.

Code	Location	Monitoring Type	Latitude	Longitude	Area Type
STN1_603	Boragaon, IASST Campus	Non PM <sub>2.5</sub>	26.11635	91.68338	Residential
STN2_596	Khanapara, Central Diary	Non PM <sub>2.5</sub>	26.0831	91.8171	Residential
STN3_519	Gopinath Nagar, ITI Building	Non PM <sub>2.5</sub>	26.160962	91.752542	Residential
STN4_541	Santipur, Prajyotish College	Non PM <sub>2.5</sub>	26.165391	91.7276	Residential
STN5_602	Guwahati University campus	Non PM <sub>2.5</sub>	26.15793	91.66312	Residential
STN6_193	Bamunimaidan, PCBA HQ	PM <sub>2.5</sub>	26.185165	91.788334	Residential

### 3. Methods

Daily average concentration data (1096 data points) for PM<sub>10</sub> (μg m<sup>-3</sup>), CO (ppm), NO<sub>2</sub> (ppb), and SO<sub>2</sub> (ppb) were collected in respect of all the six air quality monitoring stations for three years 2016-2018 from State Pollution Control Board (SPCB) office located at Guwahati. The three years (2016-2018) daily climate data (1096 data points) for ambient temperature (AT, °C), relative humidity (RH, %), wind speed (WS, ms<sup>-1</sup>), rainfall (RF, millimeter) was acquired from Regional Meteorological Department, located at Guwahati.

#### 3.1. Data Treatment

A few missing values were observed in respect of daily average concentration data for PM<sub>10</sub>, CO, NO<sub>2</sub>, and SO<sub>2</sub> for the 2016-2018 time-series data. Linear interpolation technique was used in the imputation of missing values.

$$f(x)=f(x_0)+\frac{f(x_1)-f(x_0)}{x_1-x_0} \times (x-x_0) \quad (1)$$

where x = independent value,  $x_1$  and  $x_0$  = known values of the independent variable and f(x) = value of the dependent variable for the value of the independent variable. However, there was no missing value for climate data (1096 data points).

#### 3.2. Assessment of PM<sub>10</sub> pollution across Indian cities

Approximately 120 English language articles published between 2011-2019 on monitoring and control of ambient particulate matter (PM<sub>10</sub> and PM<sub>2.5</sub>) related to Indian cities have been reviewed. We included the journal papers from computer searches and bibliographic databases (e.g., Google Scholar, Pub Med, Academia, ResearchGate) in the analysis. We used the keywords such as PM<sub>10</sub>, PM<sub>2.5</sub>, air pollution, and Indian city to search for the relevant literature for the purpose in hand. The main aim of the review was to assess the PM<sub>10</sub> status in various cities of India to understand the severity of the problem of air pollution across Indian cities. The problem severity automatically would establish the need for correctly predicting PM<sub>10</sub> concentrations at least one or two days in advance to mitigate ill effects through minimizing the exposure.



### 3.3. Descriptive statistics and analysis of time series

Descriptive statistics of the climate data, PM<sub>10</sub>, and gaseous pollutants for the period 2016-2018 (1096 data points) and time series analysis were also worked out in respect of air quality monitoring station 6 to understand the characteristics and correlation of different variables over the period of the study. Station 6 was found to be a representative one out of six air quality monitoring stations of the city due to reasons like the completeness of data sets and common reflection of land-use patterns of the city. Multiple time series charts were produced with time on the horizontal axis and PM<sub>10</sub> concentrations, climate variables, and gaseous variables (RH, SO<sub>2</sub>, CO, NO<sub>2</sub>) on the vertical coordinate axes.

### 3.4. Predictive models development and validation

We have used MLR analysis, MLP class of ANN, and CART for forecasting of one day ahead PM<sub>10</sub> concentration for all the six air quality monitoring stations of Guwahati city.

#### 3.4.1. Multiple Linear Regression (MLR)

In MLR analysis, the mathematical model was built up to forecast the dependent variable, i.e. next day PM<sub>10</sub> based on the inputs of independent variables comprising of climate variables and gaseous elements. In MLR, the coefficient of determination (R<sup>2</sup>) indicates the overall capability of the model to handle variance in data. The regression model was composed following equation 2 [42,43].

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} + \varepsilon_i \quad (2)$$

where Y is the dependent variable,  $\beta_i$  is the regression coefficients,  $X_i$  is the independent variables and  $\varepsilon$  is a stochastic error associated with the regression. This relationship was used in this study to develop a mathematical equation model to predict the next day PM<sub>10</sub> concentrations of the six ambient air monitoring stations of Guwahati with input variables like meteorological parameters, PM<sub>10</sub>, and gaseous pollutants. MLR assumes that the residuals have a normal distribution with a zero mean, uncorrelated and constant variance. The stepwise multiple linear regression procedure was used here to derive the mathematical equation [43]. Variance inflation (VIF) was used in this study to evaluate the multicollinearity effect on the variance of the estimated regression coefficient. The equation for VIF (Equation 3) is as follows:

$$\text{VIF} = \frac{1}{1-R^2} \quad (3)$$

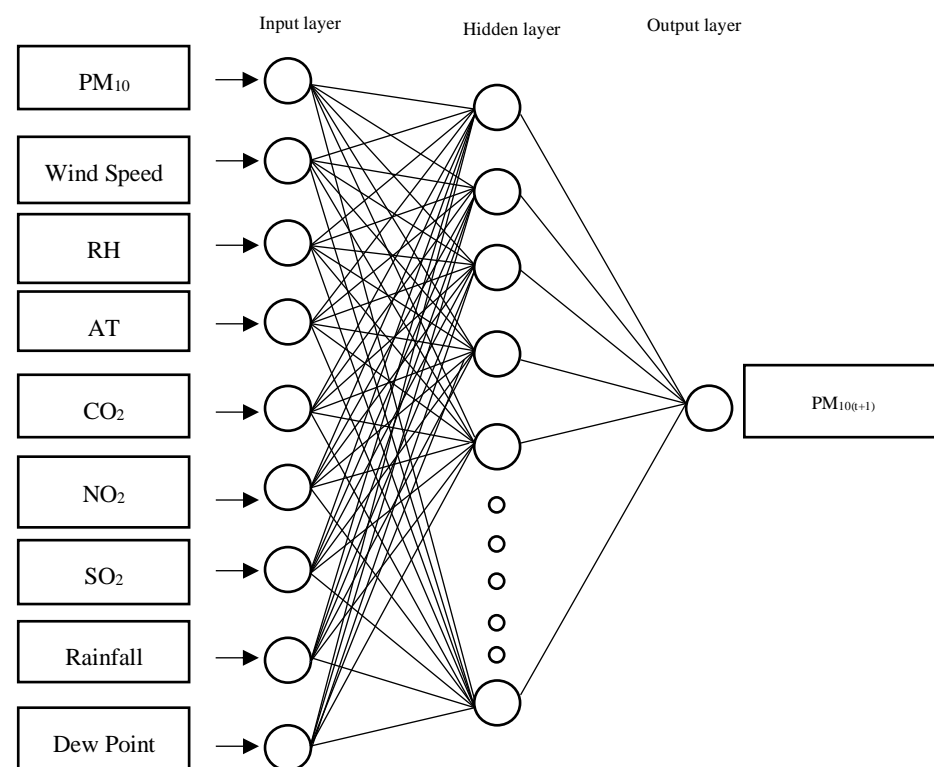
#### 3.4.2. Multi-Layer Perceptron (MLP) Model

ANN is considered to be a robust data modeling technique capable of handling the nonlinear relationship between variables and hence found suitable for the prediction of PM<sub>10</sub> which requires exploration of the complex relationship between particulate matters, meteorological variables, and gaseous pollutants present in the atmosphere [44]. MLP architecture of ANN with FFBP topologies is the most common and effective model out of several models available. We have used MLP in this study to create predictive models for each of six ambient monitoring stations of Guwahati using nonlinear combinations of the input variables (meteorological parameters, PM<sub>10</sub>, PM<sub>2.5</sub>, and gaseous pollutants) to predict the next day PM<sub>10</sub> concentrations. MLP forms a network of functionally interconnected neurons, also known as perceptron [45]. ANN scores more than MLR because of its ability to predict the dependent variable of a built-up model more accurately [46]. MLP has a simple structure consisting of three layers: the input layer, hidden layer, and output layer. One hidden layer was considered in our study, as it was suggested to be sufficient to achieve the optimum model capacity [47]. The number of neurons or the nodes, in the input layer, was equal to the number of input variables introduced in the model. The relevant input variables, i.e. observed meteorological

parameters,  $PM_{10}$ , and gaseous pollutants, are fed in the model as signals to the input layer of the model, which then passed on to the hidden layer. The neurons do the computations to detect features of the input variables and introduce them to the input layer with requisite weights. The weights are assigned to input variables based on their relative importance. The hidden layer does the critical function of nonlinear transformations of the inputs entered into the network through a predefined activation function. The neuron sums up information, including bias, in the hidden layer. The bias does the job of providing a trainable constant value to every neuron in addition to its normal value. The mathematical formulation of the MLP model is as shown below:

$$Y = F\left(\sum_{j=1}^m W_{kj} \cdot F\left(\sum_{i=1}^n W_{ji} X_i + B_j\right) + B_k\right) \quad (4)$$

where  $Y$  = output,  $F$  = transfer function,  $W_{kj}$  = weights between hidden and output layers,  $W_{ji}$  = weights between input and hidden layers,  $X_i$  = input variables,  $m$  = number of neurons in a hidden layer,  $n$  = number of neurons in an input layer,  $B_j$  = bias values of the neurons in the hidden, and  $B_k$  = bias values of the neurons in the output layers.



