Remote Sensing of Environment 335 (2026) 115268

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

AeroTrans: Hourly AOD retrieval over land from MSG-1/SEVIRI imagery integrating Transformer and transfer learning

Yulong Fan [a], Zhanqing Li [b,*], Lin Sun [a,*], Oleg Dubovik [c], Zhihui Wang [d], Fan Cheng [e], Xiaohang Shi [f], Jing Wei [g,*]

[a] College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao, China
[b] Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[c] Laboratoire d'Optique Atmosphérique, Université de Lille, CNRS, Lille, France
[d] University of Science and Technology of China, Hefei, China
[e] School of Environmental and Sustainable Engineering, Eastern Institute of Technology Ningbo, Ningbo, China
[f] State Key Laboratory of Climate System Prediction and Risk Management, School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing, China
[g] MEEKL-AERM, College of Environmental Sciences and Engineering, Institute of Tibetan Plateau, and Center for Environment and Health, Peking University, Beijing, China

ARTICLE INFO

Edited by Menghua Wang

Keywords:
MSG-1/SEVIRI
Transformer
Transfer learning
Dust storm
Wildfire
Urban haze

ABSTRACT

Geostationary Earth Orbit (GEO) satellites offer unique capabilities for capturing diurnal variations and providing valuable insights into aerosol cycles. However, publicly available hourly aerosol products with sufficient accuracy remain scarce across Europe, Africa, and West Asia, primarily due to the lack of shorter-wavelength (< 0.6 μm) channels on the Meteosat Second Generation (MSG) satellite series. Therefore, we developed a novel "AeroTrans" deep learning framework to retrieve hourly aerosol optical depth (AOD) at 550 nm over land in 2021 from MSG-1/SEVIRI imagery, which offers wider spatial coverage through Indian Ocean Data Coverage (IODC). This framework integrates an advanced time-sequence Transformer architecture with transfer learning, utilizing pre-training and fine-tuning techniques. The eXplainable Artificial Intelligence (XAI) analysis revealed that satellite observations across multiple wavelengths contribute 38% to the AOD retrieval, followed by viewing geometry (34%). In comparison with ground-based AOD measurements, our model achieves high accuracy, with an average ten-fold cross-validation (CV) R^2 of 0.88 and a root mean square error (RMSE) of 0.079. Additionally, our model maintains strong predictive performance in areas and periods lacking ground-based measurements, as evidenced by strong spatial- and temporal-based CV-R^2 values ranging from 0.71 to 0.86. The model performance is significantly improved when withholding each continent, with accuracy increasing markedly (R = 0.71–0.78) compared to models trained without transfer learning (R = 0.23–0.47). Using the generated reliable 3-km-resolution AOD datasets, we capture pronounced diurnal aerosol variations, characterized by a gradual increase after sunrise, a peak around 10:00 UTC, and a decline by late afternoon, with average magnitude changes of approximately 26% ± 15% relative to the daily mean level (0.22 ± 0.14) on an annual scale, especially during the Northern Hemisphere summer, reaching 30% ± 19%. More importantly, we successfully tracked the rapid dispersion of aerosols and their transport process throughout the day during highly polluted events, driven by both natural and anthropogenic emissions, including dust storms, wildfires, and urban haze. These findings emphasize the unique value of our study for advancing aerosol research over under-monitored regions, particularly focusing on diurnal variations during extreme events.

1. Introduction

Aerosols, consisting of fine solid particles and liquid droplets suspended in the atmosphere, can both directly scatter and absorb solar radiation and indirectly alter cloud formation processes and thus impact weather and climate (Ramanathan et al., 2001; Stier et al., 2007; Li

* Corresponding authors.
E-mail addresses: zhanqing@umd.edu (Z. Li), sunlin6@126.com (L. Sun), jingwei@pku.edu.cn (J. Wei).

https://doi.org/10.1016/j.rse.2026.115268
Received 5 August 2025; Received in revised form 12 January 2026; Accepted 22 January 2026
Available online 11 February 2026
0034-4257/© 2026 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

et al., 2016; Watson-Parris and Smith, 2022). Aerosol Optical Depth (AOD), which quantifies aerosol loading in the atmosphere and exhibits significant temporal variability, is a critical parameter for understanding such impacts (Dong et al., 2024; Ha et al., 2020; Yang et al., 2023). While polar-orbiting satellites, such as the widely used Terra and Aqua MODIS (Hsu et al., 2013; Levy et al., 2013; Lyapustin et al., 2018), the S-NPP VIIRS (Hsu et al., 2019; Sawyer et al., 2020), and the Sentinel series (Mei et al., 2018; Chen et al., 2024), have enabled global monitoring of AOD, they observe the Earth only once during daytime, limiting their ability to capture the pronounced diurnal variability of aerosols, especially during dynamic pollution events such as dust storms, biomass burning, and urban haze episodes (Arola et al., 2013; Wang et al., 2015; Zhang et al., 2012).

In contrast, geostationary-orbiting (GEO) satellites offer unprecedented opportunities for high-frequency monitoring that enhance temporal resolution from daily to hourly/minutely (Ceamanos et al., 2023; Fan et al., 2023; Kang et al., 2022; She et al., 2024). Recently, an increasing number of GEO satellites are available to track the diurnal aerosol variations and transport processes covering specific regions, including GOES-R series over the Americas and Pacific (Zhang et al., 2020), Himawari-8 over East Asia and the Western Pacific (Yoshida et al., 2018), and FengYun-4 over China (Si et al., 2024). Currently, operational aerosol products are available for these satellites with aerosol retrieval algorithms tailored for applications with specific sensors aboard geostationary satellites operated by the National Oceanic and Atmospheric Administration (NOAA), Japan Meteorological Agency (JMA), and China Meteorological Administration (CMA). These products are open-access and available to all users, making them particularly valuable and beneficial to a wide range of applications, including but not limited to air pollution monitoring and climate research, especially for estimating surface-level particulate matter concentrations (Wei et al., 2021; Xia et al., 2023; Zhang et al., 2022).

The Meteosat geostationary satellite series, including the Meteosat Second Generation (MSG) and the Meteosat Third Generation (MTG) missions, has been operated over Europe, Africa, and West Asia, collectively covering approximately 45% of the global land surface. A primary sensor aboard MSG is the Spinning Enhanced Visible and InfraRed Imager (SEVIRI), which is uniquely valuable for monitoring air quality across this vast region where air pollution is driven by diverse emission sources and meteorological processes. For example, many European countries, such as Poland, Italy, Bulgaria, and Spain, have suffered poor air quality primarily due to intense anthropogenic emissions and frequent wildfires (Chowdhury et al., 2024; de Hoogh et al., 2018; Dupuy et al., 2020; Guerreiro et al., 2014). In Africa, air pollution often results from a combination of persistent wildfires in the Congo Basin, the world's second-largest rainforest (Wimberly et al., 2024a), and substantial anthropogenic emissions from densely populated countries such as South Africa (Wei et al., 2023). Besides, northern Africa, especially the Saharan Desert, is especially vulnerable to frequent dust storms that impact surrounding regions and exacerbate air pollution in nearby metropolitan areas such as Agadir, Nouakchott, and Timbuktu (Engelstaedter et al., 2006). Therefore, enhanced monitoring of aerosol loadings and their diurnal variations across this region is crucial for advancing air quality and environmental research and management.

Many attempts have been made to retrieve high-temporal-resolution AOD from MSG/SEVIRI satellite observations. Popp et al. (2007) introduced a background analysis approach that uses the minimum apparent reflectance at 0.64 μm over a 31-day moving window to estimate surface reflectance and retrieve hourly AOD at a 25 km resolution for central Europe. Bernard et al. (2011) retrieved hourly AOD at a 3 km resolution over Europe by assuming surface reflectance. Two-channel methods were also developed and applied to retrieve AOD at 5 km over Central-eastern Europe (Zawadzka and Markowicz, 2014), which were later refined to 3 km over the entire region (Xie et al., 2020). Luffarelli and Govaerts (2019) employed the Combined Inversion of Surface and Aerosol (CISAR) algorithm, which continuously varies all state variables in the solution space, for the joint retrieval of surface reflectance and aerosol properties from instantaneous SEVIRI imagery. Ceamanos et al. (2023) further refined the AERUS-GEO algorithm, originally developed by Carrer et al. (2014), by enhancing the temporal resolution of AOD retrievals from daily to instantaneous at 3 km resolution, across Europe, Africa, and the Middle East. However, these studies may suffer from large uncertainties in AOD retrievals, especially over bright surfaces (e.g., urban and deserts), primarily due to the lack of short-wavelength channels (e.g., blue or deep-blue), which are particularly important for improving aerosol-surface contrast in such high-reflectance regions (Hsu et al., 2013; Lyapustin et al., 2018).

To address these limitations, we developed a hybrid deep-learning model, AeroTrans, which significantly advances hourly AOD (550 nm) retrievals over land from MSG-1/SEVIRI imagery, based on recent studies using deep learning to retrieve AOD with GEO satellites (Fan et al., 2023; She et al., 2024; Tang et al., 2025; Yeom et al., 2022), and our study using the Landsat high-resolution Imagery of long-term records (Wei et al., 2024). It is applied to all MSG-1 data from 2021 over the Indian Ocean Data Coverage (IODC) region to retrieve hourly AOD at 550 nm. This model mitigates the limitations of sparse ground monitoring stations by employing an advanced time-sequence Transformer architecture that fully leverages the multi-wavelength, high-frequency MSG-1/SEVIRI observations and incorporates transfer learning to improve performance in regions and periods with limited measurements, using prior knowledge learned from simulated model outputs. The model is first pre-trained on physically consistent AOD reanalysis data to embed prior information about aerosol distribution patterns and then fine-tuned using ground-based AOD measurements to capture region- and sensor-specific characteristics, thereby enhancing generalization and predictive capability across different locations and time periods. Designed specifically for land, the model addresses the greater challenges of aerosol retrieval over complex terrestrial surfaces and atmospheres. It was trained exclusively with land-based measurements and is not intended for ocean conditions due to limited monitoring data. Independent spatiotemporal validation shows that AeroTrans-MSG achieves superior accuracy in hourly AOD retrievals and provides critical insights into diurnal AOD evolution and the transport dynamics of major pollution episodes. Section 2 introduces the data used; Section 3 describes the developed deep-learning framework; Section 4 presents and discusses the results; and Section 5 summarizes the study.

2. Materials

2.1. MSG-1/SEVIRI observations

MSG-1 (Meteosat-8), the first MSG satellite, began observing the Earth in 2002 and underwent three orbit changes, with the final adjustment made in 2020 (Ceamanos et al., 2021). The MSG-1 satellite is selected in our study because it is located at 41.5° E and has a wider spatial coverage, i.e., named as Indian Ocean Data Coverage (IODC), covering Europe, Africa, and especially West Asia (which is very important for studying its air quality), compared to other MSG satellites. The SEVIRI sensor carried by MSG-1 observes approximately a quarter of the Earth's surface (67.5°N–67.5°S, 22°W–113°E) every 15 min, with a spatial resolution of ~3 km at the nadir. However, it provides observations only across wavelengths from red (0.64 μm) to longwave infrared (13.4 μm) (Table S1), lacking shorter-wavelength channels (e.g., blue). In this study, we collected all available hourly MSG-1/SEVIRI images during daylight hours (~16 h per day) in 2021, totaling about 5840 images, which were used to develop the aerosol retrieval algorithm due to the substantial data volume. For independent validation, data from 2020 and 2022 were also included. Unsuitable pixels were filtered out using the cloud-snow-ice mask products from the Nowcasting and Very Short Range Forecasting Satellite Application Facility (NWCSAF, version 2018; Kniffka et al., 2014), as potentially cloud-contaminated observations could introduce uncertainty and lead to AOD

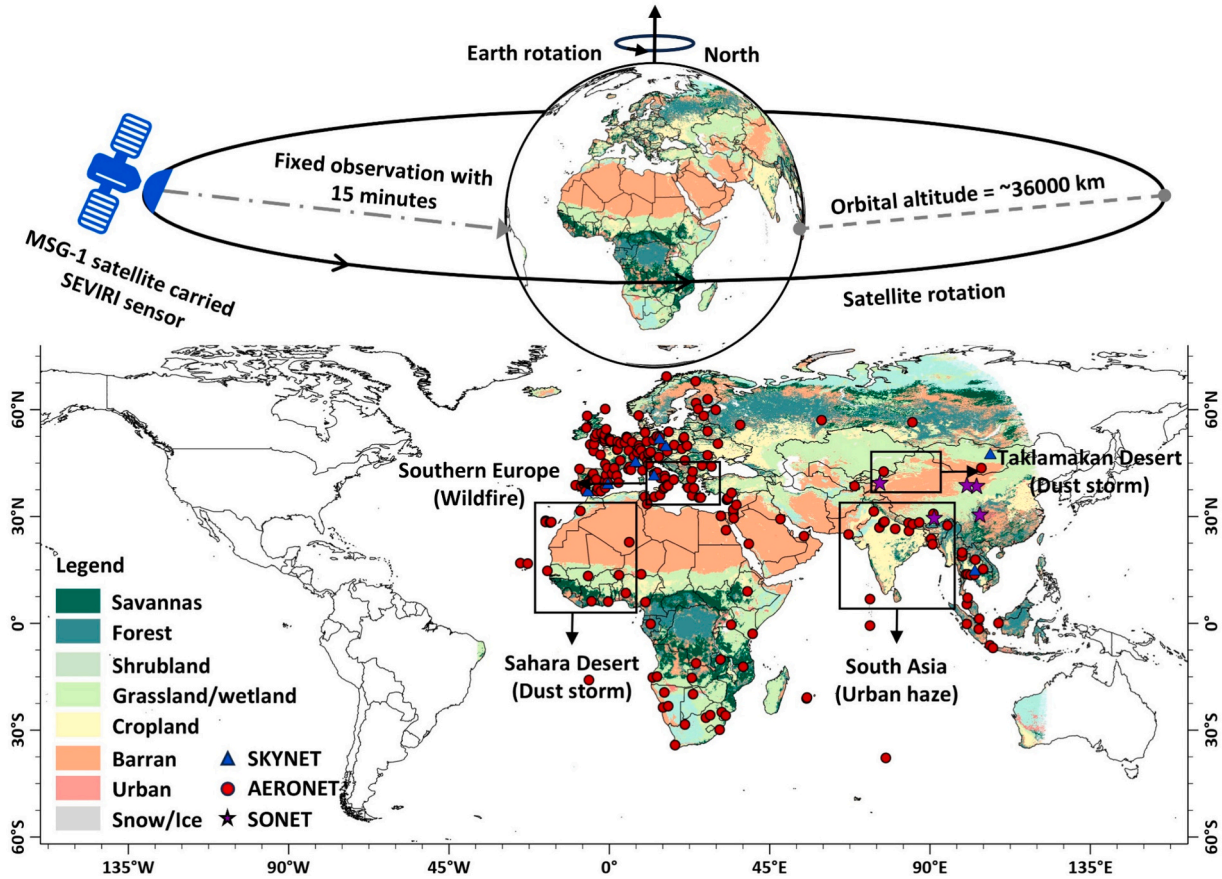

Fig. 1. Study region observed by the geostationary MSG-1/SEVIRI satellite, covering Europe, Africa, and West Asia, along with the locations of ground-based aerosol monitoring networks, including AERONET (red dots), SONET (purple stars), and SKYNET (blue triangles). The rectangles show selected regions experiencing extreme environmental events, including dust storms from the Sahara and Taklimakan deserts, urban haze in South Asia, and wildfires in southern Europe and central Africa. The base map is the MODIS land-use cover (MCD12Q1) product (Friedl et al., 2010). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

overestimation.

