



Assessment of Multi-Pollutant Concentrations Across Indian Urban Centers: A Case Study Towards Use of Sustainable Materials for Improvement in Air Quality

Deepa Naik 

Department of Computer Science and Engineering, Dr. B. C. Roy Engineering College, Durgapur 713206, West Bengal, India

Article History

Received: April 04, 2026

Accepted: July 30, 2026

Published: September 01, 2026

Abstract

Deteriorating air quality is a critical public health challenge worldwide, and India ranks among the nation's most severely affected. This study presents a multi-pollutant spatial analysis of air quality across 346 monitoring stations in India, drawing on 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—converting discrete station-level measurements into continuous pollution surfaces spanning Indian territory. Pearson's correlation analysis, multiple linear regression, and k-means clustering ($k = 4$) were used to identify inter-pollutant relationships and to classify monitoring regions into distinct pollution regime clusters. The results show that $PM_{2.5}$ and PM_{10} consistently exceed World Health Organization (WHO) guidelines across India, particularly within the Indo-Gangetic Plain (IGP), whereas SO_2 and CO remain within permissible limits across most regions. A strong linear correlation ($r = 0.81$) between $PM_{2.5}$ and PM_{10} points to shared sources, including road dust, construction activity, and biomass burning. Cross-referencing with the Ministry of Road Transport and Highways' VAHAN Sewa Dashboard reveals 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, and sustainable material-based mitigation strategies—including self-healing bio-concrete, reinforced PLA packaging films, and fiber-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 ranks as a leading cause of premature mortality worldwide, second only to hypertension among risk

factors 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].

As the world's most populous nation, India aspires to grow its economy from USD 4 trillion to USD 5 trillion, a target achievable through expanded construction, industrial production, energy generation, and transport infrastructure. These same activities, however, inherently generate environmental degradation that extends across

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 $50.6 \mu\text{g}/\text{m}^3$ —ten times the WHO annual guideline of $5 \mu\text{g}/\text{m}^3$ —placing it among the world's most polluted nations (Table 1) [4,5].

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

Country	$PM_{2.5}$ ($\mu\text{g}/\text{m}^3$)	Multiple of WHO Guideline ($5 \mu\text{g}/\text{m}^3$)
Chad	91.8	18×
Bangladesh	78.0	15×
Pakistan	73.7	14×
DR Congo	58.2	11×
India	50.6	10×

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) makes spatial modeling possible—a technique that reveals 'pollution corridors' formed as prevailing winds are channeled by topography, often crossing state boundaries [2]. India's varied topography, from the Himalayan foothills to the Deccan Plateau and coastal plains, governs the 'ventilation coefficient' of each city, meaning pollution severity is a product of both emissions and meteorological trapping rather than emissions alone [5]. This study applies spatial interpolation using Inverse Distance Weighting (IDW) and Ordinary Kriging to convert discrete point measurements into continuous pollution surfaces, thereby 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 in [6].

1.2.1. 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 relative to existing AQI and GIS-based studies:

- Multi-state, multi-pollutant spatial fusion: Unlike most studies, which focus on a single city or region, this analysis integrates 346 stations across all Indian states to produce a nationwide multi-pollutant spatial map at a resolution not previously reported in the open literature.
- Cross-domain data fusion (environmental and transport): The 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 through 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 the clustering and regression analysis reported in the results.

2. Literature Review—Atmospheric Dynamics and Regional Trends

The Indo-Gangetic Plain (IGP) is extensively documented as India's most severely polluted airshed, and researchers have characterized it as a 'topographic sink' in which pollutants accumulate owing to the blocking effect of the Himalayas and temperature inversions during the winter months [7,8]. Cool air trapped near the ground during the post-monsoon season produces 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 record rising ambient NO₂ 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 useful diagnostic: a high CO/PM ratio signals primary combustion emissions, whereas photochemically derived secondary pollutants point instead to regional atmospheric chemistry [11].

The distinction between PM₁₀ and PM_{2.5} is analytically critical. PM_{2.5} originates predominantly from combustion processes, whereas PM₁₀ is heavily influenced by 'fugitive dust' from construction, unpaved roads, and natural soil deflation. In arid states such as Rajasthan, natural dust significantly inflates PM₁₀ values, potentially masking underlying industrial pollution signals. The PM_{2.5}/PM₁₀ ratio therefore serves as a useful proxy for distinguishing anthropogenic (combustion) sources from lithogenic (crustal dust) sources [6].

A well-established link exists between air quality and economic productivity: 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, reflecting a persistent 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 obtained from the Continuous Ambient Air Quality Monitoring System (CAAQMS) managed by the Central Pollution Control Board (CPCB) of India ([https://www.](https://www.data.gov.in/)

[data.gov.in/](https://www.data.gov.in/)). The dataset comprises daily average concentrations recorded in February 2026 across 346 monitoring stations spanning 28 states and 8 Union Territories, covering seven criteria pollutants: PM_{2.5}, PM₁₀, CO, SO₂, NO₂, NH₃, and O₃.

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

Dataset characteristics: Approximately 3000 station-pollutant observations across 346 unique monitoring stations, with geographic coverage spanning latitude 8° N to 37° N and longitude 68° E to 97° E.