**Figure 2.** The architecture of the MLP network.

### 3.4.3. Classification and Regression Trees (CART)

CART is a non-parametric regression technique that can be employed for the prediction of an independent variable when the distribution of independent variables is not known. Typically, therefore, the CART method tries to ascertain the distribution pattern of the outcome (dependent) variable using the independent variables through their linear or nonlinear relationship with the outcome variable. CART builds up a decision tree through a hierarchy of binary decisions. Each binary decision will involve splitting of a target variable into two alternative and mutually exclusive branches (groups) depending upon the variation/values of the explanatory variable leading to the

most considerable possible reduction in post-split variations/values of the target variable. In other words, splitting stops when there is no additional gain by further splitting can be achieved [48, 49]. CART allows easy visualization of the process until the terminal node is reached and can handle any type variables (numeric, binary, categorical, etc.). In this study, we built up predictive CART models for all the six air quality monitoring stations of the city with meteorological parameters, PM<sub>10</sub>, and gaseous pollutants as the independent variables and next day PM<sub>10</sub> concentration as the dependent variable.

#### 3.4.4. Model validation

MLR, MLP, and CART equations have been validated by computing net absolute error (NAE), mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), index of agreement (IA), coefficient of determination (R<sup>2</sup>) [50, 51, 52].

**Table 3.** Performance indicators for model validation.

Sl. No.	Performance indicators	Equations
1	Net absolute error	$NAE = \frac{\sum_{i=1}^n  P_i - O_i }{\sum_{i=1}^n O_i}$
2	Mean absolute error	$MAE = \frac{1}{n} \sum_{i=1}^n  P_i - O_i $
3	Mean square error	$MSE = \frac{\sum_{i=0}^n (P_i - O_i)^2}{n}$
4	Root mean square error	$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}}$
5	Index of agreement	$IA = 1 - \left[ \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n ( P_i - \bar{O}  +  O_i - \bar{O} )^2} \right]$
6	Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$

SPSS 25 has been used for computation of MLR and MLP while computation for CART SPSS modeler 18 has been used in this study.

## 4. Result and Discussion

### 4.1. PM<sub>10</sub> across Indian cities: A review

The effort towards air quality improvement is not at all easy for a country like India as the country policymaker cannot forego the objective of faster economic development to sustain its vast population. Ambient particulate matter (PM<sub>2.5</sub> and PM<sub>10</sub>) is being contributed increasingly by many diverse sources where the transportation sector is one of the most important amongst others. The rapid pace of urbanization, increasing affluence level, and inefficiency of the public transportation



system have been jointly contributing towards the addition of a vast number of new automobiles in the cities every year. The number of automobiles registered in India has already reached 0.21 billion in 2015 from a mere 58.92 million during the year 2002. If the whole of India is to be considered then about 53,929 automobiles hit the roads of India every day [53]. All led to the current status of air pollution in India more notable in the world. Table 4 below presents a summary of the output of a number of studies (quantitative and qualitative) conducted by different researchers in the context of 28 Indian cities and their reported level of PM<sub>10</sub> concentrations. Amongst 28 cities considered here, 7 are tier-I, 13 tier-II, and 8 tier-III Indian cities. Contribution of specific sources can vary significantly across the cities, and meteorological parameters prevailed in the respective cities also played an important role in building the ambient PM<sub>10</sub> concentrations. It can be seen from Table 4 that Kolkata, a tier-1 city in India, even clocked PM<sub>10</sub> concentration of as high as 445±210 µg m<sup>-3</sup> during the wintertime [54]. Annual PM<sub>10</sub> concentrations in New Delhi were reported to be 222 ±142 µg m<sup>-3</sup> while an earlier study reported summer and winter mean concentrations were of 95.1 ± 22.2 µg m<sup>-3</sup> and 182 ± 32.5 µg m<sup>-3</sup> respectively [55, 56]. Bengaluru also registered a high annual mean PM<sub>10</sub> concentration level of 349.8 ± 205.8 µg m<sup>-3</sup> during the year 2015 [12]. In comparison, lower concentrations have been reported for Hyderabad and Mumbai where mean PM<sub>10</sub> concentrations for the period 2005-2012 were 174.4 ± 86.6 µg m<sup>-3</sup> and 54.4 ± 25.2 µg m<sup>-3</sup> respectively [57].

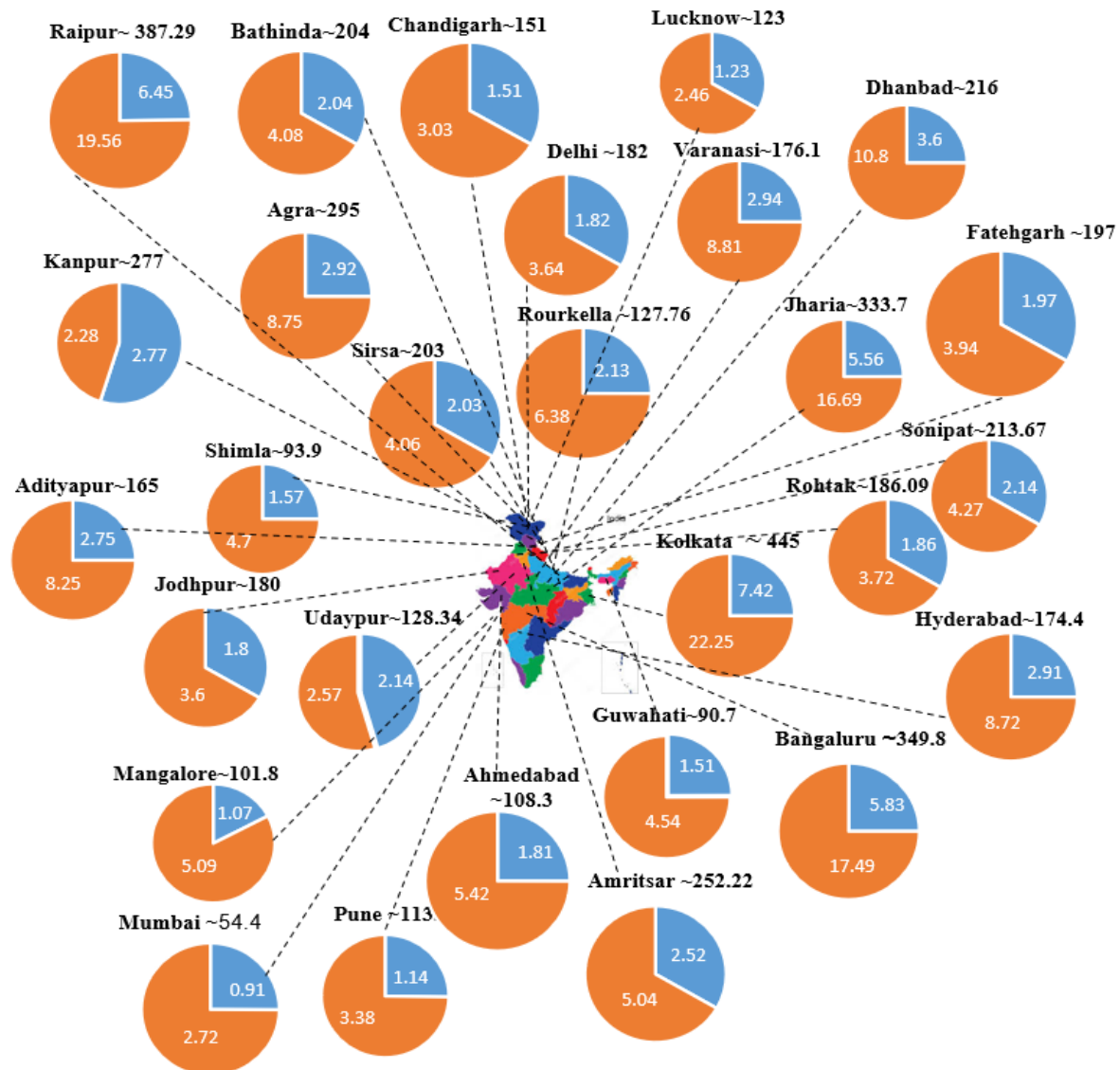
The tier-II cities are also not lagging far behind the tier-I cities of India in terms of PM<sub>10</sub> pollution. It can be seen from Table 4 that Raipur had mean PM<sub>10</sub> concentrations of 387.29±76.85 µg m<sup>-3</sup> from October 2008 to September 2009 while another city Kanpur recorded mean PM<sub>10</sub> concentrations of 277±117.61 µg m<sup>-3</sup> from October 2002 to February 2003 [58, 59]. Amongst the tier-III cities, the reported mean PM<sub>10</sub> concentrations of some specific cities like Jharia and Sonipat were also on the higher side with 333.7±17.86 µg m<sup>-3</sup> and 213.67±151.49 µg m<sup>-3</sup> for the period March 2011 to February 2012 and 03 December to 06 December 2016 respectively [60, 61].

**Table 4.** Summary of ambient PM<sub>10</sub> concentrations from several cities across India (mass concentrations in µg m<sup>-3</sup>).

