https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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 (satellite and ground-based) is crucial for integrating diverse data sources and generating spatially consistent representations of aerosol distribution for accurate climate impact assessment. In this study, a two-stage Universal Kriging (UK) framework is employed. In the first stage, UK is used for spatial interpolation to fill missing values in individual satellite datasets (MODIS and MISR). In the second stage, Kriging is formulated as a fusion model by incorporating spatial covariance structures derived from variogram models of the satellite data, thereby producing fused AOD estimates from both satellite and ground-based (ARFINET) observations. Following this, seasonal fused AOD maps are generated for winter, pre-monsoon, and post-monsoon periods. Leave-one-out cross-validation (LOOCV) shows that the 95% confidence interval (±2σ) of the fused AOD values accommodate more than 80% of the ground-based observations, effectively capturing regional variations. This also highlights the influence of number of ground measurement points in the generation of fused map. To address this, a Residual Kriging with Machine Learning (RK-ML) approach is explored. The RK-ML framework captures stable spatial patterns and yields LOOCV scores comparable to those of the UK method, even under sparse ground-based coverage. These findings demonstrate the suitability of both UK and RK-ML approaches (with adequate ground-based observations) for producing reliable and near-instantaneous fused AOD fields over the Indian region.

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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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). Some studies have reported that if the correlations between the AOD from multiple sensors are sufficiently strong (Liu et al., 2004; Jiang et al., 2007; Prasad and Singh, 2007), then the ground-based and space-based observations can be used together for optimal characterization of aerosol features over a broader region. In this context, 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).

Efforts have been undertaken to fuse satellite products from various sensors using multiple techniques, such as Single Scanning spread Function (Gupta et al., 2008), Maximum Likelihood Estimation (Leptoukh et al., 2007; Nirala, 2008), adaptive weighted estimate algorithm (Guo et al., 2013), semi-empirical optical algorithm (Xu et al., 2012), Spatial Statistical Data Fusion (Nguyen et al., 2012; Jinnagara Puttaswamy et al., 2014); Ensemble Kalman Filter (Li et al., 2020), Bayesian Maximum Entropy (Tang et al., 2016), Neural Network Model (Qifang et al., 2005) etc. Among the various data fusion techniques, Kriging has gained significant attention for its applicability under Geostatistcal framework which accounts spatial autocorrelation (Stein & Corsten, 1991; Zhao et al., 2017). The Kriging outcomes are also found to be comparable with those from Deep Neural Networks (DNN) (Chen et al., 2020; Kadow et al., 2020). Hence, the Kriging methodology 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).

In this study, we have adapted the Kriging technique to produce optimal fused AOD products over India. 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 contrast,

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

Kriging 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) and has been further extended to multi-sensor fusion methods (Lilla and Castrignanò, 2019). While Kriging approaches have been previously applied to fuse AOD over northern India, the amount of ground data included in their studies was limited (Singh and Venkatachalam, 2014; Singh et al., 2016). In this study, AOD measurements carried out from more than 40 ground-based observatories of the Aerosol Radiative Forcing over India Network (ARFINET) are primarily used to supplement the satellite-based observations from Moderate Resolution Imaging Spectrometer (MODIS) and the Multi-angle Imaging Spectro-Radiometer (MISR). Additionally, ground-based AOD data from the Aerosol Robotic Network (AERONET) are utilized to enhance the robustness of the database. In addition to the above, a hybrid approach of Machine Learning and Residual Kriging (RK-ML) has been adapted to counter nonlinear optimal estimation and more reliable fused estimation even in presence of less number of ground based observation.

2. Data and Methodology

2.1 Ground-based AOD

The ground-based AOD is primarily obtained from ARFINET observations. ARFINET is the largest network of aerosol observatories over south Asia, having continuous measurements across the region since 1985. The spectral AOD measurements in the ARFINET observatories are carried out using a Multi-wavelength solar Radiometer (MWR) and the handheld Microtops 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 for the entire day is calculated from the forenoon and afternoon data as

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

Where, T_{FN} and T_{AN} are the durations of MWR measurements during 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:

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

The typical error in the retrieved AOD is ~ 0.02-0.03 at different wavelengths. The variance of the Langley intercept typically causes an uncertainty of ~ 5%. The instrument details, AOD retrieval method, and error budget have been discussed elsewhere (Gogoi et al., 2009; Kompalli et al., 2010; Moorthy et al., 2007).

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

Apart from MWR, AOD is obtained from handheld Microtops sun-photometer (Solar Light Company, USA) at five wavelengths (440, 500, 675, 870, and 936 nm). Microtops 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 AErosol RObotic NETwork (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. (Liu et al., 2004):

$$\ln\left(\tau_{\lambda_1}/\tau_{\lambda_2}\right) = -\alpha \ln\left(\lambda_1/\lambda_2\right) \quad (1)$$

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 a linear least squares fit to the logarithmic values of AOD measured at various wavelengths.

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). Sensitivity studies across diverse land surfaces employing various algorithms have validated that integrating the Dark Target and Deep Blue 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, the AOD measurement with an expected error of $0.05 \pm 0.15 \times AOD$ over the land and $0.03 \pm 0.05 \times AOD$ over the ocean is validated in most studies throughout the years of algorithm up gradations (Levy et al., 2005; Sayer et al., 2013; 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.

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 section 2.1 and 2.2) with varying resolutions, accuracies, and types of measurements. The aim is to enhance the overall

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

understanding and prediction of spatial variables (e.g., AOD) to produce a more accurate and comprehensive representation of columnar AOD. For this, we have adapted universal Kriging methodology, where data interpolation relies on unknown functions (e.g., satellite derived AOD) represented as trend models with spatial autocorrelation through variogram analysis. The overview of the fusion method is presented in **Fig. 1**, followed by a detailed description of each step in the following sections.

