Eco Cities 2025, 6(1), 3182.
https://doi.org/10.54517/ec3182

Article

Influence of different air pollutants on concentration of $PM_{2.5}$ in national capital region (NCR), India

Abhishek Prasad, Pooja Maibam, Samiran Saikia, Kurme Uday Teja, Subhadeep Biswas*

Civil Engineering Department, National Institute of Technology Silchar, Assam 788010, India
*** Corresponding author:** Subhadeep Biswas, sdeep0819@gmail.com; subhadeep@civil.nits.ac.in

CITATION

Prasad A, Maibam P, Saikia S, et al. Influence of different air pollutants on concentration of $PM_{2.5}$ in national capital region (NCR), India. Eco Cities. 2025; 6(1): 3182. https://doi.org/10.54517/ec3182

ARTICLE INFO

Received: 26 December 2024
Accepted: 9 April 2025
Available online: 23 April 2025

Abstract: Air pollution has become a hot topic, especially among megacities throughout the globe. In this context, the national capital region (NCR) of India, including Delhi and its adjoining areas, deserves special mention. Air pollutants such as sulfur dioxide (SO_2), nitrogen oxides (NO_x), ground-level ozone (O_3), carbon monoxide (CO), and particulate matter ($PM_{2.5}$, PM_{10}) are some of the important elements often making adverse impacts on the city air in recent times. Especially $PM_{2.5}$ and PM_{10} in the NCR have come up several times in the news reports due to extremely high concentrations. The present article focuses on finding out the correlation between the concentration of $PM_{2.5}$ and other pollutants in two regions named Indirapuram and Noida, prominent locations in the NCR. The results (R^2) indicate that correlation fits better when the concentration of two pollutants is compared against $PM_{2.5}$ rather than compared with a concentration of a single pollutant. In most of the cases, a positive correlation is observed, while in a few cases, a negative correlation is attained. Finally, the model is tested against some known values of independent and dependent variables.

Keywords: air pollution; $PM_{2.5}$; NCR; Indirapuram; Noida

1. Introduction

Among different cities of the world, the National Capital Region (NCR) of India, including Delhi, is one of the most highlighted spots of recent times due to its alarming levels of air pollution. As per the World Air Quality Report published in 2021, Delhi has been enlisted as one of the most severely polluted cities in the world [1]. NCR comprises various towns and villages in three states, namely Haryana, Uttar Pradesh, and Rajasthan, which share borders with Delhi. It is considered to be one of the rapidly urbanizing regions and plays a vital role in commercial growth. However, with the economic growth and expansion of industries, the region also witnessed a drastic deterioration in the quality of the ambient air over the recent past. Air quality in this region is predominantly affected by particulate matter (PM), mainly $PM_{2.5}$ and PM_{10}, and gaseous pollutants such as nitrogen oxides (NO_x), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3). These pollutants pose a great threat to human health and the environment, both independently and in combination [2,3]. Some of the severe human health effects due to air pollution include the occurrence of asthma, nausea, high blood pressure, birth defects, etc. The intensity of the impacts depends upon the nature and duration of exposure to the pollutants. Various articles discussed the effects of different types of air pollutants in this area in contemporary times. Selokar et al. described the dreadful impacts of $PM_{2.5}$ pollution affecting the Delhi-NCR zone [4]. In another study, Kaushik and Das highlighted the influence of several meteorological parameters responsible for the variation of NO_x concentration at Anand Vihar, one of the prominent hotspots in this place [5]. Among various factors, temperature and humidity have been reported by the authors as the major influential ones in governing NO_2 concentration levels.

Several parameters actively contribute to the deteriorating air quality, including a huge number of vehicular, industrial, and coal-based power plant emissions. Different construction activities and roadside biomass burning are also some of the crucial factors. As Delhi records a large number of vehicles on the road among different cities in the world, its pollution level from the vehicles is obviously expected to be high. As per the records of 2018–2019, the number of cars plying on highways in Delhi was more than that in other Indian metropolitan cities like Kolkata, Mumbai, and Chennai. Moreover, the number of diesel-fueled vehicles, one of the triggering factors behind the rise of air pollution, is also increasing day by day [4]. Although the pollution concentration remains at an elevated concentration in the varying range throughout the year, the condition becomes even worse in the post-monsoon period, particularly in the winter months. Especially in the winter months, the Indo-Gangetic plain, of which Delhi is a major portion, is covered by smog [6]. It occurs due to the capture of the smoke arising from different sources in the lower atmosphere due to low wind speed. Despite the policies implemented in the NCR to reduce air pollution, the ambient $PM_{2.5}$ concentration in Delhi and surrounding regions often remains higher than the annual permissible limit prescribed by national ambient air quality standards (NAAQS) (40 $\mu g/m^3$) [2]. Often, during the nights of Diwali, the $PM_{2.5}$ concentration in Delhi's atmosphere crosses 300–400 $\mu g/m^3$.

Due to severe health impacts, among different air pollutants, particulate matters (PM_{10} and $PM_{2.5}$) occupy a significant position. $PM_{2.5}$ is a severely impactful pollutant as it has the ability to penetrate deep inside the human body. Its thickness is much less than that of the human hair. It can cause various health problems such as cardiovascular diseases and respiratory problems, which may even lead to lung cancer and reproductive problems, including premature mortality [7]. According to the data, exposure to it resulted in around 4.2 million deaths and more than 100 million disabilities induced in the world scenario in the year 2015 [8].

Different articles of contemporary times presented the dreadful aspects of the high concentration of $PM_{2.5}$ over the NCR in very recent times. Vaishali et al. explored the relationship between the concentration of $PM_{2.5}$ and the relative humidity and moisture content of the atmosphere [1]. Sharma and Mandal investigated the composition of $PM_{2.5}$ and also the source attributed to the elevated concentration in the ambient atmosphere [9]. Gupta et al. nicely reported the inorganic composition, their formation mechanism, and pH levels while creating $PM_{2.5}$ in the atmosphere of NCR [10]. Being motivated by these promising studies of the present day, the current work is aimed at predicting the relation between the concentration of $PM_{2.5}$ and the concentration of other gaseous and particulate pollutants present in the atmosphere. Bera et al., in one of their recent works, used multilinear regression analysis (MLR) and artificial neural networks (ANN) techniques for predicting the concentration of $PM_{2.5}$ over the atmosphere of Kolkata during the COVID-19 period [11]. In another work, Gokul et al. utilized AI techniques for the prediction of $PM_{2.5}$ in the ambient atmosphere of Hyderabad, another important Indian city [12]. Prasad et al. carried out a study to predict the concentration of $PM_{2.5}$ for the city of Vishakapatnam using the

MLR technique [13]. Continuous monitoring of air pollution in urban areas like NCR has become pertinent in modern times. On the other hand, regression analysis of these large data sets is also equally important. Moreover, due to some unavoidable circumstances, data collection is sometimes not feasible. During these periods, regression models can be quite helpful. In the present study, Indirapuram and Noida have been selected as the target stations for air pollution data analysis. Indirapuram is an urban area located in the Ghaziabad district of Uttar Pradesh. It is primarily a residential area in close proximity to Delhi. On the other hand, Noida is both a residential and industrial center. For both places, the concentration of $PM_{2.5}$ was chosen as the dependent variable, whereas the concentration of other pollutants was taken as the independent one. The MLR technique has been utilized to investigate the correlation between $PM_{2.5}$ and other pollutant concentrations. Finally, a model is formulated to predict the concentration of $PM_{2.5}$ and validated with a known set of data. This is in accordance with the previously conducted studies [11–13]. The study is extremely useful in the present scenario, as it may help to formulate a proper air quality management plan for the NCR.

2. Methodology

In the current study, MLR analysis is explored to find out the relation between the concentration levels of $PM_{2.5}$ and other particulate and gaseous pollutants existing in the ambient air. MLR is a statistical technique that analyzes the relationship between two or more variables and uses the information to estimate the value of the dependent variables. In multiple regression, the objective is to develop a model that describes a linear relationship between the dependent variable (y) and multiple independent variables (x_1, x_2, x_3…). Here, $PM_{2.5}$'s concentration is chosen as the dependent variable, whereas the concentration of PM_{10}, CO, SO_2, etc., has been taken as the independent variable. The MLR analysis has been performed using Microsoft (MS) Excel.