3.2. Data Preprocessing and Quality Control

Data integrity is essential for reliable spatial computational modeling. The following preprocessing steps were therefore applied:

- **Outlier exclusion:** Negative concentration values—physically impossible and 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 the 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, with the following libraries: Pandas 2.0, NumPy 1.24, GeoPandas 0.13, PyKrig 1.7, SciPy 1.11, Matplotlib 3.8, and Folium 0.15.
- **The Air Quality Index (AQI) was computed** using Indian sub-index transformation equations, in which each pollutant is weighted according to its health-risk potential, following the CPCB's standard methodology. Descriptive statistics—mean, median, standard deviation, and 98th percentile—were calculated 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 only; to estimate pollution levels at unmon-

itored locations, IDW interpolation was applied. IDW rests on Tobler's First Law of Geography, which holds that observations closer together geographically tend to be more similar than those farther apart. Under this approach, the estimated value at an unmonitored location is a weighted average of surrounding station values, with weights assigned inversely to distance:

$$\hat{Z}(x_0) = \Sigma [w_i \times Z(x_i)] / \Sigma w_i; \text{ where } w_i = 1/d_i^p \quad (1)$$

Here, $\hat{Z}(x_0)$ 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 chosen because ambient pollutant concentrations decay approximately with the square of distance from the source, consistent with Gaussian dispersion principles governing ground-level sources. At $p = 1$, weight decay is linear and overstates the influence of distant stations; at $p = 2$, the model more accurately captures 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:** With Ghaziabad $PM_{2.5} = 211 \mu g/m^3$ (5 km from the target point) and Greater Noida $PM_{2.5} = 196 \mu g/m^3$ (20 km), and $p = 2$, the weights are $w_1 = 1/5^2 = 0.04$ and $w_2 = 1/20^2 = 0.0025$, giving an estimated value of $(211 \times 0.04 + 196 \times 0.0025) / (0.04 + 0.0025) = (8.44 + 0.49) / 0.0425 = 210.1 \mu g/m^3$.

3.3.2. Ordinary Kriging

- Ordinary Kriging is a geostatistical interpolation method that, unlike IDW, yields not only an estimated value but also an estimation variance (kriging error), allowing prediction uncertainty to be quantified. Kriging derives optimal weights by modeling 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 plus microscale variation), C_1 is the partial sill, and a is the range. The fitted parameters were nugget $C_0 = 120$, sill $C_0 + C_1 = 850$, and range $a = 380$ km.

The Kriging system solves for weights λ_i subject to the unbiasedness constraint $\Sigma \lambda_i = 1$:

$$\begin{aligned} \Sigma_{j=1}^n \lambda_j \gamma(x_i, x_j) + \mu &= \gamma(x_i, x_0) \\ \text{for all } i; \Sigma_{j=1}^n \lambda_j &= 1 \end{aligned} \quad (2)$$

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

To assess interpolation accuracy, leave-one-out cross-validation (LOOCV) was performed: each station was sequentially withheld, and its concentration was predicted from the remaining stations. The resulting performance metrics are reported in [Table 2](#).

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:

$$r = \frac{\sum [(X_i - \bar{X})(Y_i - \bar{Y})]}{\sqrt{[\sum (X_i - \bar{X})^2 \times \sum (Y_i - \bar{Y})^2]}} \quad (3)$$

Here, r ranges from -1 (perfect negative correlation) to $+1$ (perfect positive correlation), with $r = 0$ indicating 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 takes the form:

$$PM_{2.5} = \beta_0 + \beta_1(CO) + \beta_2(NO_2) + \beta_3(PM_{10}) + \beta_4(NH_3) + \beta_5(SO_2) + \epsilon \quad (4)$$

Results: Adjusted $R^2 = 0.66$, with PM_{10} 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 likewise not significant ($p = 0.33$), 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, whereas the latter is a simple bivariate association.

Table 2: LOOCV performance: RMSE and MAE ($\mu\text{g}/\text{m}^3$) for IDW versus ordinary Kriging across key pollutants. Kriging consistently outperforms IDW.

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

3.4.3. K-Means Clustering

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

- Cluster 1—High Dust Zone (PM₁₀-dominant): Rajasthan, Gujarat, and the arid belt, characterized by a high PM₁₀/PM_{2.5} ratio (>2.5) indicating crustal dust dominance.
- Cluster 2—High Traffic Zone (NO₂ and CO elevated): Maharashtra, Karnataka, and Tamil Nadu metros, showing a CO–NO₂ co-elevation pattern consistent with fossil fuel combustion.
- Cluster 3—Industrial Combustion Zone (PM_{2.5}, SO₂ elevated): the eastern belt, comprising Jharkhand, West Bengal, and Odisha, where higher SO₂ 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 shaped 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, which amplifies concentrations even without any increase in emissions. By incorporating meteorological variables, the model distinguishes cities with genuinely high emissions from those that are simply '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 the 346 stations, while Figure 1 illustrates the average concentration of each pollutant across all stations.

