Indian Journal of
Modern Research and Reviews

This Journal is a member of the *'Committee on Publication Ethics'*

Online ISSN:2584-184X

Research Article

Scalable Spatiotemporal Estimation of Particulate Matter via Multi-Source Data Fusion and Deep Learning: A Comprehensive Review

Dhwani Chheda [1], **Reshma Jaiswar** [2], **Iqra Miyaji** [3*], **Arundhati Niwatkar** [4]
[1-4] Usha Mittal Institute of Technology S.N.D.T. Women's University Mumbai, India

Corresponding Author: *Iqra Miyaji

DOI: https://doi.org/10.5281/zenodo.19233715

Abstract

Estimating and forecasting fine particulate matter ($PM_{2.5}$ and PM_{10}) accurately is critical for public health and protecting the environment. Ground-based measurement networks have high accuracy and precision, but for the rapid urbanisation scenario in India, the coverage is too sparse. To address this problem, the research community has more and more started to combine satellite-based Aerosol Optical Depth (AOD) data and meteorological reanalysis data using Machine Learning (ML) methods. This paper is the first to review the evolution of this technique, outlining the fusion of various data sources. It analyses the use of the low-earth orbiting MODIS and VIIRS sensors, and the more recently used high- frequency geostationary meteorological satellites such as INSAT- 3DR. The review also analyses the atmospheric physics involved in estimating surface PM from column-integrated AOD. The paper also outlines the predictive modelling techniques using traditional statistical regression and more advanced ensemble models such as Random Forest, XGBoost and Long Short-Term Memory networks (LSTM), which are used for spatiotemporal deep learning. Lastly, this paper analyses the extensive research needed in vertical aerosol profiling, spatial nonstationarity and computational time.

Manuscript Information

- **ISSN No:** 2584-184X
- **Received:** 20-01-2026
- **Accepted:** 25-02-2026
- **Published:** 26-03-2026
- **MRR:**4(3); 2026: 352-356
- **©2026, All Rights Reserved**
- **Plagiarism Checked:** Yes
- **Peer Review Process:** Yes

How to Cite this Article

Chheda D, Jaiswar R, Miyaji I, Niwatkar A**.** Scalable spatiotemporal estimation of particulate matter via multi-source data fusion and deep learning: A comprehensive review. Indian J Mod Res Rev. 2026;4(3):352-356.

Access this Article Online

www.multiarticlesjournal.com

KEYWORDS: Aerosol Optical Depth (AOD), Air Quality Index (AQI), Deep Learning, INSAT 3DR, LSTM, Particulate Matter.

1. INTRODUCTION

The World Health Organisation places air pollution as one of the biggest threats to the environment and human health. Globally, air pollution contributes to 8 million premature deaths every year [1] [2]. An example of air pollution that has been studied extensively, because of its negative health impacts, is particulate matter (PM). PM is commonly separated into coarse (PM_{10}, 10 μm) and fine ($PM_{2.5}$, 2.5 μm) particles. PM2.5 aggravates cardiovascular diseases, chronic respiratory disease, and neurodegenerative diseases [4]. The United States Environmental Protection Agency and the Central Pollution Control Board of India both use AQI to simplify complicated multi-pollutant data to help the public. Several criticisms have been levelled against the AQI system (see Table I).

[2]. Critically, the breakpoint concentrations that define each system's health risk categories differ widely. In addition, air quality assessment has largely relied on a small number of air quality monitoring stations. [6] [7]. Even though air quality monitoring stations provide very precise measurements, the cost of installation and maintenance of air quality monitoring stations has led to a very uneven. distribution [8] [9] [10]

Table 1: Comparison of AQI Categories and PM2.5 Breakpoints (US EPA VS. INDIAN CPCB) [2]

US EPA Descriptor	**Indian CPCB Descriptor**	**AQI Range**	**PM2.5 Range** ($\mu g/m^3$)
Good	Good	0 50	0.0 30.0
Moderate	Satisfactory	51 100	31.0 60.0
Unhealthy for Sensitive	Moderately Polluted	101 150	61.0 90.0
Unhealthy	Poor	151 200	91.0 120.0
Very Unhealthy	Very Poor	201 300	121.0 250.0
Hazardous	Severe	301 500	> 250.0

2. Satellite Remote Sensing and AOD Physics

AOD is the basic measure used in monitoring space-based air quality. According to Beer Lambert Bouguer law, AOD is a unitless ratio of the summative light extinction of sunlight (scattering and absorption) by aerosols along a vertical column of the atmosphere. [2] [11].

