# Assimilation of high-resolution Ocean Color Monitor (OCM) aerosol optical depth in WRF-Chem improves $PM_{2.5}$ forecasts over the Indian region

**Prafull P. Yadav**

prafullyadav1996@gmail.com

Indian Institute of Tropical Meteorology, Ministry of Earth Sciences, Pune, India

**Sachin D. Ghude**

Indian Institute of Tropical Meteorology, Ministry of Earth Sciences, Pune, India

**Gaurav Govardhan**

Indian Institute of Tropical Meteorology, Ministry of Earth Sciences, Pune, India

**Rajesh Kumar**

National Science Foundation National Center for Atmospheric Research

**Rajmal Jat**

Indian Institute of Tropical Meteorology, Ministry of Earth Sciences, Pune, India

**Shivani Shah**

Space Applications Centre

**B. P. Shukla**

Space Applications Centre

**Manoj K. Mishra**

Space Applications Centre

**Deepak Putrevu**

Space Applications Centre

**Imran Girach**

Space Applications Centre

**P. K. Thapliyal**

Space Applications Centre

**Rashmi Sharma**

Space Applications Centre

---

**Article**

**Keywords:** Ocean Color Monitor (OCM), WRF-Chem, satellite data assimilation

**Posted Date:** April 28th, 2025

**DOI:** https://doi.org/10.21203/rs.3.rs-6425765/v1



**Additional Declarations:** No competing interests reported.

# Abstract

This study investigates the impact of assimilating high resolution (770 m) Aerosol Optical Depth (AOD) retrieval derived from the Ocean Color Monitor (OCM) sensor into the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem) for the first time, aiming to improve fine particulate matter ($PM_{2.5}$) forecasts over India. AOD assimilation leads to substantial improvements in model accuracy, reducing $PM_{2.5}$ biases by 30–70% and lowering root mean square error (RMSE) across critical regions such as Delhi, Punjab, Bihar, and West Bengal. The assimilation substantially improves initial conditions of surface $PM_{2.5}$ estimates by approximately 60 μg/m³. Forecast accuracy is the highest on the first day, with an RMSE of 21.35 μg/m³ and a correlation coefficient (R) of 0.75, followed by increasing RMSE values of 30.40 μg/m³ on Day 2 and 32 μg/m³ on Day 3, with correlations of 0.73 and 0.70, respectively, reflecting degradation of assimilation benefits by model uncertainties over time. With MODIS nearing phase-out, high-resolution OCM retrieval provides a reliable alternate choice for future AOD assimilation in the AIRWISE forecasting system over India.


# 1. Introduction

India's rapid urbanization, industrialization, and energy demand have led to rising anthropogenic emissions, worsening air quality nationwide[1,2]. Studies report widespread $PM_{2.5}$ pollution driven by vehicular, industrial, and agricultural emissions, coupled with unfavourable meteorology[3–5]. This poses a significant environmental and public health threat. Accurate short-term air quality forecasts can help reduce health risks by enabling timely interventions[6,7]. A reliable air quality forecasting is essential for Indian cities due to the severe and persistent air pollution challenges driven by rapid urbanization, industrialization, and vehicular emissions. With frequent episodes of high pollution levels, accurate forecasting helps provide timely Air Quality Index (AQI) information to the public, enabling precautionary measures to reduce health risks, especially for vulnerable populations[8].

The Air Quality Early Warning System (AQEWS), developed under the Ministry of Earth Sciences (MoES) for the National Capital Region (NCR), has emerged as a critical tool for addressing air pollution challenges[9–11]. Over time, AQEWS has evolved into a comprehensive framework through the integration of a Decision Support System (DSS) [12]. This integration has resulted in the Air Quality Warning and Integrated Decision Support System for Emissions (AIRWISE)[13], a cutting-edge tool that combines high-resolution numerical modeling based air quality forecasts with actionable source attribution insights for air quality management. This system helps policymakers to implement the Graded Response Action Plan (GRAP) in Delhi-NCR, which imposes temporary predefined restrictions on pollution sources based on forecast data. The integrated system helps in managing air quality during critical periods like Diwali (a major Indian festival celebrated with fireworks and lights) and other high pollution events, providing crucial support for public health and environmental management. Utilizing advanced chemical-transport models such as the WRF-Chem and employing a three-dimensional variational (3DVAR) data assimilation framework, AIRWISE integrates diverse observational datasets. These include ground-based measurements and satellite-derived AOD retrievals, particularly from (Moderate Resolution Imaging Spectroradiometer) MODIS. The assimilation of MODIS AOD data has significantly enhanced AIRWISE's ability to predict $PM_{2.5}$ concentrations during critical pollution episodes, enabling policymakers to take pre-emptive actions[14] However, with the MODIS platform nearing the end of its operational lifespan[15] the need to explore alternative satellite systems has become increasingly urgent to ensure continuity and advancement in data assimilation efforts. In this context, the OCM onboard the Indian Space Research Organization's (ISRO) Oceansat-3 satellite EOS-06, launched in 2022, presents a promising alternative. Although primarily designed for oceanographic applications, OCM extends its capabilities to aerosol monitoring over land, offering enhanced spatial and temporal resolution for air quality modeling. The transition from MODIS to OCM provides an opportunity to maintain and improve the accuracy of AIRWISE by ensuring the consistency of AOD retrievals in the data assimilation pipeline. This integration of OCM AOD retrievals into AIRWISE forms the core of the present study. By leveraging the advanced features of OCM, the research aims to refine $PM_{2.5}$ predictions during peak pollution periods and support evidence-based decision-making to mitigate air quality crises. The findings are expected to contribute to the ongoing evolution of air quality forecasting systems, emphasizing the critical role of satellite-based data in tackling complex environmental challenges in India and beyond.

In this paper, we highlight the impact of OCM data assimilation on air quality forecasting and assess how it performs during peak pollution periods compared to MODIS. Our analysis shows that the OCM-assimilated forecast falls within the expected uncertainties and is reliable for issuing timely warnings, demonstrating its suitability alongside MODIS for operational air quality forecasting.

## 2. Data and methods

# 2.1 Satellite Retrievals for Data Assimilation

This research utilizes the WRF-Chem model configuration, emissions, parameterization schemes as employed in the AIRWISE framework [9,13], focusing with 10 km grid spacing over the entire Indian region as the domain of analysis. The meteorological initial and boundary conditions were derived from the analysis and forecast product of the Indian Institute of Tropical Meteorology-Global Forecasting System (IITM-GFS, T1534). This system employs Ensemble Kalman Filtering at a horizontal resolution of 12.5 km, providing data every three hours [16].