City	City type	Sampling year	Type	PM <sub>10</sub> µg m <sup>-3</sup>	Ref.
Kolkata	I	Dec 2013–Jan 2014	Winter mean	445±210	[54]
Pune	I	Jan–Dec 2016	Mean	62.5	[62]
		June 2011- May 2012	Mean	113.8	[11]
Hyderabad	I	2005 – 2012	Mean	174.4 ± 86.6	[57]
		June 2004–May 2005	Mean	135.1 ± 37.92	[63]
Mumbai	I	2005 - 2012	Mean	54.4 ± 25.2	[57]
Ahmedabad	I	2005 - 2012	Mean	108.3 ± 69.8	[57]
Delhi	I	April to June 2008	Summer mean	95.1 ± 22.2	[55]
		Nov 2007- Jan 2008	Winter mean	182 ± 32.5	[55]
		Sept 2010-Aug 2012	Mean	222 ±142	[56]
Bengaluru	I	2011	Annual mean	221.4 ± 187.5	[12]
		2012	Annual mean	275.6 ± 180.8	[12]
		2013	Annual mean	314.3 ± 213.4	[12]
		2014	Annual mean	333.6 ± 216.3	[12]
		2015	Annual mean	349.8 ± 205.8	[12]

		2011–2015.	Mean (5 year)	298.94±200.76	[12]
		2005 - 2012	Mean	80.4 ± 21.9	[57]
Jodhpur	II	Aug – Sept 2011	Mean Monsoon	180	[64]
Varanasi	II	Mar 2013 - Feb 2014	Annual mean	176.1±85.	[65]
Agra	II	2000 – 2016	Range	175 to 295	[66]
		April 2010 - Jan 2011	Mean (Semi rural)	230.5	[67]
		April 2010 - Jan 2011	Mean (Roadside)	242	[67]
Guwahati	II	July 2013 - 30 June 2014	Annual mean	90.7 ± 59.7	[68]
Raipur	II	Oct 2008 - Sept 2009	Annual mean	387.29±76.85	[69]
Mangalore	II	Jan 2013 – Oct 2016	Mean	101.8	[70]
Simla	II	2005 - 2012	Mean	93.9 ± 58.7	[57]
Amritsar	II	9Nov - 15Nov 2016	Winter mean	252.22±108.14	[61]
Rourkella	II	Jan 2011 – Dec 2011	Mean(Four seasons)	127.755	[71]
Dhanbad	II	Mar 2014 – Feb 2015	Mean (Summer, post	216±82	[72]
Lucknow	II	Mar -June 2012	Summer Mean	123± 13	[73]
Kanpur	II	Oct 2002 – Feb 2003	Mean	80	[59]
		Oct 2002 – Feb 2003	Mean	277±117.61	[59]
Chandigarh	II	27Oct – 03Nov, 2016	Winter mean	151.45±106.40	[61]
Fatehgarh	III	3Nov – 9Nov, 2016	Winter mean	197.07±61.35	[61]
Bathinda,	III	16Nov – 21Nov, 2016	Winter mean	204.04±70.80	[61]
Sirsa	III	21Nov – 26Nov, 2016	Winter mean	203.12±83.28	[61]
Rohtak	III	26Nov – 03Dec, 2016	Winter mean	186.09±78.33	[61]
Sonipat	III	3Dec – 06Dec, 2016	Winter mean	213.67±151.49	[61]
Jharia	III	Mar 2011 – Feb 2012	Mean	333.7±17.86	[60]
Udaypur	III	July 2017 – June 2018	Mean	128.34	[74]
Adityapur	III	1 July 2013 – 30 June 2014	Mean	165 ± 43.93	[75]

The PM<sub>10</sub> concentrations of the 28 Indian cities mentioned in Table 4 also grossly violated the air quality standard values of 20 µg m<sup>-3</sup> and 60 µg m<sup>-3</sup> enacted by the WHO and NAAQS of India respectively. PM<sub>10</sub> times NAAQS and PM<sub>10</sub> times WHO standards in different Indian cities considered in this study can be seen in Figure 1. Amongst the 7 tier I cities considered here, Kolkata topped the list with 22.25 times WHO standard and 7.42 times NAAQS followed by Bengaluru (17.49 times WHO standard and 5.83 times NAAQS), Delhi (11.1 times WHO standard and 3.7 times NAAQS), Hyderabad (8.72 times WHO standard and 2.91 times NAAQS), Ahmedabad (5.42 times WHO standard and 1.81 times NAAQS) and Pune (5.69 times WHO standard and 1.9 times NAAQS). In the case of 13 tier-II cities, Raipur city's PM<sub>10</sub> concentration was almost 19.56 times the WHO mandated values and 6.45 times the NAAQS values followed by Kanpur (13.85 times WHO standard and 4.62 times NAAQS) and Agra (11.53 times WHO standard and 3.84 times NAAQS). The tier-III city Jharia, known for its underground coal fires, violated the WHO standard by 16.69 times and NAAQS by 5.56 times while the industrial city Sonipat overshoot the WHO standard by 10.68 times and NAAQS by 5.23 times.



**Figure 3.** PM<sub>10</sub> concentrations ( $\mu\text{g m}^{-3}$ ), PM<sub>10</sub> times NAAQS, and PM<sub>10</sub> times WHO standards in different Indian cities.

#### 4.2. Descriptive statistics and PM<sub>10</sub> concentration of Guwahati

##### 4.2.1. Descriptive statistics

The mean values and standard deviations of the meteorological parameters, PM<sub>10</sub>, and gaseous pollutants of the respective air quality monitoring stations of the city under consideration are provided in Table 5. High variability was observed in the PM<sub>10</sub> level. During 2016-18 the daily average PM<sub>10</sub> concentration varied. Across the six air quality monitoring stations, the maximum and minimum mean PM<sub>10</sub> concentration was  $133.32 \mu\text{g m}^{-3}$  and  $51.41 \mu\text{g m}^{-3}$  respectively. The highest daily average PM<sub>10</sub> recorded was  $259.39 \mu\text{g m}^{-3}$ , while the lowest was  $40.67 \mu\text{g m}^{-3}$  during the period 2016-18. The average RH level of the city was found to be on the higher side while wind speed on the lower side. The time-series data reveals the maximum temperature of  $34^\circ\text{C}$  recorded during the summer season while the minimum was  $14^\circ\text{C}$  during the winter season. Guwahati received rainfall due to southwest monsoon and the highest rainfall occurred from June to August.

