

Article

Spatial Gap-Filling of Himawari-8 Hourly AOD Products Using Machine Learning with Model-Based AOD and Meteorological Data: A Focus on the Korean Peninsula

Youjeong Youn ¹, Seoyeon Kim ¹, Seung Hee Kim ² and Yangwon Lee ^{1,*}

¹ Major of Geomatics Engineering, Division of Earth Environmental System Sciences, Pukyong National University, Busan 48513, Republic of Korea; dbwjd0757@pukyong.ac.kr (Y.Y.); ksyvv08@pukyong.ac.kr (S.K.)

² Institute for Earth, Computing, Human and Observing (ECHO), Chapman University, Orange, CA 92866, USA; sekim@chapman.edu

* Correspondence: modconfi@pknu.ac.kr

Abstract: Given the complex spatiotemporal variability of aerosols, high-frequency satellite observations are essential for accurately mapping their distribution. However, optical remote sensing encounters difficulties in detecting Aerosol Optical Depth (AOD) over cloud-covered regions, creating data gaps that limit comprehensive environmental analysis. This study introduces a spatial gap-filling method for Himawari-8/Advanced Himawari Imager (AHI) hourly AOD data, using a Random Forest (RF) model that integrates meteorological variables and model-based AOD data. Developed and validated over South Korea from 1 January to 31 December 2019, the model effectively improved data coverage from 6% to 100%. The approach demonstrated high performance in blind tests, achieving a root mean square error (RMSE) of 0.064 and a correlation coefficient (CC) of 0.966. Meteorological analysis indicated optimal model performance under cold, dry conditions (RMSE: 0.047, CC: 0.956), compared to humid conditions (RMSE: 0.105, CC: 0.921). Validation against Aerosol Robotic Network (AERONET) ground observations showed that, while the original Himawari-8 data exhibited higher accuracy (RMSE: 0.189, CC: 0.815, $n = 346$), the gap-filled dataset maintained reasonable precision (RMSE: 0.208, CC: 0.711) and significantly increased the number of valid data points ($n = 4149$). Furthermore, the gap-filled dataset successfully captured seasonal AOD patterns, with values ranging from 0.245–0.300 in winter to 0.381–0.391 in summer, providing a comprehensive view of aerosol dynamics across South Korea.

Keywords: aerosol optical depth (AOD); Himawari-8; gap-filling; machine learning



Citation: Youn, Y.; Kim, S.; Kim, S.H.; Lee, Y. Spatial Gap-Filling of Himawari-8 Hourly AOD Products Using Machine Learning with Model-Based AOD and Meteorological Data: A Focus on the Korean Peninsula. *Remote Sens.* **2024**, *16*, 4400. <https://doi.org/10.3390/rs16234400>

Academic Editors: Muhammad Bilal and Janet E. Nichol

Received: 12 October 2024

Revised: 17 November 2024

Accepted: 20 November 2024

Published: 25 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Aerosols, which are solid and liquid particles suspended in the atmosphere, influence global radiation through both direct effects, such as scattering or absorbing sunlight, and indirect effects, such as altering cloud brightness and longevity [1,2]. Aerosols are typically categorized by size, with Particulate Matter (PM) classified by an aerodynamic diameter of $\leq 10 \mu\text{m}$ (PM₁₀) and finer particulate matter with a diameter of $\leq 2.5 \mu\text{m}$ (PM_{2.5}) [3]. PM reduces visibility, degrades air quality, and poses significant health risks to humans [4,5]. Due to their complex sources and variable atmospheric residence times, aerosols contribute significantly to the uncertainty in climate change predictions [6].

Consequently, quantitative observations of aerosol distributions and physical properties are essential. These observations are conducted using various platforms, including ground-based stations, satellites, ships, and aerial systems [7]. Ground-based observations provide highly accurate data on the physical and chemical properties of aerosols, making them invaluable for validating satellite-derived Aerosol Optical Depth (AOD) products and analyzing aerosol characteristics [8]. However, ground-based observations alone are insufficient for providing a comprehensive understanding of global aerosol characteristics

due to the diverse spatial and temporal distributions of aerosols [9]. As a result, satellite observations have become indispensable for obtaining quantitative aerosol information. These satellite-based measurements are especially valuable because the spatiotemporal variability of aerosols is crucial for understanding their role in climate change and for predicting ground-level PM concentrations in regions lacking direct measurements [10–15].

AOD, the most widely used metric for assessing aerosol characteristics from satellites, quantifies the amount of aerosols present in the atmosphere. It is defined as the integral of aerosol extinction coefficients along the atmospheric column, i.e., from the Earth's surface to the top of the atmosphere [16]. AOD is a dimensionless measure that represents the optical thickness of aerosols, indicating the degree of light attenuation in the atmosphere. Over the years, satellite-based aerosol observations have advanced significantly. Current instruments, such as the Moderate Resolution Imaging Spectroradiometer (MODIS), Multi-angle Imaging Spectroradiometer (MISR), and Visible Infrared Imaging Radiometer Suite (VIIRS), provide valuable insights into global aerosol distributions and their impacts [17–19]. These instruments have been employed in numerous studies that analyze the effects of aerosols on climate change and air quality, estimate ground-level PM concentrations, and assess surface visibility [20–26]. However, many of these sensors are aboard polar-orbiting satellites, which typically offer only one or two daily observations of a specific area. This limited temporal resolution presents challenges in tracking rapid aerosol changes [27], which is essential for conducting atmospheric and environmental research effectively.

Recent advancements in satellite technology have led to the development of next-generation geostationary meteorological satellites, offering significantly improved temporal and spatial resolution [28]. The Himawari-8 satellite, launched in 2014, represented a major advancement in the aerosol observation capabilities over the Asia-Pacific region, including the Korean Peninsula. The Himawari-8 Advanced Himawari Imager (AHI) provides high-frequency observations, producing AOD data at hourly intervals. This capability is especially valuable for studying diurnal variations in aerosol concentrations and their effects on local air quality, addressing the limitations of previous observation systems. The AHI's multi-spectral imaging capabilities also enhance the reliability of AOD retrievals by improving the discrimination between aerosols and clouds [29,30]. We selected Himawari-8 data for this study due to its advanced features, well-established AOD products, and consistent, well-calibrated data since its launch [31]. These characteristics make Himawari-8 a suitable platform for this study, offering high-quality, high-frequency data essential for conducting detailed aerosol analysis over the Korean Peninsula.

Despite advancements in satellite technology, optical sensor-based AOD retrieval still faces challenges in areas with high surface reflectance, cloud or snow cover, and regions with high aerosol concentrations [32,33]. These limitations result in data gaps that can affect the accuracy and availability of environmental analyses [34,35]. To overcome the limitations, researchers have explored various methods for estimating AOD in gap areas, aiming to provide more complete AOD datasets for comprehensive environmental studies. These methods range from traditional statistical approaches to advanced machine learning techniques [36–42]. However, despite progress, most previous studies have focused on low temporal resolution data, such as monthly or daily averages. The need for high temporal resolution data remains crucial for accurate air pollution monitoring, especially in regions with rapidly changing aerosol conditions [27].

Among various approaches, machine learning methods have shown particular promise in addressing the AOD gap-filling challenge. Recent advancements have transitioned from traditional methods to more sophisticated machine learning techniques. Zhao et al. [41] demonstrated the effectiveness of traditional machine learning by applying Random Forest (RF) algorithms to estimate daily AOD over the Beijing-Tianjin-Hebei region. Building on this foundation, deep learning approaches have emerged as powerful alternatives, with Chen et al. [31] successfully implementing deep neural networks for high-frequency AOD estimation. More recently, hybrid approaches have been developed to leverage the strengths of multiple techniques. For example, Chen et al. [42] integrated RF-based

spatial interpolation with deep learning-based temporal estimation, showing the potential of combined methodologies. While these studies have achieved promising results, there remains a need for efficient and practical approaches that balance computational complexity with estimation accuracy, especially for operational applications in regions with unique atmospheric conditions.

In this study, we address the critical need for high-resolution temporal data in aerosol research by developing an advanced machine learning model to fill gaps in hourly AOD products. We utilize data from Himawari-8 AHI, which provides hourly AOD observations across the Asia-Pacific region, including the Korean Peninsula. This high-frequency temporal resolution is essential for capturing the dynamic nature of aerosol distributions and concentrations, overcoming the limitations of previous studies that relied on lower temporal resolution data. Our model incorporates 12 relevant variables, including model-based AOD and meteorological data, to develop a robust gap-filling method for a $0.05^\circ \times 0.05^\circ$ geographic grid over South Korea's land area. We focused on South Korea's land area for several reasons: (1) to contribute to air quality monitoring in regions heavily impacted by human activity, (2) to address the specific challenges posed by the region's complex topography and diverse land cover, and (3) to refine our methodology in a well-defined terrestrial environment before expanding to include marine areas in future studies. This approach not only preserves the frequent temporal resolution required for detailed aerosol analysis but also fills in missing data points, offering a more complete understanding of aerosol dynamics. We conducted our experiments using 2019 data, providing a comprehensive dataset to test and validate our gap-filling model under various seasonal and meteorological conditions typical of the Korean Peninsula. By focusing on hourly data and employing appropriate machine learning techniques, our study aims to overcome the limitations of previous low-temporal resolution studies, potentially enhancing air quality monitoring, improving pollution forecasting, and contributing to a deeper understanding of aerosol impacts on local and regional scales. The remainder of this paper is organized as follows: Section 2 describes the data sources and methodology, including details of the gap-filling model development and validation approach. Section 3 presents the results of the model's performance evaluation and its validation against Aerosol Robotic Network (AERONET) observations across various conditions and locations. Section 4 discusses the model's variable importance, seasonal patterns, and spatiotemporal characteristics of gap-filled AOD data. Finally, Section 5 summarizes the key conclusions and suggests directions for future research.

2. Materials and Methods

2.1. Data

2.1.1. Himawari-8 AOD Data

The Japan Meteorological Agency (JMA) launched Himawari-8, a next-generation geostationary meteorological satellite, on 7 October 2014, with meteorological products becoming available on 7 July 2015 (<https://www.eorc.jaxa.jp/ptree/>, accessed on 10 March 2021). Himawari-8's AHI is equipped with 16 spectral bands. The AOD product is generated using an algorithm developed by the Japan Aerospace Exploration Agency (JAXA) [43], which employs different methodologies for land and ocean retrievals. For land, the focus of this study, the algorithm utilizes five channels (0.47, 0.51, 0.64, 0.86, and 1.6 μm) and the Normalized Difference Vegetation Index (NDVI) to retrieve Level 2 AOD [44]. The process involves Rayleigh atmospheric correction (to remove scattering effects from atmospheric molecules), pixel compositing, and top-of-atmosphere (TOA) reflectance simulation using the System for the Transfer of Atmospheric Radiation (STAR) series [45,46]. A Lookup Table (LUT) is used for efficient calculations, incorporating parameters such as aerosol models for fine and coarse particles [47,48].

In this study, we used the Level 3 AOT_Merged product (a spatiotemporal interpolation of AOT_Pure), derived from Level 2 AOD data through optimal interpolation. This product was chosen over the Level 3 AOT_Pure version (L2 AOT with strict cloud screening) due to its higher quality and lower frequency of missing values [49,50]. The dataset, provided in NetCDF format, includes a 500 nm AOD, the Ångström Exponent (AE), and a Quality Analysis (QA) flag, all mapped onto a 0.05° spatial grid. The 500 nm AOD is derived from the multi-channel measurements of the AHI sensor using the AE [51]. The algorithm for the Level 3 hourly data incorporates aerosol and cloud spatiotemporal variability characteristics for quality control and performs hourly interpolation of the Level 2 AOD products.

Our analysis focused on eight hourly images per day, spanning from 00 UTC to 07 UTC, which corresponds to daylight hours in Korea. This temporal selection aimed to capture diurnal AOD variations while maintaining high data quality. Using this approach, we aim to conduct a comprehensive analysis of AOD patterns and their diurnal fluctuations across the Korean Peninsula. To provide a clear overview of our data and methodology, Table 1 summarizes the key characteristics of the Himawari-8 AOD data used in this study, including details on product resolution, AHI channels, and the AOD conversion method.

Table 1. Description of Himawari-8 AOD data used in this study.

Aspect	Description
Product Used	Level 3 AOT_Merged
Data Parameter	500 nm AOD, Ångström Exponent (AE), Quality Analysis (QA) flag
Spatial Resolution	0.05° × 0.05° grid (in longitude and latitude)
Temporal Resolution	Hourly (Level 3)
Daily Coverage	8 hourly images per day (00 UTC to 07 UTC, corresponding to daylight hours in Korea)
Key AHI Channels Used	0.47 µm, 0.51 µm, 0.64 µm (visible), 0.86 µm, 1.6 µm (near-infrared)
Conversion to 500 nm AOD	Interpolation based on Ångström exponent

2.1.2. CAMS Reanalysis AOD

The Copernicus Atmosphere Monitoring Service (CAMS), a key component of the European Union's Earth observation initiative, provides comprehensive environmental data by integrating modeling, satellite observations, and in-situ measurements. CAMS utilizes the Integrated Forecasting System (IFS), developed by the European Centre for Medium-Range Weather Forecasts (ECMWF). The data assimilation process incorporates aerosol information from various satellite sources, including the Advanced Along-Track Scanning Radiometer (AATSR) aboard the Envisat satellite (2003–2012) and the MODIS instruments on NASA's Terra and Aqua satellites (2003–present) [52,53]. Notably, CAMS does not use bias-adjustment algorithms in its data assimilation process, nor does it incorporate ground-based AOD measurements. This makes ground-based networks such as the AERONET particularly valuable as independent validation sources for CAMS data [54]. For this study, we used AOD values at a 550 nm wavelength from the CAMS reanalysis database, focusing on data from the year 2019. The dataset is provided at a spatial resolution of approximately 80 km (0.75° × 0.75° grid). While CAMS offers 3-hourly data, we aggregated these into daily values. The processed daily AOD data were then downloaded using Google Earth Engine (GEE) (<https://earthengine.google.com/>, accessed on 15 September 2024).

2.1.3. Merra-2 Reanalysis AOD

The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), represents a significant advancement over its predecessor by integrating the MERRA reanalysis dataset with the Gridpoint Statistical Interpolation (GSI) assimilation system and the Goddard Earth Observing System (GEOS) model. This comprehensive AOD product incorporates bias-adjusted data from multiple satellite sensors, including the Advanced Very High-Resolution Radiometer (AVHRR), MODIS, and MISR, as well as ground-based AERONET AOD observations [55,56]. For this study, we utilized MERRA-2 AOD data at a spatial resolution of $0.5^\circ \times 0.625^\circ$. The hourly AOD values were retrieved through GEE.

2.1.4. Meteorological Data

The Local Data Assimilation and Prediction System (LDAPS), a numerical weather prediction model developed by the Korea Meteorological Administration (KMA), is a regionally optimized version of the Unified Model (UM) from the UK Met Office. LDAPS produces forecasts every three hours (at 00, 03, 06, 09, 12, 15, 18, and 21 UTC, corresponding to 09, 12, 15, 18, 21, 00, 03, and 06 KST, respectively) for the Korean Peninsula with a spatial resolution of 1.5 km, using the Lambert Conformal Conic (LCC) projection (<https://data.kma.go.kr/cmmn/main.do>, accessed on 15 September 2024).