We have performed three experiments. The first was a control run without data assimilation, referred to as the CNTL experiment; the second included assimilation using MODIS AOD data, defined as the MODISDA experiment; and the third incorporated assimilation using OCM AOD data, termed the OCMDA experiment. All simulations were performed with daily forecast outputs generated for 72 hours. The meteorological parameters are reinitialized using global weather forecasts from IITM-GFS and chemistry was recycled from the previous day's forecast to generate background initial conditions at the start of each forecast cycle. The analysis period spanned from 01 November 2023 to 14 November 2023. To ensure consistency and comparability, all parameterization schemes and model configurations were kept identical across the three experiments.

To assimilate either MODIS or OCM AOD in WRF-Chem, the 3DVAR data assimilation framework within the Gridpoint Statistical Interpolation (GSI) system is utilized. AOD assimilation leverages the observational data to optimize model initial conditions of aerosol species by minimizing a cost function J, expressed as:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(H(x) - y)^T R^{-1}(H(x) - y)$$

1

Here, $x$ is the state vector representing model variables (aerosol species and meteorological parameters required for AOD calculation), $x_b$ is the background model state, B is the background error covariance matrix, H is the forward operator transforming model variables (aerosols) to observational (AOD) space, $y$ represents the satellite AOD observations, and $R$ is the observation error covariance matrix. The forward operator employs the Community Radiative Transfer Model (CRTM) to compute AOD from modeled aerosol concentrations. The assimilation process iteratively minimizes the cost function by balancing the model background with observations, resulting in updated aerosol mass concentrations and improved analysis accuracy for initialization of air quality predictions. The background error covariance matrix is modelled using the National Meteorological Center (NMC) method through the Generalized Background Error (GEN_BE) module. This approach uses differences between two forecasts valid at the same time to estimate parameters such as variance, horizontal length scales, and vertical length scales. These parameters are derived from daily pairs of 24-hour WRF-Chem forecasts initialized under varying meteorological, anthropogenic, and biomass burning conditions, We consider a 100% uncertainty in both anthropogenic and biomass burning emissions, based on intercomparison studies of various emission inventories [18,19]. The variances determine the weight of observational innovations, while the length scales govern the spatial and vertical influence of assimilation increments [10].

The MODIS instrument, onboard NASA's Terra and Aqua satellites launched in 1999 and 2002, respectively, operates in a sun-synchronous orbit with equatorial overpass times at approximately 10:30 AM and 1:30 PM local time. It provides near-global coverage every 1–2 days with a wide swath width of approximately 2330 km. It offers AOD retrievals at spatial resolutions of 10 km, derived using the Dark Target (DT) and Deep Blue (DB) algorithms. MODIS retrievals have observation errors estimated as ±(0.05 + 0.15 × AOD) over land and ±(0.03 + 0.05 × AOD) over ocean.

In line with MODIS, the OCM, onboard ISRO's Oceansat-3 satellite launched in 2022, also operates in a sun-synchronous orbit with an equatorial overpass time near 12:00 PM local time. OCM has a narrower swath width of approximately 1420 km, providing high-resolution AOD retrievals at a spatial resolution of approximately 770 m (0.007°), making it particularly suited for aerosol monitoring over South Asia and adjoining regions. OCM retrievals, produced using the Space Applications Center AErosol Retrieval (SAER) algorithm, have a theoretical uncertainty of ±(0.06 + 0.26 × AOD) over land [19].

MODIS near real-time retrievals have a latency of approximately 3 hours, while OCM retrievals are available with a shorter latency of 2 hours, making both datasets accessible by around 16:30 IST every day. In the operational forecasting setup, downloading and processing these near real-time AOD retrievals takes about 15 minutes. Each day, the chemical fields are initialized from the previous day's WRF-Chem forecast, aerosol fields are updated through data assimilation, and meteorological fields are refreshed using the IITM-GFS forecast. This workflow ensures that applying assimilation at 09 UTC is both practical and efficient.

$PM_{2.5}$ measurements from 258 monitoring sites, located in urban centers across India, are collected by the Central Pollution Control Board (CPCB), IITM, and the US Embassies' AirNow program. These sites are distributed across 12 states: 40 sites in Delhi, 17 in Gujarat, 30 in Haryana, 2 in Jharkhand, 1 in West Bengal (Kolkata US Embassies site), 29 in Madhya Pradesh, 11 in Pune (Maharashtra, IITM sites), 19 in Odisha, 8 in Punjab, 44 in Rajasthan, and 57 in Uttar Pradesh. These locations were judiciously chosen as MODIS and OCM satellites have daily swath coverage over these regions during the study period. The geographical locations of these 12 states are shown in Fig. S1 and were used to evaluate the performance of the assimilation experiments. To ensure the reliability of $PM_{2.5}$ observations for evaluating model performance, additional quality control steps were applied alongside the standard procedures implemented by the CPCB (https://cpcb.nic.in/quality-assurance-quality-control/). First, measurements below 10 μg $m^{-3}$ and above 1,500 μg $m^{-3}$ were excluded as such extreme values were likely due to instrument malfunctions. For example, $PM_{2.5}$ concentrations below 10 μg $m^{-3}$ were often recorded immediately after instrument restarts, which is improbable in regions with significant anthropogenic emissions. Second, sporadically high $PM_{2.5}$ values, which appeared intermittently in the time series at some monitoring sites, were also filtered out.

A variety of statistical metrics were employed to evaluate the performance of the model experiments, including Pearson's correlation coefficient (R), mean bias (MB), and root mean square error (RMSE). Pearson's correlation coefficient was interpreted using its statistical significance to ensure robust insights into the relationships between observed and forecasted values. These metrics comprehensively assess the accuracy and reliability of day-1 to day-3 $PM_{2.5}$ forecasts, offering valuable insights into the model's performance.

