remote sensing

MDPI

Article

Aerosol Optical Depth Retrieval from MODIS Using a Physically Informed Machine Learning Framework

Tianchen Liang [1], Linqing Zou [2,*], Qiaoning He [1] and Lin Sun [3]

[1] School of Geography and Planning, Huaiyin Normal University, Huaian 223300, China
[2] Hubei Key Laboratory of Quantitative Remote Sensing of Land and Atmosphere, School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[3] College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China
* Correspondence: zoulinqing@whu.edu.cn

Highlights

What are the main findings?

- A physically informed random forest framework was developed for 1 km land AOD retrieval by integrating MODIS observations with surface, meteorological, topographic, and physically constrained aerosol–surface features.
- Independent evaluation for 2013–2014 achieved an MAE of 0.063, an RMSE of 0.096, and 74.93% of retrievals within the MODIS expected-error envelope, showing better agreement with AERONET than the matched MOD04 Dark Target and Deep Blue products.

What are the implications of the main findings?

- Regional and seasonal minimum-AOD priors together with 6S-constrained time-series clear-sky reflectance features improve retrieval transferability across heterogeneous land surfaces and aerosol conditions.
- The proposed framework provides spatially detailed and physically plausible 1 km AOD estimates for global aerosol monitoring under clear-sky conditions.

Abstract

Retrieving aerosol optical depth (AOD) over land remains challenging because the relatively weak aerosol signal in top-of-atmosphere (TOA) observations must be separated from strong and spatially heterogeneous surface reflectance. Here, we develop a physically informed random forest framework for global 1 km land AOD retrieval from MODIS. The framework integrates multispectral TOA reflectance, surface properties, observation geometry, meteorological conditions, topography, and physically informed aerosol–surface features. Long-term Aerosol Robotic Network (AERONET) observations from 2001 to 2017 were collocated with MODIS and ancillary datasets for model development and evaluation. Two physically informed features were introduced to improve retrieval robustness across diverse aerosol and surface conditions, including minimum AOD derived from long-term AERONET observations and time-series clear-sky reflectance (TSCR) in the blue, red, and shortwave-infrared bands derived using the 6S radiative-transfer model. Independent retrieval evaluation for 2013–2014 showed good agreement with AERONET observations, with R = 0.81, MAE = 0.063, RMSE = 0.096, and 74.93% of matched samples falling within the MODIS land expected-error envelope, although increasing underestimation was observed at high aerosol loading (AOD > 1). The proposed retrievals also showed better agreement with AERONET than the MOD04 Dark Target and Deep Blue products. These results

Check for updates

Academic Editor: Carmine Serio

Received: 3 July 2026
Revised: 15 August 2026
Accepted: 20 August 2026
Published: 24 August 2026

Copyright: © 2026 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.

demonstrate the value of incorporating physically interpretable aerosol-background and surface-reflectance information into data-driven retrievals for AOD over land surfaces.

Keywords: aerosol optical depth; MODIS; AERONET; random forest

1. Introduction

Atmospheric aerosols are complex mixtures of suspended solid and liquid particles originating from both natural processes and anthropogenic activities [1,2]. As highly heterogeneous and temporally variable components of the Earth-atmosphere system, aerosols affect the radiation budget by scattering and absorbing solar shortwave and terrestrial longwave radiation [3–6]. They also act as cloud condensation nuclei and ice nuclei, thereby modifying cloud microphysics, cloud lifetime, and precipitation processes [7]. In addition to their climatic effects, aerosols are important carriers of air pollution [8] and can degrade the accuracy of atmospheric correction and quantitative surface-parameter retrievals [9]. Reliable characterization of aerosol spatiotemporal variability is therefore fundamental for climate research, atmospheric environmental monitoring, and remote-sensing applications [10].

Aerosol optical depth (AOD) is the key column-integrated optical parameter used to quantify aerosol extinction and is one of the most widely used indicators in satellite aerosol remote sensing [11]. Ground-based sun-photometer networks, especially the Aerosol Robotic Network (AERONET), provide accurate, long-term measurements of AOD and aerosol optical properties and constitute an essential benchmark for satellite aerosol product validation and algorithm development [12–14]. However, ground-based observations remain spatially sparse and cannot provide continuous aerosol monitoring over regional to global scales. Satellite remote sensing offers broad spatial coverage, frequent revisit capability, and long-term observational continuity. In particular, the Moderate Resolution Imaging Spectroradiometer (MODIS) provides a long and stable observation record that has been widely used for global aerosol monitoring and aerosol climatology studies [15,16].

Several operational and research algorithms have been developed for AOD retrieval over land, including Dark Target (DT), Deep Blue (DB), the Multi-Angle Implementation of Atmospheric Correction (MAIAC), the Simplified Aerosol Retrieval Algorithm (SARA), and higher-resolution MODIS aerosol products [17–25]. These approaches have substantially advanced global aerosol monitoring, but most of them still rely on a sequence of assumptions regarding surface reflectance, aerosol type, and radiative transfer. Over complex land surfaces, retrieval accuracy and robustness remain constrained. Errors in surface reflectance parameterization, aerosol-model selection, or regional empirical assumptions can propagate through the retrieval chain. This problem is particularly pronounced over urban areas, bright bare soil, arid and semi-arid regions, and rapidly changing land-cover types, where surface spectral and directional reflectance are highly heterogeneous [11,26].

Machine learning provides a complementary pathway for AOD retrieval because it can learn nonlinear relationships directly from multisource observations without explicitly solving every component of the physical inversion. Recent studies using machine learning have demonstrated the potential of data-driven and hybrid physical–statistical approaches for improving AOD estimation from satellite reflectance observations [27–31]. Nevertheless, many previous machine learning-based aerosol retrieval studies have been regionally focused, trained over relatively short periods, or driven primarily by top-of-atmosphere (TOA) reflectance and limited auxiliary variables [29,32,33]. Such designs may not fully capture the seasonal variability, aerosol-type diversity, surface heterogeneity, and atmospheric

background conditions required for robust global-scale AOD retrieval [34]. There therefore remains a need for machine learning retrieval frameworks that combine long-term, globally distributed observations with physically interpretable constraints on both aerosol and surface contributions to the TOA signal. To address this need, this study develops a global 1 km land AOD retrieval framework that integrates MODIS multispectral TOA reflectance, surface properties, observation geometry, meteorological variables, and topography with physically informed features. In particular, regional- and season-specific minimum AOD (AOD_{min}) is introduced to characterize the climatological low-aerosol background, while time-series clear-sky reflectance (TSCR) provides a band-specific constraint on the surface-reflectance background under relatively weak aerosol loading. These physically informed features are intended to provide complementary aerosol and surface constraints beyond those available from instantaneous satellite observations alone. A globally distributed AERONET–MODIS collocation database spanning 2001–2017 is used to train and evaluate the resulting random forest retrieval framework.

2. Materials and Methods

2.1. Data Sources

The retrieval framework integrates satellite radiance information, ground-based AOD observations, and surface-related variables, including Global land surface satellite (GLASS) visible black-sky albedo (BSA-VIS) and MODIS Normalized Difference Vegetation Index (NDVI), meteorological background fields, terrain information, and quality-control datasets. Table 1 summarizes the input datasets and their roles in the proposed framework.

Table 1. Input datasets used for 1 km land AOD retrieval.

Dataset	Product/Source	Resolution	Data Period Used	Use in This Study	URL (Web Address, Accessed on 19 August 2026)
MODIS TOA Reflectance	Terra/MODIS MOD02 L1B, bands 1–7	1 km	2001–2017	Spectral TOA reflectance	https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD021KM
Ground AOD	AERONET Level 2.0	Site observations	2001–2017	Training and validation labels	https://aeronet.gsfc.nasa.gov/new_web/webtool_aod_v3.html
Surface albedo	GLASS visible black-sky albedo	1 km	2001–2017	Surface radiative background	https://glass-product.bnu.edu.cn/introduction/abd.html
Meteorology	ERA5 hourly data	0.25° × 0.25°	2001–2017	SP, total column ozone, and total column water vapor	https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels
Observation geometry	MOD03	1 km	2001–2017	Solar/sensor viewing geometry and geolocation	https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD03
NDVI	Terra/MODIS MOD13A2	1 km	2001–2017	NDVI	https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD13A2
Elevation	GMTED2010 DEM	~1 km	Static	Elevation	https://doi.org/10.5066/F7J38R2N

Table 1. *Cont.*

Dataset	Product/Source	Resolution	Data Period Used	Use in This Study	URL (Web Address, Accessed on 19 August 2026)
Cloud mask	Terra/MODIS MOD35_L2 Cloud Mask	1 km	2001–2017	Cloud screening	https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD35_L2
Comparison AOD	Terra/MODIS MOD04_L2 Aerosol Product	10 km	2001–2017	Comparison AOD product	https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD04_L2

2.1.1. MODIS TOA Reflectance

Terra/MODIS MOD02 Level-1B reflectance data were used as the primary satellite input. MODIS provides high temporal sampling, near-global coverage every 1–2 days, and measurements in 36 spectral bands with nominal nadir spatial resolutions of 250 m, 500 m, and 1 km [35]. In the coupled surface–atmosphere system, TOA reflectance contains both aerosol path scattering and surface-reflected radiation attenuated by the atmosphere. Aerosol-induced reflectance perturbations are generally stronger at visible wavelengths, whereas near- and shortwave-infrared bands provide complementary constraints on the surface reflectance background. Accordingly, the first seven MODIS reflective solar bands were used as spectral predictors for AOD retrieval.