The selection of meteorological variables for this study was based on their known influences on atmospheric aerosols and their potential utility in gap-filling AOD data. A comprehensive literature review [36,42,57–60] guided our selection, identifying key factors strongly correlated with AOD or critical to aerosol dynamics. The selected variables include air temperature (TMP, °C) and relative humidity (RH, %), which influence aerosol hygroscopic growth and new particle formation [61]; u-/v-components of wind speed (U_WS/V_WS, m/s), essential for aerosol transport and dispersion; boundary layer height (BLH, m), which affects vertical mixing and aerosol concentration [62]; and latent heat flux (LHFL, W/m^2), influencing atmospheric stability and aerosol vertical distribution [63]. Additionally, we included high cloud cover (HCDC, %) for cloud–aerosol interactions [64–66], downward surface shortwave flux (DSSF, W/m^2), which is affected by AOD and impacts aerosol dynamics [67], surface pressure (PRES, Pa), which affects aerosol vertical distribution and transport, and dew point temperature (DPT, °C), providing supplementary information on atmospheric moisture content.

These LDAPS variables were extracted at 00, 03, 06, and 09 UTC (09, 12, 15, and 18 KST). To align the temporal resolution with the Himawari-8/AHI hourly AOD product, we interpolated the 3-hourly LDAPS forecasts to hourly data using cubic spline interpolation, assuming gradual changes in meteorological variables between forecast times [68,69] (Figure 1). The selection of these specific variables was driven by their relevance to aerosol processes, proven utility in previous AOD studies, and potential to improve AOD gap-filling accuracy. By incorporating this comprehensive set of meteorological variables, we aim to capture the complex interactions between atmospheric conditions and aerosol dynamics, enhancing our ability to accurately estimate AOD in data-sparse regions or periods.

2.2. Methods

2.2.1. Data Preprocessing

The gap-filling experiment was conducted over South Korea, covering the area between 33.73° to 38.72° N and 125.78° to 129.77° E, resulting in a grid of 8000 pixels (100×80) at a 0.05° resolution. To build datasets for the Himawari-8/AHI hourly AOD gap-filling model specifically for the land area of South Korea, we aligned the different spatial and temporal resolutions of various data sources. All datasets were integrated into the Himawari-8/AHI AOD grid ($0.05^\circ \times 0.05^\circ$) across the study region (Table 2).

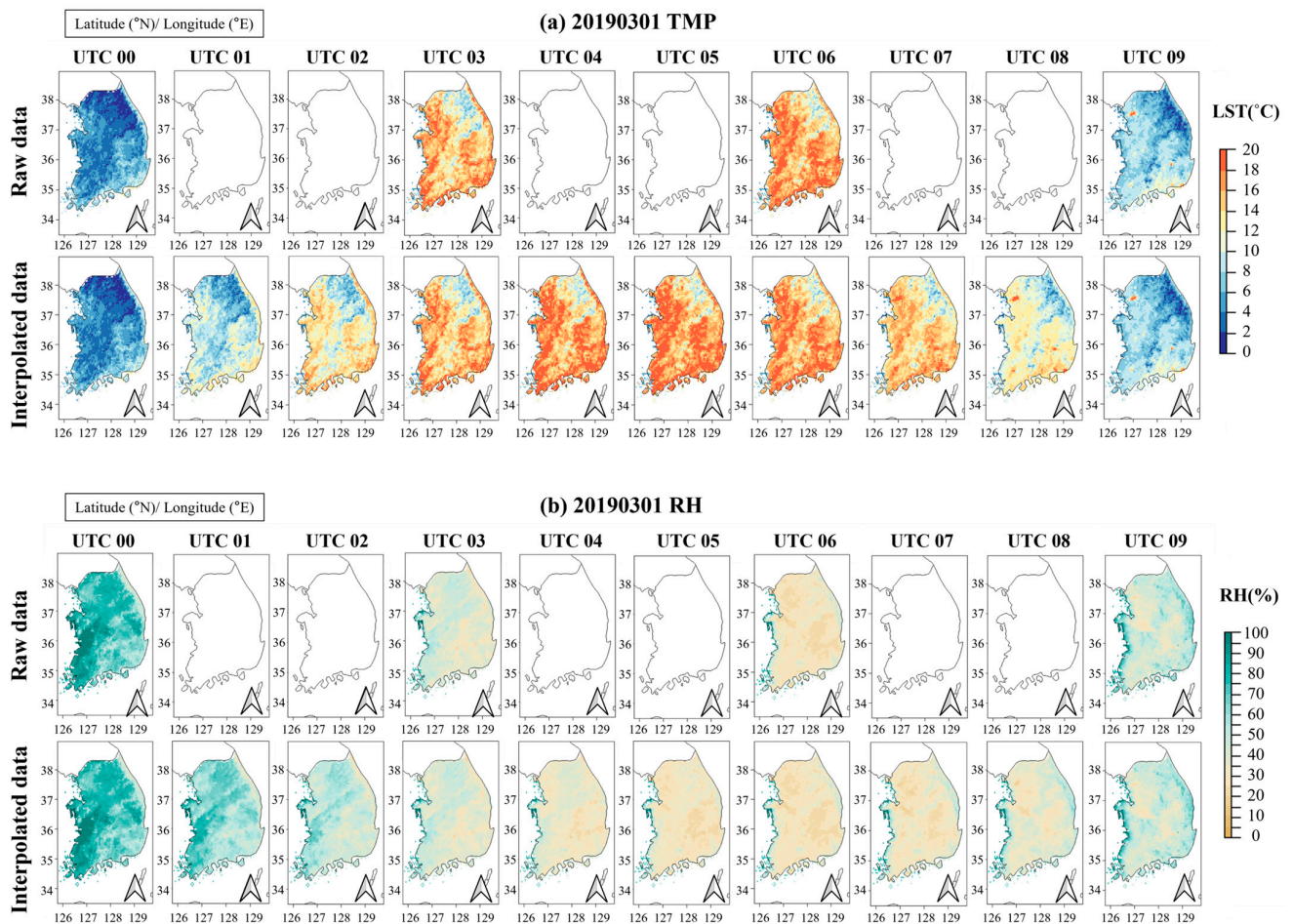


Figure 1. Examples of converting 3-hourly to hourly data using cubic spline interpolation for meteorological variables: (a) before and after temporal interpolation for the TMP on 1 March 2019 and (b) before and after temporal interpolation for the RH on 1 March 2019. The top row in each panel (a,b) shows the raw 3-hourly data, while the bottom row displays the interpolated hourly data. Each column represents a specific UTC time (00 to 09). The color scale indicates the intensity of the variables, with TMP in °C and RH in %. The interpolation process fills the temporal gaps between the 3-hourly observations, resulting in a continuous hourly dataset that matches the temporal resolution of the Himawari-8/AHI AOD product. Notably, the smooth transitions in values over time are evident in the interpolated data, particularly in the previously empty time slots (01, 02, 04, 05, 07, and 08 UTC).

Table 2. Input data used in the AOD gap-filling model: This table presents the spatial and temporal resolutions, data types, and originating organizations for each data source, including Himawari-8/AHI AOD (the target variable for gap-filling), model-based AOD estimates, and meteorological data. The Himawari-8/AHI AOD represents the primary satellite-based measurement that this study seeks to gap-fill.

Data	Source	Spatial Resolution	Temporal Resolution	
Himawari-8/AHI	AOD	JAXA	$0.05^\circ \times 0.05^\circ$	hourly
CAMS	AOD	ECMWF	$0.75^\circ \times 0.75^\circ$	3-hourly
MERRA-2	AOD	NASA	$0.5^\circ \times 0.625^\circ$ (latitude \times longitude)	hourly
LDAPS	Meteorology	KMA	$1.5 \text{ km} \times 1.5 \text{ km}$	3-hourly

The integration process involved several key steps. First, datasets with varying spatial resolutions, including LDAPS meteorological data, CAMS AOD, and MERRA-2 AOD, were resampled to match the 0.05° grid of the Himawari-8/AHI AOD product using bilinear interpolation. Next, LDAPS 3-hourly data were interpolated to an hourly resolution using cubic spline interpolation to match the temporal resolution of the AOD product. In this study, we utilized a combination of daily AOD data from CAMS and hourly AOD data from MERRA-2. This approach of integrating AOD data at different temporal resolutions alongside dynamic meteorological data provides a balance between capturing short-term atmospheric variations and maintaining a comprehensive view of aerosol patterns. By combining daily and hourly AOD data, we leveraged the strengths of both datasets, enabling a more robust analysis of aerosol dynamics across various time scales.

Additionally, only pixels with AOD reliability classified as “very good” according to the QA flag were used as references for model training and validation. The QA flag categorizes AOD reliability into four levels: “very good”, “good”, “marginal”, and “unreliable”. We selected “very good” pixels to ensure the highest data quality and reliability in our model development, as these have undergone rigorous quality checks and are least likely to be affected by cloud contamination or retrieval errors. After integrating the data, a total of 533,286 matched records were obtained.

2.2.2. Random Forest Model for AOD Gap-Filling

Among various machine learning techniques, we selected the RF model for AOD gap-filling due to several advantages. RF excels at capturing complex, non-linear relationships between variables, which is crucial for modeling the intricate interactions between aerosols and meteorological conditions. Its ensemble nature provides robust performance and resistance to overfitting through bagging (bootstrap aggregating) [70–72]. Also, RF demonstrates computational efficiency while maintaining competitive performance compared to more complex models such as deep neural networks.

The mathematical framework of our RF model for AOD gap-filling is based on ensemble learning with multiple decision trees. Each tree generates predictions independently, and the final prediction is obtained by averaging their outputs. For a given set of input variables—including meteorological parameters and model-based AOD—the RF model predicts the missing AOD values using the following mathematical formulation:

$$\hat{y} = \frac{1}{N} \sum_{j=1}^N T_j(X_j) \quad (1)$$

where \hat{y} represents the predicted AOD value, N is the total number of trees in the forest (set to 100 in our implementation), X_j is the vector of input features used for prediction, and T_j denotes the j -th tree through the bootstrap sampling process. Each decision tree T_j is constructed using a different bootstrap sample of the training data, ensuring diversity in the ensemble through random feature selection at each split node.

For gap-filling of Himawari-8/AHI AOD Level 3 hourly data (00, 01, 02, 03, 04, 05, 06, 07 UTC) from 1 January to 31 December 2019, we developed an RF model using 12 variables: CAMS AOD, MERRA-2 AOD, TMP ($^\circ\text{C}$), U_WS (m/s), V_WS (m/s), BLH (m), LHFL (W/m^2), RH (%), HCDC (%), DSSF (W/m^2), PRES (Pa), and DPT ($^\circ\text{C}$). The model was implemented using the `h2o` library in R (Figure 2). To provide optimal balance between model complexity and performance, the model’s hyperparameters were set to $N = 100$ and depth = 20.

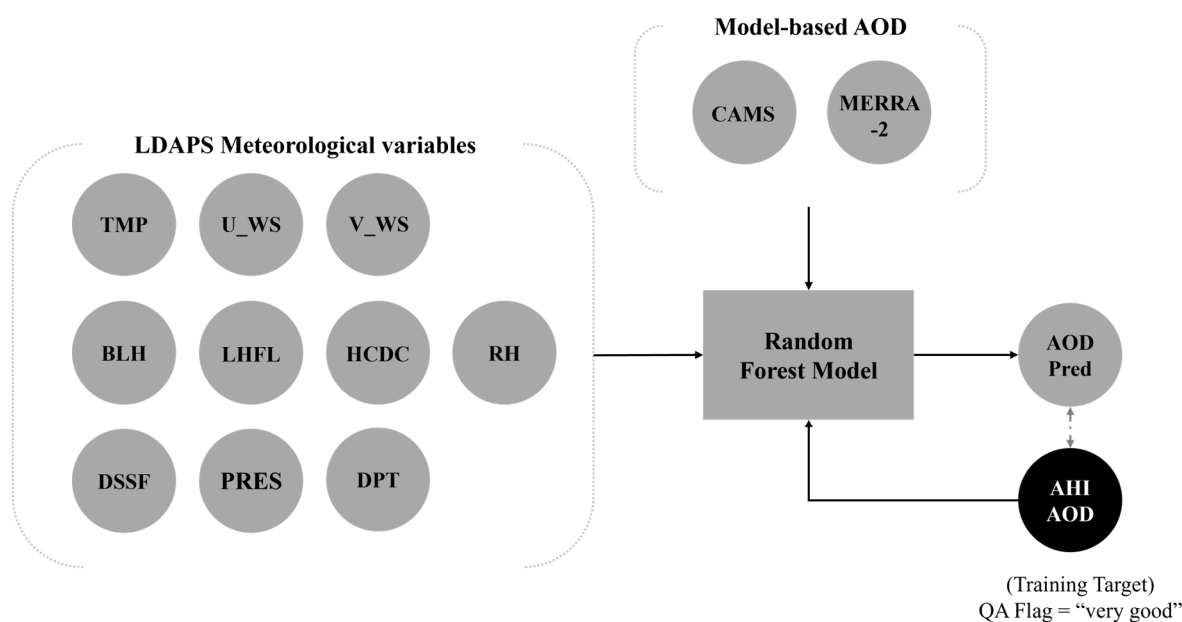


Figure 2. Schematic diagram of the RF model used for gap-filling the Himawari-8/AHI 1-hourly AOD product. The model integrates LDAPS meteorological variables (TMP, U_WS, V_WS, BLH, LHFL, HCDC, RH, DSSF, PRES, DPT) and model-based AOD data (CAMS, MERRA-2) as input features. The output is the predicted AOD (AOD Pred), which is compared with the AHI AOD for quality assurance. Only AHI AOD data classified as “very good” quality were used for model training and validation. Solid lines represent the flow and processing of data, while dashed lines indicate comparisons with reference data.

2.2.3. Model Training and Validation

The model training and verification process consisted of two main components. First, the complete dataset was split into a training set (433,286 records) and an independent test set (100,000 records). During training, the gap-filling model learns the relationship between the input variables and the AHI AOD values classified as “very good” quality, as indicated by the QA flag. These high-quality AHI AOD measurements serve as the target variable (ground truth) for model training. The gap-filling model is trained to predict these AHI AOD values using 12 input variables, which include meteorological factors and model-based AOD data.

The model evaluation involved both 5-fold Cross Validation (CV) and independent test set validation. In the 5-fold CV, the training dataset was randomly divided into five subsets. For each round, one subset was used for evaluation while the model was trained on the remaining four subsets. The results from the five rounds were then averaged to produce the CV outcome. This process helped minimize bias and prevent overfitting during the model development. Second, the independent test set was used to validate the model (Figure 3). This step ensured that the final results were more reliable, as the test set was not involved in the training process and did not influence model development.

For both verification processes, the model’s performance was quantitatively evaluated using the following metrics: Mean Bias Error (MBE, the average difference between predicted and observed values), Mean Absolute Error (MAE, the average magnitude of errors), Root Mean Square Error (RMSE, the square root of the average of squared errors), and Correlation Coefficient (CC, a measure of the linear correlation between predicted and observed values). These metrics provide a comprehensive assessment of the model’s accuracy, precision, and overall performance in gap-filling AOD data, and are defined as follows:

$$MBE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i) \quad (2)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (4)$$

$$\text{CC} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

where x_i and y_i represent the observed and predicted AOD values respectively, \bar{x} and \bar{y} are their respective means, and n is the total number of samples.