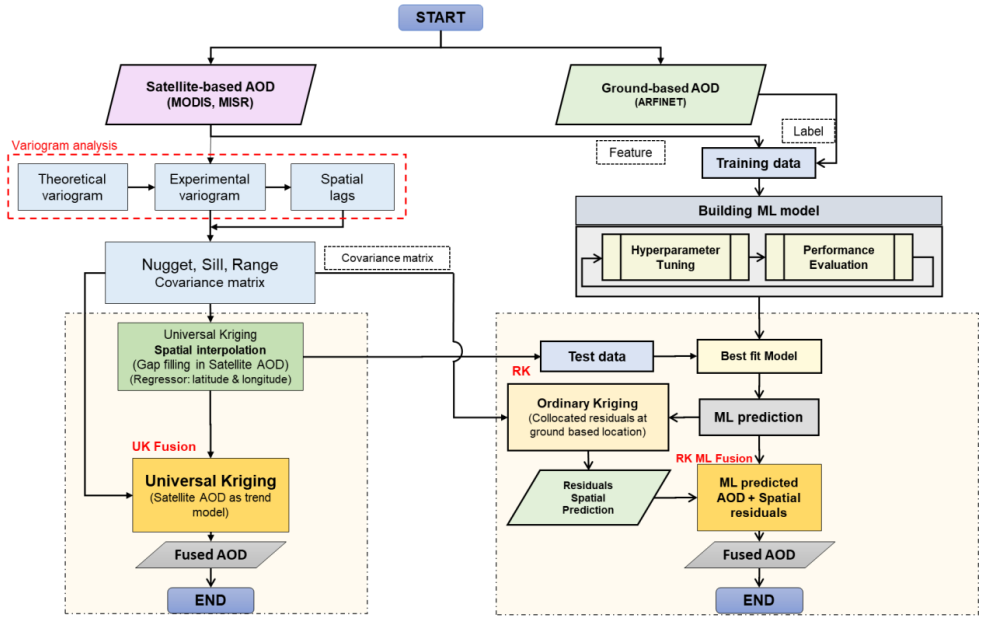

Figure 1. Flowchart of fusion methodology: Universal Kriging (UK) and Residual Kriging Machine Learning (RK-ML). The machine learning best model is designed based on the long term MODIS and MISR data which were collocated with ground-based observations.

2.3.1 Correlation analysis

As a first step of the fusion processes, the correlation analysis between the satellite and ground-based AOD is 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. Satellites offer broad spatial coverage, while ground-based measurements provide representative data at specific point locations with high temporal coverage. To improve the efficacy of AOD measurements from both satellite and ground-based sources, a statistical approach has been employed, incorporating key concepts introduced in the analysis reported elsewhere (Basart et al., 2009; Chu et al., 2002; Filonchyk et al., 2019; Ichoku et al., 2002). The spatial grid over the study domain (south Asia) is considered as 0.5° and the temporal window of 30 min which accommodates 14 to 15 measurements from MWR (data frequency 2 min) and 1 to 2 measurements from CIMEL (data frequency 15 min) observations. For this, we averaged all satellite-based AOD over a 0.5° grid box and compared it against ground-based AOD averaged over $\pm$15 min around terra-satellite overpass time. This AOD values are averaged over a certain month to represent the monthly

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

mean AOD. From this, the difference between satellite and ground AOD is assessed for different periods of the year over the study region. The results that emerged from the correlation analysis of monthly mean AODs are included in section 3.2.

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/ geophysical data often contain large-scale spatial trends, having elevation or latitude/longitude dependencies, and systematic gradients due to physical or geographical factors. These trends violate the stationarity assumption which can lead to unbounded variogram. In such a situation, detrending separates the large-scale trend from the random spatial variations. In order to validate this assumption, we have obtained (**Fig-S1**) the frequency distribution of satellite AOD and their residuals 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).

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:

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

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:

$$h_x = r \cos^{-1}(\sin\varphi_i \sin\varphi_j - \cos\varphi_i \cos\varphi_j \cos(\theta_i - \theta_j)) \quad (3)$$

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:

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

$$\gamma_{theo}(h_x) = \begin{cases} 0 & h_x = 0\ ; \\ \sigma_n^2 + \sigma_b^2\left(1 - \exp\left(-{h_x}/{l}\right)\right) & h_x > 0 \end{cases} \quad \text{(Exponential)} \tag{4}$$

$$\gamma_{theo}(h_x) = \begin{cases} 0 & h_x = 0\ ; \\ \sigma_n^2 + \sigma_b^2\left({3h_x}/{2l} - {1h_x^{\,3}}/{2l^3}\right) & 0 < h_x \le l\ ; \\ \sigma_n^2 + \sigma_b^2 & h_x > l \end{cases} \quad \text{(Spherical)} \tag{5}$$

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

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:

$$C_{ij} = \sigma^2 - \gamma_{theo} \tag{7}$$

2.3.3 Universal Kriging (UK)

Universal Kriging, 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 universal Kriging method uses both the spatial autocorrelation structure and the deterministic trend to make predictions. The universal Kriging model can be expressed as:

$$Z = M_z\beta + \in \quad \text{Or, } \in = Z - M_z\beta \tag{8}$$

Where, $Z = [Z(x_1), Z(x_2), \ldots\ldots, Z(x_n)]^T$ represent the values of the variables of interest at locations $x_1, x_2, \ldots.. x_n$, 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.

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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

$$M_z = \begin{pmatrix} 1 & MODIS_{AOD_1} & MISR_{AOD_1} \\ 1 & MODIS_{AOD_2} & MISR_{AOD_2} \\ \vdots & \vdots & \vdots \\ 1 & MODIS_{AOD_n} & MISR_{AOD_n} \end{pmatrix} \quad (9)$$

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

$$\hat{Z}(x_s) = {m_s}^T\hat{\beta} + C_{zs}^T C_{zz}^{-1}(Z - M_z\hat{\beta}) \quad (10)$$

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 (7). The unknown coefficient $\hat{\beta}$ can be expressed as the generalized least squares (GLS) estimator from the covariance matrix,

$$\hat{\beta} = \left({M_z}^T {C_{zz}}^{-1} M_z\right)^{-1} {M_z}^T {C_{zz}}^{-1} Z \quad (11)$$

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,

$$\begin{bmatrix} C_{zz} & M_z \\ {M_z}^T & 0 \end{bmatrix}\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_{zs} \\ m_s \end{bmatrix} \quad (12)$$

Here, M_z $(n \times p)$ and ${M_z}^T(p \times n)$ are trend models of AOD given by equation (9); 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:

$$\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_{zz} & M_z \\ {M_z}^T & 0 \end{bmatrix}^{-1}\begin{bmatrix} C_{zs} \\ m_s \end{bmatrix} \quad (13)$$

$$\lambda = \left\{{C_{zz}}^{-1} - {C_{zz}}^{-1} M_z\left({M_z}^T {C_{zz}}^{-1} M_z\right)^{-1} {M_z}^T {C_{zz}}^{-1}\right\} C_{zs} + {C_{zz}}^{-1} M_z\left({M_z}^T {C_{zz}}^{-1} M_z\right)^{-1} m_s \quad (14)$$

$$\lambda^T Z = {C_{zs}}^T {C_{zz}}^{-1}(Z - M_z\beta) + {m_s}^T\beta \quad (15)$$