2.1.2. Ground-Based AOD Observations

Ground-based AOD observations were obtained from AERONET. AERONET provides globally distributed, quality-assured sun-photometer measurements and has been widely used as a reference for satellite aerosol validation [14]. In this study, AERONET Level 2.0 observations from 2001 to 2017 were spectrally interpolated to 550 nm using the Ångström exponent and used as supervised-learning labels. Only cloud-screened, calibrated, and quality-controlled records were retained. The global distribution of the AERONET sites used in the collocation database is shown in Supplementary Figure S1.

2.1.3. Surface Albedo

BSA-VIS fromGLASS product suite was included to characterize the surface radiative background [36,37]. BSA-VIS is derived from the surface BRDF and represents directional-hemispherical reflectance under direct illumination at a given solar zenith angle; therefore, it was not treated as a geometry-independent surface reflectance. In the retrieval framework, BSA-VIS was used together with solar zenith angle, sensor zenith angle, relative azimuth angle, and scattering angle, allowing the model to account for variations in both surface radiative background and acquisition geometry. GLASS albedo products provide long-term global coverage and were used here for 2001–2017.

2.1.4. Meteorological Variables

Atmospheric-state variables were obtained from the ERA5 reanalysis produced by the European Centre for Medium-Range Weather Forecasts [38]. Three variables were extracted: surface pressure (SP), total column ozone (TCO3), and total column water vapor (TCWV). These variables characterize the pressure background, ozone absorption, and water-vapor conditions, respectively, thereby improving model adaptability under different atmospheric states.

2.1.5. Ancillary Datasets and Quality Control

Ancillary datasets included observation geometry, vegetation condition, topography, cloud masking, and MODIS aerosol products for comparison. Solar zenith angle, sensor zenith angle, relative azimuth angle, scattering angle, and geolocation were obtained from MOD03. Vegetation condition was represented using the 16-day 1 km MOD13A2 NDVI product [39]. Terrain elevation was obtained from Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) [40,41]. Because optical aerosol retrieval is highly sensitive to cloud contamination, MOD35_L2 was used to mask cloudy pixels; only pixels flagged as confident clear were retained [42]. MODIS MOD04_L2 DT and DB aerosol products were used as comparison datasets after quality-flag screening [19].

2.2. Predictor Variables for AOD Retrieval

Global land AOD at 550 nm was retrieved at a spatial resolution of 1 km by establishing a nonlinear relationship between AERONET AOD observations and collocated multisource satellite and ancillary variables. The predictor set comprised multispectral TOA reflectance, surface properties, observation geometry, atmospheric state, topography, and physically informed aerosol–surface features:

$$AOD550 = f\,(TOA_{1–7},\ BSA_{VIS},\ NDVI,\ Observation\ geometry,\ Atmosphere,\ DEM,\ AOD_{min},\ TSCR_{blue},\ TSCR_{red},\ TSCR_{SWIR})$$

where $TOA_{1–7}$ denotes MODIS TOA reflectance in the first seven reflective bands. Surface variables include BSA_{VIS} and NDVI, while the observation geometry variables comprise solar zenith angle, viewing zenith angle, relative azimuth angle, and scattering angle, which were included to represent illumination- and viewing-dependent variations in the coupled surface–atmosphere signal. Meteorological variables include TCWV, TCO3, and SP, and terrain information is represented by DEM-derived elevation. The physically informed features include AOD_{min} and TSCR in the blue, red, and shortwave-infrared bands. Prior to model training and retrieval, all predictor variables were spatially referenced to a common 1 km MODIS retrieval grid. Most spectral, surface, geometric, and topographic predictors were available at a native or nominal spatial resolution of 1 km (Table 1). ERA5 meteorological variables were spatially matched to the retrieval grid by assigning each ERA5 grid-cell value to all 1 km MODIS pixels located within that cell. The overall retrieval workflow is shown in Figure 1.

2.3. Construction of Physically Informed Features

Over land, TOA reflectance represents a coupled response of aerosol scattering and absorption, gaseous absorption, observation geometry, and surface-reflected radiation. This coupling makes the separation of aerosol and surface contributions inherently challenging, particularly across heterogeneous land surfaces and varying aerosol conditions [43,44]. To reduce the reliance of the statistical retrieval on empirical relationships learned from instantaneous observations alone, physical information was introduced at the predictor-construction stage through two complementary variables: AOD_{min} and TSCR. AOD_{min} was designed to represent the climatological low-aerosol background, whereas TSCR was designed to provide a band-specific constraint on the surface-reflectance background under relatively weak aerosol influence. Unlike simple algebraic or empirical transformations of instantaneous input variables, these predictors incorporate information from long-term aerosol observations, time-series satellite measurements, aerosol-type priors, and radiative-transfer calculations.

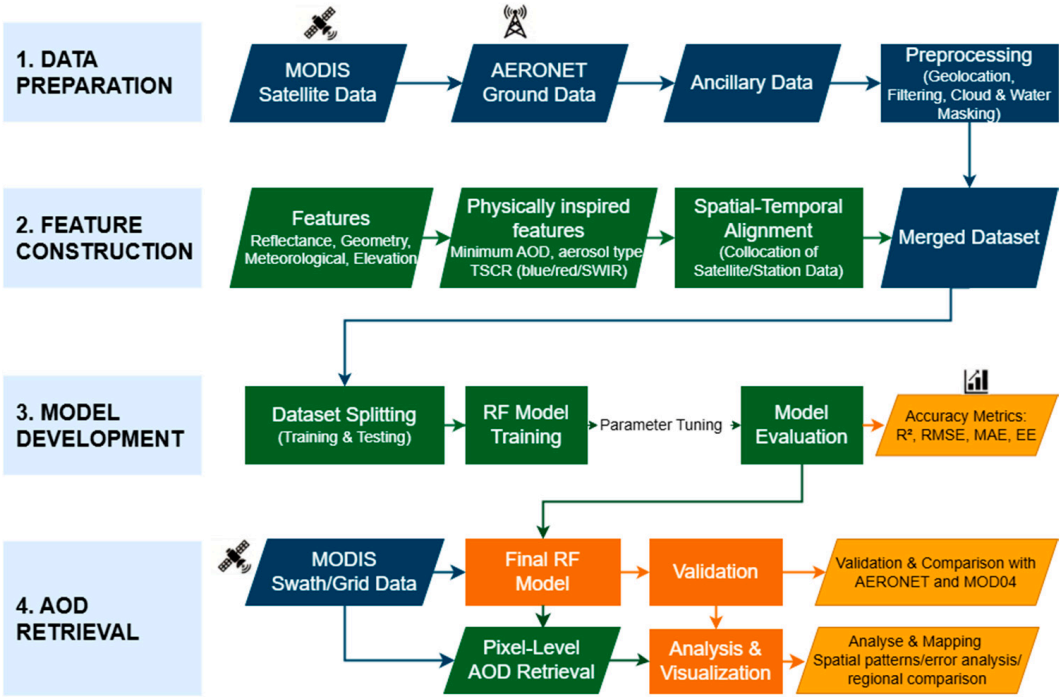


Figure 1. Workflow of the proposed 1 km aerosol optical depth retrieval framework.

AOD_{min} was derived to characterize the climatological low-aerosol background. Because a single global minimum-AOD value cannot adequately represent the diversity of aerosol environments over land, the global land area was divided into six broad regions according to the large-scale spatial and seasonal characteristics of AOD (Supplementary Figure S2). AOD_{min} was then estimated independently for each region and season using long-term, quality-controlled AERONET observations at 550 nm. For each AERONET site, observations were first grouped by year and season, and the minimum valid AOD was identified for each station–year–season combination. These annual seasonal minima were then averaged across all available years to obtain a station-level climatological low-AOD value for each season. Finally, the station-level values within each geographical region were averaged using equal station weighting to derive the corresponding regional–seasonal AOD_{min}. Station–year–season combinations without valid observations were excluded. This procedure produced 24 AOD_{min} values for six regions and four seasons (Supplementary Table S1). Each value was assigned to retrieval pixels according to their region and season and subsequently used as an aerosol-background prior in the TSCR estimation.