2.2. Ground-based measurements

Three ground-based networks, AERONET, Sun-sky radiometer Observation NETwork (SONET), and SKY Radiometer NETwork (SKYNET), are utilized for the development and validation of our algorithm. AERONET and SONET AOD measurements serve as ground-truth AOD values for constructing the training samples. AERONET is a well-known global aerosol observation network (Holben et al., 1998) that uses solar photometers to measure AOD for a long period of time, for over 20 years at many stations of high frequency (typically 15 min) at multiple channels. In this study, we utilized quality-assured level 2.0 AOD data with an average uncertainty of approximately ±2% (Giles et al., 2019) from 205 sites across the study domain, including 109, 44, and 52 sites located in Europe, Africa, and Asia (Fig. 1). SONET, a regional network similar to AERONET, comprises long-term observation sites across China (Fig. 1), which uses solar photometers to measure AOD, with an average maximum uncertainty of ±0.5% across different spectral bands (Li et al., 2018). Similarly, we collected the SONET AOD data from 5 sites within our study domain from central to western China to supplement the training samples in Asia, especially western China, where AERONET coverage is sparse.

Furthermore, SKYNET AOD measurements are employed as independent validation samples to evaluate model performance. SKYNET utilizes standardized instruments, like sun-sky radiometers, to monitor AOD at 105 sites across various regions, including Europe and Asia (Nakajima et al., 2020), following specific analysis protocols to produce standard aerosol products (Takamura et al., 2009; Campanelli et al., 2012). Here, we selected SKYNET AOD measurements from six sites in Europe and two sites in Asia (Fig. 1), which have a minimal average difference of approximately ±0.3% compared to AERONET measurements (Nakajima et al., 2020). However, these aerosol networks do not provide AOD measurements at 550 nm, so we converted the available 440 nm measurements to 550 nm using the Angström Exponent method (Wei et al., 2019a, 2019b).

2.3. Atmospheric reanalysis data

MERRA-2 hourly aerosol product in 2021 is collected to provide AOD simulations over land for transfer learning as pre-training in our study (GMAO, 2015). MERRA-2 is the NASA Global Modeling and Assimilation Office (GMAO)'s current advanced global atmospheric reanalysis for the satellite era, produced with the Goddard Earth Observing System (GEOS-5.12.4) data-assimilation system (Randles et al., 2017). Aerosols are simulated via the Goddard Chemistry Aerosol Radiation and Transport (GOCART) aerosol module (Chin et al., 2000) and assimilate observations from MODIS, MISR, and other satellite and in situ measurements (Gelaro et al., 2017). It shows good global accuracy, with typical biases of ~0.2%–6% and RMSE of ~0.108–0.159 compared with AERONET, though uncertainties may be larger in regions with fewer observations (Su et al., 2023). In particular, MERRA-2 can provide initial background aerosol information by providing more samples with richer scenes than ground-based measurements, thereby enhancing the

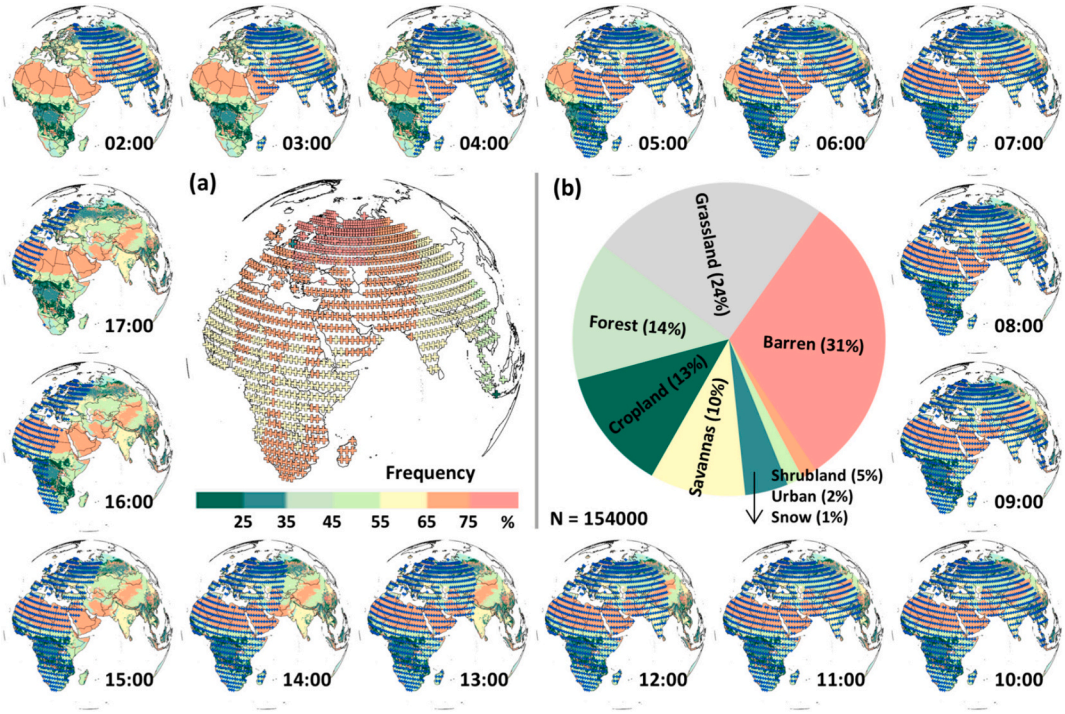

Fig. 2. Spatial distribution of pre-training samples with (a) different sampling frequency and (b) percentages for different land types during 02:00–17:00 UTC in 2021. The land cover types are sourced from the MCD12Q1 product.

generalization ability of deep learning models.

2.4. Other auxiliary data

During the deep learning modeling, we also incorporated multiple auxiliary factors that influence aerosol retrieval, accounting for varying atmospheric conditions and land surface characteristics. First, two key gases, hourly total ozone (TCO_3, unit: Dobson) and total precipitable water vapor (TCWV, unit: kg m^{-2}), were included, both collected from the MERRA-2 reanalysis. The Digital Elevation Model (DEM) from the ETOPO Global Relief Model 2022 at 0.015° × 0.015° resolution was employed to account for altitude-related effects (She et al., 2024). Additionally, the normalized differential vegetation index (NDVI) in the shortwave infrared ($NDVI_{swir}$) is not sensitive to aerosol changes and is employed to capture surface information, which is calculated using TOA reflectance at 0.81 μm and 1.64 μm, for the absence of 2.3 μm on MSG-1/SEVIRI (Vermote et al., 2016).

3. Methods

3.1. MSG aerosol retrieval framework

This study proposes a novel hybrid framework for aerosol retrieval, combining a Transformer-based deep learning model for MSG-1/SEVIRI imagery (AeroTrans-MSG), driven by the transfer-learning technology to provide prior knowledge, supporting and improving the model's adaptability in hourly AOD retrieval over land. In our framework, the Transformer architecture and transfer learning serve complementary but distinct roles, and their combination is essential for robust AOD retrieval from geostationary satellite observations. Specifically, the Transformer is used to model the nonlinear relationship between TOA reflectance and AOD. Its self-attention mechanism is well-suited for high-frequency geostationary observations, allowing the model to capture long-range dependencies and complex temporal-spectral patterns. Transfer learning addresses the limited and uneven distribution of ground-based AOD observations. The model is first pretrained using AOD reanalysis to learn a generalized reflectance-AOD relationship, and then fine-tuned with ground-based measurements to correct biases and improve realism. Overall, the Transformer provides strong representation and sequence-learning capability, while transfer learning improves model generalization under data-scarce conditions.

3.1.1. Deep learning: Transformer

Deep learning relies on multi-layer artificial neural networks that learn hierarchical feature representations from data. Its core components include neurons, stacked layers, nonlinear activation functions, and training via backpropagation (LeCun et al., 2015). Transformer, a powerful time-series deep learning model, excels in processing temporal sequential data, demonstrating significant advantages in capturing long-range dependencies and feature patterns through its unique self-attention mechanism (Vaswani et al., 2017). Additionally, it incorporates advanced residual connections and layer normalization mechanisms, which enhance both training efficiency and inference accuracy when processing large-scale structured data (She et al., 2024). This is particularly valuable for datasets with high-frequency imagery, long time periods, and large spatial distances, as demonstrated in our previous study on AOD retrieval from Landsat imagery, which provides 16-day data records spanning several decades at a very high 30 m resolution (Wei et al., 2024). More importantly, this capability holds even greater potential for geostationary satellites like MSG/SEVIRI, which provide much higher-frequency imagery, observing the Earth every 15 min and delivering around 96 images per day. This robust time series of images allows for continuous observations throughout the day under varying weather conditions, effectively capturing dynamic AOD changes. This makes it the ideal model for our study and is expected to enhance AOD retrieval accuracy for GEO satellites.

3.1.2. Transfer learning

Machine learning has proven to be highly effective in capturing nonlinear relationships between target values and predictors in air pollutant modeling. While models perform well at observation sites, their estimation accuracy in unobserved regions declines as the distance

from these locations increases (Fu et al., 2024; Tang et al., 2025; Wei et al., 2023), with the degree of degradation strongly influenced by station density. This is particularly evident for AOD, given the sparser and more uneven distribution of ground-based monitoring stations (Fig. 2), which can significantly limit the model's performance, especially in remote areas. To minimize this challenge, we integrate transfer learning into our aerosol retrieval framework, facilitating the transfer of acquired knowledge (e.g., learned features and model weights) from source domains to target tasks, thereby improving model generalization in unobserved areas (Fu et al., 2024; Tan et al., 2018; Tang et al., 2025). Transfer learning is a machine-learning framework that improves performance on a target task by leveraging knowledge gained from a related source task or domain. Unlike conventional methods that assume identical data distributions across training and test sets, transfer learning allows domains, tasks, or distributions to differ (Pan and Yang, 2010). Common strategies include using a pretrained model as a fixed feature extractor or partially fine-tuning its layers. This prior knowledge is highly valuable, as it enables the model to learn aerosol dynamics across complex atmospheric and surface conditions by incorporating data such as AOD simulations from atmospheric reanalysis, thereby providing essential background constraints on air pollution in regions lacking ground-based observations. More importantly, transfer learning can effectively extract and transfer this prior knowledge through its two-stage operation, pre-training and fine-tuning, enhancing the model's spatiotemporally predictive capabilities. Fine-tuning is the process of continuing the training of a pretrained network on a target dataset, with a lowered learning rate and selective layer updates, to adapt general representations to specific task characteristics (Zheng et al., 2025). In this study, we pre-train the Transformer on MERRA-2 AOD simulations and subsequently fine-tune it using in situ AOD measurements, thereby enhancing its generalization and enabling continuous, accurate AOD retrieval.

the gas-phase influence in the short-wavelength channels relevant to aerosol retrieval, while contributions from other gases are much weaker and can be negligible (Hsu et al., 2013; Wei et al., 2019a, 2019b). The DEM is included to account for topographic variation impacts on aerosol retrieval, while the $NDVI_{swir}$ is used to represent land surface characteristics and changes (Wei et al., 2024).

3.1.3.2. Pre-training. In our approach, the Transformer is first pre-trained in the source domain using hourly MERRA-2 AOD at 550 nm as the supervisory labels. For each MSG-1/SEVIRI observation time (02:00–17:00 UTC), we construct an input vector x = [S, Θ, G, L], including multi-channel reflectance S, solar-sensor geometry Θ, atmospheric gases G (e.g., water vapor, and ozone), and land-surface descriptors L (DEM, and $NDVI_{swir}$) (Eq. 1). All features are min-max normalized, and samples with invalid observations (e.g., cloud contamination or extreme viewing geometry) are excluded. Extreme view geometry refers to observations in which either the satellite or solar zenith angle exceeds 80°. Such geometries increase atmospheric path length, scattering, and absorption, and may introduce projection distortions and topographic shading, reducing the reliability of radiance measurements and cloud masks (Laszlo and Liu, 2016). Machine-learning models cannot fully account for these extreme-view cases because training samples under such conditions are scarce, limiting model performance and generalization. Therefore, excluding these observations improves the quality of training and validation datasets, supporting better overall model performance. Pre-training is optimized with a standard regression loss (MSE) and scalable training strategies, including distributed optimization, learning-rate scheduling, gradient accumulation, and mixed-precision. This large-scale pre-training compensates for the scarcity of ground observations and provides physically informed representations of AOD-reflectance relationships.

$$AOD_{MERRA-2} = Transformer_{pre-training}([TOA_{0.6-1.6}, BT_{3.9}, Angles, Gases, DEM, NDVI_{swir}]) \quad (1)$$

3.1.3. AeroTrans-MSG framework

3.1.3.1. Feature selection. Aerosols exhibit different effects on top-of-the-atmosphere (TOA) reflectance at various wavelengths, especially at shorter wavelengths (Hsu et al., 2013). However, the MSG-1/SEVIR lacks the widely used deep blue (typically around 412 nm) or blue band, which is highly sensitive to aerosol variations and widely employed in aerosol remote sensing, such as in MODIS (Hsu et al., 2013). To address this, we employ multiple channels, including the available shortest-wavelength channel (red: 0.64 μm) as the primary predictor, along with near-infrared (NIR: 0.81 μm) and shortwave infrared (SWIR: 1.64 μm) channels. These channels are converted to TOA reflectance ($TOA_{0.6-1.6}$) via radiometric calibration provided by the SATMOS (Service for Archiving and Meteorological Processing of Satellite Observations) data and services center. Additionally, the mid-infrared channel (MIR: 3.92 μm) is included to enhance aerosol detection, particularly for coarse-mode particles like dust (Zhang et al., 2006), and is converted to brightness temperature ($BT_{3.9}$) using SATMOS calibration data. This multi-channel information not only facilitates aerosol detection but also enables the differentiation of various aerosol types over land surfaces (Wei et al., 2018). Observation geometry (*Angles*) plays a crucial role in aerosol retrieval, with five main angles considered: solar zenith angle (SZ), solar azimuth angle (SA), viewing zenith angle (VZ), viewing azimuth angle (VA), and scattering angle (SCA). Furthermore, the effects of two key gases (*Gases*), water vapor and ozone, on TOA reflectance are corrected using TCWV and TCO_3 data. We focused on water vapor and ozone because their absorptions dominate

Using 5850 MSG-1 images, we generated ~1.5 billion training examples that span diverse surface types (i.e., barren: 31%, grassland: 24%, forest: 14%, cropland: 13%; Fig. 2), enabling the model to learn the radiative response to varying atmospheric and surface conditions. The pre-training samples were extracted by matching MERRA-2 AOD with MSG-1/SEVIRI images on an hourly, 2° × 2° grid. The spatial distribution of pre-training samples varies diurnally because they are constructed from hourly collocation of MSG-1/SERIVI observations with MERRA-2 AOD during 02:00–17:00 UTC. This pattern is largely driven by diurnal changes in the solar zenith angle, as only valid daytime observations are retained. Regions at high latitudes and near the edge of the MSG-1 field of view have fewer training samples due to limited availability of valid satellite observations (e.g., low sun angles or geometric constraints). The pre-training dataset spans a broad range of surface types, which is important because machine-learning models often inherit surface-type biases from the limited land-cover diversity of ground monitoring stations, which can reduce performance in regions with different surface characteristics. Our large-scale pre-training collects samples covering nearly all land-cover types within the MSG-1 region, improving the model's robustness and generalization in areas without ground monitoring stations.

