Synergistic Fusion of Aerosol Optical Depth over India from Multi-Sensor Satellite Retrievals with Ground-based Measurements

Shiba Shankar Gouda[1,2], Mukunda M. Gogoi[1], S Suresh Babu[1]

[1]Space Physics Laboratory, Vikram Sarabhai Space Centre, ISRO, Thiruvananthapuram, 695022, India
[2]Research Centre, Department of Physics, University of Kerala, Thiruvananthapuram, 695034, India
Correspondence: Dr. Mukunda M Gogoi (dr_mukunda@vssc.gov.in)

Abstract

Synergistic fusion of aerosol parameters from multi-sensor measurements is crucial for integrating diverse data sources and generating consistent representations of aerosol distribution for accurate climate impact assessment. In this study, satellite observations from MODIS (Moderate Resolution Imaging Spectroradiometer) and MISR (Multi-angle Imaging SpectroRadiometer) are combined with ground-based measurements from Multi-Wavelength solar Radiometer (MWR) and CIMEL sun-photometers from the ARFINET and AERONET respectively to generate fused Aerosol Optical Depth (AOD) fields over India. The primary focus of this study is to develop a fusion framework over India, involving the evaluation and comparison of two approaches Universal Kriging (UK) and novel hybrid geostatistical-machine learning approach (RK-ML). Both methods share the same geostatistical foundation (variogram-based spatial-modelling) but differ in how the mean structure of AOD is estimated. In UK, satellite-derived AOD serves as deterministic trend for spatial prediction and is effective when ground based observations are well distributed, whereas RK-ML considers SVR-predicted AOD as prior and applies Ordinary Kriging to interpolate residuals from real-time ground observations, maintaining a near-zero residual mean away from observations which reduces distortion under sparse and uneven data conditions. Our results highlight seasonal fused AOD maps (winter, pre-monsoon, and post-monsoon) over India. Leave-One-Out Cross-Validation (LOOCV) is adopted as an evaluation strategy for assessing model performance, showing that the 95% confidence interval (±2σ) of the fused AOD captures over 80% of ground observations, indicating effectiveness in capturing regional aerosol variability. RK-ML demonstrates more stable spatial patterns and improved LOOCV performance compared to UK, particularly in regions with limited ground-based coverage.

Keywords: Aerosol Optical Depth, Universal Kriging, RK-ML, ARFINET, AERONET, MODIS, MISR

1. Introduction

Atmospheric aerosols play a significant role in introducing uncertainties into climate change projections. Although various factors such as microphysical parameters and chemical composition are important, aerosol optical depth (AOD), quantified by the total amount of columnar aerosol loading in the atmosphere, is the most critical parameter for understanding their climate forcing effects. With advances in technology and retrieval-algorithms, the number of satellites and ground-based observations of AOD is increasing. Although satellites are known to capture spatial heterogeneity of AOD, there could be bias or uncertainty (Huang et al., 2021) compared to ground-based measurements. Even if different satellites observe the same aerosol load over the same region nearly at the same time, the retrieved AOD differs due to the differences in algorithms, calibration, and resolution of the sensors (Kinne, 2009; Schutgens et al., 2020). The geographical complexity also challenges satellites to accurately retrieve AOD over highly heterogeneous land surfaces. On the other hand, data from ground-based sensors, though sparsely distributed, are more reliable than satellite measurements due to improved accuracy of measurement and retrieval procedure (Holben et al., 1998; Moorthy et al., 2007). Thus, the discrepancy between various satellite measurements and between satellite- and ground-based measurements of AOD is a serious concern in accurately characterizing aerosol loading over different parts of the globe (Wong et al., 2013; Sogacheva et al., 2020). In this context, there is a considerable effort in improving aerosol retrieval accuracy using approaches such as synergy processing of sun-photometer and lidar observations (Jin et al., 2025), synergistic retrieval from multi-mission space-borne measurements (Litvinov et al., 2025), gap-filling based on improved tensor-flow-based method (Bai et al., 2024), and the application of physics-informed deep-learning framework to multi-angle polarimetric measurements (Tao et al., 2023). Several studies have reported that if the correlations between the AOD from multiple ground-based and space-based sensors are sufficiently strong (Liu et al., 2004; Jiang et al., 2007; Prasad and Singh, 2007), then these observations can be used together for optimal characterization of aerosol features over a broader region. Thus, there is a growing demand for fused products to address limitations and achieve an optimal outcome, thereby strengthening reliability of aerosol database (Kahn et al., 2023).

Several approaches have previously been developed for multi-sensor data fusion involving satellite-to-satellite and satellite-to-ground observations. One notable method is the use of point spread function (PSF) modeling for single scanning footprints (Gupta et al., 2008). While PSF-based techniques are widely applied in image fusion, they face challenges in achieving accurate spatiotemporal colocation across different satellite platforms and don't show applicability regarding ground based AOD fusion rather solely on satellite footprint as a weighting factor for the merging of AOD from different sensors, such as MODIS (Moderate Resolution Imaging Spectroradiometer) and MISR (Multi-angle Imaging SpectroRadiometer) and Clouds and the Earth's Radiant Energy System (CERES). Statistical approaches such as Maximum Likelihood Estimation (Kim et al., 2024; Nirala, 2008) and Bayesian Maximum Entropy (Tang et al., 2016) have been applied to integrate satellite and ground-based observations. These methods explicitly account for uncertainty; but, their practical implementation is often limited by high computational demands, as they require large datasets for effective sampling and detailed pixel-level uncertainty characterization to produce reliable

fused products. Similarly, approaches such as the Ensemble Kalman Filter (Li et al., 2020) improve uncertainty quantification and have been applied at the global scale; however, their application is constrained by substantial computational cost and data requirements. These limitations pose challenges for near-real-time applications and for achieving high regional accuracy, particularly in regions with limited ground-based observational support. Simpler least-squares-based approaches, including adaptive weighted estimation (Guo et al., 2013) and semi-empirical optical algorithms (Xu et al., 2012), offer computational efficiency; however, their validation and broader applicability remain uncertain. More recently, machine learning techniques, particularly deep neural networks (DNN) (Kim et al., 2024), have demonstrated comparable performance, but their dependence on large training datasets and challenges in generalization limit their practical deployment.

In this study, we have adapted the Kriging technique to produce optimal fused AOD products over India. Among the various data fusion techniques, Kriging has gained significant attention for its applicability under geostatistical framework and has long been recognized as a robust and effective geostatistical technique for spatial estimation (Zimmerman et al., 1999; Shi et al., 2007; Prafull Singh & Verma, 2019; Stein & Corsten, 1991; Zhao et al., 2017). Notably, though geostatistical approaches provide a promising framework for data fusion, they are constrained by high computational demands, particularly when incorporating both spatio-temporal autocorrelation and covariance matrix inversion. Hence, reduced-rank methods such as Spatial statistical data fusion (SSDF) (Puttaswamy et al., 2014; Nguyen et al., 2012) alleviate computational burden but may introduce overfitting due to more number of parameters. Under such a scenario, Universal Kriging (UK) offers more stable AOD estimates near domain boundaries owing to its simpler and more robust formulation (Puttaswamy et al., 2014). Consequently, UK has been widely adopted for multi-sensor fusion integrating satellite and ground-based observations (Chatterjee et al., 2010; Jinnagara Puttaswamy et al., 2014; Lilla and Castrignanò, 2019), although it does not explicitly account for sensor-specific uncertainties. It has been extensively applied and validated across diverse domains within atmospheric research. It has also been utilized for spatial mapping of nutrients over oceans (Zhou et al., 2014) and as well as in mining, hydrology, electro-magnetic field mapping, and remote sensing image processing (Rossi et al., 1994). The Kriging outcomes are also found to be comparable with those from DNN (Chen et al., 2020; Kadow et al., 2020).

Previous research over the Indian region has estimated fused AOD from ground and satellite based observations using Cressman method, which employs inverse distance weighting (IDW), a widely used Geostatistical approach (Pathak et al., 2019). In this study, AOD measurements carried out from more than 40 ground-based observatories of the Aerosol Radiative Forcing over India Network (ARFINET; **Fig.-S1a**), which constitutes the national network of aerosol observatories across India and the largest such network in South Asia, are primarily used to integrate with the satellite-based observations from MODIS (Moderate Resolution Imaging Spectrometer) and the MISR (Multi-angle Imaging Spectro-Radiometer) to generate fused AOD using UK framework. Additionally, ground-based AOD data from the AErosol RObotic NETwork (AERONET; **Fig.-S1b**) are utilized to enhance the robustness of the database. While Kriging approaches have been previously applied to generate fused AOD

over northern India, the amount of ground data included in their studies was limited (Singh and Venkatachalam, 2014; Singh et al., 2016).

While satellite and ground-based AOD measurements generally exhibit a linear correlation, regional and environmental factors introduce biases, noise and nonlinear dependencies between explanatory and response variables. Although nonlinear extensions within the UK framework are possible, they require sophisticated techniques to achieve optimal performance, making the hybrid approach a compelling alternative. While the trend component in UK framework is conventionally modeled using low-order polynomials (e.g., first or second degree), studies exploring non-linear trend modeling are still relatively rare. For instance, Snepvangers et al. (2003) incorporated a logarithmic trend to improve prediction of soil water content using net precipitation as an auxiliary variable. Freier and Lieres (2015) proposed a Taylor-based linearization technique combined with iterative parameter estimation to capture non-linear trend functions in UK. Freier et al. (2017) further extended this approach to interpolate low-density, irregular biocatalytic data. These techniques are effective when the functional form of the non-linearity is known a priori. However, in most practical scenarios, such explicit formulations are unavailable due to complex, unknown interactions between design factors and responses. In this context, machine learning (ML) models, especially kernel-based methods such as Support Vector Regression (SVR), offer an effective alternative for capturing nonlinear and implicit relationships from data without the need of predefined functional forms. Considering the usefulness of prior spatial information on AOD across the domain, a hybrid Residual Kriging with Machine Learning (RK-ML) framework is adopted in this study, where SVR is used to generate an initial prediction of AOD, which serves as a prior estimate. The preference for SVR over decision-tree–based algorithms arises from its effectiveness for problems involving a small number of features and limited datasets, enabling more reliable fused estimates even when ground-based observations are sparse. While UK involves weighted regression with spatial covariance structures (for spatial predictions), RK-ML employs ML based regression and spatial covariance structure to produce more efficient and stable spatial patterns.

In this study, we primarily implement the UK framework over the Indian region to generate monthly fused AOD. Additionally, we evaluate and compare both the geostatistical–machine learning fusion frameworks (UK and RK-ML) while integrating satellite and ground-based AOD observations over India. We further assess the sensitivity of the fusion to the density of ground-based observations, demonstrating how sparse networks can introduce artifacts.

2. Data and Methodology

2.1 Ground-based AOD

The ground-based AOD is primarily obtained from ARFINET observations, having continuous measurements across the Indian region since 1985 maintained under ISRO-GBP (Gogoi et al., 2009; Babu et al., 2013). The spectral AOD measurements in the ARFINET observatories are carried out using a Multi-Wavelength solar Radiometer (MWR) and the handheld MICROTOPS-II Sun photometer. Both these instruments have been extensively inter-compared, and their consistencies have been established (Kompalli et al., 2010). The MWR is built on the principle of filter wheel radiometry. The measurements of direct solar flux using MWR are made at ten narrow wavelength bands centered at 380, 400, 450, 500, 600, 650, 750, 850, 935, and 1025 nm. The AOD is estimated following the Langley

Technique (Shaw, 1973; Moorthy et al., 2007; Moorthy et al., 2007) after subtracting the contribution due to molecular scattering and absorption due to O_3 and water vapour from total optical depth. For this, the MWR raw data (voltage readings corresponding to the time of acquisition) for the entire day are split into forenoon and afternoon. If the data span during each half of the day is more than 3 hours, the Langley plot is made separately for both forenoon and afternoon following cloud screening criteria. In order to estimate instantaneous AOD corresponding to each MWR measurement, the time-weighted Langley Intercept (LI) for the entire day is calculated from the forenoon and afternoon data as

$$LI_{daily} = (LI_{FN} * T_{FN} + LI_{AN} * T_{AN}) / (T_{FN} + T_{AN}) \quad (1)$$

Where, T_{FN} and T_{AN} are the durations of MWR measurements in the forenoon and afternoon. Based on this, the instantaneous AOD (after correcting the contributions due to Rayleigh scattering, Ozone, and water vapor) is estimated as:

$$AOD_{ins} = \{(LI_{daily} - \ln V)/ m\} - (\tau_R + \tau_{O3} + \tau_{WV}) \quad (2)$$

The accuracy of AOD estimates from MWR is based on the accuracy of the estimate of LI. Since, LI is also a parameter of indirect calibration of the instrument, the temporal variability of LI is examined to ensure performance of the system and qualify usable data. Typically LI varies within 5% of the mean and up to 10% in worst cases. Fluctuations are more pronounced at shorter wavelengths than at longer ones. Owing to these variations, total AOD uncertainty ranges from 0.02 to 0.03, increasing at shorter wavelengths (<500 nm) and during high AOD conditions (>0.5), which are mainly limited to the pre-monsoon season. Importantly, these errors are primarily statistical and uncorrelated across channels, rather than systematic (e.g., dark current, detector offsets, and molecular scattering/absorption modelling which are <0.1%). The instrument details, AOD retrieval method, and error budget have been discussed elsewhere Gogoi et al., 2009; Babu et al., 2013; Kompalli et al., 2010; Moorthy et al., 2007).

Apart from MWR, AOD is obtained from handheld MICROTOPS-II sun-photometer (Solar Light Company, USA) at five wavelengths (440, 500, 675, 870, and 936 nm). MICROTOPS-II can achieve AOD estimates with accuracy comparable to CIMEL Sun photometers used in the AERONET network, with uncertainties ranging from 0.01 to 0.02, as reported by Ichoku et al. (2002). In addition to ARFINET measurements, simultaneous AOD products (version 3, level 2.0) available within the study region from AERONET measurements are used. The CIMEL sun-photometers in AERONET measure AOD at 340, 380, 440, 500, 675, 870, and 1020 nm in a time interval of 5 to 15 min for cloud-free conditions with an uncertainty ~ 0.01 - 0.02 (Eck et al., 1999; Holben et al., 1998; Giles et al., 2019). To use the ARFINET and AERONET AOD in the fusion experiment, the AOD values are interpolated to 550 nm (corresponding to MODIS and MISR AOD) using the methodology of Liu et al. (2004):

$$\ln\left(\tau_{\lambda_1}/\tau_{\lambda_2}\right) = -\alpha \ln\left(\lambda_1/\lambda_2\right) \quad (3)$$

Where, τ_{λ_1} and τ_{λ_2} are AODs at wavelengths λ_1 and λ_2, respectively and α is Angstrom Exponent. α is determined by applying linear least squares fit to the logarithmic values of AOD measured at various wavelengths. For this study, values of α were estimated from the

wavelength dependent relationship of AOD at 450 and 650 nm for MWR, and 440 and 670 nm pair for CIMEL. Using this, AOD at 550 nm was estimated; where the base AOD was taken at 500 nm (in case missing AOD at 500 nm, base AOD at 440 or, 450 nm was considered).

2.2 Satellite retrieved AOD

The satellite-based AOD for this study is obtained from MODIS and MISR. MODIS data (Collection 6.1 Level-2 AOD at 550 nm; 'AOD_550_Dark_Target_Deep_Blue_Combined'; spatial resolution of 10 km; over land) is obtained from NASA's Level-1 and Atmosphere Archive and Distribution System Distributed Active Archive Center (LAADS DAAC). The merged AOD product combines only high-quality Dark Target (DT; QA = 3 over land, QA > 0 over ocean) and Deep Blue (DB; QA = 2 and 3) retrievals to provide global 10 km coverage. Over land, selection is based on Normalized Difference Vegetation Index (NDVI), with DB used for bright (arid, semi-arid) surfaces (NDVI $\leq$ 0.2), DT for vegetated (darker) regions (NDVI $\geq$ 0.3).In transitional zones, the higher-QA retrieval or their average is applied, while over ocean only DT is used. Although this approach improves spatial coverage and usability, uncertainties may arise in averaged regions and due to assumptions about algorithm performance across surface types (Sayer et al., 2014). Sensitivity studies across diverse land surfaces employing various algorithms have validated that integrating the DT and DB methods yields enhanced accuracies, but errors persistently emerge over South Asia (Gao et al., 2021; Tian et al., 2018; Wei et al., 2019). Furthermore, the performance of the product has been evaluated across different seasons (Sharma et al., 2021). Overall, an expected error of 0.05 ± 0.15 × AOD for DT and 0.05 ± 0.20 × AOD for DB over the land and 0.03 ± 0.05 × AOD over the ocean is reported in most studies (Levy et al., 2005; Sayer et al., 2013; Tian et al., 2018; Tian and Gao, 2019; Wei et al., 2019).

