Environmental Science & Technology

pubs.acs.org/est

Article

Deep Learning with Pretrained Framework Unleashes the Power of Satellite-Based Global Fine-Mode Aerosol Retrieval

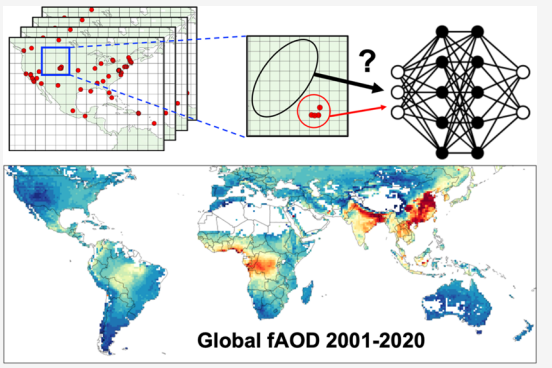
Xing Yan, Zhou Zang,* Zhanqing Li,* Hans W. Chen, Jiayi Chen, Yize Jiang, Yunhao Chen, Bin He, Chen Zuo, Terry Nakajima, and Jhoon Kim

Cite This: *Environ. Sci. Technol.* 2024, 58, 14260−14270

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Fine-mode aerosol optical depth (fAOD) is a vital proxy for the concentration of anthropogenic aerosols in the atmosphere. Currently, the limited data length and high uncertainty of the satellite-based data diminish the applicability of fAOD for climate research. Here, we propose a novel pretrained deep learning framework that can extract information underlying each satellite pixel and use it to create new latent features that can be employed for improving retrieval accuracy in regions without in situ data. With the proposed model, we developed a new global fAOD (at 0.5 μm) data from 2001 to 2020, resulting in a 10% improvement in the overall correlation coefficient (R) during site-based independent validation and a 15% enhancement in non-AERONET site areas validation. Over the past two decades, there has been a noticeable downward trend in global fAOD (-1.39×10^{-3}/year). Compared to the general deep-learning model, our method reduces the global trend's previously overestimated magnitude by 7% per year. China has experienced the most significant decline (-5.07×10^{-3}/year), which is 3 times greater than the global trend. Conversely, India has shown a significant increase (7.86×10^{-4}/year). This study bridges the gap between sparse in situ observations and abundant satellite measurements, thereby improving predictive models for global patterns of fAOD and other climate factors.



KEYWORDS: *fAOD, deep learning, pretrained framework, global trend, MODIS*

INTRODUCTION

Atmospheric fine-mode aerosols are predominately small particles (typically radius <1 μm) which are mostly anthropogenic and have considerable impacts on cloud microphysics and Earth's radiative balance. Fine-mode aerosol optical depth (fAOD) is a vital proxy for these concentrations of anthropogenic aerosols in the atmosphere. However, until now large uncertainty remains about the global fAOD changing in recent decades, especially over land where most anthropogenic aerosols are located (see Section S1).

Although fAOD can be obtained from point-scale Aerosol Robotic Network (AERONET) measurements, the spatial coverage is very limited. To monitor spatial coverage fAOD, satellite-based remote sensing combined with in situ data has been widely adopted in both physically- and machine-learning-based fAOD modeling. However, methods for bridging the gap between "point measurements" made at specific locations and "spatially continuous" satellite data remain inefficient. Figure 1 presents in situ measurements that correspond to pixel information from a satellite image at the same longitude and latitude. These matched data, also referred to as labeled data, are utilized in constructing an fAOD retrieval model.[1−4] However, due to the sparsity and limited coverage of ground-based monitoring stations,[5−8] not every satellite pixel data has a corresponding ground-observed fAOD (such unmatched satellite pixel data is also known as unlabeled data). As a consequence, there is often a wealth of satellite measurements never used for developing models. The vast potential of these unused data remains greatly unexplored and underexploited for model development. One of the most vital problems to tackle in earth environment modeling is how to make full use of these disregarded data.

The integration of satellite and in situ data has improved simulations of the fAOD, especially via the machine-learning approach.[9−13] Machine learning builds past or present-day relationships between ground-based observed fAOD and satellite information to develop models. In this process, in situ fAOD are used in machine learning to train the model and validate its results. However, a machine-learning model may incur large errors at locations distant from any in situ station.[14−18] Thus, there is an urgent need to improve the

Received: March 18, 2024
Revised: July 23, 2024
Accepted: July 24, 2024
Published: August 3, 2024

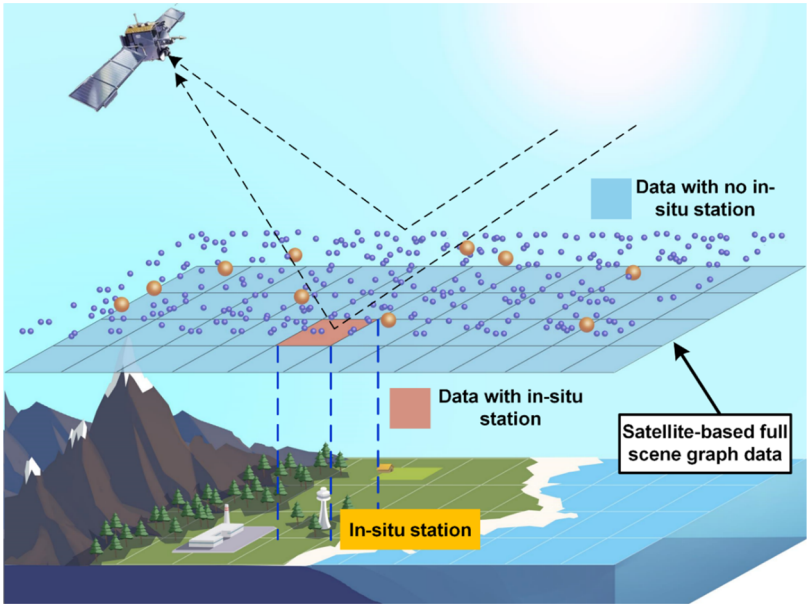


Figure 1. Satellite-based remote sensing in combination with in situ data modeling.

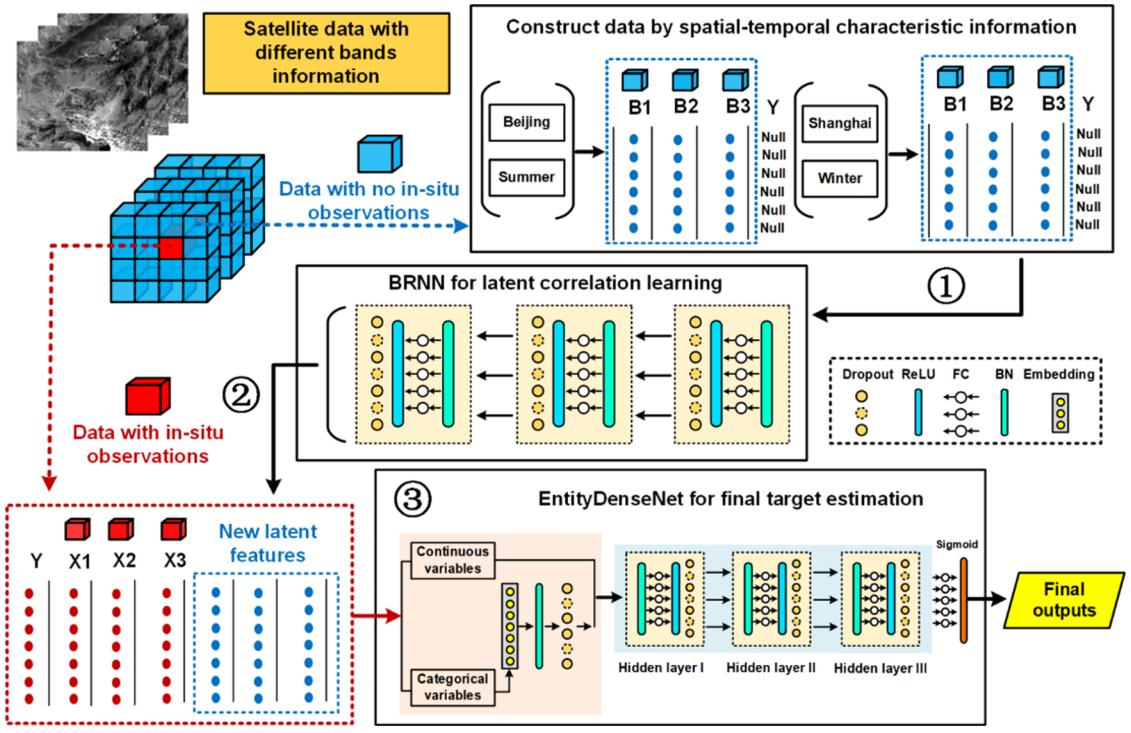


Figure 2. Schematic of the DLFE-Satellite model. The model contains three major parts. The first part in the top row shows the construction of data using spatiotemporal characteristic information from satellite data with no in situ observations. The second part in the middle row shows the use of the BRNN to learn the latent correlation from previously constructed data, producing new latent features. The third part in the last row shows the final target estimations in the deep-learning EntityDenseNet model using a combination of satellite data with in situ observations and new latent features.

accuracy of modeling in areas where no ground-based sites are available to assist in model training.[19,20]

In recent years, the successful application of pretrained models in natural language processing and global weather forecasting has attracted significant attention.[21] The core concept of pretrained model is that by exposing the model to a vast pool of unlabeled data, it can acquire generalizable and universal data representations.[22] These representations encapsulate the intrinsic structure and relationships within the data, enabling the model to generalize more effectively and perform well on subsequent tasks. For instance, in the field of natural language processing, pretrained language models such as generative pre-trained transformer-4 (GPT-4) in ChatGPT undergo pretraining on extensive text data (referred to as unlabeled data) without specific task-oriented guidance.[23] The acquired representations capture grammar and semantic

https://doi.org/10.1021/acs.est.4c02701
Environ. Sci. Technol. 2024, 58, 14260−14270

information, which can then be utilized as input for subsequent tasks like text classification, enhancing the model's overall performance. Therefore, We are in immediate need of developing pretrained models similar to GPT, which can learn useful representations of satellite signals from unlabeled satellite data and capture potential intrinsic relationships and patterns, enabling them to better capture information related to fAOD in downstream tasks. However, there is currently limited research on how to construct such pretrained models.