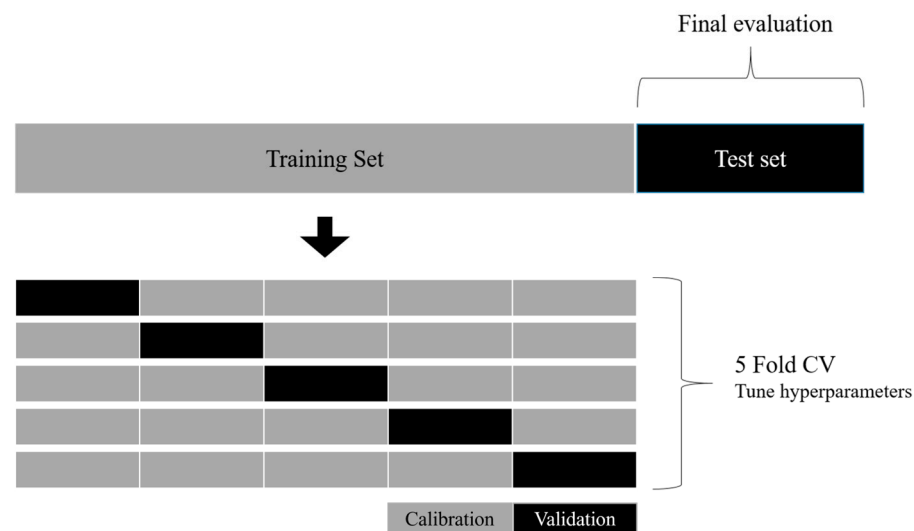


Figure 3. Schematic representation of the model validation process: The validation process includes a training set used for 5-fold CV to tune hyperparameters and a separate test set for final evaluation. The training set is divided into five subsets for CV, with each subset serving as a validation set once, while the remaining four subsets are used for training. The test set is kept entirely separate to independently evaluate the model's performance in the final step.

2.2.4. AERONET Comparison Methodology

To ensure the robustness of our gap-filling results, we conducted additional comparisons with ground-based AERONET observations (<https://aeronet.gsfc.nasa.gov/>, accessed on 1 September 2024). Spatially, we used the Himawari-8 AOD pixel containing the coordinates of each AERONET site. Temporally, all available AERONET measurements within each hour were averaged to align with the hourly temporal resolution of Himawari-8 observations. We used AERONET Level 2.0 data, which are cloud-screened, quality-assured, and fully calibrated, ensuring the highest data quality for validation. Since Himawari-8 provides AOD at 500 nm, direct comparison with AERONET 500 nm AOD was possible without wavelength interpolation.

3. Results

3.1. Gap-Filling Performance

The gap-filling model showed robust performance throughout the experiments. Based on the 5-fold CV results of the training set (81.25% of the data) and the independent blind test (18.75% of the data), our model showed consistent and reliable accuracy in AOD prediction. Table 3 presents the combined results from both the 5-fold CV and the blind test, showing various statistical metrics to provide a comprehensive assessment of model performance. The MBE was 000 for the 5-fold CV and the blind test, indicating that the

model has virtually no systematic bias in its predictions. The MAE values were 0.046 for CV and 0.044 for the blind test, suggesting that, on average, the model's predictions deviated from the observed values by less than 0.05 AOD units. The RMSE values were 0.066 for CV and 0.064 for the blind test, indicating that the typical magnitude of prediction errors were between 0.06 and 0.07 AOD units. The similarity in RMSE values between CV and the blind test demonstrates that the model generalizes well to unseen data. The CC, which measures the strength of the linear relationship between predicted and observed AOD values, was 0.963 for CV and 0.966 for the blind test. These values demonstrate an exceptionally strong positive correlation, with the R-squared (R^2) indicating that the model explains approximately 92.7% to 93.3% of the variance in observed AOD values.

Table 3. Performance metrics comparing the gap-filling model's 5-fold CV results ($n = 433,286$) and independent blind test results ($n = 100,000$).

Validation Method	n	MBE	MAE	RMSE	CC
CV	433,286	0.000	0.046	0.066	0.963
Blind test	100,000	0.000	0.044	0.064	0.966

Figure 4a,b present density scatter plots comparing the gap-filling model's predicted AHI hourly AOD values with observed AOD values for both the 5-fold CV set (number of samples (n) = 433,286) and the blind test set ($n = 100,000$). The plots illustrate the distribution of data points along the 1:1 line, with color gradients representing the density of points. Fitted line equations and performance metrics are displayed on each plot, confirming the values reported in Table 3.

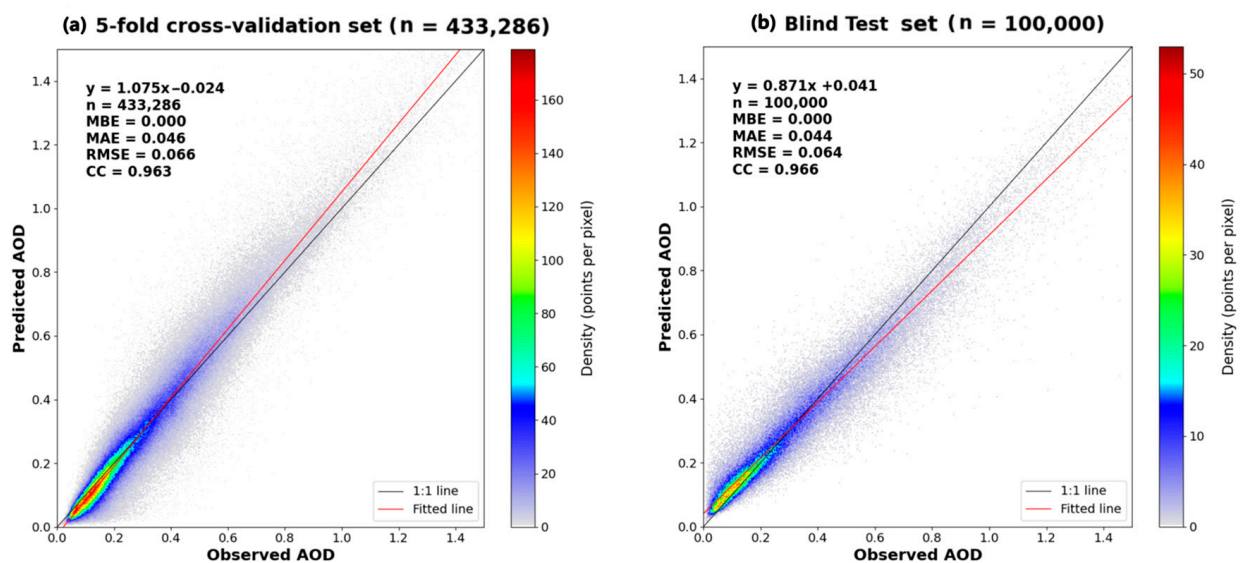


Figure 4. Density scatter plots comparing observed versus predicted AHI hourly AOD values using the gap-filling model with all features for (a) the 5-fold CV set ($n = 433,286$) and (b) the blind test set ($n = 100,000$). The color scale represents the number of data points per pixel, with warmer colors indicating higher densities. The 1:1 line (indicating perfect prediction) is in black, while the red line represents the fitted regression line. Statistical metrics (MBE, MAE, RMSE, CC) and the fitted line equations are provided in the top left corner of each plot.

For the 5-fold CV results (Figure 4a), the fitted line equation is $y = 1.075x - 0.024$, suggesting a nuanced relationship between predicted and observed AOD values. The slope greater than 1 (1.075) indicates that the model tends to overestimate AOD for higher values, while the small negative intercept (-0.024) implies a slight underestimation for very low AOD values, creating a crossover point where the model transitions from underestimation

to overestimation. For the blind test validation results (Figure 4b), the fitted line equation is $y = 0.871x + 0.041$, indicating that the model slightly underestimates AOD for higher observed values (as shown by the slope less than 1, 0.871). The positive intercept (0.041) suggests a small positive bias for very low AOD values. Both plots demonstrate a strong correlation between predicted and observed AOD values, with most points clustering along the 1:1 line, particularly for lower AOD values. The density of points decreases at higher AOD levels, indicating fewer observations at those values. The slight differences in the fitted line equations between CV and blind test results reflect some variability in model performance across different datasets, which is common in statistical modeling. Despite this, both results indicate a strong overall predictive capability, as evidenced by the high CC values (0.963 and 0.966, respectively) and low error metrics.

3.2. Model Behavior and Accuracy Under Different Conditions

3.2.1. Seasonal Variations in AOD Estimation

The model's performance showed notable variations across seasons, as illustrated in Table 4. In spring, the model achieved the highest CC of 0.970, along with a relatively low RMSE of 0.068. Summer exhibited the lowest CC (0.939) but also the highest RMSE (0.082) among all seasons. In fall, the model performed strongly, with a CC of 0.964 and the second-lowest RMSE (0.058). The winter performance was similarly robust, with a CC of 0.946 and the lowest RMSE (0.051). In terms of bias, all seasons displayed very low MBE values, ranging from -0.002 in winter to 0.002 in spring, indicating minimal systematic errors across seasons. MAE values were consistently low, with winter having the lowest (0.036) and summer the highest (0.060). The full-year statistics (CC: 0.966, RMSE: 0.064, MAE: 0.044, MBE: 0.000; Table 3) suggest that the model performs well overall, balancing seasonal variations. These results demonstrate that while the model's performance varies slightly by season, it maintains a high level of accuracy throughout the year.

Table 4. Seasonal performance metrics of the gap-filling model for AOD prediction, based on the blind test set ($n = 100,000$) using selected features. The metrics include then, MBE, MAE, RMSE, and CC for each season, as well as for the full year.

Season	n	MBE	MAE	RMSE	CC
Spring	45,091	0.002	0.048	0.068	0.970
Summer	11,627	0.000	0.060	0.082	0.939
Fall	10,625	0.001	0.038	0.058	0.964
Winter	32,657	-0.002	0.036	0.051	0.946

Notably, despite the substantial difference in sample sizes between spring ($n = 45,091$) and fall ($n = 10,625$), both seasons demonstrated comparable model accuracy, with strong CCs of 0.970 and 0.964, and similar RMSE values of 0.068 and 0.058, respectively. This suggests that our gap-filling model maintains stable predictive power, highlighting the robustness of the selected feature set in capturing seasonal AOD patterns.

3.2.2. Model Performance by Missing Pixel Rates and AOD Ranges

For a more comprehensive evaluation, we examined model performance based on the rates of missing pixels. Within the randomly selected test dataset of 100,000 cases, missing pixel rates were classified into two categories, low ($<60\%$) and high ($>60\%$), to represent different levels of data availability. The low missing rate group comprised 37,707 cases, while the high missing rates group included 62,293 cases. Both groups showed similar performance metrics, with an MAE of 0.069 and CC of 0.941 for low missing rate conditions and an MAE of 0.058 and CC of 0.930 for high missing rate conditions (Table 5).

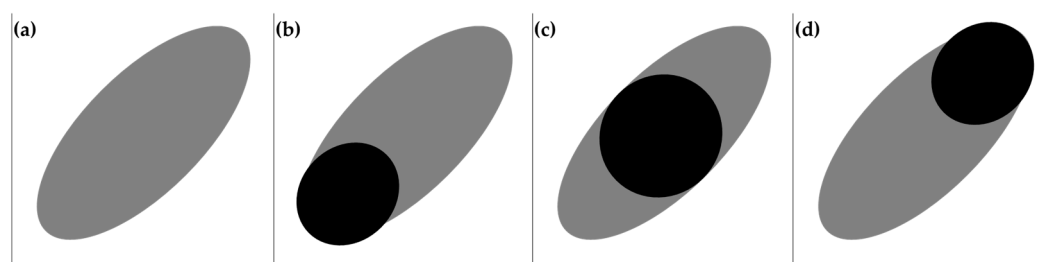
Table 5. Performance metrics for the gap-filling model across different missing pixel rates.

Missing Pixel Rates	n	MBE	MAE	RMSE	CC
Low (<60%)	37,707	−0.003	0.069	0.097	0.941
High (>60%)	62,293	0.008	0.058	0.083	0.930

Further analysis was conducted based on AOD ranges. AOD values were divided into low (<0.3), medium (0.3–0.8), and high (>0.8) categories, following widely used thresholds in aerosol studies. From the test dataset of 100,000 cases, most data points fell into the low (n = 58,550) and medium (n = 36,945) AOD ranges, with fewer samples in the high AOD range (n = 4505). Model performance was relatively poor for the high AOD group, with a higher MAE (0.166) compared to the low and medium groups (0.052 and 0.068, respectively) (Table 6). However, this discrepancy arises because the high AOD category, with values reaching up to 1.5 or higher, tends to exhibit larger errors due to the scale dependency of MAE. To mitigate this limitation of MAE, we calculated the normalized RMSE (NRMSE) by dividing the RMSE by the mean values. As a result, the NRMSE for the low AOD range was 47.0, while those for medium and high AOD ranges were 19.1 and 20.9, respectively. Estimating low AOD values proved more challenging than for medium and high values, reflecting similar challenges encountered in satellite retrieval of low AOD values. When AOD values are low, the aerosol concentration in the atmosphere is minimal, resulting in a weak aerosol signal. This low signal-to-noise ratio (SNR) makes it challenging to distinguish the aerosol signal from background signals, such as molecular scattering or surface reflection. The CC values were relatively lower for the three groups (low, medium, and high) compared to the entire test dataset. This is expected, as subsets divided by value range tend to exhibit a more dispersed pattern in scatterplots compared to the full dataset containing all values (Figure 5).

Table 6. Performance metrics for the gap-filling model across different AOD ranges. The normalized RMSE (NRMSE) was calculated by dividing the RMSE by the mean values.

AOD	n	MBE	MAE	NRMSE	CC
Low (<0.3)	58,550	0.042	0.052	47.0	0.711
Medium (0.3–0.8)	36,945	−0.046	0.068	19.1	0.800
High (>0.8)	4505	−0.163	0.166	20.9	0.784

**Figure 5.** Difference in Correlation Coefficient (CC) between the entire test dataset and the subsets divided by value ranges. A higher CC is observed for (a) the entire test dataset, while lower CC values are observed for (b) the low-value group, (c) the medium-value group, and (d) the high-value group. Black circles represent the portions of data corresponding to each group, while gray ellipses show the entirety of the data.

3.2.3. Meteorological Effects on AOD Estimation

This section examines how meteorological conditions influence the accuracy of AOD estimation, focusing on three meteorological variables (DPT, DSSF, and TMP) that showed significant influence on the model performance (Table 7). For a detailed evaluation, we

divided each variable into three ranges based on their meteorological significance. For DPT, we categorized conditions as low ($<-10\text{ }^{\circ}\text{C}$), representing dry atmospheric conditions with minimal water vapor; medium ($-10\text{ to }0\text{ }^{\circ}\text{C}$), representing intermediate moisture levels; and high ($>0\text{ }^{\circ}\text{C}$), indicating humid conditions with substantial atmospheric water vapor content. DSSF was divided into low ($<300\text{ W/m}^2$) for cloudy or twilight conditions, medium ($300\text{--}600\text{ W/m}^2$) for typical daytime, and high ($>600\text{ W/m}^2$) for clear-sky, peak solar radiation. TMP ranges were set as low ($<0\text{ }^{\circ}\text{C}$) for below-freezing conditions, medium ($0\text{--}20\text{ }^{\circ}\text{C}$) for moderate conditions, and high ($>20\text{ }^{\circ}\text{C}$) for warm conditions.