The MISR AOD (version V23) is obtained from Atmospheric Science Data Centre (ASDC). MISR V23 products provide aerosol information with a spatial resolution of 4.4 km × 4.4 km (Garay et al., 2017; Sayer et al., 2020; Witek et al., 2018, 2021). Theoretical sensitivity studies and performances for MISR (Kahn et al., 2001; Tao et al., 2020) have projected standard deviations of the measurement error associated with optical depth to be ± (0.05 + 20%$AOD_{AERONET}$), showing a consistently narrower range over ocean compared to bright land surfaces.

The MODIS and MISR datasets used in this study are both acquired from the Terra satellite platform and therefore have nearly identical overpass times. This temporal consistency ensures improved compatibility in the fusion process and minimizes uncertainties associated with diurnal variability in aerosol loading. In contrast, inclusion of MODIS observations from the Aqua satellite, which has a different overpass time, would introduce additional variability related to diurnal aerosol evolution that requires explicit treatment. Addressing such effects is beyond the scope of the present methodology-focused study and will be considered in future work. In the fusion approach, MODIS AOD represents high-quality retrievals (QA = 2, 3), while MISR exhibits minimal retrieval uncertainties (0.02–0.08) over ground stations. Additional screening or filtering was not applied beyond these criteria, as it may attenuate the inherent systematic bias between ground- and satellite-based observations. Quality-assured and expected-error-based filtering can be considered as part of the future scope of the study to enable more accurate inferences.

2.3 Fusion methodology

The geostatistical data fusion method used in this study combines spatial data from multiple sources (satellite and ground-based, as detailed in Sections 2.1 and 2.2) with varying resolutions, accuracies, and types of measurements. The aim is to enhance the overall understanding and prediction of spatial variables (e.g., AOD) to produce a more accurate and comprehensive representation of columnar AOD, with an emphasis on reducing inter-sensor biases through integration with ground-based observations. For this, we have adapted UK framework, where data interpolation relies on unknown functions (e.g., satellite derived AOD) represented as trend models with spatial autocorrelation through variogram analysis. Building on this framework, the fusion methodology is designed to operate under practical observational constraints, such as differences in sensor characteristics (e.g., spatial coverage, revisit frequency, and collocation with ground observations), which limit consistent data availability at daily timescales. Hence, the analysis is conducted at the monthly scale to improve spatial representativeness, reduce sampling gaps, and enhance statistical robustness. Notably, the monthly satellite AOD products also retain sensor-specific biases and inter-product inconsistencies. Thus, the fusion approach presented here is not primarily aimed at gap-filling, but at generating a more accurate and internally consistent AOD dataset by integrating complementary information from multiple sensors and ground-based observations. Thus, even at the monthly scale, the proposed method adds value by reducing retrieval uncertainties and improving the reliability of aerosol distributions, which is critical for climate studies and radiative forcing assessments. The overview of the fusion method is presented in **Fig. 1**, followed by a detailed description of each step in the following sections.

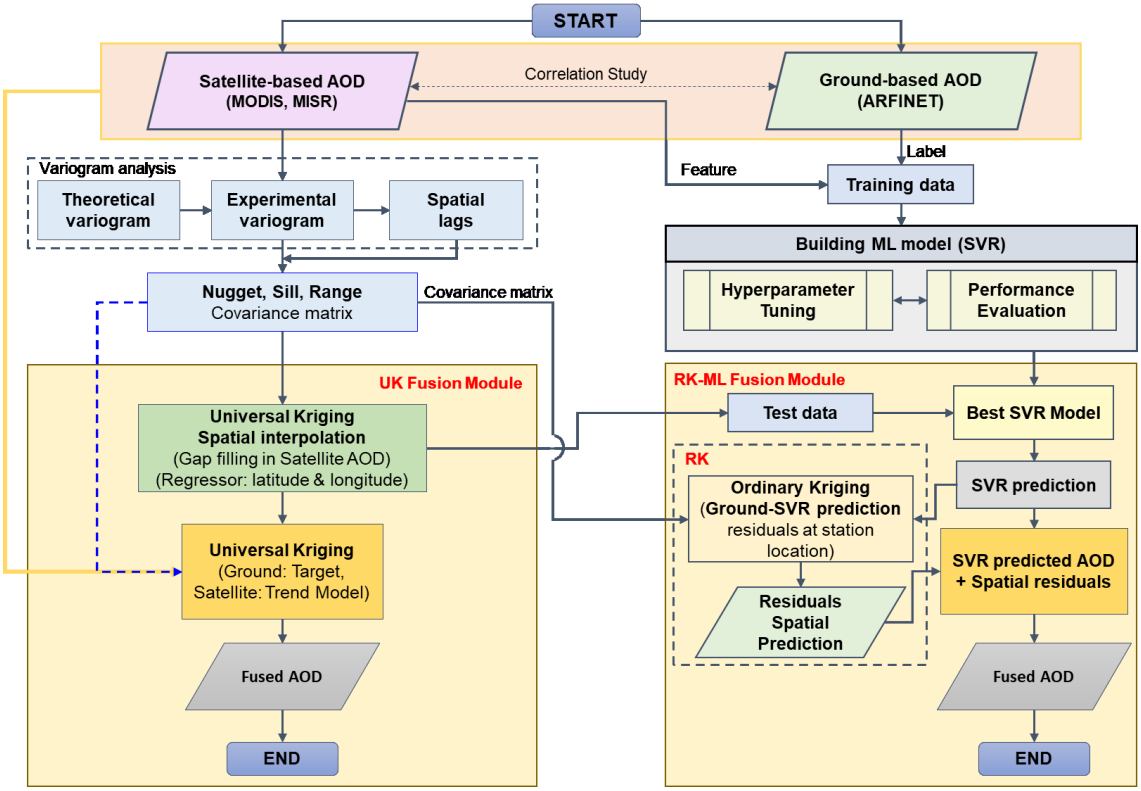


Figure 1. Flowchart of fusion methodology: Universal Kriging (UK) and Residual Kriging Machine Learning (RK-ML). The machine learning best model is designed based on the long term MODIS and MISR data which were collocated with ground-based observations.

2.3.1 Correlation analysis

As a first step of the fusion processes, the correlation analysis between the satellite and ground-based AOD was made to understand the association/ biases between the two data sets at different spatiotemporal scales. This is useful to understand the requirement of multi-sensor data fusion. For this, a statistical spatio-temporal matching approach similar to those reported elsewhere by Basart et al., (2009);Chu et al., (2002); Filonchyk et al., (2019); Ichoku et al., (2002) was applied, in which satellite observations were spatially averaged at 0.5° spatial resolution and compared with ground-based AOD averaged within a 30 minute time window around the overpass time of the TERRA satellite which accommodated 14 to 15 measurements from MWR (data frequency 2 min) and 1 to 2 measurements from CIMEL (data frequency 15 min) observations. Although satellite products such as MODIS (~10 km) and MISR (~4.4 km) provide higher spatial resolution potentially capturing finer regional variability in aerosol distributions, yet their direct comparison with ground-based point measurements introduces representativeness errors due to scale mismatch. Aggregating the data to a coarser grid (0.5°) reduces this mismatch by ensuring that both satellite and ground observations represent comparable spatial scales, thereby improving the robustness of validation and fusion. Thus, the choice of 0.5° represents an optimal choice, yielding higher correlation and lower root mean square error (RMSE) (**Figs. S2a, b**), in addition to retaining regional variability and ensuring sufficient data density within each grid cell for stable statistical estimation and fusion. The consideration of 0.5° resolution is in line with approach adopted by Tandule et al. (2026) for retrieving AOD from satellite observations, ensuring improved representativeness and temporal consistency in comparisons between satellite-derived and ground-based AOD. In addition, generating AOD at this resolution provides a valuable reference dataset for comparison and validation against reanalysis products and model outputs of AOD, where satellite observations are commonly assimilated as primary inputs. Further, AOD observations in this study were aggregated to a monthly scale to ensure more consistent spatial coverage and improve the reliability of multi-sensor fusion analysis. Due to the differences in spatial coverage and revisit characteristics of MODIS and MISR as well as temporal gaps in data availability from ground-based instruments (**Figs. S3a, 3b, 3c**), daily datasets often contained substantial spatial gaps over the study domain.

2.3.2 Variogram Analysis

Variogram analysis is used to quantify and model the spatial autocorrelation (i.e., spatial dependence) of a dataset. It evaluates how the spatial variability between data points changes as a function of lag distance, the distance separating two sample points in space. To capture the spatial dependency of the data, geographical parameters such as latitude, longitude, and elevation are often incorporated as covariates in the trend function, thereby incorporating the spatial context of the sampling locations. This approach has been widely applied in studies involving meteorological parameters (Chua and Bras, 1982; Holdaway, 1996; Nalder and Wein, 1998). In the present context, spatial representation of AOD is fairly represented as a trend function comprising of latitude, longitude, and elevation, which serve as proxies for underlying spatial variations of geographical and atmospheric influences that significantly affect aerosol distribution. However, it is important to note that most geostatistical methods, such as Kriging, assume the underlying field to follow second order or, intrinsic stationarity (mean is constant, and the covariance or, variance of increments depends only on spatial lag).

However, real-world environmental and geophysical data often exhibit large-scale spatial trends driven by physical and geographical factors, such as latitude, longitude, and elevation. In the case of AOD, these variables act as key spatial predictors that capture dominant regional gradients and can be used to model and remove the large-scale spatial trend. In the presence of strong spatial trends, variogram may become unbounded or exhibit unrealistically large ranges. These spatial trends violate the stationarity assumption which can lead to unbounded variogram. To address this spatial detrending of the data is performed, which isolates the local fluctuations or residuals from the spatial data set. This serves as an essential step in geostatistical analysis to ensure a well-defined and bounded variogram, enabling reliable estimation of sill, nugget, and range parameters for spatial covariance modeling. To validate this assumption, we obtained the frequency distribution of satellite AOD and their residuals (**Fig. S4**) after detrending. A nearly symmetric histogram of detrended residuals indicates that the trend component has been effectively removed, which is a prerequisite for second-order stationarity (Tang et al., 2016). Since the detrending in our study is purely a spatial operation, the temporal dimension is not explicitly considered and is effectively treated as constant during the detrending process. Consequently, the approach does not involve long-term datasets or explicitly account for seasonal variability (e.g., all four seasons).

The semivariance, which measures the degree of spatial variability between pairs of sample points as a function of their separation distance, known as the lag-distance (h), is calculated as:

$$\gamma(h_x) = {}^{1}/_{2n(h)} \left(\sum [z(x_i) - z(x_i + h)]^2\right) \quad (4)$$

Where $z(x_i)$ and $z(x_i+h)$ are the values of the variables of interest at locations x_i and $x_i + h$ (= x_j), respectively; n(h) is the number of pairs of points separated by the lag-distance h, which is given as:

$$h_x = r \cos^{-1}(\sin \varphi_i \sin \varphi_j - \cos \varphi_i \cos \varphi_j \cos(\theta_i - \theta_j)) \quad (5)$$

Where $\varphi_{i,j}$ represent longitudes of locations x_i and x_j, and $\theta_{i,j}$ represent latitudes of locations x_i and x_j; r is the mean radius of the earth. Following this, the empirical variogram is calculated from the actual observational data, showing the relationship between semivariance and lag distance for each set of observations. The experimental variogram is obtained after binning semivariance at certain lags of the empirical variogram. The experimental variogram is then fitted with a theoretical model to describe the spatial continuity of the variable. The theoretical models considered in the present study include Exponential, Spherical, and Matheron models; the mathematical expressions are given as:

$$\gamma_{theo}(h_x) = \begin{cases} 0 & h_x = 0 \ ; \\ \sigma_n^2 + \sigma_b^2 \left(1 - \exp\left(-{}^{h_x}/_{l}\right)\right) & h_x > 0 \end{cases} \quad \text{(Exponential)} \quad (6)$$

$$\gamma_{theo}(h_x) = \begin{cases} 0 & h_x = 0 \ ; \\ \sigma_n^2 + \sigma_b^2 \left({}^{3h_x}/_{2l} - {}^{1h_x{}^3}/_{2l^3} \right) & 0 < h_x \le l \ ; \\ \sigma_n^2 + \sigma_b^2 & h_x > l \end{cases} \quad \text{(Spherical)} \quad (7)$$

$$\gamma_{theo}(h_x) = \begin{cases} 0 & h_x = 0; \\ \sigma_n^2 + \sigma_b^2\left(1 - \exp\left(-{h_x}^2/l^2\right)\right) & h_x > 0; \end{cases} \quad \text{(Matheron)} \qquad (8)$$

In the above equations, $\sigma^2 = \sigma_n^2 + \sigma_b^2$, represents the total variance observed in AOD data at larger lag distances (spatially uncorrelated AOD data). σ_n^2 is nugget (y-intercept of the variogram), which represents the semivariance at a very small lag distance, approaching zero. Nugget (spatial variation at distances smaller than the smallest sampling interval) is indicative of the presence of measurement error or noise in the data. A large nugget relative to the sill (i.e., the semivariance value where the variogram levels off, representing maximum variability or correlation between data points at a given spatial distance) suggests significant measurement error or unresolved variability. This can indicate potential issues with data quality. On the other hand, a small nugget implies that the data is relatively free of noise and that most of the spatial variability is due to the structured spatial process. σ_b^2 is variance in spatially correlated data, and this parameter gradually increases with increasing lag distances until it reaches sill. l is the range parameter, the distance at which the semivariance reaches the sill; up to this distance, data are spatially correlated with each other. The higher the range, the more similar the values are at greater distances from each other. The spatial covariance function can be derived from the variogram model as:

$$C_{ij} = \sigma^2 - \gamma_{theo} \qquad (9)$$

2.3.3 Universal Kriging

Universal Kriging (UK) also referred to as Kriging with a trend model, extends Ordinary Kriging by incorporating a deterministic trend component alongside the stochastic spatial component. This approach is useful when there is an underlying trend in the data that varies across the study area. The UK method uses both the spatial autocorrelation structure and the deterministic trend to make predictions. The UK model can be expressed as:

$$Z = M_z\beta + \in \quad \text{Or, } \in = Z - M_z\beta \qquad (10)$$

Where, $Z = [Z(x_1), Z(x_2), \ldots\ldots, Z(x_n)]^T$ represent the values of the variables of interest at locations $x_1, x_2, \ldots\ldots x_n$, respectively. M_z is the deterministic trend component of the model $(n \times p)$ where p is representing the number of regressors; and β is the unknown drift coefficient $(p \times 1)$ to be estimated; $\in$ is the stochastic component or stochastic residuals $(n \times 1)$, i.e., mean zero random fields.

In the present study, the trend component M_z for fusion is defined as

$$M_z = \begin{pmatrix} 1 & MODIS_{AOD_1} & MISR_{AOD_1} \\ 1 & MODIS_{AOD_2} & MISR_{AOD_2} \\ \vdots & \vdots & \vdots \\ 1 & MODIS_{AOD_n} & MISR_{AOD_n} \end{pmatrix} \qquad (11)$$

This is similar to a multiple regression model, which is described through a combination of a constant term and two sensor measurements that act as regressors to predict AOD at estimation locations. The first component of this trend model represents the overall offset (i.e., the mean of the portion of the AOD distribution that is not captured by MISR and

MODIS). This constant term thereby represents any systematic offset between the combined (MISR and MODIS) satellite-retrieved AOD and the ground-measured AOD.

Following equation (10), the expected value at prediction locations (x_s) can be expressed as the best linear unbiased prediction (BLUP):

$$\hat{Z}(x_s) = {m_s}^T\hat{\beta} + C_{zs}^T C_{zz}^{-1}(Z - M_z\hat{\beta}) \quad (12)$$

Here, $C_{zs}(n \times s)$ is the spatial covariance matrix of the residuals between the sample location (i.e., measurement locations) and prediction locations (i.e., estimation locations) and $C_{zz}(n \times n)$ is the spatial covariance matrix of the residuals between the sample locations (i.e., measurement locations) as obtained from equation (9). The unknown coefficient $\hat{\beta}$ can be expressed as the generalized least squares (GLS) estimator from the covariance matrix,

$$\hat{\beta} = \left({M_z}^T {C_{zz}}^{-1} M_z\right)^{-1} {M_z}^T {C_{zz}}^{-1} Z \quad (13)$$

Alternatively, minimizing the mean square error (MSE) of all predictions among the predictors of the form $\lambda^T Z$ subjected to unbiasedness constraint, i.e., $E(\lambda^T Z) = E(Z(x_s))$ for all β, which is identical to $\lambda^T M_z\hat{\beta} = {m_s}^T\hat{\beta}$ and under conditions for minimizing variance ($\lambda^T Z - Z$), Lagrange multipliers ($\mu(p \times s)$) are used to solve the linear constraint equations as given below,

$$\begin{bmatrix} C_{zz} & M_z \\ {M_z}^T & 0 \end{bmatrix}\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_{zs} \\ m_s \end{bmatrix} \quad (14)$$

Here, M_z $(n \times p)$ and ${M_z}^T(p \times n)$ are trend models of AOD given by equation (11); m_s $(p \times s)$ is trend model at s estimation locations; $\lambda(n \times s)$ are the Kriging weights, μ is the Lagrange multiplier.