3.1.3.3. Fine-tuning. In this stage, the in situ AOD measurements at 550 nm ($AOD_{in\ situ}$) serve as target values, with the same input features fed into the Transformer model to fine-tune the last two hidden layers of the pre-training model while keeping the other model parameters

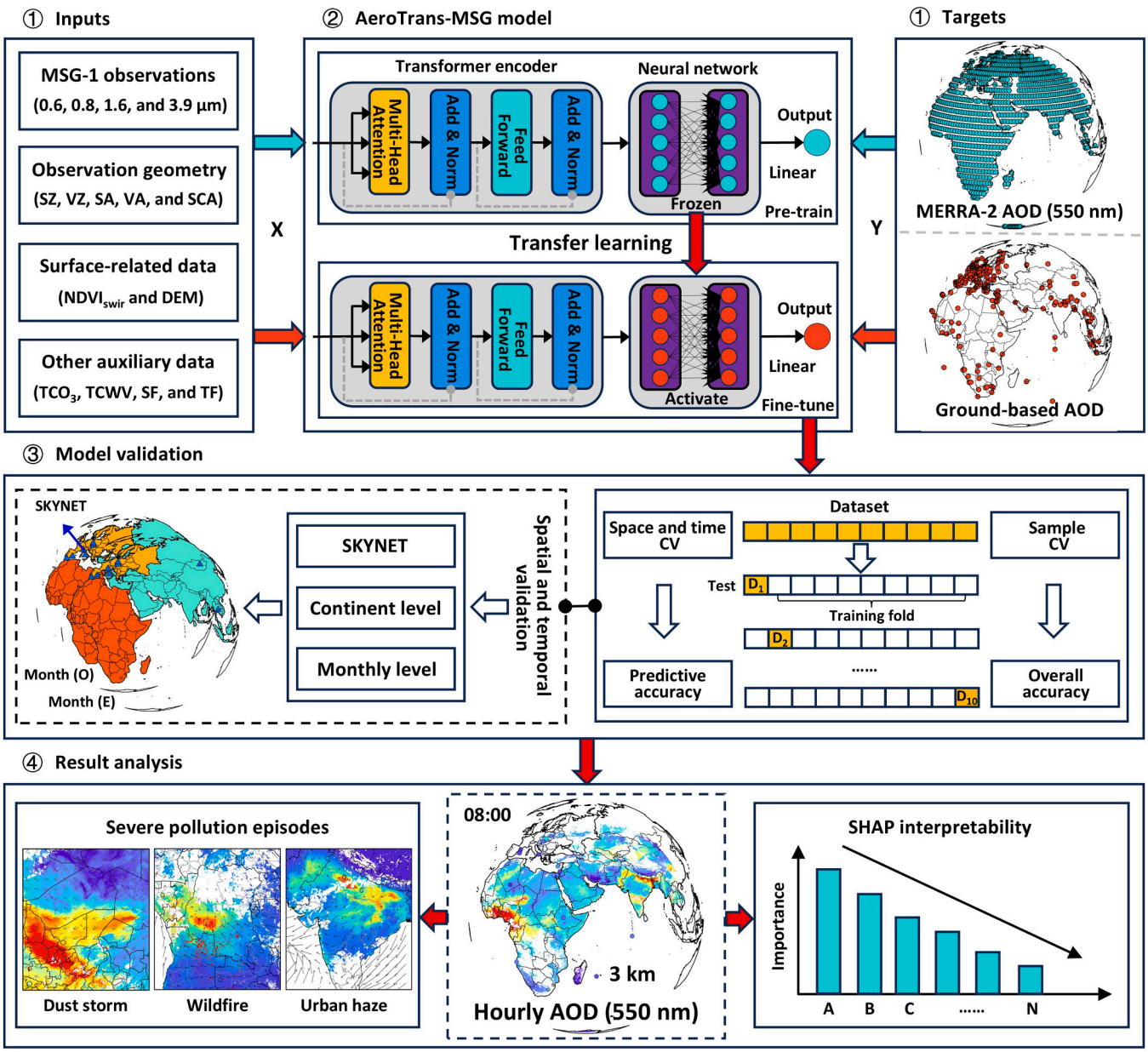

Fig. 3. Flowchart of the developed AeroTrans-MSG model for hourly AOD (550 nm) retrieval over land from MSG-1/SEVIRI imagery, combining the deep-learning Transformer model and transfer learning.

unchanged. In comparison, the in situ AOD measurements constitute a much smaller sample size (only ~0.15 billion samples) with dense coverage in urban areas (31%) but sparse coverage in barren regions (9%), which dominate much of North Africa and West Asia, underscoring the importance of pre-training (11 times more data samples). Finally, the nonlinear relationships between the actual AOD values and the input predictors (x_{all}) same with in the pre-training stage are refined and constructed by the Transformer based on these data samples:

$$AOD_{in\ situ} = Transformer_{fine-tuning}\{Model_{pre-training}([x_{all}])\} \tag{2}$$

Here, we use an hourly time series of 13 predictors to generate AOD retrievals across a 17-h daytime period. To further enhance spatiotemporal modeling, we replace the conventional positional encoding with a multi-dimensional spatiotemporal encoding scheme (SF and TF) developed in our previous study (Wei et al., 2024). In this work, we introduce minor improvements by incorporating three additional helix-shaped, hour-based trigonometric sequence vectors to more effectively capture diurnal AOD variations:

$$T_{hour} = \left[\frac{HOD}{N_h}, \cos\left(2\pi\frac{HOD}{N_h}\right), \sin\left(2\pi\frac{HOD}{N_h}\right)\right] \tag{3}$$

where HOD is the hour of the day, and N_h is the total hours of one day used for aerosol retrieval in 2021, equal to 17 in this study. The trigonometric terms are used solely to generate time-dependent features as inputs for the model. They do not constrain AOD to follow any specific functional form, but encode diurnal timing and solar-geometry information (e.g., hour-of-day effects) to help model capture time-related variations. Transformer remains free to learn arbitrary mappings from all inputs, including satellite signals, atmospheric state, land-cover, and these time features, to AOD. These trigonometric bases simply provide continuous and differentiable features that facilitate the learning of diurnal patterns.

Subsequently, our model employs the Transformer encoder architecture of Vaswani et al. (2017), enabling efficient extraction of latent

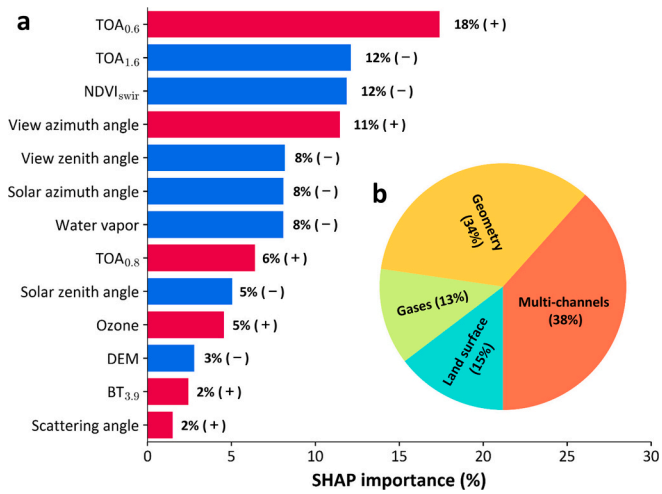

Fig. 4. SHAP-based XAI analysis of the AeroTrans-MSG framework for hourly AOD retrieval, showing (a) SHAP feature importance (%), with "(+)" and "(−)" after the black numbers indicating positive or negative impacts, and (b) the total contributions aggregated by four predefined feature categories.

aerosol features from the multi-dimensional input space while mitigating overfitting (Wei et al., 2024). The encoder comprises stacked layers, each containing a masked multi-head self-attention mechanism (to prevent access to future information), an encoder–decoder cross-attention module (to incorporate relevant input representations), and a position-wise feed-forward network, with residual connections and layer normalization applied to each sub-layer. The key hyperparameters, including the number of neurons and layers, number of attention heads, and feed-forward hidden size, were determined through systematic trial-and-error to balance model capacity and computational cost (Fan et al., 2023; Yeom et al., 2022). Rectified Linear Unit (ReLU) activation functions are applied between hidden layers to enhance learning efficiency and improve sparse feature extraction. Mean Squared Error (MSE) is used as the loss function to quantify the discrepancy between predicted and target AOD values, ensuring effective optimization and stable gradient propagation during backpropagation. Table S2 provides the hyperparameter settings for our AeroTrans-MSG model. In this study, we selected an hourly temporal resolution for AOD retrievals, as it aligns with the MERRA-2 hourly AOD simulations used for pre-training. Moreover, hourly AOD retrievals are sufficient for studying the diurnal variations of AOD within the study region.

3.2. Validation approaches

We implement two distinct validation frameworks: 10-fold cross-validation (CV) and independent spatiotemporal validation. The 10-fold CV method incorporates three specific validation strategies to comprehensively assess model performance: sample-based, temporal-based, and spatial-based CV (Wei et al., 2023). These strategies involve randomly partitioning (1) all data samples, (2) temporal units at increasing time scales (hourly, daily, and monthly), and (3) spatial units at increasing space intervals (grids of $1^\circ \times 1^\circ$, $5^\circ \times 5^\circ$, and $10^\circ \times 10^\circ$) into ten subsets. In each iteration, nine subsets are used for model training, while the remaining subset serves as the validation set. This process is repeated ten times, ensuring each subset is used exactly once for validation, thereby providing a robust evaluation of the model's overall estimation accuracy, as well as its temporal and spatial predictive accuracy.

The independent spatiotemporal validation consists of three complementary components: monthly, continental, and SKYNET-based validation. For monthly validation, we employ a temporal splitting strategy where odd-numbered months are used for training and even-numbered months for testing alternately, thereby evaluating the model's temporal extrapolation capability. Continental validation assesses spatial generalization by designating each specific continent (Europe, Africa, or Asia) as an independent test region. The model is then trained using one region (e.g., Asia) and validated on the others. We performed continent-holdout validation to assess the model's spatial generalizability. Training on some continents and testing on another exposes the model to different aerosol regimes, surface types, and viewing conditions, revealing whether it has learned broadly applicable relationships or only region-specific patterns. Although not fully representative of the operational domain, this approach provides valuable diagnostic information. Additionally, we leverage ground-based measurements from the SKYNET network as a fully independent validation dataset, further assessing the model's ability to generalize across different networks and spatiotemporal conditions.

To assess model performance, we employ the correlation coefficient (R), coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean bias (MB). Furthermore, we incorporate the expected error (EE) of AOD retrievals from the MODIS DB algorithm over land, defined as $\pm\,(0.05 + 20\% \times \mathrm{AOD}_\tau)$ (Hsu et al., 2013). We also involve the stringent accuracy requirements for AOD retrievals set by the Global Climate Observation System (GCOS), defined as $\pm\,\max(0.03,\ 10\% \times \mathrm{AOD}_\tau)$ (GCOS, 2010). Notably, we applied the selection criteria by constraining AOD retrievals to the physically plausible range of 0 to 5, thereby removing outliers before validation and subsequent analysis. Fig. 3 illustrates the flowchart of our study, showcasing the proposed AeroTrans-MSG model for hourly AOD retrieval from MSG-1/SEVIRI imagery.