A. Orbital Platforms: Polar vs. Geostationary

The choice of satellite sensor introduces a fundamental tradeoff between spatial detail and temporal frequency (Table II) [2], [12].

- **Polar Orbiting Satellites:** Instruments like the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard Terra and Aqua provide global coverage with refined spatial resolutions of up to 1 km using MAIAC algorithms [2] [13]. However, their low Earth orbits restrict observations to one or two daily overpasses, failing to capture diurnal pollution fluctuations [8] [14].
- **Geostationary Satellites:** The Indian National Satellite System (INSAT 3D/3DR) maintains a fixed position relative to the Earth. The INSAT imager calculates AOD at a high temporal frequency (every 30 minutes) over the Indian subcontinent [15] [16] [17]. This temporal density is indispensable for tracking dynamic events such as dust storms from the Thar Desert and regional biomass burning plumes [18] [19].

Table 2: Overview of Key Satellite Sensors for Aerosol Monitoring [2]

MODIS MISR	Polar Orbiting Polar Orbiting	1 km 10 km 1 2 times daily 17.6 km 1 2 times daily (multi-angle)
VIIRS	Polar Orbiting	750 m 6 km 1 2 times daily
INSAT 3DR	Geostationary	~4 km 10 km Every 30 minutes

B. Retrieval Algorithms and Surface Reflectance Challenges

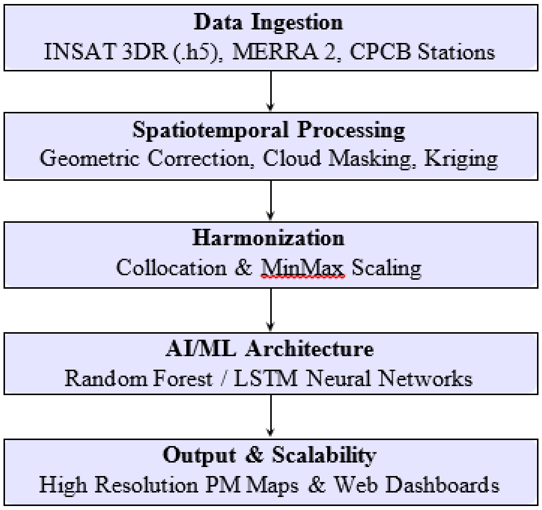
Retrieving AOD involves separating the atmospheric aerosol signal from the surface reflectance. MODIS uses the Dark Target (DT) algorithm for dense vegetation and the Deep Blue (DB) algorithm for bright surfaces like deserts. [2]. The DT algorithm presupposes the presence of low-predictable surface reflectance in the vegetation areas where the aerosol signal can be isolated more reliably. By contrast, the DB algorithm takes advantage of lower blue wavelengths in which even bright surfaces are comparatively darker, resulting in better aerosol retrieval in arid and urban regions. Conversely, INSAT 3D employs a "clear sky composite" method [15].

One notable issue of the application of INSAT AOD retrieval is that it is very responsive to surface albedo on reflective surfaces. Small errors in aerosol retrieval results will be disproportional because of the few spectral channels and simplified surface characterisation. Recent comparative analyses have shown that uncalibrated INSAT-3D AOD is not highly correlated with ground truth AERONET sites at high levels of retrieval (R 2 = 0.01), which indicates a significant amount of noise and retrieval bias. However, spurious variations were removed, and data reliability was greatly improved by applying a strong Error Envelope (EE) method, which is used to filter a retrieval outside physically plausible uncertainty ranges. This optimisation pushed the correlation coefficient to 0.66, and INSAT becomes competitive with the notations of MODIS, especially in the application of high-temporal resolution of aerosols monitoring across the Indian subcontinent. [8] [20].