Here, to better understand the global fAOD changing in recent decades, we need to address two fundamental key questions: How can satellite data without corresponding in situ observations be used in fAOD modeling, and can this hitherto-disregarded data improve model performance? To solve these issues, we developed a novel pretrained deep-learning framework called the "deep learning framework of latent features extraction for satellite data with missing in situ observations" (DLFE-satellite) to make full use of each pixel value measured by satellites to train a model. Data-use efficiency is significantly improved by including all satellite data in the modeling process. In this study, we conducted a series of experiments at the national and global scales to test the new method and demonstrate how it leads to substantial accuracy improvements in areas where no in situ observations are available.

METHODS

Proposed New Pretrained Framework: DLFE-Satellite. To fully use the information from the unused satellite measurements, we developed a new pretrained framework called DLFE-Satellite.

Figure 2 shows the overall flow of the DLFE-Satellite framework. It consists of three main parts. The first part is to convert satellite image information into tabular data, which arranges data elements in vertical columns (features) and horizontal rows (samples). The tabular data were separated into two groups: one group containing information from satellite pixels with corresponding in situ data (labeled data; see the red boxes of X1, X2, X3··· in Figure 2), and the other group containing information from satellite pixels without it (unlabeled data; see the blue boxes of B1, B2, B3··· in Figure 2). The latter group of data was constructed using spatiotemporal characteristic information, a step needed for the next part of the approach. For example, as shown in Figure 2, information from satellite pixels located over Beijing in the summer was converted into one set of tabular data.

The second part of the DLFE-Satellite approach is pretraining by self-supervised learning for the satellite data with no in situ observations. We introduced the value imputation and mask estimation (VIME) approach[24] into the DLFE-Satellite which provides two novel pretext tasks for self-supervised learning. In DLFE-Satellite, a large amount of unlabeled data is exploited to train a model f (Figure S1, deep learning model 1). Then, the pretrained model f gets the ability to capture high-level representations of the unlabeled data and is transferred together with the labeled data to a supervised downstream task (deep learning model 2) for fAOD retrieval. The detailed processes to train f by Deep learning model 1 in DLFE-Satellite are:

Define the Pretext Tasks. Based on a certain VIME objective, two pretext tasks are used: (1) Mask vector estimation and (2) feature vector estimation (Figure S2). The purpose of Mask vector estimation is to predict the probability of which data have been masked. The output data is between 0 and 1, with values closer to 0 indicating that the data has not been masked, and values closer to 1 indicating that the data has been masked. The purpose of feature vector estimation is to predict the values of the data that have been corrupted.

Data Corruption and Masking. The constructed tabular data (satellite data with no in situ observations) in the first part of the DLFE-Satellite are used for generating corrupted data $\tilde{X}$

$$\tilde{X} = \overline{X} \odot M + X \odot (1 - M) \tag{1}$$

where $\odot$ is element-wise matrix multiplication; X is the original matrix from the constructed tabular data; M is a mask matrix from a Bernoulli distribution (value is 0 or 1); and $\overline{X}$ is randomly shuffled the original data within each feature.

Model Training and Optimization. A batch normalization and robust neural network (BRNN)[11] was used as the Deep learning model 1 in DLFE-Satellite for learning the latent correlation from X. The loss function of Mask vector estimation l_M is defined as

$$l_M = -\frac{1}{N}\left[\sum_{i=1}^{N} y_i \times \log(p_i) + (1 - y_i) \times \log(1 - p_i)\right] \tag{2}$$

where y_i is the ith corresponding true mask value 1 or 0; p_i is the ith probability of 1 from BRNN output; $1 - p_i$ is the probability of 0, and N is the number of rows.

The loss function of Feature vector estimation l_F, was defined as

$$l_F = \frac{1}{N}\sum_{i=1}^{N} (y_i^m - y_i^o)^2 \tag{3}$$

where y_i^m is the ith true value; y_i^o is the ith BRNN output estimate; and N is the number of rows.

The BRNN is final optimization by a combination of these two loss functions

$$\min(l_M + w \times l_F) \tag{4}$$

where w is an adjustment parameter for the two loss functions, determined by the training process. The pretrained BRNN generates new latent features for supervised downstream tasks (deep learning model 2).

The third part of the DLFE-Satellite is combining the newly generated latent features with satellite pixels having corresponding in situ data to Deep learning model 2 for fAOD retrieval. We introduced EntityDenseNet[25] as the deep learning model 2 for the final DLFE-Satellite fAOD retrieval. EntityDenseNet employs an optimized neural network capable of capturing intricate nonlinear relationships. Each hidden layer consists of a ReLU layer, a BN layer,[26] and a dropout layer.[27] Moreover, EntityDenseNet leverages the entity embedding method for the direct processing of categorical variables,[28] enabling comprehensive information extraction from such variables within the neural network.

Input Data for Global Satellite-Based fAOD Retrieval. The DLFE-Satellite model was utilized to estimate global 20-year (2001 to 2020) fAODs. For the purpose of model training and validation using ground-truth data, we collected the AERONET Version 3 data set,[29] which includes multispectral AOD data and fine-mode fraction (FMF) data derived via the spectral deconvolution algorithm (SDA).[30,31] The respective uncertainties for AOD and FMF stand at 0.02 and 0.1,

https://doi.org/10.1021/acs.est.4c02701
Environ. Sci. Technol. 2024, 58, 14260−14270

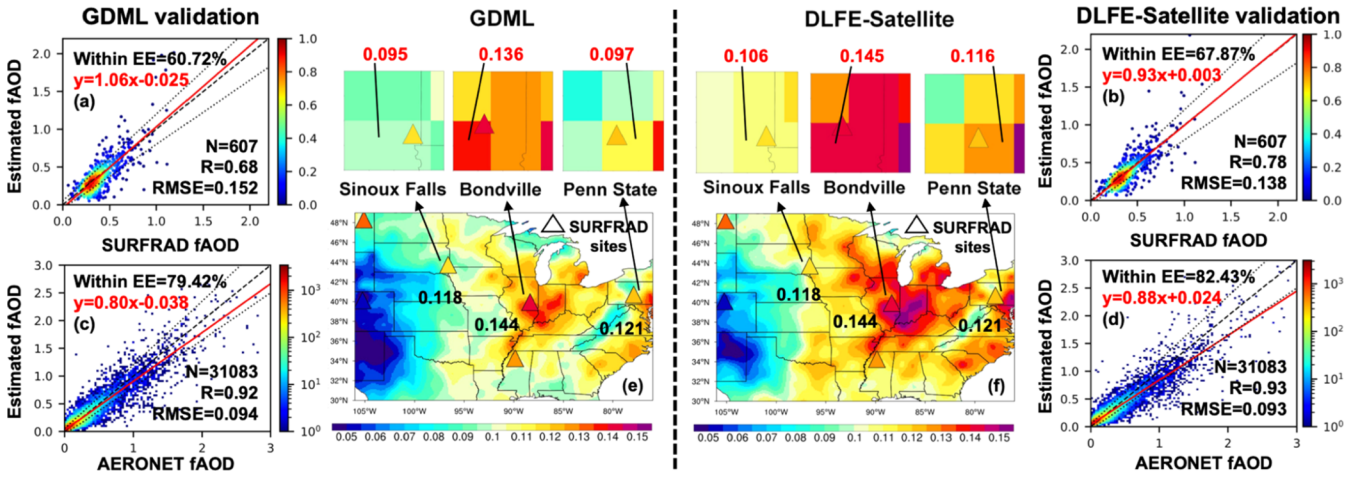


Figure 3. Validation global fAOD at SURFRAD (independent validation) and AERONET from 2001 to 2020 by GDLM and DLFE-Satellite. (a, b) Density scatterplots of GDLM and DLFE-Satellite fAOD estimate against SURFRAD fAODs for independent validation. (c, d) Density scatterplots of GDLM and DLFE-Satellite fAOD estimate against AERONET fAODs (testing data sets). The black dashed line signifies the 1:1 relationship, the red solid line represents the linear fitting line, and two black dotted lines delineate the expected error (EE) envelope of ± (0.05 + 15%). (e, f) Annual mean fAOD estimates (colored background) and SURFRAD fAODs (triangles, with black numbers indicating the observed fAOD) for the year 2009, with three zoomed-in regions showing the annual mean fAOD estimates (red numbers).

respectively[30,31] To ensure sufficient training data, we used Level 1.5 fAOD data from 2001 to 2020 for modeling.