4. Results and discussion

4.1. Model interpretability

We applied the SHapley Additive exPlanations (SHAP) approach (Lundberg and Lee, 2017), which leverages gradient information within deep learning architectures, as part of the eXplainable Artificial Intelligence (XAI) technology to interpret our AeroTrans-MSG model. The results revealed that TOA observations from the MSG-1/SEVIRI satellite at 0.64 μm show the largest positive impact (i.e., the feature increases the model's output relative to the baseline) on aerosol retrieval, contributing approximately 18% to the model output (Fig. 4). In contrast, the TOA observation at 1.64 μm has a negative influence and

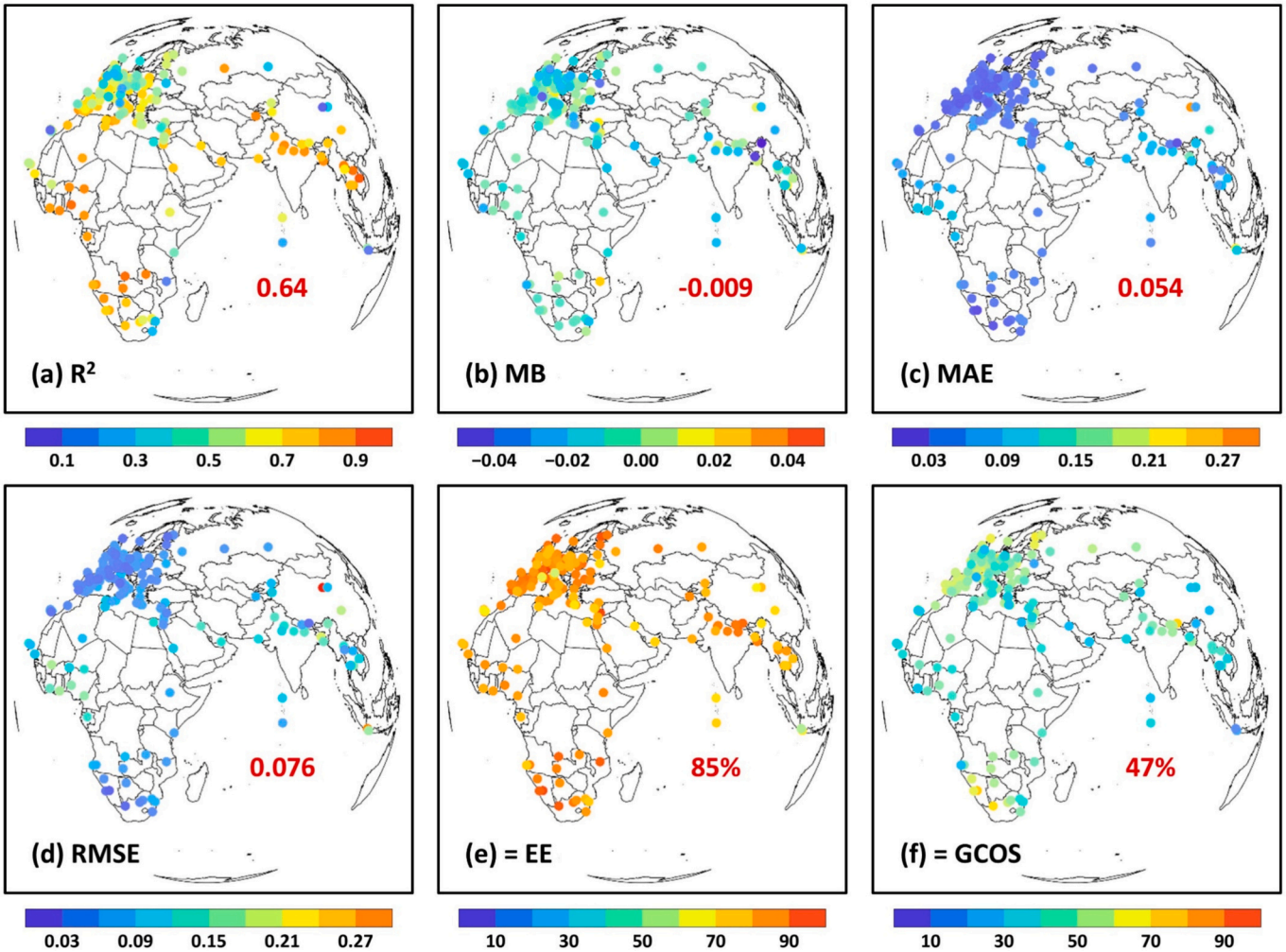

Fig. 5. Spatial pattern of overall accuracy and uncertainty of hourly AOD retrievals from MSG-1/SEVIRI imagery using the AeroTrans-MSG model, assessed through the sample-based ten-fold cross-validation approach in terms of (a) coefficient of determination (R^2), (b) mean bias (MB), (c) mean absolute error (MAE), (d) root mean square error (RMSE), percentage of retrievals within the (e) expected error (EE), and (f) Global Climate Observing System (GCOS). Red numbers indicate mean values (number of samples, $N = 168{,}333$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contributes about 12%. The negative influence of the 1.64 μm channel occurs in conditions where surface reflection or cloud contamination dominates, such as low-AOD scenes over bright surfaces or near clouds, reducing or reversing the channel's sensitivity to aerosols. This effect can be further modulated by viewing and solar geometry. Overall, TOA reflectance and BT across four channels together account for about 38% of the total feature importance. Angle information is important for sun photometers and radiative-transfer-based satellite retrievals, and it also plays a significant role in our deep-learning approach, collectively accounting for ~34% of feature importance. Both solar and viewing angles remain physically important, as they modulate the observed radiative signal and drive aerosol information content, contributing approximately 13% and 19%, respectively. In addition, $NDVI_{swir}$ has a strong negative effect on the model, contributing approximately 12%. Water vapor positively while ozone negatively affects AOD retrieval, together contributing ~13%, likely due to increased aerosol hygroscopicity under high humidity conditions and ozone's absorption effects of solar radiation at shorter wavelengths (Cheng et al., 2025; Li et al., 2014). The SHAP results confirm the physical basis of AeroTrans-MSG and highlight key auxiliary inputs that could inform future radiative-transfer-based retrieval algorithms.

4.2. Model validation

4.2.1. Overall accuracy

We first evaluated the model performance at the individual site scale across the domain using the sample-based 10-CV approach (Fig. 5). Our model effectively estimated hourly AODs across most sites, showing good agreement with ground-based measurements at ~77% of the sites, which had moderate to high sample-based cross-validation coefficients of determination (CV-$R^2 > 0.5$; Fig. 5a). Mean biases were generally low, within ±2% for about 77% of the sites (Fig. 5b). Notably, higher accuracy was observed in densely populated regions with elevated pollution, including central and southern Africa and South Asia (CV-$R^2 > 0.8$), which are of greater public concern. The lower correlation over Europe is primarily attributed to generally lower AOD loadings (average = 0.13), reflecting cleaner conditions compared with other continents such as Asia (average = 0.35) and Africa (average = 0.19). Retrieval uncertainties were consistently low at most sites, with ~73% and 77% of sites exhibiting MAE and RMSE values below 0.06 and 0.09, respectively (Fig. 5c-d). Exceptions occurred at a few individual sites along the eastern boundary of the domain, primarily due to limited satellite observations (e.g., large viewing angles). Overall, more than 95% of the sites demonstrated substantial retrieval accuracy, with over 70% of the retrievals falling into the EE envelope (Fig. 5e). In addition, approximately 73% of the sites had at least 40% of retrievals meeting the GCOS

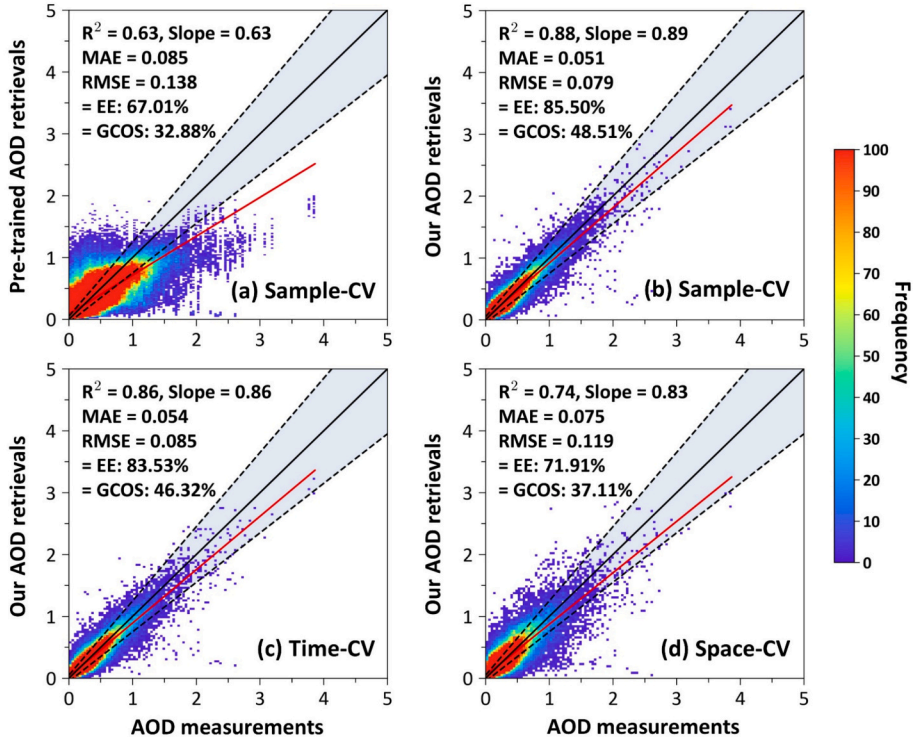

Fig. 6. Density scatter plots of our (a) pre-trained (N = 1,492,079) and (b-d) fine-tuned (N = 168,333) hourly AOD retrievals against ground-based AOD measurements, evaluated using (a-b) sample-based (9 folds of all samples for training, 1 fold for validation), (c) hour-based (time-CV; 9 folds of hours for training, 1 fold for validation), and (d) site-based (space-CV; 9 folds of stations for training, 1 fold for validation) ten-fold cross-validation methods, respectively. In each panel, the black dotted line represents the expected error range, the black solid line indicates the 1:1 line, and the red solid line represents the linear regression line. The numbers in the top-left corner of the subfigure represent the coefficient of determination (R^2), slope of the regression line, mean absolute error (MAE), root mean square error (RMSE), percentage of retrievals within the expected error (EE), and Global Climate Observing System (GCOS). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

requirements (Fig. 5f).

At the domain level, our model, initially pre-trained on MERRA-2 AODs, showed limited capability in retrieving AOD values from in situ measurements, yielding an average sample-based CV-R^2 of 0.63 (slope = 0.63) and RMSE (MAE) of 0.138 (0.085). Only 67% of its retrievals fell within the EE envelope, and just 33% met the GCOS threshold (Fig. 6a). However, after fine-tuning, the model performance improved significantly, achieving a sample-based CV-R^2 of 0.88 with a strong slope of 0.89, and reduced MAE (0.051) and RMSE (0.079) values. Approximately 86% and 49% of the AOD retrievals conform to the EE and GCOS criteria, respectively (Fig. 6b). In comparison, the directly trained model (without pre-training and fine-tuning) showed poorer performance (e.g., CV-R^2 = 0.85 and RMSE of 0.087), along with a smaller proportion of retrievals within EE (83%) and GCOS (47%) ranges (Fig. S1a). Notably, our model demonstrated enhanced training efficiency, achieving convergence in approximately 100 epochs (about twice as fast as the direct model) and maintaining lower loss values (6‰ vs. 8‰) throughout the training process (Fig. S1b). These results highlight the advantages of transfer learning in accelerating training and improving convergence stability.

4.2.2. Spatiotemporal prediction accuracy

Subsequently, we evaluated the model's predictive performance using a range of spatiotemporally independent validation strategies. Temporally, our model demonstrated strong predictive capability during hours without available measurements, achieving an hour-based CV-R^2 of 0.86 and an RMSE of 0.085; additionally, over 83% and 46% of retrievals met the EE and GCOS criteria, respectively (Fig. 6c), which were slightly inferior to those observed in sample-based cross-validation (Fig. 6b). More importantly, as the time interval used for cross-validation increased from hourly to daily to monthly scales, our model's accuracy showed only a slight decline (CV-R^2 from 0.83 to 0.77) while consistently outperforming the direct model (Fig. S2). Furthermore, by withholding data from all odd and even months separately to minimize monthly correlations and seasonal cycles between training and validation sets, our model continued to perform robustly in retrieving AODs during the remaining months, showing considerable accuracy (e.g., R = 0.87–0.88, RMSE = 0.11–0.12) and improvements over the direct models (Fig. S3).

The model's improvements were more pronounced at the spatial level: it demonstrated strong ability in predicting AODs in unmonitored areas, with a station-based CV-R^2 of 0.74, RMSE of 0.119, and approximately 72% and 37% of results meeting the EE and GCOS criteria, respectively (Fig. 6d). In addition, our model's prediction accuracy (CV-R^2) slightly decreased from 0.72 to 0.71 as the grid size increased from $1^\circ \times 1^\circ$ to $10^\circ \times 10^\circ$, showing significant improvements compared to the direct model, which decreased from 0.67 to 0.56 (Fig. S4). More importantly, by withholding data from three continents separately, the direct model exhibited very poor performance in retrieving AODs from the remaining continent, with low R values ranging from 0.23 to 0.47 and large uncertainties (e.g., RMSE = 0.15–0.41) (Fig. 7a-c). In contrast, our model demonstrated much better performance across diverse continents, increasing the R by 51–230% and reducing the RMSE values by 46–49% (Fig. 7d-f). Additionally, the fractions within the EE and GCOS criteria were 1.3–1.7 and 1.3–1.6 times higher. This discrepancy arises because conventional machine-learning models often struggle to generalize from conditions learned in one region to another, leading to significant performance degradation. In contrast, our integrated transfer-learning and deep-learning framework leverages prior knowledge from MERRA-2 AOD simulations, substantially improving predictions in extensive regions with limited or no observations.