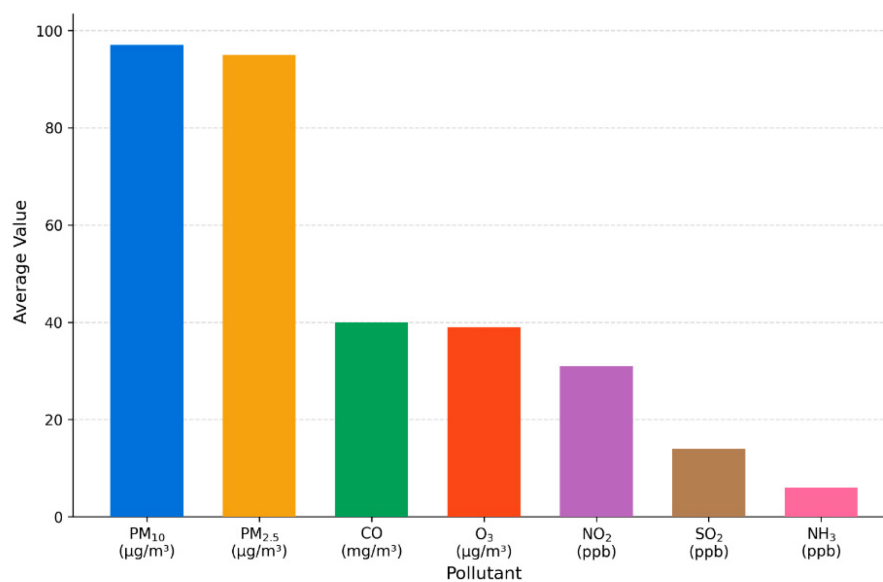


Figure 1: Average concentration of each of the seven monitored pollutants across all 346 CPCB-CAAQMS stations.

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/m^3 , while the remaining pollutants are reported in $\mu\text{g}/\text{m}^3$ (PM_{10} , $\text{PM}_{2.5}$, O_3) or ppb (NO_2 , SO_2 , NH_3). Values are therefore not directly comparable across rows without accounting for this difference.

Pollutant	Mean	Median	Std Dev	98th Percentile
PM_{10} ($\mu\text{g}/\text{m}^3$)	97.9	87.4	52.1	226.3
$\text{PM}_{2.5}$ ($\mu\text{g}/\text{m}^3$)	95.0	84.6	49.8	215.8
CO (mg/m^3)	40.9	31.0	38.4	142.5
O_3 (Ozone) ($\mu\text{g}/\text{m}^3$)	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

PM_{10} and $\text{PM}_{2.5}$ 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 mean values (~ 40 in their respective units), yet CO displays substantially more extreme outliers, likely reflecting heavy-traffic corridors and localized industrial events. SO_2 and NH_3 remain comparatively low, 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 $\text{PM}_{2.5}$ 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.

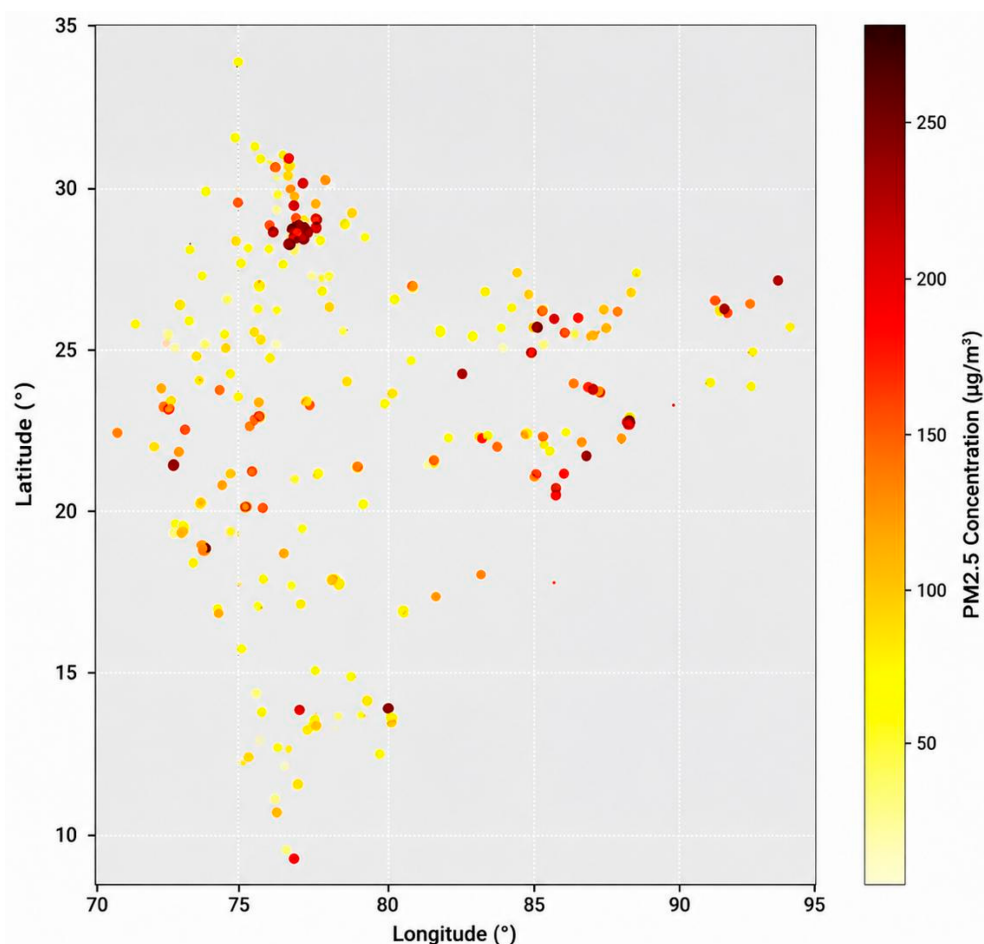


Figure 2: Geographic distribution of $\text{PM}_{2.5}$ across India.

