Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 44 No. 12
December 2023

Statistical Analysis of Particulate Matter Over an Urban Area, India

TejasTurakhia[1,4*],NirmiTurakhia[3], Dr. M.N. Patel[3], Tejas V Shah[2], Deepali H Shah[2]
[1]Research Scholar, Gujarat Technological University (GTU), Ahmedabad, Gujarat, India
[2]Gujarat Technological University (GTU), Ahmedabad, Gujarat, India
[3]Department of Statistics, USSC, Gujarat University, Ahmedabad, Gujarat, India
[4]Department of Physics and Electronics, St. Xavier's College (Autonomous), Ahmedabad, Gujarat, India

Abstract- In recent years India is developing rapidly and this development has serious impact on our environment. The rapid growth in air pollution events is becoming more and more serious. This study aims to explore the impact of meteorological factors on $PM_{2.5}$ and PM_{10} concentrations in Delhi, India. Delhi is an appropriate place for study about long range transport pollutants and the correlation between particulate concentrations and meteorological conditions as it is located in the northern part of India is capital of our country and has been affected by pollutants from within the city and outside the city areas.The particulate matter concentrations and meteorology data from April, 2018 to November, 2021 were collected. There is a significant correlation between the explanatory variables and to remove the multi-collinearity and to determine independent explanatory variables we have used Principal Component Analysis (PCA). Using the component scores obtained from PCA, we have done the regression analysis for prediction of $PM_{2.5}$ and PM_{10}. To analyze the seasonal variation, we separated the groups for analysis by creating a grouping variable. There was a statistically significant difference between groups as determined by one-way ANOVA. A Tukey post hoc test revealed that there was a significant difference in the average values of PM concentrations ($PM_{2.5}$ and PM_{10}) between most of all the pairs except the season's summer and Post Monsoon.
Keywords- PM concentration ($PM_{2.5}$ and PM_{10}), Meteorological parameters, Correlation, Principal Component Analysis (PCA), Seasonal variation

1. Introduction

Asia is considered to be one of the prime regions of the world in the environment of atmospheric aerosols loading because of the existence of growing economies like India, China and other Asian countries.In India, pollution has crossed all levels and especially the air pollution because of the enhanced anthropogenic activities like burning fossil fuels [1]. According to air quality status, some of the Indian cities are considered to be amongst the most polluted cities in the world [2, 3, 18, 28]. In current years, atmospheric pollution due to dust particles in urban areas has attracted concern since vulnerability to increased levels of particulate matter concentration is related to mortality and number of respiratory effects [1, 16, 24]. Atmospheric aerosols mainly impact on air quality and human health and lives, it can reduce visibility and causes epidemiological diseases such as lung injuries and cardiovascular diseases. It has been reported that due to atmospheric pollution there are hundreds of thousands of cases of respiratory illnesses every year alone in Delhi, India [7].

WHO made a survey of air quality of 1650 world cities and noted that the air quality in Delhi, the capital of India, is the worst of any major city in the world [29, 30].So, several districts around Delhi are also affected. Air pollution in India is estimated to kill about 2 million people every year; it is the fifth largest killer in India. India has the world's highest death rate from chronic respiratory diseases and asthma, according to the WHO. In Delhi, poor quality air irreversibly damages the lungs of 2.2 million or 50 percent of all children.

Data released by the Central Pollution Control Board shows that during January to September air quality index of Delhi remains at Moderate (101–200) level, and then during October to December it drastically deteriorates to Very Poor (301–400), Severe (401–500) or Hazardous (500+) levels due to stubble burning, road dust, vehicle pollution and cold weather and many more factors. But it has been observed that in Delhi, trucks and road dust are bigger pollutants than cars.