2.1. Data collection

All the data on the concentration of various pollutants like particulate matter ($PM_{2.5}$), carbon monoxide (CO), particulate matter (PM_{10}), nitrogen dioxide (NO_2), ground-level ozone (O_3), and sulfur dioxide (SO_2) are collected from the official website of the Central Pollution Control Board (CPCB) that comes under the Ministry of Environment, Forest and Climate Change, Government of India [14]. The daily data on the concentration of various pollutants are collected from two of the air quality monitoring stations of Ghaziabad, Uttar Pradesh, namely Indirapuram and Noida, for the years 2022 and 2023. Our study area focuses on NCR, so these stations satisfy the criteria. For both years (2022 and 2023), data at 3 pm was used for analysis. The website provides the maximum, minimum, and average concentrations of a particular pollutant at a definite time. Hence, taking the data for every day for two years also covers the daily variation. The data collected will be used in the MLR method to predict the future behavior, fluctuation, and seasonal pattern of the various pollutants and their individual and co-dependent contribution to the overall Air Quality Index for the region.

2.2. Data analysis

The temporal characteristics of concentrations of various pollutants are done for the Indirapuram and Noida regions for the years 2022 and 2023 by using regression analysis and Analysis of Variance (ANOVA). The coefficient of determination (R^2) is used as the standard for describing the nature of the fit. Its value ranges from 0 to +1. A higher R^2 value indicates a good fit and vice versa. Consequently, considering the best correlating independent variable, a model is developed for predicting $PM_{2.5}$ concentration. It was compared with the known set of data. Further, Pearson's correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE) were computed to check the suitability of the model.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \bar{y}_i|,$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y}_i)^2}.$$

Here, y_i and $\bar{y}_i$ represent observed $PM_{2.5}$ and estimated $PM_{2.5}$, respectively.

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(x_i - \bar{x})^2(y_i - \bar{y})^2}},$$

Here, y_i represents the value of dependent variables, x_i represents the values of independent variables, and $\bar{x}$ and $\bar{y}$ represent the mean of the independent variables and dependent variables of the samples.

3. Results and discussion

3.1. Trend analysis

3.1.1. Seasonal variation of particulate matters

From the data set collected from the CPCB website of various particulate matters like $PM_{2.5}$ and PM_{10}, the variation of both the particulate matters in all the months throughout the year 2022 was studied. To get an overall idea of the variation in the behavior of $PM_{2.5}$ and PM_{10}, we have plotted the graph of the maximum, mean, and standard deviation of these pollutants for all the months of the year 2022, which are shown below. **Figure 1** shows the variation of mean values throughout the year, while in **Figure 2**, the maximum value is presented. On the other hand, **Figures 3** and **4** represent the plots for the year 2023. The trends of both figures (**Figures 1** and **2**) are almost the same. Except in the monsoon time (July-September), the average PM level is always high (~ ≥ 300 μg/m^3). In **Figure 2**, the maximum value reaches 500 μg/m^3 during the worst months. High vehicular emissions, road dust, construction, and other industrial activities may be responsible for this. However, during the monsoon season, rain acts as a natural scrubber to reduce PM levels in the atmosphere.

The data analyzed has been compared with some of the other recent studies. In one of the studies, Gupta et al. described the seasonal variation of $PM_{2.5}$ for a period of nine years over this area [15]. The lowest concentration of $PM_{2.5}$ was reported during monsoon times, followed by the highest spiking in the post-monsoon duration. Six sources, including dust, vehicular emissions, biomass burning, combustion, emissions from industrial units, and Br-rich sources, are the major causes for the generation of high concentrations of $PM_{2.5}$ in Delhi-NCR. Chetna et al. reported the variation in $PM_{2.5}$ concentration over a long period of time (2007–2021) [16]. During this span of 15 years, the average concentration of $PM_{2.5}$ remained in the range of 125 ± 86 μg/m^3. The maximum value for the monthly average was observed in November, while a minimum value was obtained during August. Bawase et al., in their study [17], reported the chemical composition of $PM_{2.5}$ and PM_{10} during the period 2016–2017 in Delhi-NCR. It was seen that in winter months, the concentration reached its peak, while the minimum value was observed during summer months. Both $PM_{2.5}$ and PM_{10} have been found to be dominated by organic matter and sulfate nitrate ammonium ions.

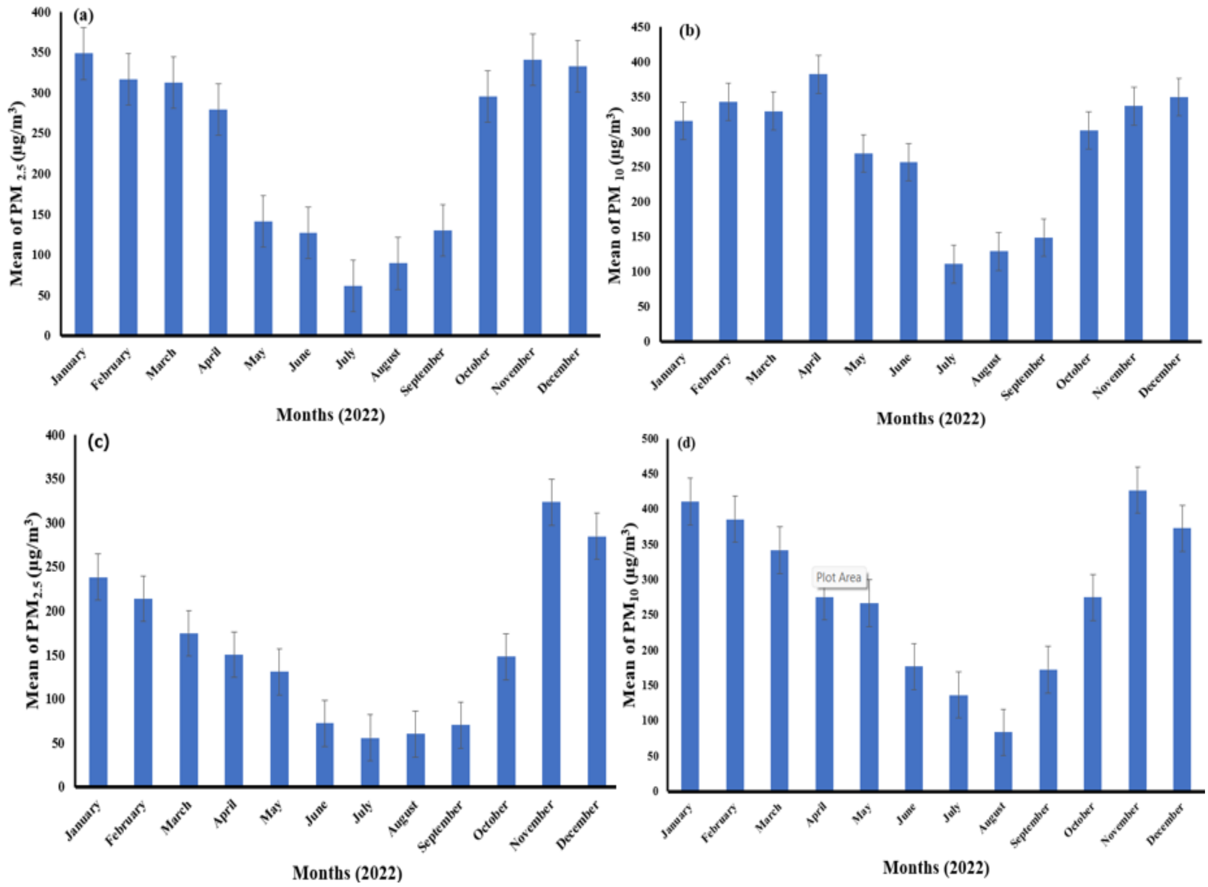


Figure 1. Monthly variation of concentration of $PM_{2.5}$ & PM_{10} (mean) in μg/m^3 of: **(a,b)** Indirapuram; **(c,d)** and Noida for the year 2022.

