Assessment of Multi-Pollutant Concentrations Across Indian Urban Centres: A case study towards use of sustainable materials for improvement in Air Quality

Deepa Naik[1]

[1]Department of Computer Science and Engineering, Dr. B. C. Roy Engineering College, Durgapur, West Bengal, India.

Abstract

Deteriorating air quality is a critical public health challenge globally, and India ranks among the most severely affected nations. This study presents a multi-pollutant spatial analysis of air quality across 346 monitoring stations in India, using high-resolution data from the Central Pollution Control Board's Continuous Ambient Air Quality Monitoring System (CPCB-CAAQMS). Precise geographic coordinates (latitude and longitude) were used to apply two spatial interpolation methods — Inverse Distance Weighting (IDW, power parameter $p = 2$) and Ordinary Kriging — to transform discrete station-level measurements into continuous pollution surfaces across Indian territory. Pearson's correlation analysis, multiple linear regression, and k-means clustering ($k = 4$) were employed to identify inter-pollutant relationships and classify monitoring regions into pollution regime clusters. Results reveal that $PM_{2.5}$ and PM_{10} consistently exceed World Health Organization (WHO) guidelines across India, particularly in the Indo-Gangetic Plain (IGP), while SO_2 and CO remain within permissible limits in most regions. A strong linear correlation ($r = 0.81$) between $PM_{2.5}$ and PM_{10} suggests shared sources including road dust, construction activity, and biomass burning. Cross-referencing with the Ministry of Road Transport and Highways' VAHAN Sewa Dashboard establishes a statistically significant association between cumulative vehicle registration density and elevated ambient NO_2 and CO levels. A Digital Twin framework is proposed for real-time atmospheric monitoring, sensor fault detection, and spatial anomaly identification. Sustainable material-based mitigation strategies — including self-healing bio-concrete, reinforced PLA packaging films, and fibre-reinforced polymer composites — are recommended as source-level interventions directly linked to the dominant pollutant types identified in this analysis.

Keywords

Air Quality Index (AQI); particulate Matter ($PM_{2.5}$, PM_{10}); spatial interpolation; kriging; inverse distance weighting; urban pollution; Indo-Gangetic plain; VAHAN dashboard; digital twin; sustainable materials.

1. Introduction

1.1 Air Pollution: A Global and Indian Perspective

Air pollution is a leading cause of premature mortality worldwide, second only to hypertension as a risk factor for death [1]. Nitrogen dioxide (NO_2) from vehicular exhaust is a primary trigger of childhood asthma, while elevated ground-level ozone is strongly associated with Chronic Obstructive Pulmonary Disease (COPD). Fine particulate matter ($PM_{2.5}$) penetrates deep into lung tissue and enters the bloodstream directly, posing severe cardiovascular and respiratory risks [2].

India, the world's most populous nation, aspires to grow its economy from USD 4 trillion to USD 5 trillion, a goal achievable through expanded construction, industrial production, energy generation, and transport infrastructure. However, these activities inherently produce environmental degradation that spans all of India's diverse geographic zones [1][3]. According to IQAir's 2024 World Air Quality Report, India records an annual average $PM_{2.5}$ concentration of 50.6 µg/m³ — ten times the WHO annual guideline of 5 µg/m³ — ranking it among the world's most polluted nations (Table-1) [4-5].

Country	$PM_{2.5}$ (µg/m³)	Multiple of WHO Guideline(5 µg/m³)
Chad	91.8	18x
Bangladesh	78.0	15x
Pakistan	73.7	14x
DR Congo	58.2	11x
India	50.6	10x

Table 1. Top 5 most polluted countries by annual average $PM_{2.5}$ concentration (IQ Air 2024 World Air Quality Report [5])

1.2 The Case for Spatial Mapping

Conventional air quality monitoring in India relies on fixed-point sensors managed by the Central Pollution Control Board (CPCB). While these provide accurate localized readings, they leave large geographic blind spots. Correlating station-level measurements with geographic coordinates (latitude and longitude) enables Spatial Modelling- a technique that reveals 'Pollution Corridors' formed by topographic channelling of prevailing winds, often crossing state boundaries [2]. India's diverse topography, from the Himalayan foothills to the Deccan Plateau and coastal plains, governs the 'Ventilation Coefficient' of each city, meaning that pollution severity is a product of both emissions and meteorological trapping [5]. Spatial interpolation using Inverse Distance Weighting (IDW) and Ordinary Kriging was applied in this study to transform discrete point measurements into continuous pollution surfaces, enabling Trans-boundary Air Quality Management. This spatial approach also complements the first comprehensive state-wise estimates of air pollution's impact on health loss and life expectancy reduction in India, reported by [6]

Research Objectives

This study aims to:

1. Quantify the concentration of seven criteria pollutants ($PM_{2.5}$, PM_{10}, CO, SO_2, NO_2, NH_3, O_3) across 346 major Indian urban monitoring stations.
2. Apply and compare IDW and Ordinary Kriging spatial interpolation to model continuous pollution surfaces from discrete station data.
3. Identify inter-pollutant dependencies and classify monitoring stations into pollution regime clusters using statistical and machine learning methods.
4. Correlate ambient pollutant concentrations with vehicular registration density using the VAHAN Sewa Dashboard.
5. Propose a Digital Twin framework for real-time atmospheric monitoring and sensor fault detection.
6. Evaluate source-specific sustainable material interventions linked directly to the dominant pollutant sources identified in the data.

1.3 Novelty and Scientific Contribution

This work makes the following distinct contributions over existing AQI and GIS-based studies:

- Multi-state, multi-pollutant spatial fusion: Unlike most studies focusing on a single city or region, this analysis integrates 346 stations across all Indian states, producing a nationwide multi-pollutant spatial map at a resolution not previously reported in the open literature.
- Cross-domain data fusion (environmental + transport): Systematic integration of CPCB ambient sensor data with VAHAN vehicle registration data is a

methodological innovation that enables computational correlation between mobile emission sources and ambient pollutant levels - an approach distinct from typical GIS-only AQI studies.

- Digital Twin with sensor predictive maintenance: The proposed Digital Twin framework extends beyond visualization to include sensor fault detection using instrument signal range validation (4–20 mA), bridging IoT sensor engineering with atmospheric monitoring.
- Sustainability linkage grounded in results: The sustainable materials recommendations are not generic - each material strategy is explicitly linked to the specific dominant pollutant source categories identified through clustering and regression analysis in the results.

2. Literature Review – Atmospheric Dynamics and Regional Trends

The Indo-Gangetic Plain (IGP) has been extensively documented as India's most severely polluted airshed, with researchers characterizing it as a 'Topographic Sink' where pollutants accumulate due to blocking by the Himalayas and temperature inversions during winter months [7][8]. Trapped cool air near the ground during the post-monsoon season results in dramatic $PM_{2.5}$ spikes across North Indian cities.

Although India has implemented Bharat Stage VI (BS-VI) vehicular emission norms — equivalent to Euro VI standards — states with high vehicular densities, such as Maharashtra and Karnataka, continue to show rising ambient NO_2 and CO concentrations [9].Studies have attributed night-time CO spikes in Tier-1 cities to last-mile delivery network operations [10]. The ratio of gaseous pollutants to particulate matter serves as a diagnostic indicator: a high CO/PM ratio signals primary combustion emissions, while photochemically derived secondary pollutants indicate regional atmospheric chemistry [11].