## 3. Results

# 3.1 Effect of Aerosol Optical Depth Assimilation on $PM_{2.5}$ Forecast

To evaluate the effectiveness of OCMDA in improving WRF-Chem simulated AOD and to understand how chemical data assimilation enhances aerosol initial conditions, we have compared the CNTL and OCMDA experiments against collocated OCM AOD retrievals on 10 November 2023 at 09 UTC (Fig. 1). OCM provides higher spatial resolution data, which can be critical for detailed regional studies that require finer spatial granularity. The CNTL experiment (Fig. 1 (b)) underestimates AOD substantially when compared with OCM observations (Fig. 1 (a)), with simulated AOD values ranging between 0.1–0.4, while observed AOD values range from 0.8–1.0. After assimilating OCM AOD, the OCMDA experiment exhibits significant improvements. The enhancements observed in OCMDA demonstrate that OCM-based assimilation effectively corrects model-simulated AOD, bringing it in line with satellite-retrieved values. The most notable improvements occur over the Central and lower Indo-Gangetic Plain (IGP), regions heavily impacted by crop residue burning in early November (Fig S1). The assimilation increases AOD by approximately 1.5 to 2.0 times compared to CNTL, highlighting the effectiveness of data assimilation in adjusting aerosol mass concentrations and refining the model's chemical initial conditions. To further analyze the impact of OCMDA, frequency distribution plots of AOD at 550 nm using all data from the study period illustrate how data assimilation improves model performance (Fig. 2 (b)). The OCMDA experiment captures higher AOD frequency distributions beyond 0.5, accurately representing more intense aerosol events that CNTL often misses. In contrast, CNTL overestimates aerosol concentrations in the 0–0.5 AOD range, leading to a higher frequency distribution compared to both observed and assimilated data. This discrepancy highlights the inability of the CNTL experiment to adequately capture high aerosol loading events, whereas OCMDA demonstrates enhanced sensitivity to aerosol variability. The one to one comparison of OCMDA with MODISDA, which has been evaluated previously in AIERWISE[13,20], shows that improvements by OCMDA is comparable with that of MODISDA (figures from Fig. 1 (f) and Fig. 1 (i)). For instance, enhancement over IGP by OCMDA falls in close range with that by MODISDA (0.8-1.0).

Additionally, we analyzed the impact of OCMDA on surface $PM_{2.5}$ concentrations by comparing $PM_{2.5}$ differences between the OCMDA and CNTL experiments, averaged over 1–14 November 2023 at the initialization of each forecast cycle (Fig. 3). On average, $PM_{2.5}$ concentrations increased by approximately 35–45 μg $m^{-3}$ over the IGP due to OCM data assimilation. These increments from OCMDA are generally smaller than those from MODISDA (Fig. 3a) over the IGP region, which is likely due to the higher observational error associated with OCM. Both are generated within the GSI assimilation framework, which adjusts the mass concentrations of all aerosol species in GOCART (Goddard Global Ozone Chemistry Aerosol Radiation and Transport,[21–23]) chemistry model based on assimilation-driven changes in WRF-Chem AOD. This process enhances the initial conditions for more accurate $PM_{2.5}$ predictions.

The forecasted 72 hour time series of $PM_{2.5}$ concentrations were extracted from the CNTL, MODISDA, and OCMDA runs at 258 observation sites across 12 states of India. The average hourly $PM_{2.5}$ values were then estimated from each of these runs and compared with the hourly mean $PM_{2.5}$ observations across the same sites for the first day of the forecast (Fig. 4). The CNTL simulation consistently underestimates $PM_{2.5}$ concentrations throughout the period, while both the OCMDA and MODISDA significantly improves agreement with observations. The improvements in OCMDA and MODISDA highlight the effectiveness of satellite AOD assimilation in enhancing model performance, particularly during the biomass burning period during the first ten days of November. A sharp decrease in $PM_{2.5}$ concentration was observed between 10 and 11 November, attributed to rainfall in parts of India. However, limitations in satellite AOD retrieval during cloud cover affected the dataset. Despite this, both assimilation runs successfully captured the overall trend of decreasing aerosol concentrations, as reflected in the observed data on 12 November. MODISDA shows slightly higher $PM_{2.5}$ concentrations compared to OCMDA with the average difference between the two assimilation runs being relatively small, around 8 µg $m^{-3}$. The impact of assimilation is also analysed for localized events during the simulation period. For example, a sharp spike in $PM_{2.5}$ concentrations was observed in the early morning of 13 November, following the Diwali celebrations on 12 November, where the observed average $PM_{2.5}$ levels reached over 400 µg $m^{-3}$. This increase may be attributed to the use of firecrackers during the Diwali festival across India, even after the ban on firecracker use is imposed in major Indian cities. The influence of firecracker usage is evident in the elevated $PM_{2.5}$ pollution observed in multiple regions. OCMDA simulations show the tendency to capture this event, although with an underestimation of approximately 200 µg $m^{-3}$, whereas the CNTL simulation completely missed this episode. Similar tendency is noticed in MODISDA time series for this event. This underperformance highlights the localized nature of the event, which presents challenges for capturing such spikes in regional-scale models. Nevertheless, the assimilation of MODIS and OCM AOD data significantly improves $PM_{2.5}$ predictions compared to the CNTL run.

The performance of all three experimental setups, OCMDA, MODISDA, and CNTL, is systematically assessed using key statistical metrics: RMSE (µg $m^{-3}$), MB (µg $m^{-3}$), and R (Table 1). This evaluation aims to quantify the impact of OCM-based data assimilation on the accuracy and reliability of air quality forecasts across multiple regions. Additionally, we have compared the OCMDA experiment with MODISDA to assess whether OCMDA is comparable in performance. Overall, the application of OCM data assimilation significantly improved the forecasts compared to the CNTL simulation, demonstrating notable enhancements in RMSE, MB, and R across the states. These improvements highlight the effectiveness of assimilating OCM AOD in refining model predictions and reducing biases in simulated $PM_{2.5}$ concentrations.

In Bihar, OCMDA significantly reduced both the mean bias (MB) and root mean square error (RMSE) compared to the control (CNTL) simulation, achieving improvements comparable to those observed with MODISDA. For instance, on Day 1, the RMSE in the CNTL run was substantially high, but OCMDA reduced it by nearly 20 µg $m^{-3}$, which is similar to the reduction achieved by MODISDA. Furthermore, while OCMDA demonstrated the lowest MB and RMSE among all configurations, its correlation coefficient (R) was also close to that of MODISDA. However, MODISDA outperformed all other configurations in terms of R, indicating slightly better spatial-temporal agreement with observations.

In Delhi, the CNTL simulation consistently underestimated $PM_{2.5}$ concentrations, particularly during periods of high pollution. Both OCMDA and MODISDA addressed this underestimation, with OCMDA showing a slightly better correlation (R = 0.5) on Day 1, indicating its effectiveness in aligning forecasts with observations.

