Indian Journal of Pure & Applied Physics
Vol. 62, December 2024, pp. 1138-1149
DOI: 10.56042/ijpap.v62i12.13527

Artificial Intelligence Based Investigation for the Impact of High $PM_{2.5}$ Concentration on Cloud Parameters over the Polluted Central IGP location, Kanpur

Pradeep Kumar Verma[a,b*], A K Srivastava[a], S P Shukla[b], V Pathak[c], Amarendra Singh[d], Bharat Ji Mehrotra[e] & Manoj K Srivastava[e]

[a]Indian Institute of Tropical Meteorology, Ministry of Earth Sciences, New Delhi 110 060, India

[b]Rajkiya Engineering College, Banda, UP 210 201, India

[c]Institute of Engineering and Technology, Lucknow, UP 226 021, India

[d]Indian Institute of Technology, New Delhi 110 016, India

[e]Department of Geophysics, Banaras Hindu University, Varanasi, UP 221 005, India

Received 8 September 2024; accepted 6 November 2024

This research focuses on using artificial neural network (ANN) models to assess daily surface $PM_{2.5}$ concentrations by incorporating aerosol optical depth (AOD) and cloud parameters from the Moderate Resolution Imaging Spectroradiometer (MODIS), along with meteorological data, for the period from January 2017 to December 2021 over Kanpur. For this exercise, three ANN models were utilized: ANN1 (1 Layer, 14 Neurons), ANN2 (2 Layers, 14, 28 Neurons), and ANN3 (1 Layer, 14, 28, 14 Neurons). Statistical tests such as FAC2, MGE, NMB, MAPE, RMSE, R, and COE were conducted to validate the models. Initial results show that the ANN1 performed the best. The study also examined spatial and temporal changes to observe variations in $PM_{2.5}$, AOD, and various cloud properties, including water vapor (WV), cloud effective radius (CER), cloud fraction (CF), cloud liquid water path (CLWP), cloud optical depth (COD), cloud top pressure (CTP), and cloud top temperature (CTT) on a seasonal and annual basis, as well as during high $PM_{2.5}$ concentration conditions. During the study, the average daily $PM_{2.5}$ was found to be approximately 100 μg/m³ (ranging from 0.45 to 470.23 μg/m³), while the average AOD was 0.79 (ranging from 0.09 to 3.55). High $PM_{2.5}$ concentrations (three to five times higher than the NAAQS annual limit) significantly influenced crucial cloud microphysical properties. The research findings aid in estimating $PM_{2.5}$ using satellite-retrieved AOD and meteorological data, providing insights into aerosol and cloud properties variability during high pollution events in the heavily polluted city of Kanpur, India.

Keywords: $PM_{2.5}$; MODIS; AOD; Cloud properties; Artificial neural network

1 Introduction

Over the past two decades, a significant amount of global research has shown that atmospheric particles, whether they arise from natural or anthropogenic sources, have a direct impact on environmental degradation and human health. Due to rapid urbanization and economic growth over the past two decades, air quality has deteriorated significantly, and posed a major threat to human health. The particulate matter ($PM_{2.5}$), one of the major contributors to air quality, directly affects the health of populations in areas with poor air quality. High $PM_{2.5}$ levels can also alter cloud properties by serving as cloud condensation nuclei (CCN) that influence cloud formation and alter visibility. Research studies done in the past advocate that monitoring $PM_{2.5}$ is important for assessment of the air quality and conducting environmental health research. For understanding air quality, the assessment of $PM_{2.5}$ is an important parameter for any polluted region. However, the ground based $PM_{2.5}$ observation network is not dense enough to assess the impact. Estimating $PM_{2.5}$ by satellite-retrieved aerosol optical depth (AOD) can fill the data gap left by ground level monitors and widen coverage to rural and forest areas.

Additionally, $PM_{2.5}$ has a significant impact on global climate change by altering the Earth's radiation balance, affecting cloud formation, and causing changes in temperature and precipitation patterns[1-3]. The problems caused by increased urbanization as a result of population growth, industrialization, and inadequate implementation of various emission rules are exacerbated by the fact that these factors are all occurring simultaneously[4]. The trend analysis of

*Corresponding author: (E-mail: pkverma.iet@gmail.com)]

$PM_{2.5}$ concentrations in several Indian cities from 1998 to 2015 using satellite data found a rapid increase in $PM_{2.5}$ exposure over this period, with the highest concentrations observed in the northern and eastern regions of India[5]. It is important to highlight that $PM_{2.5}$ is the most closely monitored air pollutant in the ambient atmosphere since they are frequently found to exceed national guidelines[6,7]. Among all the air pollutants, $PM_{2.5}$ was found to induce significant health impacts in addition to numerous climate-related impacts, and it was anticipated to have a substantial association with the daily mortality rate[8]. Several studies report that $PM_{2.5}$ is a major environmental peril in most parts of the world, particularly in South Asian countries like India, Pakistan, and Bangladesh. More than 2 million deaths each year can be attributed to outdoor air pollution, with $PM_{2.5}$ being a significant factor[9-10].

Aerosols are tiny liquid or solid particles suspended in the atmosphere and are a major component of smog, which degrades air quality. They originate from urban and industrial sources, as well as natural events like volcanic eruptions, dust storms, and forest fires. The Indo-Gangetic Plain (IGP) is a region of high aerosol loading, recognized as a global hotspot owing to the intense agricultural residue and biomass burning and transport, as well as locally generated mineral dust[11,12]. Kanpur, a significantly polluted location in the middle of the IGP region, experiences significant challenges from both local and regional air quality issues, mostly in winter season[13]. So, the high aerosol loading can expand contrary to acute and chronic health problems due to the inhalation of tiny aerosol particles in the human respiratory system[14]. The increase in aerosols results in a rise in cloud condensation nuclei (CCN) concentrations, which in turn leads to the formation of clouds with a higher density of droplets. This causes a reduction in the size of the individual droplets while maintaining the same amount of liquid water in the cloud, and is commonly referred to as the first indirect effect or the Twomey effect[15]. The presence of smaller cloud droplets can alter the efficiency of rainfall, leading to an increase in liquid water content, cloud cover, and cloud lifetime, referred to as the second indirect effect or the cloud lifetime effect[15]. Over the years, numerous studies have been conducted to understand the potential impact of atmospheric aerosols on cloud properties, and consequently, on the Earth's climate system[3,16-19]. High aerosol concentrations can lead to the formation of more intense clouds and increased convection through a feedback mechanism[20,21]. The ability of aerosol particles to influence cloud formation is dependent on their chemical composition and particle size, smoke and haze generally lead to cloud inhibition while polluted continental aerosols can contribute to cloud invigoration[22]. Although ground measurements are the precise observations of pollutant concentration, it is quite challenging to monitor over the vast area of India, which covers 3.287 million sq. km. with a ground network of instruments[23], the application of earth observation satellite data is more economical than ground measurements.