The distinction between PM_{10} and $PM_{2.5}$ is analytically critical. $PM_{2.5}$ predominantly originates from combustion processes, while PM_{10} is heavily influenced by 'Fugitive Dust' from construction, unpaved roads, and natural soil deflation. In arid states like Rajasthan, natural dust significantly inflates PM_{10} values, potentially masking industrial pollution signals. The $PM_{2.5}/PM_{10}$ ratio serves as a proxy to distinguish anthropogenic (combustion) from lithogenic (crustal dust) sources [6].

A well-established link between air quality and economic productivity exists: high particulate matter exposure is associated not only with respiratory failure but also with cognitive decline and reduced workforce productivity [9]. India's National

Clean Air Programme (NCAP) targets a 20–30% reduction in PM levels, yet $PM_{2.5}$ concentrations remain 'stubbornly high' in major cities due to an implementation gap between national policy and city-level execution [12].

3. Materials and Methods

3.1 Data Acquisition and Dataset Description

Primary Source: High-resolution cross-sectional data were acquired from the Continuous Ambient Air Quality Monitoring System (CAAQMS) managed by the Central Pollution Control Board (CPCB) of India (https://www.data.gov.in/). The dataset represents daily average concentrations recorded in February 2026 across 346 monitoring stations in 28 states and 8 Union Territories. Seven criteria pollutants are included: $PM_{2.5}$, PM_{10}, CO, SO_2, NO_2, NH_3, and O_3.

Supplementary Source: State-level vehicle registration and transaction data were obtained from the Ministry of Road Transport and Highways' VAHAN Sewa Dashboard (https://vahan.parivahan.gov.in/) to correlate mobile emission sources with ambient pollutant concentrations.

Dataset characteristics: Approximately 3,000 station-pollutant observations; 346 unique monitoring stations; geographic coverage: latitude 8°N to 37°N, longitude 68°E to 97°E.

3.2 Data Preprocessing and Quality Control

Data integrity is essential for spatial computational models. The following preprocessing steps were applied:

- Outlier exclusion: Negative concentration values (physically impossible; indicative of instrument power failure or sensor malfunction, equivalent to a signal below 4 mA DC) were treated as missing values and excluded.
- Extremum filtering: Observations exceeding three standard deviations above the state-level mean for each pollutant were flagged as anomalous and excluded from interpolation analysis.
- Station eligibility: Stations with more than 30% missing pollutant readings were excluded from spatial interpolation but retained for descriptive statistics.
- Software environment: Python 3.11; libraries: Pandas 2.0, NumPy 1.24, GeoPandas 0.13, Pykrige 1.7, Scipy 1.11, Matplotlib 3.8, Folium 0.15.
- The Air Quality Index (AQI) was computed using Indian sub-index transformation equations, where each pollutant is weighted according to its health-risk potential, as per the CPCB standard methodology. Descriptive statistics mean, median, standard deviation, and 98th percentile were

computed for each pollutant across all stations[13].

- **3.3 Spatial Interpolation Methods**
- **3.3.1 Inverse Distance Weighting (IDW)**
- Fixed monitoring stations provide point measurements. To estimate pollution levels at unmonitored locations, IDW interpolation was applied. IDW is based on Tobler's First Law of Geography: observations that are geographically closer are more similar than observations that are farther apart. The estimated value at an unmonitored location is a weighted average of surrounding station values, where weights are inversely proportional to distance:
- $\hat{Z}(x_0) = \Sigma\ [w_i \times Z(x_i)] / \Sigma\ w_i$; **where** $w_i = 1 / d_i^p$**-Eq(1)**
- where $\hat{Z}(x_0)$ is the estimated pollutant concentration at unmonitored point x_0, $Z(x_i)$ is the measured concentration at station i, d_i is the Euclidean distance between x_0 and station i, and p is the power parameter.
- **Power parameter justification (p = 2):** A power of p = 2 was selected because ambient pollutant concentrations decay approximately with the square of distance from the source, consistent with Gaussian dispersion principles for ground-level sources. With p = 1, the weight decay is linear and overestimates the influence of distant stations. With p = 2, the model more accurately reflects the rapid attenuation of PM concentrations observed near localized sources such as industrial stacks, construction sites, and high-traffic corridors.
- **Worked example — Ghaziabad to Greater Noida:** Ghaziabad $PM_{2.5}$ = 211 µg/m³ (5 km from target point); Greater Noida $PM_{2.5}$ = 196 µg/m³ (20 km). With p = 2: $w_1 = 1/5^2 = 0.04$; $w_2 = 1/20^2 = 0.0025$. Estimated value = $(211 \times 0.04 + 196 \times 0.0025) / (0.04 + 0.0025) = (8.44 + 0.49) / 0.0425 =$ 210.1 µg/m³.
- **3.3.2 Ordinary Kriging**
- Ordinary Kriging is a geostatistical interpolation method that, unlike IDW, provides not only an estimated value but also an estimation variance (kriging error), enabling quantification of prediction uncertainty. Kriging derives optimal weights by modelling the spatial autocorrelation structure of the data through a variogram.
- **Variogram model:** A spherical variogram model was fitted to the experimental variogram of $PM_{2.5}$ data: $\gamma(h) = C_0 + C_1\ [1.5(h/a) - 0.5(h/a)^3]$ for $h \leq a$, and $\gamma(h) = C_0 + C_1$ for $h > a$, where C_0 is the nugget (measurement error + microscale variation), C_1 is the partial sill, and a is the range. Fitted parameters: nugget $C_0 = 120$, sill $C_0 + C_1 = 850$, range a = 380 km.

The Kriging system solves for weights λ_i subject to the unbiasedness constraint $\Sigma\lambda_i$

= 1:

$$\sum_{\chi=1}^{n} \lambda_i \gamma(x_i, x_\chi) + \mu = \gamma(x_i, x_0) \text{ for all i; and } \sum \lambda_i = 1 \text{-Eq.2}$$

3.3.3 Validation — Leave-One-Out Cross-Validation (LOOCV)

To assess interpolation accuracy, LOOCV was performed: each station was sequentially withheld, and its concentration was predicted from the remaining stations. Performance metrics are reported in Table 2.

Pollutant	IDW RMSE	IDW MAE	Kriging RMSE	Kriging MAE
$PM_{2.5}$	24.3	18.1	19.8	14.6
PM_{10}	31.5	23.4	26.2	19.1
NO_2	8.7	6.2	7.4	5.3
CO	11.2	8.0	9.5	6.9

Table 2. LOOCV performance: RMSE and MAE (μg/m³) for IDW vs Ordinary Kriging across key pollutants. Kriging consistently outperforms IDW.