Table 7. Summary of gap-filling model performance metrics across different ranges of meteorological variables (DPT, DSSF, and TMP).

Variable	Range	n	MBE	MAE	RMSE	CC
DPT	Low	15,551	0.001	0.033	0.047	0.956
	Medium	45,894	−0.001	0.062	0.087	0.940
	High	38,555	0.000	0.077	0.105	0.921
DSSF	Low	339	−0.009	0.049	0.074	0.964
	Medium	46,348	−0.001	0.057	0.083	0.933
	High	53,313	0.001	0.068	0.096	0.938
TMP	Low	2324	0.001	0.023	0.033	0.943
	Medium	65,668	0.000	0.058	0.083	0.934
	High	32,008	−0.001	0.077	0.106	0.929

The performance metrics across these ranges reveal several key patterns. DPT shows a clear performance gradient, with the highest accuracy in cold, dry conditions (RMSE: 0.047, CC: 0.956) and gradually decreasing performance in more humid conditions (RMSE: 0.105, CC: 0.921). This pattern suggests more reliable AOD estimation in dry atmospheric conditions. The DSSF-based analysis indicates optimal performance in low-radiation conditions (CC: 0.964); however, the small sample size ($n = 339$) calls for careful interpretation. The performance remains stable across medium and high radiation conditions (CC: 0.933–0.938). TMP-based metrics show excellent performance in low-temperature conditions (RMSE: 0.033, MAE: 0.023), with slightly reduced accuracy at higher temperatures, aligning with the DPT pattern and suggesting generally better model performance in cold conditions. Across all variables, MBE values remain consistently near zero (−0.009 to 0.001), and high correlation coefficients ($\text{CC} > 0.92$) are maintained across all meteorological ranges, demonstrating the model's robust performance under various atmospheric conditions.

3.3. Comparisons with AERONET Observations

We compared the results of our gap-filling model with ground-based AERONET observations from six sites across South Korea, representing urban, forest, and coastal environments (Figure 6). The validation demonstrated that our gap-filling method substantially improved data availability while largely maintaining data quality. The number of valid data points increased from 346 to 4149, with a CC of 0.815 for the original data and 0.711 for the gap-filled data. Figure 7a shows only the original Himawari-8 data points matched with AERONET observations, and 7b represents the comprehensive gap-filled dataset excluding the original data points, specifically to evaluate the performance of the gap-filling model. While the MAE remained stable at 0.156, the RMSE increased slightly from 0.189 to 0.208. This is because the gap-filling model inherently introduces prediction uncertainty due to the indirect relationships between predictor variables and AOD. Additionally, the original dataset includes only clear-sky conditions, where satellite measurements were feasible, whereas the gap-filled dataset encompasses predictions for conditions that prevented satellite observations.

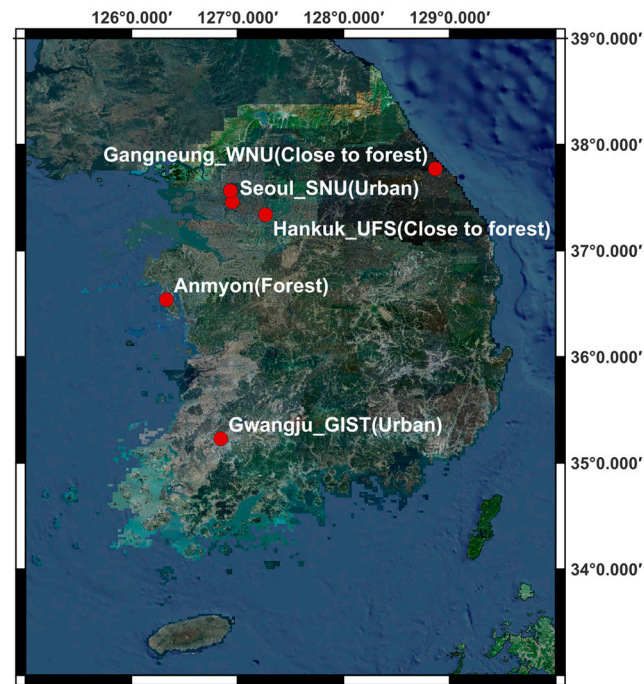


Figure 6. Geographical distribution of the six AERONET sites used for validation in South Korea. The sites include Gangneung_WNU (forest-adjacent), Seoul_SNU (urban), Hankuk_UFS (forest-adjacent), Anmyon (forest), and Gwangju_GIST (urban). The map illustrates the diverse environmental settings of these sites, covering urban areas, forests, and coastal regions across the Korean Peninsula.

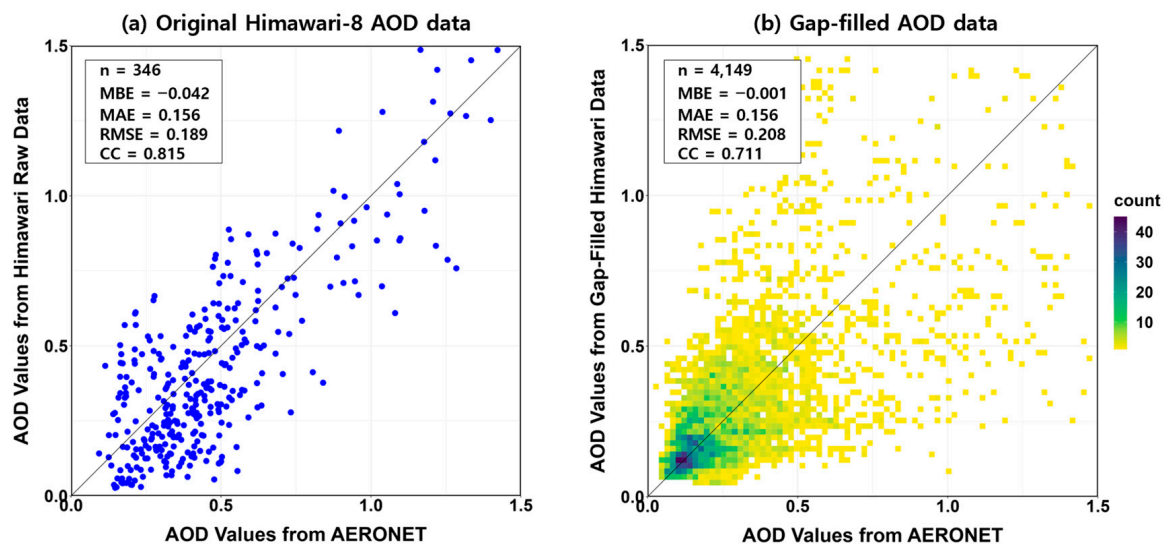


Figure 7. Scatter plots comparing (a) original Himawari-8 AOD data (only matched pairs between Himawari-8 and AERONET observations) and (b) gap-filled AOD data (only predicted values from the gap-filling model, excluding original Himawari-8 data) against AERONET AOD observations. Plot (a) was represented as a scatter plot because of smaller samples ($n = 346$), whereas plot (b) was illustrated as a density plot due to larger sample size ($n = 4149$) with blue tones indicating higher densities. The 1:1 line is shown in black. Statistical metrics, including MBE, MAE, RMSE, and CC, are also provided in the top left corner.

Seasonal analysis demonstrated varying performance across different times of the year (Table 8). The most substantial increase in data availability occurred in winter (40 to 1213 points), with minimal changes in error metrics (RMSE: 0.167 to 0.170; CC: 0.599 to 0.561). Fall showed notable improvements, with RMSE decreasing from 0.198 to 0.134

and CC increasing from 0.310 to 0.613. However, spring and summer exhibited some trade-offs, with increased RMSE and decreased CC, likely due to complex atmospheric conditions including Asian dust events and monsoon activity. The better performance in fall can be attributed to more stable atmospheric conditions, while the relatively poor performance in spring and summer reflects the challenges of predicting AOD values under complex conditions.

Table 8. Seasonal and site-specific validation of original and gap-filled AOD data against AERONET observations.

Category	Subcategory	Data Type	n	MBE	MAE	RMSE	CC
Seasonal	Spring	Original	197	−0.101	0.169	0.203	0.872
		Gap-filled	1152	−0.047	0.172	0.270	0.640
	Summer	Original	68	−0.001	0.124	0.149	0.768
		Gap-filled	1016	−0.025	0.194	0.295	0.653
	Fall	Original	41	0.125	0.161	0.198	0.310
		Gap-filled	768	0.096	0.114	0.134	0.613
	Winter	Original	40	0.006	0.144	0.167	0.599
		Gap-filled	1213	−0.010	0.109	0.170	0.561
Site-specific	Anmyon (Non-Urban/Forest)	Original	60	0.042	0.121	0.143	0.880
		Gap-filled	757	0.030	0.168	0.240	0.721
	Gangneung_WNU (Non-Urban/Close to Forest)	Original	4	0.080	0.092	0.114	−0.634
		Gap-filled	840	0.053	0.102	0.135	0.486
	Gwangju_GIST (Urban)	Original	30	−0.187	0.234	0.271	0.850
		Gap-filled	211	−0.044	0.133	0.230	0.670
	Hankuk_UFS (Non-Urban/Close to Forest)	Original	116	−0.096	0.144	0.169	0.880
		Gap-filled	777	−0.068	0.177	0.276	0.620
	Seoul_SNU (Urban)	Original	71	−0.029	0.159	0.194	0.819
		Gap-filled	492	−0.034	0.151	0.254	0.614
Yonsei_University (Urban)	Original	65	0.022	0.175	0.210	0.697	
	Gap-filled	1072	−0.006	0.151	0.237	0.679	

Site-specific analysis revealed distinct patterns across different environments. Data availability increased substantially across all sites, with non-urban areas showing increases ranging from 6.7-fold (Hankuk_UFS: 116 to 777) to 210-fold (Gangneung_WNU: 4 to 840), and urban areas demonstrating 7- to 16.5-fold increases (e.g., Yonsei_University: 65 to 1072). Performance metrics varied by site type: non-urban sites showed more variable results, with Gangneung_WNU demonstrating marked improvement from negative correlation (CC = −0.634, n = 4) to positive correlation (CC = 0.486, n = 840), while Anmyon and Hankuk_UFS showed increased RMSE (from 0.143 to 0.240 and from 0.169 to 0.276, respectively) and decreased CC (from 0.880 to 0.721 and from 0.880 to 0.620, respectively) despite bias improvements. Urban sites generally maintained stable error metrics with moderate decreases in CCs. Notably, the Gwangju_GIST site showed a lower RMSE (0.230) for gap-filled results compared to the RMSE (0.271) of the original data. Although uncommon, this can occur when the AOD range is relatively narrow.

These findings indicate that while our gap-filling method successfully addresses data continuity issues, its performance varies by season and location. The substantial increase in data availability across all sites suggests the method's effectiveness, though users should consider the seasonal and site-specific variations in accuracy when applying the gap-filled data.

4. Discussion

4.1. Variable Importance Analysis

Our gap-filling model demonstrated robust performance across both CV and blind test evaluations. Analysis of variable importance (Figure 8) revealed that satellite-derived AOD products played crucial roles in AOD predictions, with CAMS AOD and MERRA-2 AOD contributing 27.40% and 24.17% of the model's predictive power, respectively. These two model-derived AOD variables collectively accounted for approximately half (51.57%) of the total variable importance, underscoring the value of integrating multiple AOD data sources in the gap-filling process.

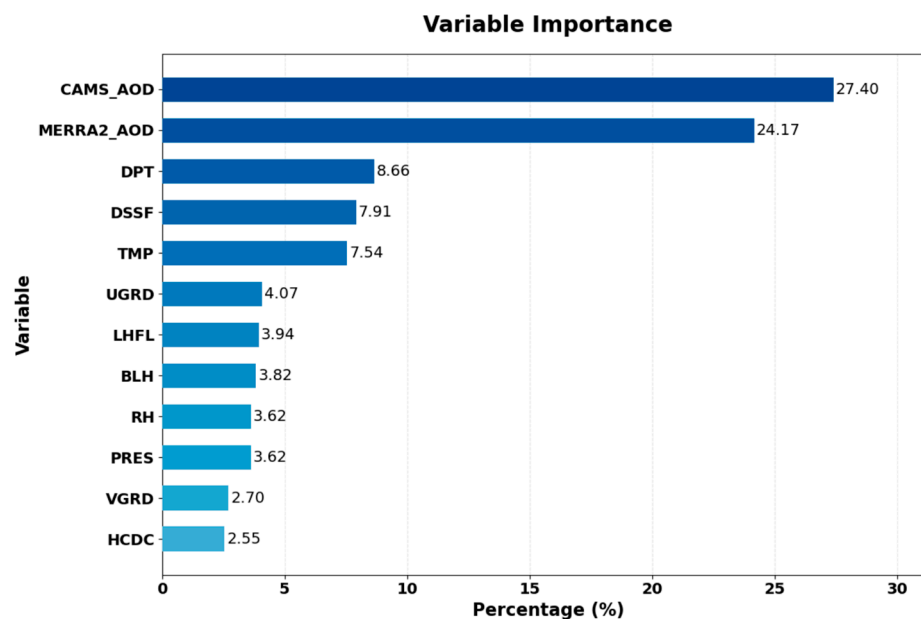


Figure 8. Variable importance of the RF model for AOD gap-filling.

The remaining 48.43% of importance is attributed to meteorological variables. Among these, DPT showed the highest importance (8.66%), indicating atmospheric water vapor content, which significantly affects aerosol hygroscopic growth, especially during Korea's humid seasons. Notably, DPT ranked higher than RH (3.62%), suggesting that the moisture state indicated by DPT may be more critical for AOD prediction in this region than relative humidity. DSSF (7.91%) and TMP (7.54%) followed in importance, reflecting their significant impacts on photochemical reactions and atmospheric mixing processes, respectively. This aligns with our understanding of aerosol dynamics, where TMP influences chemical reaction rates and atmospheric mixing, while solar radiation drives photochemical reactions and affects atmospheric stability. These three variables collectively contribute 24.11% to the model's performance, emphasizing the significant influence of temperature and radiation factors in AOD estimation.

Wind components (UGRD and VGRD), BLH, LHFL, and PRES showed moderate importance (3.62–4.07%), suggesting their roles in aerosol transport, vertical distribution, and broader weather patterns influencing AOD, though to a lesser extent than the top factors. HCDC showed the lowest importance (2.55%), suggesting that in the Korean context, other factors have a more direct influence on AOD.

Comparing our results with similar studies in East Asia, we found consistencies in the importance of certain variables, particularly meteorological factors. For instance, a study on MODIS AOD gap-filling in the Beijing-Tianjin-Hebei region [40] also reported high importance for variables such as DPT and DSSF. This consistency across regions highlights the fundamental role of meteorological variables in aerosol dynamics and AOD prediction. Overall, this analysis provides insights into the relative importance of various meteorological factors in predicting AOD over Korea, reflecting the region's unique climate and

aerosol characteristics. It highlights the complex nature of aerosol–meteorology interactions and emphasizes the need for comprehensive consideration of multiple variables in AOD prediction models for this region.