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 lie within the IGP (the Haryana–Uttar Pradesh–Bihar region), consistent with the topographic trapping hypothesis.

(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 lie within the IGP (the Haryana–Uttar Pradesh–Bihar region), consistent with the topographic trapping hypothesis.

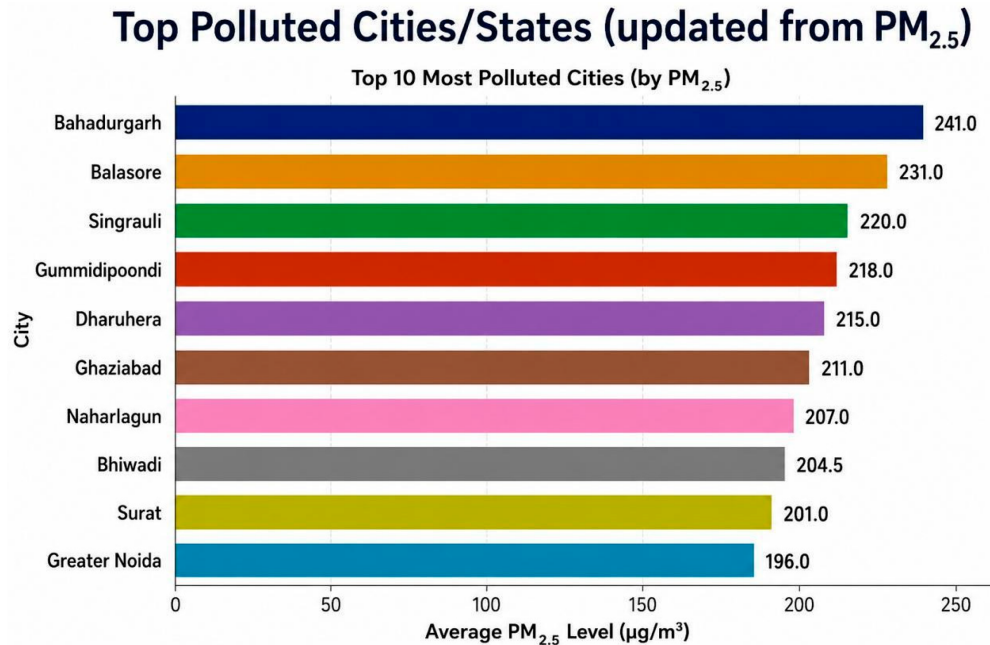


Figure 3: Top polluted cities and states by average PM_{2.5} concentration.

The geographic distribution of pollution is non-uniform and is shaped partly by the density of monitoring infrastructure. Delhi (83.7 µg/m³ average, 39 stations) consistently ranks as India's most polluted state-equivalent. Maharashtra and Uttar Pradesh, despite having the densest monitoring networks (86 and 53 stations, respectively), show lower state-wide averages, pointing to heterogeneous pollution distribution within these large states. Himachal Pradesh and Jharkhand, by contrast, appear among the top three most polluted states by average readings, yet each has only a single monitoring station, meaning these figures likely represent localized industrial hotspots rather than state-wide conditions. Data from states with fewer than three monitoring stations should accordingly be interpreted with caution [14].

4.3. Inter-Pollutant Correlation Analysis

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

- PM_{2.5} vs. PM₁₀ ($r = 0.81$, $p < 0.001$): A strong positive association indicating shared emis-

sion sources—primarily road dust, construction activity, and biomass burning. This co-variation carries clear policy relevance: measures targeting coarse dust (PM₁₀) are likely to produce proportional reductions in fine particles (PM_{2.5}) as well.

- PM_{2.5} vs. CO ($r = 0.41$) and NO₂ ($r = 0.36$): Moderate correlations with these combustion markers confirm that vehicular exhaust and fossil fuel combustion contribute substantially to fine particulate pollution, particularly in Cluster 2 (High Traffic Zone) cities.
- NH₃ vs. NO₂ ($r = 0.45$): This association points to localized agricultural or industrial influence, where both gases are co-released or co-formed.
- SO₂ and O₃: Near-zero or negative correlations with other pollutants confirm their status as 'independent' pollutants. SO₂ originates from specific industrial stacks, while ozone is instead a secondary photochemical product driven by sunlight and precursor gases rather than direct co-emission.

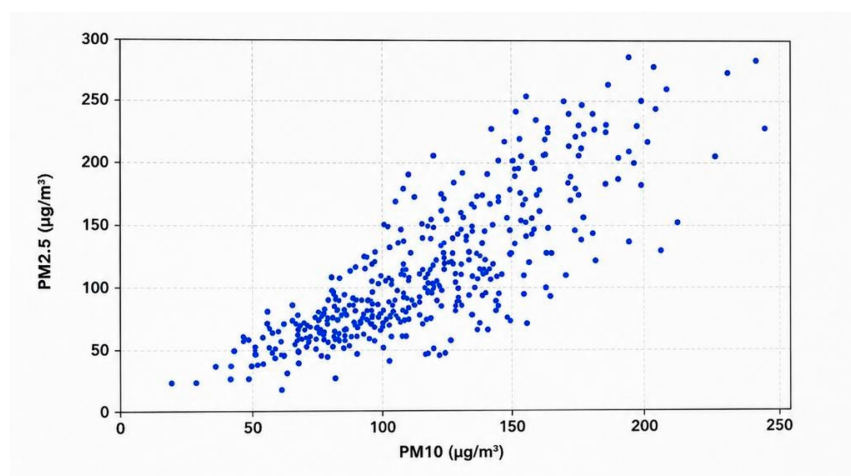


Figure 4: Scatter plot of PM_{2.5} versus PM₁₀ concentrations across all monitoring stations, illustrating their strong positive correlation ($r = 0.81$).