For Gujarat, OCMDA demonstrated substantial improvements over CNTL, reducing RMSE values for Day 1 from 35.18 $\mu g\ m^{-3}$ in CNTL to 21.35 $\mu g\ m^{-3}$, outperforming MODISDA, which achieved 29.84 $\mu g\ m^{-3}$. Similarly, correlation (R) values improved from 0.51 in CNTL to 0.75 in OCMDA and 0.72 in MODISDA, reflecting a stronger agreement with observed $PM_{2.5}$ concentrations.

A similar trend was observed in Haryana, where both assimilation experiments significantly improved RMSE and MB compared to CNTL, although a slight underestimation persisted in both MODISDA and OCMDA.

For states like Maharashtra (Pune) and Madhya Pradesh, MODISDA slightly outperformed OCMDA in reducing RMSE on Day 1, although the difference between the two was minimal. In Punjab and Rajasthan, MODISDA exhibited better performance, particularly in capturing $PM_{2.5}$ variability during pollution episodes such as Diwali. However, OCMDA also showed significant improvements, effectively reducing errors and improving correlation coefficients.

In Rajasthan, Uttar Pradesh, and Punjab both assimilation experiments demonstrated the clear benefits of data assimilation in enhancing air quality forecasts. OCMDA and MODISDA consistently reduced RMSE and improved correlation (R) compared to CNTL, indicating a better alignment between forecasts and observed trends. In Rajasthan, MODISDA significantly outperformed CNTL by correcting its poor correlation and large RMSE values. Similarly, in Uttar Pradesh, both OCMDA and MODISDA exhibited substantial improvements, with OCMDA slightly surpassing MODISDA in reducing RMSE for certain days. Punjab followed a similar pattern, where both assimilation experiments corrected CNTL's large biases and weak correlations, particularly in the early forecast periods. In Jharkhand, MODISDA significantly degraded the performance in terms of both MB and RMSE, despite showing relatively strong short-term forecast skill. Similarly, OCMDA also led to a slight degradation in RMSE; however, it notably improved the MB, indicating better alignment with observed concentration levels. Moreover, OCMDA achieved a higher correlation coefficient (R = 0.5446) over extended forecast periods, highlighting its potential to enhance the temporal evolution of $PM_{2.5}$ predictions through improved assimilation of OCM aerosol observations.

Across all states, the Day 1 forecast consistently showed better performance compared to the second day (Day 2) and third day (Day 3) forecasts, as RMSE values increased and correlation coefficients slightly decreased with increasing lead time. For example, in Odisha, OCMDA achieved an RMSE of 34.02 $\mu g\ m^{-3}$ and R = 0.80 for Day 1, while RMSE increased to 36.32 $\mu g\ m^{-3}$ and R = 0.78 by Day 2. Similarly, in West Bengal (Kolkata), OCMDA reduced Day 1 RMSE to 64.08 $\mu g\ m^{-3}$ compared to CNTL's 87.22 $\mu g\ m^{-3}$, with R improving to 0.64 from 0.56. To support these statistical evaluations, time series of $PM_{2.5}$

concentrations for each state have been plotted and are provided in the supplementary figures (S3 to S13).

Overall, the comparison between OCMDA and MODISDA confirms that OCMDA is highly effective in improving $PM_{2.5}$ forecasts and, in some cases, performs comparably to or better than MODISDA, further validating the value of OCM-based data assimilation in air quality modeling.

Thus, the improvements in statistical parameters decrease with increasing forecast lead time. This indicates that model deficiencies, such as uncertainties in emissions, meteorology, or chemical processes, begin to offset the benefits of AOD assimilation over time. Unfortunately, data on the vertical distribution of aerosols and their chemical composition were not available for the study period, limiting further analysis of how assimilation impacted aerosol distribution. The uncertainty of the OCM sensor, as calculated by the SAER algorithm, is higher than that of MODIS, indicating relatively larger retrieval errors in OCM-derived AOD. Additionally, OCM as of now does not provide data over the ocean region. If such data were available, it would allow for the assimilation of dust plumes originating from the Middle East and transported over the sea, thereby improving the representation of their impact on regional air quality. Furthermore, since MODIS fire count data will not be available in the future, having fire count data from the Oceansat satellite would be a valuable asset for air quality forecasting, ensuring continuity in fire emissions monitoring and improving model accuracy. Additionally, while the spatial resolution of OCM data is generally sufficient for broad-scale applications, its finer resolution enhances detailed, city-centric forecasting. However, its effectiveness may be influenced by temporal frequency and retrieval accuracy, particularly under challenging atmospheric conditions such as high aerosol loading or cloud cover. To address these challenges and fully exploit the advantages of OCM data for urban areas, future research should explore the benefits of integrating OCM data into higher-resolution model configurations. This advancement could significantly refine our simulation capabilities, enabling more precise predictions of aerosol dynamics and interactions. By focusing on urban environments, where aerosol impacts are often most critical, the assimilation of OCM data can lead to improved forecast accuracy that is vital for effective air quality management in metropolitan areas.

# Conclusion

The OCM refers to the Ocean Colour Monitor sensor aboard India's Oceansat-3 satellite, providing AOD at a high-resolution of 1 km since January 2023. The present study, for the first time, evaluated the utilization of AOD retrievals of OCM in the chemical data assimilation framework of AIRWISE for improving $PM_{2.5}$ air quality forecasts across India. The study is conducted during 1$^{st}$ -15$^{th}$ November 2023, keeping the entire India and surrounding region as the study domain. All simulation experiments provide $PM_{2.5}$ forecast three days in advance. For evaluating the impact of OCM AOD assimilation, three sets of simulation experiments are conducted (i) Without AOD assimilation (CNTL), (2) With OCM AOD assimilation (OCMDA) and (3) With MODIS AOD assimilation (MODISDA).The research highlights significant improvements in forecast accuracy during the selected high-pollution period. Assimilating OCM AOD data significantly enhanced WRF-Chem-simulated AOD, bringing values closer to observations

(0.8–1.0) compared to the underestimated CNTL range (0.1–0.4). The improvements in OCMDA closely match those of MODISDA, demonstrating that OCM-based assimilation is as effective as MODIS in refining aerosol distributions.

The state-wise analysis of Indian states based on ground observations of $PM_{2.5}$ revealed significant reductions in RMSE and MB in $PM_{2.5}$ forecasts underscoring the significance of OCMDA. For example, in Bihar, the Day 1 RMSE dropped from 62.44 µg/m³ in CNTL to 39.36 µg/m³ with OCMDA, showcasing the effectiveness of OCM assimilation. Similarly, in Gujarat, the RMSE for Day 1 forecasts improved from 35.18 µg/m³ in CNTL to 21.36 µg/m³ with OCMDA. Additionally, correlation coefficients (R) increased from 0.51 in CNTL to 0.75 in OCMDA, demonstrating a stronger alignment with observed $PM_{2.5}$ concentrations.