3. Data Harmonisation and The AOD to PM Conversion

One of the most important challenges that accompany the satellite-based assessment of air quality is an essential nonlinear procedure that involves converting column AOD into PM mass, which is explained by the fact that satellites can only measure the vertically integrated aerosol load and do it in a manner that fails to provide the aerosol information on the concentration of the PM on the surface of interest. This leads to the absence of uncertainty in the vertical axis of data that would establish a direct relationship between AOD and PM. Also, the relationship between AOD and PM depends on meteorological factors, including PBLH and RH, since alterations of PBLH promote aerosols in the air to be dispersed vertically, leading to changes in the PM concentration on the side but not a proportional change in columnar AOD. Also, RH has the capability of affecting the adsorption of water into and onto aerosols, resulting in an increase in the quantity and scattering of the aerosols and their corresponding scattering hy, and consequently, inaccurately portraying the AOD as indicative of the dry mass. It implies that the boundary dynamics and humidity-related corrections should be added to the estimation of surface PM used based on AOD. [2].

Fig. 1. Proposed methodology pipeline for automated multi-source data fusion and air quality forecasting.



A. Preprocessing Pipeline and Meteorological Integration

To mitigate these physical confounders, robust models integrate continuous gridded meteorological fields from atmospheric reanalysis products, such as NASA's MERRA 2 [21]. Incorporating RH, temperature inversions, and wind speed allows machine learning frameworks to dynamically adjust the AOD PM relationship [22].

As illustrated in Fig. 1, automating the ingestion of massive datasets is a profound engineering hurdle. Raw geostationary data stored in HDF5 (.h5) formats must be acquired via bulk transfer (e.g., WinSCP) and subjected to stringent geospatial subsetting to isolate regional domains like Mumbai [23]. Due to scale discrepancies between satellite pixels (10 km) and reanalysis grids (0.5 degrees), variables are systematically interpolated using geostatistical methods such as ordinary kriging [8]. Finally, features undergo MinMax Scaling to compress inputs into a range, ensuring stable gradient convergence and reduced computational overhead during neural network training.

4. Predictive Modelling: From Statistical Baseline to Deep Learning

The inherent non linearity of atmospheric physics neces- sitates a paradigm shift from traditional statistical **models to** advanced Artificial Intelligence (AI) architectures [24]– [2 The inherent non linearity of atmospheric physics necessitates a paradigm shift from traditional statistical models to advanced Artificial Intelligence (AI) architectures [24]–[26].
6].

A. Baseline Regressions and Ensemble Machine Learning

Initial investigations utilised Multiple Linear Regression (MLR). As demonstrated in recent regional studies for Mumbai, MLR effectively quantifies baseline meteorological influences, proving that maximum daytime temperatures and wind speeds dictate primary pollution dispersion ($R^2 \approx 0.628$). However, linear models struggle to map complex multivariate interactions.

Consequently, ensemble decision tree models like Random Forest (RF) and Extreme Gradient Boosting (XGBoost) are highly preferred. RF constructs multitudes of independent decision pathways and averages their outputs, rendering it highly robust against noisy satellite retrievals and overfitting [2] [27]. A recent Indian evaluation by Verma *et al.* successfully synthesised MODIS and INSAT 3D AOD with MERRA 2 meteorology using RF, achieving strong predictive accuracy ($R^2 = 0.78$ for MODIS-derived PM_{10}) [8]. Table III highlights the efficacy of various ML paradigms reported in contemporary literature.