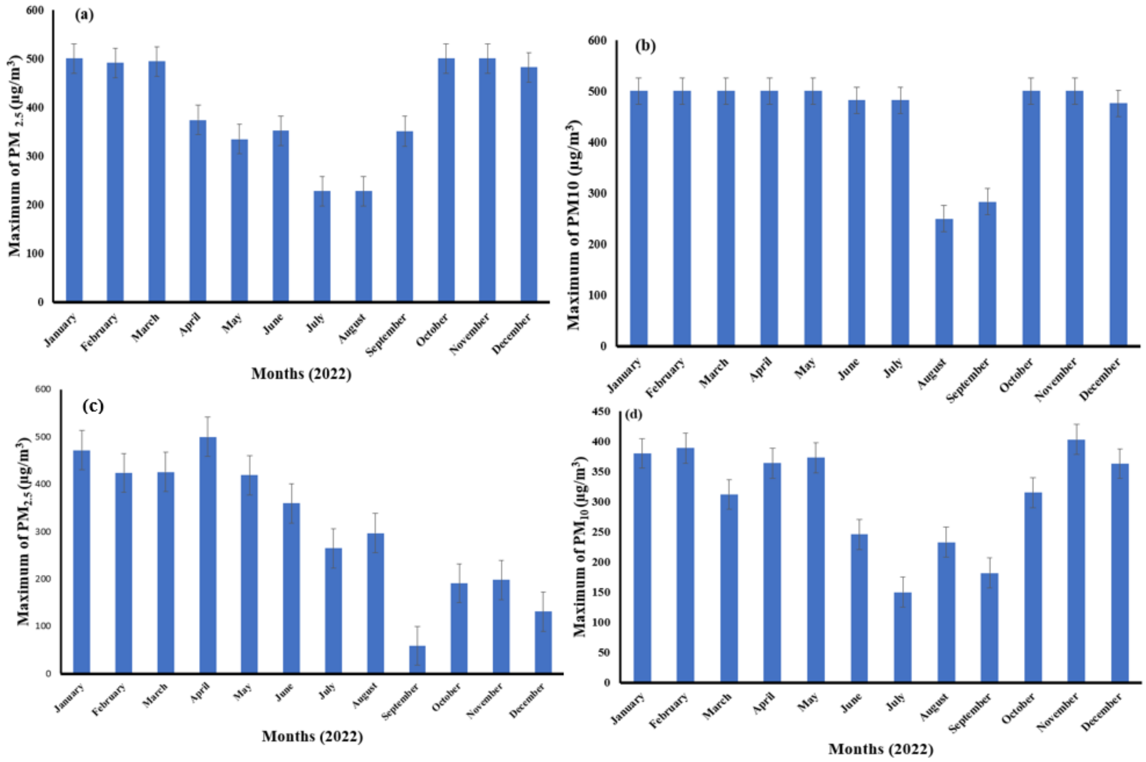


Figure 2. Monthly variation of concentration of $PM_{2.5}$ & PM_{10} (maximum) in µg/m^3 of: **(a,b)** Indira Puram; **(c,d)** and Noida for the year 2022.

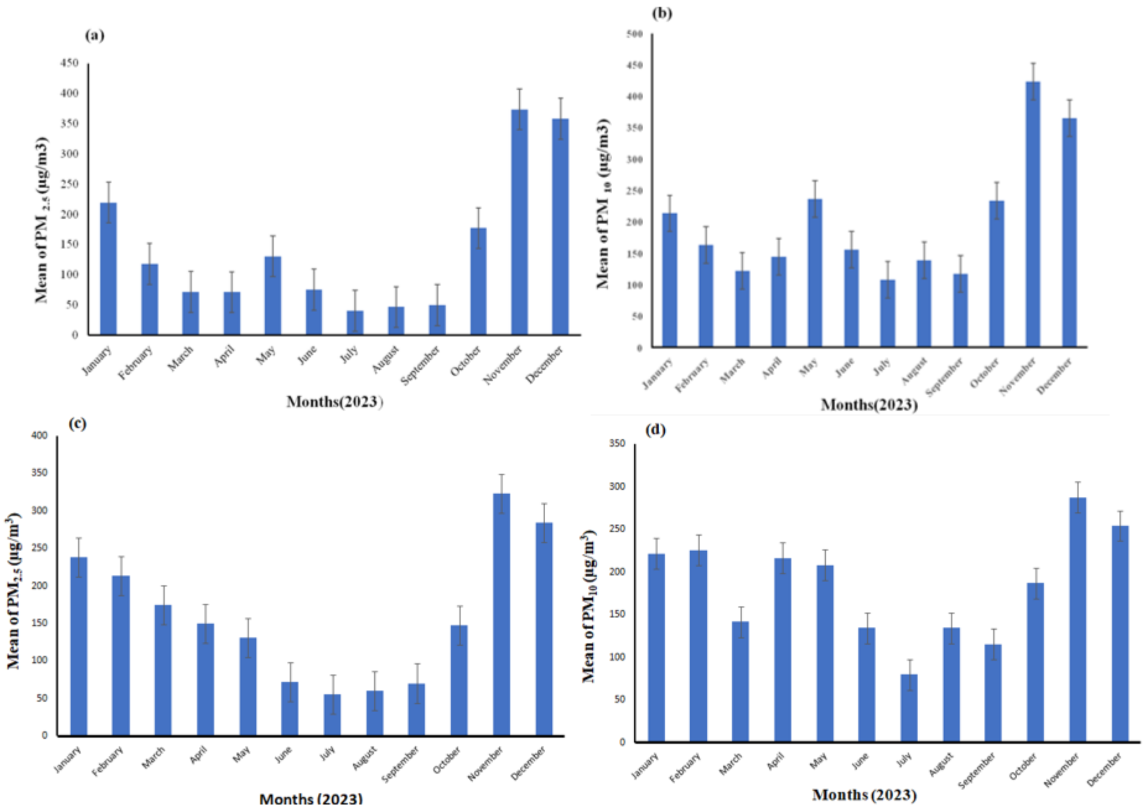


Figure 3. Monthly variation of concentration of $PM_{2.5}$ & PM_{10} (mean) in µg/m^3 of: **(a,b)** Indirapuram; **(c,d)** and Noida for the year 2023.