Due to assimilation significant reduction in underestimation over Haryana and Punjab, along with a noticeable sign change in mean bias over Delhi, highlights the role of local emission uncertainties emphasizing the importance of region-specific emission characterization for improving forecast accuracy.

The statistical analysis shows that assimilating OCM AOD is equally good as that of MODIS AOD. These findings confirm that OCM-based AOD assimilation substantially improves air quality forecasts, performing on par with or even exceeding MODIS-based assimilation in certain regions. This reinforces the potential of OCM AOD as a reliable data source for enhancing regional air quality predictions through chemical data assimilation.

Both assimilation experiments successfully captured major pollution events, such as the Diwali-induced $PM_{2.5}$ spike on November 12, 2023, when observed concentrations exceeded 400 µg/m³. This finding demonstrates the ability of data assimilation to significantly enhance the model's capability to capture regional trends in $PM_{2.5}$ pollution, though localized, short-term events remain challenging.

A key observation is that the benefits of AOD assimilation diminish with increasing forecast lead times, highlighting the growing influence of model uncertainties over longer time horizons. Moreover, the MODISDA and OCMDA experiments effectively capture higher frequency distributions beyond 0.5 AOD, reflecting their enhanced capability to accurately track more intense aerosol events, which are often missed by the CNTL. To further enhance $PM_{2.5}$ predictions, leveraging finer resolution data from the OCM and assimilating it into a higher-resolution, city-specific model configuration could be particularly beneficial. Overall, this study underscores the critical role of data assimilation in improving air quality forecasts, with the OCM sensor emerging as a promising alternative to MODIS. The results validate the potential of OCM data to ensure the continuity and advancement of air quality forecasting systems in India, enabling more accurate and actionable predictions for mitigating pollution episodes.

# Declarations

## Acknowledgments

We gratefully acknowledge the high-performance computing support provided by the Pratyush system at the Indian Institute of Tropical Meteorology (IITM), Pune. We extend our sincere thanks to the Director of IITM for their invaluable guidance and suggestions throughout this work. We acknowledge the use of surface $PM_{2.5}$ data from air quality monitoring network of the CPCB, IITM and US Embassy network India. The authors gratefully acknowledge the Space Applications Centre (SAC), Indian Space Research Organization (ISRO), for providing the Ocean Color Monitor (OCM) sensor data used in this study.The Moderate Resolution Imaging Spectroradiometer (MODIS) data, freely available through NASA's Earth Observing System Data and Information System (EOSDIS), is also acknowledged.Special thanks are due to Rajesh Kumar for his significant contributions, which were supported by the NSF National Center for Atmospheric Research, a facility funded by the U.S. National Science Foundation (agreement no. 1852977), and NASA's Health and Air Quality program (grant no. 80NSSC22K1045).

### Author contributions

All authors contributed to the research; P.P.Y and S.D.G designed the research; P.P.Y wrote the paper; P.P.Y conducted the WRF-Chem simulations and analysis; S.D.G, G.G, R.K, R.J, S.S, B.P.S, M.K.M, D.P, P.K.T and R.S formulated the research. All Authors reviewed and edited Manuscript.

### Data Availability Statement

The $PM_{2.5}$ observational data used in this study were obtained from the Central Pollution Control Board (CPCB) and are accessible at https://app.cpcbccr.com/ccr. MODIS Aerosol Optical Depth (AOD) retrievals were downloaded from NASA Earthdata at https://earthdata.nasa.gov/. Ocean Color Monitor (OCM) data from the Oceansat-3 satellite were acquired from the Space Applications Centre (SAC), ISRO, and are available at https://www.mosdac.gov.in/.

### Competing interests:

The authors declare no competing interests

## References

1. Purohit, P., Amann, M., Kiesewetter, G., Rafaj, P., Chaturvedi, V., Dholakia, H. H., Koti, P. N., Klimont, Z., Borken-Kleefeld, J., Gomez-Sanabria, A., Schöpp, W., & Sander, R. (2019). Mitigation pathways towards national ambient air quality standards in India. *Environment International, 133*, 105147. https://doi.org/10.1016/j.envint.2019.105147
2. Ghude, S. D., et al. (2016). Premature mortality in India due to PM2.5 and ozone exposure. *Geophysical Research Letters, 43*, 4650–4658. https://doi.org/10.1002/2016GL068949
3. Guttikunda, S. K., & Gurjar, B. R. (2012). Role of meteorology in seasonality of air pollution in megacity Delhi, India. *Environmental Monitoring and Assessment, 184*(5), 3199–3211. https://doi.org/10.1007/s10661-011-2182-8