The satellite products employed in this research comprised the MODIS C6.1 L1B MOD02SSH product, encompassing top-of-the-atmosphere reflectance from bands 1 to 7, and the MODIS C6.1 L3MOD09CMG product, including surface brightness temperature from bands 20, 21, 31, and 32, surface reflectance from bands 1 to 7, relative azimuth angle, solar zenith angle, and viewing zenith angle. These products were utilized at a spatial resolution of 5 km. Additionally, the MOD08_D3 product provided AODs at 550 nm, offering a spatial resolution of 1° × 1°.

The ERA5 reanalysis provided hourly meteorological data at a spatial resolution of 0.25°. For this study, we utilized the following parameters at 11 am local time (GMT+8), which aligns with the Terra satellite overpass time: 2-m air temperature, 2-m dew-point temperature, planetary boundary layer height (PBLH), surface pressure, and 10-m U/V wind components. These data were used to extract the fine-mode aerosol optical depth (fAOD) for the period from 2001 to 2020. Moreover, we computed relative humidity by employing the dew-point and surface temperatures.

Additionally, we incorporated a global digital elevation model (DEM) with a resolution of 250 m as Supporting Data. To ensure consistency, all the data employed for fAOD estimation underwent resampling to a spatial resolution of 1° × 1°, using the bilinear interpolation technique. Multiple AERONET stations were averaged when they all fell within the same 1° by 1° grid. Further information about these data sets is available in Table S1.

Modeling Validation and Comparison. We matched the images at a spatial resolution of 1° × 1° and a daily temporal resolution with AERONET SDA fAODs (at 0.5 μm) as input data. Site-based independent validation was implemented to assess the model's performance in estimating fAOD. The global AERONET stations were randomly partitioned into training (272 stations), validation (72 stations), and testing (78 stations) subsets (Figure S3). We partitioned the 20-year tabular data set (N = 301,226) by these stations into training (N = 184,324), validation (N = 56,909), and testing (N = 59,993) subsets for modeling purposes. Leveraging well-trained models, we generated 20-year global daily fAOD products and gauged their accuracy against AERONET fAOD retrievals. This evaluation involved key metrics such as the Pearson correlation coefficient (R), root-mean-square error (RMSE), mean absolute error (MAE), as well as the rate of change in RMSE and MAE (refer to supporting, Section S2 for further details).

To compare the performance over regions without AERONET measurement, we used daily mean fAODs from 2001 to 2020 retrieved from six Surface Radiation Budget (SURFRAD) sites (Figure S4 and Table S2). The multifilter rotating shadowband radiometer (MFRSR) provides spectral solar measurements at SURFRAD sites approximately 10 nm wide and the peak nominally at 415, 500, 614, 670, 870, and 940 nm,[32] thus it is able to apply the spectral deconvolution algorithm (SDA) method to calculate the fAOD (SURFRAD fAOD) for validation purposes. The details of the SDA method can be found in.[33] The SURFRAD sites were distant from those of AERONET and were not included in the modeling, so data from these sites could be used for independent validation purposes. The performances of the general deep-learning model (GDLM) and DLFE-Satellite models in estimating fAOD were compared in terms of overall agreement and spatial distributions, based on comparisons between estimated products and two sets of in situ measurements. Here, we employed EntityDenseNet as the GDLM, which is the same as the deep learning model 2 from the DLFE-Satellite. In this way, the comparison will show whether and how much the pretrained model can improve the fAOD estimation.

We also compared the DLFE-Satellite retrieved fAOD with three widely used satellite products from POLDER/GRASP, MISR and MODIS. For POLDER/GRASP, the fAOD at 490 nm from the "High-precision" product with 1° spatial resolution was used.[34−36] For MISR, we used MISR L3 daily data MIL3DAEN (available at https://asdc.larc.nasa.gov/data/MISR/) that contains fAOD at 550 nm with 0.5° spatial resolution.[37,38] For MODIS, although not recommended to use and excluded in the latest MODIS collection 6.1 (C6.1) product, daily FMF data at 550 nm can be obtained from

https://doi.org/10.1021/acs.est.4c02701
Environ. Sci. Technol. 2024, 58, 14260−14270

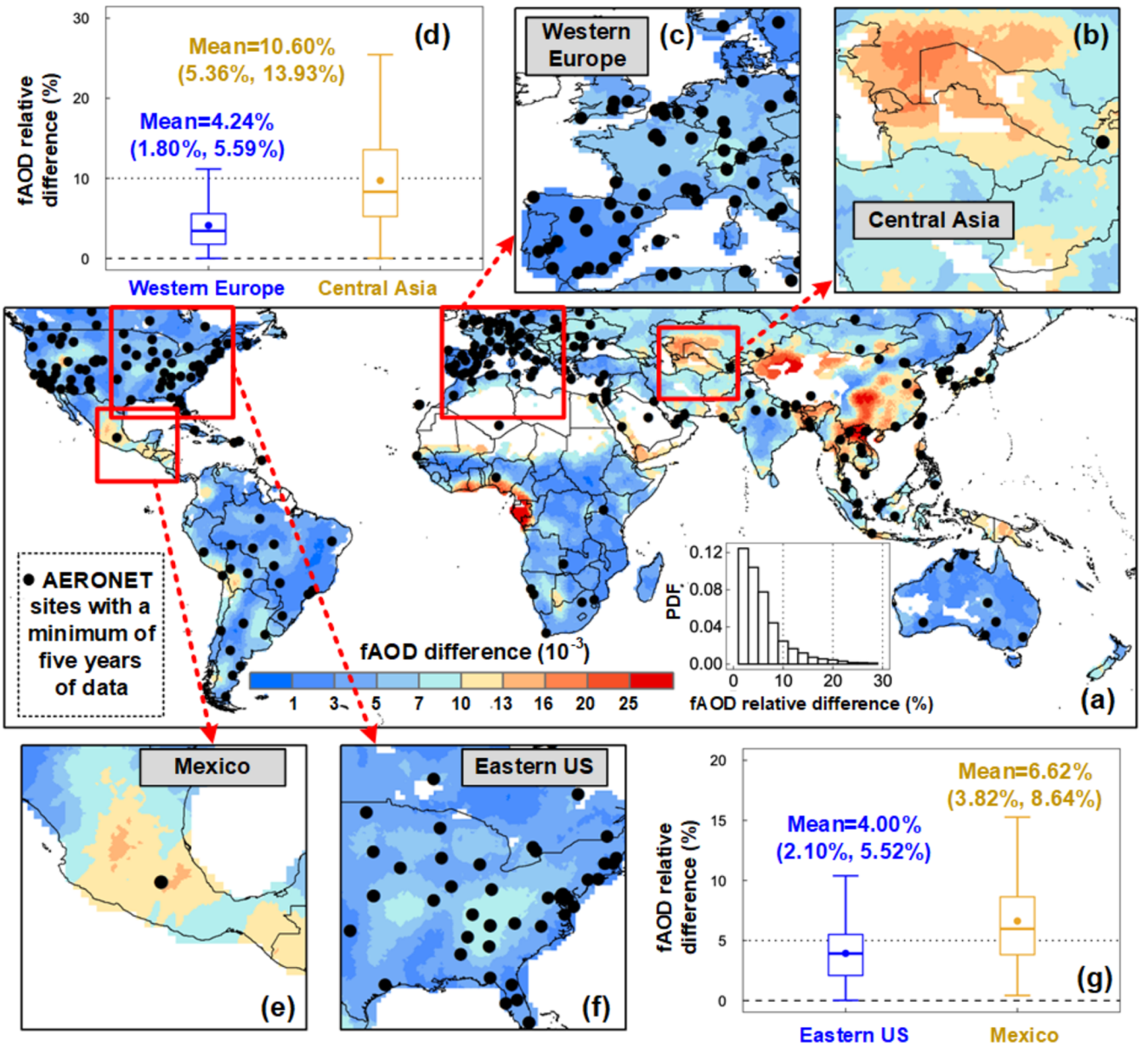


Figure 4. 20-Year global mean differences between GDLM and DLFE-Satellite fAOD estimates. (a) Distribution of 20-year global mean differences between GDLM and DLFE-Satellite fAOD estimates. Black dots indicate AERONET sites with a minimum of 5 years of data. The probability density function (PDF) inset figure shows the distribution of fAOD relative differences (%) at the global scale. (b, c) Distributions of 20-year global mean fAOD differences over Western Europe and Central Asia. (d) Boxplots of fAOD relative differences (%) over Central Asia and Western Europe. In each box, the upper, middle, and lower lines are the 75th, median, and 25th percentiles, respectively. The dot in the box represents the mean fAOD relative difference value. (e, f) Same as (b, c) but for Mexico and the Eastern US. (g) Same as (d) but for Mexico and the Eastern US.

previous MODIS C5 product at 1° spatial resolution,[39] which was used to obtain the fAOD (fAOD = FMF × AOD).