It is important to stress that these correlations indicate association rather than causation; the cross-sectional design of this study precludes any determination of temporal causality. Multiple regression analysis (Section 3.4.2) confirms that PM₁₀ is the strongest statistical predictor of PM_{2.5}, accounting for approximately 74% of the variance in PM_{2.5} concentrations (adjusted $R^2 = 0.66$) when all predictors are included.

4.4. Digital Twin Framework for Atmospheric Monitoring

A Digital Twin model was developed to serve as a virtual representation of India's atmospheric health across all monitored regions, fulfilling 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 through signal range validation, with sensors transmitting currents outside the standard 4–20 mA operational range 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 emerging pollution hotspots to be identified between reporting cycles.

It should be noted that the Digital Twin implementation described here is based on a 2026 cross-sectional dataset; 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 on their own. A regional Airshed Management Zone (AMZ) framework is therefore 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) the Northern IGP Zone, (2) the Western Industrial Corridor, (3) the Southern Peninsular Urban Cluster, (4) the Thar Desert Dust Zone, and (5) the Eastern Mining Belt.

The following interventions are linked directly to the cluster analysis results:

- Cluster 1 (High Dust Zone): Mechanized road sweeping, urban canopy expansion, and mandatory dust suppression on construction sites using bio-polymer binders.
- Cluster 2 (High Traffic Zone): Declaration of Low Emission Zones (LEZs); exclusive peak-hour access for Battery Electric Vehicles (BEVs) or hydrogen fuel-cell vehicles; and intensified Clean Cooking initiatives (LPG, electric induction) to reduce indoor-to-outdoor pollution leakage.
- Cluster 3 (Industrial Zone): Enhanced stack emission monitoring, mandatory adoption of SO₂ scrubbers, and real-time compliance reporting linked to the CPCB database.
- All zones: Deployment of low-cost sensor networks in blind spots identified through 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 ranking among the top vehicle registrants—Uttar Pradesh, Maharashtra, and Tamil Nadu—also record elevated NO₂ 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 achieved through improved fuel standards. This helps explain why PM_{2.5} remains elevated. The key computational challenge lies in modeling the diminishing returns of emission standards against rising vehicular volume—a task suited to the longitudinal modeling proposed in Section 7.