Seasonal aerosol type provided the second physical prior required for the radiative-transfer calculations. Seasonal aerosol-type fields were derived from the MOD04 aerosol product using a mode-composite approach. For each grid cell and season, all valid aerosol-type retrievals were collected, and the most frequently occurring aerosol class was assigned as the dominant type. Each class was associated with a corresponding set of aerosol optical properties, including single-scattering albedo and asymmetry factor (Supplementary Table S2), which were then used in the 6S radiative-transfer model [45]. Aerosol-type-dependent look-up tables (LUTs) were generated for the blue, red, and shortwave-infrared (SWIR) bands to describe the relationship between TOA reflectance and surface reflectance under the simulated aerosol and observation conditions.

In parallel, seasonal minimum TOA reflectance was derived from the time series of quality-controlled MODIS observations. For each pixel, season, and spectral band, the minimum valid clear-sky TOA reflectance was identified separately in the blue, red, and SWIR

bands. This quantity was used as an observational proxy for conditions with relatively weak aerosol influence and, together with AOD_{min} and the aerosol-type prior, provided the observational constraint for estimating the low-aerosol surface-reflectance background.

TSCR was then retrieved by combining the seasonal minimum TOA reflectance with the corresponding regional AOD_{min} through the 6S LUT associated with the dominant seasonal aerosol type. Specifically, for each pixel and spectral band, the appropriate aerosol-type-dependent LUT was first selected according to the seasonal aerosol-type field. The seasonal minimum TOA reflectance and the corresponding region–season-specific AOD_{min} were then used to identify the surface-reflectance value consistent with the 6S radiative-transfer relationship. The retrieved surface reflectance was defined as TSCR.

Although both BSA-VIS and TSCR describe aspects of the surface radiative background, they provide distinct information. BSA-VIS represents broadband visible directional-hemispherical albedo, whereas TSCR characterizes band-specific surface reflectance in the blue, red, and SWIR bands after accounting for the residual aerosol contribution associated with AOD_{min}. BSA-VIS therefore provides a broadband surface-albedo constraint, while TSCR retains additional spectral information relevant to aerosol–surface signal separation. The two predictor groups are physically related but complementary. The final physically informed predictor set comprised regional AOD_{min} and TSCR in the blue, red, and SWIR bands.

2.4. Model Development

A random forest (RF) regression model was used to learn the nonlinear mapping between multisource predictors and AOD [46]. RF can accommodate heterogeneous predictors without feature scaling, capture nonlinear interactions among variables, and provide stable performance at moderate computational cost. In addition, its feature-importance diagnostics facilitate the interpretation of the relative contributions of individual predictors. These characteristics make RF well suited to the large-scale AOD retrieval framework developed in this study. The RF optimal parameters were determined using out-of-bag (OOB) error. The final model consisted of 160 trees, with the maximum tree depth set to 20. All predictor variables were considered when searching for the best split, the minimum number of samples required to split an internal node was 2, and the minimum number of samples per leaf was 1. After wavelength harmonization, spatiotemporal matching, cloud screening, and invalid-value filtering, the available samples were separated temporally into a model-development dataset and an independent evaluation dataset. Samples from 2001 to 2012 were used for model development, while a subset of the 2015–2017 samples was reserved for independent temporal validation. After the model configuration was fixed, all available samples from 2001 to 2012 and 2015–2017 were combined to train the final RF model, whereas the 2013–2014 samples were excluded from model construction and used exclusively for independent retrieval evaluation.

2.5. Validation

Model performance was evaluated in two stages. In the first stage, model fitting, 10-fold cross-validation, and independent temporal validation were used to evaluate the predictive performance and temporal generalizability of the RF model. Samples from 2001 to 2012 constituted the model-development dataset. For 10-fold cross-validation, these samples were randomly partitioned into ten mutually exclusive subsets. In each iteration, nine subsets were used for model training and the remaining subset for validation, such that each sample was used once for validation. In addition, a subset of the collocated samples from 2015 to 2017 was withheld as an independent temporal validation dataset and was not

used for model fitting or parameter selection. The RF model developed using the 2001–2012 samples was subsequently evaluated against these withheld 2015–2017 observations.

In the second stage, after the RF configuration had been fixed, all available samples from 2001 to 2012 and 2015–2017, including those reserved for independent temporal validation in the first stage, were combined to train the final retrieval model. The final model was then applied to MODIS observations from 2013 to 2014 to retrieve AOD. AERONET observations from 2013 to 2014 were excluded from model construction and used only as reference measurements for evaluating the retrieved AOD. Accordingly, all subsequent retrieval assessments, including station-level evaluation, overall comparison with AERONET, comparison with MOD04 products, and analysis of retrieval errors as a function of AOD, were based on the 2013–2014 retrieval results.

For satellite–ground collocation, AERONET observations acquired within $\pm$30 min of the MODIS overpass were averaged to reduce temporal sampling differences between satellite and ground-based measurements. The retrieved AOD from the 1 km grid cell containing each AERONET site was then matched with the corresponding temporally averaged AERONET AOD. Performance was quantified using the Pearson correlation coefficient (R), root mean square error (RMSE), mean absolute error (MAE), relative mean bias (RMB), and the fraction of samples falling within the expected-error envelope (EE). The EE was defined as $\pm(0.05 + 0.15 \times \text{AOD_AERONET})$, where AOD_AERONET denotes the collocated AERONET AOD at 550 nm.

In addition to the overall sample-level evaluation, station-level validation was performed globally using the 2013–2014 retrieval results to assess regional retrieval performance under different aerosol and surface-background conditions. The proposed retrievals were also compared with the MOD04 Dark Target (DT) and Deep Blue (DB) products using collocated AERONET observations, with differences evaluated in terms of RMSE, MAE, and EE.

3. Results

3.1. Model Fitting and Validation

The RF model showed strong fitting performance during training. For the fitting samples, model-predicted AOD agreed closely with AERONET observations, with R = 0.99, RMSE = 0.037, MAE = 0.022, and 98.49% of samples falling within the EE (Figure 2a). The 10-fold cross-validation performance was lower than the fitting result but remained stable, with R = 0.81, RMSE = 0.109, MAE = 0.067, and 74.1% of the samples within the EE (Figure 2b).

To further assess the temporal generalizability of the RF model, samples from 2001 to 2012 were used for model development, while a subset of the collocated samples from 2015 to 2017 was withheld for independent temporal validation. The independent validation yielded R = 0.789, RMSE = 0.116, MAE = 0.071, with 70.84% of the samples falling within the EE (Figure 3). Although the validation performance was slightly lower than that obtained from 10-fold cross-validation, the overall agreement was comparable, indicating that the RF model maintained relatively stable predictive performance when applied to temporally independent observations from years not included in model development.

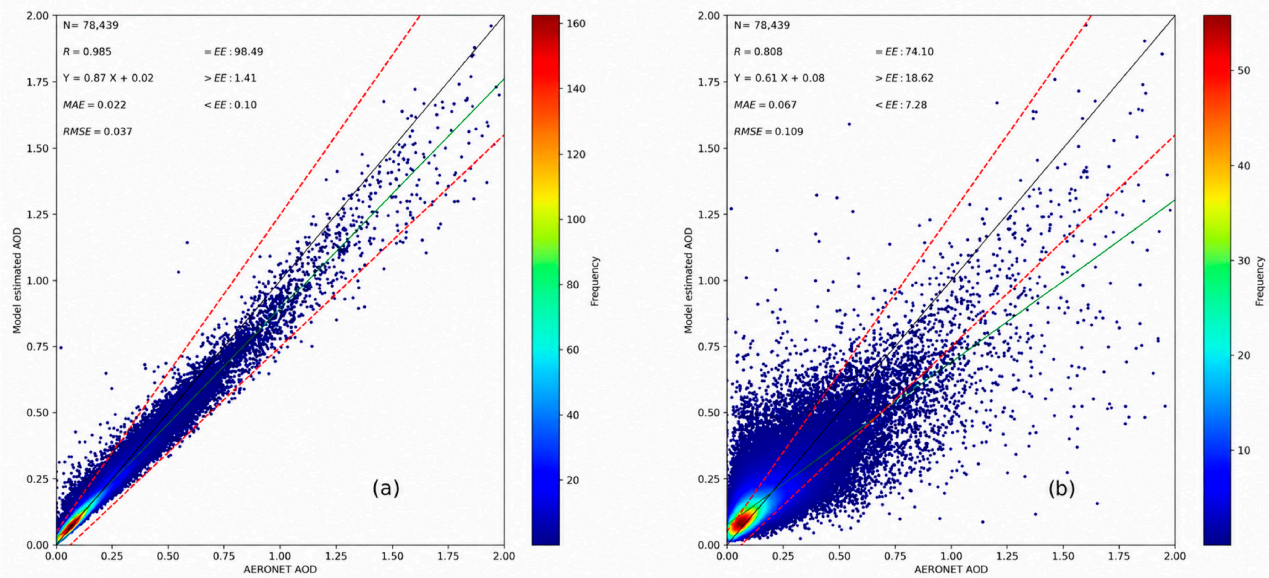


Figure 2. Random forest model performance. (**a**) Model fitting; (**b**) 10-fold cross-validation. Color scale indicates the density of observation. The red dashed line, the black and green solid lines represent the MODIS expected error, the 1:1 line and the linear regression line, respectively.