Lastly, because of the adaptability of the DLFE-Satellite framework, one can use various machine-learning models to replace the deep learning models 1 and 2 in the framework (Figure S1), according to their needs. To validate the performance of the DLFE-Satellite framework with different models, we applied two widely used machine learning models (XGBoost and LightGBM) as the Deep learning model 2 in this framework. The same input data as mentioned above for fAOD retrieval were used in this experiment.

RESULTS

Improvements by the New Deep-Learning Framework for Global fAOD. The goal of DLFE-Satellite is to fully utilize data through a pretraining process, learning latent information from the non-in situ satellite data that is ignored in the GDLM, as shown in Figure 1. This learned information from the pretrained model is then used to extract new latent features, thereby improving the fAOD retrieval accuracy (The schematic flowchart of DLFE-Satellite is illustrated in Figure S1). We performed two comprehensive validation tests for the DLFE-Satellite framework. Initially, we conducted a site-based validation using data from 78 AERONET testing stations (Figure S3) to evaluate our 20-year global fAOD product from the DLFE-Satellite model. This validation showed a strong

https://doi.org/10.1021/acs.est.4c02701
Environ. Sci. Technol. 2024, 58, 14260−14270

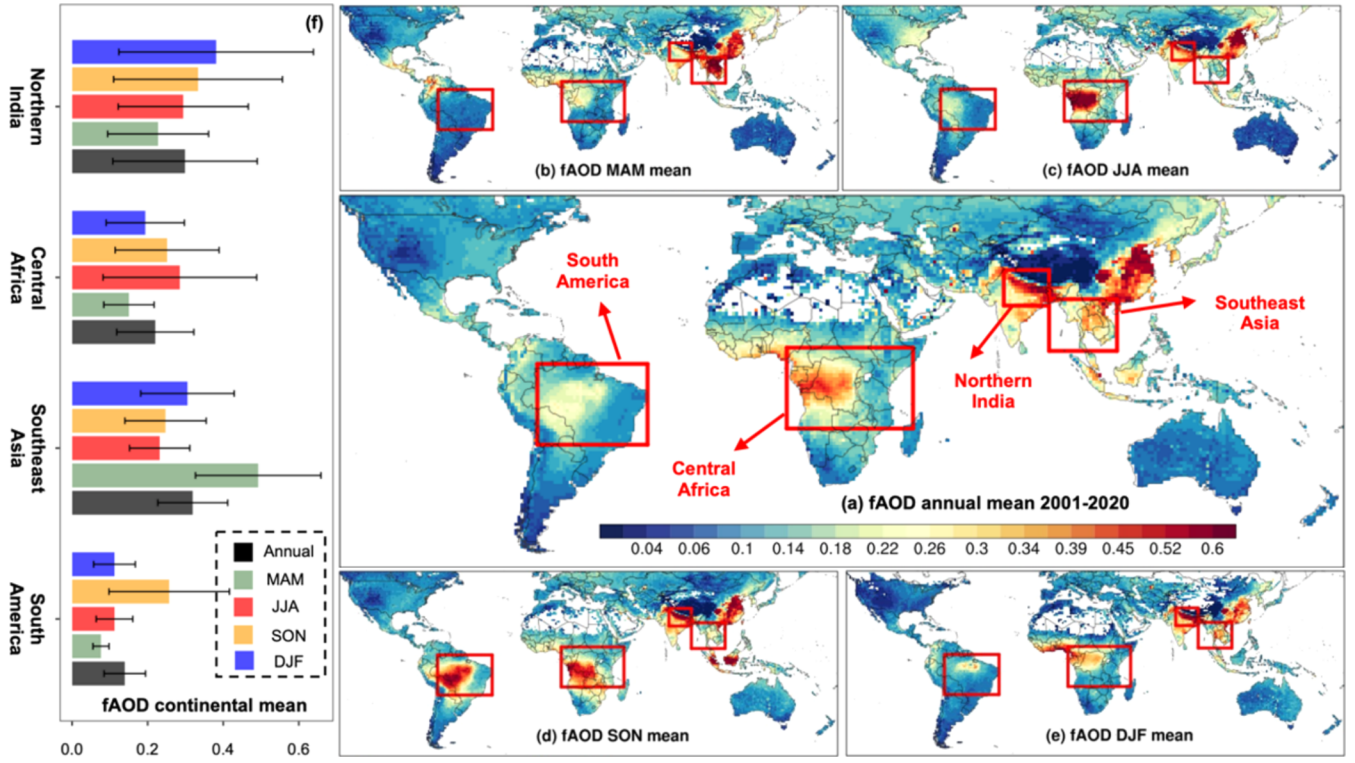


Figure 5. (a) Global spatial distribution of annual mean DLFE-Satellite fAOD in 2001−2020. The red boxes indicated four regions with clear seasonal variation of fAOD. (b−e) Same as (a) but for seasonal mean in March−April−May (MAM), June−July−August (JJA), September−October−November (SON), and December−January−February (DJF). (f) Annual and seasonal mean DLFE-Satellite fAOD averaged over four regions with clear seasonal variation of fAOD. The error bars represent the standard deviations over the space ± mean values.

correlation (R-value of 0.88) and low error rates (RMSE of 0.11 and MAE of 0.05), with 84% of retrievals falling within the expected error range ± (0.05 + 15%) (Figure S5). When compared to the GDLM fAOD retrievals, the DLFE-Satellite model exhibited a 10% improvement in R value and a 21% reduction in RMSE.

Furthermore, for a comprehensive assessment of the retrieval performance in regions without AERONET measurements, we utilized fAOD retrievals obtained from six Surface Radiation Budget Network (SURFRAD) sites in the US (Figure S4). It is worth noting that these SURFRAD sites are situated in the proximity of AERONET sites but were not employed in the validation process. As illustrated in Figure 3a,b, a comparison between the GDLM fAOD retrievals and the results obtained indicates an increase in R from 0.68 to 0.78 (a 15% improvement) and a decrease in RMSE from 0.16 to 0.14 (a reduction of 14%). Over the SURFRAD sites, GDLM-modeled fAODs (Figure 3e) were significantly underestimated over the northeastern US compared with those modeled with DLFE-Satellite (Figure 3f). Specifically, at the Sioux Falls, Bondville, and Penn State sites, GDLM-modeled fAOD values were notably underestimated compared with the DLFE-Satellite-modeled fAODs, especially at the Penn State site, where 25% of the fAOD data set values were underestimated.

To examine the contribution of variables generated by a pretrained model to the calculation of fAOD, we analyzed these variables using the feature importance within a random forest model. As illustrated in Figure S6a, some variables created by the pretrained model exhibited a stronger influence on the calculation of fAOD compared to the original variables, with one ranking second among all variables. This shows that the pretrained model effectively captured the correlations among the original variables and was able to generate new latent representations that can recover these original variables, thus having the potential to provide more general representations. Using the same input data, we rebuilt the DLFE-Satellite model with XGBoost (DLFE-XGBoost) and LightGBM (DLFE-LightGBM). In global fAOD retrievals (Figure S6b), this new pretrained framework shows that improvements can be also achieved using both DLFE-XGBoost and DLFE-LightGBM models. Especially in DLFE-XGBoost, the mean improvement of R can be achieved at 0.05.

At the global scale, significant relative differences were observed in regions with a limited number of AERONET sites (Figure 4). Relative differences were large in Mexico and Central Asia with a limited number of AERONET sites (Figure 4b,c). From 2001 to 2020, much higher relative differences were revealed over Central Asia (10.60%) than over Western Europe (4.24%) (Figure 4b,c). Similar trends were also seen when comparing relative differences between Mexico and the Eastern US (Figure 4e,f). In other regions with sparse AERONET sites, such as Bangladesh and Mississippi State, relative differences between DLFE-Satellite- and GDLM-modeled fAODs reached up to 20%.