We conducted additional validation of our model on independent

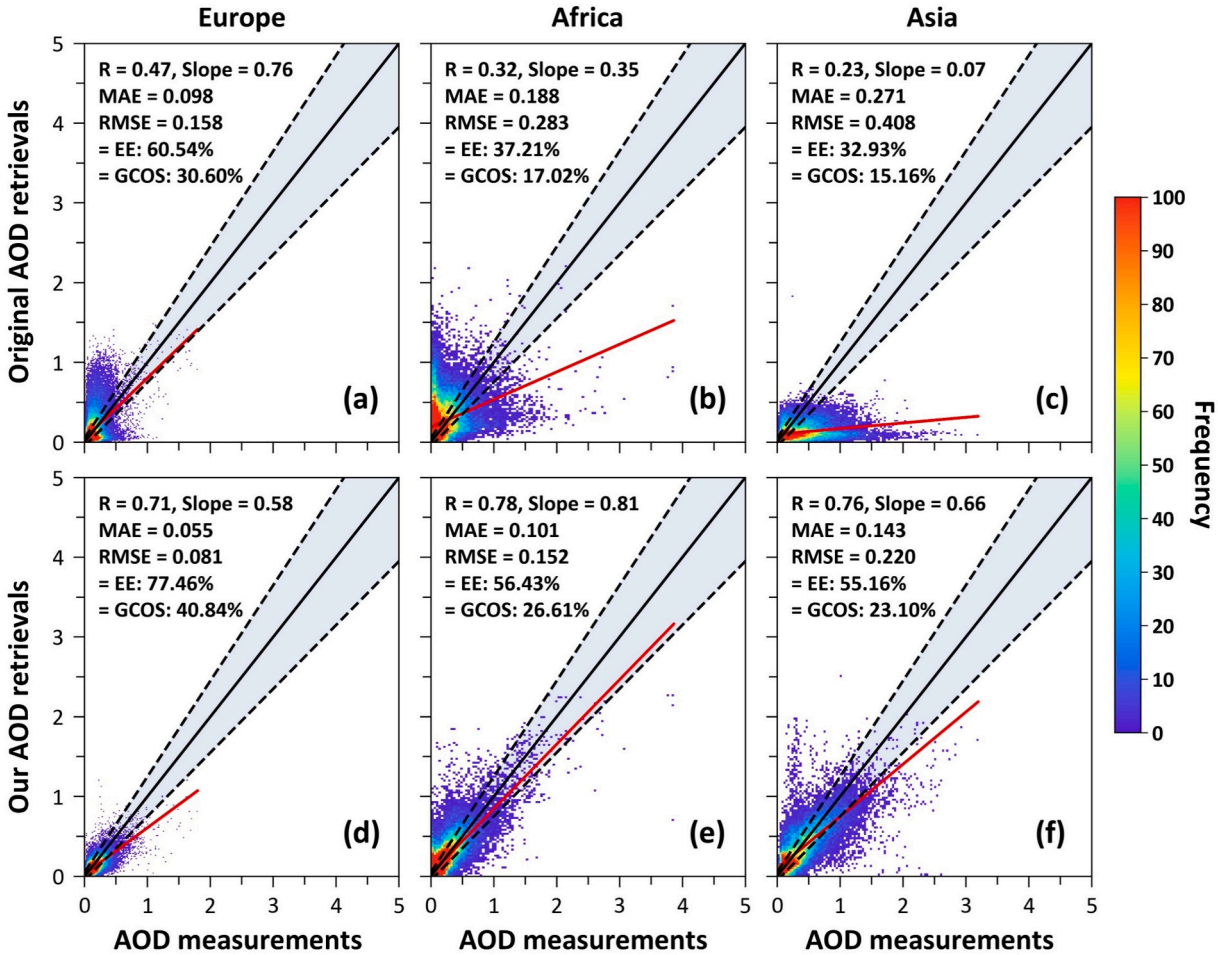

Fig. 7. Density scatter plots of (a-c) original (directly trained without pre-training) and (d-f) our (pre-trained + fine-tuned) hourly AOD retrievals against ground-based AOD measurements, with data samples from each continent, including Europe ($N = 69{,}099$), Africa ($N = 48{,}520$), and Asia ($N = 38{,}353$) withheld separately. In each panel, the black dotted line represents the expected error range, the black solid line indicates the 1:1 line, and the red solid line shows the linear regression line. The numbers in the top-left corner of the subfigure represent the correlation coefficient (R), slope of the regression line, mean absolute error (MAE), root mean square error (RMSE), percentage of retrievals within the expected error (EE), and Global Climate Observing System (GCOS). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

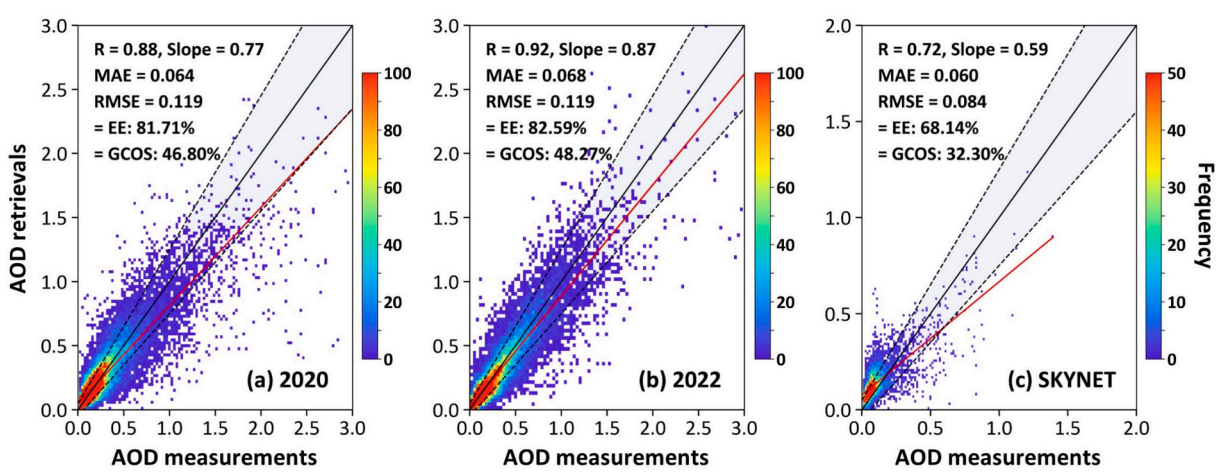

Fig. 8. Density scatter plots of AeroTrans-MSG AOD retrievals (550 nm) against ground-based AOD measurements for (a) 2020 ($N = 43{,}638$) and (b) 2022 (January–March, due to MSG-1 retirement; $N = 25{,}750$) using the model trained on 2021 data, and (c) independent validation from SKYNET ground-based measurements in 2021 ($N = 5527$). In each panel, the black dotted line represents the expected error range, the black solid line indicates the 1:1 line, and the red solid line shows the linear regression line. The numbers in the top-left corner of the subfigure represent the correlation coefficient (R), slope of the regression line, mean absolute error (MAE), root mean square error (RMSE), percentage of retrievals within the expected error (EE), and Global Climate Observing System (GCOS). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

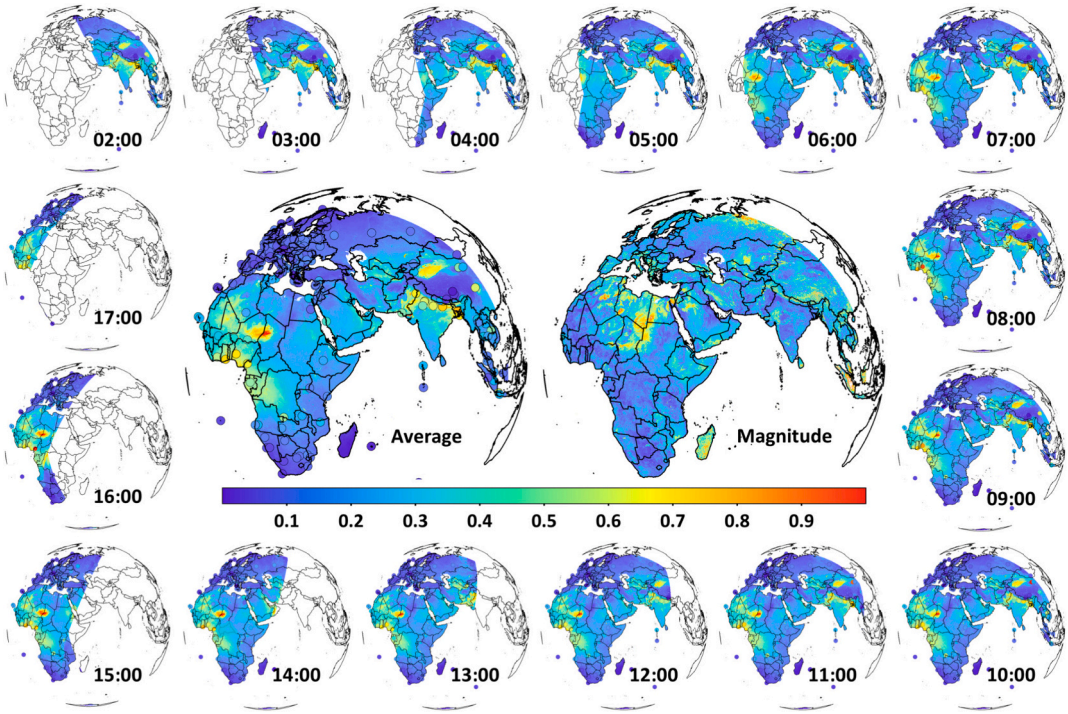

Fig. 9. Diurnal variations in AOD at a spatial resolution of 3 km from 02:00 to 17:00 UTC over Europe, Africa, and parts of Asia, along with daily average and diurnal variation magnitude maps for the entire year (2021).

years not used for training. Specifically, we evaluated AOD retrievals for 2022 and 2020 using the model trained on 2021 data (Fig. 8a-b). The model performs well on the 2020 holdout data despite the strong impacts of the COVID-19 pandemic, showing good agreement with ground-based measurements (R = 0.88, RMSE = 0.119), with ~82% of retrievals within the EE range and ~ 47% meeting the GCOS standard. Performance remains robust for 2022, with slightly higher correlation (R = 0.92), the same RMSE (0.119), ~83% of retrievals within the EE range, and 48% satisfying the GCOS criteria. These results demonstrate that our model is robust and adaptable under unseen aerosol conditions.

SEVIRI observations are affected by systematic biases of up to 10%, which can vary over time and across different MSG units (Meirink et al., 2013). It is worth noting that statistical approaches, particularly machine learning, are generally more susceptible to biases than physically based methods. In our study, the combined Transformer and transfer-learning framework mitigates some of these systematic effects by learning generalized TOA reflectance-AOD relationships from reanalysis and ground-based observations, as supported by strong independent validation results for different years (2020 and 2022). Nevertheless, residual biases may still exist, contributing to uncertainties in regions with limited ground validation.

Lastly, we validated our model's performance using the SKYNET network (Fig. 8c), providing a spatiotemporally independent assessment that further confirmed its ability to retrieve AOD at unseen locations and new times. The validation results showed a moderate correlation with SKYNET AOD measurements (R = 0.72, number of samples = 5527) and an RMSE of 0.084. Moreover, 76% of the collocated samples fell within the EE envelope, and 32% met the more stringent GCOS criteria.

Unlike traditional radiative-transfer-based approaches that rely heavily on blue-band sensitivity to aerosol scattering, our method employs a time-sequence deep-learning framework that jointly leverages information from multiple spectral bands and their temporal evolution. This allows the model to infer aerosol signals from correlated spectral-temporal features rather than relying on a single wavelength. In addition, the model is pretrained on simulated samples spanning diverse land-use types and atmospheric conditions, providing strong prior knowledge that helps mitigate the typical limitations of sensors lacking blue bands, such as SEVIRI. We further applied the same framework to Himawari-8/AHI observations and conducted experiments with and without the blue channel, demonstrating only minor differences in the accuracy of AOD retrievals (Fig. S5).

4.3. Diurnal variations in aerosol optical depth

Fig. 9 presents the annual mean diurnal distribution of AOD in 2021. Elevated AOD levels persist throughout the day over West Africa, particularly across the Sahel region, and over the Taklamakan Desert in China. These patterns are especially pronounced during the Northern Hemisphere winter (December–February) (Fig. S6) and spring (March–May) (Fig. S7), when dust transport is most active, showing substantially higher AOD loads but relatively low diurnal variation magnitudes. Capturing such diurnal cycles, as shown over the Sahara, can substantially affect weather forecasts and climate model outputs (Kocha et al., 2013), while similar considerations apply for modeling volcanic plumes (Plu et al., 2021). Elevated AOD is also evident in parts of Central Africa, such as Gabon and the Congo, where biomass burning, particularly wildfires (Giglio et al., 2018; Wimberly et al., 2024b), peaks during the Northern Hemisphere summer (June–August) (Fig. S8). In South Asia, especially in Pakistan and northern India, high aerosol loads remain high year-round, driven by continuous anthropogenic emissions and persistent pollution buildup. In contrast, Europe generally exhibits low AOD values, with slightly elevated levels in the Mediterranean and Eastern Europe during midday. Similarly, low AOD loads are observed in southern Africa and southwestern China, highlighting regional differences in emissions, land use, and atmospheric conditions.

Overall, AOD tends to rise gradually after sunrise, peak around 9:00 to 10:00 UTC, and decline by late afternoon, following the daily cycle of aerosol loading and atmospheric mixing. Similar diurnal variations are observed across different seasons (Figs. S6-S9). The magnitude of diurnal variation reveals that, while hourly changes are less dramatic than those seen in single-day analyses, certain regions, especially Central and West Africa, as well as parts of the Middle East, still exhibit

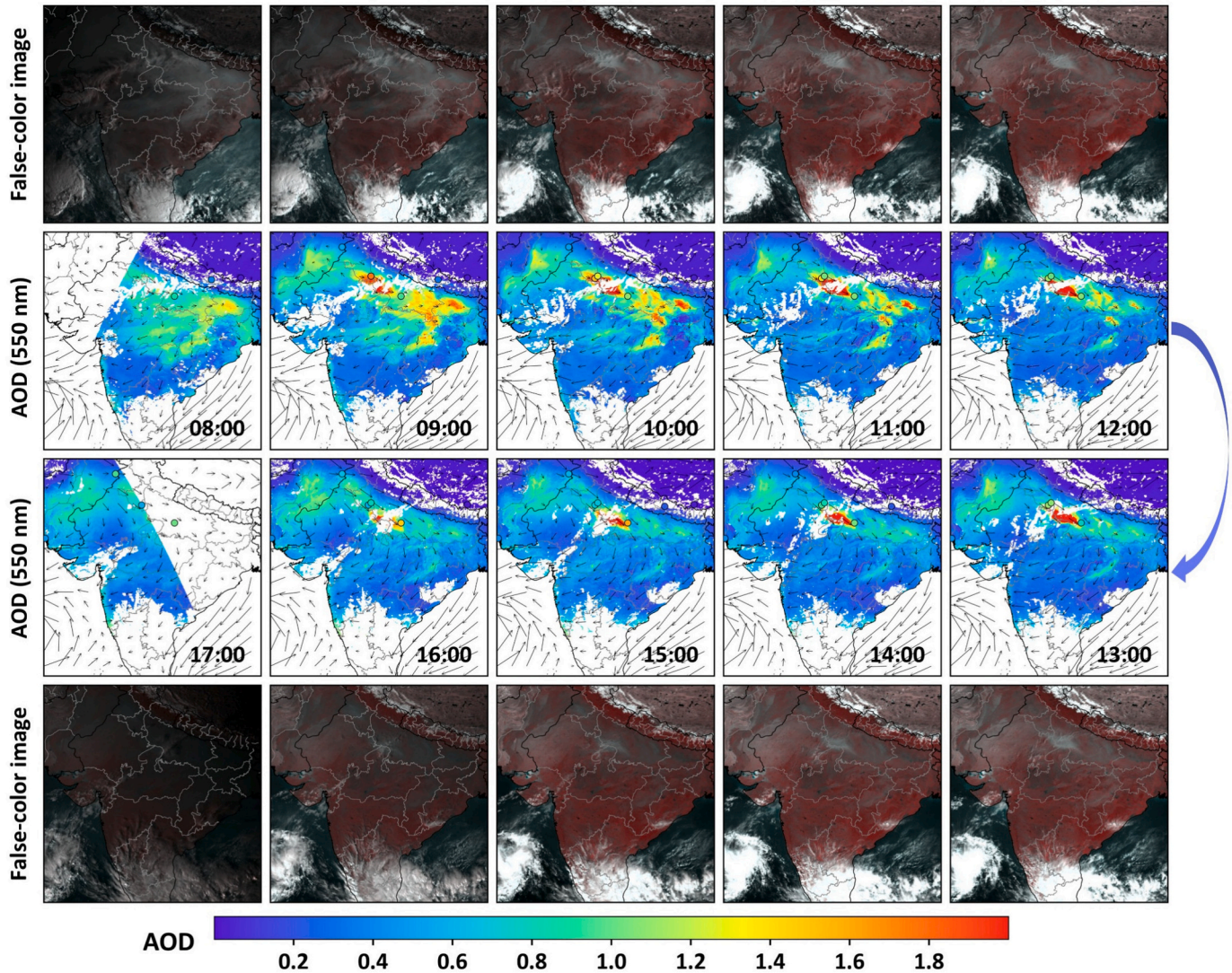

Fig. 10. Diurnal distribution and variation of AOD from 08:00–17:00 local time (UTC + 5:00) in India (70°E–90°E, 10°N–30°N) during a severe urban haze event on November 8, 2021. Colored dots represent ground-based AOD measurements, and arrows indicate 10 m wind direction from the ERA5-Land reanalysis.

notable fluctuations (> 50%) in AOD throughout the day, indicating the influence of dynamic emission sources and varying meteorological conditions. On average, the annual diurnal variation in AOD is 26% ± 15% relative to the daily mean level (0.22 ± 0.14) across the domain, with clear seasonal differences. The largest variations occur during the northern summer, contributing 30% ± 19% of the mean AOD (0.27 ± 0.19), while the northern winter exhibits the smallest fluctuations, at 24% ± 18% of the mean (0.19 ± 0.16). The average variation is 30% ± 17% (mean AOD: 0.23 ± 0.17) in northern spring and 27% ± 15% (0.20 ± 0.14) in northern autumn (September–November), reflecting the combined effects of seasonal emissions and atmospheric dynamics.