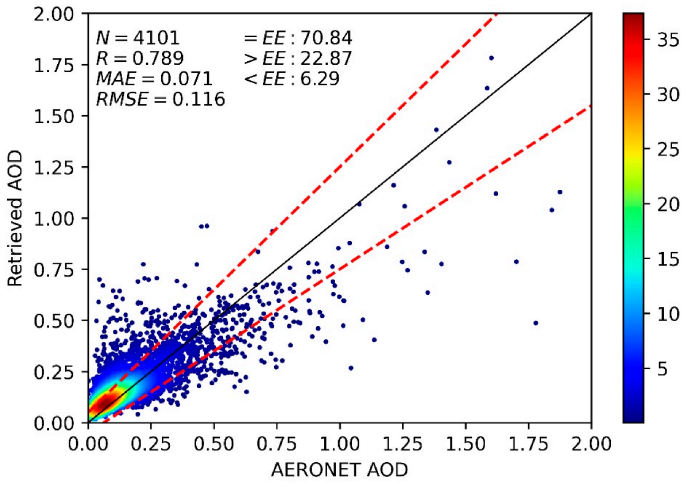


Figure 3. Density scatterplot for independent temporal validation using the 2015–2017 samples. The color scale indicates sample density, the red dashed lines show the MODIS expected-error envelope, and the black solid line represents y = x.

To quantify the contribution of individual predictor groups, controlled leave-one-group-out ablation experiments were performed using the full model as the reference (Table 2). The full model achieved R = 0.79, RMSE = 0.116, MAE = 0.071, and 70.84% of the samples within the EE (Figure 3). Removing TOA reflectance produced the largest deterioration in retrieval performance, with R decreasing to 0.665, RMSE increasing to 0.153, and the within-EE fraction decreasing to 57.80%, demonstrating that multispectral TOA observations provide the primary information for AOD retrieval. The removal of surface variables and DEM also resulted in a substantial performance loss (R = 0.75, RMSE = 0.127, and within EE = 66.91%), highlighting the importance of characterizing surface reflectance and topographic variability. The physically informed features also contributed appreciably to the retrieval. Removing the regional minimum AOD reduced R from 0.79 to 0.74 and increased RMSE from 0.116 to 0.128, while the within-EE fraction decreased from 70.84% to 66.65%. Removing TSCR resulted in R = 0.76, RMSE = 0.123, and 68.10% within EE. When

regional minimum AOD and TSCR were simultaneously excluded, the performance further degraded to R = 0.716, RMSE = 0.137, MAE = 0.083, and 63.40% within EE. The larger degradation caused by removing both feature groups indicates that regional minimum AOD and TSCR provide complementary constraints on the aerosol background and surface-reflectance state, respectively, with regional minimum AOD showing a somewhat larger marginal contribution.

Table 2. Controlled leave-one-group-out ablation results for quantifying the marginal contribution of different predictor groups.

Experiment	R	RMSE	MAE	Within EE (%)
Full model	0.79	0.116	0.071	70.84
w/o Meteorology	0.78	0.119	0.073	69.60
w/o Surface + DEM	0.75	0.127	0.078	66.91
w/o TSCR	0.76	0.123	0.075	68.10
w/o minAOD	0.74	0.128	0.078	66.65
w/o minAOD and TSCR	0.716	0.137	0.083	63.40
w/o TOA reflectance	0.665	0.153	0.094	57.80

The group-level RF feature importance shows a consistent pattern (Supplementary Figure S3). TOA reflectance accounted for the largest proportion of the total importance (30%), followed by regional minimum AOD (18%), TSCR (15%), surface variables and DEM (15%), observation geometry (14%), and meteorological variables (8%). Regional minimum AOD and TSCR together accounted for 33% of the total RF feature importance, further indicating that the physically informed predictors provide substantial information in addition to the conventional spectral, surface, geometric, and meteorological inputs.

The cumulative ablation experiments further illustrate how retrieval performance evolved as predictor groups were progressively introduced (Supplementary Table S3). The baseline model using only TOA reflectance and observation geometry yielded R = 0.50, RMSE = 0.208, MAE = 0.123, and 56.32% within EE. Adding meteorological variables produced only a modest improvement, whereas the subsequent inclusion of surface variables and DEM increased R from 0.54 to 0.67 and reduced RMSE from 0.201 to 0.158. Incorporating regional minimum AOD further increased R to 0.71 and reduced RMSE to 0.131, and the addition of TSCR resulted in the final performance of R = 0.79, RMSE = 0.116, MAE = 0.071, and 70.84% within EE. The cumulative and controlled ablation experiments therefore provide complementary evidence that the retrieval benefits from both conventional multisource predictors and the physically informed aerosol–surface constraints.

3.2. Evaluation of 2013–2014 AOD Retrieval Performance

3.2.1. Spatial Patterns of Station-Level Accuracy

This section evaluates the spatial variability in AOD retrieval performance during 2013–2014. Figure 4 presents the global station-level distributions of RMSE, MAE, RMB, and EE relative to AERONET observations. For RMSE (Figure 4a), approximately 70.2% of AERONET sites (80 of 114) exhibited values below 0.11. Sites with relatively low RMSE were concentrated mainly in North America, South America, and Europe, where AERONET coverage and sample availability were generally greater and aerosol conditions were comparatively stable. In contrast, higher RMSE values occurred primarily over India, China, and regions influenced by African desert dust. These larger errors may reflect the combined effects of high aerosol loading, frequent dust events, strong anthropogenic

emissions, limited ground-based sampling, and greater spatiotemporal variability in aerosol optical properties, all of which increase the complexity of AOD retrieval.

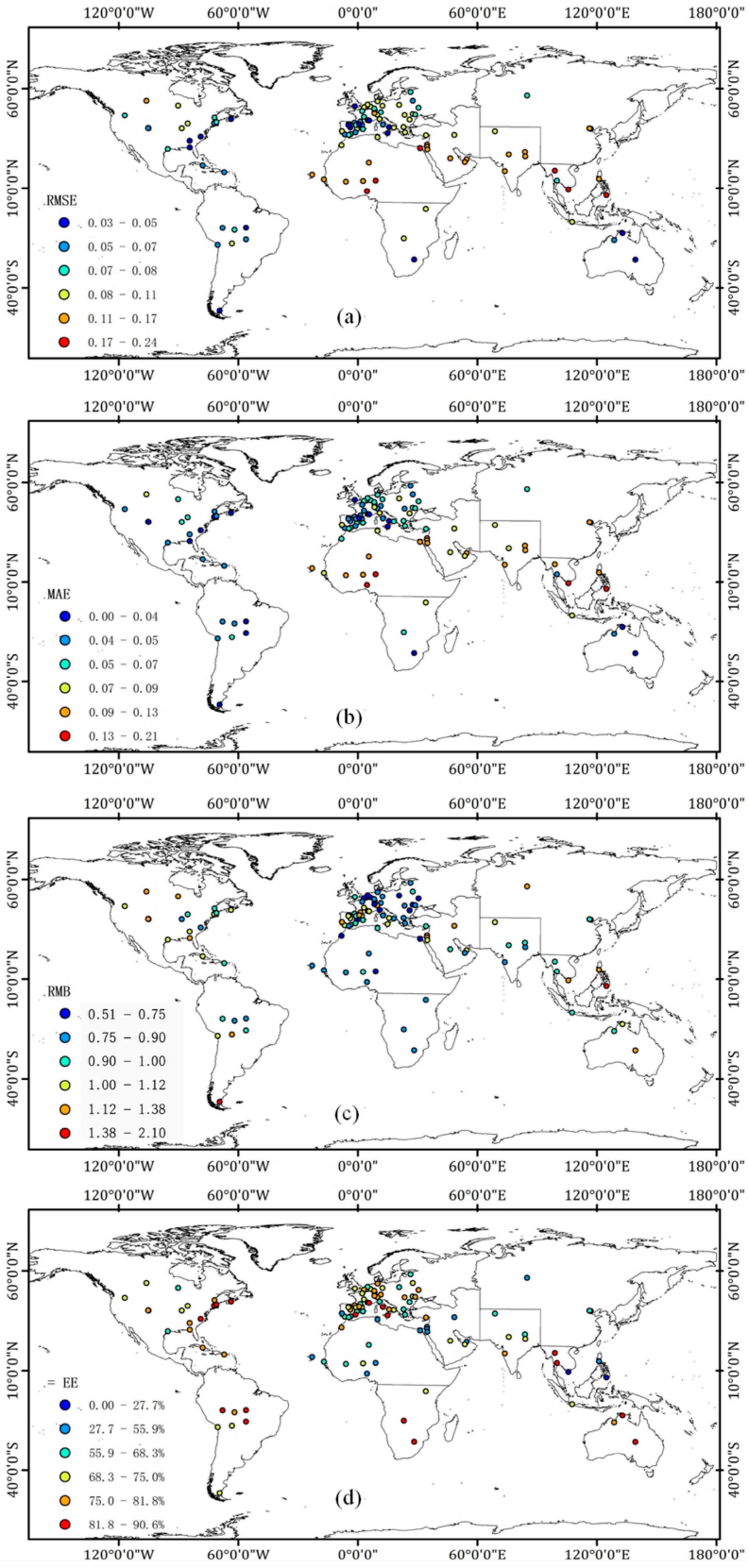


Figure 4. Station-level validation between retrieved AOD and AERONET AOD at 550 nm over land for 2013–2014. From top to bottom: (**a**) RMSE, (**b**) MAE, (**c**) RMB, and (**d**) EE.