We have also conducted a comparison among the three widely used and recognized fAOD products from MISR, MODIS, and POLDER, with DLFE-Satellite retrievals. Figure S7a−d illustrate the monthly mean fAOD product comparison with AERONET measurements from 2008−2013. We considered data points where all four fAOD products were available simultaneously at the same location and time. Overall, the DLFE-Satellite fAOD demonstrates the strongest agreement with AERONET measurements, displaying an R value of

https://doi.org/10.1021/acs.est.4c02701
Environ. Sci. Technol. 2024, 58, 14260−14270

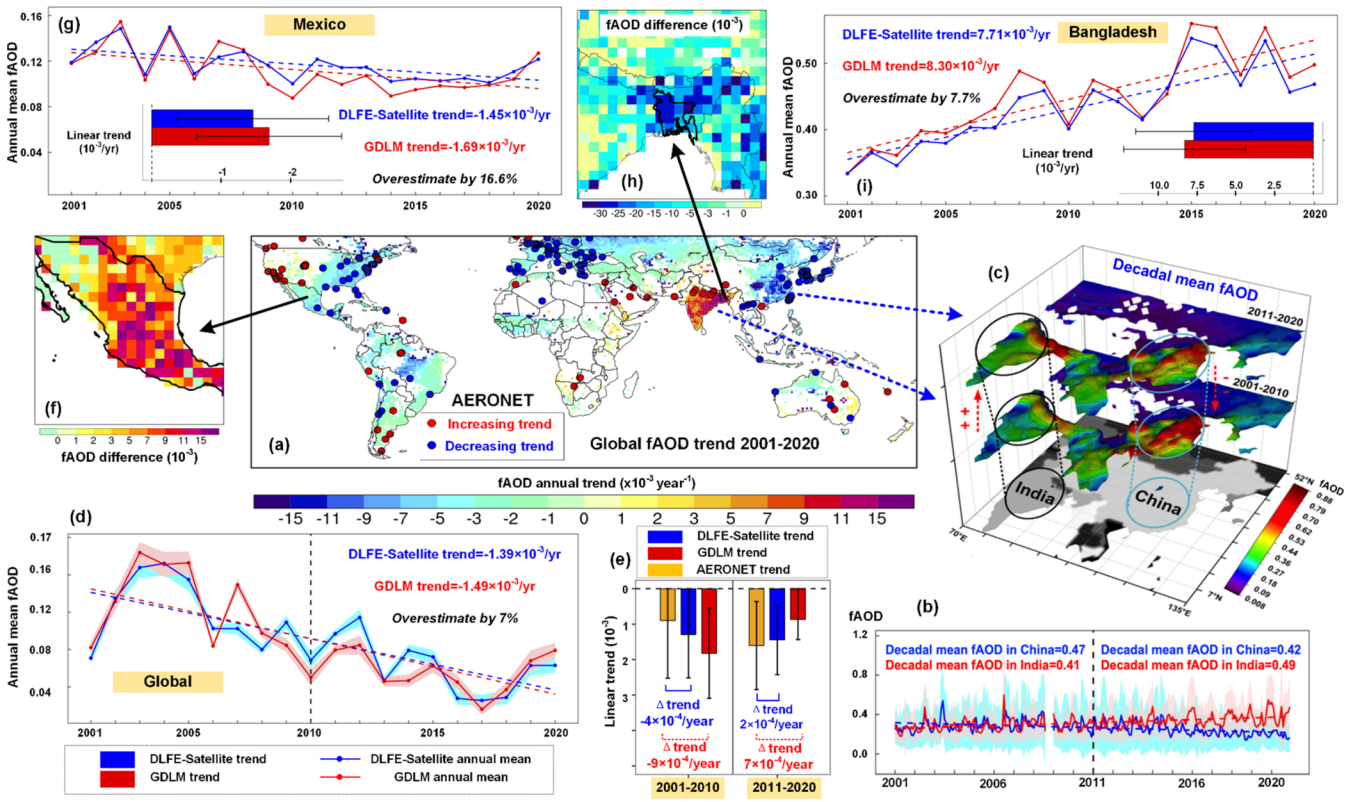


Figure 6. (a) Global DLFE-Satellite fAOD linear trends from 2001 to 2020 (at the 95% significance level). The blue and red points represent AERONET sites with decreasing and increasing linear trends at the 95% significance level. (b) Monthly mean time series of DLFE-Satellite fAOD from 2001 to 2020 in India and China. The dashed lines are the linear fitting lines, and the shaded areas represent the monthly mean fAOD ± the monthly standard deviation of fAOD. Decadal mean fAOD values from AERONET in India and China are annotated. (c) The 3D frame shows the decadal mean DLFE-Satellite fAODs averaged over the periods 2001−2010 and 2010−2020 in India and China. (d) Time series of global annual mean GDLM (red line and dots) and DLFE-Satellite (blue line and dots) fAODs from 2001 to 2020. Dashed red and blue lines represent the best-fit lines from linear regression. The shaded areas represent the likely range of annual mean fAOD ± 0.1 × (annual fAOD standard deviation). (e) Decadal linear trends of GDLM fAOD (red bars), DLFE-Satellite fAOD (blue bars), and AERONET fAOD (orange bars) from 2001 to 2010 (left side) and 2011 to 2020 (right side). Error bars represent 95% confidence intervals of the trends. (f) Differences in 20-year annual mean fAODs (DLFE-Satellite fAOD minus GDLM fAOD) over Mexico. (g) Time series of annual mean fAOD GDLM (red line and dots) and DLFE-Satellite (blue line and dots) fAODs from 2001 to 2020 over Mexico. Dashed red and blue lines represent the best-fit lines from linear regression. The inset figure shows the 20-year linear trends of GDLM fAOD (red bars) and DLFE-Satellite fAOD (blue bars), with error bars representing 95% confidence intervals of the trends. (h) Same as (f) but for Bangladesh. (i) Same as (g) but for Bangladesh.

0.95 and an RMSE of 0.04. Notably, this performance is comparable, and in some aspects, slightly superior to that of the POLDER fAOD (R = 0.82, RMSE = 0.14). By contrast, the fitting slope (0.32) for the MISR fAOD indicated an underestimation when compared to AERONET measurements, with an R value of 0.70 and an RMSE of 0.16. Meanwhile, the MODIS fAOD showed the least favorable performance (R = 0.60, RMSE = 0.23) when compared to AERONET fAOD. Therefore, it can be concluded that the overall accuracy of the DLFE-Satellite fAOD is dependable in comparison to the existing products.

Global Fine-Mode Aerosol Spatial Patterns during 2001−2020. As illustrated in Figure 5a, the DLFE-Satellite-derived fAOD data reveal a pronounced global disparity in aerosol pollution levels, identifying alarmingly high concentrations of fAOD (>0.45) particularly in eastern China and northern India. Such elevated levels are primarily attributed to the intense industrial and residential emissions prevalent in these regions, thus underscoring the profound environmental repercussions of human activities.[40−42] Similarly severe conditions in central Africa and the Amazon regions, where fAOD values exceed 0.3. This is largely due to widespread biomass burning and forest fires, which emit substantial amounts of black and organic carbon into the atmosphere.[43] By contrast, regions like central Asia, Australia, southern South America, and the western US report lower fAOD (<0.1), indicating relatively minor fine mode aerosol level. In assessing the reliability of the DLFE-Satellite-derived fAOD, we embarked on a meticulous comparison with the POLDER fAOD (only available from 2008 to 2013) to evaluate the spatial congruence and magnitude of fAOD captured by both data sets. As illustrated in Figure S9e,f, both the DLFE-Satellite and POLDER data sets proficiently identify regions of high aerosol concentration (>0.3) across Eastern China and Central Africa, as well as zones of minimal aerosol presence (<0.1) in Central Asia and the Western US.

Figure 5b−e depicts the global spatial distribution of the mean seasonal fractional fAOD for the period 2001−2020, revealing significant seasonal variations and unique regional disparities. A noteworthy observation is the dramatic increase in fAOD levels in Southeast Asia during the spring, where values surge above 0.5. This starkly contrasts with other seasons, where fAOD levels remain under 0.32. The primary driver behind this substantial springtime spike is the region's

https://doi.org/10.1021/acs.est.4c02701
Environ. Sci. Technol. 2024, 58, 14260−14270

intense period of biomass burning from January to April, which propels fAOD values to these exceptionally high levels.[44] Central Africa experiences a notable rise in fAOD during the summer and autumn months, with values surpassing 0.7, distinctly exceeding the levels observed in other seasons, which remain under 0.40. This pattern correlates with the seasonally aligned peaks in local biomass burning activities.[45] In South America, characterized by its intense biomass burning from summer through autumn,[43,46] the highest of fAOD values is reached in autumn, with measurements exceeding 0.60. This contrasts starkly with the significantly lower spring values, below 0.20, showcasing the dramatic seasonal variances. In Northern India, a unique scenario unfolds where fAOD peaks during the autumn and winter months. This is attributed to the lowering of boundary layer heights and unfavorable atmospheric dispersion conditions during these months.[47] Such environmental factors lead to the significant accumulation of anthropogenic emissions, resulting in the observed high fAOD levels.