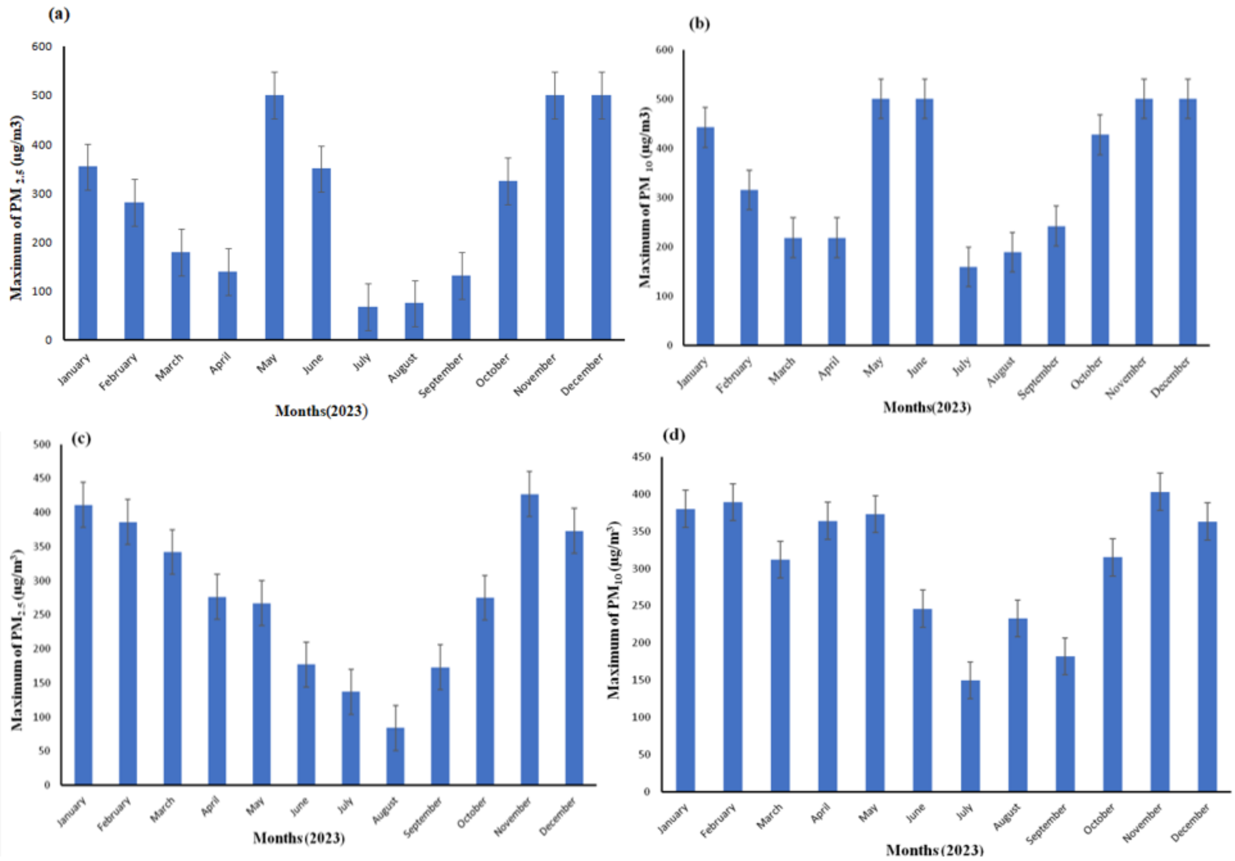
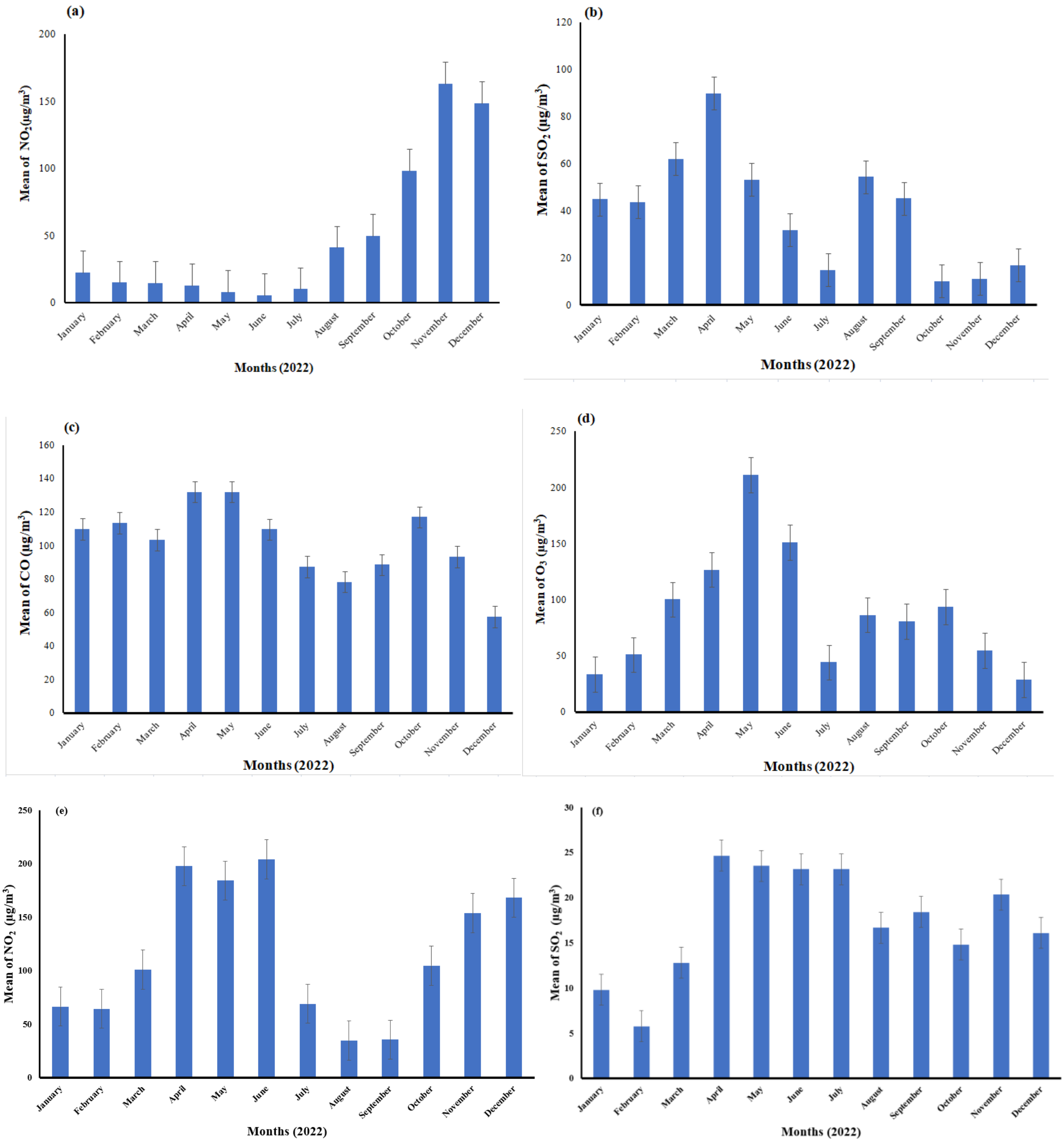


Figure 4. Monthly variation of concentration of $PM_{2.5}$ & PM_{10} (maximum) in µg/m^3 of: **(a,b)** Indirapuram; **(c,d)** and Noida for the year 2023.

3.1.2. Seasonal variation of gaseous pollutants

The concentration of four major gaseous pollutants, namely NO_2, SO_2, CO, and O_3, is studied. From the data set collected, the variation of all the gaseous pollutants is plotted for all the months throughout the year 2022. The figures (**Figures 5** and **6**) display the variation of mean concentration as well as the maximum concentration of these pollutants for the year 2022. In the case of NO_2, the average concentration is low (~≤ 100 µg/m^3) almost throughout the year except for the months of October-December. On the other hand, SO_2's concentration showed some spiking in the summer seasons. Except in the monsoon period, CO's concentration remained considerably high (≥ 100 µg/m^3).

Many articles of the current period describe the variation of gaseous pollutants in the NCR's atmosphere. Kumar et al. described the variability of ground-level ozone and NO_x in this area [18]. The maximum average concentration of NO_x was reported during the wintertime. Mandal et al. presented an overview regarding the seasonal and daily variation in the concentration of particulate matter and gaseous pollutants (PM_{10}, SO_2, NO_2, NH_3) over the Naraina industrial area of Delhi-NCR throughout the year 2017 [19]. The authors concluded that sometimes, the concentration of PM_{10} and NO_2 exceeded the prescribed limit. However, SO_2 and NH_3 concentrations have always been found within the limit. Tyagi et al. gave a detailed picture of the variability in the concentration of gaseous pollutants, namely NO_x, CO, and ground-level ozone, measured from seven air quality monitoring stations located in Delhi-NCR throughout the year 2014 [20]. The highest concentration of CO was observed during the night hours due to the nocturnal boundary conditions. On the other hand, the highest concentration of O_3 (ground level) was recorded during the daytime.