The system of equations is solved for Lagrange multiplier μ and weights λ to estimate AOD at estimation locations. This can be expressed as:

$$\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_{zz} & M_z \\ {M_z}^T & 0 \end{bmatrix}^{-1}\begin{bmatrix} C_{zs} \\ m_s \end{bmatrix} \quad (15)$$

$$\lambda = \left\{{C_{zz}}^{-1} - {C_{zz}}^{-1} M_z\left({M_z}^T {C_{zz}}^{-1} M_z\right)^{-1} {M_z}^T {C_{zz}}^{-1}\right\} C_{zs} + {C_{zz}}^{-1} M_z\left({M_z}^T {C_{zz}}^{-1} M_z\right)^{-1} m_s \quad (16)$$

$$\lambda^T Z = {C_{zs}}^T {C_{zz}}^{-1}(Z - M_z\beta) + {m_s}^T\beta \quad (17)$$

The prediction variance associated with predicted values can be represented as

$$\mathrm{Var}\left(Z(x_s) - \hat{Z}(x_s)\right) =$$

$${\sigma_{Z(x_s)}}^2 - {C_{zs}}^T {C_{zz}}^{-1} C_{zs} + \left({m_s}^T - \left({M_z}^T {C_{zz}}^{-1} C_{zs}\right)^T\right)^T \left({M_z}^T {C_{zz}}^{-1} M_z\right)^{-1}\left({m_s}^T - \left({M_z}^T {C_{zz}}^{-1} C_{zs}\right)^T\right) \quad (18)$$

The above weighting approach determines the values at prediction locations. Our foremost approach involved creating a trend model for fusion. For this purpose, we generated a complete satellite-based map of AOD from MODIS and MISR separately over the study region using the UK method. In this framework, geographical parameters such as latitude, longitude and elevation are treated as regressors (trend model), whereas observed satellite data serve as response variables to fill the gaps in individual satellite datasets. Subsequently, in the final spatial fused predictions, the ground-based AODs were treated as the response variables, where the satellite data, along with the elevation model (used as additional information), were used as regressors.

2.3.4 Residual Kriging Machine Learning (RK-ML)

The RK-ML approach is implemented in this study through a hybrid approach to generate reliable estimates of fused AOD, with improved predictive accuracy using a limited of ground-based observations. Although UK and the proposed RK-ML approach appear methodologically different, they share a common conceptual foundation. From a generalized regression perspective, both methods rely on a covariance structure to characterize spatial dependence among grid points. However, when the number of observations is limited, the statistical parameters in UK can adversely affect predictions at estimation locations. To overcome the limitations associated with limited observations, SVR is employed due to robustness of its regularized formulation in both regression and classification tasks (Sifaou et al., 2021), where predictions are learned from datasets within a historical training window. These predictions provide prior information on ground-based AOD that is independent of the spatial configuration of the current ground based-observations.

The best-performing SVR model is first identified based on a time window of five years of data (simultaneous MODIS, MISR and ground measurements) for a specific month (season) targeting consistent aerosol conditions. Subsequently, this model leverages spatially interpolated features from MODIS and MISR data (gap filled AOD by UK) to generate SVR-predicted maps to obtain a full generalized AOD map for that month. The discrepancies between SVR-predicted values and observed ground measurements are treated as residuals, which are then spatially modeled using Ordinary Kriging. The resulting residual predictions are combined with SVR outputs to produce RK-ML-based fused products. This approach provides a robust alternative to traditional UK-based data fusion techniques by capturing complex relationships between predictors and target variables. The detailed RK-ML methodology is given below.

SVR transforms features into a higher-dimensional space, making them linearly separable and helps improving the prediction of target variables like ground based AOD. The use of SVR in Kriging has been reported in previous studies to improve model predictions (Wang et al., 2008; Baisad et al., 2023). The SVR model is represented by:

$$Z_{svr} = w^T \varphi + b \tag{19}$$

where φ is the kernel transformed input features, w is the nonzero vector normal to hyperplane (the plane or decision boundary that best fits the n dimension input vectors while maintaining a margin of tolerance (ε-insensitive zone) around it) and b ϵ R. This expression assumes that Z_{svr} exist when it approximates all $w^T\varphi$ with ε precision for linearly separable data. Along with it, the concept of soft margin loss function is considered which introduces slack variable $\xi(+ve)\ and\ \xi^*(-ve)$ to allow some points lying inside the hyperplane.

Hence the optimization problem is subject to minimization of

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi + \xi^*) \quad \text{Such that} \begin{cases} Z - w^T\varphi - b \le \varepsilon + \xi \\ w^T\varphi + b - Z \ge \varepsilon + \xi^* \\ \xi_i\xi_i^* \ge 0 \end{cases} \tag{20}$$

The regularization constant C trades off between the model complexity and empirical error up to which deviations larger than ε can be tolerated. $\xi,\ \xi^*$ are regression errors. The detailed explanation of SVR model is available in the literature (Smola & Schölkopf, 2004; Brereton

& Lloyd, 2010). In the present study, best model of SVR is decided from the minimum RMSE between predicted AOD and real AOD after tuning its hyperparameters.

The residuals, which are the difference between collocated AOD of SVR predictions from ground based AOD, are estimated from the difference between $\hat{Z}_{svr}(x_s)$ and Z , *i.e.,* $\hat{Z}_{svr}(x_s) - Z$. These residuals under ordinary Kriging are modeled as $\delta(s) = \mu + \varepsilon(s)$ where μ refers to the mean values of residuals over study domain, which resembles similar mathematics of universal Kriging, except M_z & $m_s = 1$ (eq. (10) to eq. (17)). Following this, the estimated residuals at unknown locations are determined as

$$\delta(x_s) = \lambda^T \delta(x_z)$$

The weighting parameter λ is obtained from the covariance matrices as follows

$$\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_{zz} & 1 \\ 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{zs} \\ 1 \end{bmatrix} \quad (21)$$

The ordinary Kriging estimation contains the spatial relation while the SVR prediction contains the optimal estimations from features (Satellite) and labels (Ground AOD). The final fused map can then be estimated as

$$Z = \delta(x_s) + Z_{svr}(x_s) \quad (22)$$

The hyperparameters of the SVR model were optimized using a grid-search strategy as part of the training process. In this approach, predefined values for key hyperparameters - such as the regularization parameter (C), kernel type, gamma (ϒ) and epsilon (ε) - were combined to form all possible parameter configurations, and each configuration was evaluated. For every hyperparameter combination, model performance was assessed using the negative mean squared error (neg-MSE) as the evaluation metric. This metric quantifies prediction error, and allows selection of the best model by maximizing the score. Due to the limited size of the dataset, leave-one-out cross-validation (LOOCV) was used during training instead of k-fold cross-validation. Even though other cross-validation approaches, such as site-based, temporal, or sample-based validation, can also be used to assess model robustness, the LOOCV was considered the most suitable approach in this study considering the limited number of samples and the uneven spatial distribution of ground observations. In LOOCV, the model is trained repeatedly on all samples except one, which is used for validation. This procedure is repeated so that each data point serves once as the validation sample. This approach maximized the use of available data while providing an unbiased estimate of model performance. Based on this procedure, a linear kernel was identified as the optimal choice for the RK-ML models, which were subsequently evaluated on independent test sets (20% for MODIS, N=318; 10% for MISR, N=71). For MODIS AOD features, the best model configuration was C = 1, gamma = 'scale', kernel = 'linear'; for MISR AOD features, the optimal configuration was C = 100, gamma = 'scale', kernel = 'linear'. The use of a linear kernel suggests a predominantly linear relationship between satellite observations and ground-based AOD. Inclusion of the regularization parameters C and gamma controls overfitting and penalizes noisy inputs, enabling the ML framework to generate more reliable estimates. These estimates were further corrected using spatial residuals from RK, allowing RK-ML to outperform UK under conditions of limited or biased AOD observations. The final models were evaluated using correlation coefficient (R), R^2, mean absolute error (MAE), and

RMSE on both training and test sets (**Table ST1**). Results indicate that training and test performances were comparable (with training R^2 values either lower or close to test R^2), correlation coefficients were consistently high, and errors (RMSE, MAE) were low. These outcomes confirm that the SVR models did not suffer from overfitting and generalized well to unseen data, despite the limited sample size.

Although other machine learning algorithms, such as Random Forest (RF) and XGBoost, can also be integrated within the RK-ML framework, a sensitivity analysis conducted on MODIS test dataset indicated that SVR achieved comparable or better performance metrics (R, RMSE; Supplementary Fig. S5).

3. Results and discussions

3.1 Regional distribution of AOD from ground-based and satellite observations

The analysis of spatial distribution of AOD is crucial for understanding the consistency of measurements across different sensors. The large-scale spatial variations in the data help identify overall spatial trends over latitude-longitude and geographic elevation. Emphasizing spatial trends is also critical for assessing the mathematical assumptions underlying Kriging and variogram analysis, which rely on the condition of second-order stationarity within the sampled data. In view of this, ground-based AOD at 500 nm was considered for long-term comparison of MODIS and MISR AOD at 550 nm as the closest approximation. The typical AOD patterns over different regions over India from ground-based measurements, derived from 10 years of measurements from the ARFINET database using MWR and Microtops instruments, is illustrated in **Fig. 2.**

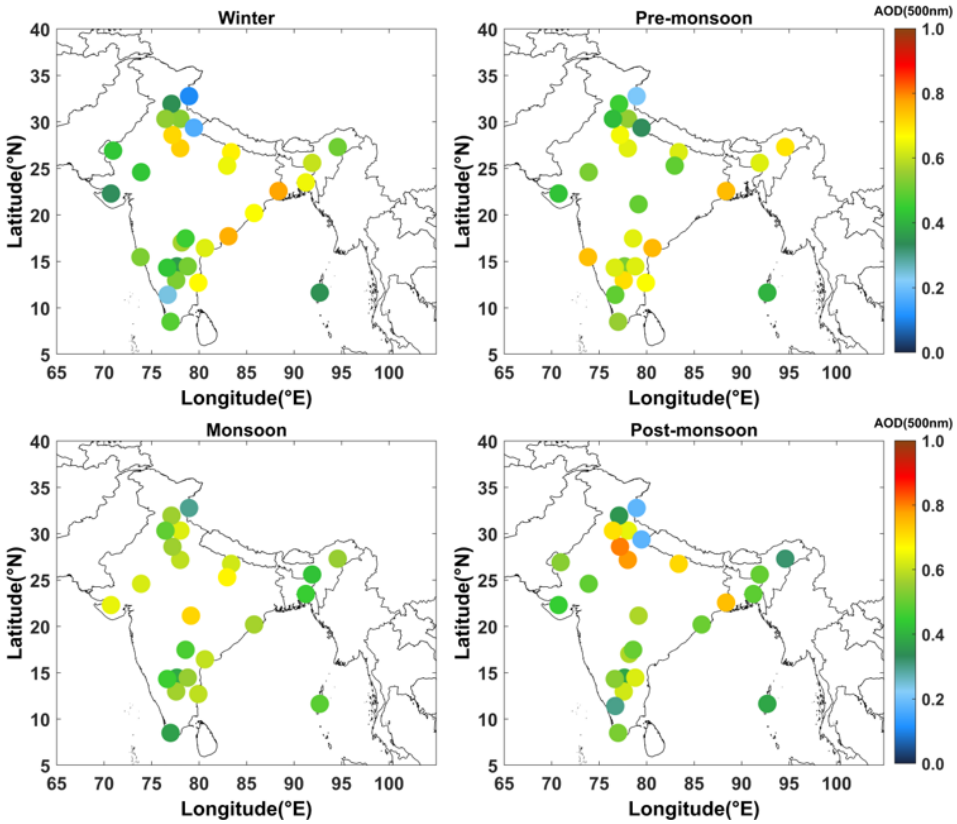


Figure 2. Long-term (2011-2020) ground-based AOD at 500 nm from MWR and MICROTOPS-II measurements in the ARFINET over the Indian region. The seasons are

winter: December, January, February (DJF); Pre-monsoon: March, April, May (MAM); Monsoon: June, July, August, September (JJAS); Post-monsoon: October, November (ON). The different regions considered for representing Indo-Gangetic plane (IGP), North-west (NW), North-east (NE), Peninsular India (PI), and Central India (CI) is provided (**Fig. S6**).

Various factors such as the dominance of natural and anthropogenic sources, local and synoptic meteorology cause observed spatio-temporal variations in AOD at a particular location. Over most of the locations in the Indo-Gangetic Plains (IGP), AOD shows consistent high values (> 0.6) throughout different seasons. This is similar to the observations reported by Lodhi et al., (2013); Singh et al., (2020); Tiwari et al., (2018). Next to the IGP, the north-eastern (NE) India experiences higher AOD with peak during the pre-monsoon season. Similar pattern is reported elsewhere (Gogoi et al., 2009). In Peninsular India (PI), AOD is highest during the pre-monsoon period, followed by a significant reduction during the summer monsoon. This is similar to the earlier studies by Kalluri et al., (2016); Kumar et al., (2009); Sinha et al., (2013); Vachaspati et al., (2018).

The spatial patterns of a decadal average satellite-based MODIS and MISR AOD (2011–2020; **Fig. 3a, 3b**) also shows persistent high AOD values in the IGP and its outflow regions across all seasons. In PI, the presence of elevated mountain ranges such as the Western and Eastern Ghats, coupled with its proximity to the Indian Ocean, results in regional-scale AOD variability. During the pre-monsoon and monsoon periods, oceanic and coastal regions exhibit higher AOD levels compared to the winter and post-monsoon periods. Though dominant spatial patterns are same in long term AOD from spaceborne sensors and their differences with ground-based AOD over same spatial grids are minimal (**Figs. S7a, b**), the discrepancy persists, especially over northern India and during monsoon period. As cloud-haze misclassifications may act as one of the factors from the observed differences between satellite and ground AOD in the monsoon periods, haze-removal criteria (following Jiao et al., 2023) to MODIS AOD was applied. A significant impact of haze over the peninsular region is seen during monsoon (Supplementary **Figs. S7c, d**), however, it shows negligible influence during the other seasons. This is clearly seen from the difference maps between the MODIS and ground-AOD over different ground locations, showing minimal changes before and after the haze removal. This exercise suggests that cloud-haze misclassification is not the primary factor driving the observed differences, except under monsoon conditions. Under such a scenario, localized discrepancies may arise due to spatial sampling limitations of the ground- based observations. As the ARFINET stations are sparsely and unevenly distributed, particularly across regions of high aerosol loading in northern India, this may result in the apparent lack of complete regional representation of ground observations. Additionally, discrepancies between MODIS and MISR AOD are also seen owing to fixed and multi angle retrievals especially in Pre-monsoon period over the NW region, where MISR AOD is significantly different from MODIS. There are also some pockets where low AOD region observed by MODIS is alternatively represented as a region of higher AOD in MISR observations, particularly in proximity to the IGP outflow. Previous studies over similar geographic regions have indicated that the frequency of observations, cloud masking, and geographical factors impact both MODIS and MISR observations, stemming from algorithm assumptions related to cloud masking and SSA. Overall, the spatial patterns of AOD from ground and satellite observations reveal the following:

- During the pre-monsoon period, northern India experiences increased AOD.
- During the winter season, cold temperatures, a low boundary layer height, and humid air create hazy conditions with high AOD (Nair et al., 2020). Along with it, winds over the IGP are mostly north-westerly, with an anti-cyclonic pattern over central India, driving aerosols to peninsular region.
- The post-monsoon AOD also remains high, similar to winter levels, particularly in the IGP due to biomass burning (Kumar et al., 2012; Lodhi et al., 2013; Subba et al., 2021).
- The spatial patterns of AOD across different seasons are well captured by both satellite and ground-based observations. However, notable differences exist between ground-based and MODIS and MISR AOD. While MODIS tends to overestimate AOD over the IGP, it generally underestimates AOD over the PI, NE, and NW regions. MISR significantly underestimates higher AOD regions.

Despite the above constraints, the general agreement in magnitude and temporal variability supports the reliability of both datasets for the fusion framework. Thus, our approach explicitly accounts for such discrepancies by integrating the broad spatial coverage of satellite observations with the higher accuracy of ground-based measurements. In this context, ground observations are treated as local constraints rather than complete spatial representations, thereby minimizing the influence of regional sampling gaps in the ground network. Consequently, the final fused AOD represents a bias-corrected satellite-derived field constrained by ground observations.