Unveiling Global and Regional Trends from Derived fAOD over the Past Two Decades. As for long-term fAOD trends, Figure 6a shows global DLFE-Satellite and AERONET fAOD linear trends from 2001 to 2020. In general, both data sets show significant decreasing trends (p-value <0.05) in fAOD over the eastern US, the Amazon rainforest, Europe, and eastern China. At each AERONET station, the DLFE-Satellite and AERONET fAOD trends agreed well with 95.5% of stations showing the same trend signs, indicating that the DLFE-Satellite fAOD captured reliable fAOD trends (Figure S8). However, recent studies have identified an increasing trend for surface $PM_{2.5}$ over central eastern China from 1998 to 2018.[48] This inconsistency is attributed to a notable opposing trend in aerosols observed between low and high altitudes over China: vertical distribution of trends in AOD appeared to be increasing near the surface (low altitude), whereas AOD at high altitudes indicated a downward tendency.[14] Strong downward trends in fAOD over eastern China, suggesting transported pollution into the western US via westerly flow, do not play a leading role in the positive fAOD trend over the western US. In the western US, aerosols are known to demonstrate a positive trend because of smoke from frequent wildfires.[49]

The most notable surge in fAOD has been observed in India (7.86×10^{-4}/year), where a striking decadal increase from 0.41 to 0.49, particularly in the northern (escalating from <0.44 to >0.53) and central regions (rising from <0.27 to ∼0.36), as illustrated in Figure 6b,c. This increase aligns with AERONET data, which confirms a growth in mean fAOD from 0.41 to 0.47 (Figure S9), primarily driven by escalating anthropogenic emissions from heightened energy consumption, agricultural residue burning, and transportation activities.[40,41,50] Conversely, China has seen a significant reduction in fAOD, particularly in the North China Plain where fAOD decreased from >0.79 to <0.70, complemented by a notable decline of 0.07 in AERONET fAODs (Figure S10). This decrease is largely attributed to effective environmental policies targeting industrial and vehicular emissions,[51] marking a contrasting trend in regional air quality dynamics. Compared to the global mean fAOD decreasing trend (-1.39×10^{-3}/year), the annual mean trend of fAOD in China is 3 times higher (-5.07×10^{-3}/year), reflecting effective control of anthropogenic emissions.

Comparing DLFE-Satellite with GDLM results for global fAOD trends analysis, although both results show significant decreasing trends, the GDLM model overestimated the global declining trend for fAODs by 7% per year compared to the DLFE-Satellite model (Figure 6d). Figure 6e shows that in the second decade (2011 to 2020), the difference in trends between GDLM and ground-based AERONET (Δtrend = 7×10^{-4}/yr) was 3 times more than that between DLFE-Satellite and ground-based AERONET (Δtrend = 2×10^{-4}/yr). At the national scale of Mexico, GMDL-modeled fAODs were generally lower than those modeled with DLFE-Satellite (Figure 6f). The decrease in fAOD over time was captured by both modeled sets (Figure 6g), corresponding well to the reduction in air pollutants in Mexico due to the establishment of the Environmental State Plan 1995−2020 (Plan Estatal de Medio Ambiente 1995−2020).[52,53] The GDLM model overestimated the declining trend over Mexico by 16.6% per year compared with that estimated by the DLFE-Satellite model (-1.45×10^{-3}/yr). GMDL-modeled fAODs over Bangladesh were generally higher than DLFE-Satellite-modeled values (Figure 6h). As one of the world's most polluted countries, Bangladesh has experienced a significant increase in air pollution.[54,55] This was captured by the times series of GMDL- and DLFE-Satellite-modeled annual mean fAODS (Figure 6i), which showed increases in fine-mode aerosols from 2001 to 2020. The GDLM model overestimated the increasing fAOD trend by 7.7% per year compared with that estimated with the DLFE-Satellite model (7.71×10^{-3}/yr). The MISR fAOD trend also shows a decreasing trend similar to the DLFE-Satellite fAOD (Figure S10). However, the magnitude (3.18×10^{-4}/yr) is smaller than the trend of DLFE-Satellite fAOD (1.39×10^{-3}/yr). This is attributed to the underestimation of MISR fAOD compared to the DLFE-Satellite fAOD and AERONET, emphasizing that an accurate estimation of fAOD is important for determining both the sign and magnitude of the global fAOD trend.

DISCUSSION

The accuracy of satellite-based fAOD modeling is constrained by a lack of in situ sites. The number of ground-based monitoring sites in networks such as the AERONET[29] is still limited and sparse in underdeveloped and developing countries with imbalanced distribution. To address this issue, we proposed a new pretrained framework for the DLFE-Satellite model (Figure 2) to fully exploit the latent information in unused satellite data during fAOD modeling, especially for regions with sparse in situ sites. Satellite measurements from pixels without monitoring sites were used for model pretraining to create new latent features. These features aim to capture the latent correlation that is ignored by traditional machine-learning models for satellite-based remote sensing, providing rich information to support a robust and generalizable model. By taking advantage of latent features, the DLFE-Satellite model expands the amount of training data in areas with sparse or no in situ sites, leading to improvements in the fitting efficiency and generating reliable spatial patterns for satellite-based remote sensing.

The DLFE-Satellite model is different from other such models in two important ways. First, predictions with the DLFE-Satellite model employ all imagery information rather than using only those pixels with in situ sites, as done in previous studies.[56−58] This allows the full use of rich information, enhancing the model's adaptability and predictive

https://doi.org/10.1021/acs.est.4c02701
Environ. Sci. Technol. 2024, 58, 14260−14270

accuracy. Second, the DLFE-Satellite model also works flexibly to fit the needs of various models for a range of satellite-based retrieval tasks. Because machine learning models generally rely on the integration of satellite and in situ data for retrievals, disparities can emerge among various machine learning models.

On a global scale, we utilized the DLFE-Satellite model to estimate fAOD (at 0.5 μm) from 2001 to 2020. The employment of the DLFE-Satellite model for fAOD estimation led to a substantial enhancement in the performance of site-based independent validation, with an increase in the correlation coefficient (R) from 0.80 to 0.88, as compared to the utilization of the GDML model. Additionally, over regions lacking AERONET measurement, the DLFE-Satellite model significantly improved the agreement with SURFRAD fAOD, demonstrating an increase in the correlation coefficient (R) from 0.68 to 0.78 (a 15% enhancement). Consequently, we have confidence in the reliability and comparability of the results based on the DLFE-Satellite fAOD when compared to existing products. Owing to the DLFE-Satellite model, the improved fAOD data has brought us a more accurate trend analysis, which is closer to the trend by AERONET stations, reducing the error (Δtrend = 2×10^{-4}/yr) by nearly 3.5 times compared to the GDLM (Δtrend = 7×10^{-4}/yr). This is important for monitoring the effectiveness of pollution control measures over time, which enables policymakers to adjust strategies as needed and track progress toward air quality goals more accurately.

The global mean fAOD trend is -1.39×10^{-3}/year indicating a general decrease. The factors contributing to this global decrease in fine mode aerosols differ across regions but are primarily linked to the enactment of air quality regulations that impose stringent emissions standards on industries and vehicles, thereby enhancing air quality. Furthermore, technological advancements have resulted in more efficient combustion mechanisms within vehicles and industrial settings,[41] coupled with a transition from coal to more environmentally friendly energy sources.[40] The rate of decline in China's fine mode aerosols (-5.07×10^{-3}/year) has reached 3 times that of the global decrease rate, demonstrating the success of policy regulation. Although India also has implemented air quality regulations, there are obviously some issues with enforcement, as the decadal mean fAOD in India increased substantially from 0.41 to 0.49. This research developed an innovative model DLFE-Satellite to fAOD estimation through the use of latent feature extraction from ignored satellite data, particularly benefiting areas with sparse ground monitoring. This method improves fAOD modeling precision and offers a more accurate depiction of global fAOD trends, supporting targeted environmental policy adjustments.

■ ASSOCIATED CONTENT

Data Availability Statement

The global land fAOD data set (2001−2020) developed by DLFE-Satellite is available at 10.5281/zenodo.6972774. The fAOD data are in the Geotiff format on a daily scale.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.4c02701.

Present global fAOD studies (Section 1); evaluation metrics (Section 2); data used for estimating fAOD (Table S1); SURFRAD sites information (Table S2); the schematic flowchart of DLFE-Satellite (Figure S1); The self-supervised learning in pretraining step of DLFE-Satellite (Figure S2); map of stations for site-based independent validation (Figure S3); AERONET and SURFRAD sites used in this study (Figure S4); validations of GDLM and DLFE-Satellite fAOD from 78 testing stations (Figure S5); exploring the correlation of DLFE-satellite generated features with fAOD and the model's generalizability across various machine models (Figure S6); validations ofDLFE-Satellite, MODIS, MISR and POLDER (Figure S7); scatter plot between DLFE-Satellite fAOD and AERONET fAOD trends (Figure S8); the decadal mean AERONET fAOD in India and China (Figure S9); global monthly mean time series of MISR fAOD from 2001 to 2020 (Figure S10) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Zhou Zang − *State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China;* orcid.org/0000-0002-8486-3246; Email: joey.zang@mail.utoronto.ca

Zhanqing Li − *Department of Atmospheric and Oceanic Science and ESSIC, University of Maryland, College Park, Maryland 20740, United States;* orcid.org/0000-0001-6737-382X; Email: zli@atmos.umd.edu

Authors

Xing Yan − *State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China;* orcid.org/0000-0001-9327-5756

Hans W. Chen − *Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg 41296, Sweden*

Jiayi Chen − *State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*

Yize Jiang − *State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*

Yunhao Chen − *State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*

Bin He − *State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*

Chen Zuo − *State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*

Terry Nakajima − *Tokyo University of Marine Science and Technology, Tokyo 108-8477, Japan*

Jhoon Kim − *Department of Atmospheric Sciences, Yonsei University, Seoul 03722, South Korea*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c02701

Author Contributions

X.Y. and Y.J. designed the research. Z.Z. and C.Z. collected and analyzed the data. Z. L. supervised the project. X.Y. and Z.Z. wrote the manuscript with support from H.W.C., Y.C., B.H.,

https://doi.org/10.1021/acs.est.4c02701
Environ. Sci. Technol. 2024, 58, 14260−14270

T.N., and J.K. All authors discussed the results and contributed to the interpretation of the final manuscript.