The presence of particulate matter in the environment of Delhi is not only due to the contribution by vehicular and industrial activities but also there is a contribution of soil originated particles and construction activities. The main sources present for the decline of air quality in New Delhi, are motor vehicle traffic, domestic fuel burning, industrial sources and power plants. The action of transportation in Indian mega cities is the major source of aerosol particles. Domestic use of biofuels (for example: wood, dried plant material, manure, biogas, biofuel) plays a crucial role for the resource of particulate matter in India as well. There are rare studies that have investigated the number concentration of particles on the Indian continent [19]. Maximum cities worldwide have evident severe air quality issue due to industries and vehicles. The analysis of air pollution revealed that the major problem influencing these cities is their eminent levels of total suspended particles (TSP) [12, 17].

It is skillfully accepted that high levels of particulate matter (PM) are remarkably related with poor health effects, eco-system damage and deteriorating visibility [8, 12]. Most of the studies on air quality performed so far are based on quantification of PM concentration in suburban air because of their health impacts [19, 23]. Though, recent studies show that the number concentration could be a strongly preferable and measure of the health effects of the particulate matter than the mass concentration [6, 19, 22]. Goyal and Sidhartha (2003) designed the ambient air quality of Delhi before and after the execution of compressed natural gas (CNG) for all transporting vehicles [12]. Various studies have released more agreeing correlation for the concentration of fine ($PM_{2.5}$) and inhalable (PM_{10}) particles with well-being than any other air pollutant [1, 25, 26]. Conversely at the same time, Osunsanya et.al, (2000), established no evidence to maintain the hypothesis that the component of particulate pollution responsible for effects on respiratory symptoms [19]. A bivariate regression model was executed on the data of air pollutants and meteorological variables collected for the four cities Delhi, Mumbai, Kolkata and Chennai for the period July-August 2001 to inspect the influence of meteorological variables on air pollutant concentrations. They suggested an air quality index using the weighted arithmetic mean method, which appear to be suitable in the evaluation of overall air quality with respect to pollutants [12, 20]. Remarkable studies have implemented statistical analysis to apportion sources and determined full year seasonal variations [1, 13, 15].

2. Material And Methodology

2.1 Study area and data availability

Delhi is the capital of India and geologically come across in the northern part of India at 28.61°N 77.23°E. It extends a peak of 318 m (1,043 ft) and is a superior characteristic of the region. The National Capital Territory of Delhi covers a region of 1,484 km^2 (573 sqm). Among which 783 km^2 (302 sqmi) is assigned to rural area while 700 km^2 (270 sqmi) for urban, which makes it the wide-reaching city in terms of area in the country. As Delhi indicates danger to major earthquakes, it is comprised in India's seismic zone-IV [22].

Delhi has an atypical version of the humid subtropical climate bordering a hot semi-arid climate. The heat season is experienced from March 21 to June 15 with an average temperature above 39°C (102°F). The winter period is from November 26 to February 9 with an average temperature below 20°C (68°F). In the prior March, the direction of wind changes from North-Western to South-Western. Hot climate occurs in the months from April to October. With an increased humidity, the rainy season is shown up at the end of June. The mild cold appears at the end of June, and is highest in the month of January with heavy fog. But in the present time of climate change, the seasonal variation is uncertain. Delhi temperature varies within the range of 2 to 47°C (35.6 to 116.6°F). The lowest noted temperature was -2.2°C (28.0°F), while the highest temperature was 48.4°C (119.1°F). The average annual temperature is 25°C (77°F). The monthly temperature on an average varies within the range of 13 to 32°C (55 to 90°F). The average annual rainfall is roughly around 886 mm (34.9 in) [23]

For this study the daily averaged concentration data of $PM_{2.5}$ and PM_{10} and meteorological data was collected from four weather stations of Delhi for the period of April 2018 to November 2021 from Central Pollution Control Board (CPCB). The data has been collected from four stations (i) AnandVihar (ii) Mandir Marg (iii) Punjabi Bagh and (iv) RK Puram for the same period. The dataset collected from this stations are:

1) PM_{10} – Particulate Matter of size less than or equal to 10 um.
2) $PM_{2.5}$ – Particulate Matter of size less than or equal to 2.5 um.
3) WS – Wind Speed in meter per second.
4) WD – Wind Direction in degree.
5) AT – Atmospheric Temperature in degree Celcius.
6) BP – Bar Pressure in mmHg.
7) RH – Relative Humidity in percentage.