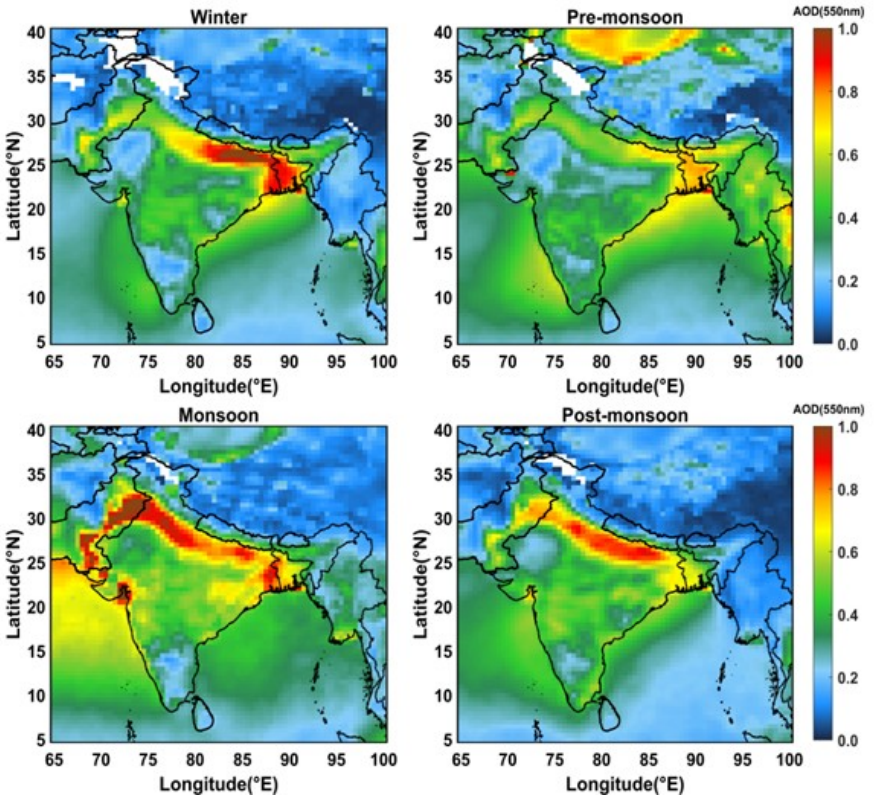


Figure 3a. Long-term (2011-2020) satellite based AOD (at 550 nm) from MODIS over south-Asian region.

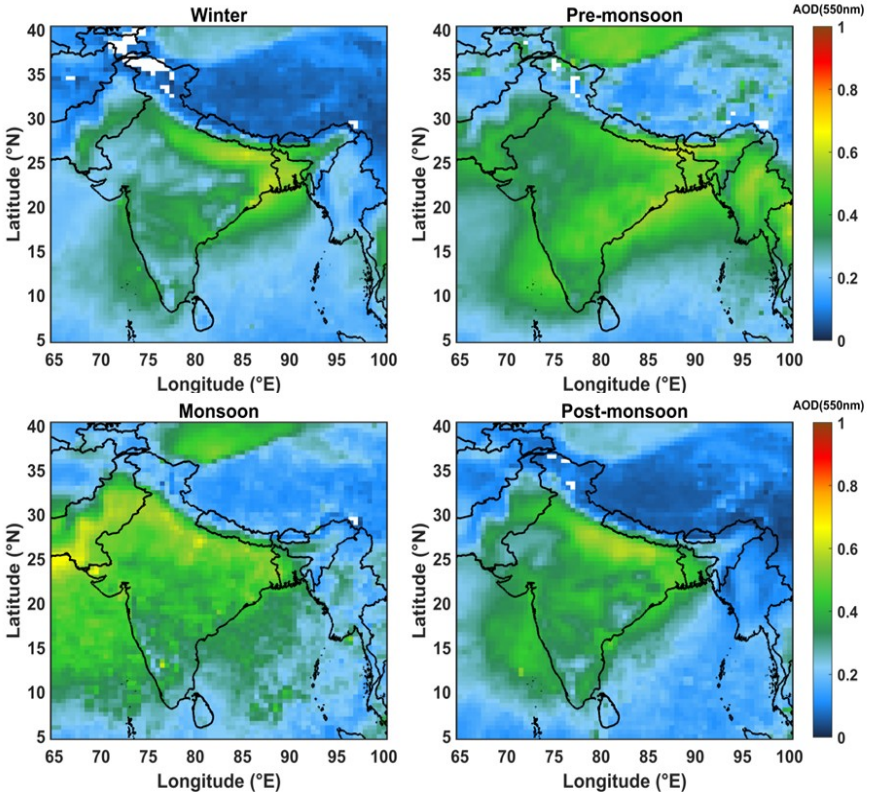


Figure 3b. Long-term (2011-2020) satellite based AOD (at 550 nm) from MISR over south-Asian region.

3.2 Inter-comparison of satellite- and ground-based AOD

Having examined the spatial distribution, a quantitative evaluation of the associations or biases between satellite and ground-based AOD at different periods is examined for the years 2012, 2016, and 2021. The three different years were selected such that way that the AOD for 2012 and 2021 provides the decadal variability, while 2016 represents an intermediate period between these two years, enabling us to better assess the progression and variability of AOD over a long period. The scatter plots (Supplementary **Figs. S8-S10;** the number of ground stations included in the correlation studies is given in **Table 1**) between MODIS/ MISR and ground-based AOD highlight moderate to strong correlations (~ 0.8-0.9) in winter (January) and post-monsoon (November), while moderate correlations (~0.54-0.77) between the two are observed in pre-monsoon (May). The RMSE between MISR and ground-based AOD is higher (≥ 0.2) during winter and post-monsoon, whereas higher RMSE values between MODIS and ground-based AOD are observed during the pre-monsoon period. The prominent locations contributing to mean errors and weak correlations with ground observations are situated in the NW and IGP regions.

The quartile-plots (**Figs. S11-S13**) highlight significant spatio-temporal variability in AOD, with both sensors displaying higher AOD over terrestrial regions, particularly in the IGP, its outflows, and South (Peninsular) and Central India. The third and fourth quartiles are more representative for AOD over land regions than in surrounding areas like oceans and elevated terrain. Data with respect to longitude and latitude show that higher AOD values are mostly confined to 20°-30°N latitude and 80°-95°E longitude. However, MODIS consistently recorded significantly higher AOD values than MISR, with notable dissimilarities in quartile patterns over northern India during May.

Both the correlation and quartile analyses highlight the advantages and limitations of MODIS and MISR observations. For example, MISR tends to underestimate high AOD conditions in urban regions compared to MODIS, even though it can effectively separate surface contributions under low aerosol loading, as also reported by Tao et al. (2020). Under high AOD conditions, the benefit of multi-angle measurements becomes limited, as thick aerosol layers smooth out surface reflectance signals (BRDF), potentially leading to an underestimation of AOD due to misattribution of aerosol contributions to surface reflectance. In contrast, when dust loading is dominated by coarse and non-spherical particles (in May), MISR demonstrates relatively better performance than MODIS (as shown by scatter plots). This difference may be attributed to the advantage of MISR's multi-angle observation capability, consistent with findings from Middle Eastern validation studies (Farahat, 2019; Garay et al., 2017).

Table 1: Number of ground stations data used in different months of the year 2012, 2016, and 2021.

2012 Jan	2016 Jan	2021 Jan	2012 May	2016 May	2021 May	2012 Nov	2016 Nov	2021 Nov
21	26	16	22	25	13	27	26	16

3.3 Fusion of satellite- and ground-based AOD

3.3.1 Variogram analysis

For the fusion of satellite- and ground-based AOD, the experimental variogram (using eq.2) is first obtained from the gridded satellite data. As mentioned in section 2.3.2, a well-fitted variogram is essential for determining appropriate parameters in geographical processes. These variogram parameters like sill, range, and nugget are not unique but depend on the theoretical models used. The choice of fitting is determined through a least square approach, selecting the best fit based on the least sum of squared errors (SSE). However, the availability of a large number of satellite data sets has made this task easier. The variogram depicted in supplementary **Figs. S14, S15, and S16** demonstrate a flattening of variance after a certain lag (interval between distances), affirming the effectiveness of our implemented detrending method. AOD values within the range highlight spatial correlation, wherein the correlated AOD values are influential in determining missing AOD values. The variogram parameters obtained from the fitted theoretical model are given in **Tables 2 and 3**.

The variogram parameters corresponding to different sensors exhibit noticeable variation across months and years, reflecting differences in their retrieval characteristics and ability to represent AOD. For instance, both MODIS and MISR show shorter spatial correlation lengths in May compared to January and November. Such reduced ranges are typically associated with long-range dust or smoke transport processes, which dominate during this period in the study region. Conversely, longer ranges indicate that the sensor retrievals capture more spatially homogeneous values, suggesting an improved ability to represent regional heterogeneity. In this study, we prioritize MODIS variograms because of their higher sill and range values, which demonstrate stronger spatial dependency (Isaaks, 1991; Vieira et al., 2009). Nevertheless, sensitivity tests indicate that using variograms derived from either MODIS or MISR produces only negligible differences (~0.01) in the fused AOD estimates. (**Fig. S17**). At this stage, it is also to be noted that geographically weighted or local variogram

approaches can better represent spatial heterogeneity, particularly over complex terrains such as the Himalayas and Western Ghats. However, in the present study, this approach was not feasible due to the limited availability of ground-based AOD observations, especially across high-altitude regions. The sparse coverage restricts the stability and generalizability of local variogram fitting, particularly at regional boundaries where different models would be required. For this reason, we adopted a single variogram model, following the approach used for large regions (e.g., eastern and western USA; Chatterjee et al. (2010)), which provides a more consistent framework for regional-scale fusion).

Table 2: Parameters obtained from variogram in different seasons (January, May, and November) of different years (2012, 2016, and 2021) from MODIS.

Year	Month	MODEL	Nugget	sill	Range (in km)
2012	Jan	Matheron	0	0.033954	848.5
2016	Jan	Exponential	0.000015	0.057399	453.3
2021	Jan	Exponential	0	0.052327	854.0
2012	May	Matheron	0	0.026499	555.6
2016	May	Matheron	0	0.029704	489.1
2021	May	Matheron	0	0.034273	378.3
2012	Nov	Matheron	0	0.030027	548.3
2016	Nov	Spherical	0	0.029582	1080.7
2021	Nov	Exponential	0	0.040153	696.2

Table 3: Parameters obtained from variogram in different seasons (January, May, and November) of different years (2012, 2016, and 2021) from MISR.

Year	Month	MODEL	Nugget	sill	Range (in km)
2012	Jan	Exponential	0.000062	0.011589	645.6
2016	Jan	Spherical	0.000177	0.017749	1060.6
2021	Jan	Matheron	0	0.027958	927.9
2012	May	Matheron	0	0.016100	371.1
2016	May	Exponential	0.001741	0.018171	463.0
2021	May	Matheron	0.000833	0.017896	533.5
2012	Nov	Exponential	0.000022	0.014509	638.7
2016	Nov	Matheron	0	0.018378	583.0
2021	Nov	Matheron	0.0000925	0.019104	516.4

3.3.2 Spatial interpolation of AOD

Monthly mean AOD gives the advantage of having full regional picture of columnar aerosol load over south-Asian region. However, it is observed that both MODIS and MISR AOD show gaps in some of the regions, either due to consistent cloud coverage or due to complex orography coupled with highly reflective land masses (e.g., snow-covered regions of the Himalayas). Hence, UK with geographic parameters as regressors applied to fill these missing areas (which are found to be ~ 2-11%) to obtain a complete spatial picture of AOD over the south-Asian region. Since kriging gives a probability map, the associated variance is higher in

the gap regions than in the regions where values exist. Thus, the interpolated values and variances are not unique, as they depend on the variogram and trend models used in the interpolation. On the other hand, the variogram can have uncertainties that stem from factors such as lag spacing, the quantity of data points, and model fitting, as highlighted by researchers (Derakhshan and Leuangthong, 1982; Koushavand and Deutsch, 2008). Hence, our proposed approach is not intended solely to enhance spatial coverage, but to generate a bias-corrected and internally consistent AOD dataset through the optimal integration of complementary satellite products.

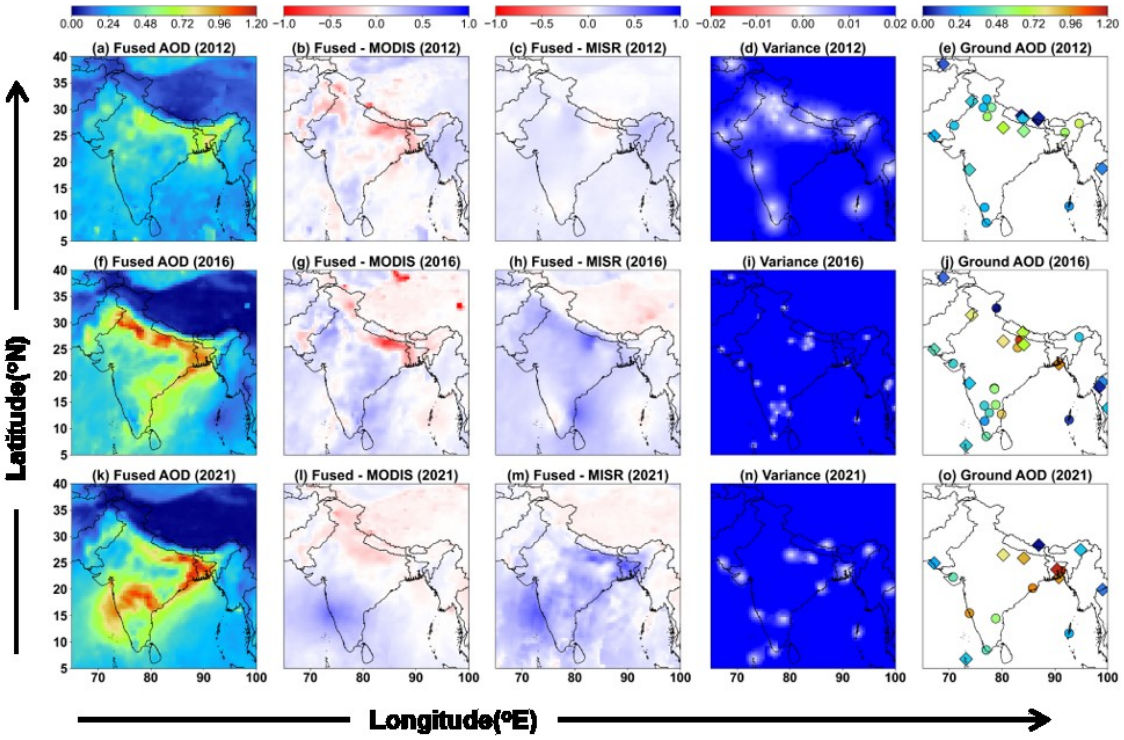
The spatial distribution of 0.5° gridded monthly mean raw and predicted (after spatial interpolation) AOD from MODIS and MISR is shown in supplementary **Figs. S18-S23**. The performance metrics (in terms of R and RMSE) of this gap-filling approach are demonstrated through sensitivity studies**,** considering different spatial gaps in the data (**Fig. S24)**. It is observed that the predicted AOD largely depends on the availability and spatial distribution of nearby observed data points. Regions such as the IGP, the Himalayan region, peninsular India, and oceanic areas generally show better performance, where smoother spatial gradients of AOD and more consistent regional aerosol patterns improve the reliability of the spatial predictions. Based on these sensitivity studies, the predicted AOD field appears to provide a reliable spatial representation with acceptable uncertainty as the interpolation (gap filling) is made over a relatively small fraction of missing values based on a large number of observed data points around the gap areas.

3.3.3 Fused AOD

The monthly fused AOD is generated using the UK fusion method, where satellite data are treated as trend model. This retains the overall spatial signatures of AOD from each satellite sensor. The optimal AOD values are determined by the weights obtained from spatial relationships, along with the trend of the satellite-based AOD at the estimation locations. **Figs. 4-6** show the fused maps of AOD at different seasons of the years 2012, 2016, and 2021. The regional average values of fused AOD, along with AOD from individual sensors, are given in **Supplementary Tables ST4, ST5,** and **ST6**.