4.2. Methodological Considerations for AOD Prediction

Our study employed a unique approach to manage input variables with varying temporal and spatial resolutions. We processed meteorological data from LDAPS to generate hourly values at a high spatial resolution, effectively capturing rapid changes in atmospheric conditions. While CAMS AOD data were available at a daily resolution and MERRA-2 AOD data provided hourly information at coarser spatial resolutions compared to LDAPS, both datasets were integral to our model due to their significant predictive value.

Model-based AOD products such as CAMS and MERRA-2 inherently contain systematic biases due to differing modeling approaches, assumptions, and data assimilation methods. As shown in Figure 9, the density scatterplot between MERRA-2 and CAMS AOD demonstrates relatively good agreement, with a CC of 0.756. The mean bias difference (MBD) of -0.019 and mean absolute difference (MAD) of 0.095 indicate that, while there are some systematic differences between the two products, they are not substantial. This moderate level of agreement suggests that both products provide complementary information about aerosol distributions. Instead of directly combining or averaging AOD values from different products, our model was designed to learn the complex, non-linear relationships between these products and the target Himawari-8 AOD values. The model treats CAMS and MERRA-2 AOD as independent input features, allowing it to automatically weight their contributions based on predictive power under different conditions. This approach enables the model to account for and reconcile systematic biases between products while preserving their unique informational content.

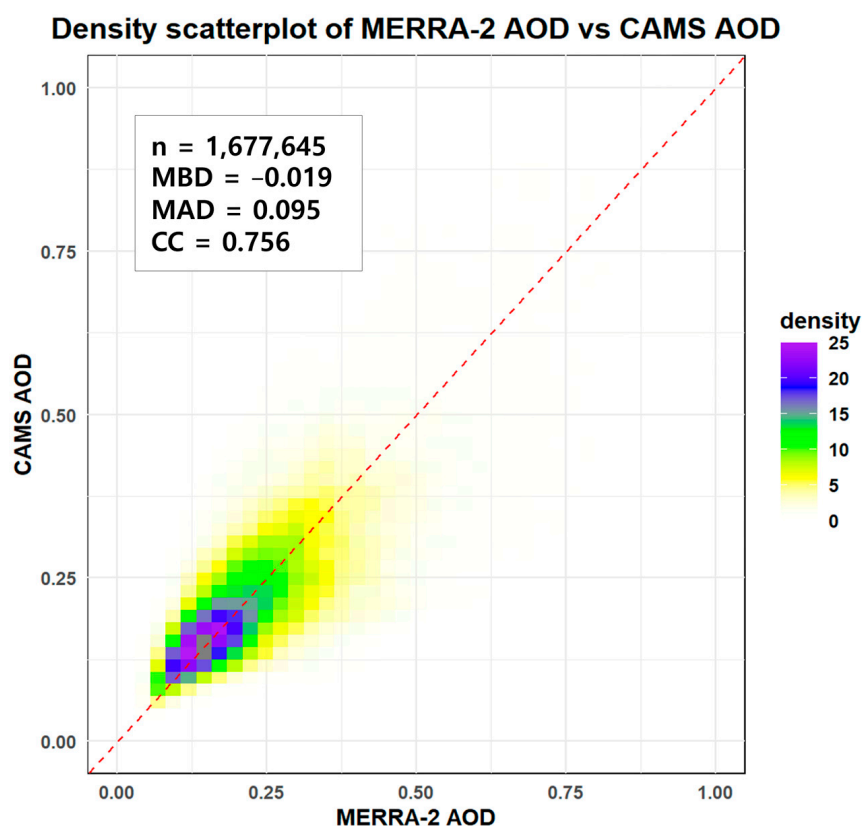


Figure 9. Comparison of daily MERRA-2 and CAMS AOD values for 2019, with the dashed red line representing the 1:1 line.

Through the integration of these diverse data sources, our model produced high-resolution AOD predictions. This approach is crucial for capturing the dynamic nature of aerosol distributions, which can vary significantly over short periods and small spatial scales. The LDAPS data allowed us to downscale the coarse AOD inputs from CAMS and MERRA-2, enabling more detailed predictions in complex terrains and urban areas. By incorporating AOD data from both sources along with dynamic meteorological data, we achieved a balance between capturing short-term atmospheric variations and maintaining a comprehensive view of aerosol patterns.

However, it is important to acknowledge the potential limitations of this approach. One consideration is the temporal interpolation of meteorological data, specifically the conversion of 3-hourly LDAPS data to hourly values using cubic spline interpolation. While necessary to match the temporal resolution of our AOD predictions, this process may introduce some uncertainties. In cases where meteorological conditions fluctuate rapidly within the 3-h intervals, our hourly interpolations may not fully capture these short-term variations. This could affect AOD predictions, particularly during periods of high atmospheric instability or unusual weather events. The assumption of relatively smooth transitions between the 3-hourly data may not always perfectly reflect real-world conditions, potentially impacting the model's accuracy in certain scenarios.

4.3. Gap-Filling Results and Statistics Summary

As shown in Table 9, prior to gap-filling, the data exhibited extremely high null pixel ratios ranging from 89.5% to 97.3% across seasons, which significantly limits the reliability of direct statistical comparisons between before and after gap-filling periods. However, the gap-filling process successfully achieved 100% coverage while maintaining physically reasonable AOD patterns.

Table 9. Seasonal statistics of Himawari-8/AHI 1-hourly AOD before and after gap-filling.

Season	Before Gap-Filling					After Gap-Filling				
	Min	Mean	Max	Standard Deviation	Null Pixel Ratio of Land (%)	Min	Mean	Max	Standard Deviation	Null Pixel Ratio of Land (%)
Spring	0.004	0.340	2.371	0.069	89.5	0.004	0.363	2.371	0.094	0.0
Summer	0.009	0.370	1.719	0.085	97.2	0.009	0.357	1.780	0.101	0.0
Fall	0.005	0.247	1.969	0.115	97.3	0.005	0.290	1.969	0.071	0.0
Winter	0.011	0.224	1.212	0.059	92.1	0.011	0.270	1.250	0.068	0.0
Full-year	0.004	0.298	2.371	0.082	94.0	0.004	0.312	2.371	0.084	0.0

After gap-filling, the seasonal AOD patterns aligned well with known atmospheric phenomena in the region. The spring season showed the highest maximum AOD value (2.371), likely corresponding to Asian dust events that typically occur during this season. The high summer values (mean: 0.357) reflect increased humidity and secondary aerosol formation, while the lower winter values (mean: 0.270) are consistent with typical seasonal patterns. The model's ability to capture these expected seasonal variations, including extreme events, while maintaining high CCs (0.970 for spring, 0.939 for summer; Table 4), suggests successful gap-filling performance.

The monthly analysis reveals crucial temporal patterns that are masked in seasonal aggregation (Table 10). While the seasonal analysis indicated spring as the season with the highest mean AOD, the monthly data showed that the highest monthly means occurred in June (0.391) and March (0.382), offering more precise insights into aerosol concentration patterns. The extreme null pixel ratios in July to September (98.9–99.0%) pinpoint the exact timing of cloud interference, likely associated with the summer monsoon period. Additionally, the monthly analysis captured sharp seasonal transitions, such as the significant increase in mean AOD from May (0.354) to June (0.391), followed by a decrease from November (0.283) to December (0.245). These abrupt changes, which are smoothed

out in seasonal averages, provide valuable insights for understanding rapid atmospheric transitions and improving predictions for high-aerosol episodes.

Table 10. Monthly statistics of Himawari-8/AHI hourly AOD before and after gap-filling.

Month	Before Gap-Filling					After Gap-Filling				
	Min	Mean	Max	Standard Deviation	Null Pixel Ratio of Land (%)	Min	Mean	Max	Standard Deviation	Null Pixel Ratio of Land (%)
January	0.024	0.224	0.822	0.060	95.2	0.024	0.264	1.020	0.067	0.0
February	0.012	0.230	1.210	0.068	91.5	0.012	0.300	1.250	0.078	0.0
March	0.006	0.378	2.371	0.100	88.3	0.006	0.382	2.370	0.100	0.0
April	0.004	0.304	2.080	0.099	86.7	0.004	0.354	2.080	0.103	0.0
May	0.007	0.338	1.250	0.116	93.2	0.007	0.354	1.250	0.082	0.0
June	0.009	0.422	1.700	0.133	93.4	0.009	0.391	1.700	0.102	0.0
July	0.016	0.416	1.780	0.100	99.0	0.016	0.381	1.780	0.117	0.0
August	0.019	0.292	1.290	0.075	98.9	0.019	0.300	1.290	0.085	0.0
September	0.007	0.223	0.966	0.075	98.9	0.007	0.240	0.966	0.067	0.0
October	0.008	0.264	1.970	0.085	98.3	0.008	0.346	1.970	0.079	0.0
November	0.005	0.242	1.550	0.065	94.6	0.005	0.283	1.550	0.067	0.0
December	0.011	0.218	1.120	0.056	89.3	0.011	0.245	1.120	0.060	0.0
Full-year	0.004	0.298	2.371	0.082	94.0	0.004	0.312	2.371	0.084	0.0

4.4. Spatio-Temporal Analysis of Gap-Filled AOD Data

Our gap-filling approach achieved complete spatial coverage, improving the AOD coverage rate from 6% to 100% for the 2019 AHI hourly AOD product. The resulting hourly AOD maps (Figure A1) captured detailed diurnal variations in aerosol patterns, enabling the observation of critical air quality events. A notable example is the severe pollution episode during 1–6 March 2019, when PM_{2.5} concentrations exceeded 50 µg/m³ in the Seoul metropolitan area, leading to emergency reduction measures. The continuous hourly data revealed the temporal evolution of this event, demonstrating the value of high-frequency observations for air quality monitoring and management.

Analysis of the monthly mean AOD (Figure 10) revealed distinct seasonal and spatial patterns. Winter months (December–February) showed relatively low AOD values (0.245–0.300), while spring months (March–April) exhibited moderate to high values (0.354–0.382). Summer months (June–July) consistently showed the highest AOD levels (0.381–0.391), followed by a sharp decrease to the annual minimum (0.240) in early autumn (September). The spatial distribution of AOD showed strong correlations with land use patterns, with consistently high AOD values observed in metropolitan areas (Seoul, Incheon, Busan, Gwangju, Daejeon, and Ulsan), coastal regions, and agricultural areas. In contrast, forested regions, particularly Gangwon-do, maintained persistently low AOD levels throughout the year. These patterns suggest that anthropogenic activities and land use characteristics significantly influence local aerosol distributions. We recognize that incorporating additional geospatial and anthropogenic factors (e.g., land use data, population density, traffic volume) could further enhance our understanding and improve AOD estimations. The enhanced temporal resolution of our gap-filled product enables better understanding of both short-term pollution events and long-term aerosol patterns, providing valuable insights for air quality management and policy development.

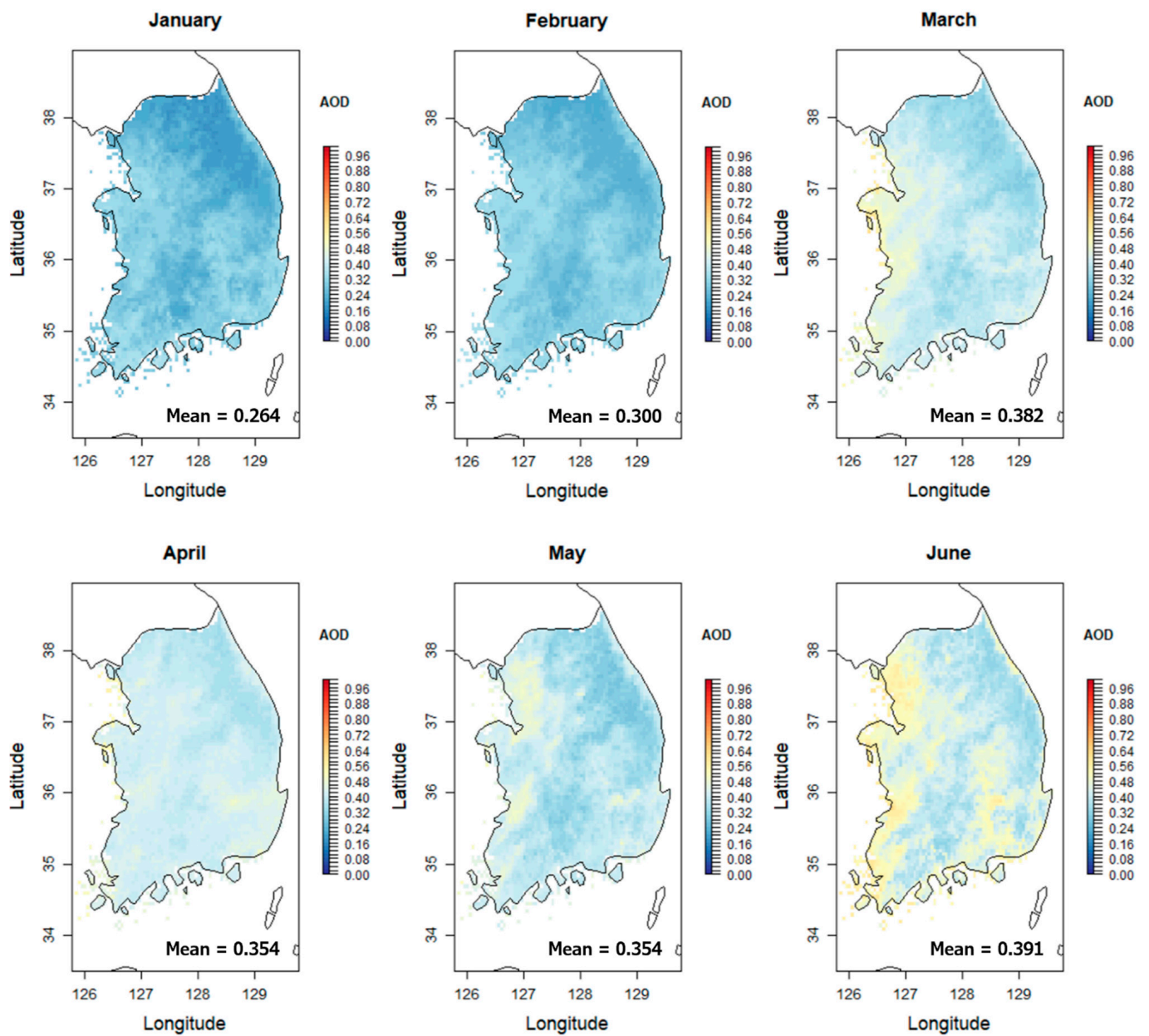


Figure 10. Cont.