3.4 Statistical Analysis

3.4.1 Pearson Correlation Analysis

Pearson's correlation coefficient (r) was computed to quantify the strength and direction of linear relationships between pollutant pairs across all 346 stations. For two pollutants X and Y measured across n stations:

$$r = \Sigma[(X_i - \bar{X})(Y_i - \bar{Y})] / \sqrt{[\Sigma(X_i - \bar{X})^2 \times \Sigma(Y_i - \bar{Y})^2]} \text{-Eq(3)}$$

r ranges from −1 (perfect negative correlation) to +1 (perfect positive correlation); r = 0 indicates no linear

relationship.

3.4.2 Multiple Linear Regression

To assess the combined predictive relationship between gaseous pollutants and $PM_{2.5}$, a multiple linear regression model was estimated with $PM_{2.5}$ as the dependent variable and CO, NO_2, PM_{10}, NH_3, and SO_2 as predictors. The model form is:

$$PM_{2.5} = \beta_0 + \beta_1(CO) + \beta_2(NO_2) + \beta_3(PM_{10}) + \beta_4(NH_3) + \beta_5(SO_2) + \varepsilon \text{-Eq(4)}$$

Results: Adjusted $R^2 = 0.66$; PM_{10} was the strongest predictor (standardized $\beta = 0.72$, $p < 0.001$), followed by CO ($\beta = 0.13$, $p < 0.001$). NO_2 ($\beta = 0.04$, $p = 0.28$) and NH_3 ($\beta = 0.07$, $p = 0.06$) were not significant predictors once PM_{10} was accounted for, and SO_2 was not a significant predictor ($p = 0.33$), consistent with its independent source profile. The standardized β for PM_{10} is numerically close to, but conceptually distinct from, the zero-order Pearson correlation reported in Section 3.4.1 ($r = 0.81$): the former reflects PM_{10}'s partial contribution within the five-predictor model, while the latter is a simple bivariate association.

profile.

3.4.3 K-Means Clustering

K-means clustering ($k = 4$, selected via elbow method and silhouette analysis; silhouette score = 0.58) was applied to classify monitoring stations into four pollution regime clusters based on the full pollutant profile:

- Cluster 1 — High Dust Zone (PM_{10}-dominant): Rajasthan, Gujarat, arid belt. Characterized by high $PM_{10}/PM_{2.5}$ ratio (> 2.5), indicating crustal dust dominance.
- Cluster 2 — High Traffic Zone (NO_2 and CO elevated): Maharashtra, Karnataka, Tamil Nadu metros. CO-NO_2 co-elevation pattern consistent with fossil fuel combustion.
- Cluster 3 — Industrial Combustion Zone ($PM_{2.5}$, SO_2 elevated): Eastern belt — Jharkhand, West Bengal, Odisha. Higher SO_2 signals industrial stack emissions from coal-based industries.
- Cluster 4 — Moderate/Clean Air Zone: Coastal and hilly states with higher ventilation coefficients.

3.5 Meteorological Normalization

Pollutant concentrations are modulated by both emissions and atmospheric conditions. Station coordinates were correlated with available meteorological data to account for Mixing Height and Ventilation Coefficient. High-latitude cities (above 28°N) frequently experience lower mixing heights during winter, leading to concentration amplification in the absence of increased emissions. By integrating meteorological variables, the model distinguishes between cities with genuinely high emissions and those that are 'Poorly Ventilated,' enabling more accurate policy targeting.

4. Results: The Spatial Landscape of Pollution

4.1 Pollutant Distribution Overview

Table 3 presents descriptive statistics for all seven monitored pollutants across 346 stations; Figure 1 illustrates the average concentration of each pollutant across all stations.

Pollutant	Mean ((μg/m^3)	Median	Std Dev	98th Percentile
PM_{10} (µg/m³)	97.9	87.4	52.1	226.3
$PM_{2.5}$ (µg/m³)	95.0	84.6	49.8	215.8
CO (mg/m³)	40.9	31.0	38.4	142.5
O_3 (Ozone) (µg/m³)	39.9	36.5	22.3	98.4
NO_2 (ppb)	31.4	14.2	15.9	68.2
SO_2 (ppb)	14.9	11.8	13.1	54.9
NH_3 (ppb)	6.6	4.3	7.8	29.1

Table 3. Descriptive statistics for seven criteria pollutants across 346 CPCB-CAAQMS stations (2026). Note: units differ by pollutant and are given in parentheses beside each pollutant name in the table; CO is reported in mg/m³, while the remaining pollutants are reported in µg/m³ (PM_{10}, $PM_{2.5}$, O_3) or ppb (NO_2, SO_2, NH_3) — values are not directly comparable across rows without accounting for this difference.

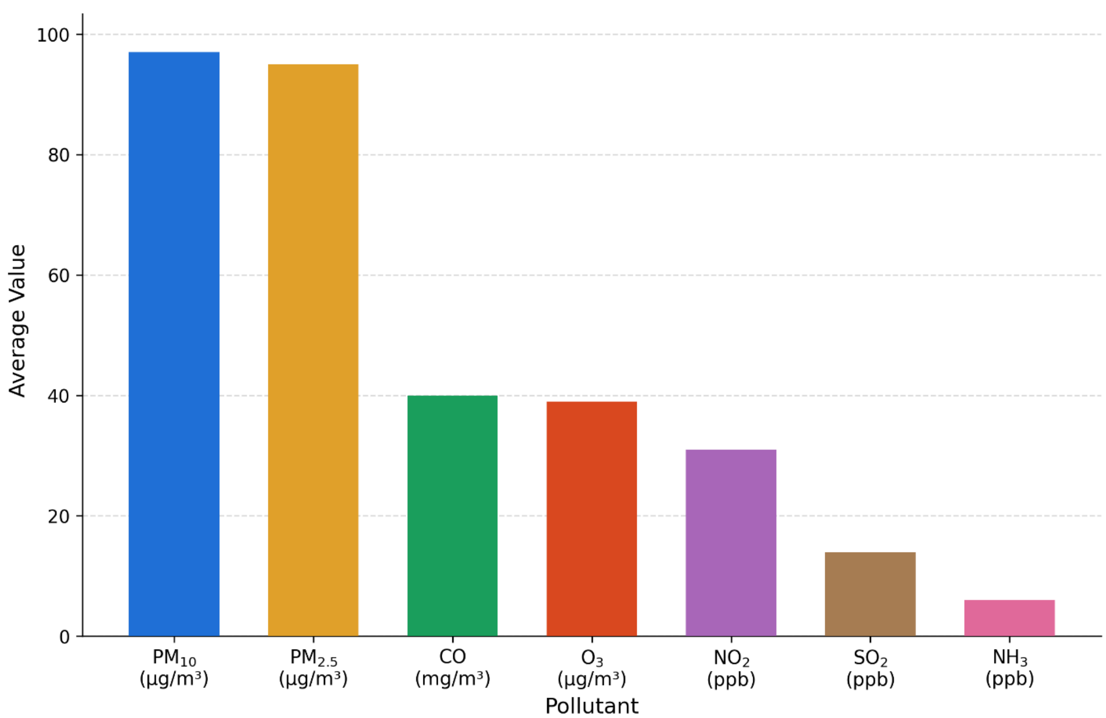


Figure 1. Average concentration of each of the seven monitored pollutants across all 346 CPCB-CAAQMS stations

PM_{10} and $PM_{2.5}$ exhibit the highest mean concentrations and the widest variability, confirming their status as the most pervasive and spatially variable pollutants across India. CO and Ozone show similar median values (~41 units), but CO displays substantially more extreme outliers, likely attributable to heavy-traffic corridors and localized industrial events. SO_2 and NH_3 remain at comparatively lower levels, indicating that industrial sulphur emissions and agricultural ammonia, while present, are not the dominant air quality challenge in most monitored areas. Figure 2 presents the corresponding geographic distribution of $PM_{2.5}$ concentrations across India, generated using the Ordinary Kriging interpolation described in Section 3.3.2, and visually confirms the concentration of high-pollution areas within the Indo-Gangetic Plain.