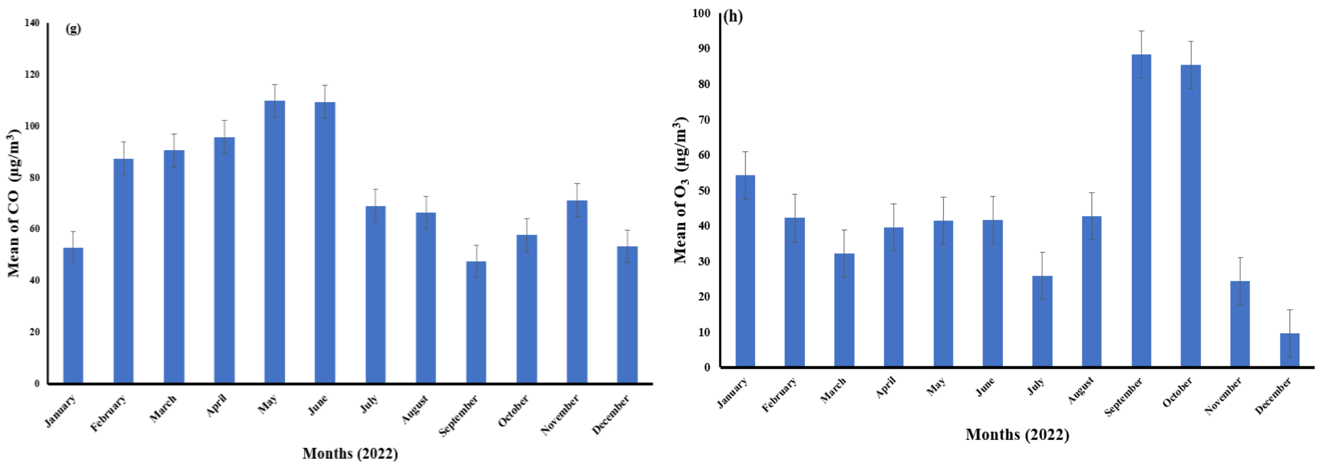
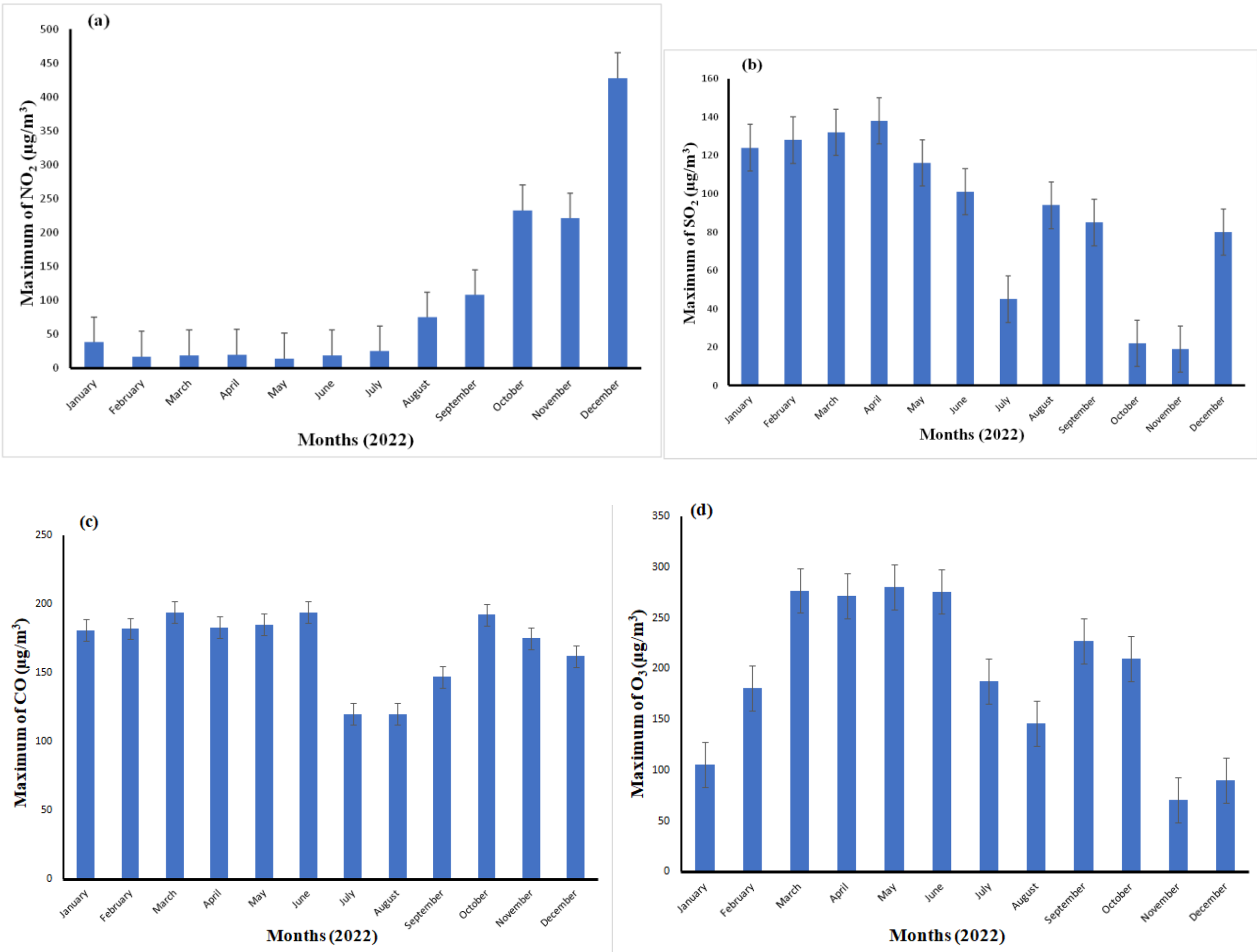


Figure 5. Mean concentration of (i) NO_2, (ii) SO_2, (iii) CO, (iv) O_3 in ($\mu g/m^3$) in 2022 (for: **(a–d)** Indirapuram; and **(e–h)** Noida).



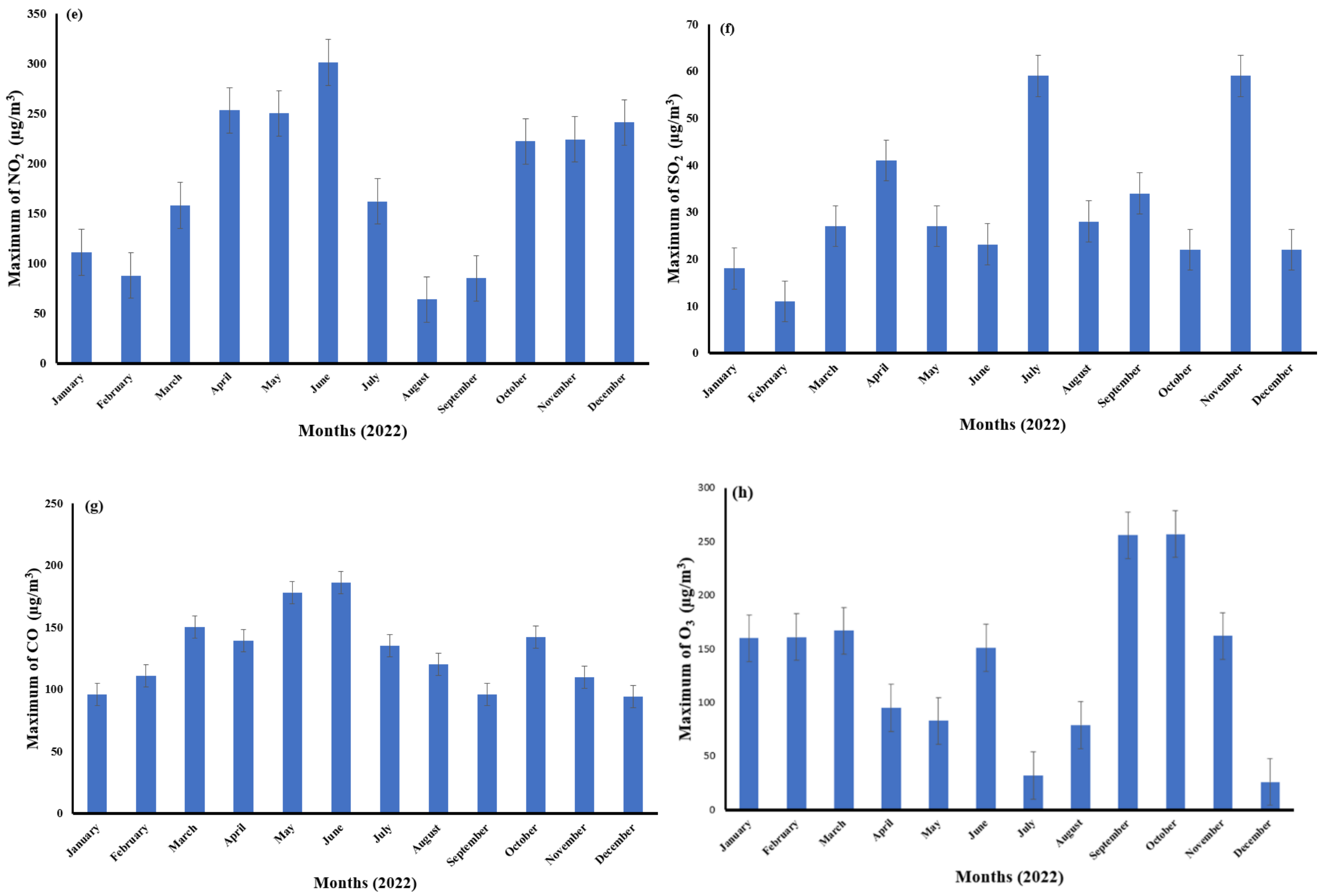


Figure 6. Maximum concentration of (i) NO_2, (ii) SO_2, (iii) CO, (iv) O_3 in (μg/m^3) in 2022 (for: **(a–d)** Indirapuram; and **(e–h)** Noida).