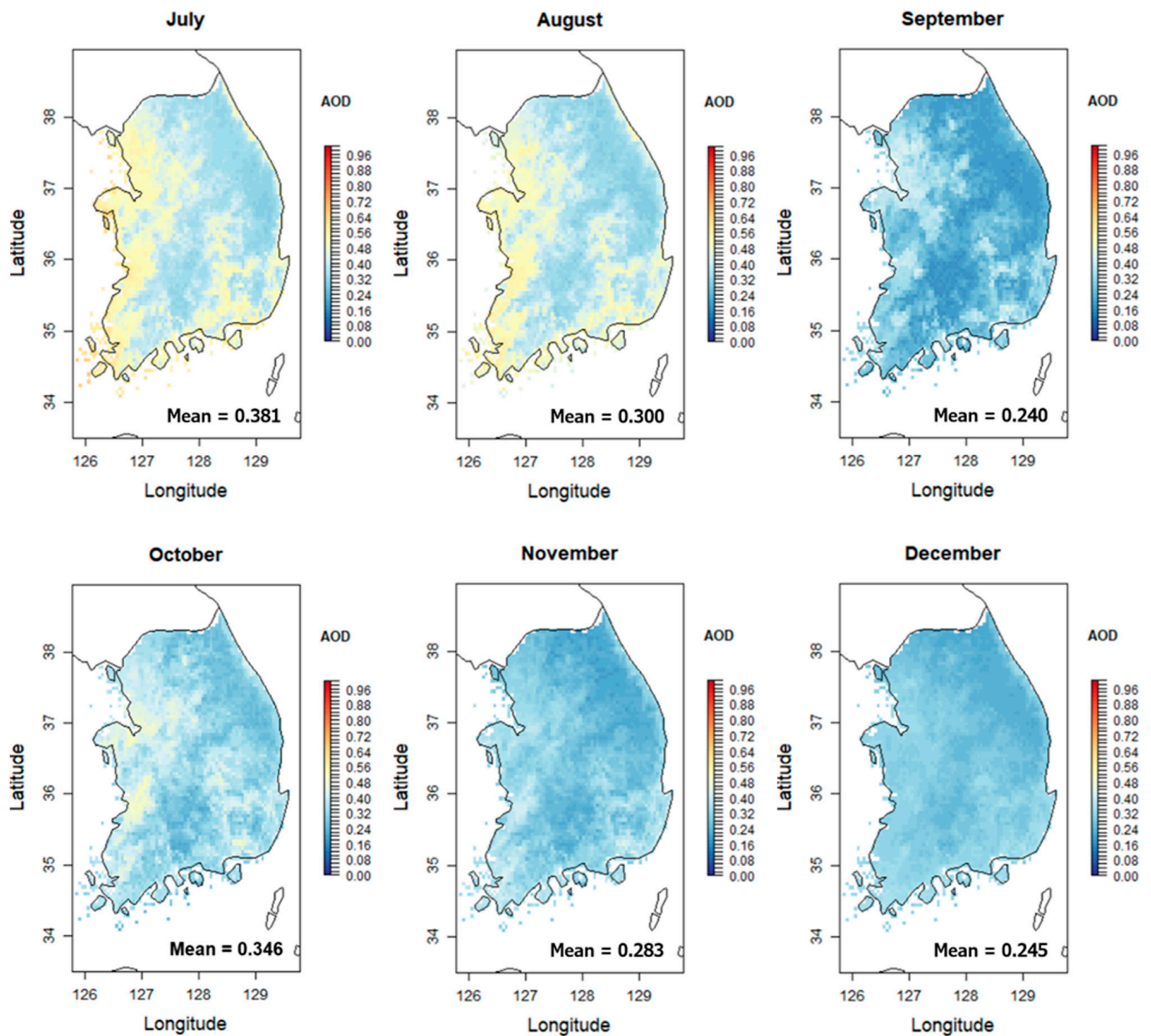


Figure 10. Monthly maps of gap-filled Himawari-8/AHI AOD for the year 2019, illustrating the spatial distribution of AOD values (ranging from 0 to 1) for each month. These maps provide a comprehensive view of the seasonal and regional variations in AOD throughout the year.

4.5. Time-Series Comparison with AERONET AOD

The performance of the hourly gap-filling model was evaluated through comparison with AERONET ground observations during 16–22 March 2019, which encompassed both typical and elevated AOD conditions. Figure 11 presents time series from three AERONET stations representing different environments: forest (Anmyon), near-forest (Hankuk_UFS), and urban (Seoul_SNU) sites. During this period, the AOD exhibited significant variations, transitioning from low concentrations (16–18 March) to high concentrations of 0.5–2.0 (19 March) before returning to low values (20–22 March). The gap-filling algorithm demonstrated robust performance in reproducing AERONET's temporal variations across these diverse conditions. The diurnal patterns during the high-AOD episode on 19 March revealed distinctive site-specific characteristics, with the Anmyon site exhibiting maximum values at 10:00 KST followed by decreasing trends, whereas the Seoul_SNU and

Hankuk_UFS sites showed sustained or increasing trends throughout the afternoon. These patterns can be attributed to the location of the observation sites and predominant wind directions. Considering Anmyon’s westernmost geographical position and the prevailing westerly winds, the delayed maxima observed at the Seoul_SNU and Hankuk_UFS sites likely reflect the aerosol transport time from west to east in the Korean Peninsula. This interpretation is corroborated by ground-based meteorological observations from adjacent Automated Synoptic Observing System (ASOS) stations (Boryeong and Hongseong), which documented persistent southwesterly winds during 09:00–14:00 KST, consistent with the observed eastward propagation of AOD maxima.

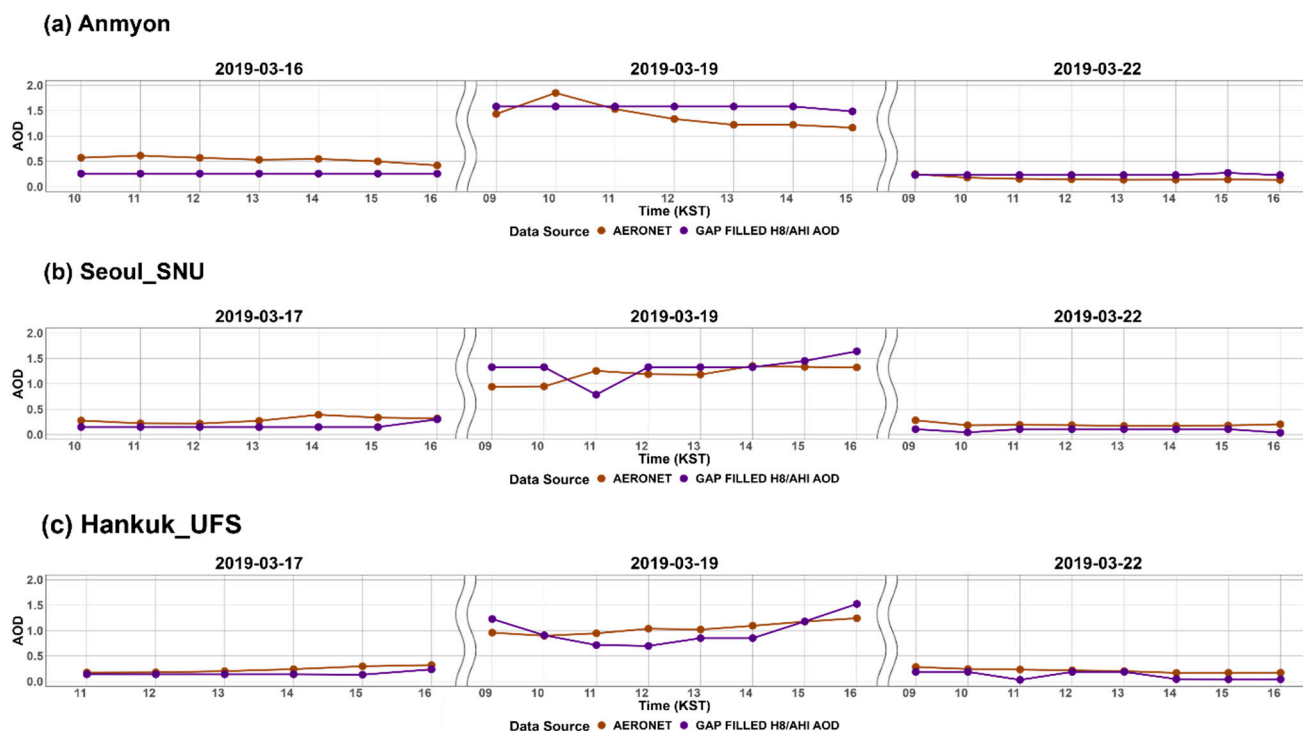


Figure 11. Time series comparison of Himawari-8 gap-filled AOD (purple) and AERONET AOD (brown) measurements during 16–22 March 2019, at different sites in South Korea: (a) Anmyon representing a forest environment, (b) Seoul_SNU representing an urban environment, and (c) Hankuk_UFS representing a near-forest environment. The middle panels (19 March) highlight the model’s performance during the high-concentration event, while the surrounding dates demonstrate its capability to capture typical conditions.

A notable case was observed at Seoul_SNU at 11:00 KST, where predicted AOD values showed a temporary decrease. This pattern appears related to the temporal interpolation characteristics of humidity-related parameters such as DPT, which exhibited the highest importance (8.66%) among meteorological variables. Particularly during this hour, DPT showed a distinctive V-shaped pattern (from -8.13 °C to -9.01 °C, then recovering to -8.78 °C). An abrupt decrease in DPT results in a drier atmosphere and then lower AOD. This case shows the importance of input data quality for AOD gap-filling, while the algorithm’s overall performance remained robust, as evidenced by the well-captured temporal variations in Figure 11.

4.6. Comparisons with Existing Gap-Filling Approaches

This study builds upon and extends previous research in AOD gap-filling using machine learning techniques. While adopting the fundamental RF approach successfully used by Zhao et al. [41] and Long et al. [73], our study refines and adapts these methods to address the specific challenges of high-temporal-resolution AOD estimation over South Korea's land area. Focusing on South Korea during 2019, with hourly estimations at a $0.05^\circ \times 0.05^\circ$ resolution, our approach strikes a balance between coverage and granularity, compared to studies such as Zhao et al. [41], which used daily estimates, and Chen et al. [42], which employed 10-min intervals.

The model performance metrics highlight the effectiveness of our streamlined RF approach. With test dataset results showing an MAE of 0.044 and CC of 0.966, our model achieves accuracy comparable to or surpassing more complex methods, such as Chen et al.'s [42] hybrid model (CC = 0.80) and Long et al.'s [73] approach (RMSE < 0.1). Extensive validation using AERONET data (4149 points), with a CC of 0.711 for predictions and 0.815 for Himawari-8, further supports these findings. A distinctive aspect of our methodology is its efficient use of input variables and moderate dataset size. While recent studies have incorporated numerous auxiliary variables, such as land use, population, and road networks, and utilized large datasets (e.g., Zhao et al.'s [41] 164 million samples), our model achieved robust performance with only 533,286 samples, relying primarily on CAMS, MERRA AOD, and LDAPS meteorological data. This efficient data selection and utilization indicate that our approach effectively captures key AOD patterns without extensive computational resources or complex data integration requirements.

These comparative insights reveal both the strengths of our approach and opportunities for future enhancement through extended temporal coverage, additional data sources, and hybrid modeling, while maintaining the model's computational efficiency and robust performance. The demonstrated success of our streamlined approach suggests its particular suitability for operational environmental monitoring applications.

5. Conclusions

This study developed and evaluated a spatial gap-filling approach for AHI hourly AOD data using an RF model over South Korea, successfully improving data coverage from 6% to 100%. Key findings include:

- (1) By incorporating model-based AOD and meteorological variables, our gap-filling model demonstrated high accuracy, achieving an MAE of 0.044 and a CC of 0.966 in the blind tests. CAMS AOD and MERRA-2 AOD were the most influential predictors, with meteorological variables (DPT, DSSF, and TMP) significantly contributing to the model's performance.
- (2) Comparisons with AERONET observations showed that the gap-filling method effectively balanced enhanced data coverage with maintained accuracy, with a stable MAE of 0.156 and a slight CC decrease from 0.815 to 0.711. The model performed well across all seasons, with particularly strong performance in fall and winter.
- (3) The gap-filled hourly AOD data revealed distinct seasonal patterns (winter: 0.245–0.300; spring: 0.354–0.382; summer: 0.381–0.391; fall: 0.240–0.346) and captured the influence of land use characteristics on AOD distributions.
- (4) Time-series analysis during 16–22 March 2019 demonstrated the model's capability to capture both typical and elevated AOD patterns (0.5–2.0) across different environments, while also revealing the critical role of input data quality in ensuring accurate gap-filling results, particularly during high-concentration episodes.

By providing a more complete and continuous AOD dataset, our method significantly enhances the capacity for comprehensive air quality monitoring and forecasting across South Korea. The improved temporal and spatial resolution enable more accurate tracking of air pollution events, including their onset, duration, and dispersion patterns, thereby contributing substantially to early warning systems and empowering authorities to implement timely mitigation strategies. The availability of gap-filled AOD data at high temporal resolutions also opens new avenues for studying diurnal variations in aerosol concentrations, providing a solid foundation for evidence-based environmental management and policy decisions.

Future research should focus on enhancing the model's performance during extreme events, such as dust storms and wildfires, which cause rapid fluctuations in aerosol distributions. Incorporating additional data sources, including high-frequency meteorological observations, real-time emission records, high-resolution Digital Elevation Models (DEMs), and detailed land-use information, could improve the model's ability to capture localized effects and extreme events. Integrating these diverse datasets would be particularly beneficial for predicting abrupt and intense changes in AOD levels, which are currently challenging to capture. Additionally, implementing hybrid deep learning models, such as Convolutional Long Short-Term Memory (ConvLSTM) networks combined with Graph Neural Networks (GNNs), could further improve model accuracy in predicting extreme events.

The proposed method can enhance our understanding of aerosol dynamics by providing continuous, high-resolution AOD data, overcoming the limitations of polar-orbiting satellites such as MODIS or VIIRS in capturing diurnal variations. This approach has potential applications for other geostationary satellites, such as the Advanced Meteorological Imager (AMI) on GeoKompsat-2A (GK2A) and the Geostationary Environment Monitoring Spectrometer (GEMS) and Geostationary Ocean Color Imager-2 (GOCI-2) on GeoKompsat-2B (GK2B), contributing to improved regional and global aerosol monitoring capabilities.

Author Contributions: Conceptualization, Y.Y. and Y.L.; methodology, Y.Y. and Y.L.; formal analysis, Y.Y.; data curation, Y.Y.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.Y., S.K., S.H.K., and Y.L.; project administration, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Korea Environment Industry & Technology Institute (KEITI) through the Project “Developing an Observation-based GHG Emissions Geospatial Information Map” funded by Korea Ministry of Environment (MOE) (RS-2023-00232066). This work was carried out with the support of the “Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ0162342024)” by the Rural Development Administration, Republic of Korea.