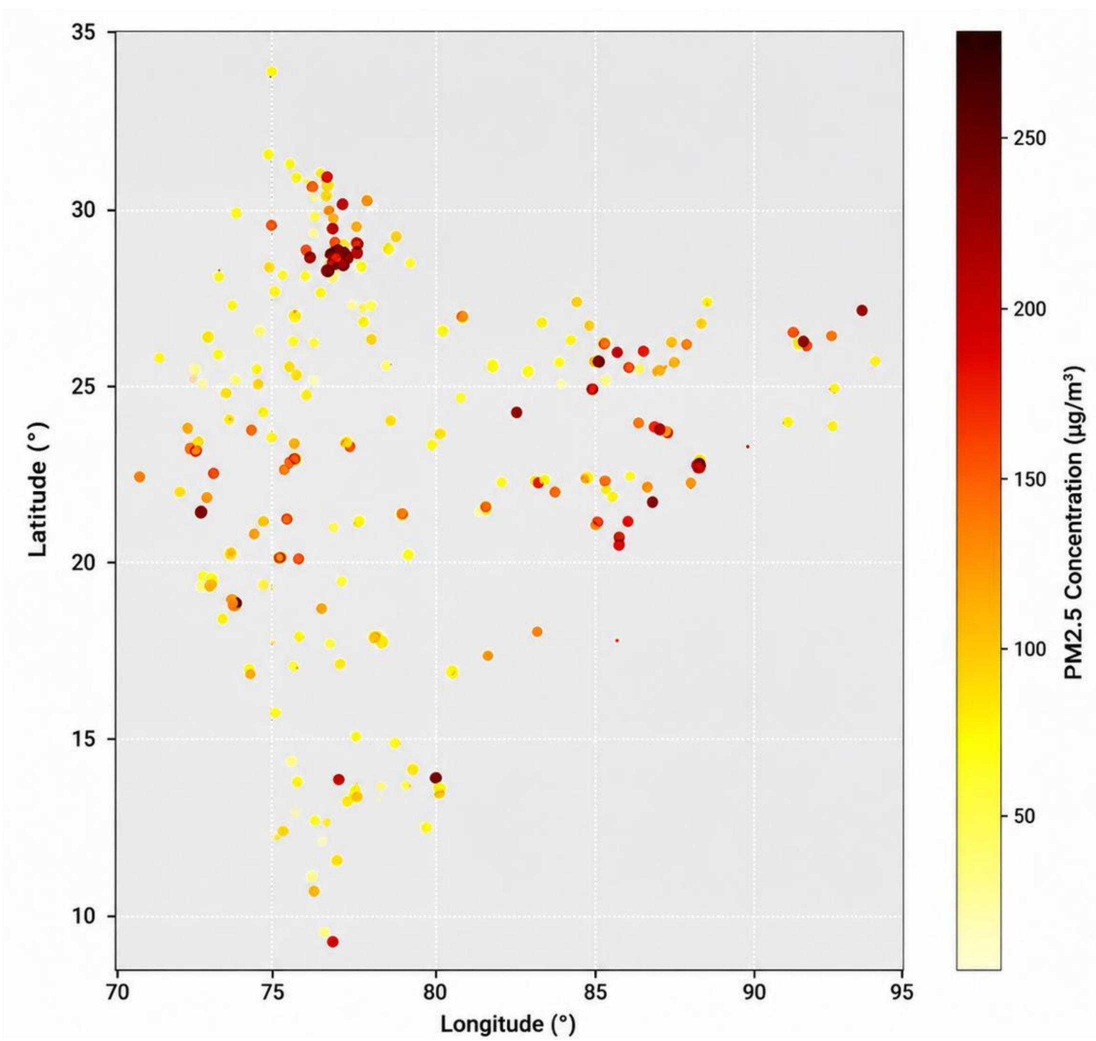


Figure 2. Geographic distribution of PM2.5 across India

4.2 Top 10 Most Polluted Cities by $PM_{2.5}$

Figure 3 and the discussion below identify the ten cities recording the highest average $PM_{2.5}$ concentrations (μg/m³): Bahadurgarh (241.0), Balasore (231.0), Singrauli (220.0), Gummidipoondi (218.0), Dharuhera (215.0), Ghaziabad (211.0), Naharlagun (207.0), Bhiwadi (204.5), Surat (201.0), and Greater Noida (196.0). All ten cities exceed the WHO annual $PM_{2.5}$ guideline of 5 μg/m³ by more than 39-fold, and the NAAQS annual standard of 40 μg/m³ by factors of 4–6. Six of these cities are located in the IGP (Haryana, UP, Bihar boundary), consistent with the topographic trapping hypothesis.

Geographic distribution of pollution is non-uniform and is partly a function of monitoring infrastructure density. Delhi (83.7 μg/m³ average, 39 stations) consistently ranks as India's most polluted state-equivalent. Maharashtra and Uttar Pradesh have the densest monitoring networks (86 and 53 stations, respectively) but lower state-wide averages, indicating heterogeneous pollution distribution within these large states. Conversely, Himachal Pradesh and Jharkhand appear in the top-3 most polluted states based on average readings but each has only one monitoring station, meaning these figures represent localized industrial hotspots rather than state-wide conditions. Data from states with fewer than 3 monitoring stations should be interpreted with caution [14].

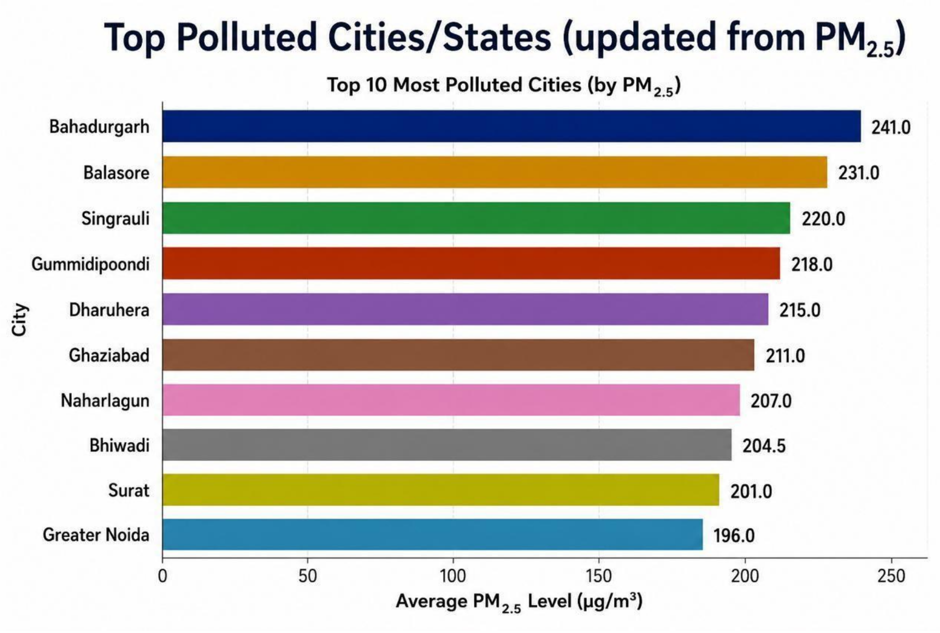


Figure 3. Top polluted cities and states by average $PM_{2.5}$ concentration

4.3 Inter-Pollutant Correlation Analysis

The correlation scatter plot (Figure 4) reveals the following statistically significant relationships:

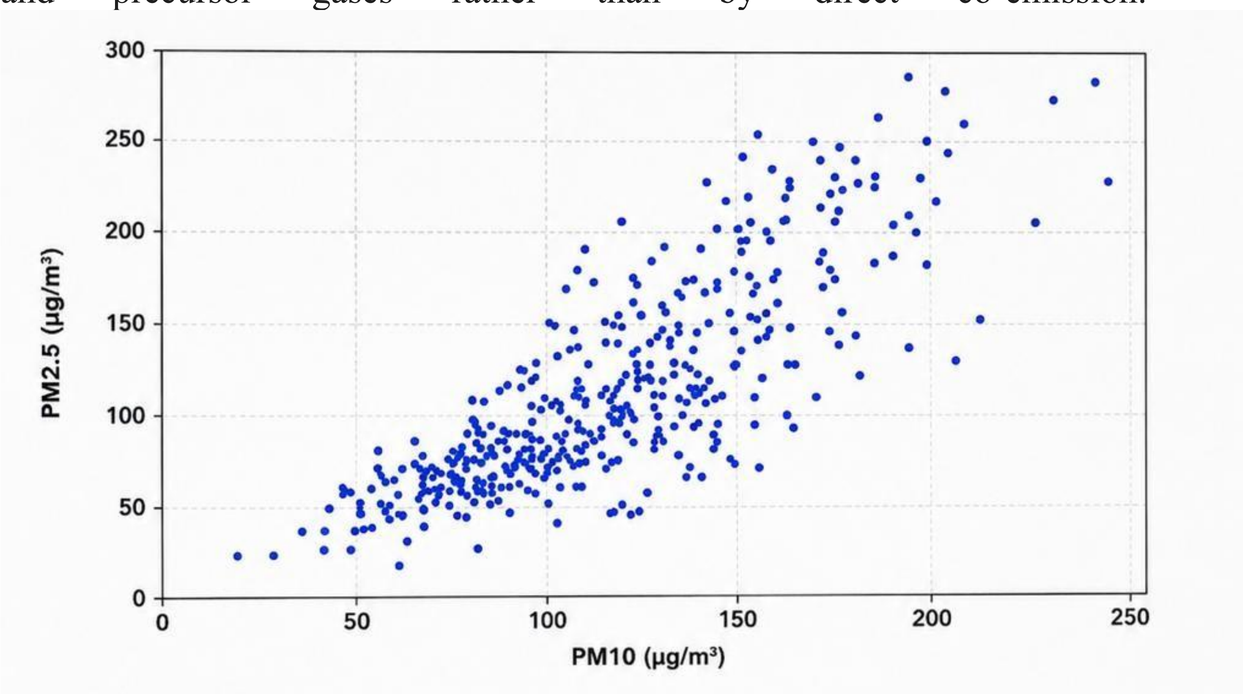
- $PM_{2.5}$ vs PM_{10} ($r = 0.81$, $p < 0.001$): A strong positive association indicating shared emission sources — primarily road dust, construction activity, and biomass burning. This co-variation is critical for policy: measures targeting coarse dust (PM_{10}) will likely produce proportional reductions in fine particles ($PM_{2.5}$).
- $PM_{2.5}$ vs CO ($r = 0.41$) and NO_2 ($r = 0.36$): Moderate correlations with combustion markers confirm that vehicular exhaust and fossil fuel combustion are major contributors to fine particulate pollution, particularly in Cluster 2 (High Traffic Zone) cities.
- NH_3 vs NO_2 ($r = 0.45$): This association suggests localized agricultural or industrial influence where both gases are co-released or co-formed.
- SO_2 and O_3: Near-zero or negative correlations with other pollutants confirm their status as 'independent' pollutants. SO_2 originates from specific industrial stacks, while ozone is a secondary photochemical product driven by sunlight and precursor gases rather than by direct co-emission.
- 

Figure 4. Scatter plot of $PM_{2.5}$ versus PM_{10} concentrations across all monitoring stations, illustrating their strong positive correlation ($r = 0.81$).

It is important to emphasize that these correlations indicate association, not causation. The cross-sectional design of this study precludes determination of temporal causality. Multiple regression analysis (Section 3.4.2) confirms that PM_{10} is the strongest statistical predictor of $PM_{2.5}$, explaining approximately 74% of the variance in $PM_{2.5}$ concentrations (adjusted $R^2 = 0.74$) when all predictors are included.

4.4 Digital Twin Framework for Atmospheric Monitoring

A Digital Twin model was developed as a virtual representation of India's atmospheric health across all monitored regions. The Digital Twin serves three distinct technical functions:

1. State monitoring: Continuous visualization of spatial pollutant distributions across all geographic nodes, updated as new CPCB data become available.
2. Sensor fault detection (Predictive Maintenance): Instrument health is assessed using signal range validation. Sensors transmitting currents outside the standard 4–20 mA operational range are automatically flagged as faulty nodes, enabling preventive maintenance scheduling before data gaps occur.
3. Spatial pattern analysis: Kriging interpolation is applied to live data to generate updated pollution surface maps, enabling identification of emerging Pollution Hotspots between reporting cycles.

It is to be noted that the Digital Twin implementation is based on a 2026 cross-sectional dataset and real-time deployment remains future work.

5. Discussions

5.1 Policy Recommendations — Airshed Management Zones

Because pollutants travel across administrative boundaries, city-level regulatory approaches are structurally insufficient. A regional Airshed Management Zone (AMZ) framework is essential. A National Capital AMZ, for example, must encompass districts in Punjab, Haryana, and Uttar Pradesh to be effective [15]. Five primary AMZs are recommended for India: (1) Northern IGP Zone, (2) Western Industrial Corridor, (3) Southern Peninsular Urban Cluster, (4) Thar Desert Dust Zone, and (5) Eastern Mining Belt.