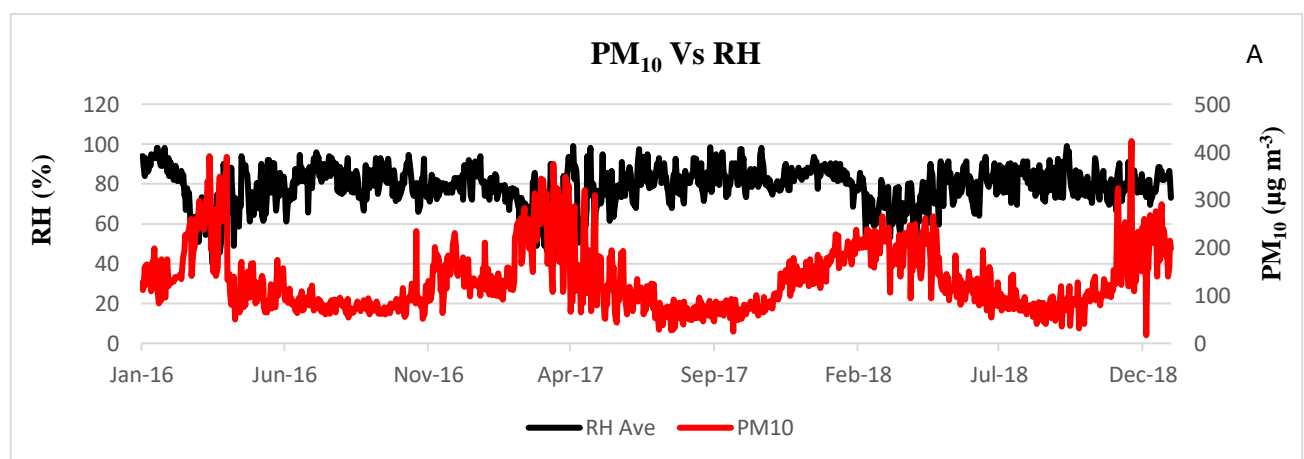
**Table 5.** Descriptive statistics

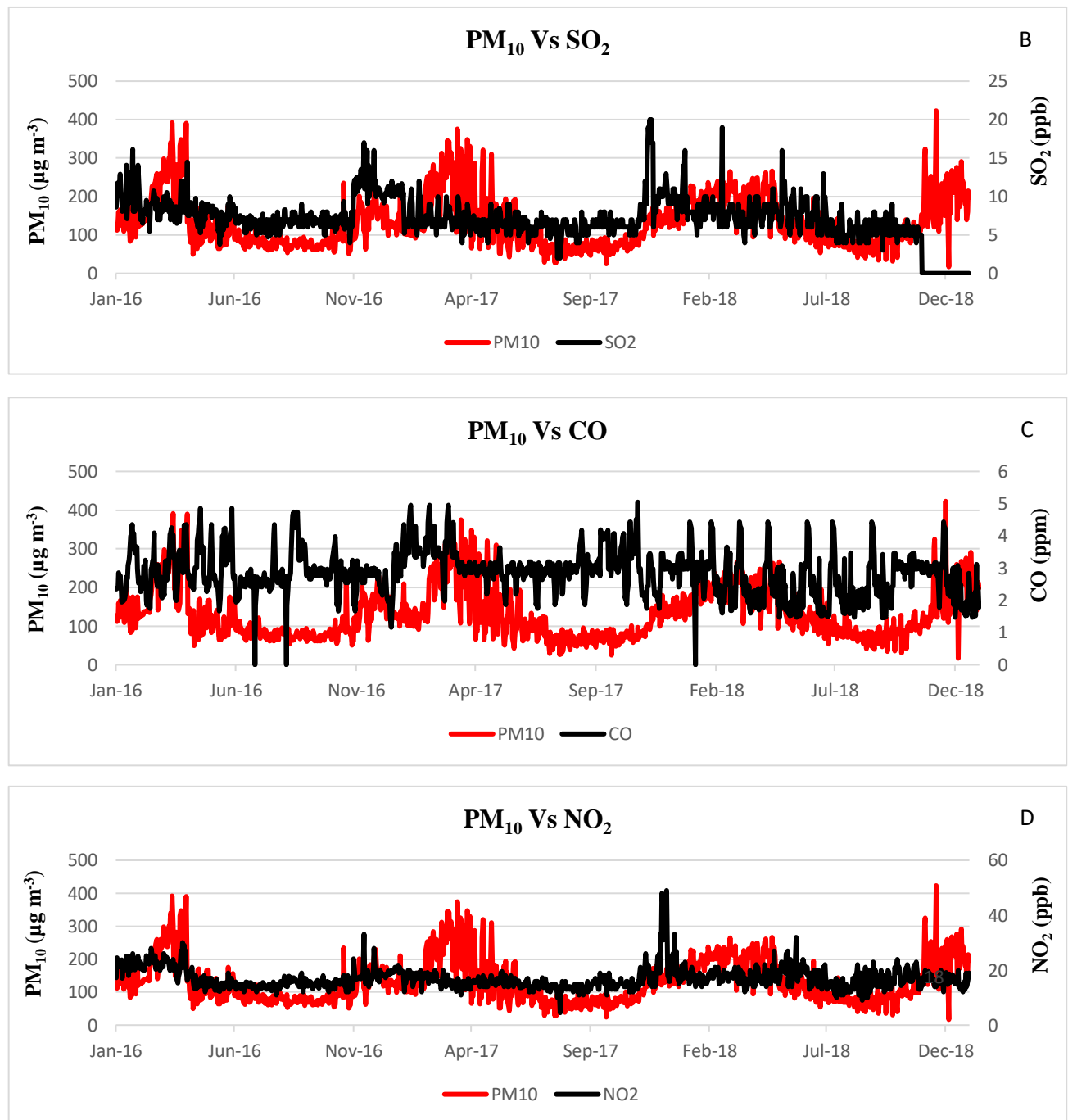
Monitoring Stations: Mean and standard deviation												
Paramet ers	S1_603 Mean	SD	S2_596 Mean	SD	S3_519 Mean	SD	S4_541 Mean	SD	S5_602 Mean	SD	S6_193 Mean	SD
RF	4.11	11.67	4.11	11.67	4.11	11.67	4.11	11.67	4.11	11.67	4.11	11.67
Temp	25.33	4.40	25.33	4.40	25.33	4.40	25.33	4.40	25.33	4.40	25.33	4.40
Dew PT	21.27	4.94	21.27	4.94	21.26	4.94	21.26	4.94	21.27	4.94	21.26	4.94
RH	78.91	9.74	78.91	9.74	78.91	9.74	78.91	9.74	78.91	9.74	78.91	9.74
WS	1.33	0.90	1.33	0.90	1.33	0.90	1.33	0.90	1.33	0.90	1.33	0.90
PM <sub>10</sub>	104.57	57.28	108.62	57.10	133.22	66.99	105.97	51.41	108.62	57.10	109.82	51.41
NO <sub>2</sub>	17.46	4.96	21.66	5.45	17.01	4.06	14.19	5.08	21.66	5.45	16.75	5.08
SO <sub>2</sub>	7.62	2.42	7.55	2.65	7.52	2.27	5.63	2.27	7.55	2.65	7.61	2.27
CO	1.64	0.72	3.25	2.38	2.85	0.71	2.88	0.71	3.25	2.38	3.43	0.72

#### 4.2.2. Correlation of PM<sub>10</sub> concentration, climate variables, and gaseous variables

In Figure 4 (A) - 4(D), the time series of the observed meteorological parameters, PM<sub>10</sub>, and gaseous pollutants are reported in respect of air quality monitoring station 6 of the Guwahati city. It can be observed from Figure 4 (A) that the site is characterized by relatively high humidity throughout the year. The time series, considered in this study shows that the concentration of PM<sub>10</sub> has maintained almost a negative correlation with relative humidity. PM<sub>10</sub> concentration behavior of the city shows a pattern of annual cycle with high concentrations during winter (December to February), possibly due to lower planetary boundary layer height, and a higher level of concentrations seems to continue up to the months of March-April as well, i.e. beyond winter.

Another peculiarity of the site is that both CO and SO<sub>2</sub> have a correlation with PM<sub>10</sub> concentrations suggesting a common source for these compounds but the correlation with SO<sub>2</sub> is stronger as shown in Figure 4 (B) and 4 (C). The PM<sub>10</sub> also shows almost a linear correlation with NO<sub>2</sub>, as evidenced in Figure 4 (D) below.





**Figure 4.** The time series 2016-18 measured in Guwahati: (A) PM<sub>10</sub> Vs RH, (B) PM<sub>10</sub> Vs SO<sub>2</sub>, (C) PM<sub>10</sub> Vs SO<sub>2</sub>, (D) PM<sub>10</sub> Vs NO<sub>2</sub>.

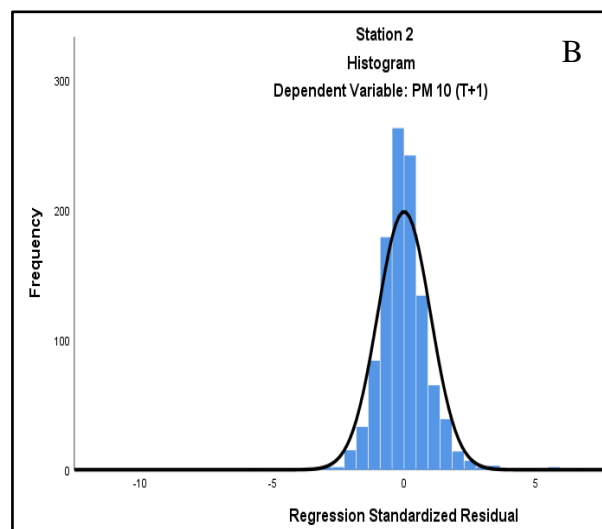
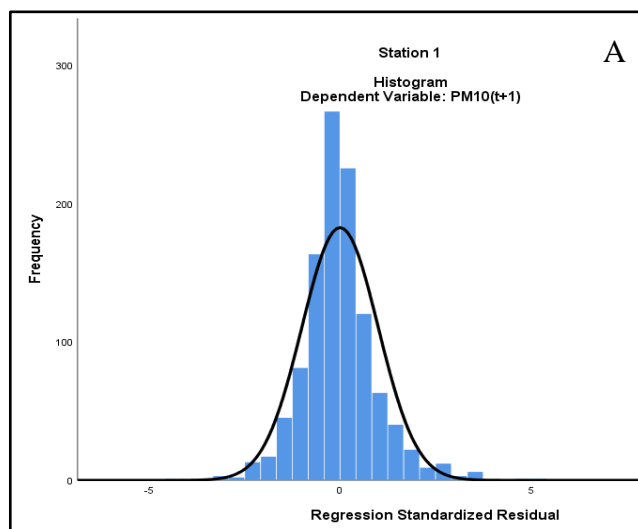
#### 4.3. Multiple Linear Regression Model for PM<sub>10</sub> forecasting

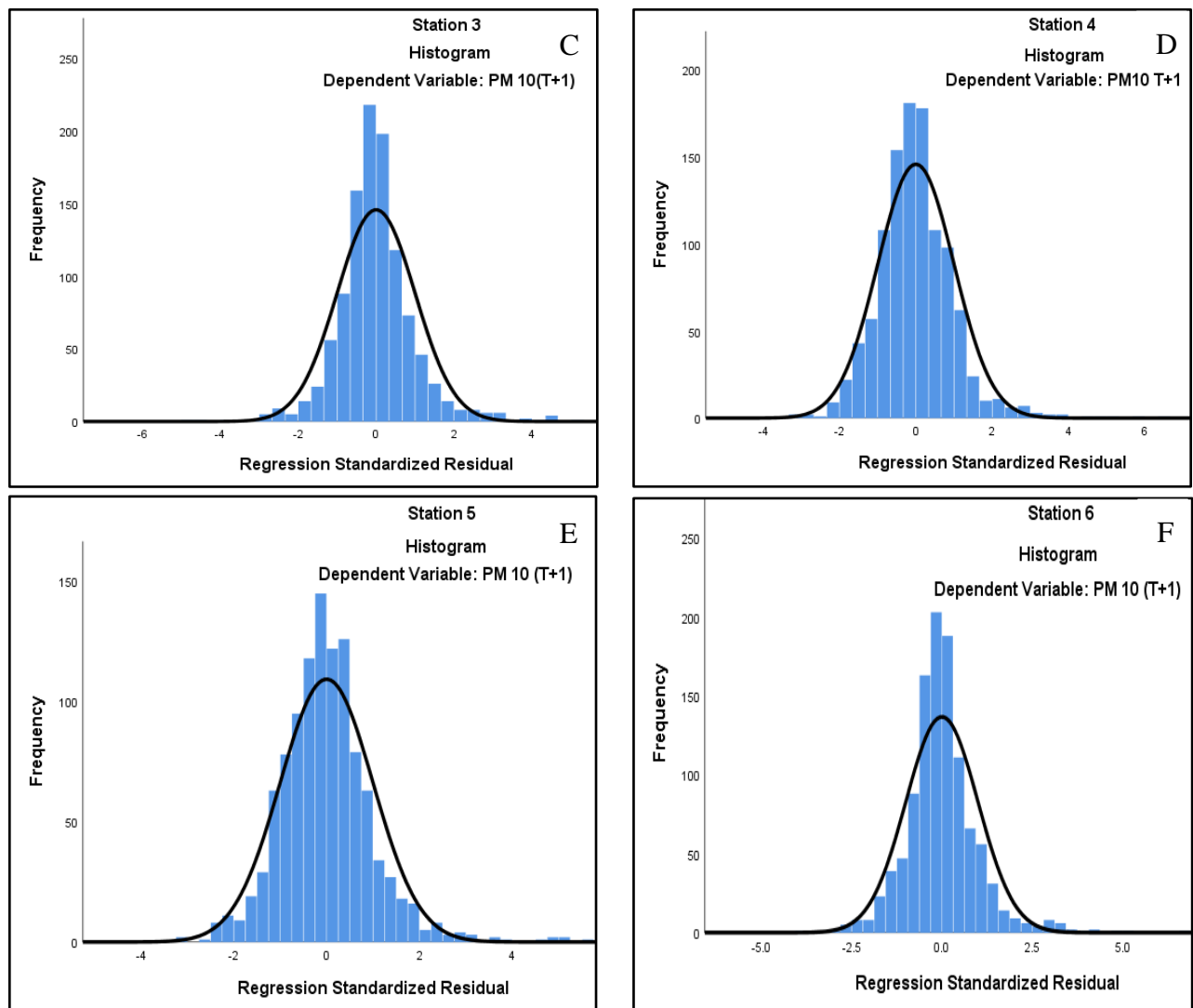
The MLR models summary, developed for all six ambient air quality monitoring stations located at Guwahati, have been placed in Table 6. The range of the Variance Inflation Factor (VIF) for the independent variables of all the six MLR models is found in order as they are below 10, showing the non-existence of multicollinearity issues in the models. Durbin Watson (D-W) statistics show that the models can accommodate the autocorrelation, as the values were in the range of 2.103–2.239. The residual (error) is critical in choosing the robustness of the factual model as linear regression is sensitive to outlier effects. Figure 5A - 5F shows the histogram plot, which indicates that the residuals are also normally distributed with zero mean and constant variance.



