https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Atmospheric Measurement Techniques Discussions

Vertical Retrieval of AOD using SEVIRI data, Case Study: European Continent

Maryam Pashayi[1], Mehran Satari[1], Mehdi Momeni Shahraki[1]

[1]Department of Geomatics Engineering, Faculty of Civil Engineering and Transportation, University of Isfahan, Isfahan, Iran

***Correspondence*:** Mehran Satari (sattari@eng.ui.ac.ir)

Abstract. Accurately determining Aerosol Optical Depth (AOD) across various altitudes with sufficient spatial and temporal resolution is crucial for effective aerosol monitoring, given the significant variations over time and space. While ground-based observations provide detailed vertical profiles, satellite data are crucial for addressing spatial and temporal gaps. This study utilizes profiles from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) and data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) to estimate vertical AOD values at 1.5, 3, 5, and 10 km layers. These estimations are achieved with spatial and temporal resolutions of 3 km × 3 km and 15 minutes, respectively, over Europe. We employed machine learning models—XGBoost (XGB) and Random Forest (RF)—trained on SEVIRI data from 2017 to 2019 for the estimations. Validation using CALIOP AOD retrievals in 2020 confirmed the reliability of our findings, emphasizing the importance of wind speed (Ws) and wind direction (Wd) in improving AOD estimation accuracy. A comparison between seasonal and annual models revealed slight variations in accuracy, leading to the selection of annual models as the preferred approach for estimating SEVIRI AOD profiles. Among the annual models, the RF model demonstrated superior performance over the XGB model at higher layers, yielding more reliable AOD estimations. Further validation using data from EARLINET stations across Europe in 2020 indicated that the XGB model achieved better agreement with EARLINET AOD profiles, with R^2 values of 0.81, 0.77, 0.71, and 0.56, and RMSE values of 0.03, 0.01, 0.02, and 0.005, respectively.

Keywords: AOD Vertical Profile, SEVIRI, Geostationary satellite, CALIOP, EARLINET, Machine Learning.

1 Introduction

Researchers acknowledge that aerosols significantly contribute to air pollution, climate change, and alterations in solar and thermal infrared radiation absorption and scattering (Hyslop, 2009; Pope et al., 2019; Li et al., 2022). Understanding their behaviour is crucial for refining atmospheric models and monitoring techniques. Aerosol Optical Depth (AOD) serves as a parameter for quantitatively estimating both the aerosol concentration and its optical properties. Recent researches highlight the pivotal role of aerosol vertical profiles in AOD retrieval uncertainties (Wang et al., 2018, Rogozovsky et al., 2021; Gupta et al., 2021; Rogozovsky et al., 2023). Moreover, understanding the vertical layering of aerosol properties enhances insights

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

into aerosol transport mechanisms, aids in source identification, elucidates atmospheric dynamics, and improves the accuracy of models tracking long-range aerosol transport (Chen et al., 2023). Consequently, continuous monitoring of the large-scale three-dimensional properties of aerosols in the atmosphere remains imperative.

Vertical AOD retrieval can be conducted through ground-based observations or inferred from remote sensing data. Ground-based LiDAR networks, such as the European Aerosol Research Lidar Network (EARLINET), provide detailed insights into various aerosol properties by offering vertical profiles of aerosol optical properties (Bösenberg et al., 2001, 2003). While these observations offer detailed vertical information, their sparse nature necessitates supplementation with satellite observations.

Satellite remote sensing emerges as the primary method for capturing temporal and spatial variations in aerosol profiles globally. While passive satellite remote sensing significantly enhances spatial coverage for aerosol monitoring, it lacks the detailed resolution of aerosol vertical layers provided by active techniques (Hsu et al., 2004; Levy et al., 2013). Spaceborne LiDAR systems, such as the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) onboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite launched in 2006, offer distributions of aerosols and clouds, along with their geometrical and optical properties. The CALIOP instrument represents the world's first operational satellite-based cloud and aerosol LiDAR (Winker et al., 2004, 2006), providing high-resolution global aerosol vertical profile data that facilitate vertical distribution studies (Winker et al., 2007). However, the CALIOP sensor encounters challenges in achieving adequate spatial and temporal coverage, with limitations in daily and global resolution (16-day temporal resolution and 5 km profile distance).

To address the limitations related to the inadequate spatial and temporal coverage of CALIOP, recent studies by Pashayi et al. (2023, 2024) have introduced Seasonal and Seasonal-Independent models. These models seek to investigate the relationship between MODIS observations, MAIAC, and CALIOP AOD for Vertical Layering of MAIAC AOD product at a spatial-temporal resolution corresponding to the MAIAC product, with a focus on the Persian Gulf region. The researchers subsequently analyse their findings using CALIOP AOD retrievals across distinct vertical layers. Despite promising outcomes, the temporal variability and transient lifespan of aerosols, particularly within vertical layers, pose challenges to the effectiveness of estimated AOD products in these studies. This limitation stems from the utilization of MODIS products aboard polar-orbiting satellites, which pass over a region approximately once a day during daylight hours, thus inadequately supporting aerosol monitoring at high temporal resolution (Wei et al., 2020).

Geostationary satellites provide observations with significantly high temporal resolutions (Wei et al., 2020). In recent years, numerous geostationary satellite sensors with enhanced radiometric, spectral, and spatial resolutions have been deployed to monitor global aerosol loading with temporal resolutions of less than an hour. Notable examples of these geostationary sensors include Himawari-8, equipped with the Advanced Himawari Imager (AHI, Da, 2015); the Advanced Baseline Imager (ABI, Kalluri et al., 2015) aboard the Geostationary Operational Environmental Satellite (GOES); and the Meteosat geostationary satellites, equipped with the Spinning Enhanced Visible and Infrared Imager (SEVIRI, Pasternak et al., 1994) instrument. These advancements facilitate observations with high temporal resolution, substantially enhancing aerosol monitoring capabilities across various regions (Schmit et al., 2018; Zhang et al., 2019; Ge et al., 2019; Tang et al., 2019; Zawadzka-Manko

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Atmospheric Measurement Techniques Discussions

et al., 2020; Witthuhn et al., 2020; Kocaman et al., 2022; Ceamanos et al., 2023). Specifically, SEVIRI offers suitable temporal and spatial resolutions, presenting a valuable opportunity to expand aerosol datasets for Europe (Stebel et al., 2021; Nicolae et al., 2021).

The retrieval of AOD values typically entails two primary approaches: physically based methods (Seidel et al., 2012; Lipponen et al., 2018; Amini et al., 2021; Mehta et al., 2022) and data mining techniques (Radosavljevic et al., 2010; She et al., 2020; Chen et al., 2022). Physically based methods rely on established principles of aerosol behaviour, utilizing models derived from physical laws to estimate AOD values. However, these methods encounter limitations due to uncertainties in inputs and the complex nature of particle phenomena. In contrast, data mining techniques offer a promising alternative by harnessing large datasets and employing machine learning algorithms to discern patterns and relationships within complex aerosol systems.

In this study, our objective is to introduce a model for enhancing the temporal resolution of AOD profile products over Europe continent by integrating SEVIRI-based information with CALIOP aerosol profile products. To achieve this, we develop a machine learning (ML) model capable of retrieving sub-hourly (approximately every 15 minutes) vertical AOD values with a spatial resolution of 3 km × 3 km. Leveraging two well-established ML models—XGBoost (XGB) and Random Forest (RF)—previously demonstrated effective in similar studies, these algorithms serve as the foundation for training layering models and assessing their seasonal independence using numerical datasets. Importantly, previous studies (Zhang et al., 2021; Lebo, 2014; Marinescu et al., 2017) have demonstrated that aerosols situated in the mid-troposphere (at altitudes ranging from 3 to 10 kilometres) significantly influence cloud characteristics (Lebo, 2014), while those in the lower troposphere have a pronounced effect on mixed-phase precipitation (Marinescu et al., 2017). Consequently, our focus is on developing region-specific models to estimate AOD values across the 1.5, 3, 5, and 10 km layers for each pixel of the SEVIRI dataset, thus notably improving the spatial-temporal resolution of AOD in these layers. Furthermore, the existence of EARLINET stations across Europe aims to validate the estimated AOD values. We organized the rest of the paper as follows: Section 2 provides a comprehensive overview of the dataset employed, while Section 3 details the necessary pre-processing steps and retrieval methodology. Subsequently, Section 4 delves into the discussion of the vertically retrieved AOD results, followed by conclusions outlined in Section 5.

2 Study Area and Data Source

2.1 Study Area

The study area encompasses a significant portion of Europe, spanning from 35°N to 71°N and -7°E to 70°E, covering approximately 10.18 million square kilometres. Despite its relatively small land area, Europe exhibits a diverse geographical landscape and complex atmospheric dynamics. Urban centers in Europe face persistent air pollution issues due to industrial activities and vehicular emissions, compounded by the effects of climate change. Various aerosol types, originating from industrial processes, transportation, biomass burning, and natural events significantly impact air quality, weather patterns, and climate dynamics across the continent. Long-range transport of aerosols, particularly from sources in Africa such as Saharan

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

dust storms, underscores the interconnectedness of atmospheric processes across continents and emphasizes the necessity of international cooperation in addressing air pollution and environmental challenges.

2.2 Data Source

2.2.1 SEVIRI

The Meteosat Second Generation (MSG) constitutes a series of four satellites managed by the Exploitation of Meteorological Satellites (EUMETSAT) and has been operational since 2004. Originally designated as MSG1 to MSG4, these satellites were subsequently rebranded as Meteosat-8 to Meteosat-11, respectively. The primary instrument onboard these satellites is the Spinning Enhanced Visible and Infrared Imager (SEVIRI), a radiometer equipped with 11 spectral channels spanning the visible to the infrared spectrum. SEVIRI provides a spatial resolution of 3 km at the sub-satellite point, with a high-resolution visible (HRV) channel offering a spatial resolution of 1 km at nadir. Strategically centered at various wavelengths, the thermal channels of SEVIRI include 6.2 and 7.3 µm (targeting strong water vapor absorption), 8.7, 10.8, and 12.0 µm (window channels), as well as 9.7 µm (for ozone absorption) and 13.4 µm (for carbon dioxide absorption). This operational system delivers full-disk Earth data, while the rapid scan service focuses on observing the upper part of the Earth's disk, covering Europe and North Africa, with a repetition time of 15 minutes (Schmetz et al., 2002; Zawadzka et al., 2014). In our study, we primarily utilize SEVIRI data from Meteosat-11, the fourth and final flight unit of the MSG program, which was launched on July 15, 2015. Meteosat-11 currently operates in geostationary orbit, positioned at 36,000 km above the equator. Its coverage extends over Europe, Africa, and the Indian Ocean, spanning from -81 to 81 degrees longitude and -79 to 79 degrees latitude. Figure 1 provides a visualization of the coverage area of SEVIRI.

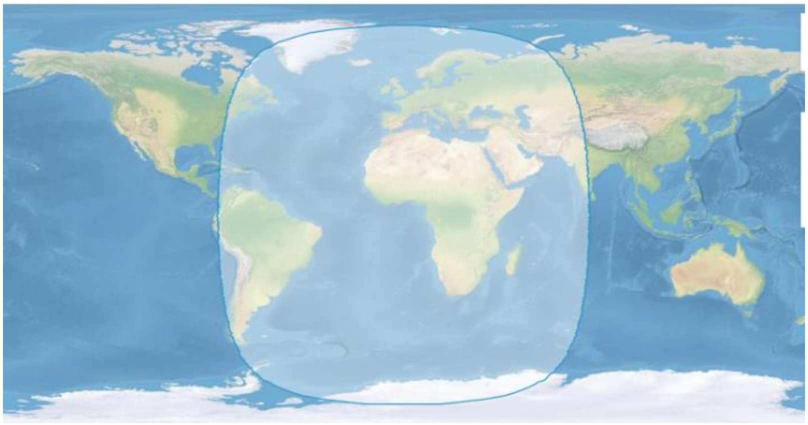


Figure 1. The area covered by the SEVIRI instrument (https://data.eumetsat.int/data/map/EO:EUM:DAT:MSG:HRSEVIRI).