Data Availability Statement: All original datasets used in this study are publicly available from the sources cited in the text. The gap-filled dataset generated during this study is available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

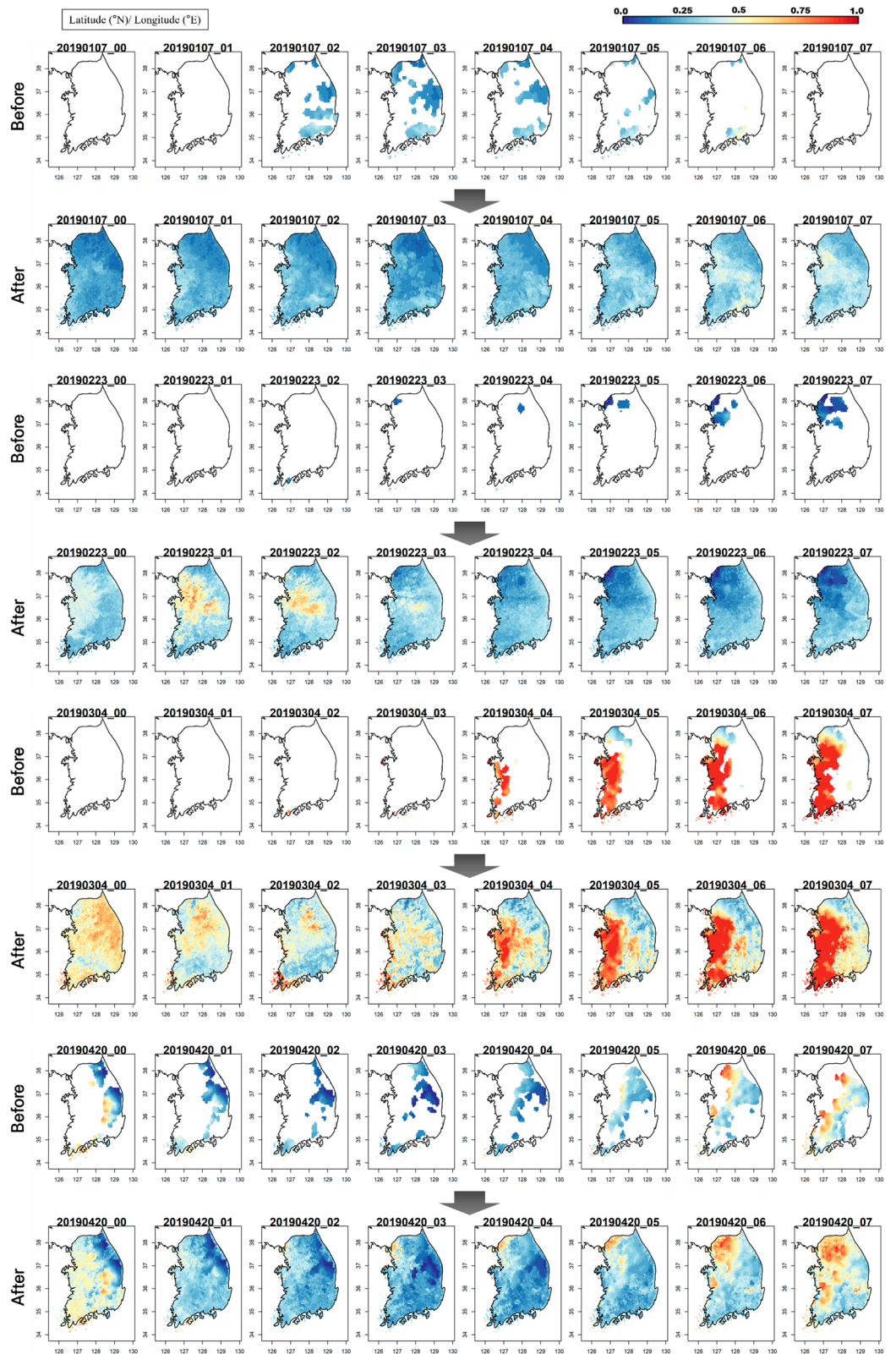


Figure A1. Cont.

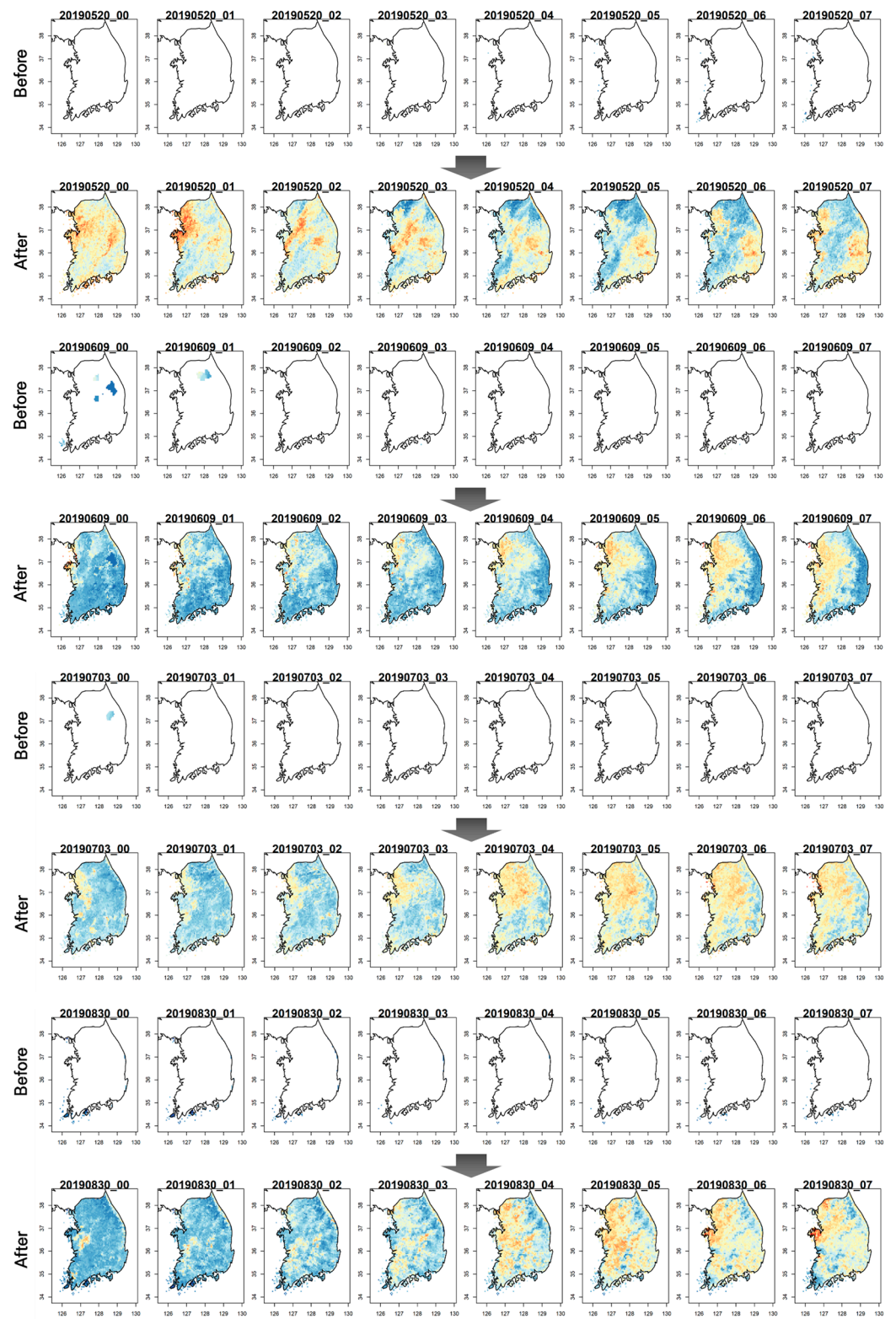


Figure A1. Cont.

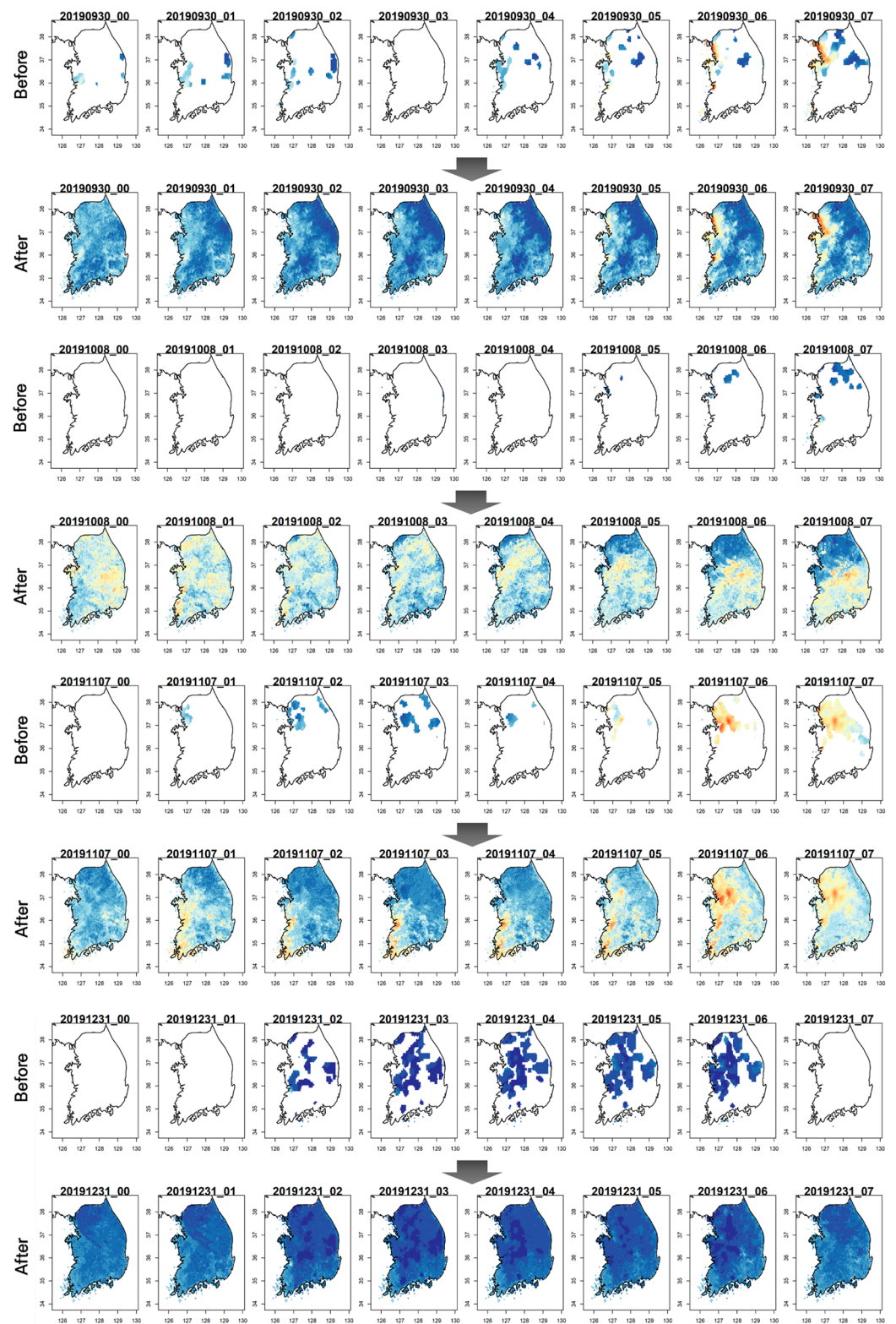


Figure A1. Representative examples of hourly AOD gap-filling results for 2019, showing sample sets from each of the 12 months (January to December) to reflect different seasonal patterns. Each row displays 8 consecutive hours (00–07 UTC) of AOD data for a selected date. The “Before” rows show the original AOD data with gaps, while the “After” rows present the gap-filled data. A color scale indicates AOD values, ranging from 0 (blue) to 1 (red).

References

1. Twomey, S. Pollution and the planetary albedo. *Atmos. Environ.* (1967) **1974**, *8*, 1251–1256. [[CrossRef](#)]
2. Albrecht, B.A. Aerosols, cloud microphysics, and fractional cloudiness. *Science* **1989**, *245*, 1227–1230. [[CrossRef](#)] [[PubMed](#)]
3. Hinds, W.C. *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles*; John Wiley & Sons: New York, NY, USA, 1999.
4. Wang, K.; Dickinson, R.E.; Liang, S. Clear sky visibility has decreased over land globally from 1973 to 2007. *Science* **2009**, *323*, 1468–1470. [[CrossRef](#)] [[PubMed](#)]
5. Zanobetti, A.; Schwartz, J. The effect of fine and coarse particulate air pollution on mortality: A national analysis. *Environ. Health Perspect.* **2009**, *117*, 898–903. [[CrossRef](#)] [[PubMed](#)]
6. Watson-Parris, D.; Smith, C.J. Large uncertainty in future warming due to aerosol forcing. *Nat. Clim. Change* **2022**, *12*, 1111–1113. [[CrossRef](#)]
7. Lee, G.-T.; Ryu, S.-W.; Lee, T.-Y.; Suh, M.-S. Analysis of AOD characteristics retrieved from Himawari-8 using sun photometer in South Korea. *Korean J. Remote Sens.* **2020**, *36*, 425–439.
8. Huebert, B.J.; Bates, T.; Russell, P.B.; Shi, G.; Kim, Y.J.; Kawamura, K.; Carmichael, G.; Nakajima, T. An overview of ACE-Asia: Strategies for quantifying the relationships between Asian aerosols and their climatic impacts. *J. Geophys. Res. Atmos.* **2003**, *108*. [[CrossRef](#)]
9. Higurashi, A.; Nakajima, T. Development of a two-channel aerosol retrieval algorithm on a global scale using NOAA AVHRR. *J. Atmos. Sci.* **1999**, *56*, 924–941. [[CrossRef](#)]
10. Voiland, A. *Aerosols: Tiny Particles, Big Impact*; NASA Earth Observatory: Greenbelt, MD, USA, 2010.
11. Kloog, I.; Nordio, F.; Coull, B.A.; Schwartz, J. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM_{2.5} exposures in the Mid-Atlantic states. *Environ. Sci. Technol.* **2012**, *46*, 11913–11921. [[CrossRef](#)]
12. Li, J.; Carlson, B.E.; Laci, A.A. How well do satellite AOD observations represent the spatial and temporal variability of PM_{2.5} concentration for the United States? *Atmos. Environ.* **2015**, *102*, 260–273. [[CrossRef](#)]
13. Di, Q.; Kloog, I.; Koutrakis, P.; Lyapustin, A.; Wang, Y.; Schwartz, J. Assessing PM_{2.5} exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol.* **2016**, *50*, 4712–4721. [[CrossRef](#)] [[PubMed](#)]
14. Stafoggia, M.; Schwartz, J.; Badaloni, C.; Bellander, T.; Alessandrini, E.; Cattani, G.; De’Donato, F.; Gaeta, A.; Leone, G.; Lyapustin, A. Estimation of daily PM₁₀ concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. *Environ. Int.* **2017**, *99*, 234–244. [[CrossRef](#)] [[PubMed](#)]
15. de Hoogh, K.; Héritier, H.; Stafoggia, M.; Künzli, N.; Kloog, I. Modelling daily PM_{2.5} concentrations at high spatio-temporal resolution across Switzerland. *Environ. Pollut.* **2018**, *233*, 1147–1154. [[CrossRef](#)]
16. Kinne, S.; Schulz, M.; Textor, C.; Guibert, S.; Balkanski, Y.; Bauer, S.E.; Bernsten, T.; Berglen, T.; Boucher, O.; Chin, M. An AeroCom initial assessment—optical properties in aerosol component modules of global models. *Atmos. Chem. Phys.* **2006**, *6*, 1815–1834. [[CrossRef](#)]
17. Kaufman, Y.; Tanré, D.; Remer, L.A.; Vermote, E.; Chu, A.; Holben, B. Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *J. Geophys. Res. Atmos.* **1997**, *102*, 17051–17067. [[CrossRef](#)]
18. Remer, L.A.; Kaufman, Y.; Tanré, D.; Mattoo, S.; Chu, D.; Martins, J.V.; Li, R.-R.; Ichoku, C.; Levy, R.; Kleidman, R. The MODIS aerosol algorithm, products, and validation. *J. Atmos. Sci.* **2005**, *62*, 947–973. [[CrossRef](#)]
19. Jackson, J.M.; Liu, H.; Laszlo, I.; Kondragunta, S.; Remer, L.A.; Huang, J.; Huang, H.C. Suomi-NPP VIIRS aerosol algorithms and data products. *J. Geophys. Res. Atmos.* **2013**, *118*, 12673–612689. [[CrossRef](#)]
20. Wang, J.; Christopher, S.A. Intercomparison between satellite-derived aerosol optical thickness and PM_{2.5} mass: Implications for air quality studies. *Geophys. Res. Lett.* **2003**, *30*. [[CrossRef](#)]
21. Engel-Cox, J.A.; Holloman, C.H.; Coutant, B.W.; Hoff, R.M. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmos. Environ.* **2004**, *38*, 2495–2509. [[CrossRef](#)]
22. Lee, K.; Kim, Y. Russian forest fire smoke aerosol monitoring using satellite and AERONET data. *J. Korean Soc. Atmos. Environ.* **2004**, *20*, 437–450.
23. Gupta, P.; Christopher, S. Seven year particulate matter air quality assessment from surface and satellite measurements. *Atmos. Chem. Phys.* **2008**, *8*, 3311–3324. [[CrossRef](#)]
24. Schaap, M.; Apituley, A.; Timmermans, R.; Koelemeijer, R.; De Leeuw, G. Exploring the relation between aerosol optical depth and PM_{2.5} at Cabauw, the Netherlands. *Atmos. Chem. Phys.* **2009**, *9*, 909–925. [[CrossRef](#)]
25. Park, J.-Y.; Kwon, T.-Y.; Lee, J.-Y. Estimation of surface visibility using MODIS AOD. *Korean J. Remote Sens.* **2017**, *33*, 171–187.
26. Yang, X.; Zhao, C.; Yang, Y.; Yan, X.; Fan, H. Statistical aerosol properties associated with fire events from 2002 to 2019 and a case analysis in 2019 over Australia. *Atmos. Chem. Phys.* **2021**, *21*, 3833–3853. [[CrossRef](#)]
27. Gao, L.; Chen, L.; Li, C.; Li, J.; Che, H.; Zhang, Y. Evaluation and possible uncertainty source analysis of JAXA Himawari-8 aerosol optical depth product over China. *Atmos. Res.* **2021**, *248*, 105248. [[CrossRef](#)]
28. Schmit, T.J.; Griffith, P.; Gunshor, M.M.; Daniels, J.M.; Goodman, S.J.; Lebar, W.J. A closer look at the ABI on the GOES-R series. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 681–698. [[CrossRef](#)]
29. Wei, J.; Li, Z.; Sun, L.; Peng, Y.; Zhang, Z.; Li, Z.; Su, T.; Feng, L.; Cai, Z.; Wu, H. Evaluation and uncertainty estimate of next-generation geostationary meteorological Himawari-8/AHI aerosol products. *Sci. Total Environ.* **2019**, *692*, 879–891. [[CrossRef](#)]