**Table 6.** Summary of the Multiple Linear Regression (MLR) models for PM<sub>10</sub> forecasting.

Monitoring Station	Model	R <sup>2</sup>	Range of VIF	D-W Statistics
S1_603	PM <sub>10, (t+1)</sub> concentration = 204.565 + 0.551 (PM <sub>10</sub> ) – 3.509 (Ambient Temperature) -0.870 (Relative Humidity)	0.628	1.277-1.895	2.216
S2_596	PM <sub>10, (t+1)</sub> concentration = 222.67 + 0.492 (PM <sub>10</sub> ) – 3.340 (Ambient Temperature) -1.131 (Relative Humidity) +1.994 (CO)	0.624	1.249-1.925	2.226
S3_519	PM <sub>10, (t+1)</sub> concentration = 214.67 + 0.618 (PM <sub>10</sub> )– 3.382 (Ambient Temperature) -0.988 (Relative Humidity)	0.674	1.330-1.904	2.180
S4_541	PM <sub>10, (t+1)</sub> concentration = 254.38 +0.417(PM <sub>10</sub> ) – 4.934 (Ambient Temperature) -0.782 (Relative Humidity) – 1.026 (SO <sub>2</sub> )	0.622	1.164-2.183	2.103
S5_602	PM <sub>10, (t+1)</sub> concentration = 161.91+0.564 (PM <sub>10</sub> ) – 2.725 (Ambient Temperature) – 0.652 (Relative Humidity)	0.61	1.204-1.754	2.239
S6_193	PM <sub>10, (t+1)</sub> concentration = 216.23 + 0.572 (PM <sub>10</sub> ) – 3.667 (Ambient Temperature) -0.868 (Relative Humidity) – 0.915 (SO <sub>2</sub> )	0.682	1.067-2.082	2.137





**Figure 5.** Histogram plots (A) station1, (B) station 2, (C) station 3, (D) station 4, (E) station 5 and (F) station 6.

#### 4.4. Multi-Layer Perceptron Model

The operating neurons of ANN (MLP) connect the three layers, input layer, hidden layer, and output layer, through adaptable synaptic weights which indicate the strength of the relationship between two connected neurons. Every neuron of each layer summed up all the inputs accepted from previous layers and become the output neuron as per the selected transfer function. The training data set is propagated in the forward phase, through the hidden layer, which comes out through the output layer. The error, i.e. the difference between output values and actual target output values are propagated back toward the hidden layer until the errors are reduced in successive cycles [76]. In the process the neural network learns and change weights during forward and backward phases. The normalized input variables  $PM_{10}$ , RF, T, RH, WS,  $NO_2$ ,  $SO_2$ , CO of the respective air monitoring stations were fed into the six different ANN models using the normalizing data conversion facility of the ANN module of SPSS software. We, in this study, engaged different combination of transfer functions like sigmoid/ hyperbolic tangent, sigmoid /linear, sigmoid /sigmoid and hyperbolic tangent/linear functions to compare and pick up the optimum  $R^2$  values as shown in Table 7. In this way, The network structure, transfer functions of each of the models, and performance indicators can be seen in Table 7 below. The optimum  $R^2$  values are also marked 'bold' in Table 7.

**Table 7.** Predictive MLP models with a network structure, transfer functions, and performance indicators.

Target	Stn.	Network structure	Transfer function:	R <sup>2</sup>	NAE	MAE	MSE	RMSE	IA
		Input:Neurons:	hidden/output layer						
		Output							
PM <sub>10</sub> , (t+1)	S1	08: 07:01	Sigmoid /Hyperbolic tangent	0.626	0.15	16.02	497.86	22.31	0.95
			Sigmoid /Linear	<b>0.650</b>					
			Sigmoid /Sigmoid	0.640					
			Hyperbolic tangent/Linear	0.646					
PM <sub>10</sub> , (t+1)	S2	08: 07:01	Sigmoid /Hyperbolic tangent	<b>0.635</b>	0.22	23.80	1200.21	34.64	0.88
			Sigmoid /Linear	0.634					
			Sigmoid /Sigmoid	0.630					
			Hyperbolic tangent/Linear	0.629					
PM <sub>10</sub> , (t+1)	S3	08: 07:01	Sigmoid /Hyperbolic tangent	0.674	0.20	26.11	1408.45	37.53	0.90
			Sigmoid /Linear	<b>0.685</b>					
			Sigmoid /Sigmoid	0.679					
			Hyperbolic tangent/Linear	0.672					
PM <sub>10</sub> , (t+1)	S4	08: 07:01	Sigmoid /Hyperbolic tangent	0.626	0.21	22.57	962.49	31.02	0.88
			Sigmoid /Linear	0.621					
			Sigmoid /Sigmoid	0.627					
			Hyperbolic tangent/Linear	<b>0.637</b>					
PM <sub>10</sub> , (t+1)	S5	08: 07:01	Sigmoid /Hyperbolic tangent	0.635	0.22	23.40	1158.79	34.04	0.88
			Sigmoid /Identify	0.630					
			Sigmoid /Sigmoid	0.623					
			Hyperbolic tangent/Identify	<b>0.648</b>					
PM <sub>10</sub> , (t+1)	S6	08: 07:01	Sigmoid /Hyperbolic tangent	<b>0.687</b>	0.22	21.87	1007.72	31.74	0.90
			Sigmoid /Identify	0.687					
			Sigmoid /Sigmoid	0.686					
			Hyperbolic tangent/Identify	0.686					

#### 4.5. Predictive CART Model

By using CART analysis, several decision trees were developed based on different combinations of observed meteorological parameters, PM<sub>10</sub>, and gaseous pollutants for the three years (2016-2018). As typical in machine learning, out of the total data points of the respective independent and dependent variables, 60% used as trained set while 40% as the test set. The optimum models were produced for each of the six air quality monitoring stations of Guwahati when they had the least relative errors in respective cases given by equation 5 below.

$$\text{Relative error of CART} = \frac{S(K)}{S(O)} \quad (5)$$

where S (K) is equal to the sum of the squared residuals at the terminal node and S (O) is the sum of squared errors of the dependent error around its mean in the root node. The predictive CART models and performance indicators are given in Table 8.

**Table 8.** Predictive CART models with input variable PM<sub>10</sub>, RF, T, RH, WS, NO<sub>2</sub>, SO<sub>2</sub>, CO, and performance indicators.

Target	Stn.	Set	R <sup>2</sup>	NAE	MAE	MSE	RMSE	IA
PM <sub>10</sub> , (t+1)	S1	Training	0.653	0.26	22.91	1156.72	34.01	0.89
		Testing	0.621	0.25	25.30	1157.21	34.02	0.88
PM <sub>10</sub> , (t+1)	S2	Training	0.501	0.23	27.20	1586.38	39.83	0.81
		Testing	0.533	0.22	27.67	1577.61	39.72	0.82
PM <sub>10</sub> , (t+1)	S3	Training	0.729	0.20	24.11	1189.59	34.49	0.92
		Test	0.656	0.21	27.74	1640.89	40.51	0.89
PM <sub>10</sub> , (t+1)	S4	Training	0.642	0.22	22.01	916.44	30.27	0.88
		Test	0.567	0.24	25.54	1219.22	34.92	0.85
PM <sub>10</sub> , (t+1)	S5	Training	0.635	0.23	21.03	794.11	28.18	0.88
		Test	0.598	0.24	22.06	912.96	30.22	0.86
PM <sub>10</sub> , (t+1)	S6	Training	0.689	0.20	21.61	953.06	30.87	0.90
		Test	0.616	0.21	23.84	1332.57	36.50	0.87

#### 4.6. Model Comparison

All six performance indicators were put to use for comparing the one day ahead PM<sub>10</sub> prediction performances of three methods, i.e. MLR, ANN (MLP), and CART to isolate the best model, as shown in Table 8. NAE, MAE, MSE, and RMSE were used to find the error of the model, where a value closer to 0 indicated a better model. The other two performance indicators, namely, IA and R<sup>2</sup>, were used to check the accuracy of the model result, where higher accuracy is given by a value closer to 1. The values for performance indicators provide specific information regarding predictive performance efficiencies [33]. RMSE wise comparison between models is best desired when the researchers want to avoid large prediction error.