Figure 1. Map of Delhi, India with selected CPCB stations

2.2 Methodology

Firstly, the CPCB data of PM concentrations and meteorological conditions was collected for the period of April, 2018 to November, 2021 over the four stations of Delhi: (i) AnandVihar (ii) Mandir Marg (iii) Punjabi Bagh and (iv) RK Puram as given in section 2.1. After that the data was arranged station-wise and year wise and a monthly average was calculated for the whole database. In section 3.1 the correlation analysis is conducted between PM concentrations and meteorological conditions. The principal component analysis was carried out to determine the set of independent explanatory variables for different meteorological variables in section 3.2. Section 3.3 covers the multivariate regression analysis for PM concentrations and meteorological factors based on the independent explanatory variables obtained from principal component analysis done in section 3.2. Finally, in section 3.4 the seasonal analysis was experimented in order to understand the effect of four season's winter, summer, monsoon and post monsoon on $PM_{2.5}$ and PM_{10} as a whole based on the data of all the situations. Section 4 deals with conclusion of the work.

3. Result And Discussion

3.1 Correlation analysis

Correlation analysis was carried out to construct the relationship between PM concentrations and meteorological conditions over Delhi, India. As we know that PM mass concentrations have a direct relationship with meteorological variables. The mixture of emissions and meteorological conditions is critical when examining the properties of PM [5, 10]. The emissions may not experience divergence. While the air quality at any location changes from time to time, since the air quality fully rely on the dynamics of the atmosphere and meteorological conditions plays a major role in controlling the hazard of air pollutants. Corresponding correlations may detect the regularity of the sources for PM emissions [5]. The results of correlation analysis between the PM concentrations and meteorological conditions (that is, relative humidity, wind direction, wind speed, bar pressure and atmospheric temperature) are displayed in Table 1. The values in the brackets indicates the significant values (p-values). The p-value in the output is studied to test the significance relationship between the two variables of the data. If p-value is less than the level of significance (0.05), we fail to prove that there is no relationship between the variables and thus we conclude that there is some statistically significant relationship among the variables under study.

Table1: Correlation Coefficient between the variables.

	$PM_{2.5}$	PM_{10}	RH	WD	WS	BP	AT
$PM_{2.5}$	1	0.792 (0.000)	0.020 (0.258)	0.066 (0.000)	-0.436 (0.000)	0.034 (0.049)	-0.484 (0.000)

PM_{10}	0.792 (0.000)	1	-0.224 (0.000)	0.137 (0.000)	-0.296 (0.000)	0.116 (0.000)	-0.199 (0.000)
RH	0.020 (0.258)	-0.224 (0.000)	1	-0.049 (0.004)	-0.199 (0.000)	-0.009 (0.623)	-0.412 (0.000)
WD	0.066 (0.000)	0.137 (0.000)	-0.049 (0.004)	1	-0.205 (0.000)	0.711 (0.000)	-0.088 (0.000)
WS	-0.436 (0.000)	-0.296 (0.000)	-0.199 (0.000)	-0.205 (0.000)	1	0.047 (0.006)	0.370 (0.000)
BP	0.034 (0.049)	0.116 (0.000)	-0.009 (0.623)	0.711 (0.000)	0.047 (0.006)	1	0.031 (0.072)
AT	-0.484 (0.000)	-0.199 (0.000)	-0.412 (0.000)	-0.088 (0.000)	0.370 (0.000)	0.031 (0.072)	1

From the above table we can say that the correlation between the variables $PM_{2.5}$ and Relative Humidity; Bar Pressure and Relative Humidity, and Bar Pressure and Atmospheric Temperature is not significant. That is there is no any linear effect of Relative Humidity on $PM_{2.5}$, Relative Humidity on Bar Pressure and Atmospheric Temperature on Bar Pressure.