4.4. Diurnal AOD variations during extreme events

Fig. 10 presents the hourly variation of AOD at 550 nm from 08:00 to 17:00 local time (UTC + 5:00) during a severe urban haze episode over northern India on November 8, 2021. In the early morning (08:00–10:00), the AOD values generally increased, peaking near 10:00–11:00, especially over the densely populated states of Uttar Pradesh and Bihar. This morning's peak maximum reflects the combined effects of boundary-layer dynamics, local emission patterns, and meteorological conditions (e.g., winds) that limit aerosol dispersion. As the day progressed (10:00–17:00), the AOD distribution expanded and shifted, while overall haze diminished. AOD values decreased significantly, and the aerosol plume spread northwestward into eastern Punjab and Haryana, and northeastward into West Bengal. The 10-m wind vectors reveal a clear shift from weak and variable winds in the morning to more organized southeasterly and easterly flows around midday. This shift appears to facilitate the horizontal advection of haze from eastern Uttar Pradesh and Bihar toward Delhi and neighboring regions, exacerbating pollution in already burdened urban regions. The haze plume maintains broad spatial coverage, and there is a weak but discernible westward extension of aerosols into Haryana and parts of eastern Pakistan, consistent with continued but relatively slow east-to-west transport indicated by the wind field. During this haze event, our retrievals capture the morning increase and afternoon decline in AOD, peaking around 11:00 UTC, with strong agreement with ground-based measurements (R = 0.86; Fig. 11a). A similar heavy haze pattern and transport pathway was observed in Pakistan on November 20, 2021 (Fig. S10).

Fig. 12 shows the diurnal evolution of AOD during a severe wildfire event on August 4, 2021, over central and western Africa. In the morning

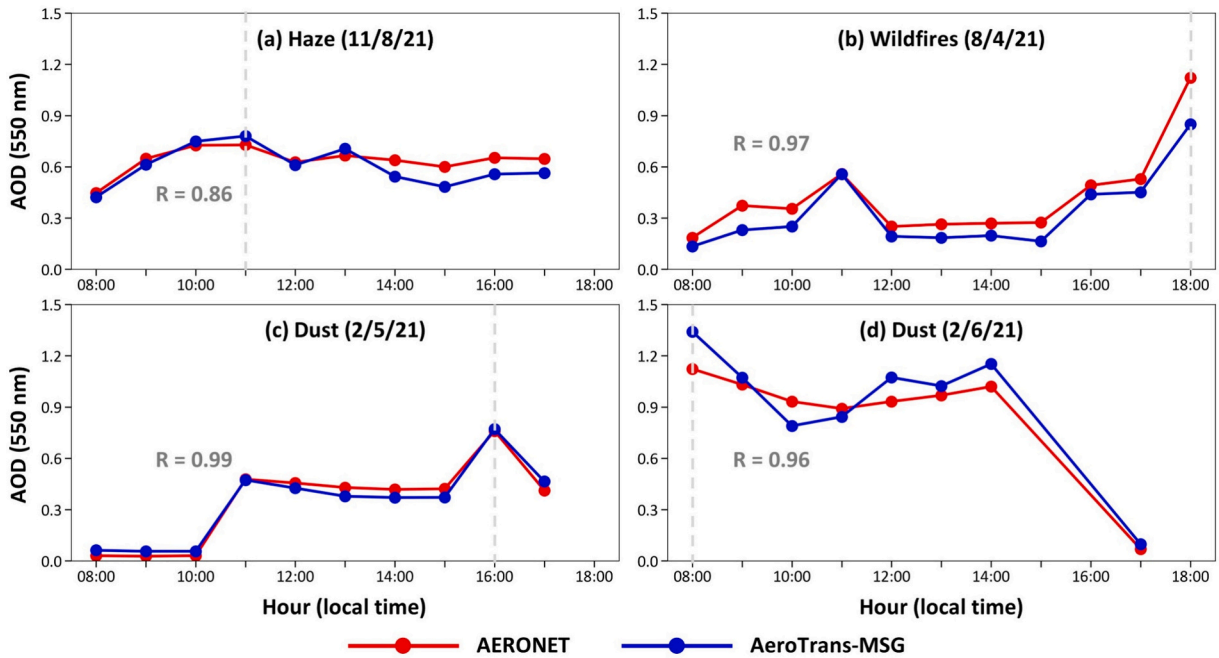

Fig. 11. Time series of our AeroTrans-MSG AOD (550 nm) retrievals (bule) and AERONET AOD measurements (red) from 08:00–18:00 local time for four representative events: (a) a severe urban haze event over India (67°E–89°E, 11°N–33°N) on November 8, 2021; (b) a typical wildfire event over central and western Africa (10°E–35°E, 17°S–8°N) on August 4, 2021; and (c-d) typical dust-storm events over the Sahara Desert (2°E–22°E, 7°N–24°N) on February 5 and 6, 2021. The gray dashed lines indicate the hour of peak daytime AOD. Statistics are derived from all available stations in the region using 3 × 3-pixel windows for AeroTrans-MSG images. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hours (08:00–09:00), the region experiences generally low AOD values, with high concentrations found in central regions around the dense fires. However, an intense wildfire broke out, causing a sharp increase in AOD values around the fire location. Thickening smoke plumes, visible in the false-colour satellite images, are indicated by red fire markers from MODIS fire products, and the and the plume continued to grow and emit smoke during 10:00–11:00. Prevailing winds drove the smoke northwestward, and by afternoon (14:00–17:00), the smoke plume gradually dissipated, leading to a significant reduce in AOD, which reached its minimum in the late afternoon (around 17:00). However, wind-driven smoke caused a significant increase in aerosol loading in the northwest during the afternoon (14:00–17:00), despite the shrinking fires. For this wildfire event, our model reproduces the observed bimodal pattern with peaks near 11:00 and 18:00 UTC, showing excellent correlation with ground measurements (R = 0.97; Fig. 11b). A similar spatial pattern and temporal evolution were observed during another wildfire event in Greece on August 8, 2021 (Fig. S11). These results demonstrated our model's ability to capture localized aerosol diurnal variations driven by both anthropogenic and natural emission sources.

Fig. 13 illustrates the diurnal variation of AOD during a typical dust storm event over the Sahara Desert spanning February 4–7, 2021. On February 4, the region exhibits generally low AOD values, indicating relatively clean air conditions. By February 5, however, high AOD values (> 1) appear over the southwestern part of the domain between 10:00 and 15:00, suggesting strong surface dust emissions. In addition, a new dust source emerges in the eastern part during the morning hours, with AOD values beginning to rise around 10:00, intensifying by 13:00, and further spreading westward by 15:00. This evolution coincides with northeasterly winds that facilitate the westward transport of the newly lofted dust toward the central and southwestern regions, contributing to elevated AOD levels. On February 6, the dust intensity gradually weakens throughout the day, particularly in the southwestern region, where concentrations drop to nearly half of the previous day's levels. Despite this reduction, the spatial extent of the dust-affected area remains broad, continuing to cover much of the domain, which may reflect persistent long-range transport or lingering suspended dust. By February 7, both the intensity and spatial coverage of the dust storm exhibit a marked decrease. The central region shows significant cleaning, with AOD values falling below 0.4, indicating that much of the dust has dissipated. Only a few localized areas, such as the southwestern corner and portions of the eastern dust source region, still retain relatively high AOD values. For the Sahara dust event (Fig. 11c-d), showing an almost perfect correlation on February 5, 2021 (R = 0.99) and similarly high agreement on February 6, 2021 (R = 0.96), accurately capturing both the daytime increase and the abrupt decline in AOD. A similar sequence of dust generation, outbreak, transport, and subsequent weakening was also observed over the Taklimakan Desert in China on March 6, 2021 (Fig. S12). We further compared our retrievals with MERRA-2 AOD and found that the two results do not appear overly similar, indicating that the model does not simply replicate the MERRA-2 data, even though it has more merit under such unfavorable conditions for remote sensing of AOD due to low contrast (Fig. S13). To some extent, our approach enhances aerosol retrieval by providing higher spatial resolution and more detailed AOD information over deserts.

Lastly, we compared our results with AOD retrievals from MSG-4/SEVIRI AERUS-GEO (Carrer et al., 2014), Terra and Aqua MODIS DB, and NOAA-20/VIIRS DB for the extreme events (Fig. S14). In addition, the available daily-average SEVIRI AODs are substantially lower than the instantaneous MODIS retrievals, likely due to strong diurnal variability; moreover, AERUS-GEO AOD is measured at 635 nm, while other products are at 550 nm, so lower values are expected at the longer wavelength. Our retrievals are generally consistent with the polar-orbiting products at comparable observation hours, while explicitly capturing the full diurnal cycle and accurately tracking the temporal evolution of these extreme events (Fig. 10–13). Nevertheless, our model is pretrained on MERRA-2, which has relatively low spatial resolution and higher uncertainty in regions with sparse ground-based measurements. Consequently, in such under-monitored regions, the model may rely more heavily on MERRA-2 AOD reanalysis (Fig. S15), resulting in increased variability in performance.

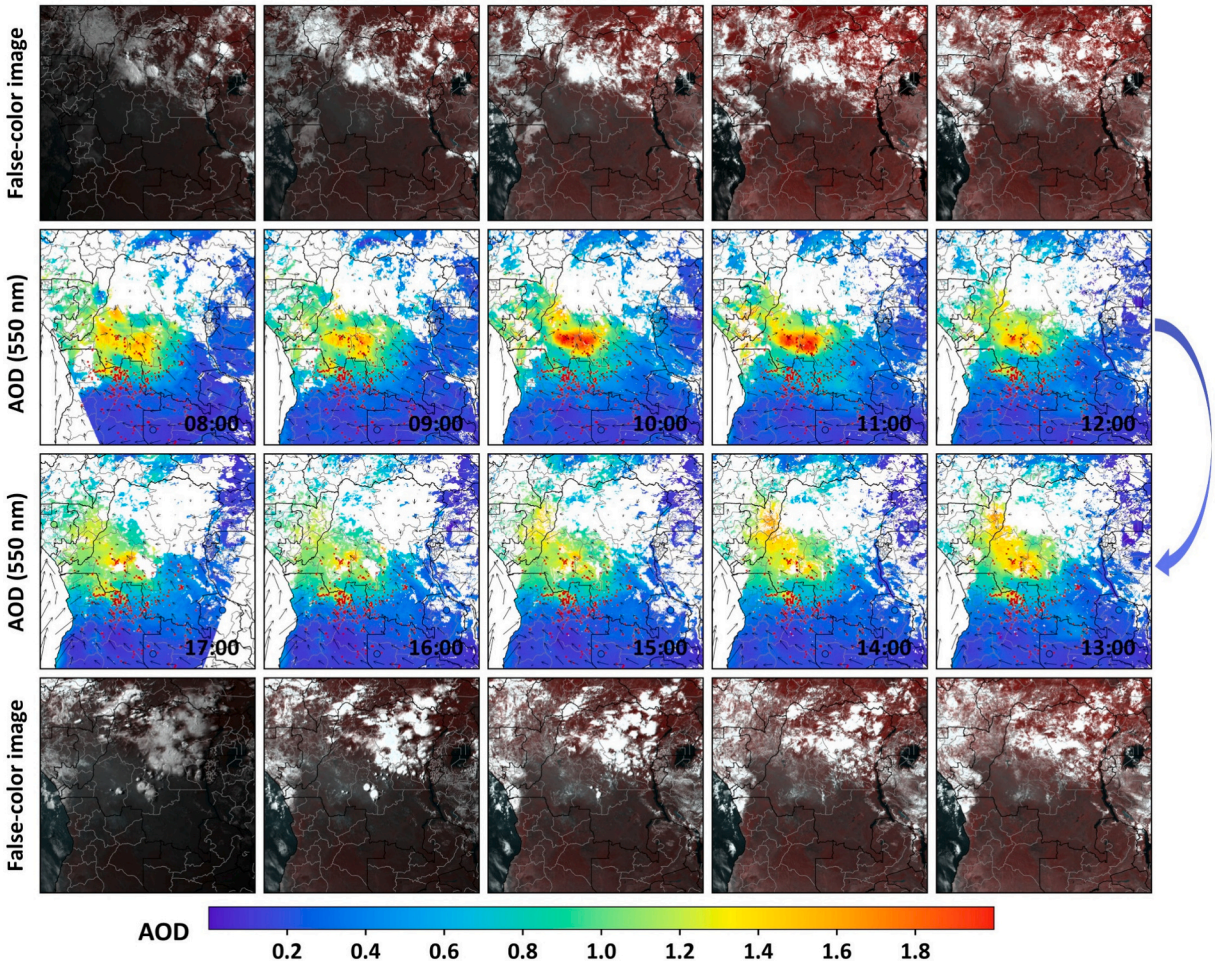

Fig. 12. Diurnal distribution and variation of AOD from 08:00–17:00 local time (UTC + 1:00) in central and western Africa (10°E–35°E, 17°S–8°N) during a typical wildfire event on August 4, 2021. Red points denote daily active fire detections from the MCD14A1 product. Colored dots represent ground-based AOD measurements, and arrows indicate 10 m wind direction from the ERA5-Land reanalysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Summary and conclusions

Geostationary satellites are highly valuable for monitoring diurnal variations of aerosols. In this study, we developed a novel hourly aerosol optical depth (AOD) retrieval framework for MSG-1/SEVIRI imagery by integrating an advanced deep-learning Transformer model with transfer learning (AeroTrans-MSG). This model is first pre-trained using MERRA-2 AOD to learn broad aerosol background patterns, and then fine-tuned with in situ AOD measurements to more accurately capture real-world aerosol–radiance relationships, ultimately enhancing its spatiotemporal predictive capabilities. Explainable artificial intelligence analysis using SHAP (SHapley Additive exPlanations) reveals that multi-band spectral channels (~38%), observation geometry (~34%), and key gaseous variables (~13%) play dominant roles in hourly aerosol retrieval. Finally, the model performance is thoroughly assessed using a variety of spatiotemporal validation strategies.