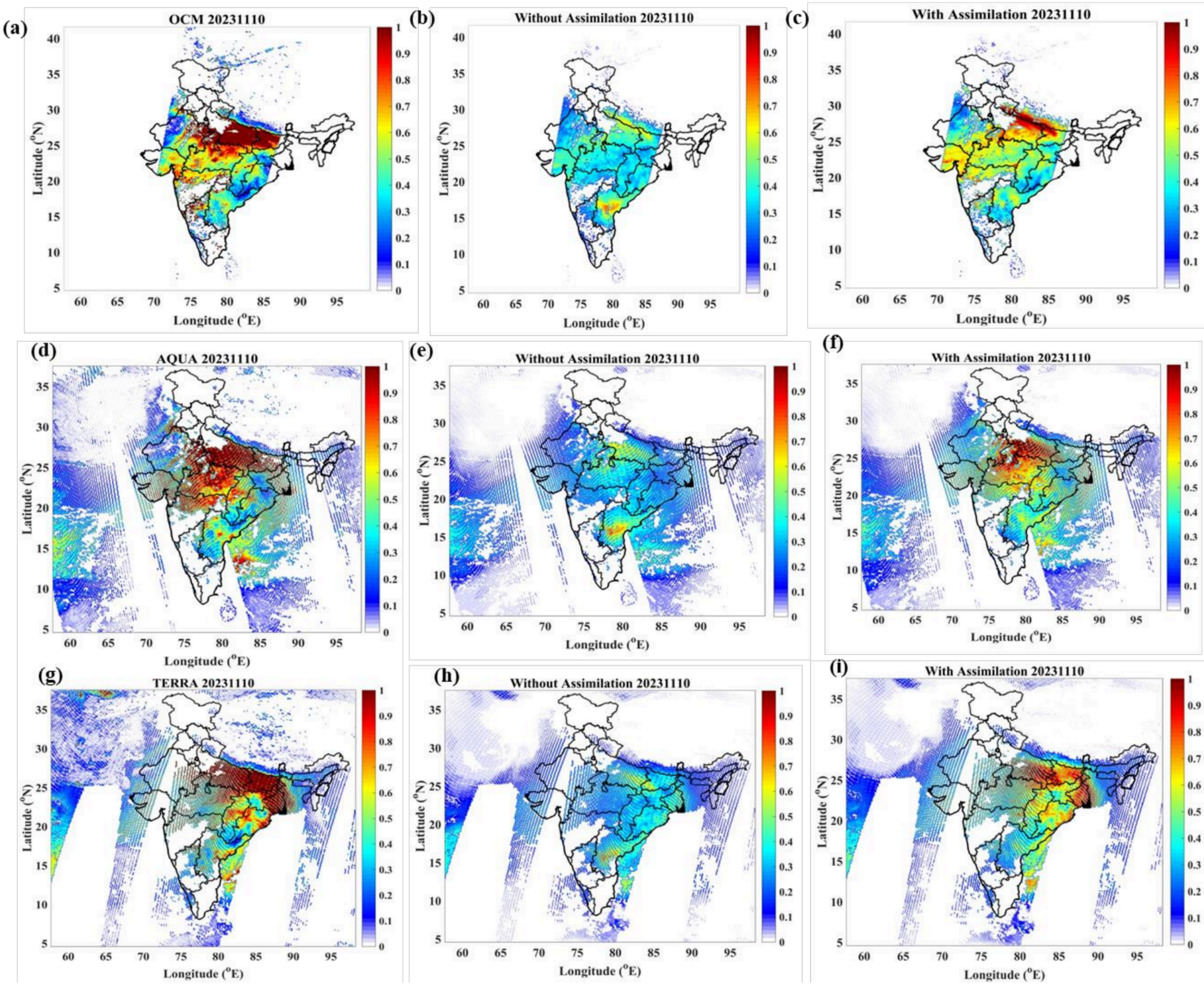
4. Kulkarni, S. H., Ghude, S. D., Jena, C., Karumuri, R. K., Sinha, B., Sinha, V., Kumar, R., Soni, V. K., & Khare, M. (2020). How much does large-scale crop residue burning affect the air quality in Delhi? *Environmental Science & Technology, 54*(8), 4790–4799. https://doi.org/10.1021/acs.est.0c00329
5. Sengupta, A., Govardhan, G., Debnath, S., Yadav, P., Kulkarni, S. H., Parde, A. N., Lonkar, P., Dhangar, N., Gunwani, P., Wagh, S., Nivdange, S., Jena, C., Kumar, R., & Ghude, S. D. (2022). Probing into the wintertime meteorology and particulate matter (PM2.5 and PM10) forecast over Delhi. *Atmospheric Pollution Research, 13*, 101426. https://doi.org/10.1016/j.apr.2022.101426
6. Ghude, S. D., Kumar, R., Govardhan, G., Jena, C., Nanjundiah, R. S., & Rajeevan, M. (2022). New Delhi: Air-quality warning system cuts peak pollution. *Nature, 602*, 211. https://doi.org/10.1038/d41586-022-00332-y
7. Jat, R., Gurjar, B. R., Ghude, S. D., & Yadav, P. P. (2024). Wintertime source apportionment of PM2.5 pollution in million plus population cities of India using WRF-Chem simulation. *Modeling Earth Systems and Environment, 10*(5), 6065–6082.
8. Kaginalkar, A., Ghude, S. D., Mohanty, U. C., Mujumdar, P., Bhakare, S., Darbari, H., et al. (2022). Integrated Urban Environmental System of Systems for Weather Ready Cities in India. *Bulletin of the American Meteorological Society, 103*, E54–E76. https://doi.org/10.1175/BAMS-D-20-0279.1
9. Ghude, S. D., et al. (2020). Evaluation of PM2.5 forecast using chemical data assimilation in the WRF-Chem model: A novel initiative under the Ministry of Earth Sciences Air Quality Early Warning System for Delhi, India. *Current Science, 118*, 1803–1815.
10. Jena, C., Ghude, S. D., Kumar, R., Debnath, S., Govardhan, G., Soni, V. K., Kulkarni, S. H., Beig, G., Nanjundiah, R. S., & Rajeevan, M. (2021). Performance of high-resolution (400 m) PM2.5 forecast over Delhi. *Scientific Reports, 11*, 4104. https://www.nature.com/articles/s41598-021-83467-8
11. Kalita, G., Yadav, P. P., Jat, R., Govardhan, G., Ambulkar, R., Kumar, R., Gunwani, P., Debnath, S., Sharma, P., Kulkarni, S., & Kaginalkar, A. (2023). Forecasting of an unusual dust event over western India by the air quality early warning system. *Atmospheric Environment, 311*, 120013. https://doi.org/10.1016/j.atmosenv.2023.120013
12. Govardhan, G., Ghude, S. D., Kumar, R., Sharma, S., Gunwani, P., Jena, C., Yadav, P., Ingle, S., Debnath, S., Pawar, P., Acharja, P., Jat, R., Kalita, G., Ambulkar, R., Kulkarni, S., Kaginalkar, A., Soni, V. K., Nanjundiah, R. S., & Rajeevan, M. (2024). Decision Support System version 1.0 (DSS v1.0) for air quality management in Delhi, India. *Geoscientific Model Development, 17*, 2617–2640. https://doi.org/10.5194/gmd-17-2617-2024
13. Ghude, S. D., Govardhan, G., Kumar, R., Yadav, P. P., Jat, R., Debnath, S., & Rajeevan, M. (2024). Air Quality warning and Integrated decision support system for emissions (AIRWISE): Enhancing Air Quality management in megacities. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-23-0181.1
14. CAQM (Commission for Air Quality Management). (2022). *Policy to curb air pollution in the National Capital Region*. https://caqm.nic.in/WriteReadData/LINKS/0031dcb806e-8af7-4b38-a9bc-65b91f2704cd.pdf