For MAE (Figure 4b), approximately 71.9% of stations (82 of 114) had MAE values no greater than 0.08, again mainly in North America, South America, and Europe. Higher MAE values were concentrated in India and China, implying that retrieval uncertainty increases in regions with complex land surfaces, high surface reflectance, and strong anthropogenic influence. RMB was used to evaluate systematic retrieval bias, with RMB = 1 indicating an unbiased retrieval, RMB < 1 indicating underestimation, and RMB > 1 indicating overestimation. The mean station-level RMB was 1.07, indicating slight overall overestimation (Figure 4c), while approximately 42.1% of stations showed underestimation. Underestimation was evident over parts of Europe, Africa, Asia, and eastern North America, whereas overestimation was observed in western North America, South America, and parts of South Asia.

For EE (Figure 4d), approximately 65.7% of stations (75 of 114) had more than 70% of samples falling within the EE. These stations were mainly located in eastern North America, South America, Europe, and Oceania. In contrast, EE values were relatively low over North Africa, South Asia, India, and China, where surface–atmosphere interactions are more complex, aerosol sources are more diverse, and aerosol loading can vary rapidly.

3.2.2. Overall Retrieval Accuracy Against AERONET

To further assess the overall retrieval accuracy during the independent 2013–2014 period, the retrieved AOD values were compared with collocated AERONET AOD observations at the sample level. Figure 5 presents the density scatterplot between the retrieved AOD and AERONET AOD, together with the 1:1 line and EE. The retrieved AOD generally showed good agreement with AERONET observations, with most samples distributed close to the 1:1 line. Approximately 74.93% of the matched samples fell within the EE. The overall error statistics further confirm the retrieval performance, with MAE = 0.063 and RMSE = 0.096 for the independent assessment. The relatively low MAE and RMSE suggest that the proposed physically informed random forest framework can effectively retrieve AOD over global land areas under clear-sky conditions.

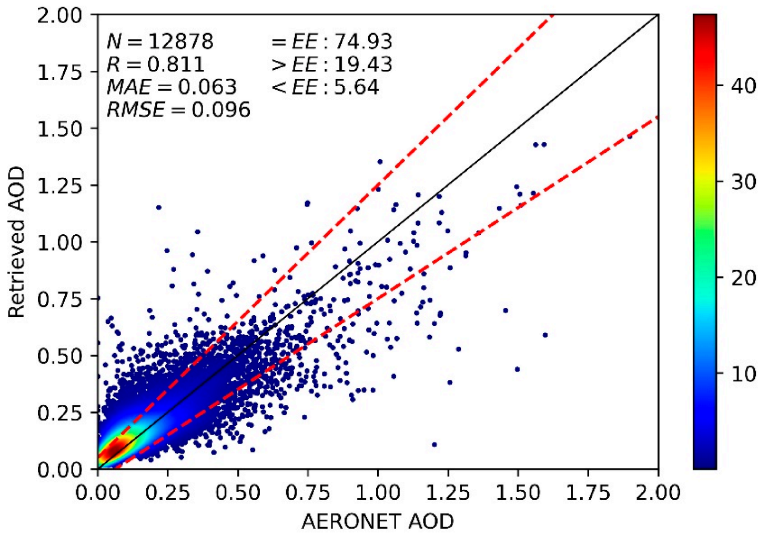


Figure 5. Density scatterplot of retrieved AOD and AERONET AOD observations. The color scale indicates sample density, the red dashed lines show the MODIS expected-error envelope, and the black solid line represents y = x.

3.2.3. Comparison with MOD04 DT and DB Products

The proposed retrievals were compared with MOD04 DT and DB aerosol products using matched AERONET observations. DB retrievals agreed reasonably well with

AERONET, with more than 67.44% of data points falling within the EE and MAE and RMSE values below 0.086 and 0.14, respectively (Figure 6a). The retrievals from this study showed stronger agreement with AERONET, with approximately 72.58% of samples satisfying the accuracy requirement and lower MAE and RMSE values (0.06 and 0.10, respectively). DB exhibited more frequent overestimation. MOD04 10 km DT also showed overestimation; approximately 71.75% of data points fell within EE, with RMSE = 0.10 and MAE = 0.07 (Figure 6b). In comparison, the proposed retrieval achieved a higher EE fraction (77.57%) and lower MAE and RMSE values (0.06 and 0.09, respectively). Overall, the proposed 1 km AOD retrievals showed better agreement with AERONET than the MOD04 DT and DB products in these validation experiments.

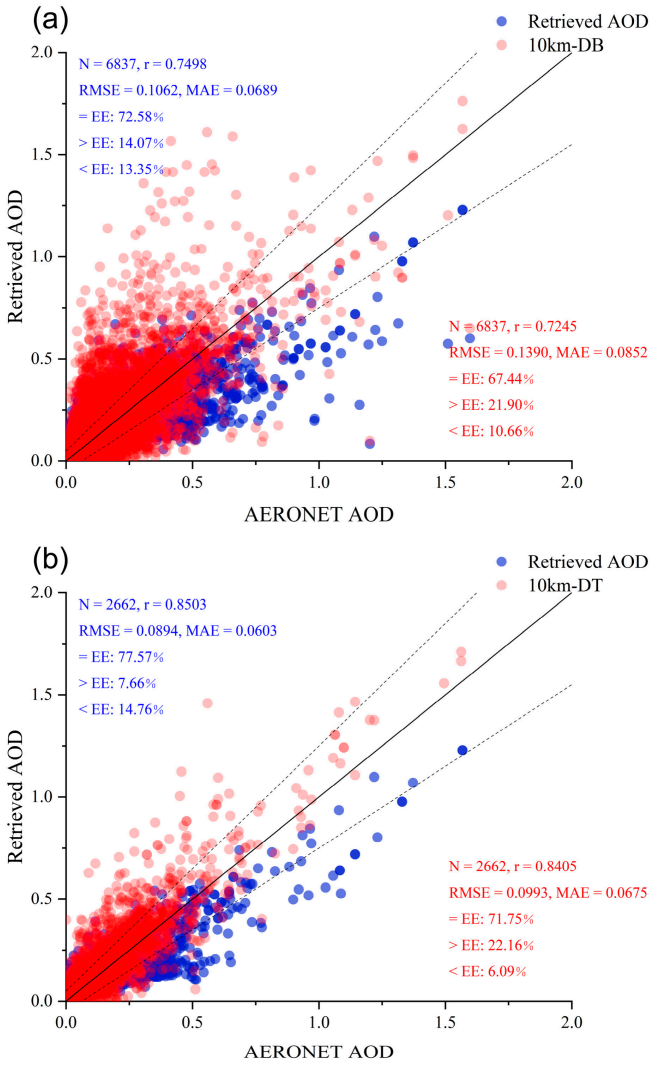


Figure 6. Validation of the retrieved AOD and MOD04 10-km AOD products against collocated AERONET observations. (**a**) Retrieved AOD (blue) and MOD04 Deep Blue (DB) AOD (red). (**b**) Retrieved AOD (blue) and MOD04 Dark Target (DT) AOD (red). The solid black line denotes the 1:1 line, while the dotted lines represent the expected error (EE) envelope. Validation statistics are shown in each panel.

3.3. Spatial Distribution of Retrieved AOD

3.3.1. Global Spatial Distribution of Retrieved AOD

Figure 7 shows the global distribution of annual mean AOD retrieved at 1 km spatial resolution over land. The retrievals exhibit clear large-scale spatial variability while preserving substantial regional detail. Relatively high AOD values are evident over several major

aerosol-affected regions, particularly South and East Asia, including the Indo-Gangetic Plain and eastern China. Elevated AOD is also observed over parts of western and central Africa, especially regions influenced by persistent biomass-burning and dust-related aerosol loading. Moderate-to-high AOD occurs over portions of northern Eurasia and Central Asia, although the spatial extent and magnitude are generally lower than those observed over the major Asian and African hotspots. In contrast, relatively low annual mean AOD dominates large areas of North America, South America, Europe, southern Africa, and Australia. These continental-scale contrasts are accompanied by pronounced spatial gradients within individual regions, particularly across South Asia, East Asia, and Africa, illustrating the ability of the 1 km retrieval to resolve subregional aerosol variability that may be obscured in coarser-resolution products. Overall, the retrieved AOD field reproduces the expected broad-scale geographical pattern of aerosol loading while providing spatially continuous information over diverse land-surface conditions.