Overall, our AeroTrans-MSG model demonstrated superior performance in retrieving hourly AOD, as indicated by an average sample-based CV-R^2 of 0.88 and an RMSE of 0.079. Approximately 86% and 49% of retrievals meet the EE and GCOS uncertainty requirements, respectively. Additionally, our model exhibited strong predictive and extrapolative capabilities in regions and periods lacking ground measurements, as evidenced by strong spatial and temporal cross-validation evaluation metrics (CV-R^2 = 0.71–0.86) based on AERONET and SONET networks, as well as independent validation against the SKYNET network (R = 0.72). Our model also significantly improved predictive performance, particularly at the spatial scale, with R increasing by 51–230% when withholding each continent, compared to models trained directly without transfer learning. Using the generated hourly AOD dataset, we revealed distinct diurnal aerosol variations, characterized by a gradual rise after sunrise, a midday peak, and an afternoon decline, with a substantial 26% ± 15% change relative to the daily mean (0.22 ± 0.14), especially during the Northern Hemisphere summer (June–August: 30% ± 19% relative to the mean of 0.27 ± 0.19). More importantly, our study enabled us to elucidate the rapid dispersion of aerosols over West Asia and the Indian Ocean regions by leveraging the unique capability of the IODC satellite.

The integrated deep-learning and transfer-learning framework developed in this study will be extended to other platforms, such as MSG-2 and MTG. In addition, we plan to systematically evaluate and compare our retrievals with other satellite products to assess performance, quantify uncertainties, and examine applicability under different atmospheric and surface conditions. Future methodological improvements will focus on incorporating additional ground-based observations in data-sparse regions and exploring more advanced deep-learning architectures, such as diffusion and generative models, to better capture complex spatiotemporal aerosol patterns. Ultimately, this framework will be extended to a global scale using GEO-Ring observations to retrieve hourly global AOD and to investigate its diurnal variability over land.

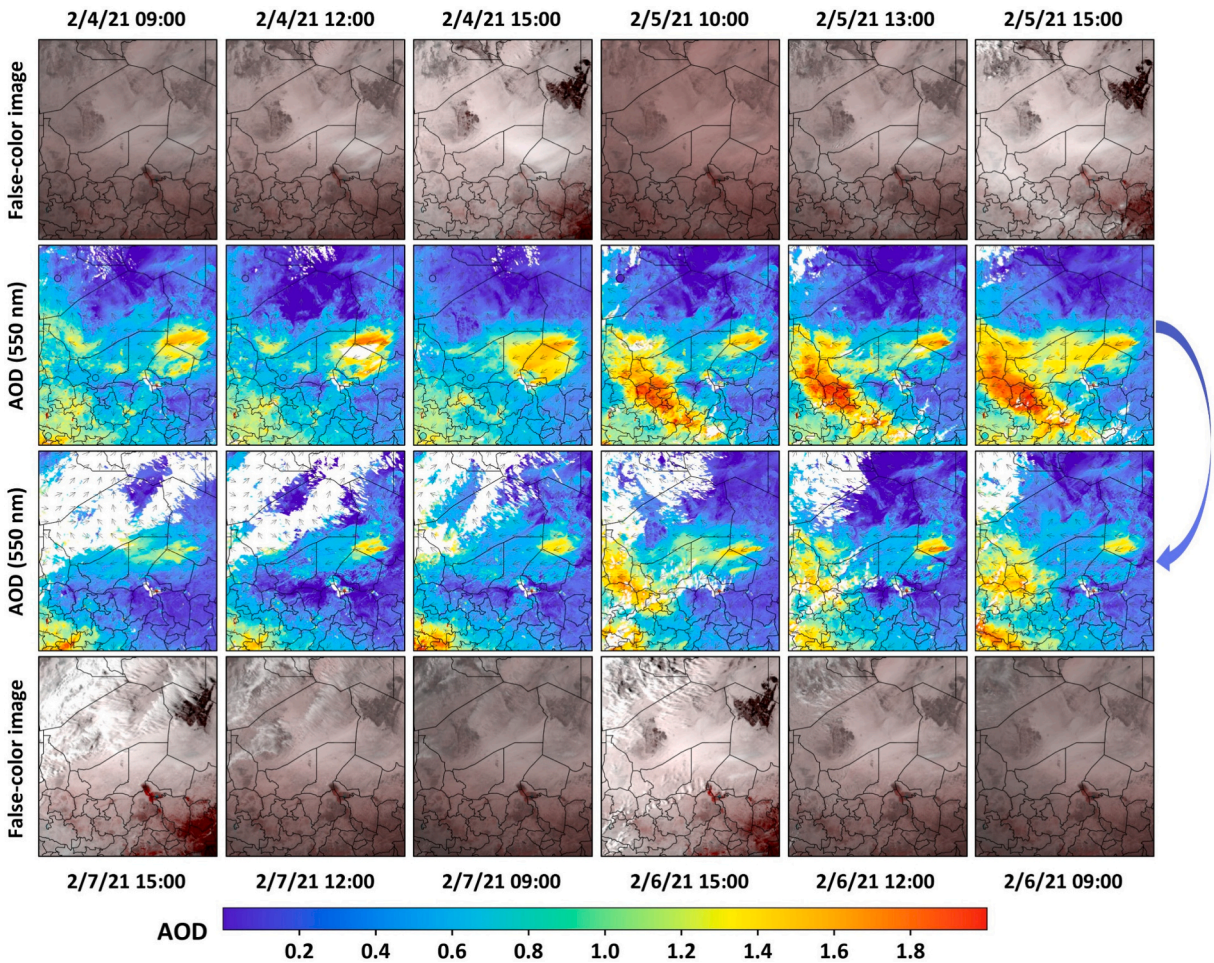

Fig. 13. Diurnal distribution and variation of AOD over the Sahara Desert (5°E–20°E and 10°N–25°N) during a typical dust storm event from February 4 to 7, 2021. Colored dots represent ground-based AOD measurements, and arrows indicate 10 m wind direction from the ERA5-Land reanalysis.

CRediT authorship contribution statement

Yulong Fan: Writing – original draft, Validation, Methodology, Formal analysis, Data curation. **Zhanqing Li:** Writing – review & editing, Supervision. **Lin Sun:** Writing – review & editing,Supervision, Funding acquisition. **Oleg Dubovik:** Writing – review & editing. **Zhihui Wang:** Validation, Formal analysis. **Fan Cheng:** Validation, Software. **Xiaohang Shi:** Validation, Software, Formal analysis. **Jing Wei:** Writing – review & editing, Supervision, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (42571461).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2026.115268.

Data availability

The ground-based AOD measurements from AERONET, SONET, and SKYNET are available at https://aeronet.gsfc.nasa.gov/, http://www.sonet.ac.cn/, and https://www.skynet-isdc.org/, respectively. The reanalysis MERRA-2 data is available at https://disc.gsfc.nasa.gov/. The MGS-1/SEVIRI satellite observation and corresponding cloud masks are available at https://test.icare.univ-lille.fr/. The ERA5-land reanalysis and multi-pressure-level data are available at https://cds.climate.copernicus.eu/. DEM is available at https://www.ncei.noaa.gov/products/. The MCD14A1 active fire product is available at https://ladsweb.modaps.eosdis.nasa.gov/. The MCD12Q1 land-cover-types product is available at https://ladsweb.modaps.eosdis.nasa.gov/. The AERUS-GEO aerosol product is available at https://test.icare.univ-lille.fr/. The MODIS DB aerosol product is available at https://ladsweb.modaps.eosdis.nasa.gov/. The VIIRS DB aerosol product is available at https://search.earthdata.nasa.gov/search/.

References

Arola, A., Eck, T.F., Huttunen, J., Lehtinen, K.E.J., Lindfors, A.V., Myhre, G., Smirnov, A., Tripathi, S.N., Yu, H., 2013. Influence of observed diurnal cycles of aerosol optical depth on aerosol direct radiative effect. Atmos. Chem. Phys. 13, 7895–7901.

Bernard, E., Moulin, C., Ramon, D., Jolivet, D., Riedi, J., Nicolas, J.M., 2011. Description and validation of an AOT product over land at the 0.6 μm channel of the SEVIRI sensor onboard MSG. Atmos. Meas. Tech. 4, 2543–2565.

Campanelli, M., Estelles, V., Smyth, T., Tomasi, C., Martìnez-Lozano, M.P., Claxton, B., Muller, P., Pappalardo, G., Pietruczuk, A., Shanklin, J., Colwell, S., Wrench, C., Lupi, A., Mazzola, M., Lanconelli, C., Vitale, V., Congeduti, F., Dionisi, D., Cardillo, F., Cacciani, M., Casasanta, G., Nakajima, T., 2012. Monitoring of Eyjafjallajökull volcanic aerosol by the new European Skynet radiometers (ESR) network. Atmos. Environ. 48, 33–45.

Carrer, D., Ceamanos, X., Six, B., Roujean, J.-L., 2014. AERUS-GEO: a newly available satellite-derived aerosol optical depth product over Europe and Africa. Geophys. Res. Lett. 41, 7731–7738.

Ceamanos, X., Six, B., Riedi, J., 2021. Quasi-global maps of daily aerosol optical depth from a ring of five geostationary meteorological satellites using AERUS-GEO. J. Geophys. Res. Atmos. 126, e2021JD034906.

Ceamanos, X., Six, B., Moparthy, S., Carrer, D., Georgeot, A., Gasteiger, J., Riedi, J., Attié, J.L., Lyapustin, A., Katsev, I., 2023. Instantaneous aerosol and surface retrieval using satellites in geostationary orbit (iAERUS-GEO) – estimation of 15 min aerosol optical depth from MSG/SEVIRI and evaluation with reference data. Atmos. Meas. Tech. 16, 2575–2599.

Chen, C., Litvinov, P., Dubovik, O., Fuertes, D., Matar, C., Miglietta, F., Pepe, M., Genesio, L., Busetto, L., Bindreiter, L., Lanzinger, V., de Graaf, M., Tilstra, G., Stammes, P., Retscher, C., 2024. Retrieval of aerosol and surface properties at high spatial resolution: hybrid approach and demonstration using Sentinel-5p/TROPOMI and PRISMA. J. Geophys. Res. Atmos. 129, e2024JD041041.

Cheng, F., Li, Z., Yang, Z., Li, R., Wang, D., Jia, A., Li, K., Zhao, B., Wang, S., Yin, D., Li, S., Xue, W., Cribb, M., Wei, J., 2025. First retrieval of 24-hourly 1-km-resolution gapless surface ozone (O_3) from space in China using artificial intelligence: diurnal variations and implications for air quality and phytotoxicity. Remote Sens. Environ. 316, 114482.

Chin, M., Rood, R., Lin, S., Müller, J., Thompson, A., 2000. Atmospheric sulfur cycle simulated in the global model GOCART: model description and global properties. J. Geophys. Res. Atmos. 105 (D20), 24671–24687.

Chowdhury, S., Hänninen, R., Sofiev, M., Aunan, K., 2024. Fires as a source of annual ambient PM2. 5 exposure and chronic health impacts in Europe. Sci. Total Environ. 922, 171314.

de Hoogh, K., Chen, J., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Klompmaker, J., Martin, R.V., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Brunekreef, B., Hoek, G., 2018. Spatial $PM_{2.5}$, NO_2, O_3 and BC models for Western Europe – evaluation of spatiotemporal stability. Environ. Int. 120, 81–92.

Dong, J., Zhang, T., Wang, L., Li, Z., Sing Wong, M., Bilal, M., Zhu, Z., Mao, F., Xia, X., Han, G., Xu, Q., Gu, Y., Lin, Y., Zhao, B., Li, Z., Xu, K., Chen, X., Gong, W., 2024. First retrieval of daily 160 m aerosol optical depth over urban areas using Gaofen-1/6 synergistic observations: algorithm development and validation. ISPRS J. Photogramm. Remote Sens. 211, 372–391.

Dupuy, J.-L., Fargeon, H., Martin-StPaul, N., Pimont, F., Ruffault, J., Guijarro, M., Hernando, C., Madrigal, J., Fernandes, P., 2020. Climate change impact on future wildfire danger and activity in southern Europe: a review. Ann. For. Sci. 77, 35.

Engelstaedter, S., Tegen, I., Washington, R., 2006. North African dust emissions and transport. Earth Sci. Rev. 79, 73–100.

Fan, Y., Sun, L., Liu, X., 2023. GOCI-II geostationary satellite hourly aerosol optical depth obtained by data-driven methods: validation and comparison. Atmos. Environ. 310, 119965.

Friedl, M., Sulla-Menashe, D., Tan, B., et al., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182.

Fu, D., Shi, H., Gueymard, C.A., Yang, D., Zheng, Y., Che, H., Fan, X., Han, X., Gao, L., Bian, J., Duan, M., Xia, X., 2024. A deep-learning and transfer-learning hybrid aerosol retrieval algorithm for FY4-AGRI: development and verification over Asia. Engineering 38, 164–174.

GCOS, 2010. Systematic Observation Requirements for Satellite-Based Products for Climate, 2011 Update, Supplemental Details to the Satellite-Based Component of the Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update). Global Climate Observing System, World Meteorological Organization, Geneva.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Zhao, B., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. 30 (14), 5419–5454.

Giglio, L., Boschetti, L., Roy, D.P., Humber, M.L., Justice, C.O., 2018. The collection 6 MODIS burned area mapping algorithm and product. Remote Sens. Environ. 217, 72–85.