15. Shah, D., Zhang, S., Sarkar, S., et al. (2024). Transitioning from MODIS to VIIRS Global Water Reservoir Product. *Scientific Data, 11*, 209. https://doi.org/10.1038/s41597-024-03028-2
16. Mukhopadhyay, P., Prasad, V. S., Krishna, R., Deshpande, M., Ganai, M., Tirkey, S., et al. (2019). Performance of a very high-resolution global forecast system model (GFS T1534) at 12.5 km over the Indian region during the 2016–2017 monsoon seasons. *Journal of Earth System Science, 128*(1), 1–18. https://doi.org/10.1007/s12040-019-1186-6
17. Granier, C., Bessagnet, B., Bond, T., D'Angiola, A., Denier van der Gon, H., Frost, G. J., et al. (2011). Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. *Climatic Change, 109*(1–2), 163–190. https://doi.org/10.1007/s10584-011-0154-1
18. Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., & Soja, A. J. (2011). The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development, 4*(3), 625–641. https://doi.org/10.5194/gmd-4-625-2011
19. Mishra, M. K., Misra, A., & Kumar, R. (2023). Operational AOD retrieval at subkilometer resolution using OceanSat-2 OCM over land: SAER algorithm, uncertainties, validation & inter-sensor comparison. *Earth and Space Science, 10*, e2023EA002896. https://doi.org/10.1029/2023EA002896
20. Kumar, R., Ghude, S. D., Biswas, M., Jena, C., Alessandrini, S., Debnath, S., et al. (2020). Enhancing accuracy of air quality and temperature forecasts during paddy crop residue burning season in Delhi via chemical data assimilation. *Journal of Geophysical Research: Atmospheres, 125*, e2020JD033019. https://doi.org/10.1029/2020JD033019
21. Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., et al. (2002). Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements. *Journal of the Atmospheric Sciences, 59*(3), 461–483. https://doi.org/10.1175/1520-0469(2002)059<0461:TAOTFT>2.0.CO;2
22. Chin, M., Rood, R. B., Lin, S.-J., Müller, J.-F., & Thompson, A. M. (2000). Atmospheric sulfur cycle simulated in the global model GOCART: Model description and global properties. *Journal of Geophysical Research: Atmospheres, 105*(D20), 24,671–24,687. https://doi.org/10.1029/2000JD900384
23. Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., & Lin, S.-J. (2001). Sources and distributions of dust aerosols simulated with the GOCART model. *Journal of Geophysical Research, 106*(D17), 20,255–20,273. https://doi.org/10.1029/2000JD000053

# Table 1

Table 1 is available in the Supplementary Files section.

# Figures




## Figure 1

Spatial distribution of AOD on 10 November 2023 at 0900 UTC. The first column represents the observed AOD from OCM, AQUA and TERRA swaths (a, d, g). The second column shows the model AOD without data assimilation (CNTL experiment) (b, e, h). The third column displays the model AOD after data assimilation: OCMDA experiment (c) MODISDA experiment (f,i).

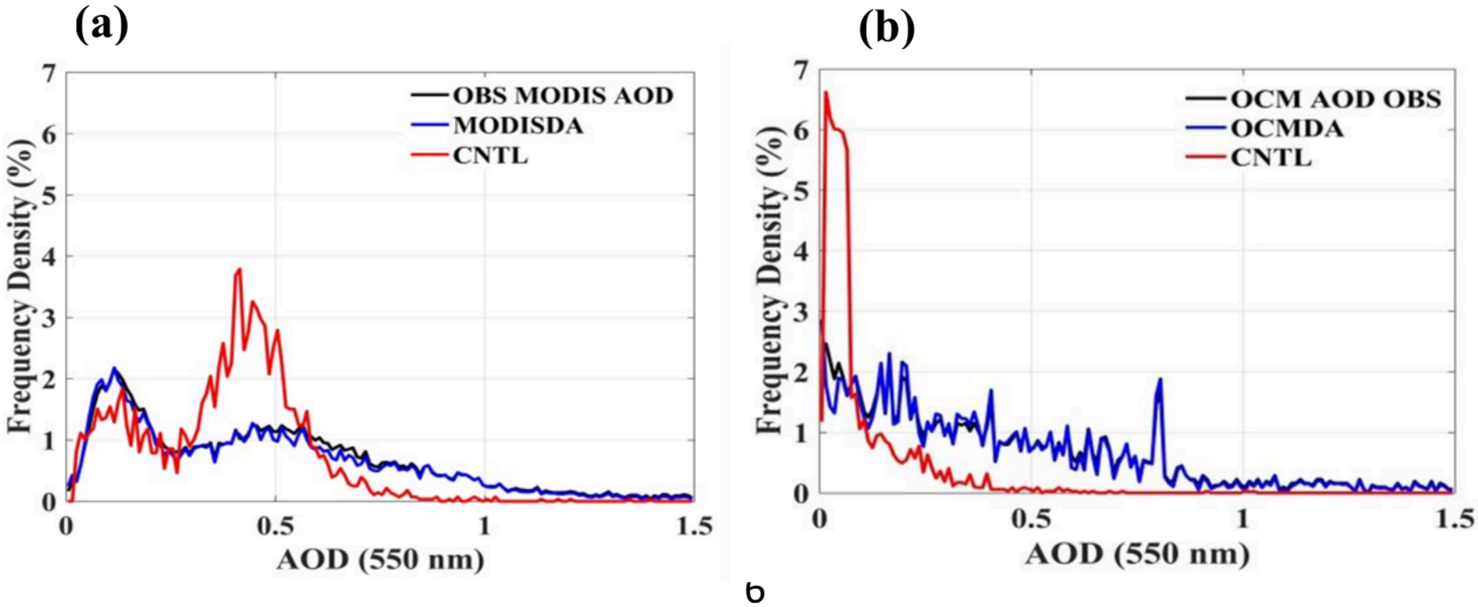



**Figure 2**

Frequency Density of Observed and Modeled AOD at 550 nm retrievals during 1 November to 14 November 2023 at 09 UTC: (a) Comparison of CNTL and MODISDA Outputs against collocated MODIS AOD Observations. (b) Comparison of CNTL and OCMDA Outputs Against OCM AOD Observations.