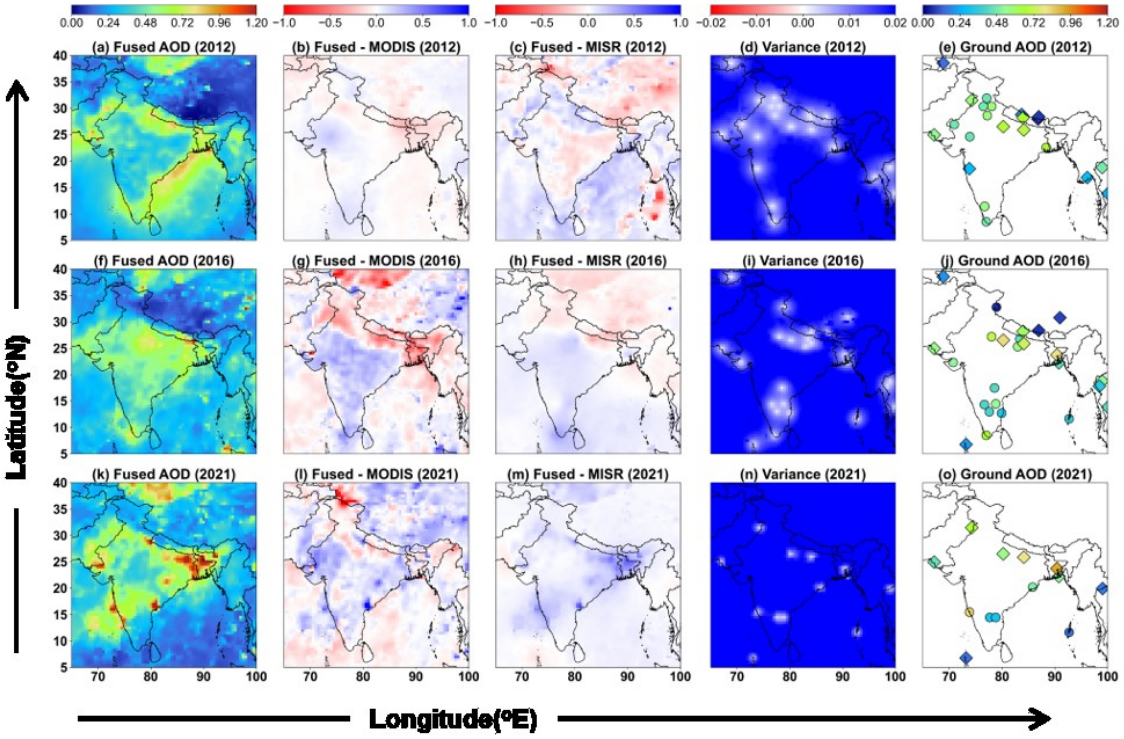
Throughout the observation period, the fusion maps highlight the significant influence of ground-based AOD on the fusion method. As shown in **Table 4**, the fused AOD is more aligned to ground-based AOD with a correlation (R) ~0.994-1 and RMSE ~0.009-0.04. On the other hand, correlation between MODIS/MISR and fused AOD shows little improvement compared to that between MODIS/MISR and ground-based AOD (**Table 5a, b**). This indicates the robustness of the fusion approach keeping ground-based AOD as an anchoring reference.

The notable outcomes from fused AOD maps are the distinct spatial features compared to those obtained from individual space-based sensors. For example, during January 2016, the significant overestimation by MODIS (AOD ~ 1.7) relative to ground-based AOD (~ 1.2) is adjusted in the fused AOD distribution, thus correcting the bias but maintaining the spatial heterogeneity. Similarly, in Jan 2021, MODIS and MISR significantly underestimated AOD over peninsular India (MODIS AOD ~ 0.36 and MISR AOD ~ 0.30) as compared to ground AOD (e.g., AOD at GOA ~ 0.95). The fused AOD corrects this bias toward values closer to ground-based observations, retaining continuous flow of aerosols toward the Arabian coast.



741

742 **Figure 4.** Monthly fused AOD (at 550 nm) maps [(a), (f) and (k)] in January for the years
743 2012, 2016, 2021; [(b), (g) and (i)] and [(c), (h) and (m)] - the deviations of fused AOD from
744 corresponding sensors, i.e., MODIS and MISR; [(d), (i) and (n)] variance; and [(e), (j) and
745 (o)] - ground-based AOD (ARFINET is represented by circle and AERONET by diamond
746 shapes) used to generate fused maps. Blue indicates overestimations, and red means an
747 underestimation of fused AOD from satellite retrieved AOD. The white dots in variance plots
748 show the ground station locations.



749

750 **Figure 5.** Similar analysis as in Fig. 4, but for the month of May in 2012, 2016, and 2021.

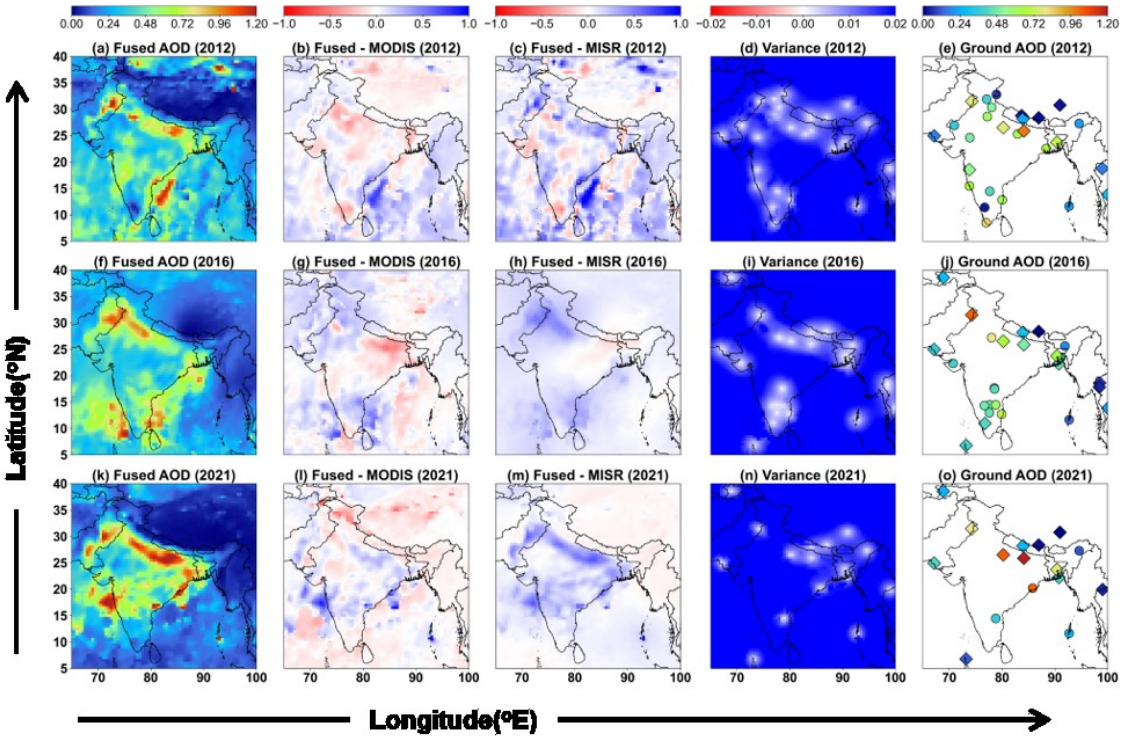


Figure 6. Similar analysis as in Fig. 4, but for the month of November in 2012, 2016, and 2021.

Table 4: Comparative analysis of fused vs ground –based AOD at ground station locations.

Year	Month	Fused & Ground				
		Correlation	RMSE	MAE	Slope	Bias
2012	Jan	0.998	0.014	0.011	1.024	0.003
	May	0.998	0.014	0.011	1.012	0.003
	Nov	0.996	0.025	0.017	1.026	0.004
2016	Jan	0.994	0.032	0.023	1.026	-0.006
	May	0.984	0.04	0.026	1.074	0.004
	Nov	0.994	0.026	0.018	0.994	-0.003
2021	Jan	0.999	0.022	0.015	1.03	0.008
	May	0.998	0.022	0.016	1.049	0.009
	Nov	1	0.009	0.007	1.003	0.002

Similar observations are also evident during the pre-monsoon period. However, during this period, the fused maps retain more of MISR spatial patterns in 2016 and 2021, while resembling MODIS in 2012. Notably, the discrepancies seen near coastal regions in May (2021), particularly across the peninsular zone, may be attributed to higher cloud fractions (**Fig. S25**), introducing greater uncertainty in aerosol-cloud discrimination (Lang et al., 2026), thereby leading to inaccuracies in satellite-derived AOD estimate. Furthermore, AOD retrievals in coastal areas may also be influenced by potential overestimation of boundary layer height (BLH) in MODIS data (Wang et al., 2025a). The fused AOD, which primarily incorporates the information from the availability and spatial proximity of ground-based measurements, tends to show higher values in these regions, effectively correcting this bias using ground-based observations.

Table 5a: Error and bias analysis of MODIS AOD with ground and fused AOD at ground station locations.

Year	Month	MODIS & Ground					MODIS & Fused				
		Correlation	RMSE	MAE	Slope	Bias	Correlation	RMSE	MAE	Slope	Bias
2012	Jan	**0.781**	**0.144**	**0.106**	**0.858**	**0.074**	**0.789**	**0.142**	**0.105**	**0.896**	**0.067**
	May	0.770	0.128	**0.102**	0.704	0.147	0.788	0.122	0.096	0.732	0.137
	Nov	0.846	0.155	0.120	0.872	0.049	0.871	0.140	0.108	0.924	0.030
2016	Jan	0.792	0.250	0.164	1.096	-0.037	0.822	0.237	0.154	1.175	-0.083
	May	0.542	0.269	0.207	0.776	0.184	0.583	0.259	0.208	0.910	0.127
	Nov	0.773	0.157	0.121	0.759	0.064	0.787	0.153	0.115	0.774	0.055
2021	Jan	**0.861**	**0.199**	**0.138**	**0.911**	**0.010**	**0.879**	**0.185**	**0.127**	**0.958**	**-0.011**
	May	0.752	0.236	0.152	1.030	0.038	0.778	0.230	0.147	1.120	0.004
	Nov	**0.944**	**0.143**	**0.120**	**0.761**	**0.082**	**0.947**	**0.139**	**0.115**	**0.766**	**0.082**

Table 5b: Error and bias analysis of MISR AOD with ground and fused AOD at ground station locations.

Year	Month	MODIS & Ground					MODIS & Fused				
		Correlation	RMSE	MAE	Slope	Bias	Correlation	RMSE	MAE	Slope	Bias
2012	Jan	**0.884**	**0.115**	**0.090**	**0.676**	**0.049**	**0.897**	**0.106**	**0.082**	**0.704**	**0.042**
	May	0.643	0.159	**0.132**	0.437	0.189	0.656	0.154	0.129	0.452	0.184
	Nov	0.760	0.223	0.175	0.466	0.098	0.780	0.212	0.168	0.492	0.089
2016	Jan	0.783	0.270	0.206	0.421	0.010	0.818	0.263	0.203	0.454	0.081
	May	0.662	0.167	0.129	0.606	0.153	0.736	0.140	0.117	0.736	0.098
	Nov	0.814	0.186	0.143	0.515	0.079	0.834	0.185	0.144	0.528	0.072
2021	Jan	**0.824**	**0.346**	**0.267**	**0.372**	**0.131**	**0.838**	**0.332**	**0.258**	**0.390**	**0.123**
	May	0.772	0.205	0.160	0.458	0.147	0.805	0.185	0.147	0.502	0.130
	Nov	**0.944**	**0.251**	**0.186**	**0.492**	**0.092**	**0.948**	**0.248**	**0.182**	**0.495**	**0.091**

During the retreating monsoon season, the association of satellite and ground AOD is good, with similar spatial distribution over northern regions, while significant differences exists over southern parts. In the spatial map of fused AOD, the bias correction is clearly seen. In November 2016, over the NW and IGP, a significant bias correction in satellite AOD is observed, which is clearly indicated by the observed difference between fused AOD and satellite measurements. Similarly, a flow pattern on the east coast of peninsular India is corrected by fused AOD, which was otherwise underestimated by MISR (though retaining the spatial pattern). Here, the fused map showed enhanced AOD attributed to observations from ground stations, viz. Chennai (CHN) and Kadapa (KDP), which were underestimated by both MODIS and MISR. Similarly, in November 2021, the flow over the IGP is modified in the fused map based on ground observations

Over the IGP, where ground-based observations are more abundant than in other regions of India, the regional mean fused AOD generally lies between MODIS and MISR values. This reflects that the fusion process balances the biases between the tendency of MODIS to overestimate and MISR to underestimate AOD in this region. In contrast, over Peninsular India, which has the second-highest number of ground stations after the IGP, the fused AOD is higher than both MODIS and MISR, suggesting that satellite based underestimation is further adjusted in this region. Over northwest India during May, when dust loading is high, the fused AOD is closer to MISR, consistent with previous studies showing that MISR performs better than MODIS in dust-dominated regions due to its multi-angle capability. However, over the NE, CI, and NW regions, the fused AOD remains higher than satellite

estimates. Fused AOD estimates over the Himalayas and oceanic regions are not analyzed in detail due to the lack of sufficiently distributed ground-based observations. Overall, the fused AOD is constrained locally on ground-based AOD, which are generally considered more accurate than satellite-based observations whereas satellite retrievals exhibit discrepancies due to variations in aerosol types and their source contributions, which likely explain the observed differences in AOD estimates (Li et al., 2025; Wang et al., 2025b).

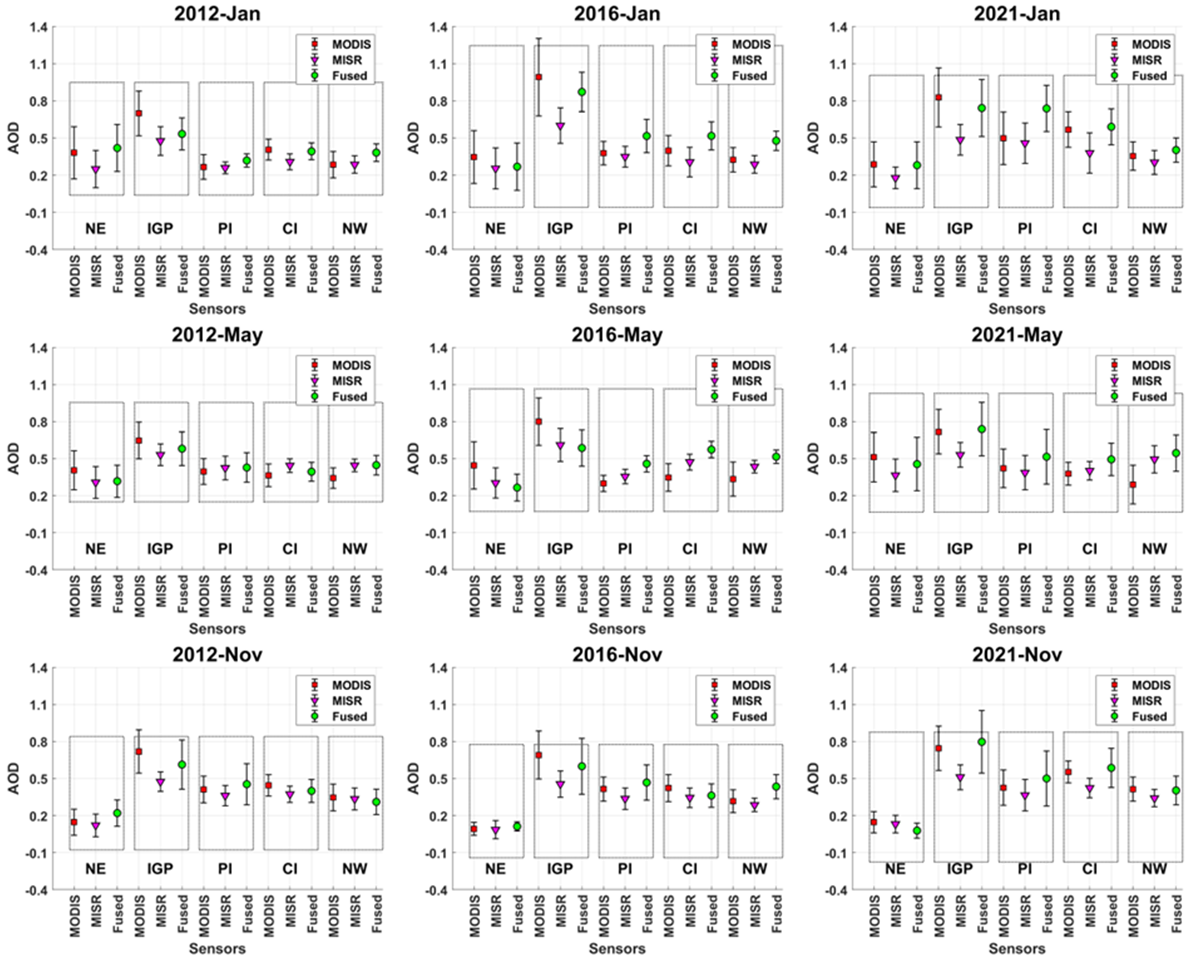


Figure 7. The regional mean values of MODIS, MISR, and fused AOD. On the y-axis, AOD values are shown as mean ± standard deviation.