PM_{10} showed significant negative correlation ($r = -0.224$, $p = 0.000$) with relative humidity during the study period, thus PM gets resolved or washed out by drizzle especially in rainy season in the case of higher relative humidity. Similar negative relationship of relative humidity and PM concentration was obtained by Hienet. al, 2002, which suggested that atmospheric particulates are separated by relative humidity and decrease the mass of re-suspended soil dust by creating the humid soil [5, 11]. $PM_{2.5}$ showed significant negative correlation ($r = -0.484$, $p = 0.000$) and PM_{10} also showed significant negative correlation ($r = -0.199$, $p = 0.000$) with atmospheric temperature. This shows that higher temperature might be against the gathering of pollutants [4, 5]. Many studies carried out in the Indian cities also obtained a negative relationship of relative humidity and temperature with PM concentrations [5, 14, 27].

Bar pressure and PM concentrations shows significant positive correlation ($PM_{2.5}$, $r = 0.034$, $p = 0.049$; PM_{10}, $r = 0.116$, $p = 0.000$), as the bar pressure remains constant during the entire study period. Wind direction has significant positive relationship with PM concentrations ($PM_{2.5}$, $r = 0.066$, $p = 0.000$; PM_{10}, $r = 0.137$, $p = 0.000$), whereas wind speed has significant negative relationship with PM concentrations ($PM_{2.5}$, $r = -0.436$, $p = 0.000$; PM_{10}, $r = -0.296$, $p = 0.000$).

Also, we observe that the variables RH and AT have not any significant effect on BP. That means RH and AT are independent of variable BP. Otherwise we observe a significant relationship between the variables. Such a significant relationship between the variables creates the multi- collinearity in the study variables.

To remove the multi-collinearity and to determine groups of independent variables we have used Principal Component Analysis in section 4. Similar study for PCA was conducted by Kulshrestha, A., et al, 2009in order to recognize important elements correlated with different sources for Agra city in India [1].

3.2 Principal component analysis

The principal components analysis (PCA) is one of the most popular multivariate statistical techniques to analyze the multivariate data. This method was developed by Karl Pearson (1901), but in 1939, Hotelling made a much more formal presentation and named the term principal component (PC) [31, 32]. It may be used for expressing the data in a way proper for highlighting their similarities and differences [34].

One of the main applications of principal component analysis is to reduce the dimensionality of the variables in a data set without loss of any information contained in the data. That is to decide how many principal components must be retained. Obviously, the maximum number of principal components that can be retained is the number of variables in the data set.

PCA is nothing but the linear transformation applied to a set of commonly correlated variables of the data, to turn them into a smaller number of uncorrelated and orthogonal variables. Thus, it is a method to determine a smaller number of variables which utilized the information contained in data set [34,35].

PCA is composed as:

$$PC_i = l_{1i}X_1 + l_{2i}X_2 + \cdots + l_{ni}X_n$$

where PC_i is the i^{th}principal component and l_{ji} is the loading of the observed variable X_j .

PCA is one of the most used statistical techniques in environmental sciences. To evaluate air quality monitoring networks (AQMN) in environmental phenomena the PCA is used [35].

Such kind of study has been made by many authors. Abdullah et al. (2018) used PCA to analyze the meteorological parameters and air pollution data. Voukantsis et al. (2011) used principal component analysis and artificial neural networks, to inter-compare air quality and meteorological data, and to forecast the concentration levels for environmental parameters. Polanco (2016) considered PCA in the evaluation of air quality monitoring networks. Recently, Nunez-Alonso e al (2019) have used PCA for air quality assessment in the Madrid region located in the Centre of the Iberian Peninsula.

We have applied PCA on the data set with variables RH, WD, WS, BP and AT. Our aim is to determine independent set of variables which can be used in prediction of $PM_{2.5}$ and PM_{10} separately for four places of Delhi and as a whole for Delhi.