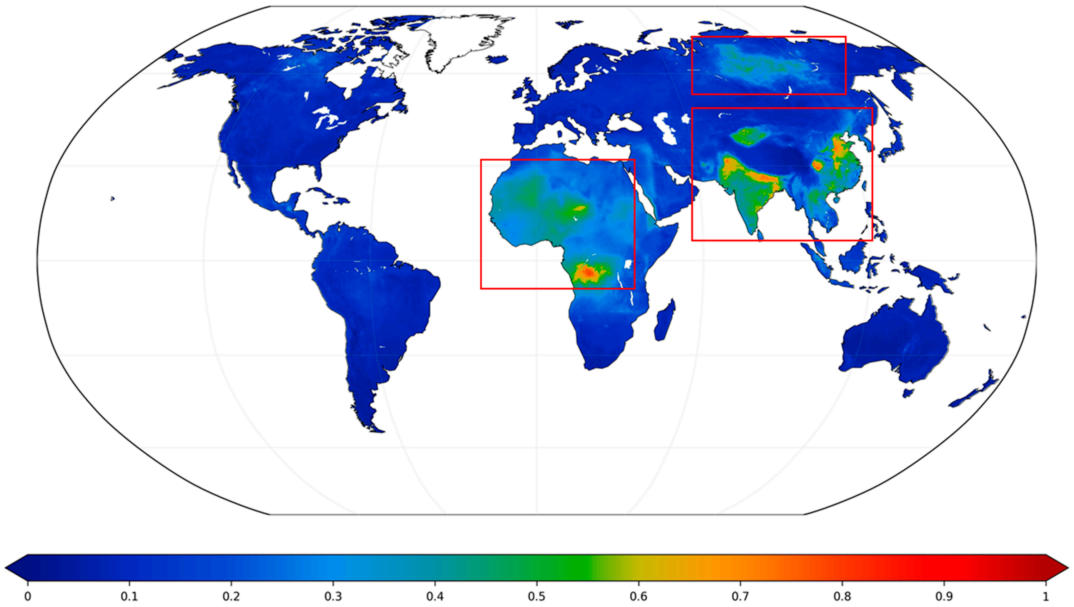


Figure 7. Global distribution of mean retrieved AOD at 550 nm at 1 km spatial resolution for 2013–2014. The red boxes indicate selected regions characterized by relatively high mean AOD.

3.3.2. Regional Spatial Patterns of Retrieved AOD

Figure 8 presents representative examples of the spatial distributions of the retrieved 1 km AOD, MAIAC 1 km AOD, and MOD04 10 km AOD for several MODIS overpass dates. The proposed retrieval generally reproduces the major aerosol spatial patterns observed in both MAIAC and MOD04, while resolving finer-scale spatial variability than the coarser MOD04 product. In particular, the 1 km retrieval captures localized AOD gradients near coastlines, urbanized areas, and heterogeneous land-cover transitions, where aerosol loading and surface reflectance can vary over short distances. Compared with MAIAC, the proposed retrieval exhibits more spatially continuous AOD fields and broader valid-pixel coverage in several of the selected scenes, with fewer retrieval gaps and a more complete representation of regional aerosol structures. Differences in AOD magnitude remain among the three products, which may reflect differences in surface-reflectance characterization, aerosol assumptions, and quality-control procedures. Overall, these examples indicate that the proposed framework can provide spatially coherent and comparatively complete 1 km AOD fields while preserving physically plausible aerosol patterns.

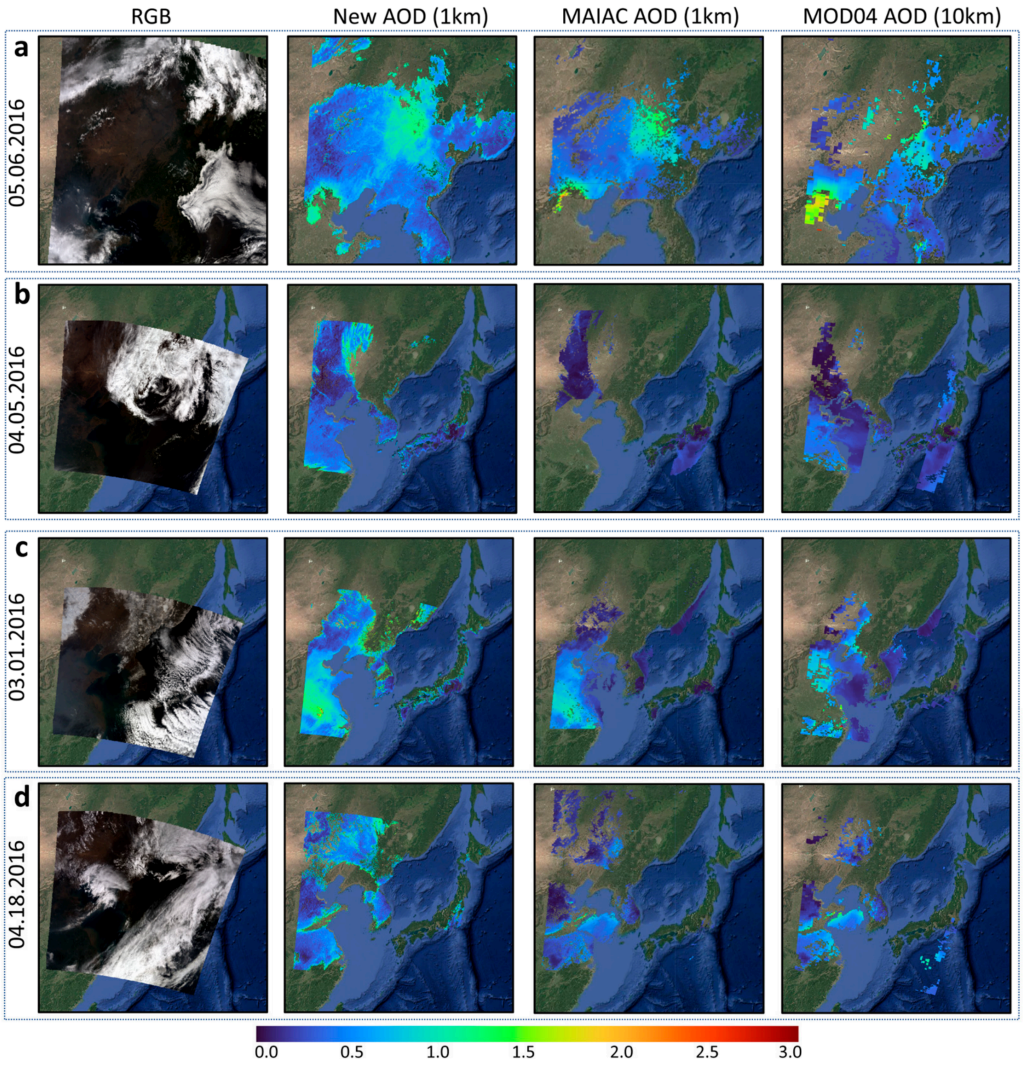


Figure 8. Spatial comparison of the retrieved 1-km AOD with MAIAC 1-km AOD and MOD04 10-km AOD for selected MODIS overpasses: (**a**) 6 May 2016, (**b**) 5 April 2016, (**c**) 1 March 2016, and (**d**) 18 April 2016. For each case, the columns from left to right show the MODIS RGB image, the 1-km AOD retrieved in this study, the MAIAC 1-km AOD, and the MOD04 10-km AOD, respectively. The RGB images are provided for visual reference, and a common color scale is used for all AOD products.

3.4. Dependence of Retrieval Error on AERONET AOD

To further evaluate the error characteristics under different aerosol loading levels, the AOD bias, defined as retrieved AOD minus AERONET AOD, and the within-EE fraction were analyzed as functions of AERONET AOD. The AERONET AOD values were divided into thirteen bins within the range of 0.0–2.0, and the corresponding bias distributions and within-EE percentages are shown in Figure 9.

The retrieval bias remained close to zero under low-to-moderate aerosol loading conditions, especially when AERONET AOD was below approximately 0.4–0.5. In these bins, the bias distribution was relatively narrow and the within-EE fraction was comparatively high, indicating stable retrieval performance for most clear-sky cases with low or moderate aerosol loading. However, as AERONET AOD increased, particularly above approximately 1, the retrievals showed an increasingly negative bias accompanied by a decline in the within-EE fraction.