Aerosol Optical Depth (AOD) measures the extinction of solar radiation by aerosol particles in the atmosphere, indicating how much light is absorbed or scattered. AOD measures the total aerosol loading along the path of sunlight through the atmosphere. AOD derived from satellites offers an alternative method to access ground level $PM_{2.5}$ concentrations on both regional and global scales[24,25]. Machine learning models, such as Artificial Neural Networks (ANN), offer a data-driven approach that adapts to the high variability in air pollution data without requiring predefined relationships, unlike traditional models. By effectively handling large datasets with multiple inputs, ANNs models improve prediction accuracy by considering the influence of weather conditions on aerosol behavior and $PM_{2.5}$ levels. These models enable fast real-time monitoring and short-term forecasting, which are crucial for effective air quality management, timely public health alerts, and informed policy-making. In recent years, many researchers have started utilizing ANN algorithms to estimate quantities of air pollutants[26,27]. The capacity of ANNs to make regressive approximations of nonlinear functions in high dimensional spaces is crucial to their success because it is absent from other statistical methods. When compared to deterministic models, ANNs require a smaller amount of input data and fewer computing resources, and their application is both straightforward and adaptable. The back-propagation neural network (BPNN) is one of the most frequently employed models and is considered to be mature, with extensive use in forecasting and investigation of air pollution[28]. Multiple Linear Regression (MLR) and ANN are commonly used in multidisciplinary modeling and forecasting. Unlike MLR, ANN effectively captures nonlinear variability in environmental datasets. The neuron set is a key

component of ANN, enabling it to process and interpret input data before passing it to the next stage for output. ANN integrates pattern recognition techniques designed specifically for environmental applications. ANN has achieved excellent results in predicting PM_{10} and $PM_{2.5}$ levels[24,25,29], exceeding the accuracy of other prediction models.

This paper seeks to address the knowledge gap in understanding the relationships between aerosol and cloud properties, with a specific focus on regional impacts over Kanpur, a metropolitan city that is marked as one of the most polluted cities both nationally and globally[29]. The key objective is to predict daily surface $PM_{2.5}$ concentrations using different ANN models over Kanpur and simultaneously investigate the variation of $PM_{2.5}$ and cloud parameters through a comprehensive analysis of inter-annual and seasonal variations from 2017 to 2021. This research further endeavors to unveil the variability of aerosol and cloud parameters during high $PM_{2.5}$ concentrations at Kanpur, contributing to a deeper understanding of this complex phenomenon.

2 Study area

The study site, Kanpur (26.45°N, 80.33°E), with an altitude of 142 m above mean sea level, a heavily polluted urban city situated in the middle of IGP in north India, which is known for its poor air quality due to a high concentration of anthropogenic emissions[30]. In 2018, Kanpur was recognized by the WHO as one of the most polluted cities in the world. According to the 2011 Indian Census, the city has a total population of 29.2 lakh people who live in an area of 403.7 km^2. Kanpur was formerly referred to as the "Manchester of India," but regrettably, through the years it has developed a reputation for being a filthy and polluted city. The city has relatively flat terrain with the Himalayan Mountains located to the north and northeast. The temperature ranges from 26 °C in January to 36 °C in May, with an annual mean of 30 °C. Meanwhile, the relative humidity varies from 45% in May to 90% in August. The boundary layer height varies with season and meteorological conditions but is generally low during the winter season, which can contribute to high levels of air pollution[31,32]. Seasons in Kanpur are defined as per IMD: Winter (December-February), Pre-monsoon/Summer (March-June), Monsoon (July-September), and Post-monsoon/Autumn (October-November). Fig. 1 shows a synoptic plot, with atmospheric temperature in °C; and relative humidity in contour line; the wind speed and direction are given by an arrow, and the study location is shown with a blue dot.

3 Data and methodology

3.1 Ground-based measurement of $PM_{2.5}$

Air pollutant monitoring data are primarily sourced from ground monitoring stations, which are often limited. In India, the Central Pollution Control Board (CPCB) monitors ground-level air pollutants at various locations as part of the National Air Quality Monitoring Programme (NAMP), following NAAQS policy standards. The continuous ambient air quality monitoring (CAAQM) stations are used by the CPCB to track the air pollution levels across various cities in India. The CAAQM stations employed monitoring techniques such as Beta Attenuation Monitors (BAMs) and tapered element oscillating microbalance (TEOM) to ensure comprehensive and accurate monitoring of $PM_{2.5}$ in ambient air[33]. For this study, the average daily mass concentration of $PM_{2.5}$ for Kanpur was collected for the period January 2017 to December 2021 (https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing/data). Fig. 1 represents the location of the monitoring station that is included in this investigation.

3.2 Satellite-based measurement of aerosol and cloud parameters

The data retrieved from Moderate Resolution Imaging Spectroradiometer (MODIS) observation

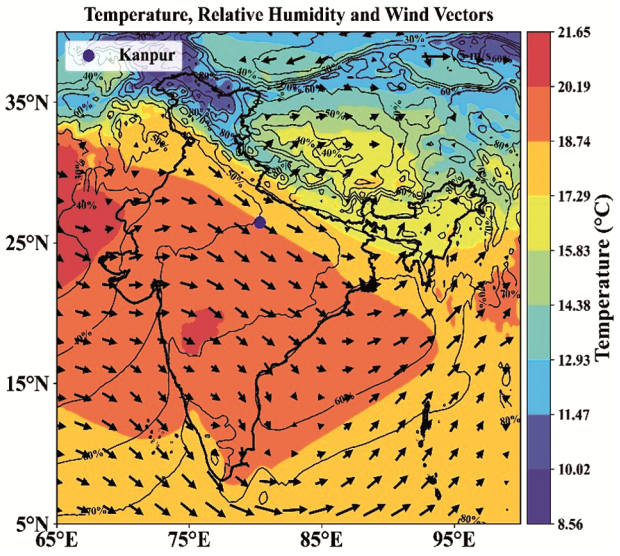

Fig. 1 — Synoptic meteorological conditions over India including the study location Kanpur (denoted by a blue circle) from January 2017 to December 2021

describe detailed information about the atmosphere, land, and ocean, making them integral to widespread applications in climate change research[34,35]. MODIS is a sensor aboard NASA's Terra and Aqua satellites, which were launched by NASA in December 1999 and May 2002, respectively. Terra and Aqua have different orbits, with Terra flying from North to South across the equator in the morning (10:30 local time (LT)) and Aqua flying from South to North over the equator in the afternoon (13:30 LT). Terra and Aqua sweep the entire surface of Earth every one to two days with a viewing swath width of 2,330 km. It consists of 36 spectral bands between 0.41 to 14.385 μm at an altitude of 705 km and three spatial resolutions of 250 m, 500 m, and 1 km. This research employed MODIS daily mean of aerosol and cloud data of level 3 and Collection 6.1 from Terra (MOD08_D3) and Aqua (MYD08_D3) platformsfrom 2017 to 2021, which are globally gridded over a 1°x1° resolution[36,37]. AOD measures the amount of light that aerosols scatter or absorb as it passes through a vertical column of the atmosphere. Higher AOD values indicate a greater concentration of aerosols, resulting in reduced visibility and increased scattering or absorption of sunlight. The AE is a parameter used to characterize the size of aerosol particles based on the wavelength dependence of AOD. The AE is inversely related to the size of aerosol particles. A smaller AE indicates a dominance of larger particles, while a larger suggests a prevalence of smaller particles. The AOD was obtained using combined Dark Target (over the ocean) and Deep Blue (over land) at 550nm while AE used only Deep Blue at 412-470 nm. However, the infrared band is adopted for retrieval of water vapor (WV) for total column, cloud Fraction (CF), and physical properties i.e., cloud top pressure (CTP) and cloud top temperature (CTT). Visible and near infrared regions are used to obtain combined cloud optical depth (COD), cloud liquid water path (CLWP), and mean of the ice and liquid cloud effective radius (CER)[38,39]. Daily mean data of aerosol and cloud parameters of MODIS were retrieved from the portal (https://giovanni.gsfc.nasa.gov).