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Atmospheric Measurement Techniques Discussions

2.2.2 CALIOP

The CALIOP instrument plays a pivotal role in the CALIPSO satellite, launched in April 2006 with the primary objective of reliably delivering high-resolution vertical profiles of global aerosol properties via an active sensing technique. Functioning as a polarization-sensitive LiDAR, CALIOP measures the depolarization ratio, serving as a degree of particle irregularity. CALIOP is specifically designed to observe aerosol optical properties during both the day and night, focusing on vertical layers at wavelengths of 532 and 1064 nm. Its Level 2 algorithm not only provides information on aerosol optical characteristics like particle depolarization ratio and color ratio but also retrieves extinction coefficients. Notably, CALIOP data offer a temporal resolution of approximately 16 days, capturing insights into aerosol dynamics over time. Sampling occurs at intervals of 333 m along the orbital track, maintaining a vertical resolution of 60 m from altitudes of -0.5 to 20 km and 180 m from 20 to 30 km within the vertical profile (Winker et al., 2004, 2006, and 2007). For this study, we employed CALIOP level 2 Version 4.2 aerosol profile products, featuring a horizontal resolution of 5 km and a vertical resolution of 60 m up to an altitude of 20.2 km. These data, spanning from 2017 to 2019, were utilized to estimate the profiles of AOD at layers of 1.5, 3, 5, and 10 km (denoted as $AOD_{1.5}$, AOD_3, AOD_5, and AOD_{10}) within the defined study region.

2.2.3 MODIS land cover data

In this research, we leveraged land cover data spanning 2017 to 2019, with a spatial resolution of 1 km, sourced from the global MODIS products (MCD12Q1 V6) covering Europe. These data, derived from both Terra and Aqua satellites, provide comprehensive land cover types annually from 2001. The dataset encompasses six classification schemes, elucidated in the downloadable User Guide available at https://ladsweb.modaps.eosdis.nasa.gov/. Each MCD12Q1 Version 6 Hierarchical Data Format 4 (HDF4) file comprises layers for Land Cover Type 1-5, Land Cover Property 1-3, Land Cover Property Assessment 1-3, Land Cover Quality Control (QC), and a Land Water Mask (Sulla-Menashe and Friedl, 2018). Our study specifically focuses on the first classification scheme, the Annual International Geosphere-Biosphere Program (IGBP) classification.

2.2.4 Meteorological data

Meteorological data were acquired from the European Centre for Medium-Range Weather Forecasts (ECMWF) dataset, accessible at (https://cds.climate.copernicus.eu/). ECMWF has been actively operational in real-time seasonal forecast systems since 1997, providing access to standard meteorological data. This dataset comprises two distinct sets of data (Copernicus Climate Change Service, Climate Data Store, (2021)). Firstly, version 2 of the Integrated Global Radiosonde Archive (IGRA) from 1978 integrates global radio sounding containing temperature, humidity, and wind data from various sources. The dataset is presented in the form of a global grid with a conventional grid resolution of 0.25° × 0.25°. Compared with previous-generation products, the temporal resolution has been increased from 6 hours to 1 hour, enabling the study of diurnal variations in the troposphere. Secondly, the Radio Sounding HARMonization (RHARM) homogenized dataset offers adjusted values for temperature, relative humidity, and wind. RHARM effectively eliminates systematic effects such as variations in measurement

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

sensors, biases induced by solar radiation, calibration drifts, station relocations, and other factors, across 700 IGRA radiosonde stations and ship-based radio soundings. RHARM includes twice-daily (0000 and 1200 UTC) radiosonde data at mandatory and standard levels, featuring essential parameters like air temperature (T, K), air pressure (P, Pa), wind speed (Ws, m/s), and wind direction (Wd, degrees from north). For this study, the global grid dataset is utilized over the European continent from 2017 to 2019, as depicted in Fig. 2.

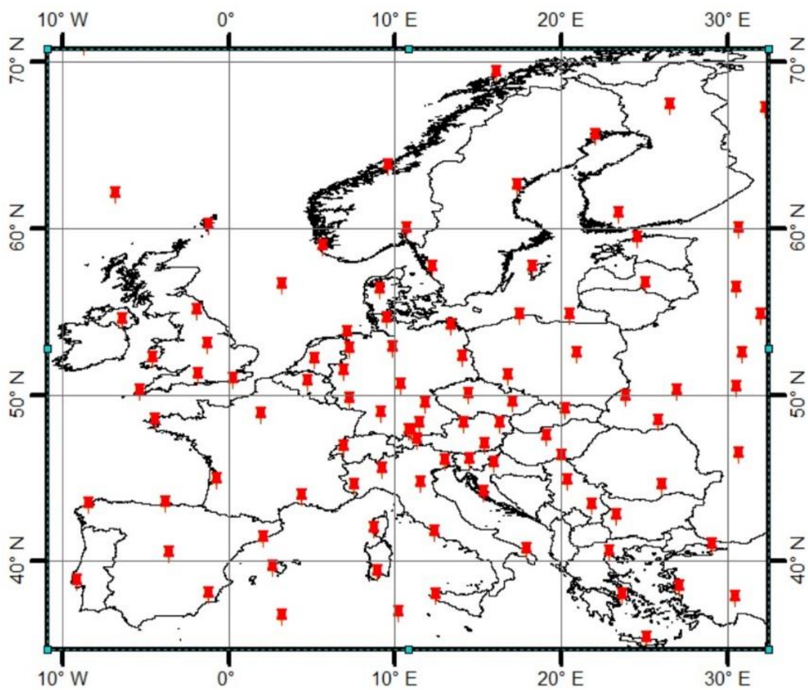


Figure 2. Map depicting the ECMWF stations for meteorological data measurements (Durre et al., 2016).

2.2.5 EARLINET

EARLINET, established in the year 2000 (Bösenberg et al., 2001, 2003), originated as a research project funded by the European Commission within the framework of the Fifth Framework Program. The primary objective of EARLINET is to generate profiles of aerosol optical properties, thereby constructing an expansive, quantitative, and statistically robust database for the continental-scale distribution of aerosols. This initiative aims to enhance network operations, facilitate research on aerosol-related processes, validate satellite sensor data, advance model development and validation, integrate aerosol data into operational models, and compile a comprehensive climatology of aerosol distribution. Currently, the network comprises 30 active stations, with the majority equipped with Raman LiDAR featuring depolarization channels. These Raman LiDAR - operating EARLINET stations typically provide profiles of aerosol extinction and backscatter coefficients without relying on significant assumptions. Figure 3 illustrates the distribution of EARLINET stations over the study area.

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

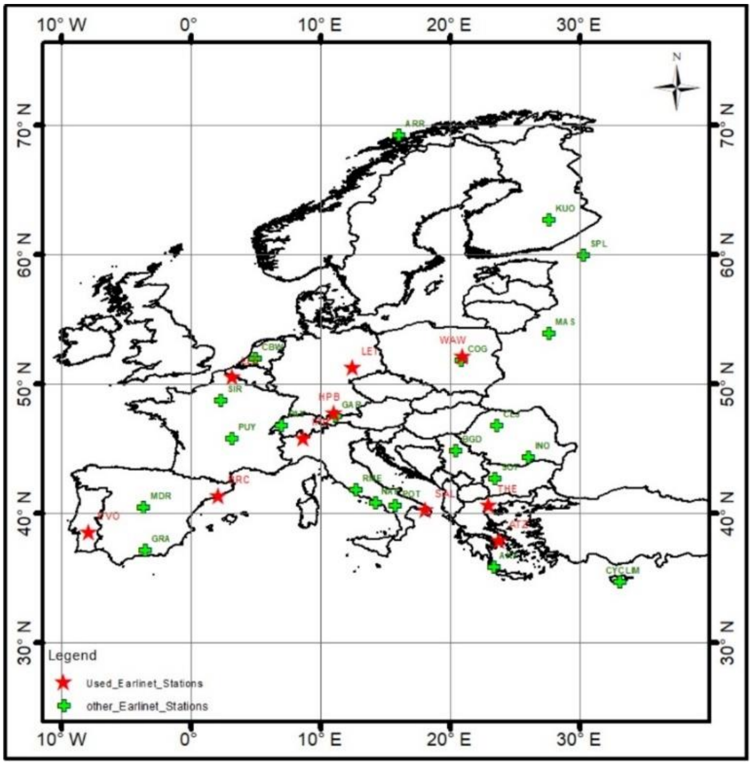


Figure 3. Map depicting currently active EARLINET stations (https://data.earlinet.org/earlinet/). The red stars indicate the geographical distribution of EARLINET LiDAR stations used in this study.

3 Methodology

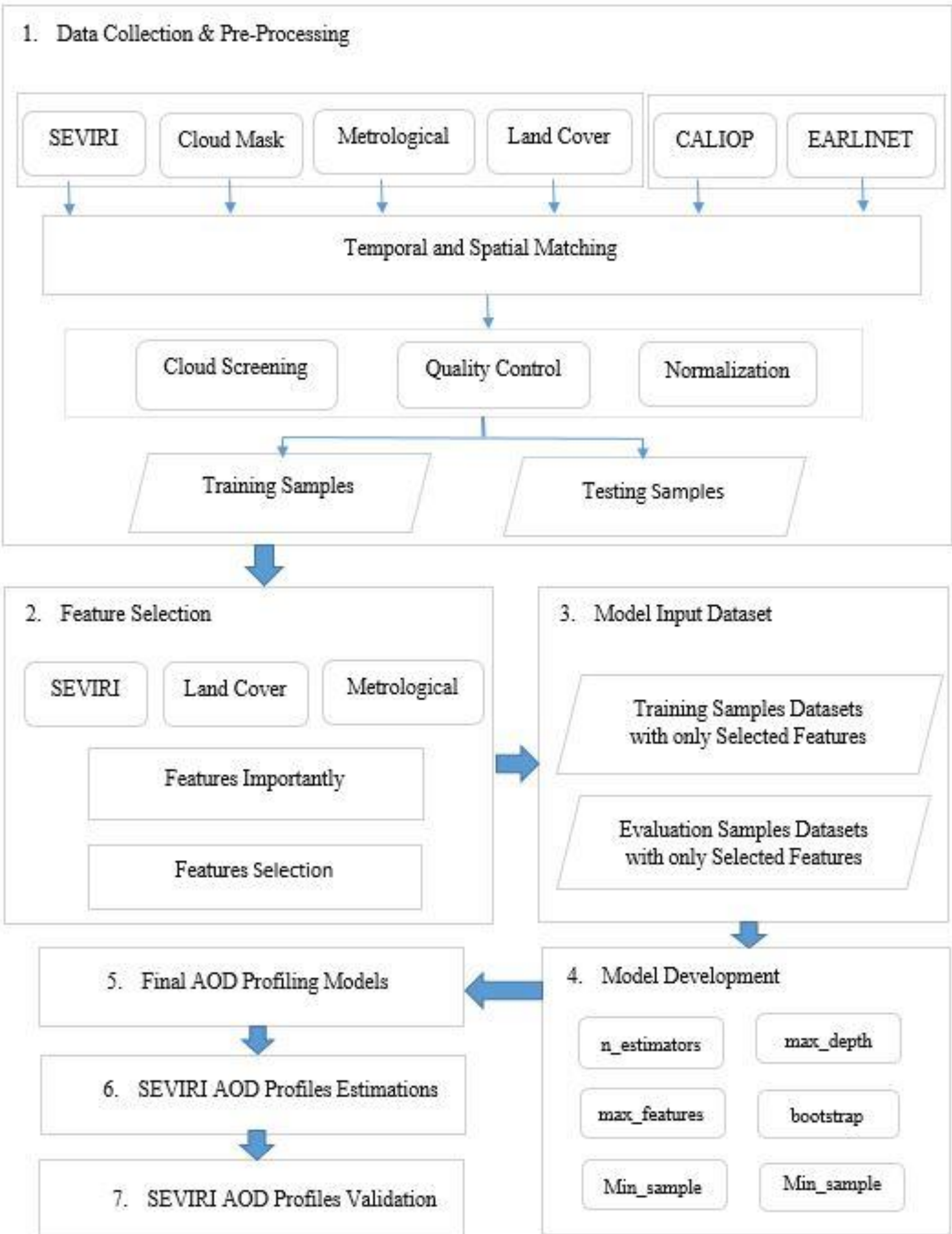
In this paper, our primary objective is to establish a robust relationship between the input variables—SEVIRI band measurements, meteorological data, and land cover products—and the vertical AOD profiles at 1.5, 3, 5, and 10 km layers. To identify this relationship, we use CALIOP AOD retrievals at these layers as reference data. Our proposed model framework for estimating AOD at the at the mentioned four distinct layers encompasses several sequential steps: data collection, preprocessing, partitioning, regression, and analysis of the performance of each regression model to ascertain the most accurate one, as illustrated in Fig. 4.