Table 3: Comparative Performance of Selected AI/ML Models for PM Estimation [2]

Model Architecture	Key Predictors Used	Performance (R^2)
Mixed Effects Model	3 km AOD, Meteorology	0.81 0.83
Random Forest (RF)	MAIAC AOD, Meteorology	0.72 0.78
XGBoost	Meteorology, Pollutants	> 0.90
ConvLSTM (Deep Learning)	AOD, Met, Ground Obs.	0.91
Geograph. Weighted Reg.	500m AOD, Land Use	0.86

B. Spatiotemporal Deep Learning: LSTM Networks

Since air pollution is fundamentally a time-dependent sequence driven by the continuous accumulation of traffic emissions, seasonal shifts, and delayed meteorological impacts, the domain has aggressively pivoted toward Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) cells [28].

Unlike vanilla RNNs, which suffer from the vanishing gradient problem over long input sequences, LSTMs utilise a complex system of memory gating mechanisms. These gates enable the network to retain crucial historical accumulation data over several days while selectively discarding irrelevant noise [29]. Implementing LSTM architectures drastically improves short-term forecasting (24 to 48 hours), laying the computational groundwork required for proactive early warning mechanisms.

5. Critical Research Gaps and Operational Challenges

Despite profound advancements, several substantial barriers impede the seamless operationalisation of global air quality frameworks:

1) **Vertical Profile Discrepancies:** There is a gap in the ability of passive satellite imagery to differentiate the source of AOD, for example, AOD could come from surface pollution or a high source aerosol [30]. There exists a data fusion problem that involves combining vertically

sparse profiles from active LiDAR, like CALIOP, which is 3D, with 2D satellite images.

2) **Spatiotemporal Non-Stationarity:** There is high variability in the chemical structure of the aerosol. A model designed for industrial emission in Northern India will most likely give poor results in places like Mumbai, a coastal city that has a high presence of marine aerosols and humidity. For accurate regional downscaling [8], the creation of hybrid models that understand space as a variable is of utmost importance.
3) **Data Scarcity and Gap Filling:** The presence of cloud cover, as well as specific surface types that highly reflect the sun, results in satellite data that is missing and is a well-recognised phenomenon. In order to produce PM concentration maps that are continuous and do not have any missing data, it is imperative to utilise advanced AI-based methods, which fill in data gaps in a spatiotemporal manner, i.e. [28].
4) **Real Time Processing Latency:** There is a high amount of engineering that is required to transition from the analysis of data that is retrospective to Near Real Time (NRT) [2]. Web-based systems that are designed to warn users of potential dangers must have the ability to automatically process data from the INSAT satellites at a high frequency, such as executables. [29]

6. CONCLUSION

The integration of satellite-based Aerosol Optical Depth, meteorological reanalysis, and ground-based data has changed how we spatially map particulate matter. INSAT-3DR and other geostationary satellites provide high-frequency data for the monitoring of quickly evolving pollution episodes in developing countries. The shift from classical statistical regressions to more refined, temporal deep learning models, e.g. LSTMs, allows researchers to capture the atmosphere's highly complex, nonlinear and time-dependent processes driving surface PM concentrations. Addressing the remaining technical issues, particularly vertical profiling of aerosols, spatially dependent non-stationarity, and latency, will result in the development of scalable, operational web-based warning systems. These systems will deliver timely information for strategically mitigating the negative health impacts of air pollution to decision makers and the general public.