3.1.3. Correlation among concentrations of different pollutants

Undoubtedly, predicting the concentration of $PM_{2.5}$ is a complex task, as it often depends upon a number of factors [11]. Hence, to have a clear idea, it is pertinent to find out the correlation among the concentrations of different pollutants. So, Pearson's method was adopted to determine the dependency of various pollutants' concentrations on each other. The result is shown in **Tables 1** and **2**. It is clear that the correlation of $PM_{2.5}$'s concentration with other pollutants (except PM_{10}) is not good.

Table 1. Correlation matrix (pearson's method) showing the association between variables for Noida (2022).

Independent variable	**$PM_{2.5}$ (μg/m^3)**	**PM_{10} (μg/m^3)**	**NO_2 (μg/m^3)**	**O_3 (μg/m^3)**	**SO_2 (μg/m^3)**	**CO (μg/m^3)**
$PM_{2.5}$ (μg/m^3)	1	0.766	0.429	−0.039	0.113	−0.01
PM_{10} (μg/m^3)	0.766	1	0.533	−0.033	0.133	0.013
NO_2 (μg/m^3)	0.429	0.533	1	0.111	0.05	0.088
O_3 (μg/m^3)	−0.039	−0.033	0.111	1	−0.025	0.002
SO_2 (μg/m^3)	0.113	0.133	0.05	−0.025	1	0.045
CO (μg/m^3)	−0.01	0.013	0.088	0.002	0.045	1

Table 2. Correlation matrix (pearson's method) showing the association between variables for Indirapuram (2023).

Independent variable	$PM_{2.5}$ (μg/m^3)	PM_{10} (μg/m^3)	NO_2 (μg/m^3)	O_3 (μg/m^3)	SO_2 (μg/m^3)	CO (μg/m^3)
$PM_{2.5}$ (μg/m^3)	1	0.903	0.748	0.286	0.232	−0.279
PM_{10} (μg/m^3)	0.903	1	0.742	0.313	0.216	−0.160
NO_2 (μg/m^3)	0.748	0.742	1	0.277	0.197	−0.057
O_3 (μg/m^3)	0.286	0.313	0.277	1	0.276	0.055
SO_2 (μg/m^3)	0.232	0.216	0.197	0.276	1	0.262
CO (μg/m^3)	−0.279	−0.160	−0.057	0.055	0.262	1

3.2. Multi-linear regression (MLR) analysis

3.2.1. Variation of concentration of $PM_{2.5}$ with concentration of other pollutants (singly)

Initially, the concentration of $PM_{2.5}$ (μg/m^3) over different time periods has been taken as the dependent variable and other pollutants (individually) as the independent one. It is quite interesting to observe that there is a good correlation between $PM_{2.5}$ and PM_{10}, a moderate correlation of $PM_{2.5}$ with SO_2 and NO_2, but a poor correlation with CO and O_3.

A good correlation between $PM_{2.5}$ and PM_{10} was observed, where the maximum value for adjusted R^2 = 0.92 during the month of September-October 2022, whereas correlation with CO and O_3 was found to be poor. **Table 3** highlights the detailed value of R^2 obtained for the correlation of $PM_{2.5}$ with each of the pollutants. Besides, plots for the data sets where $PM_{2.5}$'s concentration correlates well with the concentration of PM_{10} are shown in **Figures 7** and **8** for the years 2022 and 2023.

Table 3. Adjusted R^2 values of $PM_{2.5}$ against other independent variables.

Dependent variable	Independent variable	Duration	Adj. R^2			
			2022		2023	
			Indirapuram	Noida	Indirapuram	Noida
$PM_{2.5}$	PM_{10}	Jan–Feb	0.64	0.74	0.97	0.65
		Mar–Apr	0.34	0.58	0.74	0.16
		May–Jun	0.10	0.45	0.59	0.63
		Jul–Aug	0.50	0.51	0.26	0.32
		Sep–Oct	0.92	0.91	0.77	0.71
		Nov–Dec	0.67	0.63	−0.01	0.73
	CO	Jan–Feb	0.44	−0.02	−0.01	0.02
		Mar–Apr	0.31	−0.01	−0.01	−0.02
		May–Jun	0.30	−0.02	0.23	0.23
		Jul–Aug	0.09	0.01	0.00	0.02
		Sep–Oct	0.17	0.01	0.11	−0.02
		Nov–Dec	0.10	−0.02	−0.01	0.43

Table 3. (*Continued*).

Dependent variable	**Independent variable**	**Duration**	**Adj. R^2**			
			2022		**2023**	
			Indirapuram	**Noida**	**Indirapuram**	**Noida**
$PM_{2.5}$	SO_2	Jan–Feb	0.01	0.12	0.04	0.01
		Mar–Apr	−0.01	−0.01	0.00	0.02
		May–Jun	0.17	0.01	0.00	−0.02
		Jul–Aug	0.05	−0.03	−0.01	−0.02
		Sep–Oct	0.06	0.03	0.11	0.02
		Nov–Dec	0.07	−0.01	−0.01	0.01
	NO_2	Jan–Feb	−0.01	−0.01	0.19	0.24
		Mar–Apr	−0.008	0.05	0.62	0.07
		May–Jun	−0.004	0.32	0.30	−0.03
		Jul–Aug	0.156	0.03	−0.01	−0.01
		Sep–Oct	0.64	0.40	0.61	0.70
		Nov–Dec	0.11	0.10	0.17	0.13
	O_3	Jan–Feb	0.03	0.22	0.00	0.03
		Mar–Apr	0.01	−0.00	0.04	−0.02
		May–Jun	0.09	0.06	0.03	−0.02
		Jul–Aug	0.001	−0.02	0.45	0.06
		Sep–Oct	0.18	0.30	0.56	0.01
		Nov–Dec	0.05	−0.01	−0.02	−0.00

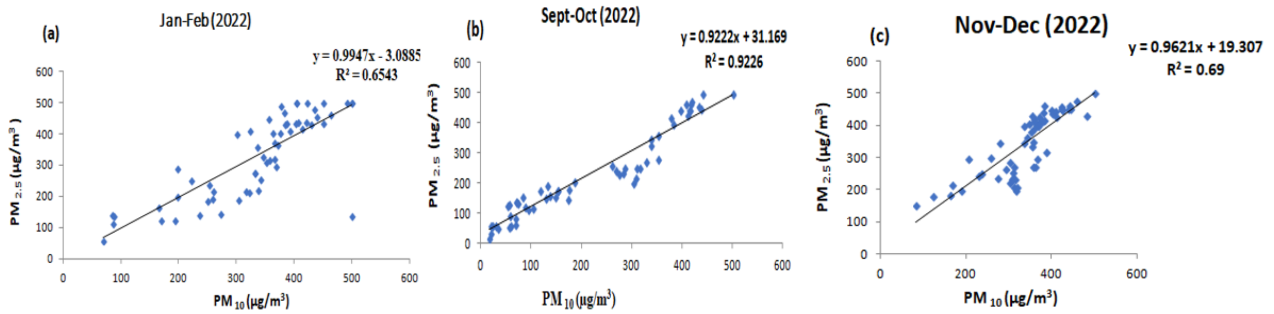


Figure 7. Linear plot of concentration of $PM_{2.5}$ vs. PM_{10} in the months of: **(a)** Jan–Feb; **(b)** Sept–Oct; **(c)** Nov–Dec (2022).