The process commences with data collection, detailed in the preceding section. Subsequently, preprocessing of both input and output data becomes imperative to ensure their suitability for subsequent analysis. The dataset is then partitioned into two subsets: training and testing, a pivotal step in ML aimed at assessing model performance and mitigating over fitting. Following data partitioning, various ML model structures are proposed and developed to capture the intricate relationships within the dataset. This phase entails selecting appropriate algorithms and architectures tailored to the specific task of AOD estimation. Finally, the performance of each model is meticulously evaluated using predefined metrics to pinpoint the most accurate and

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Atmospheric Measurement Techniques Discussions

reliable model for AOD estimation across the desired vertical layers. In the subsequent sections, we will delve into a detailed examination of each step.



180 **Figure 4.** Research framework for developing ML models to estimate SEVIRI AOD profiles.

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Atmospheric Measurement Techniques Discussions

3.1 Pre-Processing

To ensure a robust model for estimating AOD values at suitable 3D resolutions, this study integrates data from various sources, including satellites and ground-based observations. To address spatial-temporal sampling disparities, we employ a co-location approach where data from multiple sources, such as satellites and ground-based observations, are matched within a ±30 minutes timeframe and within a 3 km radius of the study area (Kittaka et al., 2011; Redemann et al., 2012; Han et al., 2017; Liu et al., 2018a). This method harmonizes disparate datasets, enhancing the reliability and comprehensiveness of our analysis. The subsequent preprocessing stages necessary for data refinement and analysis are elaborated upon in the following subsections.

3.1.1 SEVIRI

Utilizing SEVIRI data necessitates a critical preprocessing step involving co-referencing and applying geometric corrections. The Data Tailor tool, accessible at https://www.eumetsat.int/data-tailor, serves as a valuable spatial resource introduced in recent years. It simplifies the definition of coordinate systems, image systems, cutting ranges, expected output types, and requisite file extensions for the output data. Estimating AOD values requires the conversion of radiance to reflectance for the SEVIRI reflective bands (VIS06, VIS08, and NIR16), and equivalent brightness temperature for the remaining eight bands. To achieve this, we computed the Bidirectional Reflectance Factor (BRF) for the SEVIRI warm channels using Equation (1) proposed by the European Organization for the Exploitation of Meteorological Satellites (2012):

$$r_{\lambda_i} = \frac{\pi \,. R_{\lambda_i} \,.\, d^2(t)}{I_{\lambda_i} \,.\, \cos(\theta(t,x))}, \tag{1}$$

where i denotes the channel number (1 = VIS06, 2 = VIS08, 3 = NIR16, 4 = HRV), r_{λ_i} represents the Bidirectional Reflectance Factor (BRF) for channel λ_i, R_{λ_i} stands for the measured radiance in mW$\cdot m^{-2} \cdot sr^{-1} \cdot (cm^{-1})^{-1}$, d(t) signifies the Sun-Earth distance in Astronomical Unit (AU) at time t, I_{λ_i} signifies the band solar irradiance for channel λ_i at 1 AU in mW$\cdot m^{-2} \cdot sr^{-1} \cdot (cm^{-1})^{-1}$, and $\theta(t,x)$ denotes the Solar Zenith Angle in radians at time t and location x. The equivalent brightness temperature (T_b) of a satellite observation is defined as the temperature of a black body emitting the same amount of radiation. Therefore, the brightness temperature follows the form of Equation (2).

$$T_b = \frac{C_2\, v_c}{\alpha\, \log C_1\, v_c^3/R+1} - \frac{\beta}{\alpha}, \tag{2}$$

Using the observed radiances $\bar{R}$ (in $mWm^{-2}sr^{-1}(cm^{-1})^{-1}$) and radiation constants C_1 = 2hc² and C_2 = hc/k, where c, h, and k represent the speed of light, Planck's constant, and the Boltzmann constant respectively, the regression coefficients v_c, α, and β are determined through non-linear regression analysis. This analysis is conducted on a pre-calculated lookup table generated for the various SEVIRI channels, as delineated in Table 1 (Tjemkes et al., 2012).

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Table 1. Values for the regression parameters.

Channel No.	Channel ID	v_c , cm^{-1}	α	β, K
4	IR 3.9	2567.330	0.9956	3.410
5	WV 6.2	1598.103	0.9962	2.218
6	WV 7.3	1362.081	0.9991	0.478
7	IR 8.7	1149.069	0.9996	0.179
8	IR 9.7	1034.343	0.9999	0.060
9	IR 10.8	930.647	0.9983	0.625
10	IR 12.0	839.660	0.9988	0.397
11	IR 13.4	752.387	0.9981	0.578

3.1.2 CALIOP

In this study, to mitigate the impact of cloud contamination and retrieval errors on CALIOP AOD retrieval, our screening methods closely follow the guidelines established by Winker et al., 2013. We employ various quality filters to identify and filter aerosol pixels, including CAD scores, extinction QC flags, and uncertainty values. Specifically, we utilize a CAD score range outside [-100, -20] to address uncertainties in cloud-aerosol discrimination, ensuring the selection of cloud-free pixels with high confidence. Additionally, we apply extinction quality control flags with values 0 and 1 to filter extinction retrievals with high confidence. This includes constrained retrievals utilizing transmittance measurements and unconstrained retrievals where the initial LiDAR ratio remains unchanged in iterations. Furthermore, we exclusively consider daytime profiles in this study. Uncertainty flags associated with extinction coefficients are employed for data screening. Range bins with an uncertainty flag value of 99.9 km^{-1} are excluded from the analysis, following the methodology outlined by Winker et al., 2013.

3.1.3 Land Cover Product

Considering that the original MCD12Q1 product is stored in a HDF and utilizes the sinusoidal projection, several data pre-processing steps are required. These steps encompass format conversion, reprojection, resampling, image mosaicking, and sub-area masking. To execute these tasks, we employ the pyModis Free and Open-Source Python-based library. This tool enables the conversion of MODIS HDF data format into Geotiff format and facilitates the conversion of data projection from SIN to WGS84/UTM. Additionally, it facilitates image mosaicking and sub setting. Moreover, to enable comparison between the MCD12Q1 and SEVIRI datasets, the spatial resolution of MCD12Q1 is resampled at 3 km using the nearest neighbor resampling method. This method preserves the gray values of the original image, unlike bilinear interpolation or cubic convolution interpolation methods, which may alter them.

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

3.2 ML Models and parameter Tuning

In this study, our primary objective is to develop a ML model to estimate SEVIRI AOD values at various altitudes—1.5 km, 3 km, 5 km, and 10 km—using CALIOP's vertical profiles across the European continent. We employ two distinct ML algorithms, RF and XGBoost, to train layering models. Both RF and XGBoost adopt an ensemble approach, which involves constructing and aggregating multiple decision trees (Breiman, 2001; Chen and Guestrin, 2016). In RF, each tree is built using a bootstrap sample of the data, with nodes determined by the best subset of randomly selected predictors (Breiman, 2001). These trees are then averaged to obtain a final ensemble prediction. Conversely, XGBoost implements the gradient boosting method, where trees are interdependent as newly trained trees are constructed based on previous trees, incorporating their ability to predict the residuals of prior trees (Chen and Guestrin, 2016). In both RF and XGBoost, all trained trees are combined to make the final prediction.

We systematically explored various parameter combinations for each ML model. Parameters such as the number of decision trees (N_estimators), the number of variables considered for splitting at each node (max_features), and the maximum depth of each decision tree (max_depth) for RF, as well as parameters including the number of gradient boosting rounds or decision trees (n_estimators), minimum sum of instance weight (Min_sample split), maximum depth of each decision tree (max_depth), and minimum number of samples required to be at a leaf node (Min_sample leaf) for XGBoost, were optimized using a grid search algorithm. This algorithm exhaustively searches through a specified subset of the hyperparameter space. We set up a grid of possible values for each hyperparameter to be tuned, as illustrated in the "Specific Search Range" column in Table 2. For each combination of hyperparameters in the grid, the algorithm trains the model using the training data and evaluates its performance through cross-validation. The performance of each hyperparameter combination is measured using several specified evaluation metrics. Finally, the combination of hyperparameters that results in the best performance on the validation set is selected, as shown in the "Optimum Value" column in Table 2. This optimal set of parameters is then used to train the final model on the entire training dataset. For a comprehensive overview of the optimized parameters, refer to Table 2.

Table 2. The control parameter for tuning the ML models.

Model	Parameter	Specific search range	Optimum value
RF	n_estimators	50 to 150	150
	max_features	[auto, sqrt, log2]	sqrt
	max_depth	[5,10,20]	20
	bootstrap	[True, False]	False
XGBoost	n_estimators	50 to 500	100
	max_depth	[5,10,20]	20
	Min_sample split	0.1 to 1	0.3
	Min_sample leaf	3 to 10	8

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

3.3 Model Training and Evaluation

Data partitioning is a pivotal aspect of training and comprehensively assessing the performance of ML algorithms. Two widely adopted techniques for this purpose are Hold-Out and Cross-validation. Hold-Out involves dividing the dataset into training and testing sets using an 80-20 split, ensuring independent model evaluation. On the other hand, Cross-validation, often referred to as 'k-fold,' randomly partitions the data into 'k' groups, thereby enhancing generalization (Yadav and Shukla, 2016). In this study, we employed a combination of both techniques to evaluate the reliability and stability of the models across spatial and temporal domains. Initially, the dataset underwent division using Hold-Out into an 80% training set and a 20% testing set. Subsequently, a 10-fold cross-validation was conducted on the 80% training data, with eight groups randomly chosen for training and two for model validation.

During the training phase of our ML models, we leveraged datasets spanning diverse temporal periods and geographical regions where both SEVIRI and CALIOP data were accessible. However, following this training phase, the algorithms function autonomously, relying solely on SEVIRI data as their input. This advancement enables us to estimate AOD values at four specified vertical layers within each pixel of the SEVIRI dataset, based on a single SEVIRI observation along with its associated meteorological data and land cover data, covering the entire study area.

Evaluation of the AOD Profiling models involved statistical metrics such as the coefficient of determination (R^2), Pearson correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE). The selection of the optimal model was based on higher R^2 and R values, along with lower RMSE and MAE scores. Additionally, we conducted a validation analysis of AOD Profiling with EARLINET AOD profiles on a continental scale to ascertain the model's performance.

4 Results and Discussion

In this paper, our primary aim is to develop a ML model capable of retrieving AOD across four distinct vertical layers: 1.5, 3, 5, and 10 km. To accomplish this, we utilized two well-established ML models, XGB and RF, previously employed in related studies. These models were trained on SEVIRI data spanning the European continent from 2017 to 2019. Our objective was to estimate sub-hourly AOD values, approximately every 15 minutes, at a spatial resolution of 3 km × 3 km.