Giles, D.M., Sinyuk, A., Sorokin, M.G., Schafer, J.S., Smirnov, A., Slutsker, I., Eck, T.F., Holben, B.N., Lewis, J.R., Campbell, J.R., Welton, E.J., Korkin, S.V., Lyapustin, A.I., 2019. Advancements in the aerosol robotic network (AERONET) version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. Atmos. Meas. Tech. 12, 169–209.

GMAO, 2015. MERRA-2 const_2d_asm_Nx: 2d, constants V5.12.4. Earth Sciences Data and Information Services Center (GES DISC, Greenbelt, MD, USA, Goddard.

Guerreiro, C.B.B., Foltescu, V., de Leeuw, F., 2014. Air quality status and trends in Europe. Atmos. Environ. 98, 376–384.

Ha, S., Liu, Z., Sun, W., Lee, Y., Chang, L., 2020. Improving air quality forecasting with the assimilation of GOCI aerosol optical depth (AOD) retrievals during the KORUS-AQ period. Atmos. Chem. Phys. 20, 6015–6036.

Holben, B.N., Eck, T.F., Slutsker, I., Tanré, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y.J., Nakajima, T., Lavenu, F., Jankowiak, I., Smirnov, A., 1998. AERONET—A federated instrument network and data archive for aerosol characterization. Remote Sens. Environ. 66, 1–16.

Hsu, N.C., Jeong, M.-J., Bettenhausen, C., Sayer, A.M., Hansell, R., Seftor, C.S., Huang, J., Tsay, S.-C., 2013. Enhanced deep blue aerosol retrieval algorithm: the second generation. J. Geophys. Res. Atmos. 118, 9296–9315.

Hsu, N.C., Lee, J., Sayer, A.M., Kim, W., Bettenhausen, C., Tsay, S.-C., 2019. VIIRS deep blue aerosol products over land: extending the EOS long-term aerosol data records. J. Geophys. Res. Atmos. 124, 4026–4053.

Kang, Y., Kim, M., Kang, E., Cho, D., Im, J., 2022. 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. 183, 253–268.

Kniffka, Anke, Stengel, Martin, Hollmann, Rainer, 2014. SEVIRI cloud mask dataset - Edition 1 - 15 minutes resolution. Satellite Application Facility on Climate Monitoring. https://doi.org/10.5676/EUM_SAF_CM/CMA_SEVIRI/V001. https://doi.org/10.5676/EUM_SAF_CM/CMA_SEVIRI/V001.

Kocha, C., Tulet, P., Lafore, J., Flamant, C., 2013. The importance of the diurnal cycle of aerosol optical depth in West Africa. Geophys. Res. Lett. 40, 785–790.

Laszlo, I., Liu, H., 2016. EPS Aerosol Optical Depth (AOD) Algorithm Theoretical Basis Document. NOAA NESDIS Center for Satellite Applications Research, pp. 1–91.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Levy, R.C., Mattoo, S., Munchak, L.A., Remer, L.A., Sayer, A.M., Patadia, F., Hsu, N.C., 2013. The collection 6 MODIS aerosol products over land and ocean. Atmos. Meas. Tech. 6, 2989–3034.

Li, J., Han, Z., Zhang, R., 2014. Influence of aerosol hygroscopic growth parameterization on aerosol optical depth and direct radiative forcing over East Asia. Atmos. Res. 140-141, 14–27.

Li, Z., Lau, W.K.-M., Ramanathan, V., Wu, G., Ding, Y., Manoj, M.G., Liu, J., Qian, Y., Li, J., Zhou, T., Fan, J., Rosenfeld, D., Ming, Y., Wang, Y., Huang, J., Wang, B., Xu, X., Lee, S.-S., Cribb, M.C., Zhang, F., Yang, X., Zhao, C., Takemura, T., Wang, K., Xia, X., Yin, Y., Zhang, H., Guo, J., Zhai, P.M., Sugimoto, N., Babu, S.S., Brasseur, G. P., 2016. Aerosol and monsoon interactions in Asia. Rev. Geophys. https://doi.org/10.1002/2015 RG000500.

Li, Z.Q., Xu, H., Li, K.T., Li, D.H., Xie, Y.S., Li, L., Zhang, Y., Gu, X.F., Zhao, W., Tian, Q.J., Deng, R.R., Su, X.L., Huang, B., Qiao, Y.L., Cui, W.Y., Hu, Y., Gong, C.L., Wang, Y.Q., Wang, X.F., Wang, J.P., Du, W.B., Pan, Z.Q., Li, Z.Z., Bu, D., 2018. Comprehensive study of optical, physical, chemical, and radiative properties of total columnar atmospheric aerosols over China: an overview of Sun–Sky Radiometer Observation Network (SONET) measurements. Bull. Am. Meteorol. Soc. 99, 739–755.

Luffarelli, M., Govaerts, Y., 2019. Joint retrieval of surface reflectance and aerosol properties with continuous variation of the state variables in the solution space–part 2: application to geostationary and polar-orbiting satellite observations. Atmos. Meas. Tech. 12 (2), 791–809.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc., Long Beach, California, USA, pp. 4768–4777.

Lyapustin, A., Wang, Y., Korkin, S., Huang, D., 2018. MODIS collection 6 MAIAC algorithm. Atmos. Meas. Tech. 11, 5741–5765.

Mei, L., Rozanov, V., Vountas, M., Burrows, J.P., Richter, A., 2018. XBAER-derived aerosol optical thickness from OLCI/Sentinel-3 observation. Atmos. Chem. Phys. 18, 2511–2523.

Meirink, J.F., Roebeling, R.A., Stammes, P., 2013. Inter-calibration of polar imager solar channels using SEVIRI. Atmos. Meas. Tech. 6, 2495–2508.

Nakajima, T., Campanelli, M., Che, H., Estellés, V., Irie, H., Kim, S.W., Kim, J., Liu, D., Nishizawa, T., Pandithurai, G., Soni, V.K., Thana, B., Tugjsurn, N.U., Aoki, K., Go, S., Hashimoto, M., Higurashi, A., Kazadzis, S., Khatri, P., Kouremeti, N., Kudo, R., Marenco, F., Momoi, M., Ningombam, S.S., Ryder, C.L., Uchiyama, A., Yamazaki, A., 2020. An overview of and issues with sky radiometer technology and SKYNET. Atmos. Meas. Tech. 13, 4195–4218.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.

Plu, M., Bigeard, G., Sič, B., Emili, E., Bugliaro, L., El Amraoui, L., Guth, J., Josse, B., Mona, L., Piontek, D., 2021. Modelling the volcanic ash plume from Eyjafjallajökull eruption (may 2010) over Europe: evaluation of the benefit of source term improvements and of the assimilation of aerosol measurements. Nat. Hazards Earth Syst. Sci. 21, 3731–3747.

Popp, C., Hauser, A., Foppa, N., Wunderle, S., 2007. Remote sensing of aerosol optical depth over Central Europe from MSG-SEVIRI data and accuracy assessment with ground-based AERONET measurements. J. Geophys. Res. Atmos. 112.

Ramanathan, V., Crutzen, P.J., Kiehl, J.T., Rosenfeld, D., 2001. Aerosols, climate, and the hydrological cycle. Science 294, 2119–2124.

Randles, C.A., Da Silva, A.M., Buchard, V., Colarco, P.R., Darmenov, A., Govindaraju, R., Flynn, C.J., 2017. The MERRA-2 aerosol reanalysis, 1980 onward. Part I: system description and data assimilation evaluation. J. Clim. 30 (17), 6823–6850.

Sawyer, V., Levy, R.C., Mattoo, S., Cureton, G., Shi, Y., Remer, L.A., 2020. Continuing the MODIS dark target aerosol time series with VIIRS. J. Geophys. Res. Atmos. 12, 308.

She, L., Li, Z., de Leeuw, G., Wang, W., Wang, Y., Yang, L., Feng, Z., Yang, C., Shi, Y., 2024. Time series retrieval of multi-wavelength aerosol optical depth by adapting transformer (TMAT) using Himawari-8 AHI data. Remote Sens. Environ. 305, 114115.

Si, Y., Chen, L., Wang, Y., Xu, N., Zhang, X., Yang, L., Hu, X., Shi, S., 2024. An improved aerosol retrieval algorithm based on nonlinear surface model from FY-3D/MERSI-II remote sensing data. IEEE Trans. Geosci. Remote Sens. 62, 1–17.

Stier, P., Seinfeld, J.H., Kinne, S., Boucher, O., 2007. Aerosol absorption and radiative forcing. Atmos. Chem. Phys. 7, 5237–5261.

Su, X., Huang, Y., Wang, L., Cao, M., Feng, L., 2023. Validation and diurnal variation evaluation of MERRA-2 multiple aerosol properties on a global scale. Atmos. Environ. 311, 120019.

Takamura, T., Takenaka, H., Cui, Y., Nakajima, T.Y., Higurashi, A., Fukuda, S., Kikuchi, N., Nakajima, T., Sano, I., Pinker, R.T., 2009. Aerosol and cloud validation system based on SKYNET observations: estimation of shortwave radiation budget using ADEOS-II/GLI data. J. Remote Sens. Soc. Jpn. 29, 40–53.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A Survey on Deep Transfer Learning. Artificial Neural Networks and Machine Learning. Springer International Publishing, Cham, pp. 270–279.
Tang, C., Shi, C., Letu, H., Yin, S., Nakajima, T., Sekiguchi, M., Xu, J., Zhao, M., Ma, R., Wang, W., 2025. Development of a hybrid algorithm for the simultaneous retrieval of aerosol optical thickness and fine-mode fraction from multispectral satellite observation combining radiative transfer and transfer learning approaches. Remote Sens. Environ. 319, 114619.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. Neural Inf. Proc. Syst. 30, 5998–6008.
Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56.
Wang, Z., Liu, D., Wang, Y., Wang, Z., Shi, G., 2015. Diurnal aerosol variations do affect daily averaged radiative forcing under heavy aerosol loading observed in Hefei, China. Atmos. Meas. Tech. 8, 2901–2907.
Watson-Parris, D., Smith, C.J., 2022. Large uncertainty in future warming due to aerosol forcing. Nat. Clim. Chang. 12, 1111–1113.
Wei, J., Sun, L., Peng, Y., Wang, L., Zhang, Z., Bilal, M., Ma, Y., 2018. An improved high-spatial-resolution aerosol retrieval algorithm for MODIS images over land. J. Geophys. Res. Atmos. 123, 12291–12307.
Wei, J., Li, Z., Peng, Y., Sun, L., 2019a. MODIS collection 6.1 aerosol optical depth products over land and ocean: validation and comparison. Atmos. Environ. 201, 428–440.
Wei, J., Li, Z., Peng, Y., Sun, L., Yan, X., 2019b. A regionally robust high-spatial-resolution aerosol retrieval algorithm for MODIS images over eastern China. IEEE Trans. Geosci. Remote Sens. 57 (7), 4748–4757.
Wei, J., Li, Z., Pinker, R.T., Wang, J., Sun, L., Xue, W., Li, R., Cribb, M., 2021. Himawari-8-derived diurnal variations in ground-level $PM_{2.5}$ pollution across China using the fast space-time light gradient boosting machine (LightGBM). Atmos. Chem. Phys. 21, 7863–7880.
Wei, J., Li, Z., Lyapustin, A., Wang, J., Dubovik, O., Schwartz, J., Sun, L., Li, C., Liu, S., Zhu, T., 2023. First close insight into global daily gapless 1 km $PM_{2.5}$ pollution, variability, and health impact. Nat. Commun. 14, 8349.
Wei, J., Wang, Z., Li, Z., Li, Z., Pang, S., Xi, X., Cribb, M., Sun, L., 2024. Global aerosol retrieval over land from Landsat imagery integrating transformer and Google earth engine. Remote Sens. Environ. 315, 114404.
Wimberly, M., Wanyama, D., Doughty, R., Peiro, H., Crowell, S., 2024. Increasing fire activity in african tropical forests is associated with deforestation and climate change. Geophys. Res. Lett. 51, e2023GL106240.
Xia, X., Min, J., Sun, S., Chen, X., 2023. Simultaneous assimilation of Fengyun-4A and Himawari-8 aerosol optical depth retrieval to improve air quality simulations during one storm event over East Asia. Front. Earth Sci. 11.
Xie, Y., Xue, Y., Guang, J., Mei, L., She, L., Li, Y., Che, Y., Fan, C., 2020. Deriving a global and hourly data set of aerosol optical depth over land using data from four geostationary satellites: GOES-16, MSG-1, MSG-4, and Himawari-8. IEEE Trans. Geosci. Remote Sens. 58, 1538–1549.
Yang, Z., Zhang, M., Wang, L., Su, X., Qin, W., 2023. Diurnal time representation of MODIS, VIIRS, MISR, and AHI over Asia and Oceania. Remote Sens. Environ. 299, 113878.
Yeom, J.M., Jeong, S., Ha, J.S., Lee, K.H., Lee, C.S., Park, S., 2022. Estimation of the hourly aerosol optical depth from GOCI geostationary satellite data: deep neural network, machine learning, and physical models. IEEE Trans. Geosci. Remote Sens. 60, 1–12.
Yoshida, M., Kikuchi, M., Nagao, T.M., Murakami, H., Nomaki, T., Higurashi, A., 2018. Common retrieval of aerosol properties for imaging satellite sensors. J. Meteorol. Soc. Japan Ser. II 96B, 193–209.
Zawadzka, O., Markowicz, K., 2014. Retrieval of aerosol optical depth from optimal interpolation approach applied to SEVIRI data. Remote Sens. 6, 7182–7211.
Zhang, P., Lu, N.-M., Hu, X.-Q., Dong, C.-H., 2006. Identification and physical retrieval of dust storm using three MODIS thermal IR channels. Glob. Planet. Chang. 52, 197–206.
Zhang, Y., Yu, H., Eck, T.F., Smirnov, A., Chin, M., Remer, L.A., Bian, H., Tan, Q., Levy, R., Holben, B.N., Piazzolla, S., 2012. Aerosol daytime variations over north and South America derived from multiyear AERONET measurements. J. Geophys. Res. Atmos. 117.
Zhang, H., Kondragunta, S., Laszlo, I., Zhou, M., 2020. Improving GOES advanced baseline imager (ABI) aerosol optical depth (AOD) retrievals using an empirical bias correction algorithm. Atmos. Meas. Tech. 13, 5955–5975.
Zhang, H., Wei, Z., Henderson, B.H., Anenberg, S.C., O'Dell, K., Kondragunta, S., 2022. Nowcasting applications of geostationary satellite hourly surface $PM_{2.5}$. Data. Wea. Forecasting 37, 2313–2329.
Zheng, H., Shen, L., Tang, A., Luo, Y., Hu, H., Du, B., Tao, D., 2025. Learning from models beyond fine-tuning. Nat. Machine Intell. 7 (1), 6–17.