The prediction variance associated with predicted values, can be represented as

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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

$$\sigma_{Z(x_s)}{}^2 - C_{zs}{}^T C_{zz}{}^{-1} C_{zs} + (m_s{}^T - \left(M_z{}^T C_{zz}{}^{-1} C_{zs}\right)^T)^T \, (M_z{}^T C_{zz}{}^{-1} M_z)^{-1} (m_s{}^T - \left(M_z{}^T C_{zz}{}^{-1} C_{zs}\right)^T) \quad (16)$$

The above weighting approach decides the values at prediction locations. Our foremost approach involved creating a full satellite-based map of AOD over the study region using geographical parameters as the trend model where UK method acts as spatial interpolation method to get full AOD spatial map over the study domain. 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)

While satellite and ground-based AOD measurements generally exhibit a linear correlation, regional and environmental factors introduce biases, noises and nonlinear dependencies. Although nonlinear extensions within the UK framework are possible, they require sophisticated techniques to achieve optimal performance, making the hybrid approach a compelling alternative. UK depends on spatial covariance structures with regression methods to enhance spatial predictions. The trend component in UK is conventionally modeled using low-order polynomials (e.g., first or second degree). However, real-world relationships between influencing factors and response variables are often highly non-linear, making such assumptions inadequate. While UK with linear trends is widely used, studies exploring non-linear trend modeling within the UK framework 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 bio catalytic 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 like SVR—offer an alternative for capturing non-linear and implicit relationships from data without the need for predefined functional forms. In this study, we introduced a Residual Kriging with Machine Learning (RK-ML) framework that integrates Support Vector Regression (SVR) for trend estimation. The preference for using SVR over decision-based algorithms in RK-ML stems from its suitability for problems with a small number of features and limited datasets. SVR is also well known for the robustness of regularized methods in both regression and classification (Sifaou et al., 2021). Random Forest (RF), as a bagging method, and XGBoost can also be applied within RK-ML; however, a sensitivity study comparing SVR, XGBoost, and RF on 20% of the MODIS test data showed that SVR achieved comparable or better performance metrics (R, RMSE; Supplementary **Fig. S2**). Overall, unlike UK, RK-ML does not require strict scaling for stationarity or normality, making it particularly suitable for generating fused AOD.

In this study, the RK-ML is implemented through a hybrid approach to make reliable estimates of fused AOD with improve predictive accuracy with fewer numbers of ground-based observations. For this, the best-performing SVR model is first identified based on a

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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 Residual Kriging Machine Learning (RK-ML)-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:

$$Z_{svr} = w^T \varphi + b \tag{17}$$

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

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi + \xi^*) \quad \text{Such that} \begin{cases} Z - w^T\varphi - b \le \varepsilon + \xi \\ w^T\varphi + b - Z \ge \varepsilon + \xi^* \\ \xi_i\xi_i^* \ge 0 \end{cases} \tag{18}$$

The regularization constant C trades off between the model complexity and empirical error up to which deviations larger than ε can be tolerated. ξ, ξ^* 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. (8) to eq. (15)). 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

$$\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_{zz} & 1 \\ 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{zs} \\ 1 \end{bmatrix} \tag{19}$$

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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

$$Z = \delta(x_s) + Z_{svr}(x_s) \quad (20)$$

To minimize overfitting, the SVR hyperparameters were optimized using a grid-search with the neg-mean squared error metric under a leave-one-out cross-validation (LOOCV) scheme on the training data. 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 R^2, correlation coefficient (R), 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.

There are other data fusion approaches such as Deepkriging (Chen et al., 2024), Bayesian data fusion (Pilz and Spöck, 2008; Tang et al., 1955; Xue et al., 2017), ensemble Kriging (Yang, 2018) etc. The Bayesian fusion has been applied across different satellites (Tang et al., 2016), while neural network (NN) and Gaussian weighted regression (GWR) has been implemented for large number of collocated ground and satellite observations (Li et al., 2017). However, implementing these approaches for real-time fusion is challenging, as they require large datasets and substantial computational resources, particularly for procedures such as Monte Carlo Markov Chain (MCMC) simulations. Bayesian approaches additionally require the specification of prior distributions and a sufficient number of collocated observations to estimate error structures reliably. From these discussions, it appears that RK-ML is particularly well-suited for multi-sensor fusion in regions where ground datasets are sparse and limited.

3. Results and discussions

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

The analysis of spatial pattern of AOD is crucial for understanding the consistency of measurements across different sensors. The large-scale spatial variations in the data help

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

identify overall spatial trends over latitude-longitude and elevation wise. 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.

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.**

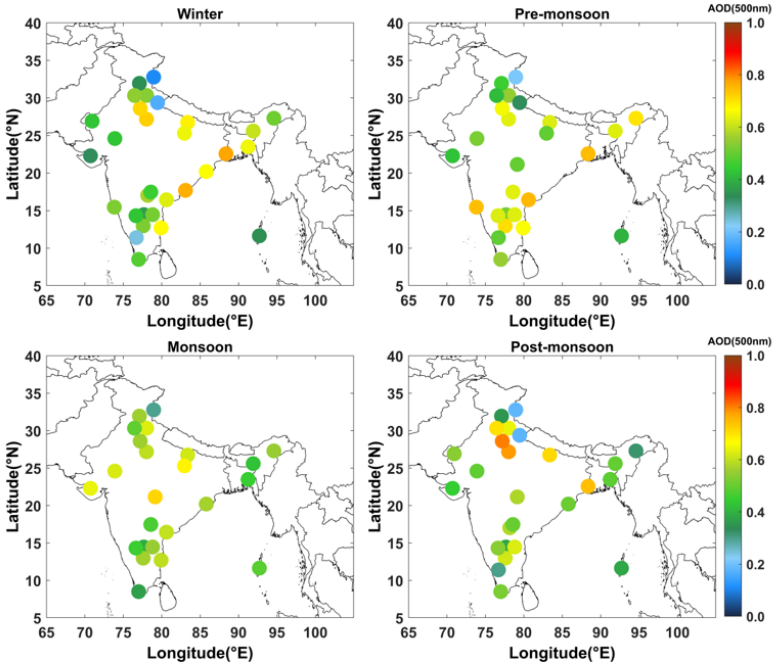

Figure 2. Long term (2011-2020) ground-based AOD at 500 nm from MWR and Microtops measurements in the ARFINET over the Indian region. The seasons are winter: December, January, February (DJF); Pre-monsoon: March, April, May (MAM); Monsoon: June, July, August, September (JJAS); Post-monsoon: October, 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. S3**).

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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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 MODIS AOD (2011–2020; **Fig. 3**) also shows persistent high AOD values in the IGP and its outflow 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.