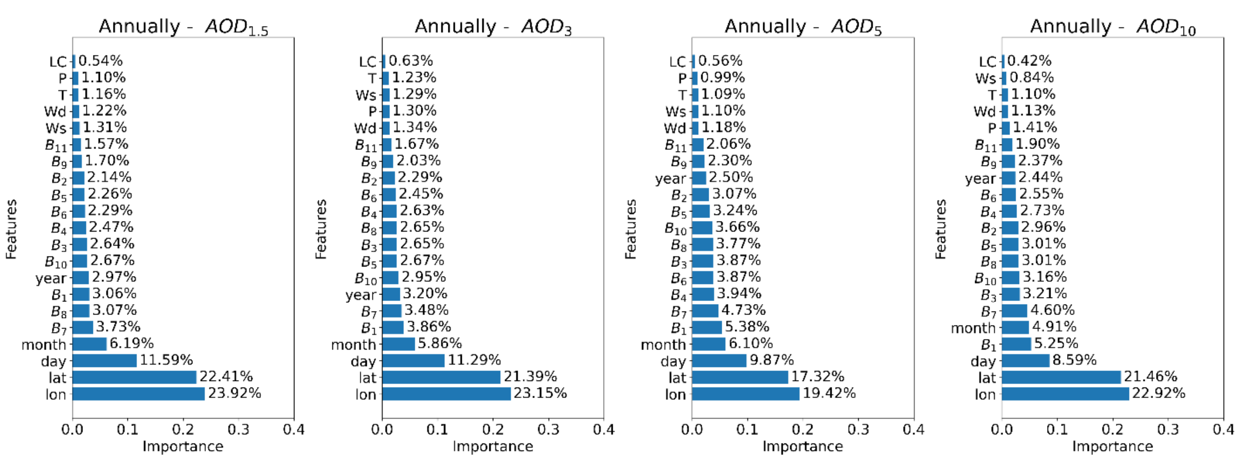
To explore the relationship between AOD and potential predictor variables, we conducted a correlation analysis experiment utilizing the Pearson Correlation Coefficient (PCC, Benesty et al., 2009). Furthermore, we evaluated the influence of land cover and meteorological data as input variables for the ML models in estimating AOD profiles from SEVIRI data, with a specific focus on identifying the most optimal model. Moreover, we conducted training and testing of the ML models across various temporal scales, including annual and seasonal analyses. Subsequently, we assessed the performance of each model using independent satellite and ground-based AOD profiles, employing evaluation metrics such as R^2, R, MAE, and RMSE. In the subsequent sections, we will provide a comprehensive review of the results derived from the aforementioned assessments.

4.1 Validation of Estimated AOD with Satellite Retrieval AOD

4.1.1 Feature Importance

According to established radiative transfer theory (Tsang et al., 1984; Zege et al., 1991), the spectral signal captured by a satellite sensor at the top of the atmosphere (TOA) is intricately shaped by various factors, including the composition, size distribution, and altitude of aerosols, as well as atmospheric molecules such as water vapor (WV). These factors directly influence the estimation of AOD values. Hence, SEVIRI reflectance and brightness temperature across bands 1 to 11 were identified as critical features for our analysis. The relationship between AOD and all candidate features was investigated through a correlation analysis experiment, as illustrated in Fig. 5, employing the PCC as the selected filter. The findings underscored that the majority of selected features in this study exhibited significance levels exceeding 1%.

Spatial features, Latitude and Longitude, alongside day, consistently demonstrate high importance to estimate the AOD values across all vertical layers. Lon is the most significant feature in all four cases, with importance ranging from approximately 19.42% to 23.92%. Lat also shows substantial importance, with values between 17.32% and 22.92%. For all four heights, the top three features remain consistent, albeit in different orders: lon, lat, and day. Although the relative importance of these features slightly decreases with height, they remain dominant. Month, year, and various spectral bands (B_1 to B_{11}) also contribute, ranging from 1.63% to 3.86%, to the model. This alignment with previous studies (Kaufman et al., 1997; Hyer et al., 2011; Chen et al., 2022) underscores the robustness of our findings. The importance of these secondary features varies slightly with the height of the AOD layer. For instance, the importance of the month feature is highest at 10 km (5.25%) and lowest at 1.5 km (3.73%). Additionally, meteorological data such as Pressure (P), Temperature (T), Wind Speed (Ws), and Wind Direction (Wd) have relatively low importance across all heights, with contributions below 2% in most cases. Land Cover (LC) is consistently the least important feature in all scenarios. Given the relative importance of meteorological data (P, T, Wd, and Ws) and LC, along with the significant influence of SEVIRI TOA measurements (B_1 to B_{11}), these were retained as input features for our AOD profile retrieval model based on machine learning techniques.



https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Figure 5. The Importance of input features in the retrieval of SEVIRI AOD profiles, as determined by the Pearson Correlation Coefficient (PCC).

4.1.2 Impact of Meteorological data and Land cover

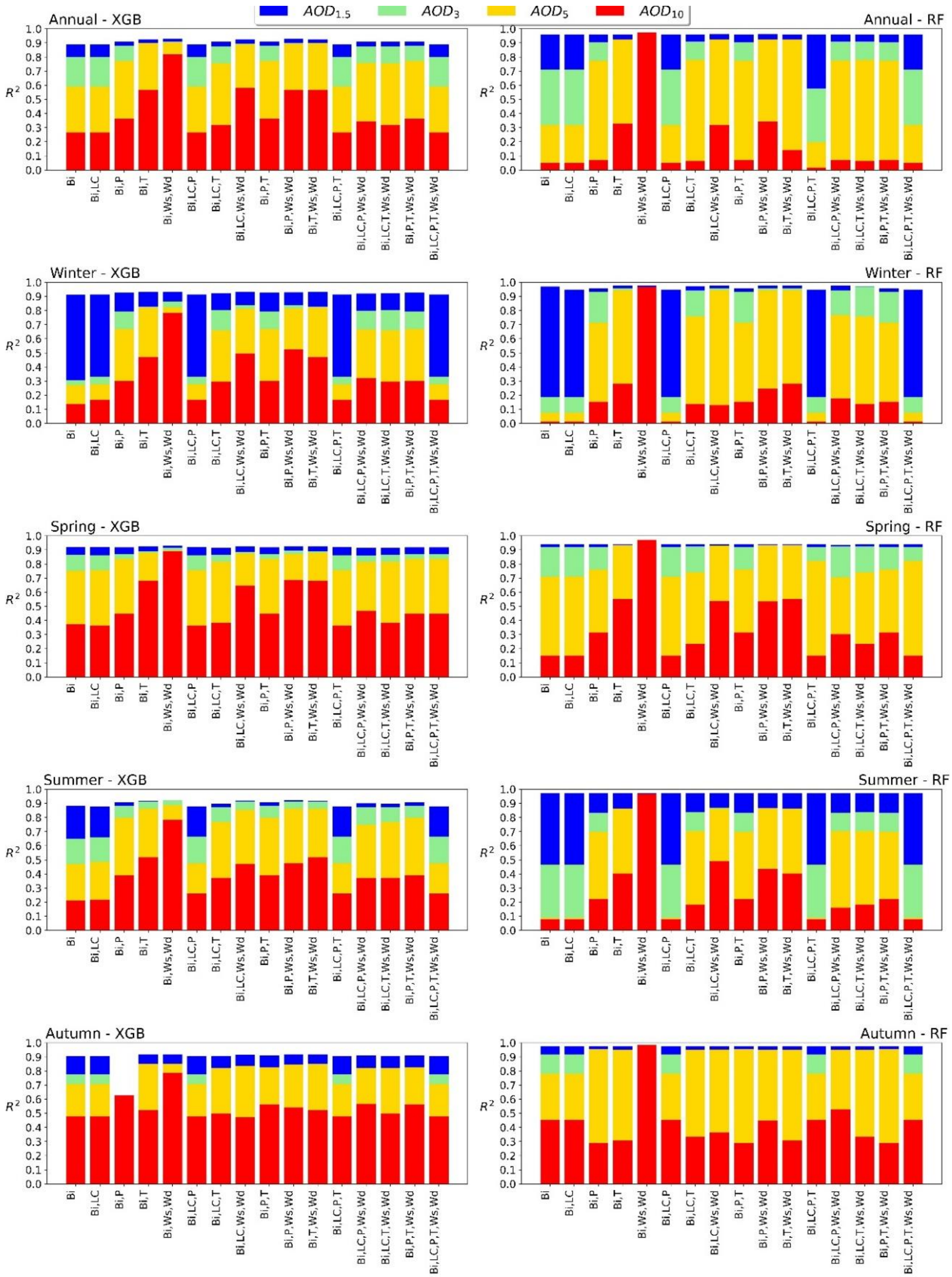
As previously noted, LC, T, P, Ws, and Wd are key features in AOD estimation. To further understand their impact on ML model performance in SEVIRI AOD profiling, we conducted 16 cases of experiments with varied feature combinations, validated using CALIOP AOD retrievals. Our analysis, depicted in Fig. 6 and supplementary Tables S1-S2, is summarized using statistical metrics like R^2, R, RMSE, and MAE.

Our findings indicate that, for most cases across annual and seasonal datasets, adding features beyond B_i has negligible impact on the 1.5 km layer. However, integrating T and P features, as seen in cases 3, 4, and 9, notably enhances AOD accuracy at 3, 5, and 10 km altitudes. This improvement is attributed to P reflecting changes in aerosol vertical layers, influencing aerosol diffusion capacity, while T is closely linked to atmospheric aerosol distribution by altering air movement dynamics. Additionally, our results highlight the superior performance of Case 5, employing Ws and Wd as input wind dynamics features. Incorporating these features significantly enhances R^2 values across models, with substantial increases ranging from 14 to 89 and 5 to 99 observed in R^2, and decreases ranging from 3.1 to 1 and 3.5 to 0.2 in RMSE for both XGB and RF models from Case 1 to Case 5 in the 10 km layer. These statistical values underscore the crucial role of wind speed and wind direction in influencing the spatial and temporal properties of atmospheric aerosols, particularly in the 10 km layer.

In comparison to P and T, wind dynamics features exhibit a greater impact on SEVIRI AOD estimation performance, particularly in the 5 and 10 km layers. This is attributed to aerosols' capacity for long-range transport within the atmosphere facilitated by wind-driven advection, particularly in these layers (Nicolae et al., 2019; Georgoulias et al., 2016; Ortiz-Amezcua et al., 2017; Granados-Muñoz et al., 2016). Conversely, LC emerges as an influential confounding feature at 3, 5, and 10 km layers in most models, resulting in a notable reduction in R^2. This phenomenon arises from the fact that the vertical distribution of aerosols across different atmospheric layers over Europa is more heavily influenced by continental and regional transport patterns, atmospheric stability, and meteorological conditions than localized land cover characteristics (Zhao et al., 2019).

Our model validation using CALIOP AOD retrievals underscores the reliability of our findings, particularly regarding the significance of Ws and Wd in improving AOD estimation accuracy. The consistency of these results across different modeling approaches (RF and XGB, Annual and Seasonal) emphasizes the significance of Ws and Wd in AOD estimation at both 5 km and 10 km layers. Consequently, we prioritize Ws and Wd, along with Bi, as the preferred input features for our models due to their demonstrated impact on improving AOD estimation accuracy.

Atmospheric Measurement Techniques
Discussions



https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Atmospheric Measurement Techniques Discussions

Figure 6. The impacts of input features on the retrieval of SEVIRI AOD profiles, represented as R^2 metrics, for the RF and XGB models. Each row displays the results for the annual period and the four seasons (Winter, Spring, Summer, and Autumn). The four colors in each bar plot indicate the R^2 values for AOD at 1.5, 3, 5, and 10 km layers.