REFERENCES

1. World Health Organisation. Health impacts of air pollution. Geneva: WHO; 2023. Available from: https://www.who.int
2. A Comprehensive Literature Survey on Monitoring Air Pollution from Space: Integrating Satellite Observations, Ground Measurements, Re-analysis Data, and Artificial Intelligence. Literature Survey Report.
3. U.S. Environmental Protection Agency. Health and environmental effects of particulate matter (PM) . Washington (DC): EPA; Available from: https://www.epa.gov
4. Ren M, et al. Concentration-response relationship between PM2.5 and daily respiratory deaths in China: A systematic review and meta-regression analysis of time-series studies. J Thorac Dis. 2017;9(12):4508–4518.
5. Central Pollution Control Board (CPCB). National Air Quality Index [Internet]. New Delhi: CPCB; Available from: https://cpcb.nic.in
6. Dey S, et al. A satellite-based high-resolution (1 km) ambient PM2.5 database for India (2000–2019). Remote Sens. 2020.
7. Sahu SK, et al. Estimating ground-level PM2.5 concentrations in India using satellite AOD and GAM. Environ Int. 2020.
8. Verma S, et al. Particulate matter estimation using satellite datasets: A machine learning approach. Environ Sci Pollut Res. 2024; 31:66372–66387.
9. Ma Z, et al. Satellite-based spatiotemporal trends in PM2.5 concentrations: China, 2004–2013. Environ Health Perspect. 2016;124(2):184–192.
10. Chelani AB, Devotta S. Estimating PM2.5 concentration from satellite-derived AOD in Indian cities. Atmos Pollut Res. 2019.
11. NASA Goddard Space Flight Centre. Aerosol Optical Depth . NASA; Available from: https://www.nasa.gov
12. Xie Y, et al. Daily estimation of ground-level PM2.5 concentrations over Beijing using 3 km resolution MODIS AOD. Environ Sci Technol. 2015;49(20):12280–12288.
13. Maheshwarkar P, et al. Population exposure across central India to PM2.5 derived from MAIAC AOD. Sci Rep. 2021.
14. Ashok S, et al. Estimation of surface PM2.5 with MODIS aerosol optical depth and source identification: A case of Hyderabad City, India. Atmos Environ. 2022.
15. Mishra MK, Rastogi G. Retrieval of aerosol optical depth from INSAT-3D imager. J Geophys Res Atmos. 2018.
16. Mishra MK, Rastogi G. Operational retrieval of aerosol optical depth over the Indian subcontinent using INSAT-3D imager. ISRO/SAC/VEDAS Technical Note. 2014.
17. MOSDAC/ISRO. Aerosol optical depth retrieval from INSAT-3DR imager: ATBD. Space Applications Centre; 2019.
18. Di A, et al. Dust aerosol optical depth retrieval and dust storm detection using INSAT-3D. Remote Sens. 2016.
19. Kumar S, et al. Comparison of INSAT-3D AOD over the Indian region with satellite and ground-based measurements. Proc SPIE. 2016.
20. Kumar A, et al. Prediction of PM2.5 concentrations over an urban region using different satellite data. J Indian Soc Remote Sens. 2024.

21. Bosilovich MG. MERRA-2: Initial evaluation of the climate. NASA Technical Memorandum TM-2015-104606; 2015.
22. Sivakumar S, Venkatesan R. Air quality index prediction using a machine learning model for IoT application. Arab J Geosci. 2022.
23. Chheda D. Analysis of spatiotemporal pollution data for Mumbai region. Technical Report. 2024.
24. Li T, et al. Estimating ground-level PM2.5 by fusing satellite and station observations. Geophys Res Lett. 2017.
25. Di Q, et al. Assessing PM2.5 exposures with high spatiotemporal resolution: A neural network–based hybrid model. 2016.
26. Ameer S, et al. Comparative analysis of machine learning techniques for predicting air quality. IEEE Access. 2019.
27. Chen G, et al. A machine learning method for estimating PM2.5 concentrations in China. Environ Sci Technol. 2018.
28. Zheng T, et al. Estimating ground-level PM2.5 using micro-satellite imagery with machine learning. Sci Total Environ. 2020.
29. Wang F, et al. Estimating high-resolution PM2.5 concentrations by fusing satellite AOD and smartphone photographs. Remote Sens. 2022.
30. Hu X, et al. Estimating ground-level PM2.5 concentrations in the southeastern US using geographically weighted regression. Environ Res. 2013.

About the corresponding author

Iqra Miyaji is a student at Usha Mittal Institute of Technology, S.N.D.T. Women's University, Mumbai, India. She is engaged in technical and interdisciplinary research, with interests in emerging technologies, data analysis, and environmental applications. She is committed to academic excellence and contributing to innovative solutions for real-world challenges through research and learning.