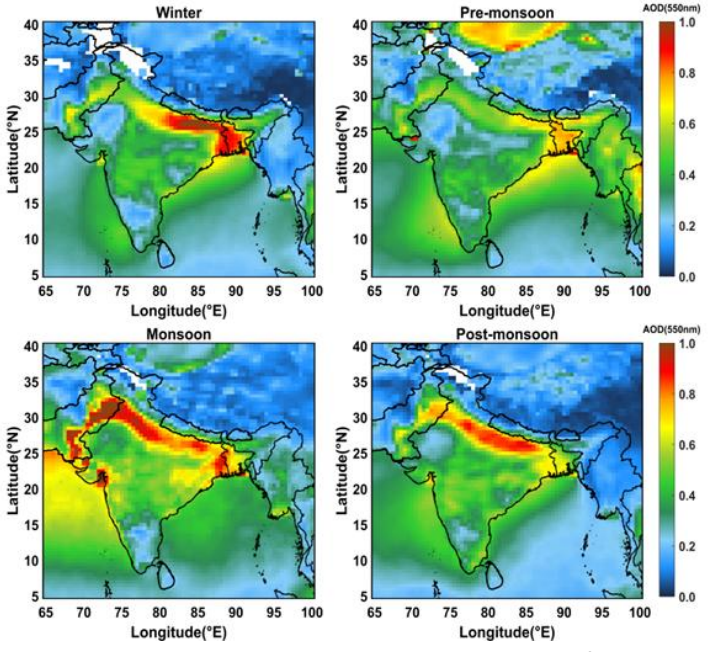

Figure 3. Long term (2011-2020) satellite based AOD (at 550 nm) from MODIS over south-Asian region.

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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

ground-based and MODIS AOD. While MODIS tends to overestimate AOD over the IGP, it generally underestimates AOD over the PI, NE, and NW regions.

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

The associations or biases between satellite and ground-based AOD at different year periods are examined for the years 2012, 2016, and 2021 (Supplementary **Figs. S4-S6**). The quartile-based plots 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 (**Figs. S7-S9**).

The spatial variability of AOD within each 0.5° spatial grid of satellite AOD values (from both MODIS and MISR) around each nearest ground station is accounted for sub-grid scale heterogeneity. Due to the dynamic nature of atmospheric conditions, a temporal window of ±15 minutes was considered for ground-based AOD measurements, and the averaged values were collocated with satellite data within a 0.5° spatial domain around each station. Sensitivity tests with varying temporal (±15, ±30, ±60 minutes) and spatial windows (0.2°, 0.5°, 1°, and 2°) indicated that spatio-temporal variability was marginal for the ±15-minute and 0.5° configuration ($\Delta MAE \approx 0.002–0.004$; $\Delta R \approx 0.002–0.01$; **Table ST2, Fig. S10**). Both MODIS and MISR showed strong correlations ($R > 0.93$) with ground-based AOD for 0.2°–0.5° resolutions, while correlations decreased at 1° (notably for MISR) and further at 2°. Correspondingly, RMSE and MAE values were lower at finer resolutions (0.2°, 0.5°) than at coarser ones (1°, 2°). These results demonstrate that the spatial domain has a stronger influence on collocation accuracy than the temporal window, and that the use of monthly-averaged datasets further enhances the correlation between satellite and ground observations by smoothing short-term variability (**Fig. S11**). The number of ground stations included in the correlation studies is given in **Table 1**. During the winter months of January, MODIS and MISR AOD show correlations of approximately 0.8 with ground observations. The association between satellite and ground-based AOD in May is weaker than in January and November. During November, both sensors showed better correlation (approximately > 0.8). The prominent locations contributing to mean errors and weak correlations with ground observations are situated in the NW and IGP regions.

The above observations provide evidence of the differences in satellite-based AOD between MODIS and MISR and ground-based measurements (as revealed in **Figs. S7-S9**). Earlier studies on comparative analyses over the same 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. The current assessment highlights that MISR outperforms in mountainous and complex regions, which is attributed to its advantages in multi-angle measurements. In these regions, MISR measurements align well within the expected range, displaying robust correlation values with respect to ground observations (Farahat, 2019).

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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 after detrending its geographical covariates, i.e., longitude, latitude, and elevation. 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. S12, S13, and S14** 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**.

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

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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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. S15**). 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 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 almost full picture 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). Universal Kriging is 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. In universal Kriging of satellite data, geographic parameters (latitude, longitude, elevation) are taken as regressors. 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).

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

The 0.5° gridded monthly mean AOD from MODIS and MISR and their spatial distribution after Spatial Prediction are shown in supplementary **Figs. S16-S21**. The variances of predictions are also shown in the figure. It is observed that compared to oceans and high-elevated surfaces like the Himalayan Mountain ranges, the spatial predictions are highly accurate over mainland regions of India, where natural and anthropogenic activities are more dominant. On the other hand, the limitations of the model over remote oceanic and high-altitude regions are expected due to significant variations in meteorological or environmental factors, which may have been overlooked. The gap regions where spatial prediction has been applied exhibit higher variance compared to the variance observed in the previously available data. Overall, using large numbers of data to interpolate a small fraction of missing values makes it possible to create a reliable picture with less associated variance.

3.3.3 Fused AOD

The fusion method 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**.