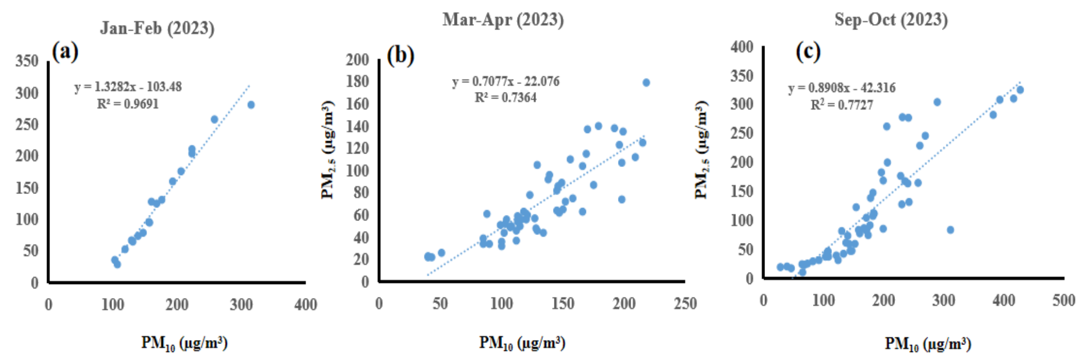


Figure 8. Linear plot of concentration of $PM_{2.5}$ vs PM_{10} in the months of: **(a)** Jan–Feb; **(b)** Mar–Apr; **(c)** Sep–Oct (2023).

3.2.2. Correlation of concentration of $PM_{2.5}$ with PM_{10} and other pollutants

Table 4. Adjusted R^2 values of $PM_{2.5}$ against other independent variables.

Dependent variable	Independent variable	Duration	Adj. R^2			
			2022		2023	
			Indirapuram	Noida	Indirapuram	Noida
$PM_{2.5}$	PM_{10} & O_3	Jan–Feb	0.64	0.77	0.82	0.64
		Mar–Apr	0.34	0.57	0.74	0.13
		May–Jun	0.19	0.54	0.53	0.63
		Jul–Aug	0.53	0.52	0.33	0.32
		Sep–Oct	0.93	0.90	0.81	0.72
		Nov–Dec	0.67	0.70	0.70	0.75
	PM_{10} & SO_2	Jan–Feb	0.64	0.77	0.81	0.63
		Mar–Apr	0.35	0.55	0.75	0.30
		May–Jun	0.20	0.49	0.61	0.63
		Jul–Aug	0.50	0.57	0.25	0.33
		Sep–Oct	0.93	0.92	0.77	0.69
		Nov–Dec	0.67	0.64	0.49	0.75
	PM_{10} & CO	Jan–Feb	0.64	0.73	0.81	0.64
		Mar–Apr	0.35	0.58	0.73	0.24
		May–Jun	0.33	0.45	0.64	0.65
		Jul–Aug	0.51	0.61	0.35	0.32
		Sep–Oct	0.96	0.91	0.77	0.71
		Nov–Dec	0.63	0.62	0.49	0.72
	PM_{10} & NO_2	Jan–Feb	0.65	0.68	0.80	0.65
		Mar–Apr	0.35	0.58	0.74	0.71
		May–Jun	0.096	0.53	0.65	0.61
		Jul–Aug	0.46	0.50	0.25	0.33
		Sep–Oct	0.96	0.91	0.78	0.78
		Nov–Dec	0.61	0.63	0.52	0.74

In the second case, an attempt was made to correlate the concentration of $PM_{2.5}$ with that of PM_{10} and other pollutants. In this case, it was seen that the correlation of $PM_{2.5}$ with PM_{10} and O_3 was found to be comparatively good, while CO and O_3 showed a poor correlation. The correlation between $PM_{2.5}$ and the combination of the other two pollutants was found using MLR analysis and has been reported in **Table 4**. Though there was a poor correlation between $PM_{2.5}$ and CO and O_3 when a correlation between $PM_{2.5}$ and a single pollutant was observed, a good correlation was observed when CO and O_3 were combined with PM_{10}.

3.2.3. Model validation

Model validation is the process of assessing how accurately the model can predict the dependent variable. This is done by using statistical measures to quantify the difference between the predicted and observed values. The reliability of air quality models is generally based on their accuracy and applicability. In this study, we mainly used MLR for model validation purposes. Apart from these, RMSE, MAE, and Pearson's correlation coefficient (R) have been used. RMSE and MAE are the measurements of the average error, and lower values are better. RMSE is the square root of the average squared difference between the predicted and the observed values, which are predicted $PM_{2.5}$ and observed $PM_{2.5}$ for this study. The obtained value of RMSE is as low as 36.34 for Indirapuram and 31.61 for Noida. Also, MAE = 25.87 and 20.38 have been obtained for Indirapuram and Noida, respectively, which proves to be a reliable method for the prediction of $PM_{2.5}$. R^2 is the proportion of the variance in the dependent variable that can be predicted from the independent variables. Higher R^2 means a stronger association between the variables [10]. It has been found that there is a strong association between $PM_{2.5}$ and PM_{10} in the year 2022 for both the stations, i.e., Indirapuram and Noida, with $R^2 = 0.93$ and 0.9, respectively. Pearson correlation coefficient (R) is used to measure the strength and direction of a linear relationship between two variables. The value of R ranges from −1 to +1. +1 and −1 indicate a perfect positive and negative correlation, respectively. 0 indicates no linear correlation. R is used to describe the linear correlation. The correlation is stronger when the value of R is closer to 1 or −1 [21]. In both stations, there is a strong correlation between $PM_{2.5}$ and PM_{10}, which is evident from the Pearson correlation coefficient (R) = 0.903 and 0.766 for Indirapuram in 2023 and Noida in 2022, respectively.

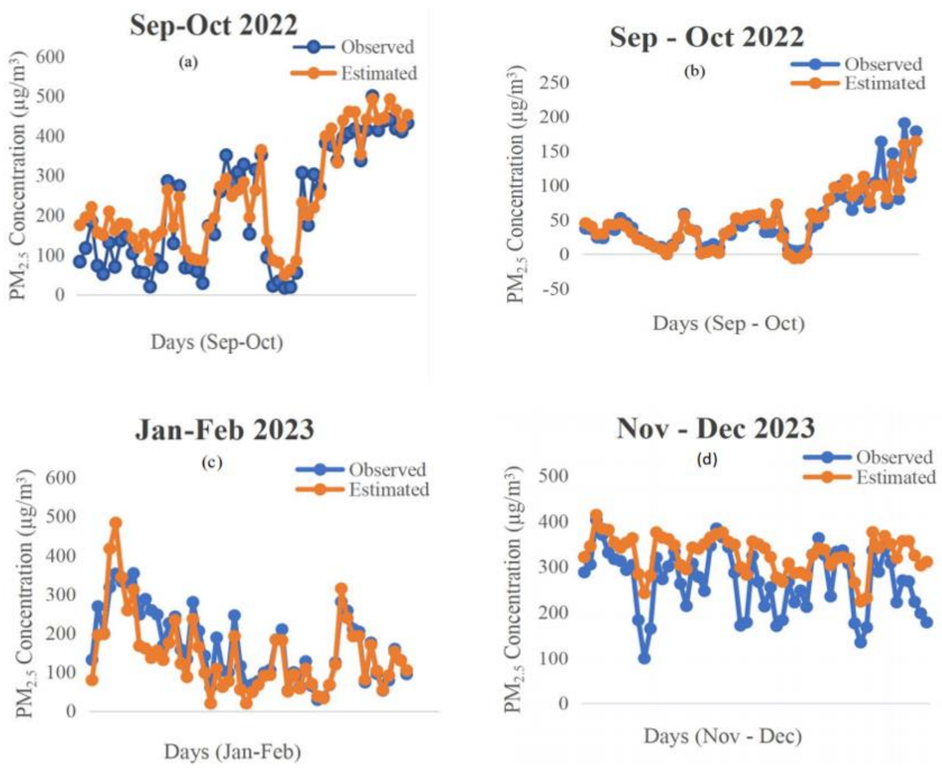


Figure 9. Observed and estimated value using MLR for: **(a)** Sep–Oct 2022, Indirapuram; **(b)** Sep–Oct 2022, Noida; **(c)** Jan–Feb 2023, Indirapuram; **(d)** Nov–Dec 2023, Noida.

In addition to these standard methods, other statistical evaluations can be used, such as bias, normalized mean absolute error (NMAE), index of agreement (IA), detection rate (DR), and false alarm rate (FAR). Bias is defined as the average error for the given set of observations and predictions. NMAE is the ratio of MAE to the mean of the observed concentrations, generally expressed as a percentage. The IA is a dimensionless indicator between zero and unity, with 1.0 representing perfect predictions. When predictions from two or more data sets with dissimilar means and variances are compared, IA is more accurate than NMAE. The detection rate (DR) is defined as the fraction of observed NAAQS exceedances correctly predicted by the model, while the false alarm rate (FAR) is the fraction of alarms in which the observed concentration was less than the alarm threshold [22]. **Figure 9** shows the graphical representation for the comparison between the observed $PM_{2.5}$ and estimated $PM_{2.5}$ using MLR (best fit).