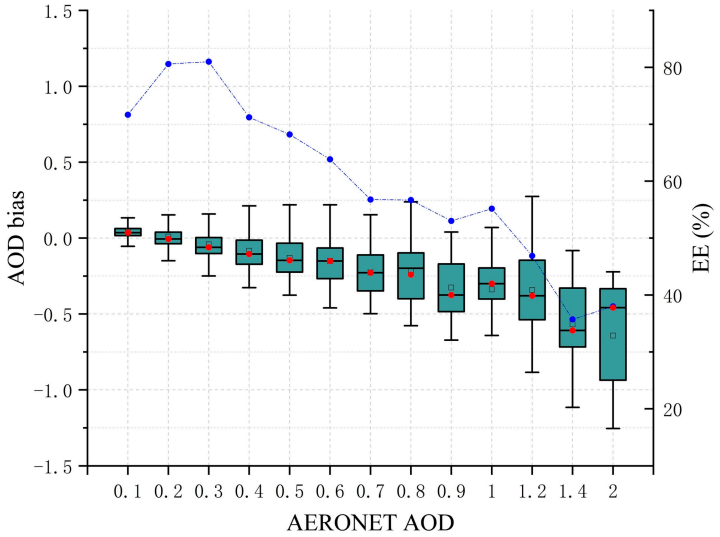


Figure 9. Boxplots of AOD bias and the percentage of samples within the expected-error envelope (EE; blue curve) as functions of AERONET AOD at 550 nm. In each box, red dots indicate mean bias, while the middle, lower, and upper horizontal lines represent the median, 25th percentile, and 75th percentile, respectively.

The sample distribution indicates that high-AOD conditions were sparsely represented in the independent evaluation dataset. Among the 12,878 collocated samples, 86.33% had AERONET AOD below 0.3 and 95.87% had AOD below 0.5, whereas only 183 samples (1.42%) had AOD $\geq$ 0.7 and only 53 samples (0.41%) had AOD $\geq$ 1.0 (Supplementary Table S4). Consequently, the statistics for the highest-AOD bins should be interpreted with caution because of their relatively small sample sizes. The observed high-AOD underestimation may reflect both limited sample representation and increased retrieval complexity under heavy aerosol loading. In particular, stronger aerosol scattering and absorption and greater uncertainty in aerosol-type representation make it more difficult to separate atmospheric aerosol effects from land-surface reflection. In addition, the limited representation of extreme aerosol conditions may restrict the ability of the RF model to adequately characterize severe pollution and dust events. Future improvements should therefore focus on increasing the representation of high-AOD cases and refining aerosol-type constraints, particularly in regions frequently affected by severe pollution and dust events. Future improvements could address the high-AOD underestimation from both physical and data-driven perspectives. Physically, more flexible aerosol-type priors and a broader range of aerosol optical properties in the 6S LUTs could improve the representation of severe pollution and dust conditions. From the machine learning perspective, AOD-stratified sampling or increased weighting of high-AOD observations could enhance the influence of these relatively rare cases during model training and reduce the regression toward the more densely sampled low-to-moderate AOD range.

4. Conclusions

This study developed a physically informed random forest framework for global 1 km land AOD retrieval from MODIS. The framework integrates multispectral TOA reflectance, surface albedo, NDVI, observation geometry, ERA5 meteorological variables, topography, and physically informed aerosol–surface features represented by regional minimum AOD and time-series clear-sky reflectance. The model showed stable predictive performance in both cross-validation and independent temporal validation. Validation using withheld samples from 2015 to 2017 yielded R = 0.789, RMSE = 0.116, MAE = 0.071, with 70.84% of samples falling within the expected-error (EE) envelope. The final model was subse-

quently applied to retrieve AOD for 2013–2014, yielding MAE = 0.063, RMSE = 0.096, and 74.93% of collocated samples within the EE when evaluated against AERONET observations. Station-level evaluation showed generally lower retrieval errors across North America, South America, and Europe, whereas larger errors occurred over India, China, and dust-influenced regions with more complex aerosol and surface conditions. Comparisons with MOD04 Dark Target and Deep Blue products further showed that the proposed retrievals achieved comparable or improved agreement with AERONET while providing AOD estimates at 1 km spatial resolution. Overall, the results demonstrate the value of integrating physically interpretable aerosol-background and surface-reflectance information into data-driven AOD retrieval. Remaining uncertainties are most evident under high aerosol loading, where retrievals tend to underestimate AOD. Future work should therefore focus on improving the representation of high-AOD conditions, refining aerosol-type priors, incorporating complementary satellite observations, and developing pixel-level uncertainty estimates.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18172862/s1, Figure S1: Distribution of the selected AERONET sites; Figure S2: Continent distribution diagram; Figure S3: Feature importance derived from the full RF model. The importance of each predictor group was calculated by summing the normalized importance scores of the individual variables belonging to that group; Table S1: Statistical results of seasonal AOD minimum values; Table S2: Aerosol type parameters; Table S3: Cumulative ablation experiments showing the progressive contribution of predictor groups to RF-based AOD retrieval; Table S4: Distribution of collocated samples across AERONET AOD intervals.

Author Contributions: T.L.: Writing—original draft, Methodology, Investigation, Formal analysis, Data curation, Visualization, Funding. L.Z.: Writing—original draft, Supervision, Investigation, Formal analysis, Conceptualization. Q.H.: Writing—review and editing. L.S.: Writing—review and editing, Resources, Project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Huaiyin Normal University Research and Innovation Fund [grant number 31LTC00].

Data Availability Statement: Data will be made available on request.

Acknowledgments: We gratefully acknowledge the NASA MODIS science and data-processing teams for providing the MODIS observations used in this study. We thank the AERONET principal investigators and site managers for establishing and maintaining the sites and for providing the ground-based aerosol observations. We also acknowledge the GLASS product team for the land-surface albedo data, the European Centre for Medium-Range Weather Forecasts/Copernicus Climate Change Service for the ERA5 reanalysis data, and the relevant data providers for the GMTED2010 elevation dataset.

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

Abbreviations

Abbreviation	Definition
AOD	Aerosol optical depth
AERONET	Aerosol Robotic Network
MODIS	Moderate Resolution Imaging Spectroradiometer
TOA	Top-of-atmosphere
RF	Random forest
GLASS	Global Land Surface Satellite
BSA-VIS	Visible black-sky albedo
ERA5	Fifth-generation ECMWF atmospheric reanalysis

SP	Surface pressure
TCO3	Total column ozone
TCWV	Total column water vapor
NDVI	Normalized difference vegetation index
DEM	Digital elevation model
TSCR	Time-series clear-sky reflectance
LUT	Look-up table
OOB	Out-of-bag
CV	Cross-validation
RMSE	Root mean square error
MAE	Mean absolute error
RMB	Relative mean bias
EE	Expected error
DT	Dark Target
DB	Deep Blue
MAIAC	Multi-Angle Implementation of Atmospheric Correction
6S	Second Simulation of a Satellite Signal in the Solar Spectrum

References

1. Carslaw, K.S.; Boucher, O.; Spracklen, D.V.; Mann, G.W.; Rae, J.G.L.; Woodward, S.; Kulmala, M. A review of natural aerosol interactions and feedbacks within the Earth system. *Atmos. Chem. Phys.* **2010**, *10*, 1701–1737. [CrossRef]
2. Pöschl, U. Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects. *Angew. Chem. Int. Ed.* **2005**, *44*, 7520–7540. [CrossRef] [PubMed]
3. Bellouin, N.; Quaas, J.; Gryspeerdt, E.; Kinne, S.; Stier, P.; Watson-Parris, D.; Boucher, O.; Carslaw, K.S.; Christensen, M.; Daniau, A.-L.; et al. Bounding Global Aerosol Radiative Forcing of Climate Change. *Rev. Geophys.* **2020**, *58*, e2019RG000660. [CrossRef] [PubMed]
4. Carslaw, K.S.; Lee, L.A.; Reddington, C.L.; Pringle, K.J.; Rap, A.; Forster, P.M.; Mann, G.W.; Spracklen, D.V.; Woodhouse, M.T.; Regayre, L.A.; et al. Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature* **2013**, *503*, 67–71. [CrossRef] [PubMed]
5. Kaufman, Y.J.; Tanré, D.; Boucher, O. A satellite view of aerosols in the climate system. *Nature* **2002**, *419*, 215–223. [CrossRef] [PubMed]
6. Myhre, G.; Samset, B.H.; Schulz, M.; Balkanski, Y.; Bauer, S.; Berntsen, T.K.; Bian, H.; Bellouin, N.; Chin, M.; Diehl, T.; et al. Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations. *Atmos. Chem. Phys.* **2013**, *13*, 1853–1877. [CrossRef]
7. Stevens, B.; Feingold, G. Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature* **2009**, *461*, 607–613. [CrossRef] [PubMed]
8. van Donkelaar, A.; Martin, R.V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P.J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, 847–855. [CrossRef] [PubMed]
9. Vermote, E.F.; Kotchenova, S. Atmospheric correction for the monitoring of land surfaces. *J. Geophys. Res. Atmos.* **2008**, *113*, D23S90. [CrossRef]
10. van Donkelaar, A.; Martin, R.V.; Brauer, M.; Hsu, N.C.; Kahn, R.A.; Levy, R.C.; Lyapustin, A.; Sayer, A.M.; Winker, D.M. Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. *Environ. Sci. Technol.* **2016**, *50*, 3762–3772. [CrossRef] [PubMed]
11. Wei, X.; Chang, N.-B.; Bai, K.; Gao, W. Satellite remote sensing of aerosol optical depth: Advances, challenges, and perspectives. *Crit. Rev. Environ. Sci. Technol.* **2020**, *50*, 1640–1725. [CrossRef]
12. Dubovik, O.; Smirnov, A.; Holben, B.N.; King, M.D.; Kaufman, Y.J.; Eck, T.F.; Slutsker, I. Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) Sun and sky radiance measurements. *J. Geophys. Res. Atmos.* **2000**, *105*, 9791–9806. [CrossRef]
13. Giles, D.M.; Sinyuk, A.; Sorokin, M.G.; Schafer, J.S.; Smirnov, A.; Slutsker, I.; Eck, T.F.; Holben, B.N.; Lewis, J.R.; Campbell, J.R.; et al. Advancements in the Aerosol Robotic Network (AERONET) Version 3 database—Automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmos. Meas. Tech.* **2019**, *12*, 169–209. [CrossRef]