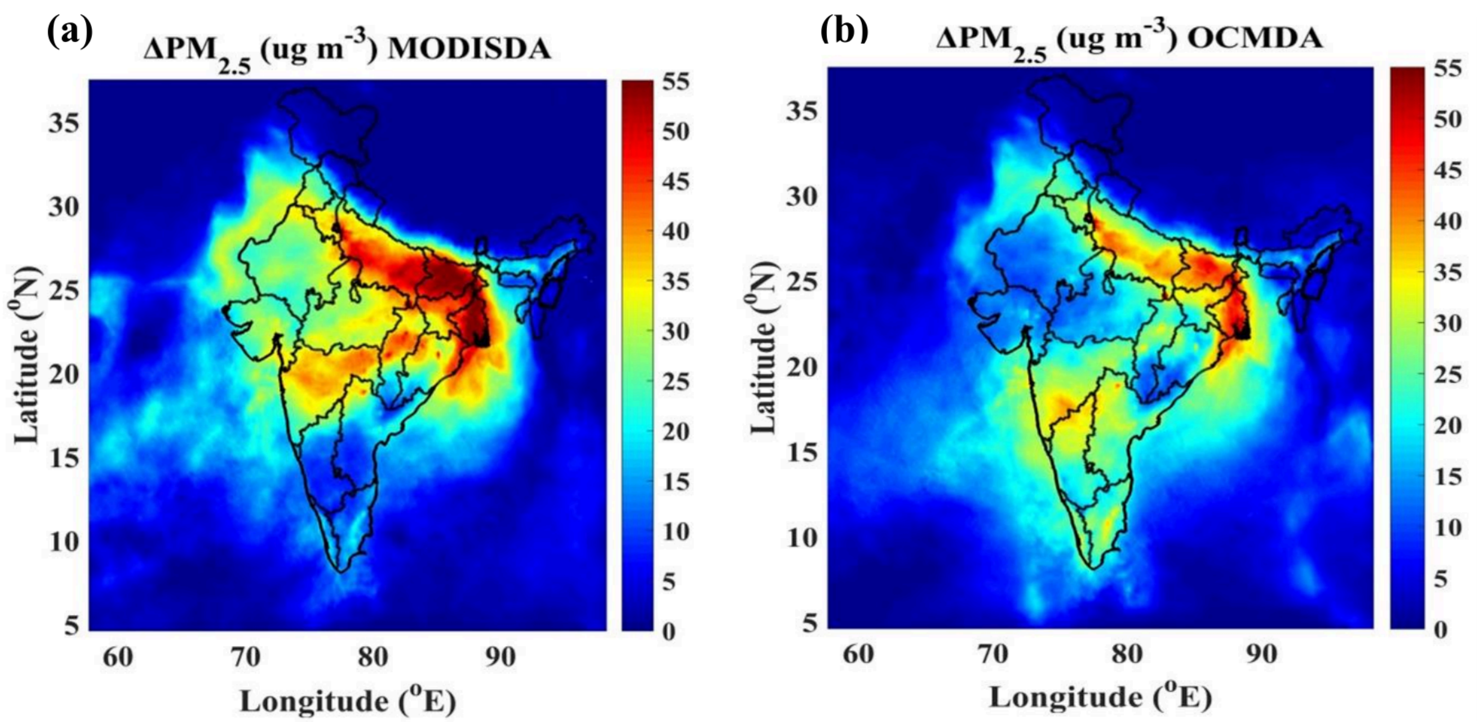



**Figure 3**

Spatial distribution of $PM_{2.5}$ analysis increment due to (a) MODISDA and (b) OCMDA.

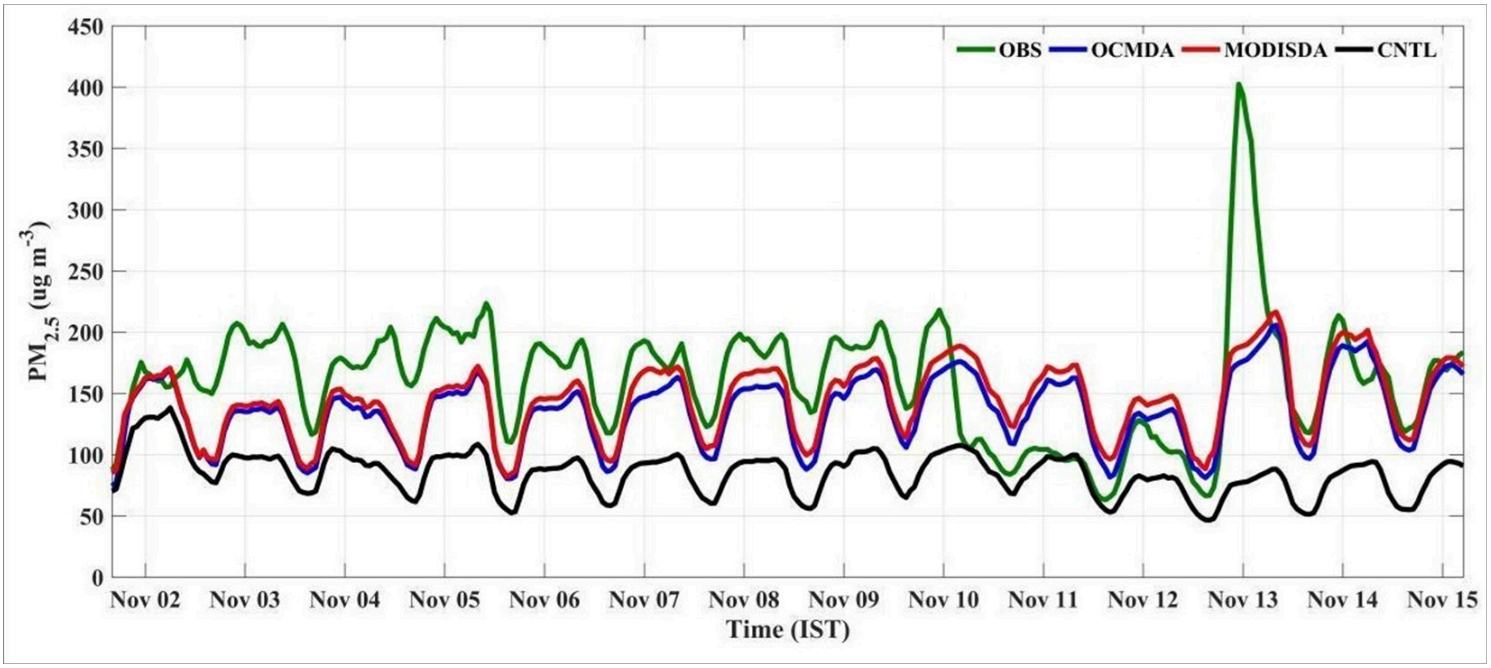



**Figure 4**

Hourly time series of $PM_{2.5}$ concentrations for the first-day forecast averaged over 258 observation sites across 12 states of India from 1 November to 15 November 2023, comparing observations (OBS) with CNTL, MODISDA, and OCMDA simulations

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx
- Table1.docx