3.2.4. Future scope

The current study primarily focuses on analyzing the seasonal variation of the concentration of particulate matter and gaseous pollutants over the atmosphere of two stations, namely Indirapuram and Noida, in the Delhi-NCR zone for the recent years of 2022 and 2023. As $PM_{2.5}$ is one of the most alarming air pollutants, an attempt has been made to find out the correlation between its concentration and other pollutants' concentrations. It is already mentioned that the MLR technique has been chosen to find the correlation as well as for further model development and prediction purposes. However, the MLR technique is best suited for data analysis when a linear relationship between the variables exists. However, different reports of recent times reveal that MLR often does not show the best fit in the case of air pollution data.

In our case also, in several cases, the data does not show a good fit with the linear trend. Hence, the future scope includes:

(1) More years shall be studied to get an idea regarding the variation in concentration of pollutants over the years.

(2) Other stations in the region may be included to get a better insight regarding the purpose of model validation.

(3) Study for model development using techniques like artificial neural networks (ANN).

(4) Algorithm based study may be conducted.

4. Conclusion

$PM_{2.5}$ is an important constituent of maintaining air quality standards in the megacity's atmosphere. Therefore, its continuous monitoring is an important research topic of contemporary times. Apart from data collection, regression analysis of the obtained data is also an important task. In this work, an attempt was made to correlate the concentration of $PM_{2.5}$ with that of other pollutants over the atmosphere in Indirapuram and Noida (NCR) with the MLR technique. The detailed analysis showed that when one pollutant is taken as the independent variable, PM_{10}'s concentration fits well with that of $PM_{2.5}$. However, when binary independent variables were chosen, PM_{10} and ground-level O_3's concentration showed the best correlation with that of $PM_{2.5}$. However, there is a lack of fit in many cases with the linear trend. Hence, other models apart from linear ones need to be studied in the near future.

Author contributions: Conceptualization, AP and PM; methodology, AP; software, PM; validation, AP, PM and KUT; formal analysis, SB; investigation, AP; resources, AP; data curation, SS; writing—original draft preparation, AP, PM, SS and KUT; writing—review and editing, PM; visualization, SS; supervision, SB. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Vaishali, Verma G, Das RM. Influence of Temperature and Relative Humidity on PM2.5 Concentration over Delhi. MAPAN. 2023; 38(3): 759-769. doi: 10.1007/s12647-023-00656-8
2. Hama SML, Kumar P, Harrison RM, et al. Four-year assessment of ambient particulate matter and trace gases in the Delhi-NCR region of India. Sustainable Cities and Society. 2020; 54: 102003. doi: 10.1016/j.scs.2019.102003
3. Chen Z, Li S, Liu C. Challenges and Potentials: Environmental Assessment of Particulate Matter in Spaces Under Highway Viaducts. Atmosphere. 2024; 15(11): 1325. doi: 10.3390/atmos15111325
4. Selokar A, Ramachandran B, Elangovan KN, et al. PM 2.5 particulate matter and its effects in Delhi/NCR. Materials Today: Proceedings. 2020; 33: 4566-4572. doi: 10.1016/j.matpr.2020.08.187
5. Kaushik N, Das RM. Investigation of NOx and related secondary pollutants at Anand Vihar, one of the most polluted area of Delhi. Urban Climate. 2023; 52: 101747. doi: 10.1016/j.uclim.2023.101747
6. Goenka R, Thakur J, Taori A, et al. The prevailing smog conditions over the Delhi-NCR during the 2022 post monsoon. Advances in Space Research. 2024; 73(5): 2609-2617. doi: 10.1016/j.asr.2023.12.018
7. Singh AK, Pathak AK, Saini G. Premature mortality risk and associated economic loss assessment due to PM2.5 exposure in Delhi, India during 2015–2019. Air Quality, Atmosphere & Health. 2024; 17(9): 1867-1883. doi: 10.1007/s11869-024-01550-1
8. Singh V, Singh S, Biswal A. Exceedances and trends of particulate matter (PM2.5) in five Indian megacities. Science of The Total Environment. 2021; 750: 141461. doi: 10.1016/j.scitotenv.2020.141461
9. Sharma SK, Mandal TK. Elemental Composition and Sources of Fine Particulate Matter (PM2.5) in Delhi, India. Bulletin of Environmental Contamination and Toxicology. 2023; 110(3). doi: 10.1007/s00128-023-03707-7
10. Gupta L, Joshi S, Habib G, et al. Characteristics and atmospheric processes of water-soluble ions in PM2.5 and PM10 over an industrial city in the National Capital Region (NCR) of India. Atmospheric Environment. 2023; 312: 120020. doi: 10.1016/j.atmosenv.2023.120020
11. Bera B, Bhattacharjee S, Sengupta N, et al. PM2.5 concentration prediction during COVID-19 lockdown over Kolkata metropolitan city, India using MLR and ANN models. Environmental Challenges. 2021; 4: 100155. doi: 10.1016/j.envc.2021.100155
12. Gokul PR, Mathew A, Bhosale A, et al. Spatio-temporal air quality analysis and PM2.5 prediction over Hyderabad City, India using artificial intelligence techniques. Ecological Informatics. 2023; 76: 102067. doi: 10.1016/j.ecoinf.2023.102067
13. Prasad NVK, Sarma MSSRKN, Sasikala P, et al. Regression Model to Analyse Air Pollutants Over a Coastal Industrial Station Visakhapatnam (India). International Journal of Data Science. 2020; 1(2): 107-113. doi: 10.18517/ijods.1.2.107-113.2020
14. Central Pollution Control Board. Central Pollution Control Board. Available online: http://cpcb.nic.in/ (accessed on 10 November 2024).
15. Gupta S, Sharma SK, Tiwari P, et al. Insight Study of Trace Elements in PM2.5 During Nine Years in Delhi, India: Seasonal Variation, Source Apportionment, and Health Risks Assessment. Archives of Environmental Contamination and Toxicology. 2024; 86(4): 393-409. doi: 10.1007/s00244-024-01070-0
16. Chetna, Dhaka SK, Longiany G, et al. Trends and Variability of PM2.5 at Different Time Scales over Delhi: Long-term Analysis 2007–2021. Aerosol and Air Quality Research. 2023; 23(5): 220191. doi: 10.4209/aaqr.220191
17. Bawase M, Sathe Y, Khandaskar H, et al. Chemical composition and source attribution of PM2.5 and PM10 in Delhi-National Capital Region (NCR) of India: results from an extensive seasonal campaign. Journal of Atmospheric Chemistry. 2021; 78(1): 35-58. doi: 10.1007/s10874-020-09412-7

18. Kumar A, Singh D, Singh BP, et al. Spatial and temporal variability of surface ozone and nitrogen oxides in urban and rural ambient air of Delhi-NCR, India. Air Quality, Atmosphere & Health. 2014; 8(4): 391-399. doi: 10.1007/s11869-014-0309-0
19. Mandal P, Sarkar R, Kamal N, et al. Diurnal and Seasonal Variation of Atmospheric Particulate Matter and Trace Gases in Industrial Area of Delhi: A Study. Bulletin of Environmental Contamination and Toxicology. 2022; 109(3): 488-494. doi: 10.1007/s00128-022-03572-w
20. Tyagi S, Tiwari S, Mishra A, et al. Spatial variability of concentrations of gaseous pollutants across the National Capital Region of Delhi, India. Atmospheric Pollution Research. 2016; 7(5): 808-816. doi: 10.1016/j.apr.2016.04.008
21. Ni XY, Huang H, Du WP. Relevance analysis and short-term prediction of PM 2.5 concentrations in Beijing based on multi-source data. Atmospheric Environment. 2017; 150: 146-161. doi: 10.1016/j.atmosenv.2016.11.054
22. Cobourn WG. An enhanced PM2.5 air quality forecast model based on nonlinear regression and back-trajectory concentrations. Atmospheric Environment. 2010; 44(25): 3015-3023. doi: 10.1016/j.atmosenv.2010.05.009