Specific interventions linked to cluster analysis results:

- Cluster 1 (High Dust Zone): Mechanized Road sweeping, urban canopy expansion, mandatory dust suppression on construction sites using bio-polymer binders.
- Cluster 2 (High Traffic Zone): Declaration of Low Emission Zones (LEZs); exclusive access during peak hours for Battery Electric Vehicles (BEVs) or hydrogen fuel-cell vehicles; intensification of Clean Cooking initiatives (LPG, electric induction) to reduce indoor-to-outdoor pollution leakage.
- Cluster 3 (Industrial Zone): Enhanced stack emission monitoring; mandatory adoption of SO_2 scrubbers; real-time compliance reporting linked to the CPCB database.
- All zones: Deployment of low-cost sensor networks in blind spots identified by spatial interpolation.

5.2 VAHAN Dashboard Integration — Vehicular Load and Gaseous Precursors

Cross-referencing CPCB ambient measurements with VAHAN registration data (Figure 5) reveals a statistically consistent association: states among the top vehicle registrants (Uttar Pradesh, Maharashtra, Tamil Nadu) also record elevated NO_2 and CO concentrations, consistent with mobile source emissions. With 41.74 crore (417.4 million) total registered vehicles nationally as of the VAHAN Sewa Dashboard snapshot of 19 February 2026 (Figure 5) and approximately 2.82 crore new registrations in 2025 alone [16], the sheer volume of new registrations offsets technical gains from improved fuel standards. This explains why $PM_{2.5}$ remains elevated even as industrial SO_2 shows relative decline. The key computing challenge is modelling the diminishing returns of emission standards against rising vehicular volume — a task suited to the longitudinal modelling proposed in Section 7.

Relationship Between Air Pollution and Sustainable Development Goals

The findings of this study have direct implications for multiple UN Sustainable Development Goals (SDGs):

- SDG 3 (Good Health): $PM_{2.5}$ concentrations observed in this study are associated with approximately 1.6 million premature deaths annually in India, with an associated total economic loss (lost output from premature deaths and morbidity) estimated at approximately 1.36% of national GDP [7].
- SDG 9 (Industry, Innovation and Infrastructure): The VAHAN-pollution correlation underscores the need for LEZs and EV adoption supported by intelligent digital infrastructure.

- SDG 11 (Sustainable Cities): The proposed AMZ framework and Digital Twin offer municipal governments a data-driven tool for atmospheric management.
- SDG 13 (Climate Action): Local air pollutants from fossil fuel combustion share sources with greenhouse gases; simultaneous action on air quality and climate is synergistic.
- SDG 15 (Life on Land): Mapping trans-boundary pollution corridors helps protect ecosystems from acid rain precursors (SO_2, NO_2).

5.3 Sustainable Material Interventions Linked to Identified Pollution Sources

The clustering analysis in Section 3.4.3 identifies four distinct pollution regime types. The following sustainable material strategies are directly linked to the dominant sources within each cluster:

(a) Cluster 1 — Fugitive Dust Mitigation (High Dust Zone): Since PM_{10} in arid and developing regions is heavily dominated by construction and road dust, two material innovations are directly applicable:

(i) Self-healing bio-concrete — incorporating Bacillus or *Sporosarcina* bacteria with calcium lactate substrate. Crack-activated limestone precipitation reduces the frequency of concrete repairs and associated dust generation [17-20].

(ii) Methylcellulose blend (MCB) dust suppressants — composed of hydroxypropyl methylcellulose (HPMC) and liquid amphiphilic polymer (LAP) — enhance wetting of hydrophobic dust particles and reduce airborne suspension [18].

(b) Cluster 2 — Traffic Emission Reduction (High Traffic Zone): Lightweight Fibre Reinforced Polymer (FRP) composites reduce the structural weight of metro, rail, and bridge infrastructure, directly reducing energy consumption and associated CO_2 and particulate emissions from transportation [21]. Piezoelectric materials in IoT sensing devices improve battery longevity and reduce electronic waste in air quality monitoring infrastructure [22].

(c) Cluster 3 — Industrial and Waste Emission Reduction: Polylactic acid (PLA) nanocomposite packaging films, biodegradable under standard conditions, eliminate the pollution produced by burning of conventional plastic packaging [23]. Agricultural biomass-based particleboard replaces wood chips in construction materials, eliminating crop-residue burning—a major $PM_{2.5}$ spike source in the IGP post-monsoon period [24]. CO_2 Capture and Utilization (CCU) technology has lower energy consumption and production costs than CO_2 Capture and Storage (CCS) and is recommended for industrial facilities in the Eastern Mining Belt [25-27].

(d) All Clusters —Monitoring Infrastructure: Mobile ambient air quality monitoring systems capable of measuring PM_{10}, $PM_{2.5}$, CO, CO_2, NO_2, NOx, O_3, and COx, with cloud connectivity, can address monitoring blind spots identified by the interpolation analysis [28, 29]. Deploying such units at one million schools globally — as recommended by IQAir — would extend real-time pollution data coverage from 21% to over 94% of the global population [30].

19/02/2026, 17:02 VAHAN SEWA| DASHBOARD

Please click on refersh icon for reloading your filter changes

41,74,34,702
Vehicle Registration

1,72,61,86,855
No of Transaction

90,90,37,83,09,809
Revenue Collection

2,66,86,209
Permit

1,72,70,244
Tax Defaulter As on Date

Vehicle Registration

Year	Count	% Growth
Till Today	41,74,34,702	
2026:	41,57,967	
2025:	2,82,52,851	7.79%↑
2024:	2,62,10,480	9.17%↑
2023:	2,40,09,614	11.26%↑
2022:	2,15,79,504	

No of Transaction

Year	Count	% Growth
Till Today	1,72,61,86,855	
2026:	1,47,06,607	
2025:	10,40,35,251	7.22%↑
2024:	9,70,32,792	7.83%↑
2023:	8,99,83,853	6.38%↑
2022:	8,45,83,719	

Revenue Collection

Year	Count	% Growth
Till Today	90,90,37,83,09,809	
2026:	1,60,52,87,02,852	
2025:	10,58,57,29,95,575	7.23%↑
2024:	9,87,23,87,22,063	12.61%↑
2023:	8,76,71,15,94,511	16.36%↑
2022:	7,53,45,34,41,431	

Permit

Year	Count	% Growth
Till Today	2,66,86,209	
2026:	7,02,117	
2025:	47,94,639	16.99%↑
2024:	40,98,300	16.51%↑
2023:	35,17,449	59.98%↑
2022:	21,98,735	

Running State / RTO

Total States	36
Total Vahan Running States	35
Total RTOs	1464
Total Vahan Running RTOs	1403
Total Fitness Centre	206