Notes

The authors declare no competing financial interest.

Code availability The codes of DLFE-Satellite and its user guide are freely available online at 10.5281/zenodo.13139532. The codes of EntityDenseNet is freely available online at https://github.com/RegiusQuant/ESIDLM, and the user guide is at https://esidlm-tutorial.readthedocs.io/en/latest/.

■ ACKNOWLEDGMENTS

This research has been supported by the National Natural Science Foundation of China (grant nos. 42030606 and 42192580), the Natural Science Foundation of Beijing (grant nos. 8222058), and the Fundamental Research Funds for the Central Universities. The authors gratefully acknowledge the European Centre for Medium Range Weather Forecasts, MODIS, SURFRAD and AERONET teams for their effort in making the data available.

■ REFERENCES

(1) 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* (7), 3762−3772.

(2) Wang, Y.; Yuan, Q.; Li, T.; Zhu, L.; Zhang, L. Estimating daily full-coverage near surface O3, CO, and NO2 concentrations at a high spatial resolution over China based on S5P-TROPOMI and GEOS-FP. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 311−325.

(3) Yang, Q.; Yuan, Q.; Yue, L.; Li, T.; Shen, H.; Zhang, L. Mapping PM2. 5 concentration at a sub-km level resolution: A dual-scale retrieval approach. *ISPRS J. Photogramm. Remote Sens.* **2020**, *165*, 140−151.

(4) Zang, Z.; Li, D.; Guo, Y.; Shi, W.; Yan, X. Superior PM2. 5 estimation by integrating aerosol fine mode data from the Himawari-8 satellite in deep and classical machine learning models. *Remote Sens.* **2021**, *13* (14), No. 2779.

(5) Bai, K.; Li, K.; Ma, M.; Li, K.; Li, Z.; Guo, J.; Chang, N.-B.; Tan, Z.; Han, D. LGHAP: the Long-term Gap-free High-resolution Air Pollutant concentration dataset, derived via tensor-flow-based multi-modal data fusion. *Earth Syst. Sci. Data* **2022**, *14* (2), 907−927.

(6) Crow, W. T.; Berg, A. A.; Cosh, M. H.; Loew, A.; Mohanty, B. P.; Panciera, R.; de Rosnay, P.; Ryu, D.; Walker, J. P. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products *Rev. Geophys.* 2012; Vol. *50* 2 DOI: 10.1029/2011RG000372.

(7) Lin, C.; Li, Y.; Yuan, Z.; Lau, A. K.; Li, C.; Fung, J. C. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM2. 5. *Remote Sens. Environ.* **2015**, *156*, 117−128.

(8) Xu, S.; Zou, B.; Lin, Y.; Zhao, X.; Li, S.; Hu, C. Strategies of method selection for fine-scale PM 2.5 mapping in an intra-urban area using crowdsourced monitoring. *Atmos. Meas. Tech.* **2019**, *12* (5), 2933−2948.

(9) Emili, E.; Popp, C.; Wunderle, S.; Zebisch, M.; Petitta, M. Mapping particulate matter in alpine regions with satellite and ground-based measurements: An exploratory study for data assimilation. *Atmos. Environ.* **2011**, *45* (26), 4344−4353.

(10) Yang, J.; Zhou, A.; Han, L.; Li, Y.; Xie, Y. Monitoring urban black-odorous water by using hyperspectral data and machine learning. *Environ. Pollut.* **2021**, *269*, No. 116166.

(11) Yan, X.; Liang, C.; Jiang, Y.; Luo, N.; Zang, Z.; Li, Z. A deep learning approach to improve the retrieval of temperature and humidity profiles from a ground-based microwave radiometer. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58* (12), 8427−8437.

(12) Yan, X.; Zang, Z.; Jiang, Y.; Shi, W.; Guo, Y.; Li, D.; Zhao, C.; Husi, L. A Spatial-Temporal Interpretable Deep Learning Model for improving interpretability and predictive accuracy of satellite-based PM2. 5. *Environ. Pollut.* **2021**, *273*, No. 116459.

(13) Zang, Z.; Guo, Y.; Jiang, Y.; Zuo, C.; Li, D.; Shi, W.; Yan, X. Tree-based ensemble deep learning model for spatiotemporal surface ozone (O3) prediction and interpretation. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, No. 102516.

(14) Dong, L.; Li, S.; Yang, J.; Shi, W.; Zhang, L. Investigating the performance of satellite-based models in estimating the surface PM2. 5 over China. *Chemosphere* **2020**, *256*, No. 127051.

(15) Li, T.; Shen, H.; Yuan, Q.; Zhang, L. A Locally Weighted Neural Network Constrained by Global Training for Remote Sensing Estimation of $PM_{2.5}$. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1−13.

(16) Wang, P.; Tang, Q.; Zhu, Y.; He, Y.; Yu, Q.; Liang, T.; Zheng, K. Spatial-Temporal Variation of AOD Based on MAIAC AOD in East Asia from 2011 to 2020. *Atmosphere* **2022**, *13* (12), No. 1983.

(17) Wang, Z.; Hu, B.; Huang, B.; Ma, Z.; Biswas, A.; Jiang, Y.; Shi, Z. Predicting annual PM2. 5 in mainland China from 2014 to 2020 using multi temporal satellite product: An improved deep learning approach with spatial generalization ability. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 141−158.

(18) Zhu, X.; Zhang, Q.; Xu, C.-Y.; Sun, P.; Hu, P. Reconstruction of high spatial resolution surface air temperature data across China: A new geo-intelligent multisource data-based machine learning technique. *Sci. Total Environ.* **2019**, *665*, 300−313.

(19) Evans, S.; Williams, G. P.; Jones, N. L.; Ames, D. P.; Nelson, E. J. Exploiting earth observation data to impute groundwater level measurements with an extreme learning machine. *Remote Sens.* **2020**, *12* (12), No. 2044.

(20) Liu, Z.; Gao, W.; Yu, Y.; Hu, B.; Xin, J.; Sun, Y.; Wang, L.; Wang, G.; Bi, X.; Zhang, G.; et al. Characteristics of PM 2.5 mass concentrations and chemical species in urban and background areas of China: Emerging results from the CARE-China network. *Atmos. Chem. Phys.* **2018**, *18* (12), 8849−8871.

(21) Han, X.; Zhang, Z.; Ding, N.; Gu, Y.; Liu, X.; Huo, Y.; Qiu, J.; Yao, Y.; Zhang, A.; Zhang, L.; et al. Pre-trained models: Past, present and future. *AI Open* **2021**, *2*, 225−250.

(22) Wang, X.; Chen, G.; Qian, G.; Gao, P.; Wei, X.-Y.; Wang, Y.; Tian, Y.; Gao, W. Large-scale multi-modal pre-trained models: A comprehensive survey. *Mach. Intell. Res.* **2023**, *20*, 447−482.

(23) Nori, H.; King, N.; McKinney, S. M.; Carignan, D.; Horvitz, E. Capabilities of gpt-4 on Medical Challenge Problems. 2023, arXiv:2303.13375. arXiv.org e-Printarchive. https://arxiv.org/abs/2303.13375.

(24) Yoon, J.; Zhang, Y.; Jordon, J.; van der Schaar, M. In *Vime: Extending the Success of Self-and Semi-Supervised Learning to Tabular Domain*, Advances in Neural Information Processing Systems 33; NeurIPS, 2020; pp 11033−11043.

(25) Yan, X.; Zang, Z.; Luo, N.; Jiang, Y.; Li, Z. New interpretable deep learning model to monitor real-time PM2. 5 concentrations from satellite data. *Environ. Int.* **2020**, *144*, No. 106060.

(26) Ioffe, S.; Szegedy, C. In *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, International Conference on Machine Learning; PMLR, 2015; pp 448−456.

(27) Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. learn. Res.* **2014**, *15* (1), 1929−1958.

(28) Guo, C.; Berkhahn, F. Entity Embeddings of Categorical Variables. 2016, arXiv:1604.06737. arXiv.org e-Printarchive. https://arxiv.org/abs/1604.06737.

(29) Holben, B. N.; Eck, T. F.; Slutsker, I.; Tanré, D.; Buis, J.; Setzer, A.; Vermote, E.; Reagan, J. A.; Kaufman, Y.; Nakajima, T.; et al. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sens. Environ.* **1998**, *66* (1), 1−16.