On the other hand, MAE casts light on the average magnitude of the error without considering their direction. The advantage of the linear score of MAE lies in the fact that all individual differences between predictions and corresponding observed values are given equal weight in the average. However, amongst all seven performance indicators, R<sup>2</sup> is the best single measure of how well the predicted values match the observed values.

In this study, the prediction of one day ahead PM<sub>10</sub> for all the six air quality monitoring stations displayed relatively good fits through the use of MLP methods (R<sup>2</sup> = 0.63–0.85 ; IA = 0.88-0.95) and smallest errors ( NAE = 0.15 – 0.22; MAE = 16-26.11 ; MSE = 497.86–1408.45 ; and RMSE = 22.31- 37.53 in comparison to its closest performer MLR method (R<sup>2</sup> = 0.61–0.68 ; IA = 0.87-0.90 ; NAE = 0.20 – 0.23; MAE = 21.37 – 26.26 ; MSE = 859.23–1461.48 ; and RMSE = 29.31- 38.23). It can be seen from Table 6 that CART as predictive method for one day ahead PM<sub>10</sub> were close to MLR but not equal in terms of model evaluation indicators with R<sup>2</sup> = 0.53–0.92; IA = 0.82-0.89; NAE = 0.21 –0.25; MAE = 0.22 – 27.67; MSE = 910.91–1636.21; and RMSE = 30.18- 40.46 in the test set results as clearly revealed in Table 9.

**Table 9.** PM<sub>10</sub> prediction model performance statistics: NAE MAE, MSE, RMSE IA, and R<sup>2</sup> between measured and estimated values for six air quality monitoring stations.

Station 1			Station 2		
MLR	MLP	CART	MLR	MLP	CART

NAE	0.23	0.15	0.25	0.22	0.22	0.25
MAE	24.37	16.00	25.24	24.02	23.73	27.61
MSE	1219.39	497.86	1154.61	1226.49	1192.59	1574.07
RMSE	34.92	22.31	33.98	35.02	34.53	39.67
R <sup>2</sup>	0.63	0.85	0.62	0.62	0.63	0.53
IA	0.87	0.95	0.88	0.87	0.88	0.82
<b>Station 3</b>			<b>Station 4</b>			
	MLR	MLP	CART	MLR	MLP	CART
NAE	0.20	0.20	0.21	0.22	0.21	0.24
MAE	26.26	26.11	27.67	23.05	22.57	25.48
MSE	1461.48	1408.45	1637.21	998.95	962.49	1216.48
RMSE	38.23	37.53	40.46	31.61	31.02	34.88
R <sup>2</sup>	0.67	0.69	0.65	0.62	0.64	0.56
IA	0.89	0.90	0.89	0.87	0.88	0.85
<b>Station 5</b>			<b>Station 6</b>			
	MLR	MLP	CART	MLR	MLP	CART
NAE	0.23	0.22	0.24	0.20	0.20	0.21
MAE	21.37	23.40	22.00	22.20	21.84	23.78
MSE	859.33	1158.79	910.91	1023.64	1007.12	1329.57
RMSE	29.31	34.04	30.18	31.99	31.74	36.46
R <sup>2</sup>	0.61	0.64	0.92	0.68	0.69	0.92
IA	0.87	0.88	0.86	0.90	0.90	0.87

## 5. Conclusion and Recommendation

A review of the literature of 28 Indian different categories of cities (tier-I, tier-II, and tier-III cities) reveals that the PM<sub>10</sub> concentrations in all of them were on the higher side. Kolkata, a tier-I city in India, even clocked PM<sub>10</sub> concentration of as high as 445±210 µg m<sup>-3</sup> during the wintertime. The tier-III cities like Raipur and Kanpur were found to be not lagging far behind the tier-I cities in terms of ambient PM<sub>10</sub> concentration. Interestingly, tier-III cities like Jharia and Sonipat were also recorded PM<sub>10</sub> concentration as high as 333.7±17.86 µg m<sup>-3</sup> and 213.67±151.49 µg m<sup>-3</sup> respectively. The PM<sub>10</sub> concentrations level in all of the 28 Indian cities grossly violated both WHO and NAAQS standards by a wide margin. Kolkata topped the list with 22.25 times more than the WHO standard and 7.42 times NAAQS followed by Bengaluru (17.49 times WHO standard and 5.83 times NAAQS), and Delhi (11.1 times WHO standard and 3.7 times NAAQS). Therefore, it is high time for some requisite actions for diminishing or preventing the build up of high ambient PM<sub>10</sub> concentration level in the cities. One way out is abatement action through short term traffic reduction in cities based on predicted PM<sub>10</sub> concentration level in advance. Therefore, it is high time for the initiation of some requisite actions for diminishing or preventing the build-up of the high ambient PM<sub>10</sub> concentration level in the cities. One way out is abatement action through short term traffic reduction in cities based on predicted PM<sub>10</sub> concentration level in advance. Therefore, it entails the correct prediction of the city level PM<sub>10</sub> concentrations at least one or two days in advance and accordingly initiating prior actions to save the city dwellers from PM<sub>10</sub> induced health hazards.

The tier-II city Guwahati recorded high variability in the observed in PM<sub>10</sub> level due to the rapid urbanization. The highest daily average PM<sub>10</sub> recorded was 259.39 µg m<sup>-3</sup>, while the lowest was 40.67 µg m<sup>-3</sup> during the period 2016-18. During 2016-18, the average daily NO<sub>2</sub>, CO, and SO<sub>2</sub> concentrations



were found to be in correlation with PM<sub>10</sub> concentrations and thereby suggesting a common source for these compounds.

Different forecasting algorithms have been used in different cities of the world to predict PM<sub>10</sub> in advance. However, the use of MLR with stepwise inclusion of input variable was found to be the most widely used tool for temporal prediction of PM<sub>10</sub> in different urban areas of India, and that too mostly applied in bigger cities of the country. This study found that in better forecasting of the next day's PM<sub>10</sub> concentrations in a tier-II city context, i.e. Guwahati the non-linear algorithm MLP with FFBN topologies of ANN class would be giving the best prediction value when compared with linear MLR and nonlinear CART model. These models were critically assessed through a comparative evaluation of performance indicators keeping in mind the end goal is to choose the best-fitted model for accurate forecasting PM<sub>10</sub> at the city level. The result of the study reveals that the one day ahead PM<sub>10</sub> for all the six air quality monitoring stations of Guwahati, prediction ability has been relatively better through the use of MLP methods ( $R^2 = 0.63-0.85$  ; IA = 0.88-0.95) and with smallest errors (NAE = 0.15 – 0.22; MAE = 16-26.11 ; MSE = 497.86–1408.45 ; and RMSE = 22.31- 37.53 in comparison to its closest performer MLR method ( $R^2 = 0.61-0.68$  ; IA = 0.87-0.90 ; NAE = 0.20 – 0.23; MAE = 21.37 – 26.26 ; MSE = 859.23–1461.48 ; and RMSE = 29.31- 38.23). It is interesting to note that CART as predictive method for one day ahead PM<sub>10</sub> were close to MLR but not equal in terms of model evaluation indicators with  $R^2 = 0.53-0.92$ ; IA = 0.82-0.89; NAE = 0.21 –0.25; MAE = 0.22 – 27.67; MSE = 910.91–1636.21; and RMSE = 30.18- 40.46.

In the backdrop of CPCB's acknowledgement that comparatively smaller tier-II cities are also facing severe air pollution, city authorities are contemplating initiating several steps for curtailing air pollution and health hazards thereof. We recommend the local authority to use the non-linear algorithm MLP (ANN) with FFBN topologies for forecasting PM<sub>10</sub> concentration in the smaller Indian cities like Guwahati too for avoiding PM induced health hazards to a great extent. 'Predict pollution and defeat concentration' could be another approach to fight air pollution menace in addition to the odd-even rule, which few Indian cities are enforcing presently to rein on air pollution through curtailment of vehicular pollution. Moreover, with this model, the local SPCB authorities can caution city dwellers of impending dangerous levels of PM<sub>10</sub>, so that they can lessen their outdoor activities for those days and thereby avoiding exposure to unhealthy levels of air quality.

## 6. Acknowledgements

This study was supported by the Graduate School Thesis Grant, Chulalongkorn University, Bangkok, Thailand. The authors also thank the State Pollution Control Board, Assam, and Regional Meteorological Department, Guwahati, for air pollution and meteorological information, respectively.