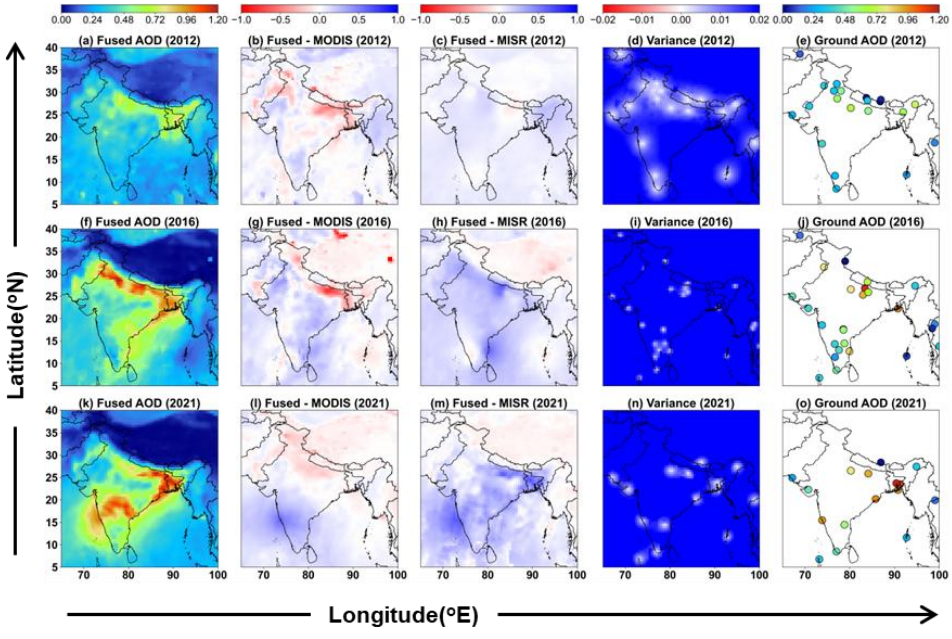

Figure 4. Fused AOD (at 550 nm) maps [(a),(f),(k)] in January for the years 2012, 2016, 2021; [(b),(g),(i)] and [(c),(h),(m)] - the deviations of fused AOD from corresponding sensors, i.e., MODIS and MISR; [(d),(i),(n)] – variance; and [(e),(j),(o)]– ground-based AOD used to generate fused maps. Blue indicates overestimations, and red means an underestimation of fused AOD from satellite retrieved AOD. The white dots in variance plots show the ground station locations.

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

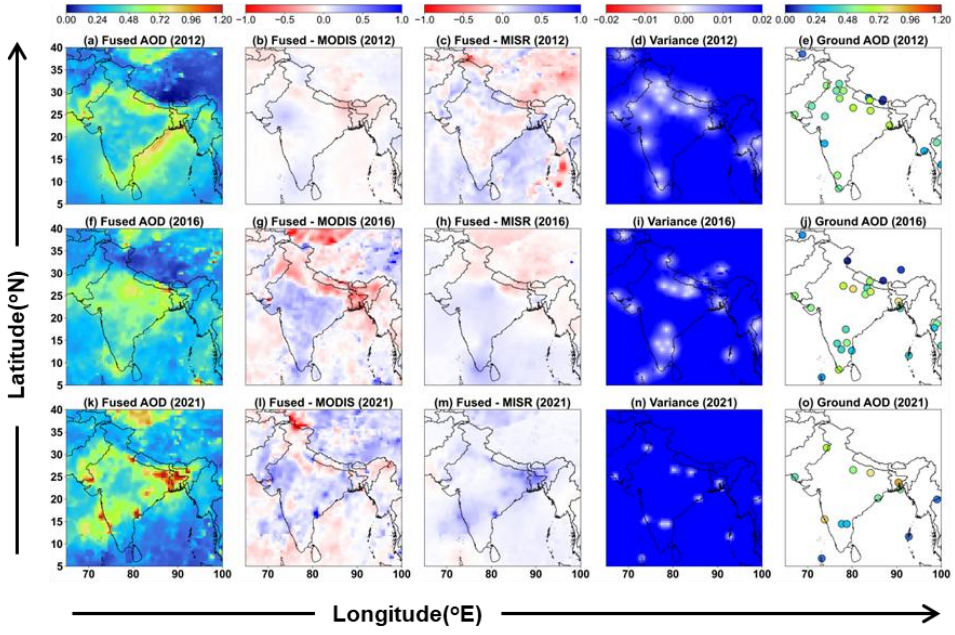

Figure 5. Similar analysis as in Figure 4, but for the month of May in 2012, 2016, and 2021.

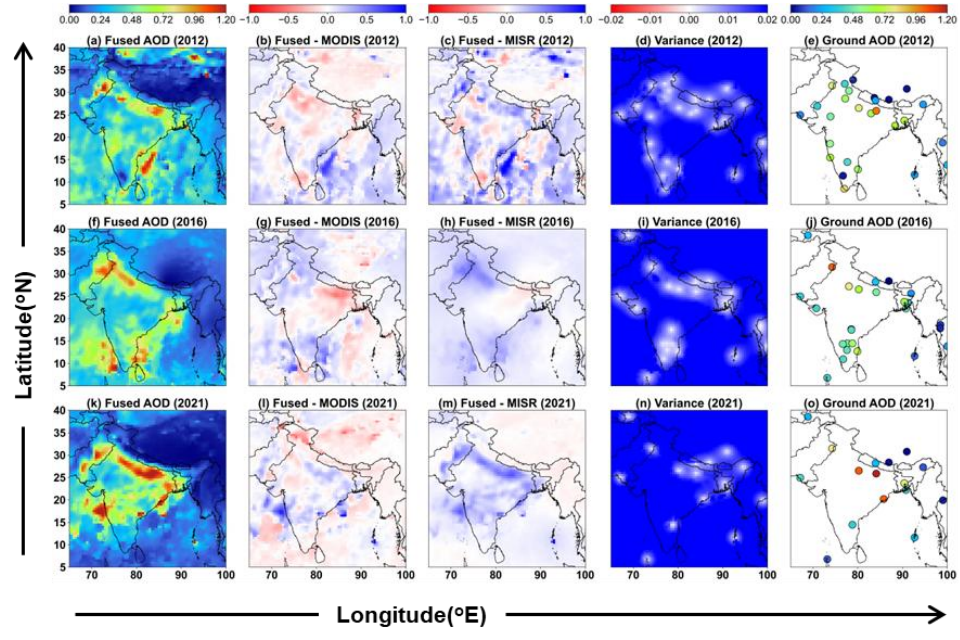

Figure 6. Similar analysis as in Figure 4, but for the month of November in 2012, 2016, and 2021.

Throughout the observation period, the fusion maps highlight the significant impact of ground-based AOD on the fusion method. Notably, the scarcity of ground-based AOD measurements directly influences the stability and reliability of spatial models, as it challenges the fulfilment of key geostatistical assumptions such as stationarity, normality, and

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

appropriate data transformation. In this study, UK is carefully applied to generate fused AOD from transformed ground data through Box-cox method using optimal λ_{bc} values. This transformed data is back transformed after Kriging method, which helps to mitigate adverse effects (e.g. negative values or undefined values) of data scarcity. This underscores that the operational applicability of UK with small datasets is dependent upon careful pre-processing of the target variable and its covariates an effort intensive and data specific process.

In the analysis conducted on a seasonal basis, it is evident that the correlation between MODIS/MISR AOD with ground AOD for May is notably weaker and shows large biases (**Table. 4, 5**) compared to other months. During fusion experiment, this discrepancy is particularly influenced by ground AOD at specific locations during certain months, as detailed in the analysis below. Overall, the fused AOD demonstrates little enhancements in the correlation between MODIS and MISR data than the ground-based measurements.

Table 4. Error and bias analysis of MODIS AOD with ground and fused AOD at ground station locations.