At the first step, to proceed with the PCA the checking of KMO and Bartlett's test values are important.

The result of KMO and Bartlett's test shown in Table 2. exhibits two tests that specify the appropriateness of our data for structure observation. It indicates the proportion of variance in our variables that might have been caused by some primary factors. Small p-values ($p\text{-value}<0.05$) concludes that the factor analysis will be appropriate for our data.

Table2: KMO and Bartlett's Test

Measure of Sampling	0.573
Approximate Chi-Square	3962.772
Degree of Freedom	10
Sig. value	0.000

Principal components (PC) are the eigenvectors of a covariance matrix or a correlation matrix, and each PC extracts a maximal share of the total variance. According to Kaiser Criterion, PC with an Eigen value greater than or equal to 1 is considered as being of statistical significance. The Eigen values refer to the total variance explained by each factor.

In our data, according to this criterion, we have two Eigen values greater than oneand hence from PCA we have obtained two independent sets of variables which are dominated by the loadings of RH, WD, WS, BP and AT.

Then we have computed the score coefficients for each variable in each component, which are shown in the Table 3.

Table3: Component Score Co-efficient Matrix

	Component	
	1	2
RH	0.183	-0.391
WD	0.468	0.269
WS	-0.278	0.305
BP	0.404	0.351
AT	-0.270	0.413

From Table 3, we get two components which we consider as independent explanatory variables for predicting $PM_{2.5}$ and PM_{10}.

F_1=0.183(RH)+0.468(WD)-0.278(WS)+0.404(BP)-0.270(AT)

F_2=0.391(RH)+0.269(WD)+0.305(WS)+0.351(BP)+0.413(AT)

Using the above two equations we can determine the component scores.

3.3 Multivariate regression analysis

In this section we have developed multivariate regression model to predict $PM_{2.5}$ and PM_{10} for four regions of Delhi separately and as a whole for Delhi.

Sharma et al (2018) considered time series regression forecasting for analyzing the pollution trends of 2016 to 2021 in Delhi [36]. Chaloulakou et al. (2003) used Neural Network and Multiple Regression Models for predicting PM10 in Athens [37].Elbayoumi (2014)utilized multivariate regression method based on PCA for predicting indoor PM_{10} and $PM_{2.5}$ in naturally ventilated schools buildings in Gaza Strip (Palestine) [38].

We have developed the regression model to predict $PM_{2.5}$ and PM_{10} considering the two factors obtained by PCA as explanatory variables. The multiple correlation coefficient (R) and its p-value and regression equations for dependent variable $PM_{2.5}$ are shown in Table 4. and for dependent variable PM_{10} are shown in Table 5.

Table 4: Multiple Correlation Coefficient and Regression Equations.

SR. NO	STATION NAME	R (p-value)	Regression Equation $PM_{2.5}$=
1	ANAND VIHAR	0.454 (0.000)	$167.978+(35.994)F_1-(34.969)F_2$
2	MANDIR MARG	0.532 (0.000)	$77.553+(49.090)F_1-(19.595)F_2$
3	PUNJABI BAGH	0.514 (0.000)	$240.493+(96.830)F_1-(65.352)F_2$
4	RK PURAM	0.506 (0.000)	$98.770+(59.142)F_1-(13.937)F_2$
5	COMBINE	0.432 (0.000)	$125.394+(30.307)F_1-(31.375)F_2$

Table 5: Multiple Correlation Coefficient and Regression Equations.