With a view to assessing the improvement in the fused AOD estimates over MODIS and MISR, further analysis is carried out on regional scale considering different sub-regions over the study domain. **Fig. 7** shows the regional mean values of fused AOD along with AOD from individual satellite sensors. It is observed that the fused AOD significantly improves the biases in AOD from individual sensors. For example, in January 2012, MODIS AOD was significantly higher than fused AOD over IGP, while MISR AOD was closer to fused AOD over NE and CI regions. This indicates that the overestimation by MODIS in the IGP and the underestimation by MISR in the NE and CI regions are effectively corrected by the fusion framework. Similarly, in 2016, MODIS significantly overestimated AOD (~ 0.99 ± 0.31) in the IGP region. However, this was corrected in the fused AOD ~ 0.87 ± 0.16, which also closely matches ground-based AOD (~ 0.86). During the pre-monsoon period, MISR AOD during 2012 and 2016 is closer to fused AOD; whereas in 2021, MODIS is closer to the fused AOD over the NE, IGP, and PI regions than MISR, except over the CI and NW regions. This

behavior is consistent with the lower RMSE of MISR relative to ground-based AOD. Post-monsoon analysis reveals MODIS AOD is closer to fused AOD over IGP with an overestimation in 2012 and 2016, and an opposite pattern in 2021. MISR AOD is lower than fused AOD in all these periods. These observations clearly indicate the regional level bias corrections in the individual satellite sensors, resulting in a more accurate representation of aerosol features in terms of fused AOD.

3.3.4 Performance analysis and cross-validation

The performance analysis carried out using the correlation analysis parameters such as RMSE and MAE, revealed a decrease in RMSE and MAE in fused data and an increase in correlation with MODIS and MISR (**Figs S26-S28; Table 4, 5**). However, the improvement is inconsistent due to factors such as the variogram model used, errors between observation values of ground observation, and MODIS, MISR. Thus, the final characteristics of fused products are influenced by their individual instruments. The accuracy of fusion can be concluded from the cross-validation analysis. This is characterized by LOOCV method.

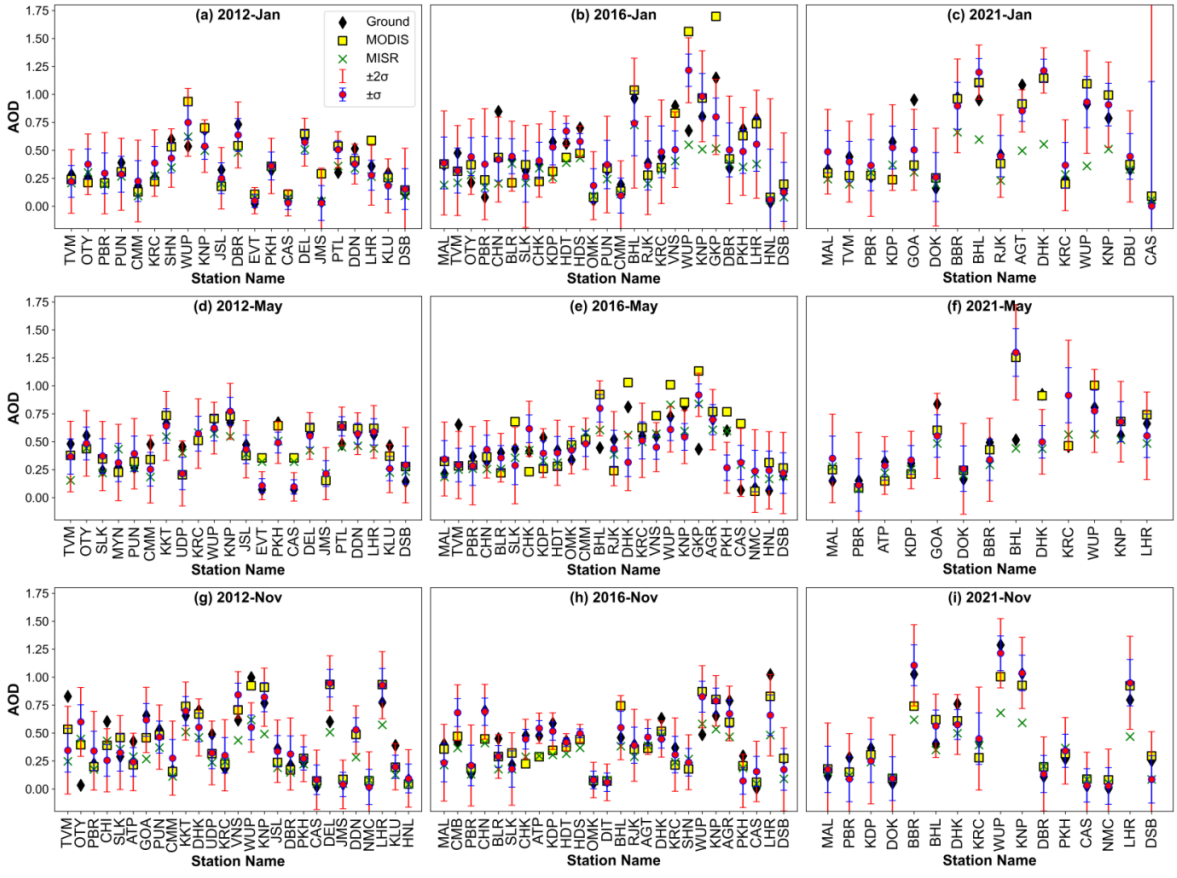


Figure 8. Predicted AOD values (as magenta points) with error bars $\pm\sigma$ (blue line), $\pm 2\sigma$ (red line), and ground observed AOD (black diamond), MODIS observation (yellow square), MISR (green cross) at different stations. For station names refer **Table ST7 and ST8**.

Figure 8 shows the predicted AOD values at each ground location during each leave-one-out process. The prediction model performances analyzed in terms of mean prediction error (MPE) and root mean square prediction error (RMSPE) are given in **Table 6**. The predicted AOD values (as magenta points) at each of the ground locations with standard error bars ± σ (blue line), ± 2σ (grey line) are also shown in **Fig. 8**, along with AOD from the ground (black diamond), MODIS (yellow triangle), MISR (red cross) observations. The figure shows that

more than 80% of ground AOD are within ± 2σ (95% Confidence interval) of predicted AOD for seven out of nine months. The highest accuracy was achieved in 2021 November and 2012 May (100%), and the lowest in 2016 May (76%), indicating the importance of the association between different sensors during the fusion process. The enhanced accuracy of the model for fused estimations required good correlation and reduced errors, as indicated in **Tables 4, 5 & 6**.

Table 6: Accuracy assessment of the predicted AOD through leave-one-out cross-validation. Here % describes how many ground AOD (actual values) are covered within the range of predicted AOD.

Year	Month	Correlation	RMSPE	MPE	Ground AOD within $\pm\sigma$ (in %)	Ground AOD within $\pm 2\sigma$ (in %)
2012	Jan	**0.852**	**0.108**	**0.091**	**76.19**	**95.24**
	May	**0.809**	**0.118**	**0.093**	**77.27**	**100**
	Nov	0.676	0.215	0.152	59.26	77.78
2016	Jan	0.708	0.213	0.160	65.38	84.62
	May	0.367	0.219	0.170	56	76
	Nov	0.793	0.146	0.109	65.38	88.46
2021	Jan	**0.893**	**0.162**	**0.121**	**81.25**	**81.25**
	May	0.451	0.302	0.212	61.54	84.62
	Nov	**0.964**	**0.108**	**0.086**	**87.50**	**100**

3.3.5 Machine Learning enhanced Geostatistical data fusion

To understand the influence of number of ground measurement points in the generation of fused map, sensitivity studies has been carried out by varying the number of ground based measurement points. The number of ground points from maximum of 27 ground locations has been subsequently reduced to 22, 13, 8, and 6 respectively. The corresponding variations in the fused outputs are provided (Supplementary **Fig. S29-S30, Table ST9**) and a special case is included in **Fig. 9**. The figure clearly explains the changes in prevailing spatial pattern of aerosols according to changes in number of data points, indicating that RK-ML method is a good alternative to UK, when the available ground measurements are limited. This study also indicates an inherent limitation in UK method, alike to multiple linear regression models, which are highly susceptible to noise in predictor variables. In contrast, RK-ML demonstrates greater robustness by first modeling the deterministic component using a machine learning regressor (in this case, SVR), followed by kriging of the residuals to capture the stochastic component. This two-step approach effectively leverages machine learning for optimized estimation under noisy conditions, while kriging incorporates spatial variability of residuals obtained from observations, resulting in more reliable spatial predictions.

Notably, when data points are fewer, UK has overestimated AOD values in mainland regions relative to RK-ML predictions. This discrepancy may arise due to the complexity of the underlying surface, where OK has been shown to outperform UK, as discussed on basis of different surface types (Zimmerman et al., 1999). Since RK is an extension of OK, it inherits these advantages, contributing to the improved performance of the RK-ML method. The sensitivity study further highlights that variations in the number of ground station data points and associated errors have a lesser impact on fused AOD from RK-ML approach in terms of retaining spatial patterns compared to the UK-approach. This demonstrates the RK-ML fusion ability to better manage observational uncertainties and data sparsely.

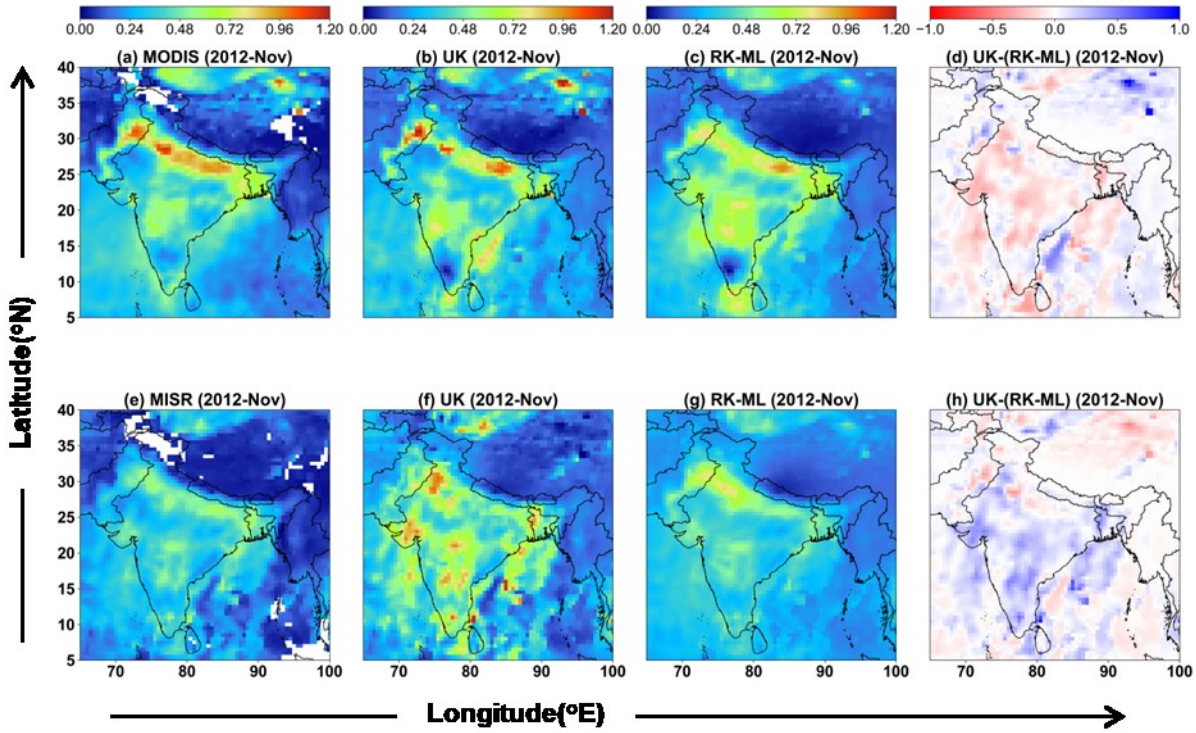


Figure 9. MODIS (a), MISR (e), and fused AOD using Universal Kriging (UK, (b), (f)), and Residual Kriging with Machine Learning (RK-ML, (c), (g)); Difference between UK and RK-ML predictions ((d), (h)). Blue color in difference map indicates where UK predictions exceed those of RK-ML, and vice versa for red. The top panels (b, c and d) of fused AOD shows estimations using 27 ground-data points (Case - 1), while the bottom panel (f, g and h) shows fused AOD derived from 6 ground-data points (Case - 2).

The UK and RK-ML approach yields comparable results when the observations are higher in number. On the other hand, RK-ML outperforms UK when the number of ground observation is significantly decreased (**Fig. 10**). Hence, the accuracy of this estimation is significantly influenced by the availability of small datasets, which indicates that the large number of ground stations are crucial for best representation of fused AOD products from Universal Kriging, while RK-ML is a good choice in case of limited number of ground observations. Hence, the applicability of RK-ML can be made for generating fused AOD on daily basis per each satellite overpasses over ground based stations.

The applicability of UK and RK-ML depends on their ability to capture realistic spatial patterns from spatially representative data points (e.g., ground-based AOD) in order to produce accurate fused AOD distributions. These estimations inherently carry uncertainties, which increase with distance from ground observation sites. Consequently, the resulting maps show elevated uncertainties in areas with sparse ground coverage. To evaluate model performance under varying ground-station availability, we employed LOOCV. Results indicate that RK-ML is generally more effective than UK when fewer ground observations are available, as it can capture nonlinear relationships and penalize erroneous AOD values. However, the Kriging step still requires a sufficient number of data points to propagate spatial residuals. Similarly, UK applied with a limited number of stations (e.g., 27 in the present study) estimates weights from collocated AOD points and satellite regressors (MODIS and MISR). Thus, accurate uncertainty representation and improved fused AOD maps require a more uniformly distributed network of ground stations. As illustrated in **Fig. S29**, increasing

station density enhances the similarity between fusion maps from both methods, as seen in the larger white patches of the deviation plots (**Fig. S29 a-d**).

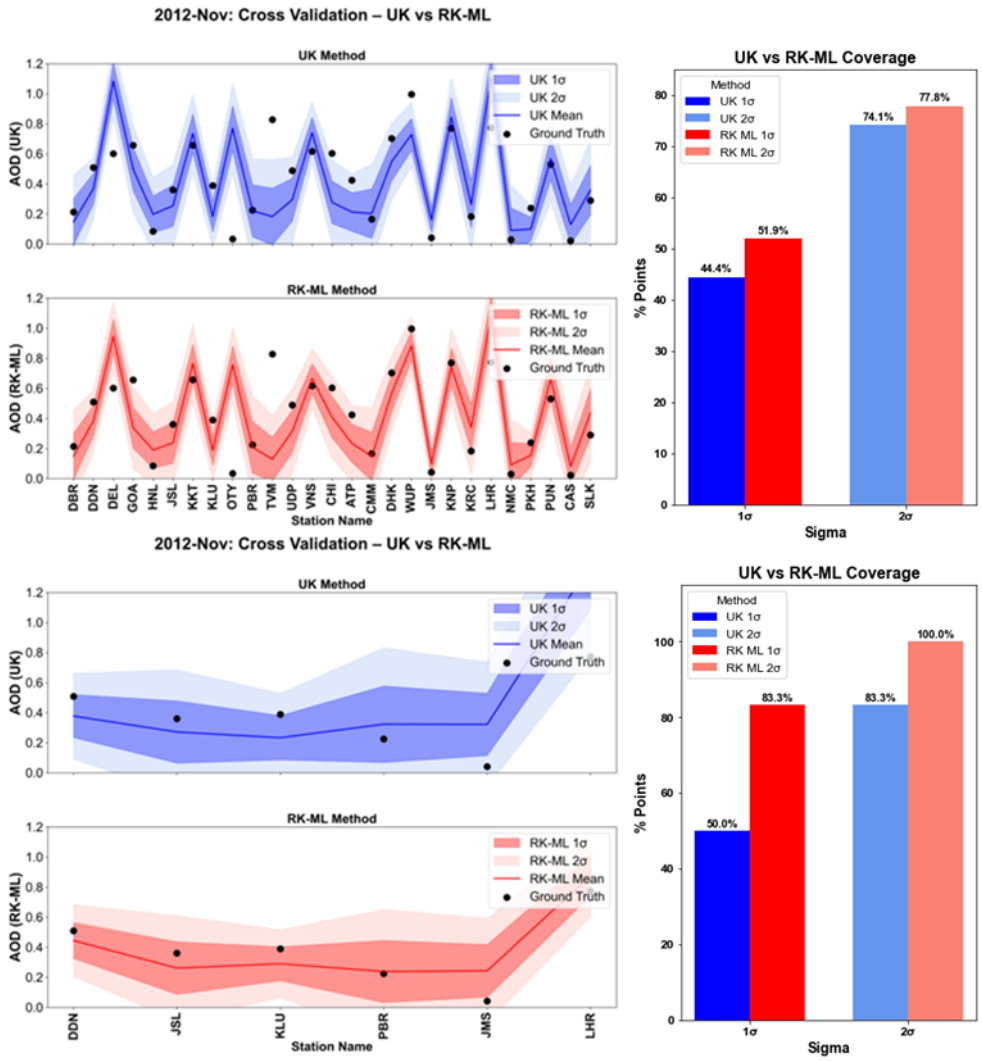


Figure 10: Line plots of LOOCV results from UK method (blue line) and RK-ML method (red line), covering the ground AOD (black dots) for Case - 1 (27 points, 1st and 2nd line plots) and case - 2 (6 points, 3rd and 4th line plots) within 1σ (dark shade) and 2σ (light shade). For station names and details, refer to supplementary **Table ST8 and ST9**.