4.1.3 Validation of Seasonal Modeling

Given the significant variations in both aerosol distribution and meteorological conditions across seasons, we aimed to investigate whether tailoring our proposed modeling approach to different seasons could enhance the precision of AOD profile retrievals. Following the methodology outlined in Section 3, we partitioned the sample dataset, derived from 2017 to 2019 data, into four segments based on seasonal distinctions: Winter (January, February, and March), Spring (April, May, and June), Summer (July, August, and September), and Autumn (October, November, and December), as detailed in Table 3. Subsequently, we trained individual ML models on these seasonal datasets. Additionally, an annual model was constructed using the entire dataset spanning 2017 to 2019. For this analysis, we separately estimated SEVIRI AOD profiles for the year 2020 using both the seasonal and annual models. The accuracy of these estimations was then assessed using CALIOP AOD retrievals. Detailed seasonal validation findings, including R^2 and RMSE metrics, are delineated in Table 4, complemented by supplementary insights and further details on R and MAE available in Table S3.

Table 3. Number of Samples Used to ML models training in this Study.

Period	All	Winter	Spring	Summer	Autumn
2017-2019	343489	82752	87358	99750	73629

The XGB model exhibited acceptable performance across different seasons, with R^2 (RMSE) values for the 1.5 km layer as follows: 0.930 (0.040 mg/m³) for spring, 0.918 (0.036 mg/m³) for summer, 0.918 (0.046 mg/m³) for autumn, and 0.932 (0.046 mg/m³) for winter. Notably, the RF model showed improvement, boasting R^2 values of 0.987, 0.992, 0.994, and 0.994, and corresponding RMSE values of 0.017, 0.011, 0.012, and 0.013 mg/m³ for spring, summer, autumn, and winter, respectively. Similarly, both the XGB and RF models demonstrated satisfactory performance across other layers, with R^2 ranging from 0.84 to 0.99, 0.82 to 0.98, and 0.78 to 0.98 for the 3 km, 5 km, and 10 km layers, respectively. Clearly, the performances of models tend to decrease in the upper layers compared to the 1.5 km layer. Due to the prevalent types and sizes of existing aerosols throughout most of the year, with aerosol distribution in Europe predominantly concentrated within the 1.5 and 3 km atmospheric layers. Consequently, R^2 and R metrics demonstrate higher values in these layers compared to the 5 and 10 km layers. Conversely, RMSE and MAE metrics are elevated at the 1.5 and 3 km layers but lower at the 5 and 10 km layers. This pattern arises from the typically higher aerosol concentrations occurring in the lower atmospheric layers, juxtaposed with lower AOD values observed in the 5 and 10 km layers.

Table 4. Seasonal performance of proposed AOD profiling ML models.

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

	Targets	Metrics	Annually	Winter	Spring	Summer	Autumn
XGB	$AOD_{1.5}$	R^2	0.928	0.932	0.930	0.918	0.918
		RMSE	0.042	0.046	0.040	0.036	0.046
	AOD_3	R^2	0.908	0.865	0.913	0.921	0.843
		RMSE	0.021	0.025	0.024	0.018	0.024
	AOD_5	R^2	0.905	0.823	0.894	0.887	0.851
		RMSE	0.015	0.022	0.018	0.016	0.018
	AOD_{10}	R^2	0.818	0.782	0.892	0.784	0.788
		RMSE	0.015	0.012	0.010	0.016	0.016
RF	$AOD_{1.5}$	R^2	0.991	0.994	0.987	0.992	0.994
		RMSE	0.015	0.013	0.017	0.011	0.012
	AOD_3	R^2	0.971	0.982	0.980	0.949	0.976
		RMSE	0.012	0.009	0.012	0.015	0.009
	AOD_5	R^2	0.983	0.986	0.982	0.956	0.979
		RMSE	0.006	0.006	0.008	0.010	0.007
	AOD_{10}	R^2	0.996	0.976	0.988	0.988	0.975
		RMSE	0.002	0.004	0.003	0.004	0.006

However, the XGB Annually model exhibited the RMSE values (0.042 mg/m³, 0.021 mg/m³, 0.015 mg/m³, and 0.015 mg/m³) and the R^2 values (0.928, 0.908, 0.905, and 0.818). Similarly, the RF Annually model demonstrated significant results, achieving R^2 values of 0.991, 0.971, 0.983, and 0.996, along with RMSE values of 0.015 mg/m³, 0.012 mg/m³, 0.006 mg/m³, and 0.002 mg/m³, respectively. In conclusion, the effectiveness of AOD profiling models exhibits slight variations in accuracy across seasons compared to annual models. Therefore, we considered the annual models as the desired models to estimate AOD profiles of SEVIRI.

4.1.4 Comparison of the models

Figure 7 presents scatterplots illustrating AOD profiles estimated using the proposed annual RF (Fig. 7a-d) and XGB (Fig. 7e-h) models at a wavelength of 530 nm, compared with CALIOP-retrieved AOD profiles over Europe in 2020. Each subplot includes the number of points and mentioned metrics i.e. R^2, R, RMSE, MAE, Bias, and linear regression equations to facilitate clear and thorough analysis.

Both models exhibit a strong correlation between the estimated values and retrievals. However, the RF model demonstrates slightly superior performance, with R^2 (R) values of 0.991, 0.971, 0.981, and 0.995 (0.995, 0.986, 0.991, and 0.998) for the 1.5, 3, 5, and 10 km layers, respectively. In comparison, the XGB model shows lower R^2 (R) values of 0.928, 0.908, 0.905,

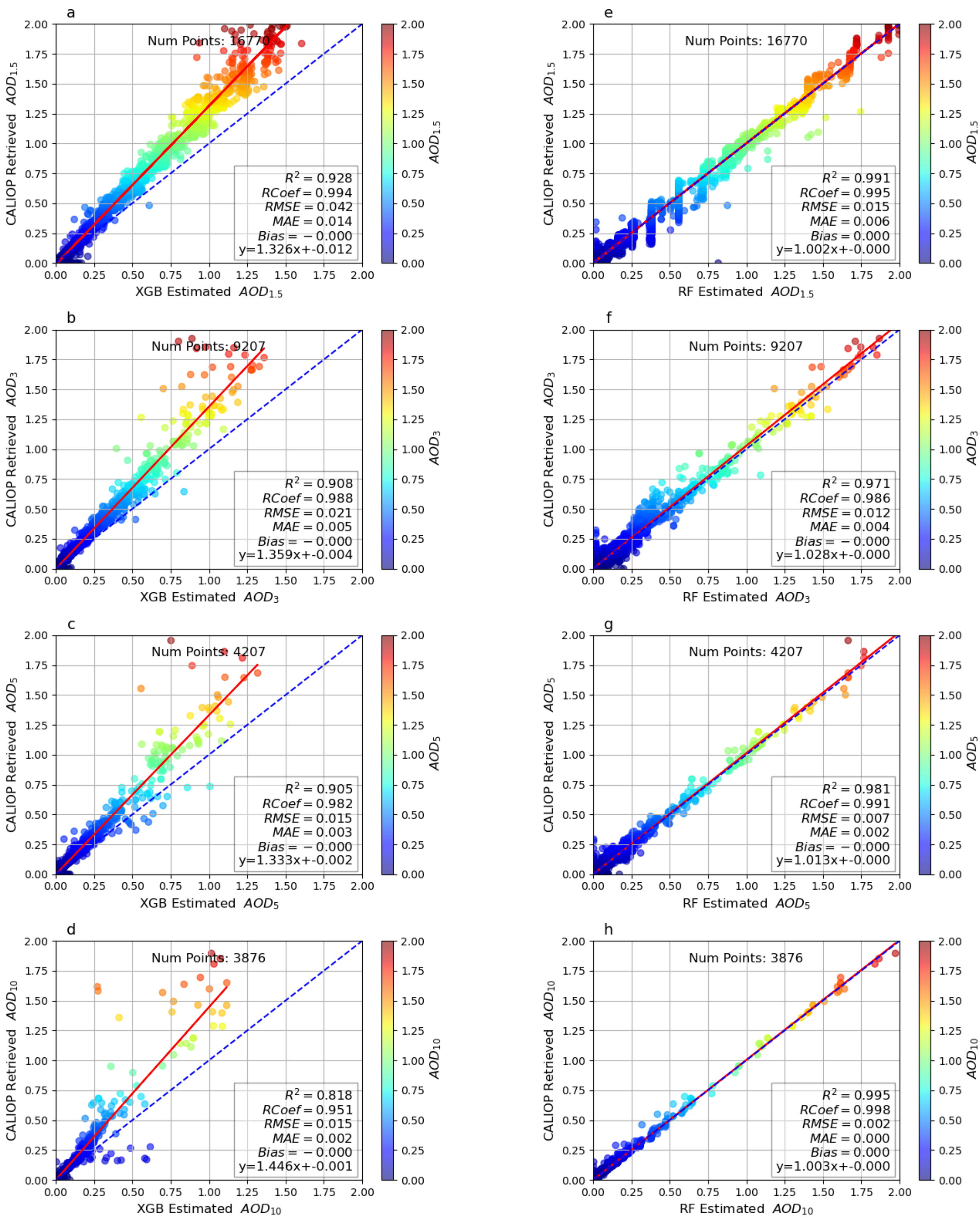
https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

and 0.818 (0.994, 0.988, 0.982, and 0.951) for the same layers. While this model demonstrates superior accuracy in estimating AOD within the 1.5 km layer compared to the other layers (3 km, 5 km, and 10 km), the RF model exhibits even greater proficiency in the upper layers, showcasing a notable discrepancy in performance when contrasted with XGB. Notably, the disparity in accuracy between the two models is more pronounced in the upper layers (3 km, 5 km, and 10 km) compared to the 1.5 km layer. This suggests that while both models offer reasonably accurate estimations of AOD in the 1.5 km layer, the RF model excels in capturing the nuances of AOD variability in the upper layers.

Overall, the minimal variation in R^2, R, RMSE, and MAE across the models suggests comparable estimation capabilities. However, a detailed analysis reveals the superior accuracy of the RF model in capturing AOD values, as evidenced by the slope values in Fig. 7e-h. In contrast, the slope values in Fig. 7a-d indicate that the XGB model tends to slightly underestimate higher AOD values. In summary, while both models demonstrate proficiency, the RF model outperforms the XGB model, particularly in its accuracy for higher altitude layers, thereby providing more reliable AOD estimations.

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Atmospheric Measurement Techniques Discussions



https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Figure 7. Scatterplots comparing the estimated SEVIRI AOD profiles derived from the proposed ML models with the CALIOP AOD profiles for the year 2020. The red line represents the linear fit between the two datasets.

4.2 Validation of Estimated AOD with Ground LiDAR Retrievals

To further validate our top-performing models, Annual XGB and RF, we conducted an extensive analysis using data from eight EARLINET stations across Europe in 2020. We focused on pixels within a 3 km radius around each station, centering our comparison and validation on the 550 nm estimated SEVIRI AOD profiles. Figure 8 visually presents the comparison between the estimated SEVIRI and retrieved EARLINET AOD profiles, utilizing a linear regression and metrics including R^2, R, RMSE, MAE, and Bias. The figure includes four scatterplots of AOD results at 1.5 km, 3 km, 5 km, and 10 km for each model. The XGB model shows better agreement with EARLINET AOD profiles, with R^2 values of 0.81, 0.77, 0.71, and 0.56, and RMSE values of 0.03, 0.01, 0.02, and 0.005, respectively. Conversely, the RF model exhibits lower correlation, with R^2 values of 0.78, 0.12, 0.43, and 0.07, and RMSE values of 0.028, 0.024, 0.022, and 0.008.

The models were trained using CALIOP data, making them expected to perform effectively when validated with data from the same source. However, significant differences between RF and XGB in feature importance ranking and extraction results (Strobl et al., 2007; Zamani Joharestani et al., 2019) suggest that the RF model may have captured specific characteristics of the CALIOP dataset, contributing to its superior performance in this context. This implies that the RF model might be biased towards the specific patterns and noise characteristics present in the CALIOP training data, leading to decreased performance when applied to the EARLINET data. In contrast, the XGB model appears to generalize to the distinct characteristics of the EARLINET data. This can be attributed to XGB's ensemble nature and its ability to reduce bias through boosting, enabling it to handle complex and diverse datasets more effectively (Ahmed et al., 2023). This adaptability allows the XGB model to perform more accurately with the EARLINET data, resulting in higher R^2 values despite the differences from the training data. In summary, the study demonstrates the superior performance of the XGB model over the RF model in retrieving AOD profiles from the EARLINET data.