3.3 Meteorological parameters

The European Centre for Medium-Range Weather Forecasts (ECMWF) has been reanalyzing historical observations with its prediction model and data assimilation system using ERA5 reanalysis data, which consists of hourly collections of atmospheric and land-surface meteorological features occurring since 1979[40]. All of the meteorological data that was utilized in this investigation was derived from the ERA5 database, which has been widely used by other researchers[41,42]. These datasets at 0.1° resolution included the following: atmospheric temperature (AT), relative humidity (RH), wind speed (WS), surface pressure (SP), and boundary layer height (BLH). ERA5 gave the following data: surface relative humidity (RH, expressed as a percentage), wind speed (u10, v10; in ms^{-1}), air temperature at a height of 2 m (t2m, in K), boundary layer height (BLH, in m) and surface pressure (SP, in Pa). These parameters have been shown to affect the concentration of $PM_{2.5}$ in a number of different research investigations[43,44]. Table 1 presents the dataset (2017–2021) utilized in this work, which consists of measured concentrations of $PM_{2.5}$ and AOD together with five meteorological parameters (i.e., AT, RH, WS, SP, and BLH).

Table1 —The descriptive statistics of daily $PM_{2.5}$, aerosol, meteorological, and cloud parameters as input data (from 01 January 2017 to 31 December 2021).

	AOD	BLH	WS	RH	SP	AT	$PM_{2.5}$	AE	WV	CER	CF	COD	CLWP	CTP	CTT
		m	m/s	%	mmHg	˚C	μg/m^3	_	cm	μm	_		g/m^2	hPa	K
Count	1420	1826	1826	1826	1826	1826	1767	1312	1676	1353	1826	1352	1299	1800	1654
Average	0.79	500.93	2.27	63.25	747.10	25.28	99.93	1.14	3.23	20.83	0.51	8.78	79.34	671.25	276.98
Standard Error	0.02	11.72	0.05	1.48	17.48	0.59	2.38	0.03	0.08	0.57	0.01	0.24	2.20	15.82	6.81
Median	0.67	436.95	2.15	65.39	746.89	27.27	68.18	1.26	2.26	20.82	0.48	5.32	50.00	739.08	288.94
Standard Deviation	0.45	225.12	0.89	17.10	5.11	6.23	84.89	0.56	2.21	8.09	0.36	9.92	86.57	245.19	27.72
Sample Variance	0.20	50678.69	0.79	292.27	26.11	38.87	7205.85	0.32	4.87	65.52	0.13	98.48	7494.77	60119.47	768.19
Skewness	1.77	0.87	1.09	-0.50	-0.04	-0.42	1.43	-0.54	0.99	0.23	0.09	2.46	2.45	-0.54	-1.03
Range	3.46	1137.94	6.56	80.29	22.93	29.32	469.78	1.80	8.81	43.57	1.00	67.43	531.50	882.75	119.47
Minimum	0.09	112.98	0.51	14.73	734.90	8.15	0.45	0.00	0.69	4.17	0.00	0.04	1.00	125.65	195.73
Maximum	3.55	1250.92	7.07	95.02	757.83	37.47	470.23	1.80	9.51	47.74	1.00	67.47	532.50	1008.40	315.20
Confidence Level (95.0%)	0.02	10.33	0.04	0.78	0.23	0.29	3.96	0.03	0.11	0.43	0.02	0.53	4.71	11.33	1.34

3.4 Model

Artificial neural networks (ANN) are networks of highly interconnected processors that run parallel computation methods. Using only a small subset of the data, it can learn to imitate more complex behaviors in the dataset without requiring the user to supply any additional code[45]. Because of this, ANN is able to acquire new knowledge and adjust to a setting that is always evolving. By adjusting the weights (connections) between its constituents, it can be trained to carry out a variety of tasks. ANN model can recognize similarities in input data, enhancing their ability to interpolate accurately, which is especially beneficial when dealing with noisy or imprecise data. When an ANN is being trained, the method is carried out in such a way that a specific input will result in a specific goal output[46]. Every network has a certain number of neurons in its input and output layers, proportional to the total number of inputs and outputs. Between the input and output layers of a multi-layer feed-forward neural network is the hidden layer, which represents a collection of parallel processing units (or nodes). The hidden layer's primary purpose is to enable the network to recognize and record important patterns in the data, as well as to carry out the sophisticated nonlinear mapping between the variables that are input and those that are output. The only job of the nodes that make up the input layer is to establish connections between the external inputs and the neurons that make up the hidden layer. The current study uses three different ANN models which are ANN1 (1 Layer, 14 Neurons), ANN2 (2 Layer, 14,28 Neurons), and ANN3 (3 Layer, 14,28,14 Neurons). As a result, the number of variables that are input is equal to the number of nodes that are input. Next, the hidden layer's outputs are sent to the output layer, which generates the network's ultimate result. The detailed structure[46]of neural interpretation is shown in Fig. 2.

3.5 Measures of accuracy applied in the model's performance

In order to assess how well the suggested nonlinear regression model performs, we use these performance indices: Fraction of predictions within a factor of 2 (FAC2), mean gross error (MGE), Normalized Mean Bias (NMB), mean absolute percentage error (MAPE), the root-mean-square error (RMSE), coefficient of efficiency (COE) and the Pearson coefficient (R). The following is the calculation for these indices:

$$FAC2 = 0.5 \leq \frac{p_i}{y_i} \leq 2.0 \quad \ldots (1)$$

$$MGE = \frac{1}{n}\sum_{i=1}^{n}|p_i - y_i| \quad \ldots (2)$$

$$NMB = \frac{\sum_{i=1}^{n}(p_i - y_i)}{\sum_{i=1}^{n} y_i} \quad \ldots (3)$$

$$MAPE = 100 * \frac{1}{n}\sum_{i=1}^{n}\frac{|p_i - y_i|}{y_i} \quad \ldots (4)$$

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2} \quad \ldots (5)$$

$$R = \frac{1}{(n-1)}\sum_{i=1}^{n}\left(\frac{p_i - \bar{p}}{\sigma_p}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right) \quad \ldots (6)$$

$$COE = 1.0 - \left(\frac{\sum_{i=1}^{n}|p_i - y_i|}{\sum_{i=1}^{n}|y_i - \bar{y}|}\right) \quad \ldots (7)$$

Where, y_i stands for the concentration that was actually measured, p_i for the value that was anticipated based on y_i, and n for the total number of samples used in the analysis. The FAC2 metric represents the fraction of points falling within 0.5 to 2 times the measured value. While RMSE quantifies absolute disagreement, MAPE measures relative discrepancy between values. COE, or Coefficient of Efficiency, indicates model performance: COE=1 signifies perfect predictions, while COE=0 indicates inability to predict observed values. The statistical

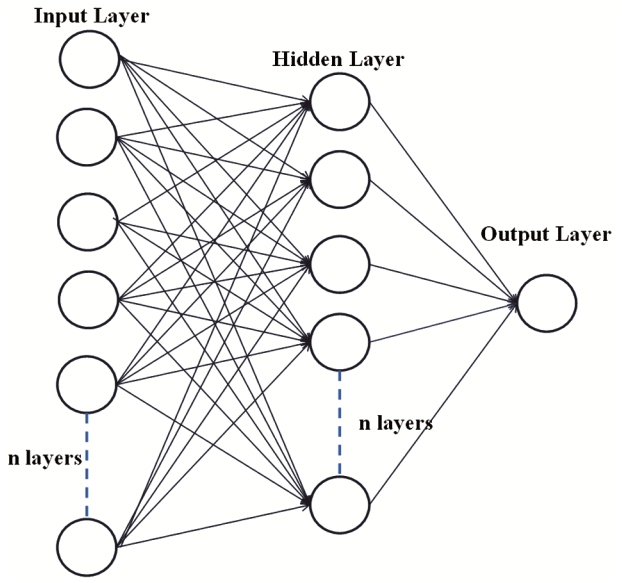

Fig. 2 — Structure of neural interpretation for predicting $PM_{2.5}$ concentrations

indices (FAC2, MGE, NMB, MAPE, RMSE, R and COE) were utilized to assess the overall quality and reliability of the created models. It is generally acknowledged that the availability of accurate meteorological projections is an essential requirement for the productive use of models in the production of accurate real-time forecasts.