Year	Month	Correlation	RMSE	MAE	Slope	Bias	Correlation	RMSE	MAE	Slope	Bias
		MODIS & Ground					MODIS & Fused				
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 5. Error and bias analysis of MISR AOD with ground and fused AOD at ground station locations.

Year	Month	Correlation	RMSE	MAE	slope	bias	Correlation	RMSE	MAE	slope	bias
		MISR & Ground					MISR & Fused				
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**

A correlation study between MISR and ground observations in January 2012 shows lower error and a better correlation coefficient compared to MODIS and ground observations, resulting in the fused map closely resembling the MISR map. In 2016, MODIS overestimated AOD in the IGP region (~ 0.99 ± 0.31), while the fused AOD corrected it to ~ 0.87 ± 0.16,

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

closely matching with ground-based AOD (~ 0.86). Ground stations in the IGP, including Gandhi College (WUP), Kanpur (KNP), Gorakhpur (GKP), and Varanasi (VNS), were underestimated by MISR, while MODIS exhibited a mix of overestimations and underestimations. Excluding WUP and GKP stations, MODIS demonstrated lower errors (RMSE ~ 0.149 and MAE ~ 0.117) and a better correlation coefficient (~ 0.85) with ground observations, leading the fused map to reflect MODIS patterns predominantly. Similarly, in 2021, MODIS and MISR observed widespread AOD over peninsular India but underestimated ground-based AOD at GOA (~ 0.95), reporting ~ 0.36 and ~ 0.30. The fused AOD captured the high values and showed continuous flow toward the Arabian coast.

Across several stations of Northwestern India and IGP, including Karachi (KRC), GKP, WUP, Dhaka (DHK), Agra (AGR), VNS, and Bhola (BHL), MODIS overestimated ground observations in May-2016. This overestimation resulted in poor correlation and higher errors with ground-AOD compared to that between MISR and ground-AOD. However, during May in 2012 and 2021, MODIS demonstrated better performance than MISR, showing a strong association with ground observations, e.g., except at BHL in 2021, with improved correlation and errors (R~0.90, RMSE ~ 0.122, MAE ~ 0.102). Consequently, deviation maps revealed white patches, indicating the influence of each sensor, which was also reflected in the fused maps. In Nov-2016, over the NW and IGP, a significant discrepancy was observed between fused AOD data and satellite measurements, as noted by ground stations. For example, LHR (Lahore) reported an AOD of approximately 1.02, while AGR recorded around 0.78. Additionally, a flow pattern on the east coast of peninsular India, detected by the MISR, was also preserved in the fused map. This map showed an enhanced AOD value attributed to observations from ground stations Chennai (CHN) and Kadapa (KDP), which had been underestimated by both MODIS and MISR. Similarly, in November 2021, the flow over the IGP was modified in the fused map based on ground observations, such as those from WUP with an AOD of 1.28 and KNP with an AOD of 1.03.

Since, different time periods have different number of ground stations; the regional analysis can only assess the improvement in the fused AOD estimates over MODIS and MISR. **Fig. 7** illustrates that the regional mean fused AOD values closely follow the variations observed in individual satellite sensor data, although the magnitude differs significantly. This highlights that ground-based AOD, though traditionally advantageous for its temporal coverage, contributes to improved spatial representation. Over the IGP, where ground-based observations are more abundant than in any other region of India, the regional mean fused AOD generally lies between MODIS and MISR values. This reflects the tendency of MODIS to overestimate and MISR to underestimate AOD in this region, with the fusion process balancing these biases. 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 observations underestimate AOD in this region. Similarly, in central India, the fused AOD exceeds both sensors. Over northwest India during May, when dust loading is high, the fused AOD is close 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 further due to the lack of sufficiently distributed ground-based observations.

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

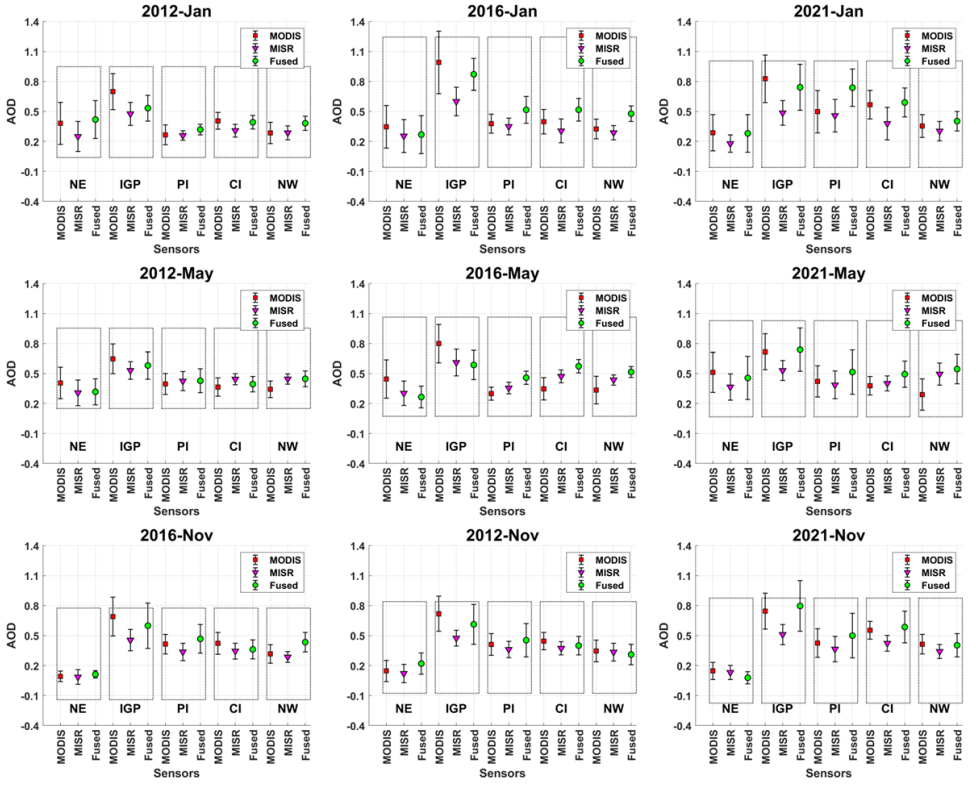

Figure 7. The regional analysis plots showing the difference between the AOD obtained from MODIS and MISR, and AOD estimated using the fusion method. On the y-axis, AOD values are shown as mean ± standard deviation.

3.3.4 Performance analysis and cross-validation

The performance analysis carried out using the correlation analysis parameters such as Root mean square error (RMSE) and mean absolute error (MAE), revealed a decrease in RMSE and MAE in fused data and an increase in correlation with MODIS and MISR (**Figs S22-S24; Table 4, 5**). Along with this, bias is also improved each time. 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** 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 **Figure-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 $\pm 2\sigma$ (95% Confidence interval) of predicted AOD for seven out of nine months. The highest accuracy was achieved in 2021 November

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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 **Table. 4, 5 & 6**.