30. Xu, W.; Wang, W.; Chen, B. Comparison of hourly aerosol retrievals from JAXA Himawari/AHI in version 3.0 and a simple customized method. *Sci. Rep.* **2020**, *10*, 20884. [[CrossRef](#)]
31. 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 Sens.* **2022**, *14*, 2967. [[CrossRef](#)]
32. Van Donkelaar, A.; Martin, R.V.; Levy, R.C.; da Silva, A.M.; Krzyzanowski, M.; Chubarova, N.E.; Semutnikova, E.; Cohen, A.J. Satellite-based estimates of ground-level fine particulate matter during extreme events: A case study of the Moscow fires in 2010. *Atmos. Environ.* **2011**, *45*, 6225–6232. [[CrossRef](#)]
33. Tao, M.; Chen, L.; Su, L.; Tao, J. Satellite observation of regional haze pollution over the North China Plain. *J. Geophys. Res. Atmos.* **2012**, *117*. [[CrossRef](#)]
34. Xiao, Q.; Wang, Y.; Chang, H.H.; Meng, X.; Geng, G.; Lyapustin, A.; Liu, Y. Full-coverage high-resolution daily PM_{2.5} estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ.* **2017**, *199*, 437–446. [[CrossRef](#)]
35. Youn, Y.; Kim, S.; Jeong, Y.; Cho, S.; Kang, J.; Kim, G.; Lee, Y. Spatial Gap-Filling of Hourly AOD Data from Himawari-8 Satellite Using DCT (Discrete Cosine Transform) and FMM (Fast Marching Method). *Korean J. Remote Sens.* **2021**, *37*, 777–788.
36. Yu, C.; Chen, L.; Su, L.; Fan, M.; Li, S. Kriging interpolation method and its application in retrieval of MODIS aerosol optical depth. In Proceedings of the 2011 19th International Conference on Geoinformatics, Shanghai, China, 24–26 June 2011; pp. 1–6.
37. Singh, M.K.; Gautam, R.; Venkatachalam, P. Bayesian merging of MISR and MODIS aerosol optical depth products using error distributions from AERONET. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 5186–5200. [[CrossRef](#)]
38. Tang, Q.; Bo, Y.; Zhu, Y. Spatiotemporal fusion of multiple-satellite aerosol optical depth (AOD) products using Bayesian maximum entropy method. *J. Geophys. Res. Atmos.* **2016**, *121*, 4034–4048. [[CrossRef](#)]
39. Zhang, R.; Di, B.; Luo, Y.; Deng, X.; Grieneisen, M.L.; Wang, Z.; Yao, G.; Zhan, Y. A nonparametric approach to filling gaps in satellite-retrieved aerosol optical depth for estimating ambient PM_{2.5} levels. *Environ. Pollut.* **2018**, *243*, 998–1007. [[CrossRef](#)]
40. Stafoggia, M.; Bellander, T.; Bucci, S.; Davoli, M.; De Hoogh, K.; De’Donato, F.; Gariazzo, C.; Lyapustin, A.; Michelozzi, P.; Renzi, M. Estimation of daily PM₁₀ and PM_{2.5} concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environ. Int.* **2019**, *124*, 170–179. [[CrossRef](#)]
41. 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. *Atmos. Environ.* **2019**, *203*, 70–78. [[CrossRef](#)]
42. Chen, A.; Yang, J.; He, Y.; Yuan, Q.; Li, Z.; Zhu, L. High spatiotemporal resolution estimation of AOD from Himawari-8 using an ensemble machine learning gap-filling method. *Sci. Total Environ.* **2023**, *857*, 159673. [[CrossRef](#)]
43. Yoshida, M.; Kikuchi, M.; Nagao, T.M.; Murakami, H.; Nomaki, T.; Higurashi, A. Common retrieval of aerosol properties for imaging satellite sensors. *J. Meteorol. Soc. Jpn. Ser. II* **2018**, *96B*, 193–209. [[CrossRef](#)]
44. Fukuda, S.; Nakajima, T.; Takenaka, H.; Higurashi, A.; Kikuchi, N.; Nakajima, T.Y.; Ishida, H. New approaches to removing cloud shadows and evaluating the 380 nm surface reflectance for improved aerosol optical thickness retrievals from the GOSAT/TANSO-Cloud and Aerosol Imager. *J. Geophys. Res. Atmos.* **2013**, *118*, 13520–13531. [[CrossRef](#)]
45. Nakajima, T.; Tanaka, M. Matrix formulations for the transfer of solar radiation in a plane-parallel scattering atmosphere. *J. Quant. Spectrosc. Radiat. Transf.* **1986**, *35*, 13–21. [[CrossRef](#)]
46. Ota, Y.; Higurashi, A.; Nakajima, T.; Yokota, T. Matrix formulations of radiative transfer including the polarization effect in a coupled atmosphere–ocean system. *J. Quant. Spectrosc. Radiat. Transf.* **2010**, *111*, 878–894. [[CrossRef](#)]
47. Omar, A.H.; Won, J.G.; Winker, D.M.; Yoon, S.C.; Dubovik, O.; McCormick, M.P. Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements. *J. Geophys. Res. Atmos.* **2005**, *110*. [[CrossRef](#)]
48. Sayer, A.; Smirnov, A.; Hsu, N.; Holben, B. A pure marine aerosol model, for use in remote sensing applications. *J. Geophys. Res. Atmos.* **2012**, *117*. [[CrossRef](#)]
49. Kikuchi, M.; Murakami, H.; Suzuki, K.; Nagao, T.M.; Higurashi, A. Improved hourly estimates of aerosol optical thickness using spatiotemporal variability derived from Himawari-8 geostationary satellite. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3442–3455. [[CrossRef](#)]
50. Kim, S.; Jeong, Y.; Youn, Y.; Cho, S.; Kang, J.; Kim, G.; Lee, Y. A Comparison between multiple satellite AOD products using AERONET sun photometer observations in South Korea: Case study of MODIS, VIIRS, Himawari-8, and Sentinel-3. *Korean J. Remote Sens.* **2021**, *37*, 543–557.
51. Zhang, W.; Xu, H.; Zhang, L. Assessment of Himawari-8 AHI aerosol optical depth over land. *Remote Sens.* **2019**, *11*, 1108. [[CrossRef](#)]
52. Benedetti, A.; Morcrette, J.-J.; Boucher, O.; Dethof, A.; Engelen, R.J.; Fisher, M.; Flentje, H.; Huneus, N.; Jones, L.; Kaiser, J.W.; et al. Aerosol analysis and forecast in the ECMWF Integrated Forecast System. Part II: Data assimilation. *J. Geophys. Res.* **2009**, *114*, D13205.
53. Morcrette, J.-J.; Benedetti, A.; Jones, L.; Kaiser, J.W.; Razinger, M.; Suttie, M.; Aumann, H.H.; Beekmann, M.; Bellouin, N.; Boucher, O.; et al. Aerosol analysis and forecast in the ECMWF Integrated Forecast System. Part I: Forward modelling. *J. Geophys. Res.* **2009**, *114*, D06206. [[CrossRef](#)]
54. Tuygun, G.; Elbir, T. Comparative analysis of CAMS aerosol optical depth data and AERONET observations in the Eastern Mediterranean over 19 years. *Environ. Sci. Pollut. Res.* **2024**, *31*, 27069–27084. [[CrossRef](#)] [[PubMed](#)]

55. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **2017**, *30*, 5419–5454. [[CrossRef](#)] [[PubMed](#)]
56. Molod, A.; Takacs, L.; Suarez, M.; Bacmeister, J. Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA-2. *Geosci. Model Dev.* **2015**, *8*, 1339–1356. [[CrossRef](#)]
57. Bi, J.; Belle, J.H.; Wang, Y.; Lyapustin, A.I.; Wildani, A.; Liu, Y. Impacts of snow and cloud covers on satellite-derived PM2.5 levels. *Remote Sens. Environ.* **2019**, *221*, 665–674. [[CrossRef](#)]
58. Li, L.; Franklin, M.; Girguis, M.; Lurmann, F.; Wu, J.; Pavlovic, N.; Breton, C.; Gilliland, F.; Habre, R. Spatiotemporal imputation of MAIAC AOD using deep learning with downscaling. *Remote Sens. Environ.* **2020**, *237*, 111584. [[CrossRef](#)]
59. Chen, B.; You, S.; Ye, Y.; Fu, Y.; Ye, Z.; Deng, J.; Wang, K.; Hong, Y. An interpretable self-adaptive deep neural network for estimating daily spatially-continuous PM2.5 concentrations across China. *Sci. Total Environ.* **2021**, *768*, 144724. [[CrossRef](#)]
60. Kianian, B.; Liu, Y.; Chang, H.H. Imputing satellite-derived aerosol optical depth using a multi-resolution spatial model and random forest for PM2.5 prediction. *Remote Sens.* **2021**, *13*, 126. [[CrossRef](#)]
61. Chen, J.; Li, Z.; Lv, M.; Wang, Y.; Wang, W.; Zhang, Y.; Wang, H.; Yan, X.; Sun, Y.; Cribb, M. Aerosol hygroscopic growth, contributing factors, and impact on haze events in a severely polluted region in northern China. *Atmos. Chem. Phys.* **2019**, *19*, 1327–1342. [[CrossRef](#)]
62. Collaud Coen, M.; Praz, C.; Haeefele, A.; Ruffieux, D.; Kaufmann, P.; Calpini, B. Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model. *Atmos. Chem. Phys.* **2014**, *14*, 13205–13221. [[CrossRef](#)]
63. Ding, A.J.; Huang, X.; Nie, W.; Sun, J.N.; Kerminen, V.-M.; Petäjä, T.; Su, H.; Cheng, Y.F.; Yang, X.-Q.; Wang, M.H.; et al. Enhanced haze pollution by black carbon in megacities in China. *Geophys. Res. Lett.* **2016**, *43*, 2873–2879. [[CrossRef](#)]
64. Ramaswamy, V.; Boucher, O.; Haigh, J.; Hauglustaine, D.; Haywood, J.; Myhre, G.; Nakajima, T.; Shi, G.; Solomon, S. *Radiative Forcing of Climate Change. Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*; Griggs, D.J., Noguier, M., van der Linden, P.J., Dai, X., Maskell, K., Johnson, C.A., Eds.; Cambridge University Press: Cambridge, UK, 2001; Volume 350, p. 416.
65. Myhre, G.; Stordal, F.; Johnsrud, M.; Kaufman, Y.; Rosenfeld, D.; Storelvmo, T.; Kristjansson, J.E.; Berntsen, T.K.; Myhre, A.; Isaksen, I.S. Aerosol-cloud interaction inferred from MODIS satellite data and global aerosol models. *Atmos. Chem. Phys.* **2007**, *7*, 3081–3101. [[CrossRef](#)]
66. Yoo, J.-W.; Park, S.-Y.; Jeon, W.; Kim, D.-H.; Lee, H.; Lee, S.-H.; Kim, H.-G. Effect of Aerosol Feedback on Solar Radiation in the Korean Peninsula Using WRF-CMAQ Two-way Coupled Model. *J. Korean Soc. Atmos. Environ.* **2017**, *33*, 435–444. [[CrossRef](#)]
67. Alam, K.; Khan, R.; Blaschke, T.; Mukhtiar, A. Variability of aerosol optical depth and their impact on cloud properties in Pakistan. *J. Atmos. Sol.-Terr. Phys.* **2014**, *107*, 104–112. [[CrossRef](#)]
68. Aires, F.; Prigent, C.; Rossow, W. Temporal interpolation of global surface skin temperature diurnal cycle over land under clear and cloudy conditions. *J. Geophys. Res. Atmos.* **2004**, *109*. [[CrossRef](#)]
69. Chen, J.M.; Deng, F.; Chen, M. Locally adjusted cubic-spline capping for reconstructing seasonal trajectories of a satellite-derived surface parameter. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2230–2238. [[CrossRef](#)]
70. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009; pp. 1–758.
71. Louppe, G. Understanding Random Forests. Ph.D. Thesis, University of Liege, Liège, Belgium, 2014.
72. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
73. Long, Z.; Jin, Z.; Meng, Y.; Ma, J. Generation of High Temporal Resolution Full-Coverage Aerosol Optical Depth Based on Remote Sensing and Reanalysis Data. *Remote Sens.* **2023**, *15*, 2769. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.