4 Results and Discussion

4.1 Daily variation of aerosol, cloud and meteorological parameters

Figure 3 depicts the daily fluctuation of the time series of air pollutants that were detected over Kanpur throughout the research from 2017 to 2021. The descriptive statistics are presented in Table 1, which provides a frame work for the general characteristics of the data. Over the study, the average daily concentration of AOD was 0.79, with a range of 0.09 to 3.55, while for $PM_{2.5}$, it was 99.9 µg/m^3, with a range of 0.45 to 470.23 µg/m^3. It was observed that the $PM_{2.5}$ concentrations were higher than their corresponding 24-hour NAAQS value ($PM_{2.5}$ = 60 µg/m^3), which was found to be the case for approximately 55.7% of the total observation days. It is not out of the ordinary for this region to have higher $PM_{2.5}$ concentrations and exceedances[47]. The daily variation of AT was found from 8.15 ˚C (Winter) to 37.47 ˚C (pre-monsoon) and RH was 14.73 % (pre-monsoon) to 95 % (monsoon), with a mean value of 25 ˚C and 63 %. WS was determined by making use of both u and v components of wind speed, and a mean value of 2.27 m/s[1] was found.

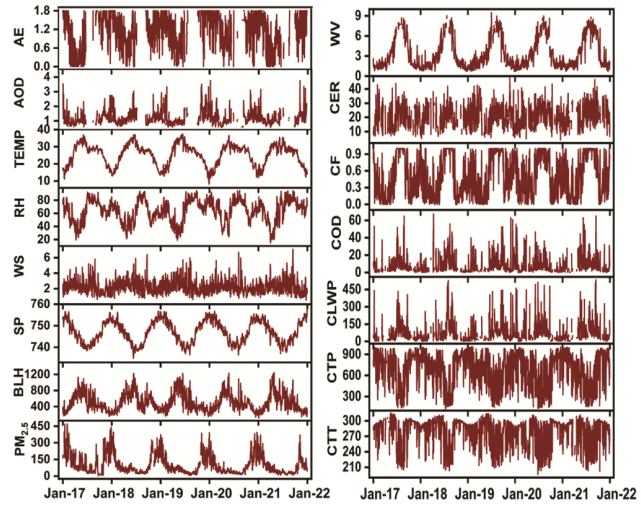

Fig. 3 — Daily variations of AOD, $PM_{2.5}$, cloud, and meteorological parameters at Kanpur during the entire study period from January 2017 to December 2021.

The daily mean values of SP were discovered to be 747 mmHg, while the BLH values were determined to be 500.93 m. The daily mean AE value was 1.14, with a range from zero to 1.80, indicating the dominance of fine mode aerosol generated from anthropogenic activities[48]. The daily variation of WV and CER was 3.23 cm and 20.83 µm respectively. The value of CF ranged from zero to one with a daily mean of 0.51. The study found that the daily mean value of COD was 8.78, while CLWP was 79.34 g/m^2, with the highest levels observed during the monsoon days. The daily mean CTP was 671.25 hPa, ranging from 125.7 to 1008.4 hPa. The daily mean value of CTT was found 277 K, varying from 195.7 to 315.2 K throughout the entire study period. Variations[49] in temperature, relative humidity, wind speed, precipitation, CER, COD, and cloud emissivity are associated with changes in the concentration and composition of $PM_{2.5}$. Moreover, these variations are linked by cloud phase and specific location.

4.2 Accuracy of training model using ANN

The pre-trained back propagation artificial neural network (ANN) was utilized to estimate the daily mean $PM_{2.5}$ concentrations, as illustrated in Figs. 4(a-c). Statistical measures using Pearson's correlation coefficient are used to assess the different ANN model's performance during the training phase for $PM_{2.5}$ forecasting models (R). The performance and regression graphs that were produced while the model was being trained were made available by MATLAB software, which was employed in the process of developing a back propagation neural network. In this model, we were provided with a training data set spanning from 2017 to 2020 and tasked with making predictions for 2021. To achieve optimal performance using ANN, we conducted experiments with three different ANN models (ANN1= 1 Layer, 14 Neurons, ANN2= 2 Layer, 14, 28 Neurons, ANN3= 1 Layer, 14, 28, 14 Neurons) with different hidden and output layers. These three models have different correlation coefficients (R) during regression analysis conducted of training, validation, test, and overall datasets. In the ANN1 model, the corresponding R values were 0.85, 0.77, 0.77, and 0.82 for training, validation, test, and all data, while for the ANN2 model R values are 0.80, 0.76, 0.74, and .078. Similarly, for the ANN3 model corresponding R values were 0.89, 0.86, 0.74, and 0.86 for training, validation, test, and all data. The study shows that a strategy very similar to this one