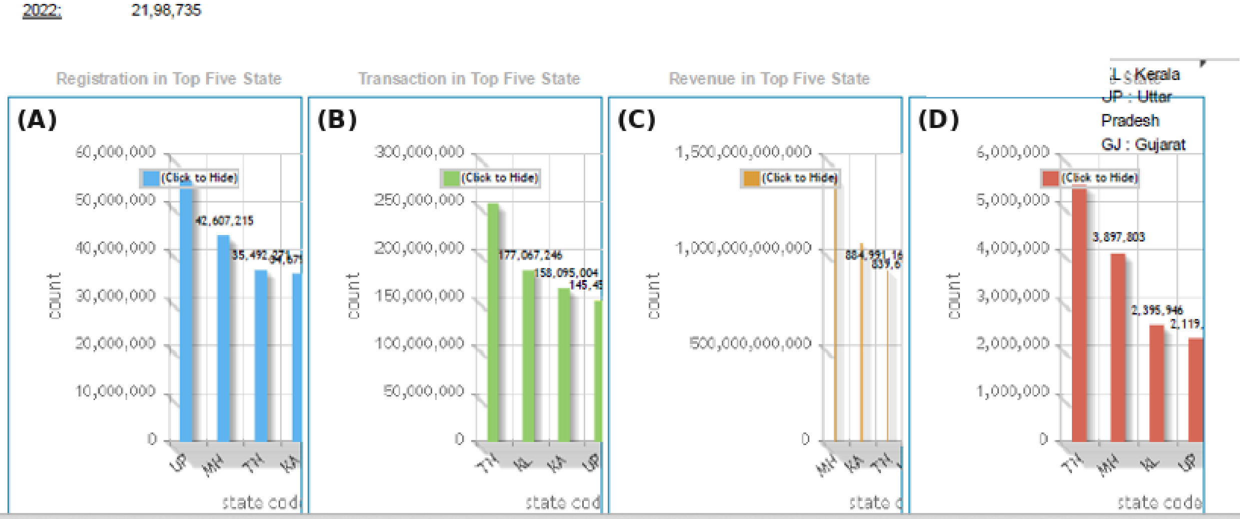


Figure 5. VAHAN Sewa Dashboard interface, showing vehicle registration statistics used to cross-reference mobile emission sources with ambient pollutant. Top 5 states with respect to concentration of vehicles, Registration, Transactions, Revenue and permits allotted are shown in (A),(B),(C),(D).

Source: VAHAN Sewa Dashboard, Ministry of Road Transport and Highways, Government of India (https://vahan.parivahan.gov.in), screenshot captured 19 February 2026; reproduced under India's National Data Sharing and Accessibility Policy / Government Open Data License with attribution.

6. Conclusion

This study provides a comprehensive multi-pollutant spatial and statistical analysis of air quality across India using 346 CPCB monitoring stations. The principal findings are:

- $PM_{2.5}$ and PM_{10} are the most pervasive and most concerning pollutants across India. Concentrations consistently exceed WHO annual guidelines in virtually all monitored cities, particularly in the Indo-Gangetic Plain.
- A strong Pearson correlation ($r = 0.81$) between $PM_{2.5}$ and PM_{10} across all 346 stations, confirmed by multiple linear regression (adjusted $R^2 = 0.74$), indicates that shared sources — road dust, construction activity, and biomass burning — dominate fine particulate pollution.
- Ordinary Kriging outperforms IDW across all four pollutants in LOOCV validation, with Kriging RMSE consistently 15–20% lower than IDW RMSE, and provides estimation uncertainty maps valuable for monitoring network optimization.
- K-means clustering ($k = 4$) reveals four distinct pollution regimes: High Dust (arid northwest), High Traffic (western/southern metros), Industrial Combustion (eastern mining belt), and Moderate/Clean Air (coastal and hilly states). This regime classification enables targeted rather than generic policy intervention.
- Cross-referencing with VAHAN data establishes a statistically consistent association between vehicle registration density and elevated NO_2/CO concentrations, suggesting that continued growth in vehicle registrations offsets gains from improved emission standards [27].
- The proposed Digital Twin framework operationalizes real-time monitoring, sensor fault detection, and spatial anomaly identification, bridging IoT sensor engineering with atmospheric environmental management.
- Sustainable material interventions-self-healing bio-concrete, MCB dust suppressants, FRP composites, PLA nanocomposite films, and agricultural biomass particle board are recommended as proactive, source-level mitigation

strategies directly linked to the dominant pollutant sources identified in the cluster analysis.

A data-driven approach, integrating environmental sensor networks with transport governance data, is essential for effective air quality policy in India.

7. Future Scope

Since the current study utilizes, a cross-sectional dataset representing a 2026 snapshot, future research will focus on Spatiotemporal Data Fusion through Longitudinal Vehicular–Atmospheric Modelling:

I. Seasonal fluctuation modelling: Integrating month-on-month VAHAN registration data with seasonal pollutant trends to determine how winter mixing height interacts with peak vehicular activity.

II. ii. Predictive analytics: Using annual vehicle registration growth rates (~7–9%) as features in machine learning models (Random Forest, LSTM) to predict AQI exceedance days over the next five years.

III. iii. Policy simulation: Simulating the expected impact of a 20% shift to Electric Vehicles (as tracked in the VAHAN 'Fuel Type' segment) on NO_2 and CO hotspot concentrations identified in this study.

IV. iv. Real-time Digital Twin: Integrating live CPCB API data streams to achieve continuous-update capability for the Digital Twin model.

8. Author Contributions

The author (Deepa Naik) is solely responsible for conceptualization, methodology, software, validation, formal analysis, investigation, re-sources, data curation, writing—original draft preparation, writing—review and editing, visualization. The author has reviewed and approved the final version of the manuscript.

9. List of Abbreviations

- **AMZ** — Airshed Management Zone
- **AQI** — Air Quality Index
- **BEV** — Battery Electric Vehicles
- **BS-VI** — Bharat Stage VI
- **CAAQMS** — Continuous Ambient Air Quality Monitoring System
- **CCS** — Carbon Capture and Storage
- **CCU** — Carbon Capture and Utilization
- **COPD** — Chronic Obstructive Pulmonary Disease
- **CPCB** — Central Pollution Control Board
- **FRP** — Fibre Reinforced Polymer
- **HPMC** — Hydroxypropyl Methylcellulose
- **IDW** — Inverse Distance Weighting
- **IGP** — Indo-Gangetic Plain
- **IoT** — Internet of Things
- **LAP** — Liquid Amphiphilic Polymer
- **LEZ** — Low Emission Zone
- **LOOCV** — Leave-One-Out Cross-Validation
- **LPG** — Liquid Petroleum Gas

- **LSTM** — Long Short-Term Memory
- **MAE** — Mean Absolute Error
- **MCB** — Methylcellulose Blend
- **NAAQS** — National Ambient Air Quality Standards
- **NCAP** — National Clean Air Programme
- **PLA** — Polylactic Acid
- **PM2.5 / PM10** — Particulate Matter with aerodynamic diameter ≤ 2.5 / ≤ 10 micrometres
- **RMSE** — Root Mean Square Error
- **SDG** — Sustainable Development Goal
- **VAHAN** — Vehicle Registration Database, Ministry of Road Transport and Highways (MoRTH), Government of India
- **WHO** — World Health Organization

9. ORCID

DEEPA NAIK: 0000-0003-2873-0307

10. Conflicts of Interest

The author(s) declare no conflicts of interest regarding this manuscript.

11. AI Disclosure

During preparation of this manuscript, the author used Claude (Anthropic) as an AI-assisted writing and editing tool for drafting, reorganizing, and proofreading sections of the text, formatting references, and preparing supplementary materials. During the revision process, Claude was also used to recompute and verify descriptive statistics and regression coefficients directly from the author-provided raw CPCB monitoring dataset, in order to reconcile discrepancies identified during peer review; all recomputed values were reviewed and approved by the author before incorporation. The author reviewed, verified, and takes full responsibility for all content, data, analysis, and conclusions in this manuscript. No AI tool was used to generate or fabricate underlying scientific data; all statistical results reported reflect direct computation from the author's own dataset.