4. Conclusion

The utilization of a universal Kriging approach, combining satellite-based measurements with ground-based observations, has demonstrated enhanced AOD estimation with reduced uncertainties compared to relying on a single instrument. Despite inherent differences among instruments, the implemented approach capitalizes on their complementary features, statistically combining the three datasets to provide robust estimates. The significant outcomes of this study are as follows:

- MODIS and MISR observations exhibit good but variable associations with ground-based AOD measurements influenced by seasonal and geographic differences.
- Variogram analysis reveals different autocorrelation length implying capability of each sensor to get the spatial variability or auto correlation structure in different periods. In some of the months, MODIS shows higher spatial range as compared to

MISR, while the opposite is seen during the rest of the months. On the other hand, sill is always higher in case of MODIS.

- Spatial interpolation of AOD through variogram analysis provides very good predictions at the missing grids of the satellite observations, emphasizing the effectiveness of the universal Kriging method.
- The fused AOD maps reveal distinct results, highlighting the significant impact of ground-based AOD on the fusion process. During the pre-monsoon period specifically, the correlation coefficient and slope between MISR and fused AOD, which is comparable to ground-based AOD at point locations, improved by ~11% and ~21%, respectively. Furthermore, the RMSE was reduced by ~16% compared to pre-fusion values. These results demonstrate a marked improvement in the spatial representation of ground AOD following the fusion process. Additionally, the near-perfect correlation (R~0.99) between fused and ground-based AOD suggests effective bias correction in MODIS or MISR datasets, which otherwise significantly overestimated or underestimated aerosol measurements across certain parts of the study region.
- Cross-validation experiments further underline the effectiveness of the models in this study, e.g., strong correlations (0.964) and low RMSPE (0.108) and MPE (0.086) errors during November 2021. Even the model is able to predict 87.50 % of ground truths within ±σ and even 100% within ±2σ. Moreover the measurements which were failed in predictions are due to highly local heterogeneity influenced by nearby ground measurements only.
- Incorporating a greater number of ground-based measurements enhances the fused results, yielding a cross-validation accuracy ranging from approximately 78% to 88% (based on 27 ground location points in November 2012 and 26 points in November 2016). However, the alternative RK-ML method can also be effective when long-term observation stations are available, even if their numbers are limited. The establishment of additional ground-based stations is recommended to strengthen the representation of air quality, especially in regions with high heterogeneity. This methodology can be implemented to get the fusion maps of finer spatiotemporal resolution.

Author Contributions Statement:

SSG - Data Curation, Software, Formal analysis, Visualization, Investigation, Writing - Original Draft and Editing; **MMG** - Methodology, Visualization, Validation, Software, Writing, Review and Editing, Supervision; **SSB** - Conceptualization, Supervision, Review and Editing, Project administration.

Data availability: ARFINET data used for this study are available upon request from Surendran Nair Suresh Babu (s_sureshbabu@vssc.gov.in). All other datasets used in this paper are open access data, and can be freely downloaded from the websites listed in the acknowledgements.

Competing interests: The contact author has declared that none of the authors has any competing interests.

Acknowledgements:

This study was carried out as part of the ARFI project of ISRO-GBP. We express our sincere thanks to the ARFINET investigators for the continued support and long-term contributions over the years in operating the network. We authors are also thankful to the AERONET team for providing AOD data (data available at http://aeronet.gsfc.nasa.gov). Additionally, we acknowledge NASA's Level-1 and Atmosphere Archive and Distribution System Distributed Active Archive Center (LAADS DAAC; data available at https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/61/MOD04_L2/) and the Atmospheric Science Data Center (ASDC; data available at https://l0dup05.larc.nasa.gov/cgi-bin/MISR/main.cgi) for making the MODIS and MISR datasets available.

References:

Babu, S. S., Krishna Moorthy, K., and Satheesh, S. K.: Temporal heterogeneity in aerosol characteristics and the resulting radiative impacts at a tropical coastal station – Part 2: Direct short wave radiative forcing, Ann. Geophys., 25, 2309–2320, https://doi.org/10.5194/angeo-25-2309-2007, 2007.

Bai, K., Li, K., Shao, L., Li, X., Liu, C., Li, Z., and Ma, M.: LGHAP v2 : a global gap-free aerosol optical depth and PM 2 . 5 concentration dataset since 2000 derived via big Earth data analytics, 2425–2448, 2024.

Baisad, K., Chutsagulprom, N., and Moonchai, S.: A Non-Linear Trend Function for Kriging with External Drift Using Least Squares Support Vector Regression, Mathematics, 11, 4799, https://doi.org/10.3390/math11234799, 2023.

Basart, S., Pérez, C., Cuevas, E., Baldasano, J. M., and Gobbi, G. P.: Aerosol characterization in Northern Africa, Northeastern Atlantic, mediterranean basin and middle east from direct-sun AERONET observations, Atmos. Chem. Phys., 9, 8265–8282, https://doi.org/10.5194/acp-9-8265-2009, 2009.

Brereton, R. G. and Lloyd, G. R.: Support Vector Machines for classification and regression, Analyst, 135, 230–267, https://doi.org/10.1039/b918972f, 2010.

Chatterjee, A., Michalak, A. M., Kahn, R. A., Paradise, S. R., Braverman, A. J., and Miller, C. E.: A geostatistical data fusion technique for merging remote sensing and ground-based observations of aerosol optical thickness, J. Geophys. Res. Atmos., 115, 1–12, https://doi.org/10.1029/2009JD013765, 2010.

Chen, Z.-Y., Jin, J.-Q., Zhang, R., Zhang, T.-H., Chen, J.-J., Yang, J., Ou, C.-Q., and Guo, Y.: Comparison of Different Missing-Imputation Methods for MAIAC (Multiangle Implementation of Atmospheric Correction) AOD in Estimating Daily PM2.5 Levels, Remote Sens., 12, https://doi.org/10.3390/rs12183008, 2020.

Chu, D. A., Kaufman, Y. J., Ichoku, C., Remer, L. A., Tanré, D., and Holben, B. N.: Validation of MODIS aerosol optical depth retrieval over land, Geophys. Res. Lett., 29, MOD2-1-MOD2-4, https://doi.org/10.1029/2001GL013205, 2002.

Chua, S. H. and Bras, R. L.: Optimal estimators of mean areal precipitation in regions of orographic influence, J. Hydrol., 57, 23–48, https://doi.org/10.1016/0022-1694(82)90101-9, 1982.

Derakhshan, H. and Leuangthong, O.: Impact of Data Spacing on Variogram Uncertainty, 1–19, 1982.

Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne, S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, https://doi.org/10.1029/1999JD900923, 1999.

Edward H. Isaaks, R. M. S.: An Introduction to Applied Geostatistics, Comput. Geosci., 17, 471–473, https://doi.org/10.1016/0098-3004(91)90055-I, 1991.

Farahat, A.: Comparative analysis of MODIS, MISR, and AERONET climatology over the

Middle East and North Africa, Ann. Geophys., 37, 49–64, https://doi.org/10.5194/angeo-37-49-2019, 2019.
Filonchyk, M., Yan, H., Zhang, Z., Yang, S., Li, W., and Li, Y.: Combined use of satellite and surface observations to study aerosol optical depth in different regions of China, Sci. Rep., 9, 1–15, https://doi.org/10.1038/s41598-019-42466-6, 2019.
Freier, L. and Von Lieres, E.: Kriging based iterative parameter estimation procedure for biotechnology applications with nonlinear trend functions, IFAC-PapersOnLine, 28, 574–579, https://doi.org/10.1016/j.ifacol.2015.05.043, 2015.
Freier, L., Wiechert, W., and von Lieres, E.: Kriging with trend functions nonlinear in their parameters: Theory and application in enzyme kinetics, Eng. Life Sci., 17, 916–922, https://doi.org/10.1002/elsc.201700022, 2017.
Gao, L., Chen, L., Li, C., Li, J., Che, H., and Zhang, Y.: Evaluation and possible uncertainty source analysis of JAXA Himawari-8 aerosol optical depth product over China, Atmos. Res., 248, 105248, https://doi.org/10.1016/j.atmosres.2020.105248, 2021.
Garay, M. J., Kalashnikova, O. V., and Bull, M. A.: Development and assessment of a higher-spatial-resolution (4.4 km) MISR aerosol optical depth product using AERONET-DRAGON data, Atmos. Chem. Phys., 17, 5095–5106, https://doi.org/10.5194/acp-17-5095-2017, 2017.
Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database - Automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, Atmos. Meas. Tech., 12, 169–209, https://doi.org/10.5194/amt-12-169-2019, 2019.
Gogoi, M. M., Krishna Moorthy, K., Suresh Babu, S., and Bhuyan, P. K.: Climatology of columnar aerosol properties and the influence of synoptic conditions: First-time results from the northeastern region of India, J. Geophys. Res. Atmos., 114, https://doi.org/10.1029/2008JD010765, 2009.
Guo, J., Gu, X., Yu, T., Cheng, T., Chen, H., and Xie, D.: Trend analysis of the aerosol optical depth over China using fusion of MODIS and MISR aerosol products via adaptive weighted estimate algorithm, Earth Obs. Syst. XVIII, 8866, 88661X, https://doi.org/10.1117/12.2024687, 2013.
Gupta, P., Patadia, F., and Christopher, S. A.: Multisensor data product fusion for aerosol research, IEEE Trans. Geosci. Remote Sens., 46, 1407–1415, https://doi.org/10.1109/TGRS.2008.916087, 2008.
Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET - A federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, https://doi.org/10.1016/S0034-4257(98)00031-5, 1998.
Holdaway, M. R.: Spatial modeling and interpolation of monthly temperature using kriging, Clim. Res., 6, 215–225, https://doi.org/10.3354/cr006215, 1996.
Huang, J., Patrick Arnott, W., Barnard, J. C., and Holmes, H. A.: Theoretical uncertainty analysis of satellite retrieved aerosol optical depth associated with surface albedo and aerosol optical properties, Remote Sens., 13, 1–21, https://doi.org/10.3390/rs13030344, 2021.
Ichoku, C., Allen Chu, D., Mattoo, S., Kaufman, Y. J., Remer, L. A., Tanré, D., Slutsker, I., and Holben, B. N.: A spatio-temporal approach for global validation and analysis of MODIS aerosol products, Geophys. Res. Lett., 29, MOD1-1-MOD1-4, https://doi.org/10.1029/2001GL013206, 2002.
Jiang, X., Liu, Y., Yu, B., and Jiang, M.: Comparison of MISR aerosol optical thickness with AERONET measurements in Beijing metropolitan area, Remote Sens. Environ., 107, 45–53, https://doi.org/10.1016/j.rse.2006.06.022, 2007.
Jiao, Y., Zhang, M., Wang, L., and Qin, W.: A New Cloud and Haze Mask Algorithm From

Radiative Transfer Simulations Coupled With, IEEE Trans. Geosci. Remote Sens., 61, 1–16, https://doi.org/10.1109/TGRS.2023.3252264, 2023.
Jin, S., Ma, Y., Li, H., Liu, B., Fan, R., Zhang, M., Lopatin, A., Dubovik, O., Hu, X., Gong, W., and Wang, L.: Characterizing Aerosol Optical Properties and Direct Radiative Effects From the Perspective of Components : A Synergy Retrieval Study Based on Sun Photometer and Lidar in Central China, https://doi.org/10.1029/2024GL113448, 2025.
Jinnagara Puttaswamy, S., Nguyen, H. M., Braverman, A., Hu, X., and Liu, Y.: Statistical data fusion of multi-sensor AOD over the Continental United States, Geocarto Int., 29, 48–64, https://doi.org/10.1080/10106049.2013.827750, 2014.
Kadow, C., Hall, D. M., and Ulbrich, U.: Artificial intelligence reconstructs missing climate information, Nat. Geosci., 13, 408–413, https://doi.org/10.1038/s41561-020-0582-5, 2020.
Kahn, R., Banerjee, P., and McDonald, D.: Sensitivity of multiangle imaging to natural mixtures of aerosols over ocean, J. Geophys. Res. Atmos., 106, 18219–18238, https://doi.org/10.1029/2000JD900497, 2001.
Kahn, R., Andrews, E., Brock, C. A., Chin, M., Feingold, G., Gettelman, A., Levy, R. C., Murphy, D. M., Nenes, A., Pierce, J. R., Popp, T., Redemann, J., Sayer, A. M., da Silva, A. M., Sogacheva, L., and Stier, P.: Reducing Aerosol Forcing Uncertainty by Combining Models With Satellite and Within-The-Atmosphere Observations: A Three-Way Street, Rev. Geophys., 61, 1–27, https://doi.org/10.1029/2022rg000796, 2023.
Kalluri, R. O. R., Gugamsetty, B., Kotalo, R. G., Nagireddy, S. K. R., Tandule, C. R., Thotli, L. R., Rajuru Ramakrishna, R., and Surendranair, S. B.: Direct radiative forcing properties of atmospheric aerosols over semi-arid region, Anantapur in India, Sci. Total Environ., 566–567, 1002–1013, https://doi.org/10.1016/j.scitotenv.2016.05.056, 2016.
Kim, M., Kim, J., Lim, H., Lee, S., Cho, Y., Lee, Y., Go, S., and Lee, K.: Aerosol optical depth data fusion with Geostationary Korea Multi-Purpose Satellite (GEO-KOMPSAT-2) instruments GEMS , AMI , and GOCI-II : statistical and deep neural network methods, 4317–4335, 2024.
Kinne, S.: Remote sensing data combinations: superior global maps for aerosol optical depth, Satell. Aerosol Remote Sens. over L., 361–381, https://doi.org/10.1007/978-3-540-69397-0_12, 2009.
Kompalli, S. K., Suresh Babu, S., and Krishna Moorthy, K.: Inter-comparison of aerosol optical depth from the Multi-Wavelength Solar Radiometer with other radiometric measurements, Indian J. Radio Sp. Phys., 39, 364–371, 2010.
Koushavand, B. and Deutsch, C. V: A Methodology to Quantify and Transfer Variogram Uncertainty through Kriging and Simulation, Cent. Comput. Geostatistics, 1–16, 2008.
Kumar, K. R., Narasimhulu, K., Reddy, R. R., Gopal, K. R., Reddy, L. S. S., Balakrishnaiah, G., Moorthy, K. K., and Babu, S. S.: Temporal and spectral characteristics of aerosol optical depths in a semi-arid region of southern India, Sci. Total Environ., 407, 2673–2688, https://doi.org/10.1016/j.scitotenv.2008.10.028, 2009.
Kumar, S., Kumar, S., Singh, A. K., and Singh, R. P.: Seasonal variability of atmospheric aerosol over the North Indian region during 2005-2009, Adv. Sp. Res., 50, 1220–1230, https://doi.org/10.1016/j.asr.2012.06.022, 2012.
Lang, Q., Zhang, M., He, Q., Jin, S., Qin, W., Luo, L., and Wang, L.: Significant uncertainties from overlooking aerosol-cloud coexistence in surface solar radiation estimates using passive satellite observations, Remote Sens. Environ., 333, 115168, https://doi.org/https://doi.org/10.1016/j.rse.2025.115168, 2026.
Levy, R. C., Remer, L. A., Martins, J. V, Kaufman, Y. J., Plana-Fattori, A., Redemann, J., and Wenny, B.: Evaluation of the MODIS Aerosol Retrievals over Ocean and Land during CLAMS, 2005.
Li, H., Zhang, M., Wang, L., Su, X., and Lu, Y.: Effects of Different Types of Aerosols on Diffuse Radiation Based on Global AERONET, J. Geophys. Res. Atmos., 130,