SR. NO.	STATION NAME	R (p-value)	Regression Equation PM_{10}=
1	ANAND VIHAR	0.404 (0.000)	$398.950+(139.041)F_1+(46.229)F_2$
2	MANDIR MARG	0.334 (0.000)	$167.428+(56.506)F_1+(30.256)F_2$
3	PUNJABI BAGH	0.351 (0.000)	$381.902+(104.825)F_1+(80.956)F_2$
4	RK PURAM	0.422 (0.000)	$187.513+(116.875)F_1+(27.716)F_2$
5	COMBINE	0.206 (0.000)	$286.666+(40.164)F_1+(46.754)F_2$

From the Table 4 and Table 5 we found that all the multiple regression co-efficient are significant (p-value < 0.05) and the two factor F_1 and F_2 have significant effect $PM_{2.5}$ and PM_{10}. So the equation can be used for prediction of $PM_{2.5}$ and PM_{10} for given value of the five variables.

3.4 Seasonal analysis

The origin of the variation in the PM concentrations between monsoon, winter, spring and summer seasons are due to the manmade activities, existing meteorology, various sources of strength, and the diffusion properties of particulate matter. The seasonal variations are principally due to the meteorological factors for example temperature, relative humidity, and wind speed which hazardously affect the diffusion of particulate matter[5]. It will be of worth interest to know the effect of four season's winter, summer, monsoon and post monsoon on $PM_{2.5}$ and PM_{10} as a whole based on the data of all the situations. For such analysis we have framed the following hypothesis and tested using the Analysis of Variance technique.

Null-Hypothesis: H_0: There is no significant difference in the average value of PM concentrations ($PM_{2.5}$ and PM_{10}) between the four seasons.

Alternative-Hypothesis: H_1: At least two of the pairs differ significantly.

To analyze the seasonal variation in the given data using SPSS software, we separated the groups for analysis by creating a grouping variable called month code and gave the values to the seasons. The value 1 was assigned for winter season (November-February), value 2 was assigned for summer season (March-June), value 3 was assigned for Monsoon season (July, August) and value 4 was assigned for Post Monsoon season (September, October).The statistical output is shown in the ANOVA table in Table 6.

Table 6: ANOVA Table

		Sum of Squares	DF	Mean Square	F value	Sig.
$PM_{2.5}$	Between Groups	13918378.300	3	4639459.433	787.763	0.000
	Within Groups	19340811.681	3284	5889.407		
	Total	33259189.981	3287			
PM_{10}	Between Groups	30378991.490	3	10126330.497	351.603	0.000
	Within Groups	94580805.563	3284	28800.489		
	Total	124959797.054	3287			

As the p- value is less than 0.05, the level of significance, ANOVA shows that the mean value significantly differs for each seasons, it would be necessary to check which of the four seasons have different effect on the mean concentrations. For such analysis we have used Post Hoc (Tukey) test for multiple comparison of all possible pair of seasons. The results are shown in the Table 7.

Table 7: Post- Hoc test for Multiple Comparisons

Dependent Variable	Season code(I)	Season code(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
$PM_{2.5}$	1.00	2.00	130.84596*	3.46356	0.000	121.9434	139.7485
		3.00	177.54926*	3.99208	0.000	167.2882	187.8103
		4.00	125.30384*	4.01350	0.000	114.9877	135.6199
	2.00	3.00	46.70330*	3.77315	0.000	37.0050	56.4016
		4.00	-5.54212	3.79581	0.462	-15.2987	4.2144
	3.00	4.00	-52.24542*	4.28353	0.000	-63.2556	-41.2353
PM_{10}	1.00	2.00	156.54442*	7.65926	0.000	136.8575	176.2314
		3.00	280.64178*	8.82803	0.000	257.9507	303.3329
		4.00	165.14325*	8.87540	0.000	142.3304	187.9561
	2.00	3.00	124.09736*	8.34389	0.000	102.6507	145.5441
		4.00	8.59883	8.39399	0.735	-12.9766	30.1743
	3.00	4.00	-115.4985*	9.47254	0.000	-139.8462	-91.1508

*The mean difference is significant at the 0.05 level.

Table 7, shows which pair of groups differed from each other. The Tukey post hoc test is generally the preferred test for conducting post hoc tests on a one-way ANOVA, but there are many others. We can see from the table 7 that there is no statistically significant difference in the average PM concentrations between the season's summer and Post Monsoon.