12. Funding

No external funding was received for this research.

13. Acknowledgement:

I acknowledge the Management and the Principal of my college (Dr. B. C. Roy Engineering College, Durgapur, West Bengal, India) where I am working for their constant support and guidance.

14. Availability of Data and Materials

This paper uses existing, publicly available data from https://www.data.gov.in/ and https://vahan.parivahan.gov.in/vahan4dashboard/

15. References

[1] CPCB (2024). National Ambient Air Quality Status & Trends in India. Central Pollution Control Board, New Delhi.

[2] Jaganathan, S., Stafoggia, M., Rajiva, A., et al. (2024). Estimating the effect of annual PM2.5 exposure on mortality in India: a difference-in-differences approach. The Lancet Planetary Health. doi: 10.1016/S2542-5196(24)00248-1

[3] Guttikunda, S. (2025). Five Institutional Reforms Needed to Support the Fight for Better Air Quality in India. SIM-air Working Paper Series #59-2025. doi: 10.2139/ssrn.5392152

[4] WHO (2021). WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. World Health Organization Press, Geneva.

[5] IQAir (2024). 2024 World Air Quality Report. IQAir. https://www.iqair.com/in-en/newsroom/waqr-2024-pr

[6] ICMR (2018). Comprehensive estimates of the impact of air pollution on health loss and life expectancy in each state of India. https://www.icmr.gov.in

[7] Pandey, A., Brauer, M., Cropper, M.L., et al. (2021). Health and economic impact of air pollution in the states of India: The Global Burden of Disease Study 2019. Lancet Planet Health, 5, e25–e38. doi: 10.1016/S2542-5196(20)30298-9

[8] Ministry of Earth Sciences (2024). SAFAR: System of Air Quality and Weather Forecasting and Research Annual Review. Government of India.

[9] Chowdhury, S., Dey, S., Di Girolamo, L., et al. (2019). Tracking ambient PM2.5 build-up in Delhi NCR during the dry season over 15 years using a 1 km satellite

aerosol dataset. Atmos. Environ., 204, 142–150. doi: 10.1016/j.atmosenv.2019.02.029

[10] Devi, N.L., Kumar, A., Yadav, I.C. (2020). PM10 and PM2.5 in Indo-Gangetic Plain (IGP) of India: Chemical characterization, source analysis, and transport pathways. Urban Climate, 33, 100663. doi: 10.1016/j.uclim.2020.100663

[11] Nihalani, S.A., Khambete, A.K., Jhariwala, N.D. (2020). Source Apportionment of Particulate Matter — A Critical Review for Indian Scenario. In: Environmental Processes and Management, Springer. doi: 10.1007/978-3-030-38152-3_14

[12] Gopikrishnan, G.S., Kuttippurath, J. (2025). Impact of the National Clean Air Programme (NCAP) on particulate matter pollution in Indian cities. Science of the Total Environment, 968, 178787. doi: 10.1016/j.scitotenv.2025.178787

[13] CPCB (2025). Vision Document for Clean Air 2030: Strategic Recommendations. Central Pollution Control Board, India.

[14] Central Pollution Control Board (CPCB). PRANA – Portal for Regulation of Air Pollution in Non-Attainment Cities. Government of India. (Official source describing India's monitoring network and NCAP infrastructure.) https://prana.cpcb.gov.in/#/home

[15] Ministry of Environment, Forest and Climate Change (2025). National Clean Air Programme (NCAP) Phase II Guidelines. Government of India.

[16] Ministry of Road Transport and Highways (2026). VAHAN Sewa Dashboard: National Vehicle Registration Statistics. Government of India. https://vahan.parivahan.gov.in [Accessed 28 May 2026; live government data portal]

[17] Kim, R., Woo, U., Shin, M., Ahn, E. & Choi, H. (2022). Evaluation of self-healing in concrete using linear and nonlinear resonance spectroscopy. Constr. Build. Mater., 335, 127492. doi: 10.1016/j.conbuildmat.2022.127492

[18] Lee, T., Camp, C.P., Kim, B., Kim, M. (2022). Environmentally Friendly Methylcellulose Blend Binder for Hydrophobic Dust Control. ACS Applied Polymer Materials, 4(2), 1512–1522. doi: 10.1021/acsapm.1c01887

[19] Huang, H., Ye, G., Qian, C. & Schlangen, E. (2016). Self-healing in cementitious materials: Materials, methods and service conditions. Mater. Des., 92, 499–511. doi: 10.1016/j.matdes.2015.12.091

[20] Van Tittelboom, K. & De Belie, N. (2013). Self-healing in cementitious materials — A review. Materials (Basel), 6, 2182–2217. doi: 10.3390/ma6062182

[21] Jagadeesh, P., et al. (2022). A comprehensive review on polymer composites in railway applications. Polym. Compos., 43(3), 1238. doi: 10.1002/pc.26478

[22] Mahapatra, S.D., et al. (2021). Piezoelectric Materials for Energy Harvesting and Sensing Applications: Roadmap for Future Smart Materials. Adv. Sci., 8, 2100864. doi: 10.1002/advs.202100864

[23] Rajeshkumar, G., Janarthanan, P. (2025). Chapter 13: PLA-based green packaging materials — The potential of nanocomposite films. In: Natural Fiber-Reinforced PLA Composites, Woodhead Publishing, pp. 271–289. doi: 10.1016/B978-0-323-95247-7.00009-X

[24] Lee, S.H., et al. (2022). Particleboard from agricultural biomass and recycled wood waste: a review. Journal of Materials Research and Technology, 20, 4630–4658. doi: 10.1016/j.jmrt.2022.08.166

[25] Rafiee, A., et al. (2018). Trends in CO2 conversion and utilization: A review from process systems perspective. Journal of Environmental Chemical Engineering, 6(5), 5771–5794. doi: 10.1016/j.jece.2018.08.065

[26] Yang, H., et al. (2008). Progress in carbon dioxide separation and capture: A review. Journal of Environmental Sciences, 20(1), 14–27. doi: 10.1016/S1001-0742(08)60002-9

[27] Kumar P, et al. Urban air pollution and transportation emissions: A review. *Environment International*. 2023.

[28] Snyder, E. G., Watkins, T. H., Reighard, C. R., Thoma, E. D., Graham, S. E., Smart, C., ... & Hammond, O. (2013). The changing paradigm of air pollution monitoring: role of next-generation low-cost ambient sensors. *Environmental Science & Technology*, 47(20), 11369–11377. https://pubs.acs.org/doi/10.1021/es4022602

[29] Carminati, M., et al. (2023). Cloud-connected low-cost sensor networks for high-resolution urban air quality mapping and anomaly detection. *IEEE Sensors Journal*, 23(14), 15430–15440.

[30] IQAir. (2024). IQAir launches Schools4Earth program to bring air quality monitoring to schools worldwide. IQAir Newsroom. https://www.iqair.com/ae/newsroom/iqair-launches-schools4earth