e2024JD042701, https://doi.org/https://doi.org/10.1029/2024JD042701, 2025.
Li, J., Kahn, R. A., Wei, J., Carlson, B. E., Lacis, A. A., Li, Z., Li, X., Dubovik, O., and Nakajima, T.: Synergy of Satellite- and Ground-Based Aerosol Optical Depth Measurements Using an Ensemble Kalman Filter Approach, J. Geophys. Res. Atmos., 125, 1–17, https://doi.org/10.1029/2019JD031884, 2020.
Lilla, R. and Castrignanò, A.: Science of the Total Environment A geostatistical approach for multi-source data fusion to predict water table depth Warning : Editor name is, Sci. Total Environ., 696, 133763, https://doi.org/10.1016/j.scitotenv.2019.133763, 2019.
Litvinov, P., Chen, C., Dubovik, O., Zhai, S., Matar, C., Li, C., Lopatin, A., Fuertes, D., Lapyonok, T., Bindreiter, L., Dornacher, M., Lehner, A., Dandocsi, A., Gasbarra, D., and Retscher, C.: Synergetic retrieval from multi-mission spaceborne measurements for enhanced aerosol and surface characterization, 7679–7716, 2025.
Liu, Y., Sarnat, J. A., Coull, B. A., Koutrakis, P., and Jacob, D. J.: Validation of Multiangle Imaging Spectroradiometer (MISR) aerosol optical thickness measurements using Aerosol Robotic Network (AERONET) observations over the contiguous United States, J. Geophys. Res. Atmos., 109, https://doi.org/10.1029/2003jd003981, 2004.
Lodhi, N. K., Beegum, S. N., Singh, S., and Kumar, K.: Aerosol climatology at Delhi in the western Indo-Gangetic Plain: Microphysics, long-term trends, and source strengths, J. Geophys. Res. Atmos., 118, 1361–1375, https://doi.org/10.1002/jgrd.50165, 2013.
Moorthy, K. K., Babu, S. S., Satheesh, S. K., Srinivasan, J., and Dutt, C. B. S.: Dust absorption over the "Great Indian Desert" inferred using ground-based and satellite remote sensing, J. Geophys. Res. Atmos., 112, 1–10, https://doi.org/10.1029/2006JD007690, 2007.
Moorthy K., K., Suresh Babu, S., and Satheesh, S. K.: Temporal heterogeneity in aerosol characteristics and the resulting radiative impact at a tropical coastal station – Part 1: Microphysical and optical properties, Ann. Geophys., 25, 2293–2308, https://doi.org/10.5194/angeo-25-2293-2007, 2007.
Nalder, I. A. and Wein, R. W.: Spatial interpolation of climatic Normals: Test of a new method in the Canadian boreal forest, Agric. For. Meteorol., 92, 211–225, https://doi.org/10.1016/S0168-1923(98)00102-6, 1998.
Nguyen, H., Cressie, N., and Braverman, A.: Spatial statistical data fusion for remote sensing applications, J. Am. Stat. Assoc., 107, 1004–1018, https://doi.org/10.1080/01621459.2012.694717, 2012.
Nirala, M.: Multi-sensor data fusion of aerosol optical thickness, Int. J. Remote Sens., 29, 2127–2136, https://doi.org/10.1080/01431160701395336, 2008.
Pathak, H. S., Satheesh, S. K., Nanjundiah, R. S., Moorthy, K. K., Lakshmivarahan, S., and Babu, S. N. S.: Assessment of regional aerosol radiative effects under the SWAAMI campaign - Part 1: Quality-enhanced estimation of columnar aerosol extinction and absorption over the Indian subcontinent, Atmos. Chem. Phys., 19, 11865–11886, https://doi.org/10.5194/acp-19-11865-2019, 2019.
Prasad, A. K. and Singh, R. P.: Comparison of MISR-MODIS aerosol optical depth over the Indo-Gangetic basin during the winter and summer seasons (2000-2005), Remote Sens. Environ., 107, 109–119, https://doi.org/10.1016/j.rse.2006.09.026, 2007.
Rossi, R. E., Dungan, J. L., and Beck, L. R.: Kriging in the shadows: Geostatistical interpolation for remote sensing, Remote Sens. Environ., 49, 32–40, https://doi.org/10.1016/0034-4257(94)90057-4, 1994.
Sayer, A. M., Hsu, N. C., Bettenhausen, C., and Jeong, M. J.: Validation and uncertainty estimates for MODIS Collection 6 "deep Blue" aerosol data, J. Geophys. Res. Atmos., 118, 7864–7872, https://doi.org/10.1002/jgrd.50600, 2013.
Sayer, A. M., Munchak, L. A., Hsu, N. C., Levy, R. C., Bettenhausen, C., and Jeong, M.-J.: MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and "merged" data sets, and usage recommendations, J. Geophys. Res. Atmos., 119,

13,913-965,989, https://doi.org/https://doi.org/10.1002/2014JD022453, 2014.
Sayer, A. M., Govaerts, Y., Kolmonen, P., Lipponen, A., Luffarelli, M., Mielonen, T., Patadia, F., Popp, T., Povey, A. C., Stebel, K., and Witek, M. L.: A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing, Atmos. Meas. Tech., 13, 373–404, https://doi.org/10.5194/amt-13-373-2020, 2020.
Schutgens, N., Sayer, A. M., Heckel, A., Hsu, C., Jethva, H., De Leeuw, G., Leonard, P. J. T., Levy, R. C., Lipponen, A., Lyapustin, A., North, P., Popp, T., Poulsen, C., Sawyer, V., Sogacheva, L., Thomas, G., Torres, O., Wang, Y., Kinne, S., Schulz, M., Stier, P., and Schutgens, N.: An AeroCom-AeroSat study: Intercomparison of satellite AOD datasets for aerosol model evaluation, Atmos. Chem. Phys., 20, 12431–12457, https://doi.org/10.5194/acp-20-12431-2020, 2020.
Sharma, V., Ghosh, S., Bilal, M., Dey, S., and Singh, S.: Performance of MODIS C6 . 1 Dark Target and Deep Blue aerosol products in Delhi National Capital Region , India : Application for aerosol studies, Atmos. Pollut. Res., 12, 65–74, https://doi.org/10.1016/j.apr.2021.01.023, 2021.
Shaw, G. E., Reagan, J. A., and Herman, B. M.: Investigations of Atmospheric Extinction Using Direct Solar Radiation Measurements Made with a Multiple Wavelength Radiometer, J. Appl. Meteorol., 12, 374–380, https://doi.org/10.1175/1520-0450(1973)012<0374:IOAEUD>2.0.CO;2, 1973.
Shi, Y., Li, L., and Zhang, L.: Application and comparing of IDW and Kriging interpolation in spatial rainfall information, in: Geoinformatics 2007: Geospatial Information Science, 67531I, https://doi.org/10.1117/12.761859, 2007.
Sifaou, H., Kammoun, A., and Alouini, M. S.: A Precise Performance Analysis of Support Vector Regression, Proc. Mach. Learn. Res., 139, 9671–9680, 2021.
Singh, M. K. and Venkatachalam, P.: Merging of aerosol optical depth data from multiple remote sensing sensors, in: 2014 IEEE Geoscience and Remote Sensing Symposium, 4173–4175, https://doi.org/10.1109/IGARSS.2014.6947407, 2014.
Singh, M. K., Gautam, R., and Venkatachalam, P.: A merged aerosol dataset based on MODIS and MISR Aerosol Optical Depth products, Remote Sens. Atmos. Clouds, Precip. VI, 9876, 987627, https://doi.org/10.1117/12.2223485, 2016.
Singh, P. and Verma, P.: A comparative study of spatial interpolation technique (IDW and Kriging) for determining groundwater quality, in: GIS and Geostatistical Techniques for Groundwater Science, Elsevier, 43–56, https://doi.org/10.1016/B978-0-12-815413-7.00005-5, 2019.
Singh, P., Vaishya, A., Rastogi, S., and Babu, S. S.: Seasonal heterogeneity in aerosol optical properties over the subtropical humid region of northern India, J. Atmos. Solar-Terrestrial Phys., 201, https://doi.org/10.1016/j.jastp.2020.105246, 2020.
Sinha, P. R., Dumka, U. C., Manchanda, R. K., Kaskaoutis, D. G., Sreenivasan, S., Krishna Moorthy, K., and Suresh Babu, S.: Contrasting aerosol characteristics and radiative forcing over Hyderabad, India due to seasonal mesoscale and synoptic-scale processes, Q. J. R. Meteorol. Soc., 139, 434–450, https://doi.org/10.1002/qj.1963, 2013.
Sivadasan Nair, V., Giorgi, F., and Keshav Hasyagar, U.: Amplification of South Asian haze by water vapour-aerosol interactions, Atmos. Chem. Phys., 20, 14457–14471, https://doi.org/10.5194/acp-20-14457-2020, 2020.
Smola, A. and Schölkopf, B.: A tutorial on support vector regression, Stat. Comput., 14, 199–222, https://doi.org/10.1023/B%3ASTCO.0000035301.49549.88, 2004.
Snepvangers, J. J. J. C., Heuvelink, G. B. M., and Huisman, J. A.: Soil water content interpolation using spatio-temporal kriging with external drift, Geoderma, 112, 253–271, https://doi.org/10.1016/S0016-7061(02)00310-5, 2003.
Sogacheva, L., Popp, T., Sayer, A. M., Dubovik, O., Garay, M. J., Heckel, A., Christina Hsu, N., Jethva, H., Kahn, R. A., Kolmonen, P., Kosmale, M., De Leeuw, G., Levy, R. C.,

Litvinov, P., Lyapustin, A., North, P., Torres, O., and Arola, A.: Merging regional and global aerosol optical depth records from major available satellite products, Atmos. Chem. Phys., 20, 2031–2056, https://doi.org/10.5194/acp-20-2031-2020, 2020.
Stein, A. and Corsten, L. C. A.: Universal Kriging and Cokriging as a Regression Procedure, Biometrics, 47, 575, https://doi.org/10.2307/2532147, 1991.
Subba, T., Gogoi, M. M., Moorthy, K. K., Bhuyan, P. K., Pathak, B., Guha, A., Srivastava, M. K., Vyas, B. M., Singh, K., Krishnan, J., Lakshmikumar, T. V. S., and Babu, S. S.: Aerosol Radiative Effects Over India from Direct Radiation Measurements and Model Estimates, SSRN Electron. J., https://doi.org/10.2139/ssrn.3986898, 2021.
Tandule, C. R., Gogoi, M. M., Gouda, S. S., and Suresh Babu, S.: Retrieval of aerosol optical depth from INSAT-3DR for accurate geostationary monitoring of regional and temporal aerosol dynamics, Atmos. Environ., 367, 121730, https://doi.org/https://doi.org/10.1016/j.atmosenv.2025.121730, 2026.
Tang, Q., Bo, Y., and Zhu, Y.: Spatiotemporal fusion of multiple-satellite aerosol optical depth (AOD) products using bayesian maximum entropy method, J. Geophys. Res., 121, 4034–4048, https://doi.org/10.1002/2015JD024571, 2016.
Tao, M., Wang, J., Li, R., Chen, L., Xu, X., Wang, L., Tao, J., Wang, Z., and Xiang, J.: Characterization of Aerosol Type Over East Asia by 4.4 km MISR Product: First Insight and General Performance, J. Geophys. Res. Atmos., 125, https://doi.org/10.1029/2019JD031909, 2020.
Tao, M., Chen, J., Xu, X., Man, W., Xu, L., Wang, L., Wang, Y., Wang, J., Fan, M., Shahzad, M. I., and Chen, L.: A robust and flexible satellite aerosol retrieval algorithm for multi-angle polarimetric measurements with physics-informed deep learning method, Remote Sens. Environ., 297, 113763, https://doi.org/https://doi.org/10.1016/j.rse.2023.113763, 2023.
Tian, X. and Gao, Z.: Validation and Accuracy Assessment of MODIS C6.1 Aerosol Products over the Heavy Aerosol Loading Area, Atmosphere (Basel)., 10, https://doi.org/10.3390/atmos10090548, 2019.
Tian, X., Liu, Q., Li, X., and Wei, J.: Validation and comparison of MODIS C6.1 and C6 aerosol products over Beijing, China, Remote Sens., 10, https://doi.org/10.3390/rs10122021, 2018.
Tiwari, S., Kaskaoutis, D., Soni, V. K., Dev Attri, S., and Singh, A. K.: Aerosol columnar characteristics and their heterogeneous nature over Varanasi, in the central Ganges valley, Environ. Sci. Pollut. Res., 25, 24726–24745, https://doi.org/10.1007/s11356-018-2502-4, 2018.
Vieira, S., Carvalho, J., Ceddia, M., and Paz-González, A.: Detrending non stationary data for geostatistical applications, Bragantia, 69, 1–8, https://doi.org/10.1590/S0006-87052010000500002, 2009.
Viswanatha Vachaspati, C., Reshma Begam, G., Nazeer Ahammed, Y., Raghavendra Kumar, K., and Reddy, R. R.: Characterization of aerosol optical properties and model computed radiative forcing over a semi-arid region, Kadapa in India, Atmos. Res., 209, 36–49, https://doi.org/10.1016/j.atmosres.2018.03.013, 2018.
Wang, H., Zhang, R., Liu, K., Wang, G., Liu, W., and Li, N.: Improved Kriging interpolation based on support vector machine and its application in oceanic missing data recovery, in: Proceedings - International Conference on Computer Science and Software Engineering, CSSE 2008, 726–729, https://doi.org/10.1109/CSSE.2008.924, 2008.
Wang, Y., You, Z., Wang, L., Wang, J., Zhou, M., Tao, M., and Kim, J.: First high temporal resolution retrievals of AOD over shallow and turbid coastal waters for Himawari-8, ISPRS J. Photogramm. Remote Sens., 228, 603–612, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2025.07.027, 2025a.
Wang, Z., Su, X., Wang, L., Lang, Q., Lu, Y., and Wang, L.: A Physics-Guided Neural Network Model to Estimate All-Sky Diffuse Solar Radiation Using Himawari-8 Data, IEEE

Trans. Geosci. Remote Sens., 63, 1–18, https://doi.org/10.1109/TGRS.2025.3543883, 2025b.
Wei, J., Li, Z., Sun, L., Peng, Y., and Wang, L.: Improved merge schemes for MODIS Collection 6.1 Dark Target and Deep Blue combined aerosol products, Atmos. Environ., 202, 315–327, https://doi.org/10.1016/j.atmosenv.2019.01.016, 2019.
Witek, M. L., Garay, M. J., Diner, D. J., Bull, M. A., and Seidel, F. C.: New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water, Atmos. Meas. Tech., 11, 429–439, https://doi.org/10.5194/amt-11-429-2018, 2018.
Witek, M. L., Garay, M. J., Diner, D. J., Bull, M. A., Seidel, F. C., Nastan, A. M., and Hansen, E. G.: Introducing the MISR level 2 near real-time aerosol product, Atmos. Meas. Tech., 14, 5577–5591, https://doi.org/10.5194/amt-14-5577-2021, 2021.
Wong, M. S., Shahzad, M. I., Nichol, J. E., Lee, K. H., and Chan, P. W.: Validation of MODIS, MISR, OMI, and CALIPSO aerosol optical thickness using ground-based sunphotometers in Hong Kong, Int. J. Remote Sens., 34, 897–918, https://doi.org/10.1080/01431161.2012.720739, 2013.
Xu, H., Xue, Y., Guang, J., Li, Y., Yang, L., Hou, T., He, X., Dong, J., and Chen, Z.: A semi-empirical optical data fusion technique for merging aerosol optical depth over China, in: 2012 IEEE International Geoscience and Remote Sensing Symposium, 2524–2527, https://doi.org/10.1109/IGARSS.2012.6350338, 2012.
Zhang, F. and O'Donnell, L. J.: Support vector regression, in: Machine Learning, Elsevier, 123–140, https://doi.org/10.1016/B978-0-12-815739-8.00007-9, 2020.
Zhao, A., Li, Z., Zhang, Y., Zhang, Y., and Li, D.: Merging MODIS and ground-based fine mode fraction of aerosols based on the geostatistical data fusion method, Atmosphere (Basel)., 8, https://doi.org/10.3390/atmos8070117, 2017.
Zhou, Y., Scavia, D., and Michalak, A. M.: Nutrient loading and meteorological conditions explain interannual variability of hypoxia in Chesapeake Bay, Limnol. Oceanogr., 59, 373–384, https://doi.org/10.4319/lo.2014.59.2.0373, 2014.
Zimmerman, D., Pavlik, C., Ruggles, A., and Armstrong, M. P.: An experimental comparison of ordinary and universal kriging and inverse distance weighting, Math. Geol., 31, 375–390, https://doi.org/10.1023/A:1007586507433, 1999.