When comparing the R^2 metrics of XGB AODs across different layers, it was found that XGB AODs exhibited lower R^2 values with EARLINET at the 10 km layer but showed significant improvement at the 1.5, 3, and 5 km layers, with R^2 values of 0.81, 0.76, and 0.71, respectively. This indicates a stronger correlation between XGB AOD estimations and EARLINET retrievals in these layers compared to the 10 km layer, which had an R^2 value of 0.56. This trend is consistent with other evaluation metrics. Closely scrutinizing Fig. 8, it becomes apparent that specific points revealing notable discrepancies between EARLINET and XGB AOD profiles are indicated by red rectangles in the subplots. To determine the root cause of these outliers, the data were color-coded based on AOD values, revealing that the majority of outliers occurred when EARLINET retrieved low AOD values in each layer. At these points, the XGB model tends to overestimate. This tendency contributed to a low R^2 value (0.56) in the linear regression for the 10 km layer, as this layer contains small AOD values (0-0.05).

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Furthermore, the slope of the regression line for XGB model exceeds 1, indicating a consistent underestimation of EARLINET AOD values at the 1.5km layer. In contrast, the slopes for the other layers fall within the range of 0.8 to 0.9. This observation implies that for every unit increase in estimated values, the EARLINET AOD values increase by less than one unit, suggesting a tendency for the estimated values to overestimate the EARLINET AOD in the remaining layers.

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

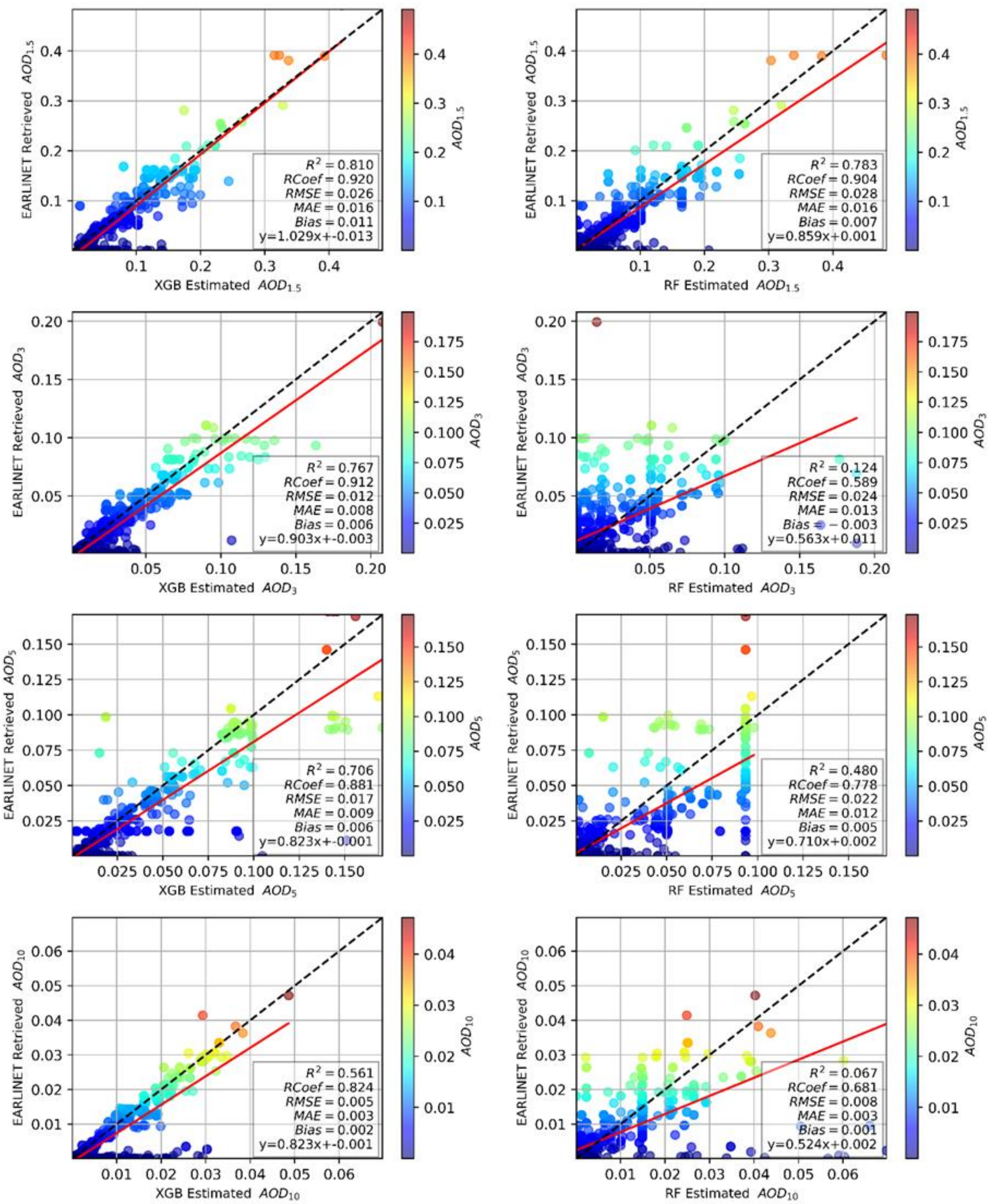


Figure 8. Scatterplots comparing the estimated SEVIRI AOD profiles derived from the proposed ML models with the EARLINET AOD profiles across 8 specified stations in Europe for the year 2020. The red line represents the linear fit

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Atmospheric Measurement Techniques Discussions

between the two datasets. Note that the scales of the subplots vary due to the different ranges of AOD values at the various vertical layers (1.5, 3, 5, and 10 km).

The statistical analysis of EARLINET AOD profiles at the eight specified stations, alongside estimated SEVIRI AOD profiles, is comprehensively presented in Table 5. The number of analyzed pairs (N) varies across stations, ranging from 12 at HPB to 387 at ATZ, providing robust validation of SEVIRI AOD profiles with EARLINET AOD profiles. Metrics such as RMSE, MAE, and Bias offer valuable insights into model performance at each station. A detailed analysis of the data reveals that the values of these metrics across the four layers are not consistently identical. Discrepancies between the XGB-estimated and retrieved EARLINET AOD values are evident in different layers for each station, as illustrated in Table 5.

Performance varies significantly across stations and layers, with notable discrepancies observed at ATZ (Greece), particularly at the 1.5 km layer, which exhibits the highest RMSE of 3.1×10^{-2} mg/m³. The substantial Bias at this station indicates that the model tends to consistently overestimate the AOD at this altitude. Conversely, performance improves at 3 km and 5 km, with RMSE values of 1.5×10^{-2} mg/m³ and 1.8×10^{-2} mg/m³, respectively. At this station, the model demonstrates the best performance at the 10 km layer, with an RMSE of 0.6×10^{-2} mg/m³, compared to the other layers. These variations are likely due to frequent forest fires in Greece, as most smoke from these fires remains in the lower layers of the atmosphere (Nicolae et al., 2019). In contrast, the XGB model generally performs well at the SAL, HPB, and LLE stations, where both RMSE and Bias are minimal. The RMSE at the 1.5 km layer, ranging from 1.1×10^{-2} mg/m³ to 2.2×10^{-2} mg/m³ at the IPR, WAW, INO, and THE stations, alongside the low RMSE and Bias values across other layers, demonstrates good overall model performance at these stations. A closer examination reveals that RMSE and Bias metrics are often elevated at the 1.5 km and 3 km layers but lower at the 5 km and 10 km layers. This pattern arises from the typically higher aerosol concentrations in the lower atmospheric layers, compared with lower AOD values retrieved in the 5 km and 10 km layers.

The discrepancies between the estimated and retrieved profiles could stem from the different measurement techniques employed by satellite and ground-based systems. EARLINET utilizes ground-based LiDAR systems to capture backscattered light from aerosols within the atmosphere by looking upward, whereas satellite measurements are performed from above, looking down. In this configuration, the lower atmospheric layers attenuate the LiDAR signal, resulting in reduced power to penetrate the upper layers. This attenuation can complicate the detection of aerosols in the upper layers (Grigas et al., 2015; Nicolae et al., 2019). Furthermore, these limitations may be attributed to the constraints associated with the utilization of CALIOP AODs, particularly their reduced precision in low aerosol concentration scenarios. This reduced precision arises from the low signal-to-noise ratio under clean weather conditions, which is often insufficient to accurately detect weak aerosol layers on the aerosol extinction vertical profile. Because both transmitted and scattered light must traverse this portion of the atmosphere, highly diffuse and/or tenuous scattering aerosol layers below the CALIOP detection threshold are ignored in CALIOP's estimates of column AOD. Consequently, weak aerosol layers that are not detected would not be retrieved, leading to decreased retrieved AODs under clean weather conditions (Liu et al., 2018a, b).

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Atmospheric Measurement Techniques Discussions

455 Finally, the efficacy of the XGB model is clearly demonstrated by its ability to reliably estimate SEVIRI AOD profiles compared to EARLINET retrieved AOD profiles across various European regions.

Table 5. Station based statistics analysis of XGB estimated SEVIRI vs. retrieved EARLINET AOD profiles.

Station ID	location	N	Layer	MAE $\times 10^{-2}$	RMSE $\times 10^{-2}$	Bias $\times 10^{-2}$
INO	Romania	13	$AOD_{1.5}$	1.6	1.8	1.3
			AOD_3	0.5	0.55	0.3
			AOD_5	1.01	2.2	1.8
			AOD_{10}	0.4	0.6	0.2
IPR	Italy	257	$AOD_{1.5}$	1.5	2.2	1.0
			AOD_3	0.8	1.1	0.6
			AOD_5	0.9	1.6	0.5
			AOD_{10}	0.3	0.5	0.2
ATZ	Greece	387	$AOD_{1.5}$	2.1	3.1	1.4
			AOD_3	1	1.5	0.8
			AOD_5	1.03	1.8	0.7
			AOD_{10}	0.3	0.6	0.3
SAL	Italy	13	$AOD_{1.5}$	0.05	0.05	0.05
			AOD_3	0.04	0.05	0.03
			AOD_5	0.008	0.01	0.008
			AOD_{10}	0.04	0.04	0.04
THE	Greece	50	$AOD_{1.5}$	1.1	1.7	0.9
			AOD_3	0.8	1.3	0.4
			AOD_5	0.5	0.7	0.4
			AOD_{10}	0.3	0.5	0.3
WAW	Poland	28	$AOD_{1.5}$	0.3	0.4	0.3
			AOD_3	0.06	0.1	0.06
			AOD_5	0.05	0.07	0.05
			AOD_{10}	0.05	0.05	0.05
HPB	Germany	12	$AOD_{1.5}$	0.08	0.1	0.07
			AOD_3	0.06	0.07	0.06
			AOD_5	0.02	0.02	0.02
			AOD_{10}	0.02	0.03	0.02
LLE	France	39	$AOD_{1.5}$	0.2	0.3	0.2
			AOD_3	0.08	0.1	0.08

	AOD_5	0.03	0.04	0.03
	AOD_{10}	0.04	0.04	0.04

5 Conclusion

This study develops a model that integrates satellite TOA reflectance data from the SEVIRI satellite, meteorological data, and land cover data to estimate vertical AOD across distinct layers of 1.5, 3, 5, and 10 km. Utilizing CALIOP AOD profiles as reference data, models employing RF and XGB were trained on a dataset spanning 2017 to 2019. Subsequently, SEVIRI AOD profiles for 2020 over Europe were estimated and compared with CALIOP and EARLINET AOD products, leading following insight.