4. Conclusion

In this study the daily averaged concentration data of $PM_{2.5}$ and PM_{10} and meteorological data was collected from four weather stations of Delhi for the period of April 2018 to November 2021 from Central Pollution Control Board (CPCB). There is a significant correlation between the explanatory variables and to remove the multi-collinearity and to determine independent explanatory variables we have used Principal Component Analysis (PCA). Using the component scores, we have done the regression analysis. To analyze the seasonal variation, we separated the groups for analysis by creating a grouping variable. There was a statistically significant difference between groups as determined by one-way ANOVA (for $PM_{2.5}$, p = 0.000; for PM_{10}, p= 0.000). A Tukey post hoc test revealed that there was a significant difference in the average values of PM concentrations ($PM_{2.5}$ and PM_{10}) between most of all the pairs except the season's summer and Post Monsoon.

Acknowledgement

The authors would like to thank Dr. Rajesh Iyer, Head, Department of Physics and Electronics, St. Xavier's College, Ahmedabad, Gujarat, India for helping us to understand the concept of Air Quality and guiding us in downloading the Air Quality and Meteorological data of Delhi. We are thankful to Dr. Shraddha C Bhatt, Assistant Professor, and Dr. KirtanParmar, Assistant Professor, Department of Statistics, St. Xavier's College, Ahmedabad, Gujarat, India for explaining us the principles of statistical analysis. The authors also wish to express gratitude to the anonymous reviewers for their valuable comments and suggestions.

References

[1] Kulshrestha, A., et al. 2009. Metal concentration of $PM_{2.5}$ and PM_{10} particles and seasonal variations in urban and rural environment of Agra, India. *Sci. of the Total Env.* 407(24):6196-6204.

[2] Central Poll. Cont. Board. Air quality status and trends in India. National Ambient Air Quality Monitoring Series. (2000). NAAQMS/14/1999-2000.

[3] Central Poll. Cont. Board. Air quality status in Delhi 1989-2000. National Ambient Air Quality Monitoring Series. (2001). NAAQMS/17/2000-2001.

[4] Mitra, A.P., et al. 2002. Indian aerosols: present status. *Chemosphere*, 49:1175-1190.

[5] TERI. Review of past and on-going work on urban air quality in India. *Tata Energy Res. Inst.,* 101. (TERI Project Report No. 2001EE41).

[6] Lee SC, et al. 2006. $PM_{1.0}$ and $PM_{2.5}$ characteristics in the roadside environment of Hong Kong. *Aerosol Sci. Tech.* 40:157-165.

[7] ROOsli, M., et al. 2001. Temporal and spatial variation of PM_{10} at urban and rural sites in the Basel area, Switzerland. *Atmos. Env*. 35:3701-3713

[8] Faiz, A., et al. 2002. New Directions: air pollution and road traffic in developing countries. *Atmos. Env.* 1:241-243.

[9] Ambient (outdoor) air pollution in cities database WHO 2014. Retrieved 31 May 2015.

[10] Monkkonen, P., et al. 2004. Relationship and variations of aerosol number and PM_{10} mass concentrations in a highly polluted urban environment- New Delhi, India. *Atmos. Env.* 38:425-433.

[11] Gupta, I., Kumar, R. 2006. Trends of particulate matter in four cities in India. *Atmos. Env.* 40:2552-2566.

[12] Mage, D., et al. 1996. Urban air pollution in megacities of the world. *Atmos. Env.* 30:681–686.

[13] Goswami, E., et al. 2002. Spatial characteristics of fine particulate matter identifying representative monitoring locations in Seattle, Washington. *J. of Air and Waste Manage. Assoc.* 52:324–333.

[14] Pope C.A. 2000. Epidemiology of Fine Particulate Air Pollution and human Health: Biological Mechanisms and Who's at Risk? *Env. Health Perspectives*, 108:713-723.