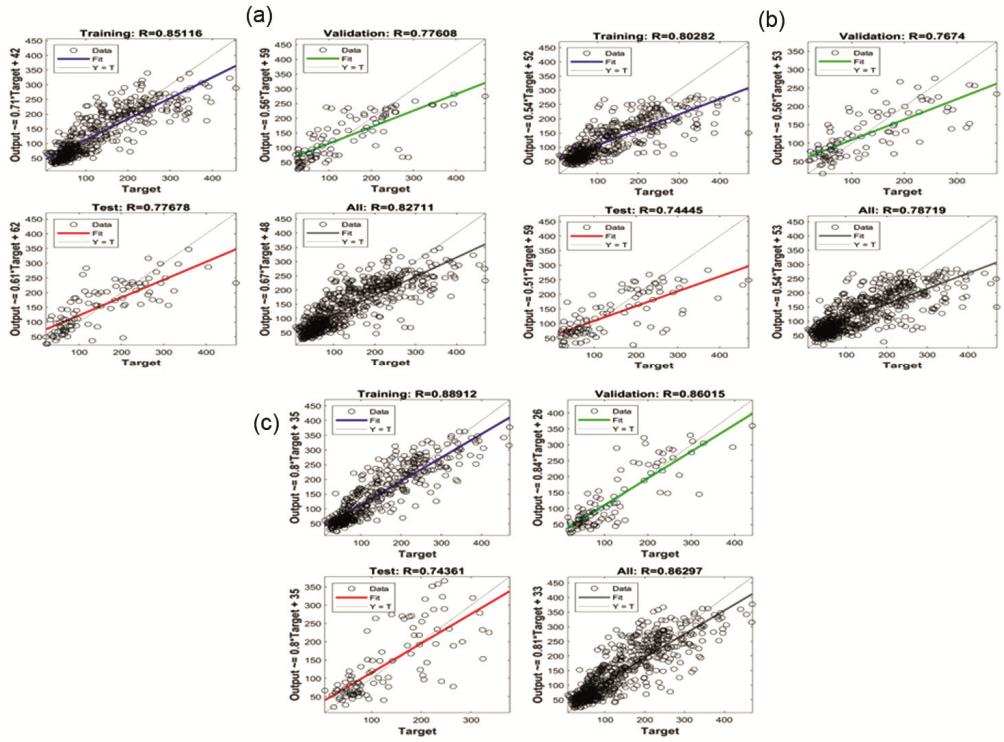

Fig. 4 — Accuracy of training model using ANN Model

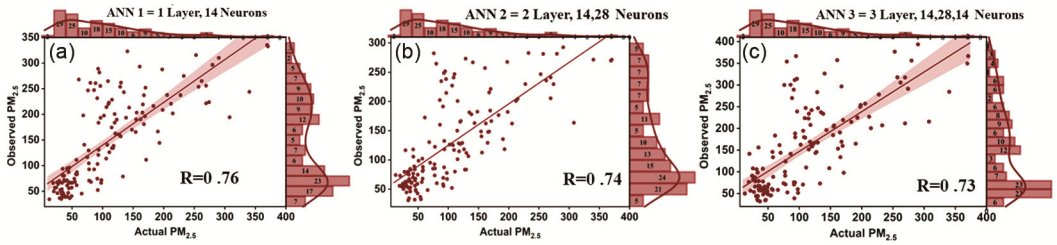

Fig. 5 — Comparison between predicted and measured $PM_{2.5}$concentration.

was employed for the prediction of PM_{10} using ANN[50]. The input parameters that they used were solar radiation, wind speed, and atmospheric pressure at Varanasi. The values of the correlation factor that were close to 0.6 were regarded to be typical within the realm of unpredictable climate fluctuations[51].

4.3 Accuracy of the predicted model

In this study, artificial neural network back propagation methods were employed to estimate ground-based $PM_{2.5}$ concentration by utilizing AOD and cloud parameters along with meteorological data. Three distinct $PM_{2.5}$ prediction models were implemented using the ANN method. Fig. 5 illustrates the outcomes for these three models. The contribution of the general models to the total variance in the data was assessed using the Pearson coefficient (R). The first model (ANN1), the second model (ANN2), and the third model (ANN3) reveal that 43%, 46%, and 42% respectively of the projected points are less than 100 µg/m^3, compared to the observed value, where only 58% of values are less than 100 µg/m^3. Model predictions exhibit a 95% level of significance, as depicted in Fig. 5. The $PM_{2.5}$ prediction accuracy for ANN1, ANN2, and ANN3 models was found to be represented by R values of 0.76, 0.74, and 0.73, respectively. The statistical errors of the models were computed based on FAC2, MGE, NMB, MAPE, RMSE, R, and COE, as detailed in Table 2. While the

Table 2 — Statistical analysis of different model performance

	FAC2	MGE	NMB	MAPE	RMSE	R	COE
ANN1	0.76	49.51	-0.26	64.32	68.54	0.76	0.35
ANN2	0.78	42.82	-0.17	61.98	59.63	0.74	0.33
ANN3	0.73	58.14	-0.30	78.84	81.61	0.73	0.30

Table 3 — Mean value with a corresponding standard deviation of $PM_{2.5}$, aerosol, and cloud parameters across various seasons from 2017 to 2021in Kanpur.

	Winter	Pre-monsoon	Monsoon	Post-monsoon	Annual
$PM_{2.5}$	183.95±62.61	66.78±27.40	37.02±18.51	129.56±67.56	101.49±73.09
AOD	0.94±0.26	0.64±0.20	0.86±0.28	0.82±0.21	0.80±0.26
AE	1.44±0.15	0.79±0.27	1.24±0.41	1.35±0.22	1.15±0.39
WV	1.55±0.19	3.15±1.43	6.77±0.88	2.53±0.93	3.5±2.22
CER	19.20±3.54	19.49±3.97	23.77±2.25	20.04±2.99	20.62±3.74
CF	0.48±0.13	0.35±0.22	0.88±0.14	0.34±0.15	0.51±0.28
COD	6.02±3.60	8.55±3.98	12.27±3.18	5.78±2.85	8.21±4.33
CLWP	49.72±26.41	71.33±38.96	130.85±34.22	53.99±23.52	76.58±9.61
CTP	808.36±66.51	670.94±69.99	456.04±114.36	776.97±84.46	672.44±160.91
CTT	283.23±5.21	282.87±10.64	257.40±13.41	287.73±8.47	277.68±15.50

first model demonstrates high R values, the lowest error is observed in the ANN2 model, with ANN3 exhibiting comparatively poorer performance. This study's results are consistent with recent research that has shown ANN models to be more effective than traditional regression methods for predicting $PM_{2.5}$ concentrations using remote sensing data[24,29,52]. The study utilized MODIS AOD data and found that the ANN model outperformed traditional regression methods[46]. Overall, the findings of this study validate the effectiveness of the ANN model in predicting PM2.5 concentrations, especially with remote sensing data, as evidenced by the higher R values achieved in comparison to other models. Many researchers have broadly used the ANN model for quantification and predicting air quality[53,54].

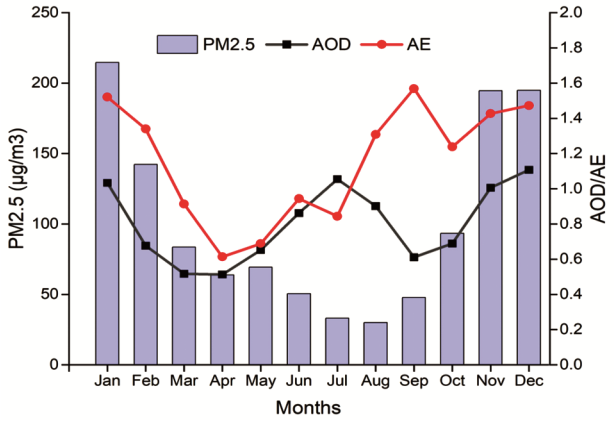

Fig. 6 — Time series plot of the monthly average of $PM_{2.5}$, AOD, and AE obtained from MODIS for period of 2017 to 2021 over Kanpur.