Both RF and XGB models demonstrate commendable accuracy in sub-hourly (approximately 15-minute intervals) SEVIRI AOD profile estimation when validated using CALIOP AOD data. However, the RF model exhibits slightly superior performance, with R^2 values ranging from 0.971 to 0.995 across the different layers. In comparison, the XGB model outperforms the RF model when compared to EARLINET retrieval AOD profiles, with R^2 values ranging from 0.561 to 0.810 across the layers. Additionally, the inclusion of meteorological data (T, P, Ws, and Wd) alongside LC data during model training enhances the performance of the proposed frameworks. These features, often overlooked in physical AOD retrieval methods relying solely on atmospheric radiation transfer models, significantly contribute to refining SEVIRI AOD profile estimates. Notably, wind speed and direction emerge as the most influential meteorological data, leading to increased R^2 values and reduced RMSE across all estimated SEVIRI AOD profiles.

In conclusion the XGB model can estimate detailed sub-hourly 3x3 km² SEVIRI AOD profiles, providing valuable insights into aerosol properties. Although our study focuses on Europe and validates the model using ground-based LiDAR data, future research should broaden its application to establish a more generalized approach for AOD retrieval, considering the utilization of an ensemble of geostationary meteorological satellites simultaneously. Additionally, our current model utilized a restricted set of features, overlooking significant factors that influence AOD values, such as precipitation, NDVI, and land use. Enhancing the model's performance by integrating these additional features is a primary focus of future research.

Data and Material Availability. Data will be made available upon request.

Code availability. Code will be made available on request.

Author Contributions. MP conducted the investigation, design, data curation, and processing, as well as the programming and evaluation. MP also wrote the original draft. MS was responsible for conceptualization, methodology, supervision, and validation, and contributed to the review and editing of the manuscript. MMSh contributed to the conceptualization, supervision, and validation of the work. All authors have read and approved the published version of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgment. The authors extend their gratitude to the SEVIRI team (https://eoportal.eumetsat.int) for their dedicated efforts in developing and continuously improving SEVIRI data products. They also acknowledge the MODIS Science Team for their invaluable work in processing and making available the Collection 6 MODIS Land Cover Dynamics (MCD12Q2) Product (www.ladsweb.modaps.eosdis.nasa.gov). We also extend our heartfelt thanks to the CALIOP (www.eosweb.larc.nasa.gov) and EARLINET (https://data.earlinet.org/earlinet/) Profile Products for their crucial contribution as benchmarks, validating our model outputs and providing invaluable reference data. Additionally, we express our gratitude to the European Centre for Medium-Range Weather Forecasts (ECMWF) for granting us access to their meteorological dataset (https://cds.climate.copernicus.eu), which greatly enhanced our understanding and modeling of atmospheric conditions.

Funding. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahmed, A., Song, W., Zhang, Y., Haque, M. A., & Liu, X.: Hybrid BO-XGBoost and BO-RF Models for the Strength Prediction of Self-Compacting Mortars with Parametric Analysis. Materials, 16(12), 4366. https://doi.org/10.3390/ma16124366, 2023.

Amini, S., Momeni, M., & Monadjemi, A.: Sensitivity analysis of Look-up table for satellite-based aerosol optical depth retrieval. Journal of Aerosol Science, 158, 105842. https://doi.org/10.1016/j.jaerosci.2021.105842, 2021.

Benesty, J., Chen, J., Huang, Y., Cohen, I.: Pearson Correlation Coefficient. In: Noise Reduction in Speech Processing. Springer Topics in Signal Processing, vol 2. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-00296-0_5, 2009.

Bishop, C. M.: Neural Networks for Pattern Recognition. Clarendon Press google schola, 2, 223-228. https://doi.org/10.1093/oso/9780198538493.001.0001, 1995.

Bösenberg, J., Ansmann, A., Baldasano, J.M., Balis, D., Bockmann, C., Calpini, B., Wiegner, M.,: EARLINET: a European aerosol research lidar network. Advances in laser remote sensing, 155, 2001.

Bösenberg, J., Matthias, V., Amodeo, A., Amoiridis, V., Ansmann, A., Baldasano, J. M., ... & Zerefos, C.: EARLINET project: A European Aerosol Research Lidar Network. Max-Planck Institute (MPI), Final Report, 348, 1-250, 2003.

Breiman, L.: Random forests. Machine learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324, 2001.

Ceamanos, X., Six, B., Moparthy, S., Carrer, D., Georgeot, A., Gasteiger, J., ... & Katsev, I.: Instantaneous aerosol and surface retrieval using satellites in geostationary orbit (iAERUS-GEO)–Estimation of 15-min AOD from MSG/SEVIRI and evaluation with reference data. Atmospheric Measurement Techniques Discussions, 2023, 1-37. https://doi.org/10.5194/amt-16-2575-2023, 2023.

Chen, B., Dong, L., Huang, J., Wang, Y., Jing, Z., Yan, W., ... & Huang, Y.: Analysis of Long-Term Trends in the Vertical Distribution and Transport Paths of Atmospheric Aerosols in Typical Regions of China Using 15 Years of CALIOP Data. Journal of Geophysical Research: Atmospheres, 128(14), e2022JD038066. https://doi.org/10.1029/2022JD038066, 2023.

Chen, B., Ye, Y., Tong, C., Deng, J., Wang, K., & Hong, Y.: A novel big data mining framework for reconstructing large-scale daily MAIAC AOD data across China from 2000 to 2020. GIScience & Remote Sensing, 59(1), 670-685. https://doi.org/10.1080/15481603.2022.2051382, 2022.

Chen, T., & Guestrin, C.: Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794). http://arxiv.org/abs/1603.02754, 2016.

Chen, Y., Fan, M., Li, M., Li, Z., Tao, J., Wang, Z., & Chen, L.: Himawari-8/AHI aerosol optical depth detection based on machine learning algorithm. Remote Sensing, 14(13), 2967. https://doi.org/10.3390/rs14132967, 2022.

Copernicus Climate Change Service, Climate Data Store.: In situ atmospheric harmonized temperature, relative humidity and wind from 1978 onward from baseline radiosonde networks. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.f101d0bf, 2021.

Da, C.: Preliminary assessment of the Advanced Himawari Imager (AHI) measurement onboard Himawari-8 geostationary satellite. Remote sensing letters, 6(8), 637-646. https://doi.org/10.1080/2150704X.2015.1066522, 2015.

Durre, I., Xungang, Y., Vose, R.S., Applequist, S., Arnfield, J.: Integrated Global Radiosonde Archive (IGRA), Version 2. NOAA National Centers for Environmental Information. DOI: 10.7289/V5X63K0Q, 2016.

European Organization for the Exploitation of Meteorological Satellites: Conversion from radiances to reflectances for SEVIRI warm channels, Rep. EUM/MET/TEN/12/0332, 8 pp. Available at: (http://www.eumetsat.int/website/home/Data/Products/Calibration/MSGCalibration/index.html), 2012.

Ge, B., Li, Z., Liu, L., Yang, L., Chen, X., Hou, W., ... & Qie, L.: A dark target method for Himawari-8/AHI aerosol retrieval: Application and validation. IEEE Transactions on Geoscience and Remote Sensing, 57(1), 381-394. DOI: 10.1109/TGRS.2018.2854743, 2018.

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Georgoulias, A. K., Alexandri, G., Kourtidis, K. A., Lelieveld, J., Zanis, P., Pöschl, U., ... & Tsikerdekis, A.: Spatiotemporal variability and contribution of different aerosol types to the aerosol optical depth over the Eastern Mediterranean. Atmospheric Chemistry and Physics, 16(21), 13853-13884. https://doi.org/10.5194/acp-16-13853-2016, 2016.

Granados-Muñoz, M. J., Navas-Guzmán, F., Guerrero-Rascado, J. L., Bravo-Aranda, J. A., Binietoglou, I., Pereira, S. N., ... & Alados-Arboledas, L.: Profiling of aerosol microphysical properties at several EARLINET/AERONET sites during the July 2012 ChArMEx/EMEP campaign. Atmospheric chemistry and physics, 16(11), 7043-7066. https://doi.org/10.5194/acp-16-7043-2016, 2016.

Grigas, T., Hervo, M., Gimmestad, G., Forrister, H., Schneider, P., Preißler, J., ... & O'Dowd, C.: CALIOP near-real-time backscatter products compared to EARLINET data. Atmospheric Chemistry and Physics, 15(21), 12179-12191. https://doi.org/10.5194/acp-15-12179-2015, 2015.

Gupta, G., Ratnam, M. V., Madhavan, B. L., Prasad, P., & Narayanamurthy, C. S.: Vertical and spatial distribution of elevated aerosol layers obtained using long-term ground-based and space-borne lidar observations. Atmospheric Environment, 246, 118172. https://doi.org/10.1016/j.atmosenv.2020.118172, 2021.

Han, B., Ding, H., Ma, Y., Gong, W.: Improving retrieval accuracy for aerosol optical depth by fusion of MODIS and CALIOP data. Teh. Vjesn. 24 (3), 791–800. https://doi.org/10.17559/TV-20160429044233, 2017.

Hsu, N. C., Tsay, S. C., King, M. D., & Herman, J. R.: Aerosol properties over bright-reflecting source regions. IEEE transactions on geoscience and remote sensing, 42(3), 557-569. DOI: 10.1109/TGRS.2004.824067, 2004.

Hyslop, N. P.: Impaired visibility: the air pollution people see. Atmospheric Environment, 43(1), 182-195. https://doi.org/10.1016/j.atmosenv.2008.09.067, 2009.

Hyer, E. J., Reid, J. S., & Zhang, J.: An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals. Atmospheric Measurement Techniques, 4(3), 379-408. https://doi.org/10.5194/amt-4-379-2011, 2011.

Kalluri, S., Gundy, J., Haman, B., Paullin, A., Van Rompay, P., Vititoe, D., & Weiner, A.: A high performance remote sensing product generation system based on a service oriented architecture for the next generation of Geostationary Operational Environmental Satellites. Remote Sensing, 7(8), 10385-10399. https://doi.org/10.3390/rs70810385, 2015.

Kaufman, Y. J., Tanré, D., Gordon, H. R., Nakajima, T., Lenoble, J., Frouin, R., ... & Teillet, P. M.: Passive remote sensing of tropospheric aerosol and atmospheric correction for the aerosol effect. Journal of Geophysical Research: Atmospheres, 102(D14), 16815-16830. https://doi.org/10.1029/97JD01496, 1997.

Kittaka, C., Winker, D., Vaughan, M., Omar, A., Remer, L.: Intercomparison of column aerosol optical depths from CALIPSO and MODIS-Aqua. Atmos. Meas. Tech. 4 (2), 131. https://doi.org/10.5194/amt-4-131-2011, 2011.

Kocaman, S., Debaecker, V., Bas, S., Saunier, S., Garcia, K., & Just, D.: A comprehensive geometric quality assessment approach for MSG SEVIRI imagery. Advances in Space Research, 69(3), 1462-1480. https://doi.org/10.1016/j.asr.2021.11.018, 2022.

Lebo, Z. J.: The sensitivity of a numerically simulated idealized squall line to the vertical distribution of aerosols. Journal of the Atmospheric Sciences, 71(12), 4581-4596. https://doi.org/10.1175/JAS-D-14-0068.1, 2014.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., & Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean. Atmospheric Measurement Techniques, 6(11), 2989-3034. https://doi.org/10.5194/amt-6-2989-2013, 2013.

Li, J., Carlson, B.E., Yung, Y.L., Lv, D., Hansen, J., Penner, J.E., Liao, H., Ramaswamy, V., Kahn, R.A., Zhang, P., Dubovik, O., Ding, A., Lacis, A.A., Zhang, L., Dong, Y.: Scattering and absorbing aerosols in the climate system. Nat. Rev. Earth Environ. 3, 363–379. https://doi.org/10.1038/s43017-022-00296-7, 2022.