(30) O'Neill, N. T.; Dubovik, O.; Eck, T. F. Modified Ångström exponent for the characterization of submicrometer aerosols. *Appl. Opt.* **2001**, *40* (15), 2368−2375.

(31) O'neill, N.; Eck, T.; Smirnov, A.; Holben, B.; Thulasiraman, S. Spectral discrimination of coarse and fine mode optical depth *J. Geophys. Res.: Atmos.* 2003; Vol. *108* D17.
(32) Harrison, L.; Michalsky, J.; Berndt, J. Automated multifilter rotating shadow-band radiometer: an instrument for optical depth and radiation measurements. *Appl. Opt.* **1994**, *33* (22), 5118−5125.
(33) O'Neill, N.; Eck, T.; Smirnov, A.; Holben, B. Spectral deconvolution algorithm (SDA) technical memo *NASA Technol. Memo* 2008.
(34) Dubovik, O.; Lapyonok, T.; Litvinov, P.; Herman, M.; Fuertes, D.; Ducos, F.; Lopatin, A.; Chaikovsky, A.; Torres, B.; Derimian, Y.et al.GRASP: a versatile algorithm for characterizing the atmosphere *SPIE Newsroom* 2014; Vol. *25* .
(35) Dubovik, O.; Li, Z.; Mishchenko, M. I.; Tanré, D.; Karol, Y.; Bojkov, B.; Cairns, B.; Diner, D. J.; Espinosa, W. R.; Goloub, P.; et al. Polarimetric remote sensing of atmospheric aerosols: Instruments, methodologies, results, and perspectives. *J. Quant. Spectrosc. Radiat. Transfer* **2019**, *224*, 474−511.
(36) Chen, C.; Dubovik, O.; Fuertes, D.; Litvinov, P.; Lapyonok, T.; Lopatin, A.; Ducos, F.; Derimian, Y.; Herman, M.; Tanré, D.; et al. Validation of GRASP algorithm product from POLDER/PARASOL data and assessment of multi-angular polarimetry potential for aerosol monitoring. *Earth Syst. Sci. Data* **2020**, *12*, 3573−3620.
(37) Kahn, R. A.; Gaitley, B. J. An analysis of global aerosol type as retrieved by MISR. *J. Geophys. Res.: Atmos.* **2015**, *120* (9), 4248−4281.
(38) Garay, M. J.; Witek, M. L.; Kahn, R. A.; Seidel, F. C.; Limbacher, J. A.; Bull, M. A.; Diner, D. J.; Hansen, E. G.; Kalashnikova, O. V.; Lee, H.; et al. Introducing the 4.4 km spatial resolution Multi-Angle Imaging SpectroRadiometer (MISR) aerosol product. *Atmos. Meas. Tech.* **2020**, *13* (2), 593−628.
(39) 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; Vol. *112* D13 DOI: 10.1029/2006JD007811.
(40) Jethva, H.; Torres, O.; Field, R. D.; Lyapustin, A.; Gautam, R.; Kayetha, V. Connecting crop productivity, residue fires, and air quality over northern India. *Sci. Rep.* **2019**, *9* (1), No. 16594.
(41) Manoj, M. R.; Satheesh, S.; Moorthy, K. K.; Gogoi, M. M.; Babu, S. S. Decreasing trend in black carbon aerosols over the Indian region. *Geophys. Res. Lett.* **2019**, *46* (5), 2903−2910.
(42) Zheng, B.; Chevallier, F.; Yin, Y.; Ciais, P.; Fortems-Cheiney, A.; Deeter, M. N.; Parker, R. J.; Wang, Y.; Worden, H. M.; Zhao, Y. Global atmospheric carbon monoxide budget 2000−2017 inferred from multi-species atmospheric inversions. *Earth Syst. Sci. Data* **2019**, *11* (3), 1411−1436.
(43) Pérez-Ramírez, D.; Andrade-Flores, M.; Eck, T. F.; Stein, A. F.; O'Neill, N. T.; O'Neill, N. T.; Lyamani, H.; Gassó, S.; Whiteman, D. N.; Veselovskii, I.; Velarde, F. Multi year aerosol characterization in the tropical Andes and in adjacent Amazonia using AERONET measurements. *Atmos. Environ.* **2017**, *166*, 412−432.
(44) Yin, S.; Wang, X.; Zhang, X.; Guo, M.; Miura, M.; Xiao, Y. Influence of biomass burning on local air pollution in mainland Southeast Asia from 2001 to 2016. *Environ. Pollut.* **2019**, *254*, No. 112949.
(45) Jiang, Y.; Zhou, L.; Raghavendra, A. Observed changes in fire patterns and possible drivers over Central Africa. *Environ. Res. Lett.* **2020**, *15* (9), No. 0940b8.
(46) Generoso, S.; Bréon, F.-M.; Balkanski, Y.; Boucher, O.; Schulz, M. Improving the seasonal cycle and interannual variations of biomass burning aerosol sources. *Atmos. Chem. Phys.* **2003**, 3 (4), 1211−1222.
(47) Ramachandran, S. Aerosol optical depth and fine mode fraction variations deduced from Moderate Resolution Imaging Spectroradiometer (MODIS) over four urban areas in India *J. Geophys. Res.: Atmos.* 2007; Vol. *112* D16 DOI: 10.1029/2007JD008500.
(48) Hammer, M. S.; van Donkelaar, A.; Li, C.; Lyapustin, A.; Sayer, A. M.; Hsu, N. C.; Levy, R. C.; Garay, M. J.; Kalashnikova, O. V.; Kahn, R. A.; et al. Global estimates and long-term trends of fine particulate matter concentrations (1998−2018). *Environ. Sci. Technol.* **2020**, *54* (13), 7879−7890.
(49) Buchholz, R. R.; Park, M.; Worden, H. M.; Tang, W.; Edwards, D. P.; Gaubert, B.; Deeter, M. N.; Sullivan, T.; Ru, M.; Chin, M.; et al. New seasonal pattern of pollution emerges from changing North American wildfires. *Nat. Commun.* **2022**, *13* (1), No. 2043.
(50) Venkataraman, C.; Brauer, M.; Tibrewal, K.; Sadavarte, P.; Ma, Q.; Cohen, A.; Chaliyakunnel, S.; Frostad, J.; Klimont, Z.; Martin, R. V. Source influence on emission pathways and ambient PM 2.5 pollution over India (2015−2050). *Atmos. Chem. Phys.* **2018**, *18* (11), 8017−8039.
(51) Yang, X.; Jiang, L.; Zhao, W.; Xiong, Q.; Zhao, W.; Yan, X. Comparison of ground-based PM2. 5 and PM10 concentrations in China, India, and the US. *Int. J. Environ. Res. Public Health* **2018**, *15* (7), No. 1382.
(52) Aguirre-López, M. A.; Rodríguez-González, M. A.; Soto-Villalobos, R.; Gómez-Sánchez, L. E.; Benavides-Ríos, Á. G.; Benavides-Bravo, F. G.; Walle-García, O.; Pamanés-Aguilar, M. G. Statistical Analysis of PM10 Concentration in the Monterrey Metropolitan Area, Mexico (2010−2018). *Atmosphere* **2022**, *13* (2), No. 297.
(53) Ipiña, A.; López-Padilla, G.; Retama, A.; Piacentini, R. D.; Madronich, S. Ultraviolet radiation environment of a tropical megacity in transition: Mexico City 2000−2019. *Environ. Sci. Technol.* **2021**, *55* (16), 10946−10956.
(54) Pavel, M. R. S.; Zaman, S. U.; Jeba, F.; Islam, M. S.; Salam, A. Long-Term (2003−2019) Air quality, climate variables, and human health consequences in Dhaka, Bangladesh. *Front. Sustainable Cities* **2021**, *52*, No. 681759, DOI: 10.3389/frsc.2021.681759.
(55) Shi, Y.; Matsunaga, T.; Yamaguchi, Y.; Li, Z.; Gu, X.; Chen, X. Long-term trends and spatial patterns of satellite-retrieved PM2. 5 concentrations in South and Southeast Asia from 1999 to 2014. *Sci. Total Environ.* **2018**, *615*, 177−186.
(56) Gupta, P.; Christopher, S. A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach *J. Geophys. Res.: Atmos.* 2009; Vol. *114* D14 DOI: 10.1029/2008JD011496.
(57) Li, T.; Shen, H.; Yuan, Q.; Zhang, X.; Zhang, L. Estimating ground-level PM2. 5 by fusing satellite and station observations: a geo-intelligent deep learning approach. *Geophys. Res. Lett.* **2017**, *44* (23), 11985−11993.
(58) Yan, X.; Zang, Z.; Li, Z.; Luo, N.; Zuo, C.; Jiang, Y.; Li, D.; Guo, Y.; Zhao, W.; Shi, W.; Cribb, M. A global land aerosol fine-mode fraction dataset (2001−2020) retrieved from MODIS using hybrid physical and deep learning approaches. *Earth Syst. Sci. Data* **2022**, *14*, 1193−1213.