4.4. Distribution of aerosol and cloud properties

Aerosol properties ($PM_{2.5}$, AOD, and AE) and cloud parameters (WV, CER, CF, COD, CLWP, CTT, and CTP) were acquired from January 2017 to December 2021 in Kanpur are presented in Table 3. The annual mean values of AOD, AE, and $PM_{2.5}$ are 0.80±0.26, 1.15±0.39 and 101.49±73.09 µg/m^3 respectively. The highest annual average $PM_{2.5}$ was estimated in the winter season (183.95 µg/m^3) while the minimum annual average was found in monsoon (37.02 µg/m^3). This is because pollution remains near the ground for longer periods due to the lower mixed boundary layer height, resulting in higher pollutant concentrations. Moreover, the winter gave the maximum annual average AOD (0.94) and lowest AOD (0.64) in the pre-monsoon. A possible reason is that AOD measures aerosols throughout the vertical column from the Earth's surface to the upper atmosphere, whereas $PM_{2.5}$ particles are suspended close to the Earth's surface and influenced by local climatic conditions, especially during winter season[36,55]. High values of WV, CER, CF, COD, and CLWP were observed in the monsoon, while high CTP and CTT were reported in the winter and post-monsoon respectively. Annual average values of WV, CER, CF, COD, CLWP, CTP, and CTT are 3.50 cm, 20.62 µm, 0.51, 8.21, 76.58 g/m^2, 672.44 hPa and 277.68 K, respectively. Monthly mean variations in $PM_{2.5}$, AOD, and AE were analyzed for the years 2017-2021 are presented in Fig. 6. Analysis shows that the months with high $PM_{2.5}$ levels are detected in November, December, January, and February, while high AOD values are observed from November to January and July. The highest $PM_{2.5}$ and AOD are observed in January (214.74 µg/m^3) and December (1.11), while the lowest $PM_{2.5}$ and AOD are measured in August (30.02 µg/m^3) and April (0.51), respectively. The atmospheric conditions such as low relative humidity, light wind, low temperature, and low rainfall are associated with $PM_{2.5}$ measured near the Earth's surface[56,57].The monthly AOD and $PM_{2.5}$ concentrations fall in the ranges from 0.44 to 1.56 and

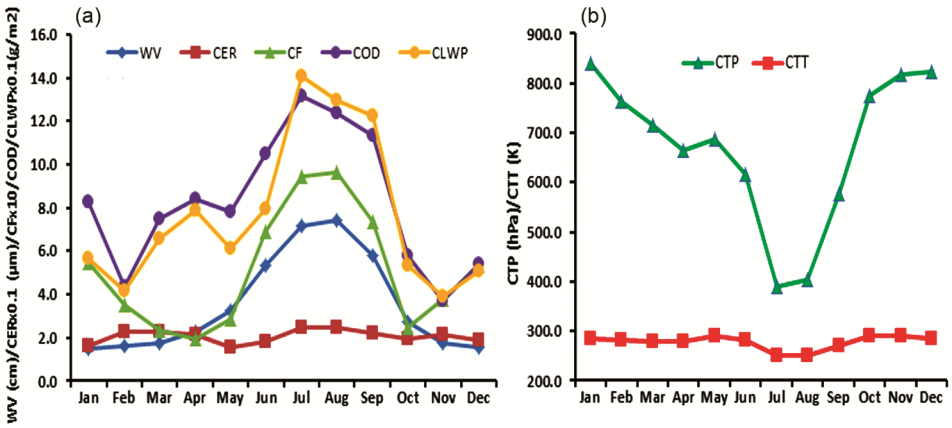

Fig.7 — Time series plot of the monthly average value of (a) WV, CER, CF, CLWP, and COD (b) CTP and CTT obtained from MODIS for a period of 2017 to 2021 over Kanpur.

from 22.60 μg/m³ to 270.78 μg/m³ respectively, while the AE ranges from 1.70 to 0.25 throughout the study period in Kanpur. The monthly mean AOD, $PM_{2.5}$, and AE are following a similar pattern except in June, August, and September[16]. Fig. 7 shows the analysis of monthly mean variations of the WV, CER, CF, COD, CLWP, CTP, and CTT data from 2017 to 2021 over Kanpur. The highest CER, COD, and CLWP were reported in July and the maximum WV and CF were found in August, while high CTP and CTT were analyzed in January and May, respectively. Additionally, monthly COD and CLWP follow similar trends as CTT and CTP throughout all months, while WV, CF, and CER exhibit a similar pattern in most months. The values of WV, CER, CF, COD, CLWP, CTP, and CTT are in the range of 7.65 to 1.17 cm, 27.25 to 12.03 μm, 0.98 to 0.10, 17.96 to 1.53, 188.83 to 13.39 g/m^2, 903.04 to 292.42 hPa and 299.07 to 236.25 K, respectively.

4.5 Variation of cloud parameters during high $PM_{2.5}$ concentrations

Severe air pollution is generally experienced in Kanpur during the period from November to February, mainly due to the combined effects of temperature inversion, stagnant air, increased emissions from vehicles and industries, burning of crop residues after harvest season, low rainfall and use of firecrackers in Diwali festival[48,58]. In the winter months, weather conditions result in pollutants being more effectively retained in the atmosphere for extended periods than in other seasons[36]. As shown in Fig. 8, Four-month average value of $PM_{2.5}$ is found to be 186.61 μg/m^3, which varied between 91.58 and 270.78 μg/m³. On the other hand, the magnitude of various cloud parameters such as WV, CER, CF,

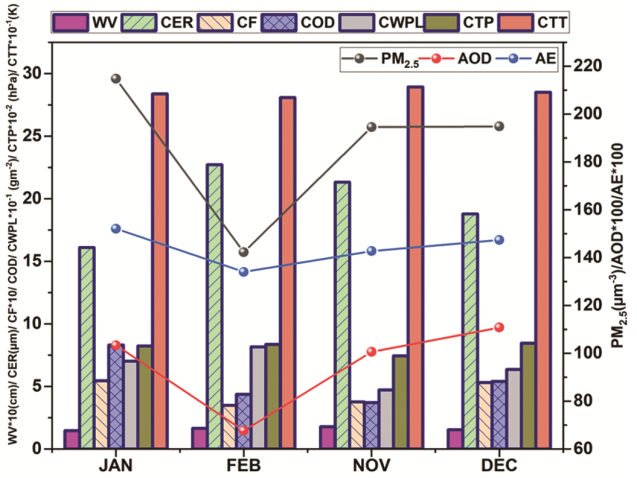

Fig. 8 — Variation of $PM_{2.5}$, AOD, AE, WV, CER, CF, CLWP, COD, CTP, and CTT during the high $PM_{2.5}$ concentration in the Kanpur region between 2017 to 2021.

COD, CLWP, CTP, and CTT are 1.61 cm, 19.73 μm, 0.45, 5.44, 46.98 g/m^2, 810.81 hPa, and 284.76 K, respectively during the high $PM_{2.5}$ conditions. In polluted areas like Kanpur during the winter months, high $PM_{2.5}$ values often lead to the formation of fog, low-level clouds, haze, and smog due to temperature inversions and reduced mixing. These conditions reduce visibility and create overcast skies, exacerbating air quality issues[55,59]. High $PM_{2.5}$ values have a significant impact on cloud properties within the atmospheric environment[36]. During the high $PM_{2.5}$ concentration of four-month study period, WV, CER, CF, COD, CLWP, CTP, and CTT demonstrate a range of values from 0.55 to 1.44 cm, 1.17 to 2.12 μm, 0.26 to 0.71, 1.53 to 15.31, 13.39 to 111.35 g/m², 685.24 to 903.04 hPa, and 274.14 to 296.93K, respectively. In some cases, stratocumulus clouds (low level clouds) may also form during high $PM_{2.5}$