Lipponen, A., Mielonen, T., Pitkänen, M. R., Levy, R. C., Sawyer, V. R., Romakkaniemi, S., ... & Arola, A.: Bayesian aerosol retrieval algorithm for MODIS AOD retrieval over land. Atmospheric Measurement Techniques, 11(3), 1529-1547. https://doi.org/10.5194/amt-11-1529-2018, 2018.

Liu, B., Ma, Y., Gong, W., Zhang, M., Wang, W., Shi, Y.: Comparison of AOD from CALIPSO, MODIS, and sun photometer under different conditions over Central China. Sci. Rep. 8, 10066. https://doi.org/10.1038/s41598-018-28417-7, 2018 a.

Liu, C., Shen, X., Gao, W.: Intercomparison of CALIOP, MODIS, and AERONET aerosol optical depth over China during the past decade. Int. J. Rem. Sens. 39 (21), 7251–7275. https://doi.org/10.1080/01431161.2018.1466070, 2018 b.

Marinescu, P. J., van den Heever, S. C., Saleeby, S. M., Kreidenweis, S. M., & DeMott, P. J.: The microphysical roles of lower-tropospheric versus midtropospheric aerosol particles in mature-stage MCS precipitation. Journal of the Atmospheric Sciences, 74(11), 3657–3678. https://doi.org/10.1175/jas-d-16-0361.1, 2017.

Mehta, M., Jain, A., & Chauhan, P.: Aerosol optical depth retrieval over land from OCEANSAT-2/OCM-2 data–A simple physics based approach. Atmospheric Pollution Research, 13(3), 101339. https://doi.org/10.1016/j.apr.2022.101339, 2022.

Nicolae, A., Alexandru, M., Horatiu, S., Andrei, R., Camelia, B., Olga, Z. M., ... & Claus, Z.: SEVIRI Aerosol Optical Depth Validation Using AERONET and Intercomparison with MODIS in Central and Eastern Europe. Remote Sensing, 13(5), 844. https://doi.org/10.3390/rs13050844, 2021.

Nicolae, V., Talianu, C., Andrei, S., Antonescu, B., Ene, D., Nicolae, D., ... & Vasilescu, J.: Multiyear typology of long-range transported aerosols over Europe. Atmosphere, 10(9), 482. https://doi.org/10.3390/atmos10090482, 2019.

Ortiz-Amezcua, P., Guerrero-Rascado, J. L., Granados-Muñoz, M. J., Benavent-Oltra, J. A., Böckmann, C., Samaras, S., ... & Alados-Arboledas, L.: Microphysical characterization of long-range transported biomass burning particles from North America at three EARLINET stations. Atmospheric Chemistry and Physics, 17(9), 5931-5946. https://doi.org/10.5194/acp-17-5931-2017, 2017.

Pashayi, M., Satari, M., & Shahraki, M. M.: Improvement of spatial-temporal resolution of aerosol profile by using multi-source satellite data over the Persian Gulf. Atmospheric Environment, 292, 119410. https://doi.org/10.1016/j.atmosenv.2022.119410, 2023.

Pashayi, M., Satari, M., Shahraki, M. M., & Amini, S.: MAIAC AOD profiling over the Persian Gulf: A seasonal-independent machine learning approach. Atmospheric Pollution Research, 102128. https://doi.org/10.1016/j.apr.2024.102128, 2024.

https://doi.org/10.5194/amt-2024-105
Preprint. Discussion started: 1 August 2024

Pasternak, F., Lorsignol, J., & Wolff, L.: Spinning enhanced visible and infrared imager (SEVIRI): the new imager for meteosat second generation. In Space Optics 1994: Earth Observation and Astronomy (Vol. 2209, pp. 86-94). SPIE. https://doi.org/10.1117/12.185247, 1994.

Pope III, C.A., Lefler, J.S., Ezzati, M., Higbee, J.D., Marshall, J.D., Kim, S.Y., Bechle, M., Gilliat, K.S., Vernon, S.E., Robinson, A.L., Burnett, R.T.: Mortality risk and fine particulate air pollution in a large, representative cohort of U.S. adults. Environ. Health Perspect. 127 (7) https://doi.org/10.1289/EHP443, 2019.

Radosavljevic, V., Vucetic, S., & Obradovic, Z.: A data-mining technique for aerosol retrieval across multiple accuracy measures. IEEE Geoscience and Remote Sensing Letters, 7(2), 411-415. DOI: 10.1109/LGRS.2009.2037720, 2010.

Redemann, J., Vaughan, M.A., Zhang, Q., Shinozuka, Y., Russell, P.B., Livingston, J.M., Remer, L.A.: The comparison of MODIS-Aqua (C5) and CALIOP (V2 & V3) aerosol optical depth. Atmos. Chem. Phys. 12 (6), 3025–3043. https://doi.org/10.5194/acp-12-3025-2012, 2012.

Rogozovsky, I., Ansmann, A., Althausen, D., Heese, B., Engelmann, R., Hofer, J., ... & Chudnovsky, A.: Impact of aerosol layering, complex aerosol mixing, and cloud coverage on high-resolution MAIAC aerosol optical depth measurements: Fusion of lidar, AERONET, satellite, and ground-based measurements. Atmospheric Environment, 247, 118163. https://doi.org/10.1016/j.atmosenv.2020.118163, 2021.

Rogozovsky, I., Ohneiser, K., Lyapustin, A., Ansmann, A., Chudnovsky, A.: The impact of different aerosol layering conditions on the high-resolution MODIS/ MAIAC AOD retrieval bias: the uncertainty analysis. Atmos. Environ. 309, 119930. https://doi.org/10.1016/j.atmosenv.2023.119930, 2023.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., & Ratier, A.: An introduction to Meteosat second generation (MSG). Bulletin of the American Meteorological Society, 83(7), 977-992. https://doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2, 2002.

Schmit, T.J.; Lindstrom, S.S.; Gerth, J.J.; Gunshor, M.M.: Applications of the 16 spectral bands on the Advanced Baseline Imager (ABI). J. Oper. Meteorol, 6, 33–46. https://doi.org/10.15191/nwajom.2018.0604, 2018.

Seidel, F. C., Kokhanovsky, A. A., & Schaepman, M. E.: Fast retrieval of aerosol optical depth and its sensitivity to surface albedo using remote sensing data. Atmospheric Research, 116, 22-32. https://doi.org/10.1016/j.atmosres.2011.03.006, 2012.

She, L., Zhang, H. K., Li, Z., de Leeuw, G., & Huang, B.: Himawari-8 aerosol optical depth (AOD) retrieval using a deep neural network trained using AERONET observations. Remote Sensing, 12(24), 4125. https://doi.org/10.3390/rs12244125, 2020.

Stebel, K., Stachlewska, I. S., Nemuc, A., Horálek, J., Schneider, P., Ajtai, N., ... & Zehner, C.: SAMIRA-Satellite based monitoring initiative for regional air quality. Remote sensing, 13(11), 2219. https://doi.org/10.3390/rs13112219, 2021.

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T.: Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC bioinformatics, 8, 1-21. https://doi.org/10.1186/1471-2105-8-25, 2007.

Sulla-Menashe, D., & Friedl, M. A.: User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. Usgs: Reston, Va, Usa, 1, 18. Available at: (https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf), 2018.

Atmospheric Measurement Techniques Discussions

Tang, D., Liu, D., Tang, Y., Seyler, B. C., Deng, X., & Zhan, Y.: Comparison of GOCI and Himawari-8 aerosol optical depth for deriving full-coverage hourly PM2. 5 across the Yangtze River Delta. Atmospheric environment, 217, 116973. https://doi.org/10.1016/j.atmosenv.2019.116973, 2019.

Tjemkes, S., Stuhlmann, R., Hewison, T., Müller, J., Gartner, V., & Rota, S.: The conversion from effective radiances to equivalent brightness temperatures. Technical report, EUMETSAT. Version, 1, Available at: (https://www-cdn.eumetsat.int/files/2020-04/pdf_effect_rad_to_brightness.pdf), 2012.

Tsang, L., Kong, J. A., & Shin, R. T.: Radiative transfer theory for active remote sensing of a layer of nonspherical particles. Radio Science, 19(02), 629-642, DOI: 10.1029/RS019i002p00629, 1984.

Wang, H., Sun, Z., Li, H., Gao, Y., Wu, J., & Cheng, T.: Vertical-distribution characteristics of atmospheric aerosols under different thermodynamic conditions in Beijing. Aerosol and Air Quality Research, 18(11), 2775-2787. https://doi.org/10.4209/aaqr.2018.03.0078, 2018.

Wei, X., Chang, N. B., Bai, K., & Gao, W.: Satellite remote sensing of aerosol optical depth: Advances, challenges, and perspectives. Critical Reviews in Environmental Science and Technology, 50(16), 1640-1725. https://doi.org/10.1080/10643389.2019.1665944, 2020.

Winker, D. M., Hunt, W., & Hostetler, C.: Status and performance of the Caliop lidar. Laser Radar Techniques for Atmospheric Sensing, 5575, 8–15. https://doi.org/10.1117/12.571955, 2004.

Winker, D. M., Hunt, W. H., & McGill, M. J.: Initial performance assessment of CALIOP. Geophysical Research Letters, 34 (19). https://doi.org/10.1029/2007GL030135, 2007.

Winker, D. M., Tackett, J. L., Getzewich, B. J., Liu, Z., Vaughan, M. A., & Rogers, R. R.: The global 3-D distribution of tropospheric aerosols as characterized by CALIOP. Atmospheric Chemistry and Physics, 13(6), 3345-3361. https://doi.org/10.5194/acp-13-3345-2013, 2013.

Winker, D. M., Vaughan, M., & Hunt, W.: The CALIPSO mission and initial results from CALIOP. Proceedings of SPIE, 6409. https://doi.org/10.1117/12.698003, 2006.

Witthuhn, J., Hünerbein, A., & Deneke, H.: Evaluation of satellite-based aerosol datasets and the CAMS reanalysis over the ocean utilizing shipborne reference observations. Atmospheric Measurement Techniques, 13(3), 1387-1412. https://doi.org/10.5194/amt-13-1387-2020, 2020.

Yadav, S., Shukla, S.: Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. February. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC). IEEE, pp. 78–83. https://doi.org/10.1109/IACC.2016.25, 2016.

Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., & Talebiesfandarani, S.: PM2. 5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. Atmosphere, 10(7), 373. https://doi.org/10.3390/atmos10070373, 2019.

Zawadzka, O., & Markowicz, K.: Retrieval of aerosol optical depth from optimal interpolation approach applied to SEVIRI data. Remote Sensing, 6(8), 7182-7211. https://doi.org/10.3390/rs6087182, 2014.

Zawadzka-Manko, O., Stachlewska, I. S., & Markowicz, K. M.: Near-real-time application of seviri aerosol optical depth algorithm. Remote Sensing, 12(9), 1481. https://doi.org/10.3390/rs12091481, 2020.

Zege, E. P., Ivanov, A. P., & Katsev, I. L.: Image transfer through a scattering medium (Vol. 349). Heidelberg: Springer-Verlag, 1991.

Zhang, M. Z., Deng, X., Zhu, R. H., Ren, Y. Z., & Xue, H. W.: The impact of aerosol vertical distribution on a deep convective cloud. Atmosphere, 12(6), 675. https://doi.org/10.3390/atmos12060675, 2021.

Zhang, Z., Wu, W., Fan, M., Tao, M., Wei, J., Jin, J., ... & Wang, Q.: Validation of Himawari-8 aerosol optical depth retrievals over China. Atmospheric environment, 199, 32-44. https://doi.org/10.1016/j.atmosenv.2018.11.024, 2019.

Zhao, C., Liu, Z., Wang, Q., Ban, J., Chen, N. X., & Li, T.: High-resolution daily AOD estimated to full coverage using the random forest model approach in the Beijing-Tianjin-Hebei region. Atmospheric environment, 203, 70-78. https://doi.org/10.1016/j.atmosenv.2019.01.045, 2019.