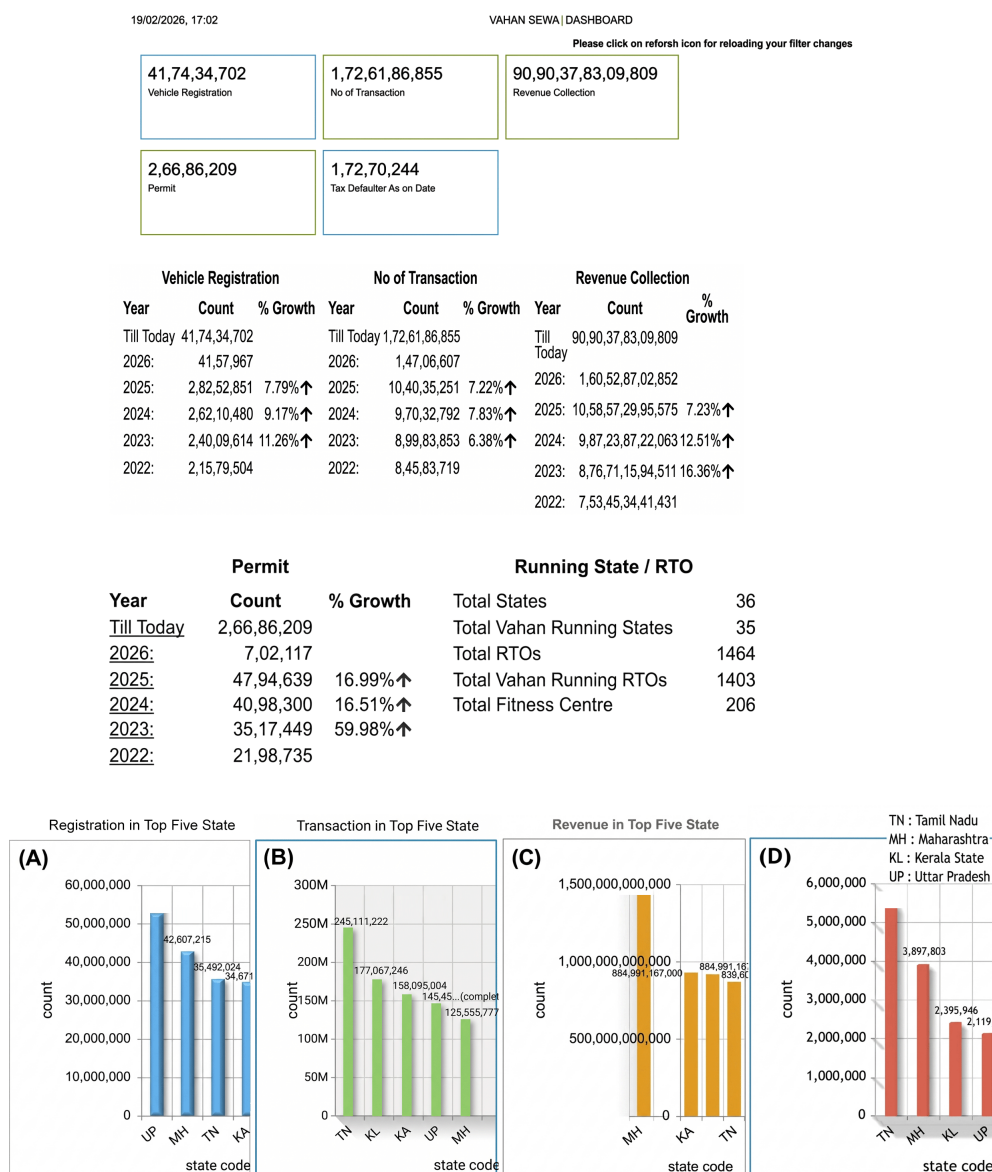


Figure 5: VAHAN Sewa Dashboard interface (snapshot 19 February 2026) used to cross-reference mobile emission sources with ambient pollutant concentrations. (A) Top five states by total registered vehicles. (B) Top five states by number of transactions. (C) Top five states by revenue collection. (D) Top five states by permits allotted. Source: VAHAN Sewa Dashboard, Ministry of Road Transport and Highways, Government of India (<https://analytics.parivahan.gov.in/analytics/publicdashboard/vahan?lang=en>), reproduced under India's National Data Sharing and Accessibility Policy / Government Open Data License with attribution [16].

5.2.1. Relationship Between Air Pollution and Sustainable Development Goals

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

- SDG 3 (Good Health and Well-Being): PM_{2.5} concentrations observed in this study are associated with a substantial fraction of the approximately 0.98 million deaths attributable to ambient particulate-matter pollution in India (part of the broader 1.67 million deaths from all air pollution sources reported by Pandey et al. [7]), with an associated total economic loss estimated at approximately 1.36% of national GDP.
- 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 together 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, making simultaneous action on air quality and climate mutually reinforcing.
- SDG 15 (Life on Land): Mapping trans-boundary pollution corridors helps protect ecosystems from acid rain precursors (SO₂, NO₂).

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):** Because PM₁₀ 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 a calcium lactate substrate. Crack-activated limestone precipitation reduces the frequency of concrete repairs and the associated dust generation [17–20].
 - (ii) Methylcellulose blend (MCB) dust suppressants—composed of hydroxypropyl methylcellulose (HPMC) and liquid amphiphilic polymer (LAP)—enhance the wetting of hydrophobic dust particles and reduce airborne suspension [18].

- (b) **Cluster 2—Traffic Emission Reduction (High Traffic Zone):** Lightweight fiber-reinforced polymer (FRP) composites reduce the structural weight of metro, rail, and bridge infrastructure, directly lowering energy consumption and the associated CO₂ and particulate emissions from transportation [21]. Piezoelectric materials in IoT sensing devices further 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 generated by burning conventional plastic packaging [23]. Agricultural biomass-based particleboard replaces wood chips in construction materials, eliminating crop-residue burning—a major source of PM_{2.5} spikes in the IGP during the post-monsoon period [24]. CO₂ Capture and Utilization (CCU) technology, which carries lower energy consumption and production costs than CO₂ Capture and Storage (CCS), 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 PM₁₀, PM_{2.5}, CO, CO₂, NO₂, NO_x, O₃, and CO_x, equipped with cloud connectivity, can help close monitoring blind spots identified by the interpolation analysis [28,29]. Deploying such units across 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].

6. Conclusions

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

- PM_{2.5} and PM₁₀ are the most pervasive and most concerning pollutants across India, with concentrations consistently exceeding WHO annual guidelines in virtually all monitored cities, particularly within the Indo-Gangetic Plain.
- A strong Pearson correlation ($r = 0.81$) between PM_{2.5} and PM₁₀ across all 346 stations, confirmed by multiple linear regression (adjusted $R^2 = 0.66$), 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 it further 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 and CO concentrations, suggesting that continued growth in vehicle registrations is offsetting gains from improved emission standards [27].
- The proposed Digital Twin framework is designed to support 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 particleboard—are recommended as proactive, source-level mitigation strategies directly linked to the dominant pollutant sources identified in the cluster analysis.

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

7. Future Scope

Since the current study relies on a cross-sectional dataset representing a 2026 snapshot, future research will focus on spatiotemporal data fusion through longitudinal vehicular–atmospheric modeling:

- Seasonal fluctuation modeling: Integrating month-on-month VAHAN registration data with seasonal pollutant trends to determine how winter mixing height interacts with peak vehicular activity.
- Predictive analytics: Using annual vehicle registration growth rates ($\sim 7\text{--}9\%$) as features in machine learning models (random forest, LSTM) to predict AQI exceedance days over the next five years.

- Policy simulation: Simulating the expected impact of a 20% shift to electric vehicles (as tracked in the VAHAN 'Fuel Type' segment) on the NO_2 and CO hotspot concentrations identified in this study.
- Real-time Digital Twin: Integrating live CPCB API data streams to achieve continuous update capability for the Digital Twin model.