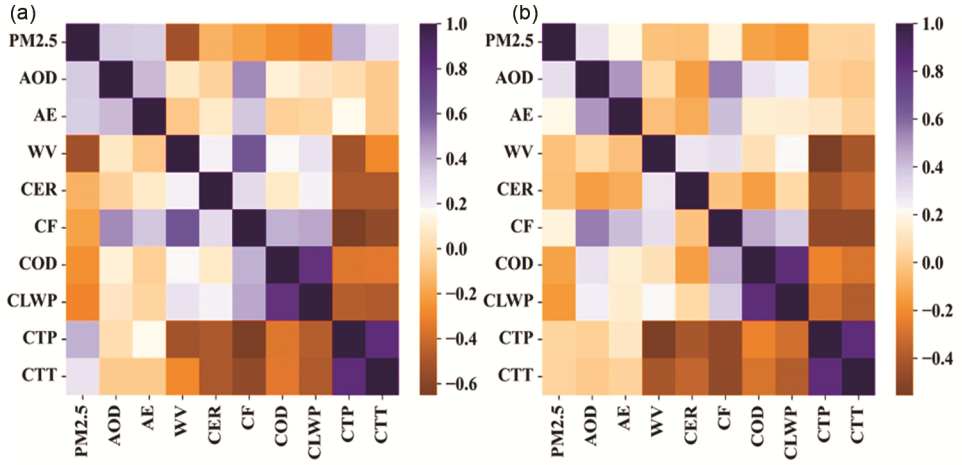

Fig. 9 — Daily correlation between $PM_{2.5}$, aerosol, and cloud properties in the Kanpur region for the period of 2017 to 2021. (a) Throughout the year (b) During high $PM_{2.5}$ concentration

levels[36], contributing to haziness and reduced visibility in the region[60].

4.6 Correlation between PM 2.5, aerosol and cloud properties

$PM_{2.5}$ and AOD indicate the concentration of aerosol particles in the atmosphere. Therefore, to understand the correlation between aerosol and cloud properties for daily data from January 2017 to December 2021 over Kanpur are presented in Figs. 9(a & b). The correlation between $PM_{2.5}$ and AOD was positive throughout the study period and also in high $PM_{2.5}$ months which is 0.34 and 0.31 respectively. The relationship between $PM_{2.5}$ and WV was found negative (-0.53) during the entire study period, while weak negative (-0.03) correlation during the high $PM_{2.5}$ concentration month. AOD and WV showed a weak positive correlation in both durations. The figure shows that the correlation between $PM_{2.5}$ with CER was weak and negative in both durations. A similar pattern was also found with AOD and CER. The analysis presented that the variability of CER with aerosol properties is consentient with the anti-Twomey effect in a contest of water vapor in cloud droplets, resulting vaporization of small droplets of cloud. The correlation between CF and $PM_{2.5}$ was negative (-0.20) during the entire study period while positive (0.17) in the high $PM_{2.5}$ concentration month but CF and AOD were positive during both periods. Several researchers also found a positive correlation between AOD and CF[16,38,61]. The correlation analysis between $PM_{2.5}$ with COD and CLWP was negative while AOD showed a positive relationship with COD and CLWP, in both durations. However, the increase in aerosol concentrations may result in a higher number of cloud droplets due to a reduction in droplet size at constant CLWP, leading to an increase in COD[49]. Analysis observed that CTP and CTT positively correlated with $PM_{2.5}$ during the entire study period, while weak positively correlated in high $PM_{2.5}$ concentration months. This effect may result from the inhibition of precipitation by extending cloud lifetime, which in turn influences cloud albedo and alters CTP[3]. The analysis shows that the relationship of CTP, CTT with AOD closed to zero in both periods. Observations suggest that the positive relationship between AOD and CTT is attributed to climate patterns and meteorological conditions[16].

Conclusion

This research aimed to develop a model for estimating surface $PM_{2.5}$ and to investigate the variability in aerosol and cloud parameters during high $PM_{2.5}$ concentrations in Kanpur, a heavily polluted city in India's Indo-Gangetic Plain (IGP) region. The average daily AOD was 0.79, ranging from 0.09 to 3.55. For $PM_{2.5}$, the average concentration was 99.9 µg/m³, with a range of 0.45 to 470.23 µg/m³. PM2.5 concentrations exceeded the 24-hour NAAQS limit of 60 µg/m³ on approximately 55.7% of observation days. The accuracy of $PM_{2.5}$ predictions using the ANN models (ANN1, ANN2, and ANN3) was represented by R values of 0.76, 0.74, and 0.73, respectively. The highest $PM_{2.5}$ concentration was recorded in winter (183.95 µg/m³), while the lowest was observed during the monsoon season (37.02 µg/m³). For AOD, the maximum value (0.94) occurred in winter while the minimum (0.64) was in the pre-monsoon season. The highest values of various cloud parameters, including WV, CER, CF, COD, and CLWP, were observed during the

monsoon. In contrast, the highest CTP was observed in winter, while the highest CTT was noted in the post-monsoon season. During periods of high $PM_{2.5}$ levels, concentrations ranged from 91.58 to 270.78 μg/m³. High surface $PM_{2.5}$ concentrations enhance the formation of low-level stratus clouds by increasing the availability of cloud condensation nuclei (CCN).This affects cloud microphysics by raising the number of cloud droplets while reducing their size, which leads to changes in cloud albedo, optical thickness, precipitation efficiency, cloud dynamics, and overall cloud lifetime. These stratus clouds often result in gloomy, overcast days with occasional drizzle or mist.

During winter in highly polluted areas like Kanpur, high $PM_{2.5}$ concentrations contribute to the formation of fog, low-level clouds, haze, and smog due to temperature inversions and limited air mixing. These conditions reduce visibility, create overcast skies, worsen air quality, and significantly impact cloud properties. The findings can be used by the local community and the government to support policy decisions and develop plans for improving air quality at regional and national levels.

Data availability

The paper provides information regarding the sources of data utilized. All datasets mentioned are openly available for access from the website https://giovanni.gsfc.nasa.gov/giovanni/ and https://airquality.cpcb.gov.in/ccr/#/caaqm-dashboard/caaqm-landing/data.

Acknowledgements

The authors acknowledge MODIS (Terra and Aqua) and ERA5 for providing online data access for this research. We also thank to Central Control Room (CCR) for Air Quality Management by CPCB Delhi for the online availability of $PM_{2.5}$ data. The authors express their gratitude to the Director, IITM for his continuous encouragement and support during the study period. The authors also extend their appreciation to the anonymous reviewers for their valuable comments and suggestions aimed at enhancing the manuscript.