## References

1. CPCB Bulletin - July 2016. Central Pollution Control Board, Delhi. 2016. Available online: [https://www.cpcb.nic.in/openpdffile.php?id=TGF0ZXN0RmlsZS9MYXRlc3RfMTIzX1NVVTU1BUllfQk9PS19GUy5wZGY=. \[AQ5\]](https://www.cpcb.nic.in/openpdffile.php?id=TGF0ZXN0RmlsZS9MYXRlc3RfMTIzX1NVVTU1BUllfQk9PS19GUy5wZGY=. [AQ5]) (accessed on 8 January 2020).
2. WHO. Concentration occurrence or they should stay away from the high-risk areas. WHO, Geneva; 2018. Available online: [http://www.who.int/phe/health\\_topics/outdoorair/](http://www.who.int/phe/health_topics/outdoorair/) (accessed on 10 March 2020).
3. Myllyvirta, L.; Dahiya, S.; Sivalingam, N. Out of sight: how coal burning advances India's air pollution crisis. Greenpeace Environment Trust, Bengaluru; 2016, Available online: <http://www.greenpeace.org/india/Global/india/cleanairmation/Reports/Out%20of%20Sight.pdf>. (accessed on 26, February 2020).
4. Apte, J.S.; Marshall, J.D.; Cohen, A.J.; Brauer, M. Addressing global mortality from ambient PM<sub>2.5</sub>. *Environ. Sci. Technol.* **2015**, *49*(13), 8057–8066.

5. Sharma, S.; Nayak, H.; Lal, P. Post-Diwali morbidity survey in a resettlement colony of Delhi. *Indian J. Burns*. **2015**, *23*(1), 76–80.
6. Ferreira, T.M.; Forti, M.C.; de Freitas, C.U.; Nascimento, F.P.; Junger, W.L.; Gouveia, N. Effects of particulate matter and its chemical constituents on elderly hospital admissions due to circulatory and respiratory diseases. *Int. J. Environ. Res. Public Health*. **2016**, *13*(10), 947.
7. Chen, K.; Glonek, G.; Hansen, A.; Williams, S.; Tuke, J.; Salter, A.; and Bi, P. The effects of air pollution on asthma hospital admissions in Adelaide, South Australia, 2003–2013: time-series and case–crossover analyses. *Clin. Exp. Allergy*. **2016**, *46*(11), 1416–1430.
8. Ostro, B.; Chestnut, L.; Vichit-Vadakan, N.; Laixuthai, A. The impact of particulate matter on daily mortality in Bangkok, Thailand. *J. Air Waste Manag. Assoc.* **1999**, *49*(9), 100–107.
9. Guttikunda, S.K. Clearing the Air Seminar Series, 'Filling the Knowledge Gap on Air Quality in Indian Cities' Initiative on Climate, Energy and Environment (ICEE) at the Centre for Policy Research (CPR). Delhi, 4 December 2017.
10. Gurjar, B.R.; Jain, A.; Sharma A.; Agarwal A.; Gupta P.; Nagpure A.S.; Lelieveld, J. Human health risks in megacities due to air pollution. *Atmos. Environ.* **2010**, *44*(36), 4606–4613.
11. Yadav, S.; Satsangi, P.G. Characterization of particulate matter and its related metal toxicity in an urban location in South West India. *Environ. Monit. Assess.* **2013**, *185*, 7365–7379.
12. Guttikunda, S.K.; Nishadh, K.A.; Gota, S.; Singh, P.; Chanda, A.; Jawahar, P.; Asundi, J. Air quality, emissions, and source contributions analysis for the Greater Bengaluru region of India, *Atmos. Pollut. Res.* **2019**, *10*(3), 941–953.
13. Czernecki, B.; Póhrolniczak, M.; Kolendowicz, L.; Maros., M.; Kendzierski, S.; and Pilguy, N. Influence of the atmospheric conditions on PM<sub>10</sub> concentrations in Poznań, Poland. *J. Atmos. Chem.* **2017**, *74*(1), 115–139.
14. Carnevale, C.; Pisoni, E.; Volta, M. A non-linear analysis to detect the origin of PM<sub>10</sub> concentrations in Northern Italy. *Sci. Total. Environ.* **2010**, *409*(1), 182–191.
15. Wang, W. Progress in the impact of polluted meteorological conditions on the incidence of asthma. *J. Thorac. Dis.* **2016**, *8*(1), E57–E61.
16. Hooyberghs, J.; Mensink, C.; Dumont, G.; Fierens, F.; Brasseur, O. A neural network forecast for daily average PM<sub>10</sub> concentrations in Belgium. *Atmos. Environ.* **2005**, *39*(18), 3279–3289.
17. Shahraiyni, H.T.; Sodoudi, S. Statistical Modeling Approaches for PM<sub>10</sub> Prediction in Urban Areas; A Review of 21st-Century Studies. *Atmosphere*. **2016**, *7*, 15.
18. Jiang, P.; Dong, Q.; Li, P. A novel hybrid strategy for PM<sub>2.5</sub> concentration analysis and prediction. *J. Environ. Manage.* **2017**, *196*, 443–457.
19. Cabaneros, S.M.; Calautit, J.K.; Hughes, B.R. A review of artificial neural network models for ambient air pollution prediction. *Environ. Model Softw.* **2019**, *119*, 285–304.
20. Goyal, P.; Chan, A.T.; Jaiswal, N. Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmos. Environ.* **2006**, *40*(11), 2068–2077.
21. Nagendra, S.M.S.; Khare, M. Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. *Ecol. Model.* **2006**, *190*(1–2), 99–115.
22. Tikhe Shruti, S.; Khare, K.C.; Londhe S.N. Forecasting criteria air pollutants using data driven approaches; An Indian case study, *Journal Of Environmental Science, Toxicology And Food Technology (IOSR-JESTFT)*. **2013**, *3*(5), 1–8.

23. Chelani, A.B.; Gajghate, D.G.; Hasan, M.Z. Prediction of ambient PM<sub>10</sub> and toxic metals using artificial neural networks. *J. Air Waste Manag. Assoc.* **2002**, *52*(7), 805–810.
24. Kumari, P.R.; Avisetty, R.V.S.D.S.P.; Akkala, P.; Subash, K.V.V.; Manideep, K.S.; Bojja, P.; Aruna, B. Prediction and Estimation of Pm10 and SO<sub>2</sub> Concentrations in the Ambient Air At Vijayawada Station using Artificial Neural Networks Computing. *Int. J. Recent Technol. Eng.* **2019**, *7*(6C2), 790-793.
25. Kaur, M.; Mandal, A. "PM<sub>2.5</sub> Concentration Forecasting using Neural Networks for Hotspots of Delhi," 2020. International Conference on Contemporary Computing and Applications (IC3A), Lucknow, India, 5-7 February 2020; pp. 40-43
26. Dutta, A.; Jinsart, W. Risks to health from ambient particulate matter (PM<sub>2.5</sub>) to the residents of Guwahati city, India: An analysis of prediction model. *Hum. Ecol. Risk Assess.* **2020**.
27. Gocheva-Ilieva, S.G.; Stoimenova, M.P. PM<sub>10</sub> Prediction and Forecasting Using CART: A Case Study for Pleven, Bulgaria. World Academy of Science, Engineering and Technology, *International Journal of Environmental and Ecological Engineering.* **2018**, *12*(9), 572-577.
28. Slini, T.; Kaprara, A.; Karatzas, K.; Moussiopoulos. N. PM<sub>10</sub> forecasting for Thessaloniki, Greece. *Environ. Modell. Softw.* **2006**, *21*(4), 559–565.
29. Barman, N.; Gokhale, S. Urban black carbon-source apportionment, emissions and long-range transport over the Brahmaputra River Valley. *Sci. Total Environ.* **2019**, *693*, 133577.
30. Pant, P.; Lal, R.M.; Guttikunda, S.K.; Russell, A.G.; Nagpure, A.S.; Ramaswami, A.; Peltier, R.E. Monitoring particulate matter in India: recent trends and future outlook. *Air Qual. Atmos. Health.* **2019**, *12*(1), 45–58.
31. Prakash, A.; Kumar, U.; Kumar, K.; Jain, V.K. A wavelet-based neural network model to predict ambient air pollutants' concentration. *Environ. Model. Assess.* **2011**, *16*(5), 503–517.
32. Singh, K.P.; Gupta, S.; Kumar, A.; Shukla, S.P. Linear and nonlinear modeling approaches for urban air quality prediction. *Sci. Total Environ.* **2012**, *426*, 244–255.
33. Singh, K.P.; Gupta, S.; Rai, P. Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmos. Environ.* **2013**, *80*, 426–437.
34. Kottur, S.V.; Mantha, S.S. An integrated model using artificial neural network (ANN) and kriging for forecasting air pollutants using meteorological data. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE).* **2015**, *4*(1), 146-152.
35. Mishra, D.; Goyal, P.; Upadhyay, A. Artificial intelligence based approach to forecast PM<sub>2.5</sub> during haze episodes: A case study of Delhi, India. *Atmos. Environ.* **2015**, *102*, 239-248.
36. Gogikar, P.; Tyagi, B.; Gorai, A.K. Seasonal prediction of particulate matter over the steel city of India using neural network models. *Model. Earth Syst. Environ.* **2019**, *5*, 227–243.
37. Yadav, V.; Nath, S. Novel hybrid model for daily prediction of PM<sub>10</sub> using principal component analysis and artificial neural network. *Int. J. Environ. Sci. Technol.* **2019**, *16*(6), 2839–2848.
38. Masood, A.; Ahmad, K. A model for particulate matter (PM<sub>2.5</sub>) prediction for Delhi based on machine learning approaches. *Procedia Comput. Sci.* **2020**, *167*, 2101-2110.
39. Agarwala, S.; Sharma, S.; Suresh, R.; Rahman, M.H.; Vranckx, S.; Maiheu, B.; Blyth, L.; Janssen, S.; Gargava, P.; Shukla, V.K.; Batra, S. Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions. *Sci. Total Environ.* **2020**, *735*, 139-454.
40. Nadeem, I.; Ilyas, A.M.; Uduman, P.S.S. Forecasting Ambient Air Quality Of Chennai City In India. *Geogr. Environ. Sustain.* **2020**, *0*.