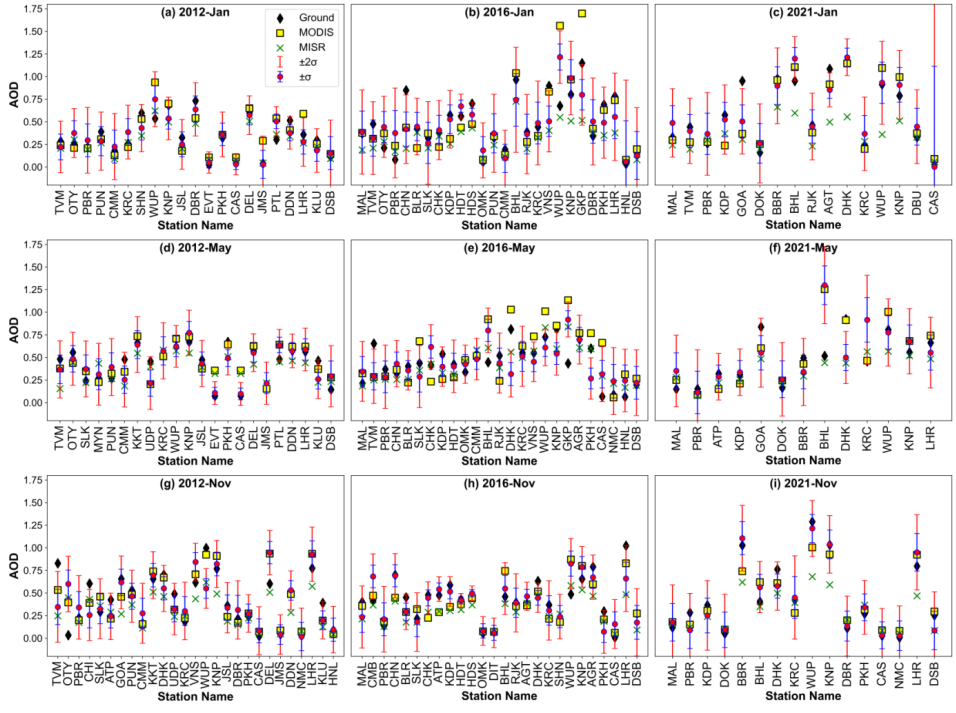

Figure 8. Predicted AOD values (as magenta points) with error bars $\pm\sigma$ (blue line), $\pm 2\sigma$ (red line), and ground observed AOD (black diamond), MODIS observation (yellow square), MISR (green cross) at different stations. For station names refer **Table ST7 and ST8**.

Table 6. Accuracy assessment of the predicted AOD through leave-one-out cross-validation.

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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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. S25-S26**) 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.

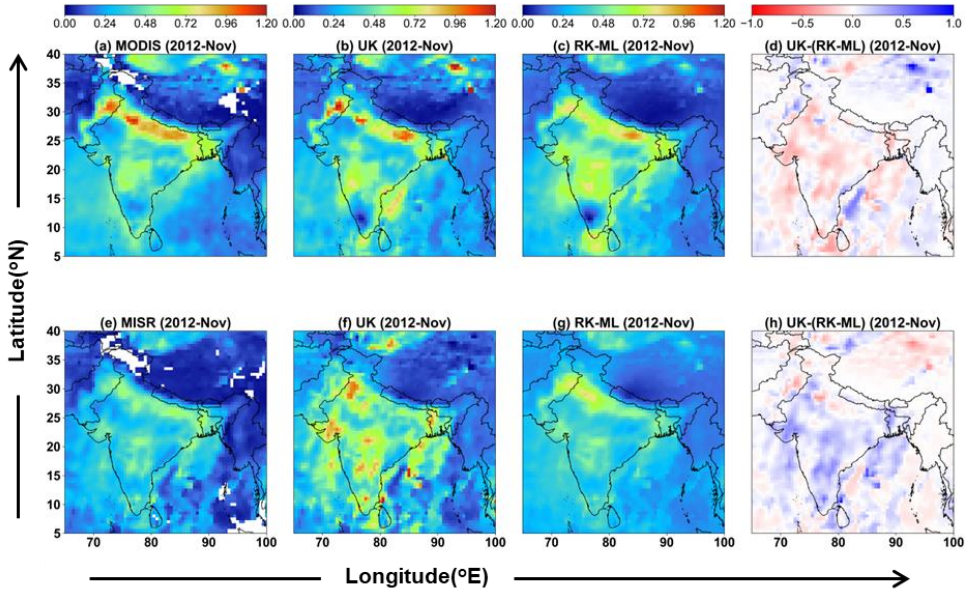

Figure 9. MODIS (a), MISR (e), and fused AOD using Universal Kriging (UK, (b), (f)), and Residual Kriging with Machine Learning (RK-ML, (c), (g)); Difference between UK and RK-ML predictions ((d), (h)). Blue color indicates where UK predictions exceed those of RK-ML, and vice versa. The top panels of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

Notably, when data points are fewer, UK has overestimated AOD values in mainland regions relative to RKML predictions. This discrepancy may arise due to the complexity of the underlying surface, where Ordinary Kriging (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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

demonstrates the RK-ML fusion ability to better manage observational uncertainties and data sparsity.

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.