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	Fiber-reinforced Polymer
GDP	Gross Domestic Product
GIS	Geographic Information System
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	Liquefied 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
$\text{PM}_{2.5}/\text{PM}_{10}$	Particulate Matter with aerodynamic diameter $\leq 2.5/\leq 10$ 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

Author Contributions

The author (Deepa Naik) is solely responsible for conceptualization, methodology, software, validation, formal

analysis, resources, data curation, writing—original draft preparation, writing—review and editing, and visualization. The author has read and agreed to the final version of the manuscript.

Availability of Data and Materials

This paper uses existing publicly available data from <https://www.data.gov.in/> and the VAHAN Sewa Dashboard (<https://analytics.parivahan.gov.in/analytics/publicdashboard/vahan?lang=en>). The VAHAN snapshot used in the analysis was captured on 19 February 2026.

Conflicts of Interest

The author declares no conflicts of interest.

Funding

The study did not receive any external funding and was conducted using only institutional resources.

Acknowledgments

The author gratefully acknowledges the Management and the Principal of Dr. B. C. Roy Engineering College, Durgapur, West Bengal, India, for their constant support and guidance.

AI Declaration

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.

Supplementary Materials

Supplementary material associated with this article has been published online and is available at: <https://doi.org/10.69709/ESM.2026.551166>.

References

[1] CPCB. *National Ambient Air Quality Status & Trends in India*; Central Pollution Control Board: New

Delhi, India, 2024. Available online: <https://cpcb.nic.in/about-namp/> (accessed on 19 February 2026).

- [2] Jaganathan, S.; Stafoggia, M.; Rajiva, A.; Mandal, S.; Dixit, S.; de Bont, J.; Wellenius, G.A.; Lane, K.J.; Nori-Sarma, A.; Kloog, I.; et al. Estimating the Effect of Annual PM_{2.5} Exposure on Mortality in India: A Difference-in-Differences Approach. *Lancet Planet. Health* **2024**, *8*, e987–e996. [CrossRef] [PubMed]
- [3] Guttikunda, S. *Five Institutional Reforms Needed to Support the Fight for Better Air Quality in India*; SIM-air Working Paper Series #59-2025; UrbanEmissions. Info: New Delhi, India, 2025. [CrossRef]
- [4] WHO. *WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*; World Health Organization: Geneva, Switzerland, 2021; Available online: <https://www.who.int/publications/i/item/9789240034228> (accessed on 19 February 2026).
- [5] IQAir. 2024 World Air Quality Report. 2024. Available online: <https://www.iqair.com/in-en/newsroom/waqr-2024-pr> (accessed on 19 February 2026).
- [6] ICMR. Comprehensive Estimates of the Impact of Air Pollution on Health Loss and Life Expectancy in Each State of India. 2018. Available online: https://www.icmr.gov.in/icmr/object/custom_data/1702896411_press.pdf (accessed on 19 February 2026).
- [7] Pandey, A.; Brauer, M.; Cropper, M.L.; Balakrishnan, K.; Mathur, P.; Dey, S.; Turkoglu, B.; Kumar, G.A.; Khare, M.; Beig, G.; et al. Health and Economic Impact of Air Pollution in the States of India: The Global Burden of Disease Study 2019. *Lancet Planet. Health* **2021**, *5*, e25–e38. [CrossRef] [PubMed]
- [8] Ministry of Earth Sciences. *SAFAR: System of Air Quality and Weather Forecasting and Research Annual Review*; Government of India: New Delhi, India, 2024. Available online: <https://safari.tropmet.res.in> (accessed on 19 February 2026).
- [9] Chowdhury, S.; Dey, S.; Di Girolamo, L.; Smith, K.R.; Pillarisetti, A.; Lyapustin, A. Tracking Ambient PM_{2.5} Build-Up in Delhi National Capital Region During the Dry Season over 15 Years Using a High-Resolution (1 km) Satellite Aerosol Dataset. *Atmospheric Environ.* **2019**, *204*, 142–150. [CrossRef]
- [10] Devi, N.L.; Kumar, A.; Yadav, I.C. PM₁₀ and PM_{2.5} in Indo-Gangetic Plain (IGP) of India: Chemical Characterization, Source Analysis, and Transport Pathways. *Urban Clim.* **2020**, *33*, 100663. [CrossRef]
- [11] Nihalani, S.A.; Khambete, A.K.; Jhariwala, N.D. Source Apportionment of Particulate Matter—A Critical Review for Indian Scenario. In *Environmental Processes and Management*; Singh, R., Shukla, P., Singh, P., Eds.; Environmental Processes and Management, Water Science and Technology Library, vol 91.; Springer: Cham, Switzerland, 2020. [CrossRef]
- [12] Gopikrishnan, G.; Kuttippurath, J. Impact of the National Clean Air Programme (NCAP) on the Partic-