41. Bhardwaj, R.; Pruthi, D. Evolutionary techniques for optimizing air quality model. *Procedia Comput. Sci.* **2020**, *167*, 1872–1879.
42. Vlachogianni, A.; Karppinen, A.; Kassomenos, P.; Karakitsios, S.; Kukkonen, J. Evaluation of a multiple regression model for the forecasting of the concentrations of NO<sub>x</sub> and PM<sub>10</sub> in Athens and Helsinki. *Sci. Total Environ.* **2011**, *409*(8), 1559–1571.
43. Abdullah, S.; Ismail, M.; Ahmed, A.N.; Abdullah, A.M. Forecasting Particulate Matter Concentration Using Linear and Non-Linear Approaches for Air Quality Decision Support. *Atmosphere* **2019**, *10*, 667.
44. Feng, X.; Li, Q.; Zhu, Y.; Hou, J.; Jin, L.; Wang, J. Artificial neural networks forecasting of PM<sub>2.5</sub> pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos. Environ.* **2015**, *107*, 118–128.
45. Vemuri, V. *Artificial neural networks: theoretical concepts*; IEEE Computer Society Press Washington DC, United States, (1988), pp. 145; ISBN: 978-0-8186-0855-1.
46. Gardner, M.W.; Dorling, S.R.; Artificial neural networks (the multilayer perceptron)- a review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*(14-15), 2627-2636.
47. Bishop, C.M. *Neural Networks for Pattern Recognition*, Oxford Univ. Press: Oxford, NY, USA, 1995; ISBN 978-0-19-853864-6.
48. Moisen, G.G.; Frescino, T.S. Comparing five modelling techniques for predicting forest characteristics. *Ecol. Model.* **2002**, *157*(2-3), 209-225.
49. Mckenney, D.W.; Pedlar, J.H. Spatial models of site index based on climate and soil 701 properties for two boreal tree species in Ontario, Canada. *Forest Ecol. Manage.* **2003**, *175*, 497-507.
50. Willmott, C.J.; Matsuura, K.; Robeson, S.M. Ambiguities inherent in sums-of-squares-based error statistics. *Atmos. Environ.* **2009**, *43*(3): 749–752.
51. Jinsart, W.; Sripraparkorn, C.; Siems, S.T.; Hurley, P.J.; Thepanondh, S. Application of the air pollution model (TAPM) to the urban air shed of Bangkok, Thailand. *Int. J. Environ. Pollut.* **2010**, *42*(1-3), 68–84.
52. Grzesiak, W.; Zaborski, D. *Examples of the use of data mining methods in animal breeding*. Data mining applications in engineering and medicine. Adem Karahoca, IntechOpen, Croatia. 2012; pp. 303-324, Available online: <https://www.intechopen.com/books/data-mining-applications-in-engineering-and-medicine/examples-of-the-use-of-data-mining-methods-in-animal-breeding>.
53. Dutta, G.; Dutta, A. Indian Growth Story of Automobile Sector and Atmospheric Emission Projection, *Pollut. Res.* **2018**, *37*(1), 131-143.
54. Das, R.; Khezri, B.; Srivastava, B.; Datta, S.; Sikdar, P.K.; Webster, R.D. Trace Element Composition of PM<sub>2.5</sub> and PM<sub>10</sub> from Kolkata - A Heavily Polluted Indian Metropolis. *Atmos. Pollut. Res.* **2015**, *6*(5), 742-747.
55. Singh, D.P.; Gadi, R.; Mandal, T.K. Characterization of particulate-bound polycyclic aromatic hydrocarbons and trace metals composition of urban air in Delhi, India. *Atmos. Environ.* **2011**, *45*, 7653–7663.
56. Tiwari, S.; Bisht, D.S.; Srivastava, A.K.; Pipal, A.S.; Taneja, A.; Srivastava, M.K.; Attri, S.D. Variability in atmospheric particulates and meteorological effects on their mass concentrations over Delhi, India. *Atmos. Res.* **2014**, *145–146*, 45–56.
57. Dholakia, H.H.; Bhadra, D.; Garg, A. Short term association between ambient air pollution and mortality and modification by temperature in five Indian cities. *Atmos. Environ.* **2014**, *99*, 168-174.

58. Deshmukh, D.K.; Deb, M.K.; Tsai, Y.I.; Mkomu, S.L. Water Soluble Ions in PM<sub>2.5</sub> and PM<sub>1</sub> Aerosols in Durg City, Chhattisgarh, India. *Aerosol Air Qual. Res.* **2011**, *11*, 696-708.
59. Sharma, M.; Maloo, S. Assessment of ambient air PM<sub>10</sub> and PM<sub>2.5</sub> and characterization of PM<sub>10</sub> in the city of Kanpur, India. *Atmos. Environ.* **2005**, *39*(33), 6015–6026.
60. Roy, D.; Singh, G.; Seo, Y.C. Carcinogenic and non-carcinogenic risks from PM<sub>10</sub> and PM<sub>2.5</sub>-bound metals in a critically polluted coal mining area. *Atmos. Pollut. Res.* **2019**, *10*(6), 1964-1975.
61. Ravindra, K.; Singh, T.; Mor, S.; Singh, V.; Mandal, T.K.; Bhatti, M.S.; Gahlawat, S.K.; Dhankhar, R.; Mor, S.; Beig, G. Real-time monitoring of air pollutants in seven cities of North India during crop residue burning and their relationship with meteorology and transboundary movement of air. *Sci. Total Environ.* **2019**, *690*, 717-729.
62. Gawhane, R.D.; Rao, P.S.P.; Budhavant, K.; Meshram, D.C.; Safai, P.D. Anthropogenic fine aerosols dominate over the Pune region, Southwest India. *Meteorol. Atmos. Phys.* **2019**, *131*, 1497–1508.
63. Gummeneni, S.; Yusup, Y.B.; Chavali, M.; Samadi, S.Z. Source apportionment of particulate matter in the ambient air of Hyderabad city, India. *Atmos. Res.* **2011**, *101*(3), 752–764.
64. Sudheer, A.K.; Aslam, M.Y.; Upadhyay, M.; Rengarajan, R.; Bhushan, R.; Rathore, J.S.; Singh, S.K.; Kumar, S. Carbonaceous aerosol over semi-arid region of western India: Heterogeneity in sources and characteristics. *Atmos. Res.* **2016**, *178–179*, 268–278.
65. Murari, V.; Kumar, M.; Barman, S.C.; Banerjee, T. Temporal variability of MODIS aerosol optical depth and chemical characterization of airborne particulates in Varanasi, India, *Environ. Sci. Pollut. Res.* **2015**, *22*, 1329–1343.
66. De, S. Long-term ambient air pollution exposure and respiratory impedance in children: A cross-sectional study, *Respir. Med.* **2019**.
67. Pipal, A.S.; Jan, R.; Satsangi, P.; Tiwari, S.; Taneja, A. Study of Surface Morphology, Elemental Composition and Origin of Atmospheric Aerosols (PM<sub>2.5</sub> and PM<sub>10</sub>) over Agra, India. *Aerosol Air Qual. Res.* **2014**, *14*, 1685-1700.
68. Tiwari, S.; Dumka, U.C.; Gautam, A.S.; Kaskaoutis, D.G.; Srivastava, A.K.; Bisht, D.S.; Chakrabarty, R.K.; Sumlin, B.J.; Solm, F. Assessment of PM<sub>2.5</sub> and PM<sub>10</sub> over Guwahati in Brahmaputra River Valley: Temporal evolution, source apportionment and meteorological dependence, *Atmos. Pollut. Res.* **2017**, *8*, 13-28.
69. Deshmukh, D.K., Deb, M.K. & Mkomu, S.L. Size distribution and seasonal variation of size-segregated particulate matter in the ambient air of Raipur city, India. *Air Qual. Atmos. Health.* **2013**, *6*, 259–276.
70. Kalaiarasan, G.; Balakrishnan, R.M.; Sethunath, N.A.; Manoharan, S.; Source apportionment studies on particulate matter (PM<sub>10</sub> and PM<sub>2.5</sub>) in ambient air of urban Mangalore, India. *J. Environ. Manag.* **2018**, *217*, 815–824.
71. Kavuri, N.C.; Paul, K.K. Chemical Characterization of Ambient PM<sub>10</sub> Aerosol in a Steel City, Rourkela, India. *Res. J. Recent Sci.* **2013**, *2*(1), 32-38.
72. Jena, S.; Singh, G. Human health risk assessment of airborne trace elements in Dhanbad, India. *Atmos. Pollut. Res.* **2017**, *8*(3), 490-502.
73. Lawrence, A.; Fatima, N. Urban air pollution & its assessment in Lucknow City — The second largest city of North India. *Sci. Total. Environ.* **2014**, *488–489*, 447–455.



74. Yadav, M.; Soni, K.; Soni, B.K.; Singh, N.K.; Bamniya, B.R. Source apportionment of particulate matter, gaseous pollutants, and volatile organic compounds in a future smart city of India. *Urban Climate*. **2019**, *28*:100470.
75. Shubhankar, B.; Ambade, B. Chemical characterization of carbonaceous carbon from industrial and semi urban site of eastern India. *Springer Plus*. **2016**, *5*, 837.
76. Ul-Saufie, A.Z.; Yahaya, A.S.; Ramli, N.A.; Rosaida, N.; Hamid, H.A. Future daily PM<sub>10</sub> concentrations forecasting by combining regression models and feedforward backpropagation models with principal component analysis (PCA). *Atmos. Environ*. **2013**, *77*, 621–630.