14. Holben, B.N.; Eck, T.F.; Slutsker, I.; Tanré, D.; Buis, J.P.; Setzer, A.; Vermote, E.; Reagan, J.A.; Kaufman, Y.J.; Nakajima, T.; et al. AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sens. Environ.* **1998**, *66*, 1–16. [CrossRef]
15. Remer, L.A.; Kaufman, Y.J.; Tanré, D.; Mattoo, S.; Chu, D.A.; Martins, J.V.; Li, R.-R.; Ichoku, C.; Levy, R.C.; Kleidman, R.G.; et al. The MODIS Aerosol Algorithm, Products, and Validation. *J. Atmos. Sci.* **2005**, *62*, 947–973. [CrossRef]
16. Remer, L.A.; Kleidman, R.G.; Levy, R.C.; Kaufman, Y.J.; Tanré, D.; Mattoo, S.; Martins, J.V.; Ichoku, C.; Koren, I.; Yu, H.; et al. Global aerosol climatology from the MODIS satellite sensors. *J. Geophys. Res. Atmos.* **2008**, *113*, D14S07. [CrossRef]
17. Hsu, N.C.; Jeong, M.-J.; Bettenhausen, C.; Sayer, A.M.; Hansell, R.; Seftor, C.S.; Huang, J.; Tsay, S.-C. Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *J. Geophys. Res. Atmos.* **2013**, *118*, 9296–9315. [CrossRef]
18. Levy, R.C.; Remer, L.A.; Mattoo, S.; Vermote, E.F.; Kaufman, Y.J. Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *J. Geophys. Res. Atmos.* **2007**, *112*, D13211. [CrossRef]
19. 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. *Atmos. Meas. Tech.* **2013**, *6*, 2989–3034. [CrossRef]
20. Remer, L.A.; Levy, R.C.; Mattoo, S.; Tanré, D.; Gupta, P.; Shi, Y.; Sawyer, V.; Munchak, L.A.; Zhou, Y.; Kim, M.; et al. The Dark Target Algorithm for Observing the Global Aerosol System: Past, Present, and Future. *Remote Sens.* **2020**, *12*, 2900. [CrossRef]
21. Sayer, A.M.; Hsu, N.C.; Bettenhausen, C.; Jeong, M.-J. Validation and uncertainty estimates for MODIS Collection 6 "Deep Blue" aerosol data. *J. Geophys. Res. Atmos.* **2013**, *118*, 7864–7872. [CrossRef]
22. Lyapustin, A.; Martonchik, J.; Wang, Y.; Laszlo, I.; Korkin, S. Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *J. Geophys. Res. Atmos.* **2011**, *116*, D03210. [CrossRef]
23. Lyapustin, A.; Wang, Y.; Laszlo, I.; Kahn, R.; Korkin, S.; Remer, L.; Levy, R.; Reid, J.S. Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *J. Geophys. Res. Atmos.* **2011**, *116*, D03211. [CrossRef]
24. Bilal, M.; Nichol, J.E.; Bleiweiss, M.P.; Dubois, D. A Simplified high resolution MODIS Aerosol Retrieval Algorithm (SARA) for use over mixed surfaces. *Remote Sens. Environ.* **2013**, *136*, 135–145. [CrossRef]
25. Remer, L.A.; Mattoo, S.; Levy, R.C.; Munchak, L.A. MODIS 3 km aerosol product: Algorithm and global perspective. *Atmos. Meas. Tech.* **2013**, *6*, 1829–1844. [CrossRef]
26. Gupta, P.; Remer, L.A.; Levy, R.C.; Mattoo, S. Validation of MODIS 3 km land aerosol optical depth from NASA's EOS Terra and Aqua missions. *Atmos. Meas. Tech.* **2018**, *11*, 3145–3159. [CrossRef]
27. Liang, T.; Sun, L.; Li, H. MODIS aerosol optical depth retrieval based on random forest approach. *Remote Sens. Lett.* **2021**, *12*, 179–189. [CrossRef]
28. Su, T.; Laszlo, I.; Li, Z.; Wei, J.; Kalluri, S. Refining aerosol optical depth retrievals over land by constructing the relationship of spectral surface reflectances through deep learning: Application to Himawari-8. *Remote Sens. Environ.* **2020**, *251*, 112093. [CrossRef]
29. 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 Sens.* **2020**, *12*, 4125. [CrossRef]
30. Fan, Y.; Li, Z.; Sun, L.; Dubovik, O.; Wang, Z.; Cheng, F.; Shi, X.; Wei, J. AeroTrans: Hourly AOD retrieval over land from MSG-1/SEVIRI imagery integrating Transformer and transfer learning. *Remote Sens. Environ.* **2026**, *335*, 115268. [CrossRef]
31. Wei, J.; Wang, Z.; Li, Z.; Li, Z.; Pang, S.; Xi, X.; Cribb, M.; Sun, L. Global aerosol retrieval over land from Landsat imagery integrating Transformer and Google Earth Engine. *Remote Sens. Environ.* **2024**, *315*, 114404. [CrossRef]
32. Kang, Y.; Kim, M.; Kang, E.; Cho, D.; Im, J. Improved retrievals of aerosol optical depth and fine mode fraction from GOCI geostationary satellite data using machine learning over East Asia. *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 253–268. [CrossRef]
33. Fu, D.; Shi, H.; Gueymard, C.A.; Yang, D.; Zheng, Y.; Che, H.; Fan, X.; Han, X.; Gao, L.; Bian, J.; et al. A Deep-Learning and Transfer-Learning Hybrid Aerosol Retrieval Algorithm for FY4-AGRI: Development and Verification over Asia. *Engineering* **2024**, *38*, 164–174. [CrossRef]
34. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]
35. Barnes, W.L.; Pagano, T.S.; Salomonson, V.V. Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1088–1100. [CrossRef]
36. Liang, S.; Zhao, X.; Liu, S.; Yuan, W.; Cheng, X.; Xiao, Z.; Zhang, X.; Liu, Q.; Cheng, J.; Tang, H.; et al. A long-term Global LAnd Surface Satellite (GLASS) data-set for environmental studies. *Int. J. Digit. Earth* **2013**, *6*, 5–33. [CrossRef]
37. Liu, Q.; Wang, L.; Qu, Y.; Liu, N.; Liu, S.; Tang, H.; Liang, S. Preliminary evaluation of the long-term GLASS albedo product. *Int. J. Digit. Earth* **2013**, *6*, 69–95. [CrossRef]
38. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]

39. Didan, K. *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006*; NASA EOSDIS Land Processes Distributed Active Archive Center DAAC Data Set; USGS Earth Resources Observation and Science (EROS) Center: Sioux Falls, SD, USA, 2015. [CrossRef]
40. Gesch, D.B.; Verdin, K.L.; Greenlee, S.K. New land surface digital elevation model covers the Earth. *Eos Trans. Am. Geophys. Union* **1999**, *80*, 69–70. [CrossRef]
41. Danielson, J.J.; Gesch, D.B. *Global Multi-Resolution Terrain Elevation Data 2010 (GMTED2010)*; Open-File Report; U.S. Geological Survey: Reston, VA, USA, 2011. [CrossRef]
42. Ackerman, S. *MODIS Atmosphere L2 Cloud Mask Product*; NASA MODIS Adaptive Processing System; Goddard Space Flight Center: Greenbelt, MD, USA, 2015. [CrossRef]
43. Dubovik, O.; Holben, B.; Eck, T.F.; Smirnov, A.; Kaufman, Y.J.; King, M.D.; Tanré, D.; Slutsker, I. Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations. *J. Atmos. Sci.* **2002**, *59*, 590–608. [CrossRef]
44. Levy, R.C.; Remer, L.A.; Dubovik, O. Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land. *J. Geophys. Res. Atmos.* **2007**, *112*, D13210. [CrossRef]
45. Vermote, E.F.; Tanre, D.; Deuze, J.L.; Herman, M.; Morcette, J.-J. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 675–686. [CrossRef]
46. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [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.