- ulate Matter Pollution and Associated Reduction in Human Mortalities in Indian Cities. *Sci. Total. Environ.* **2025**, 968, 178787. [CrossRef] [PubMed]
- [13] Central Pollution Control Board. *PRANA: Portal for Regulation of Air Pollution in Non-Attainment Cities*. Available online: <https://prana.cpcb.gov.in/#/about> (accessed on 19 February 2026).
- [14] Central Pollution Control Board. *PRANA: Portal for Regulation of Air Pollution in Non-Attainment Cities*. Government of India. (Official Source Describing India's Monitoring Network and NCAP Infrastructure). Available online: <https://prana.cpcb.gov.in/> (accessed on 19 February 2026).
- [15] Ministry of Environment, Forest and Climate Change. *NCAP Targets to Achieve Reductions up to 40% of PM10 Concentrations by 2025-26*; Press Information Bureau, Government of India: New Delhi, India, 2023. Available online: <https://pib.gov.in/PressReleasePage.aspx?PRID=1914423> (accessed on 19 February 2026).
- [16] Ministry of Road Transport and Highways. *VAHAN Sewa Dashboard: National Vehicle Registration Statistics*; Government of India: New Delhi, India, 2026. Available online: <https://analytics.parivahan.gov.in/analytics/publicdashboard/vahan?lang=en> (accessed on 19 February 2026).
- [17] Kim, R.; Woo, U.; Shin, M.; Ahn, E.; Choi, H. Evaluation of Self-Healing in Concrete Using Linear and Nonlinear Resonance Spectroscopy. *Constr. Build. Mater.* **2022**, 335, 127492. [CrossRef]
- [18] Lee, T.; Camp, C.P.; Kim, B.; Kim, M. Environmentally Friendly Methylcellulose Blend Binder for Hydrophobic Dust Control. *ACS Appl. Polym. Mater.* **2022**, 4, 1512–1522. [CrossRef]
- [19] Huang, H.; Ye, G.; Qian, C.; Schlangen, E. Self-Healing in Cementitious Materials: Materials, Methods and Service Conditions. *Mater. Des.* **2016**, 92, 499–511. [CrossRef]
- [20] Van Tittelboom, K.; De Belie, N. Self-Healing in Cementitious Materials—A Review. *Materials* **2013**, 6, 2182–2217. [CrossRef] [PubMed]
- [21] Jagadeesh, P.; Puttegowda, M.; Oladijo, O.P.; Lai, C.W.; Gorbatyuk, S.; Matykiewicz, D.; Rangappa, S.M.; Siengchin, S. A Comprehensive Review on Polymer Composites in Railway Applications. *Polym. Compos.* **2022**, 43, 1238–1251. [CrossRef]
- [22] Das Mahapatra, S.; Mohapatra, P.C.; Aria, A.I.; Christie, G.; Mishra, Y.K.; Hofmann, S.; Thakur, V.K. Piezoelectric Materials for Energy Harvesting and Sensing Applications: Roadmap for Future Smart Materials. *Adv. Sci.* **2021**, 8, 2100864. [CrossRef] [PubMed]
- [23] Rajeshkumar, G.; Janarthanan, P. Chapter 13: PLA-Based Green Packaging Materials—The Potential of Nanocomposite Films. In *Natural Fiber-Reinforced PLA Composites*; Chandrasekar, M., Krishnasamy, S., Kumar Thiagamani, S.M., Ishak, M.R.B., Eds.; Woodhead Publishing: Cambridge, UK, 2025; pp. 271–289. [CrossRef]
- [24] Lee, S.H.; Lum, W.C.; Boon, J.G.; Kristak, L.; Antov, P.; Pędzik, M.; Rogoziński, T.; Taghiyari, H.R.; Lubis, M.A.R.; Fatriasari, W.; et al. Particleboard from Agricultural Biomass and Recycled Wood Waste: A Review. *J. Mater. Res. Technol.* **2022**, 20, 4630–4658. [CrossRef]
- [25] Rafiee, A.; Khalilpour, K.R.; Milani, D.; Panahi, M. Trends in CO₂ Conversion and Utilization: A Review from Process Systems Perspective. *J. Environ. Chem. Eng.* **2018**, 6, 5771–5794. [CrossRef]
- [26] Yang, H.; Xu, Z.; Fan, M.; Gupta, R.; Slimane, R.B.; Bland, A.E.; Wright, I. Progress in Carbon Dioxide Separation and Capture: A Review. *J. Environ. Sci.* **2008**, 20, 14–27. [CrossRef] [PubMed]
- [27] Lakshmanan, S.; Upadhyay, A.; Kumar, N.; Bhat-tacharya, S. Region-wise and state-wise synthesis of vehicular emissions in India and their mitigation due to vehicular emissions standards. *Sci. Total Environ.* **2023**, 900, 165838. [CrossRef] [PubMed]
- [28] Snyder, E.G.; Watkins, T.H.; Solomon, P.A.; Thoma, E.D.; Williams, R.W.; Hagler, G.S.W.; Shelow, D.; Hindin, D.A.; Kilaru, V.J.; Preuss, P.W. The Changing Paradigm of Air Pollution Monitoring. *Environ. Sci. Technol.* **2013**, 47, 11369–11377. [CrossRef] [PubMed]
- [29] Zaidan, M.A.; Xie, Y.; Motlagh, N.H.; Wang, B.; Nie, W.; Nurmi, P.; Tarkoma, S.; Petäjä, T.; Ding, A.; Kulmala, M. Dense Air Quality Sensor Networks: Validation, Analysis, and Benefits. *IEEE Sens. J.* **2022**, 22, 23507–23520. [CrossRef]
- [30] IQAir. IQAir Launches Schools4Earth Program to Bring Air Quality Monitoring to Schools Worldwide. 2025. Available online: <https://www.iqair.com/ae/newsroom/iqair-launches-schools4earth> (accessed on 19 February 2026).

Disclaimer/Publisher's Note: The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the publisher or editors. The publisher and editors assume no responsibility for any injury or damage resulting from the use of information contained herein.

Maps/Territorial Claims: Scifiniti Publishing maintains a neutral stance on territorial or jurisdictional claims depicted in maps and on the institutional affiliations stated by authors.