[15] Donaldson, K., et al. 2001. Ultrafine particles. *Occupational Env. Medicine* 58:211-216.

[16] Peters, A., et al. 1997. Respiratory effects are associated with the number of ultrafine particles. *American J. of Respiration and Critical Care Medicine.* 155:1376-1383.

[17] Goyal, P., et al. 2003 Present scenario of air quality in Delhi: a case study of CNG implementation. Atmos. Env. 37:5423–5431.

[18] Schwarta, J. 1994. What are people dying of an high air pollution days. *Env. Res*. 64:26-35.

[19] Schwartz, J., et al. 1996. Is daily mortality associated specifically with fine particles? *Air Waste Manage. Assoc.* 46:927-939.

[20] Osunsanya, T., et al. 2001. Acute respiratory effects of particles: mass or number? *Occupational Env. Medicine* 58:154-159.

[21] Mouli, P.C., et al. 2004. Monitoring of air pollution in Indian Metropolitan cities: modeling and quality indexing. *Int. J. Env.* Poll. 21:365–382.

[22] Karar, K., Gupta AK. 2006. Seasonal variations and chemical characterization of ambient PM_{10} at residential and Industrial sites of an urban region of Kolkata (Calcutta), India. *Atmos. Res*. 81:36-53.

[23] Khillare PS, et al. 2004. Spatial and temporal variation of heavy metals in atmospheric aerosol of Delhi. *Env. Monit Assess*. 90:1-21.

[24] https://en.wikipedia.org/wiki/Delhi

[25] https://en.wikipedia.org/wiki/Climate_of_Delhi

[26] http://cpcb.nic.in/real-time-air-quality-data/

[27] Dhananjay, K., et al. 2012. Size distribution and seasonal variation of size-segregated particulate matter in the ambient air of Raipur city, India. *Air Qual Atmos Health* 6:259-276.

[28] Gupta AK, et al. 2004. A statistical analysis of particulate data sets of Jawaharlal Nehru port and surrounding harbor region in India. *Environ Moint Assess.* 95:295-309.

[29] Hien, PD, et al. 2002. Influence of meteorological conditions on $PM_{2.5}$ and

$PM_{2.5\text{-}10}$ concentrations during the monsoon season in Hanoi, Vietnam. *Atmos Env.* 36:3473-3484.

[30] Deshmukh DK, et al. 2011. Water soluble ions in $PM_{2.5}$ and PM_1 aerosols in Drug city, Chhattisgarh, India. *Aerosol Air Qual Res* 11:696-708.

[31] Katiyar SC, et al. 2002. Assessment of breathing level ambient air quality of Shilong: Part 1: total suspended particulate loadings. *Indian J Environ Prot.* 22:115-122

[32] Singh, R., et al. 2009. Seasonal air quality profile of inorganic ionic composition of PM_{10} near TajMahal in Agra, India. *Env. Moint Assess*

[33] Abdi, H. & Williams, L. J. 2010, Principal Component Analysis. *Wiley Interdisc. Rev.: Comp. Stat.* 2(4):433–459.

[34] Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *J. of Educational Psychology* 24(6):417–441.

[35] Jolliffe IT. 2002. Principal component analysis. 2nd ed. *New York: Springer*.

[36] Wilks, D. S. 1995. Statistical Methods in the Atmospheric Sciences. *Academic Press, London.*

[37] Sharma, N., et al. 2018. Forecasting air pollution load in Delhi using data analysis tools. *Procedia computer science*. 132:1077-1085.

[38] Chaloulakou, A., et al. 2003. Network and Multiple Regression Models for PM_{10} Prediction in Athens: A Comparative Assessment, J. of the Air & Waste Manage. Assoc. 53:10, 1183-1190.

[39] Elbayoumi, M. 2014. Multivariate methods for indoor PM_{10} and $PM_{2.5}$ modelling in naturally ventilated schools buildings. *Atmos. Env.* 94:11-21.

[40] Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Phil. Mag.* 2(11):559–572.