References

1 Lelieveld J, Klingmüller K, Pozzer A, Burnett R T, Haines A & Ramanathan V, *Proc Natl Acad Sci,* 116 (2019) 7192.
2 Trenberth K E*, J Energy Nat Res,* 36 (2018, 463.
3 Verma P K, Srivastava A K Shukla, S P & Pathak V, *J Atmos Sol Terr Phys*, 258 (2024) 106233.
4 Kaushik C P, Ravindra K, Yadav K, Mehta S & Haritash A K, *Environ Monit Assess*, 122 (2006) 27.
5 Pal R, Chowdhury S, Dey S & Sharma A R, *Aerosol Air Qual Res*, 18 (2018) 2332.
6 Singh A, Srivastava A K, Pathak V & Shukla A K, *Atmos Environ,* 270 (2021) 118893.
7 Mukherjee A & Agrawal M, *Atmos Res*, 213 (2018) 275.
8 Chowdhury S, Pozzer A, Haines A, Klingmueller K, Muenzel T, Paasonen P, *et al., Environ Int,* 159 (2022) 107020.
9 Cohen A J, Brauer M, Burnett R, Anderson H R, Frostad J, Estep K, Balakrishnan K, Brunekreef B, Dandona L, Dandona R & Feigin V, *Lancet*, 389 (2017) 1907.
10 Guo H, Kota, S H, Chen K, Sahu S K, Hu J, Ying Q, Wang Y & Zhang H, *Atmos Chem Phys*, 18 (2018) 15219.
11 Ansari K & Ramachandran S, *Atmos Environ*, 293 (2023) 119434.
12 Ramachandran S, Rupakheti M & Lawrence M G, *Environ Int*, 142 (2020) 105814.
13 Chakraborty R P, Mandariya A K & Gupta T, *Atmos Res*, 178 (2016) 506.
14 Dandona L, Dandona R, Kumar GA, Shukla D K, Paul V K, Balakrishnan K, Prabhakaran D, Tandon N, Salvi S, Dash AP & Nandakumar A, *Lancet*, 390 (2017) 2437.
15 Twomey S, *J Atmos Sci,* 34 (1977) 1149.
16 Kumar K R, Boiyo R, Madina A & Kang N, *J Atmos Sol Terr Phys*, 179 (2018) 55.
17 Boiyo R, Kumar K R & Zhao T, *Int J Climatol*, 38 (2018) 1221.
18 Adesina A J, Kumar K R & Sivakumar V, *Aerosol Air Qual Res*, 16 (2016) 195.
19 He Q, Zhang M & Huang B, *Atmos Environ*, 129 (2016) 79.
20 Sarangi C, Kanawade V P, Tripathi S N, Thomas A & Ganguly D, *Nat Commun*, 9 (2018) 3754.
21 Jiang J H, Su H, Huang L, Wang Y, Massie S, Zhao B, Omar A & Wang Z, *Nat Commun*, 9 (2018) 3874.
22 Jose S, Mishra A K & Singh S, *Atmos Res*, 262 (2021) 105796.
23 Lamsal L N, Duncan B N, Yoshida Y, Krotkov N A, Pickering K E, Streets D G & Lu Z, *Atmos Environ*, 110 (2015) 130.
24 Bilal M, Nichol J E & Spak S N, *Aerosol Air Qual Res*, 17 (2017) 356.
25 Yeganeh B, Hewson M G, Clifford S, Knibbs L D & Morawska L, *Environ Model Softw*, 88 (2017) 84.
26 Nieto P G, Lasheras F S, García-Gonzalo E & de Cos Juez F, *J Sci Total Environ*, 621 (2018) 753.
27 Xu Y, Du P & Wang J, *Environ Pollut*, 223 (2017) 435.
28 Qin W, Wang L, Lin A, Zhang M & Bilal M, *Remote Sens*, 10 (2018) 1022.
29 Bera B, Bhattacharjee S, Sengupta N & Saha S, *Environ Chall*, 4 (2021) 100155.
30 Ram K, Singh S, Sarin M M, Srivastava A K & Tripathi S N, *Atmos Res*, 174 (2016) 52.
31 Jain S, Sharma S K, Vijayan N & Mandal T K, *Environ Pollut*, 262 (2020) 114337.
32 Rai P, Chakraborty A, Mandariya A K & Gupta T, *Atmos Res*, 178 (2016) 506.
33 CPCB, 2011 Guidelines for the measurement of ambient air pollutants, P R Division, Central Pollution Control Board, New Delhi, India.

34 Shaik D S, Kant Y, Sateesh M, Sharma V, Rawat D S & Chandola H C, *Atmos Remote Sens,* (2023) 121.

35 Vohra K, Vodonos A, Schwartz J, Marais E A, Sulprizio M P & Mickley L J, *Environ Res*, 195 (2021) 110754.

36 Pilahome O, Nissawan W, Jankondee Y, Masiri I & Kumharn W, *Adv Space Res*, 71 (2023) 3166.

37 Levy R C, Mattoo S, Munchak L A, Remer L A, Sayer A M, Patadia F & Hsu N C, *Atmos Meas Techn*, 6 (2013) 2989.

38 Alam K, Khan R, Blaschke T & Mukhtiar A, *J Atmos Sol Terr Phys*, 107 (2014) 104.

39 Jin M & Shepherd J M, *J Geophys Res Atmos*, 113 (2008) 24.

40 Jiang Q, Li W, Fan Z, He X, Sun W, Chen S, Wen J, Gao J & Wang J, *J Hydrol*, 595 (2021) 125660.

41 Venkataraman C, Bhushan M, Dey S, Ganguly D, Gupta T, Habib G, Kesarkar A, *et al*., *Bull Am Meteorol Soc*, 101 (2020) 1052.

42 Hu J, Wang H, Zhang J, Zhang M, Zhang H, Wang S & Chai F, *Atmos*, 10 (2019) 121.

43 Wang Y, Zou L, Wu T, Xiong L, Zhang T, Kong L, Xue Y & Tang M, *Ecotoxicol Environ Saf,* 169 (2019) 863.

44 Gui K, Che H, Wang Y, Wang H, Zhang L, Zhao H, Zheng Y, Sun T & Zhang X, *Environ Pollut,* 247 (2019) 1125.

45 Anderson J A, An introduction to neural networks, *Cambridge, MA: MIT press*, (1995).

46 Ahmad M, Alam K, Tariq S, Anwar S, Nasir J & Mansha, *Atmos Environ,* 219 (2019) 117050.

47 Khan S & Hassan Q, *Lecture Notes in Civil Engineering, Springer Singapore*, 58 (2020).

48 Singh A, Singh S, Srivastava A K, Payra S, Pathak V & Shukla A K, *Environ Monit Assess,* 194 (2022) 827.

49 Belle J H, Chang H H, Wang Y, Hu X, Lyapustin A & Liu Y, *Int J Environ Res Public Health,* 14 (2017) 1244.

50 Yadav V & Nath S, *Int J Environ Sci Technol*, 16 (2019) 2839.

51 Ul-Saufie A Z, Yahaya A S, Ramli N A, Rosaida N & Hamid H A, *Atmos Environ*, 77 (2013) 621.

52 Sharma V, Ghosh S, Dey S & Singh S, *Annals GIS,* 29 (2023) 415.

53 Suleiman A, Tight M R & Quinn A D, *Atmos Pollut Res*, 10 (2019) 134.

54 Alimissis A, Philippopoulos K, Tzanis C G & Deligiorgi D, *Atmos Environ,* 191 (2018) 205.

55 Mhawish A, Banerjee T, Sorek-Hamer M, Bilal M, Lyapustin A I, Chatfield R & Broday D M, *Environ Sci Technol,* 54 (2020) 7891.

56 Zhang Q, Zheng Y, Tong D, Shao M, Wang S, Zhang Y, Xu X, *et al.,Proc Natl Acad Sci*, 116 (2019) 2446.

57 Cheng Z, Luo L, Wang S, Wang Y, Sharma S, Shimadera H, Wang X, Bressi M, *et al., Enviro Int*, 89 (2016) 212.

58 Kumar S, Singh A, Srivastava A K, Sahu S K, Hooda R K, Dumka U C & Pathak V, *Urban Clim,* 38 (2021) 100880.

59 Liu Y, Zhang J, Zhou P, Lin T, Hong J, Shi L, Yao F, Wu J, Guo H & de Leeuw G, *Atmos Chem Phys*, 18 (2018) 18187.

60 Eck T F, Holben BN, Kim J, Beyersdorf A J, Choi M, Lee S, Koo J H, Giles D M, *et al., Atmos Environ,* 232 (2020) 117530.

61 Gopal K R, Reddy K R O, Balakrishnaiah G, Arafath S M, Reddy N S K, Rao T C, *et al., J Atmos Sol Terr Phys,* 146 (2016) 38.