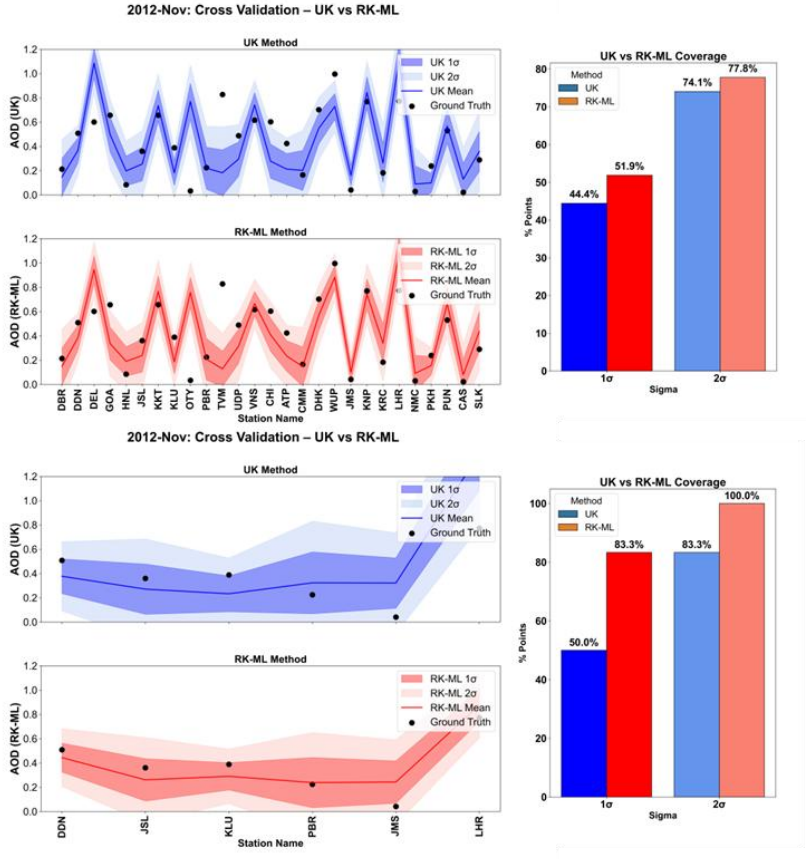

Figure 10. Line plots of LOOCV results from UK method (blue line) and RK-ML method (red line), covering the ground AOD (black dots) for Case-1 (27 points) and case -2 (6 points) within 1σ (dark shade) and 2σ (light shade). For station names, refer to supplementary **Tables – ST7 and ST8**.

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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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. S25**, increasing station density enhances the similarity between fusion maps from both methods, as seen in the larger white patches of the deviation plots (**Fig. S25a–d**).

3.3.6 Limitations of this Study and Future Scope

The present study encounters a few limitations due to:

1. A limited number of ground-based observations across geographically complex regions, further affected by data gaps due to cloudy conditions, instrumental and operational issues. Hence, fusion using UK requires careful data transformations to avoid spurious values (e.g., negatives) that may arise from unconstrained regression in UK.

2. Our collocation approach assumes spatially uniform aerosol variation within the selected spatio-temporal window and grid resolution. At instantaneous collocation scales, the variability depends on grid size, with approximate values of 0.05 for 0.1°–0.25° resolution, 0.1 for 0.5°, and 0.15 for 1°. This variability is substantially reduced when averaged to monthly means. Sensitivity tests indicate that variability increase noticeably beyond 1°, while resolutions between 0.2° and 0.5° yield comparable metrics. The typical spatial representativeness errors are in the range of ~0.01–0.09.

   In our fusion methodology, representativeness errors are not explicitly corrected at individual locations but are considered at the domain level. These errors are generally estimated from the nugget effect of collocated points. However, due to limited availability of such data, this estimation was not applied here, which likely increased the nugget effect and produced smoothed averages rather than fully resolving local variability. We therefore acknowledge that spatial representativeness error is a limitation of our current approach.

3. Uncertainties arising from the use of uniform gridding criteria have been indicated through nugget errors. The application of single variogram/trend model across diverse geographic regions and seasons may cause bias, but using geographical weighted or local variograms were impractical to create a best realization of fused AOD over a large domain while it may help when we have sufficient ground points well distributed in different regions. Though various sensor variograms can affect inaccuracies of ~0.01.

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

4. The availability of low number of ground stations and LOOCV as metric to quantify the model's performance can have limitations in regions with highly heterogeneous nearby observations, which lowers LOOCV scores.

Despite these challenges, the fusion methodology employed in this study underscores the importance of creating a fused AOD data product from space-borne and ground-based observations for large-scale applications. While our analysis does not claim to fully capture the real-world scenario, it provides optimally fused results based on the available ground observational points serving as the primary influence. When a sufficient number of simultaneous ground-based and satellite observations are available, the fusion technique can better capture fine-resolution continuous processes while representing simplified trend expressions across geographically distinct regions. With advancements in computational power and the increasing robustness of machine learning models applied to long-term observations across various sensors, real-time product generation using multiple sensors presents a potential extension of this work.

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. Especially in pre-monsoon period, the correlation coefficient and slope between MISR and fused AOD (comparable to ground-based AOD at the point locations) improved by ~11% and ~21% respectively, with reduction in RMSE by ~16% as compared to before fusion.
- 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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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.

Acknowledgments: 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) and the Atmospheric Science Data Center (ASDC) 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.

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, W., Li, Y., Reich, B. J., and Sun, Y.: Deepkriging: Spatially Dependent Deep Neural Networks for Spatial Prediction, Stat. Sin., 34, 291–311,

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

https://doi.org/10.5705/ss.202021.0277, 2024.
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.
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.

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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.

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.

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.

Leptoukh, G., Zubko, V., and Gopalan, A.: Spatial aspects of multi-sensor data fusion: Aerosol optical thickness, Int. Geosci. Remote Sens. Symp., 3119–3122, https://doi.org/10.1109/IGARSS.2007.4423505, 2007.

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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, 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.

Li, T., Shen, H., Zeng, C., Yuan, Q., and Zhang, L.: Point-surface fusion of station measurements and satellite observations for mapping PM2.5 distribution in China: Methods and assessment, Atmos. Environ., 152, 477–489, https://doi.org/10.1016/j.atmosenv.2017.01.004, 2017.

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.

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.

Pilz, J. and Spöck, G.: Why do we need and how should we implement Bayesian kriging methods, Stoch. Environ. Res. Risk Assess., 22, 621–632, https://doi.org/10.1007/s00477-007-0165-7, 2008.

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.

Qifang, X., Obradovic, Z., Bo, H., Yong, L., Braverman, A., and Vucetic, S.: Improving aerosol retrieval accuracy by integrating AERONET, MISR and MODIS data, 2005 7th Int. Conf. Inf. Fusion, FUSION, 1, 654–660, https://doi.org/10.1109/ICIF.2005.1591916, 2005.

Rossi, R. E., Dungan, J. L., and Beck, L. R.: Kriging in the shadows: Geostatistical

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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., 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

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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.
Tang, Q., Bo, Y., and Zhu, Y.: Journal of geophysical research, Nature, 175, 238, https://doi.org/10.1038/175238c0, 1955.
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.
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.
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.
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-

https://doi.org/10.5194/egusphere-2025-5824
Preprint. Discussion started: 23 December 2025

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.

Xue, J., Leung, Y., and Fung, T.: A bayesian data fusion approach to spatio-temporal fusion of remotely sensed images, Remote Sens., 9, https://doi.org/10.3390/rs9121310, 2017.

Yang, D.: Spatial prediction using kriging ensemble, Sol. Energy, 171, 977–982, https://doi.org/10.1016/j.solener.2018.06.105, 2018.

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.