

Article

Wide and Deep Learning Model for Satellite-Based Real-Time Aerosol Retrievals in China

Nana Luo ¹, Junxiao Zou ², Zhou Zang ³, Tianyi Chen ¹ and Xing Yan ^{2,*} 

¹ School of Geomatics and Urban Information, Beijing University of Civil Engineering and Architecture, Beijing 102616, China; luonana@bucea.edu.cn (N.L.)

² Innovation Research Center of Satellite Application (IRCSA), Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

³ Department of Geography and Planning, University of Toronto, Toronto, ON M5S 1A4, Canada; joey.zang@mail.utoronto.ca

* Correspondence: yanxing@bnu.edu.cn

Abstract: Machine learning methods have been recognized as rapid methods for satellite-based aerosol retrievals but have not been widely applied in geostationary satellites. In this study, we developed a wide and deep learning model to retrieve the aerosol optical depth (AOD) using Himawari-8. Compared to traditional deep learning methods, we embedded a “wide” modeling component and tested the proposed model across China using independent training (2016–2018) and test (2019) datasets. The results showed that the “wide” model improves the accuracy and enhances model interpretability. The estimates exhibited better accuracy ($R^2 = 0.81$, root-mean-square errors (RMSEs) = 0.19, and within the estimated error (EE) = 63%) than those of the deep-only models ($R^2 = 0.78$, RMSE = 0.21, within the EE = 58%). In comparison with extreme gradient boosting (XGBoost) and Himawari-8 V2.1 AOD products, there were also significant improvements. In addition to higher accuracy, the interpretability of the proposed model was superior to that of the deep-only model. Compared with other seasons, higher contributions of spring to the AOD concentrations were interpreted. Based on the application of the wide and deep learning model, the near-real-time variation of the AOD over China could be captured with an ultrafine temporal resolution.

Keywords: Himawari-8; deep learning; AOD



Citation: Luo, N.; Zou, J.; Zang, Z.; Chen, T.; Yan, X. Wide and Deep Learning Model for Satellite-Based Real-Time Aerosol Retrievals in China.

Atmosphere **2024**, *15*, 564. <https://doi.org/10.3390/atmos15050564>

Academic Editors: Alessia Sannino, Alejandro Rodríguez-Gómez and Simone Lolli

Received: 30 March 2024

Revised: 26 April 2024

Accepted: 30 April 2024

Published: 30 April 2024



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

1. Introduction

Aerosols are liquid and solid particles suspended in the atmosphere, and are one of the least known components of the climate system [1–3]. The spatiotemporal characteristics of atmospheric aerosols are very important for weather forecasting [4,5], plant growth, and air quality evaluation [6,7]. Therefore, the accurate and efficient determination of aerosol conditions is one of the biggest challenges in research on the Earth–atmosphere system. The aerosol optical depth (AOD) represents the amount of light extinction by aerosols within the atmosphere, which is a good measure of the aerosol loading [8]. Over the past decade, the AOD has been retrieved using space-based sensors and ground-based instruments. Ground-based measurements, for example, those of the aerosol robotic network (AERONET) provide accurate AOD information; however, monitoring sites are sparse in the world, which complicates the spatial AOD coverage. In contrast, satellite-based observations can be used to create spatial AOD maps.

Many satellite sensors have provided official AOD products, such as the moderate resolution imaging spectroradiometer [9–11], Himawari-8 [12], and multi-angle imaging spectroradiometer (MISR) [13,14]. These official products were obtained using individual AOD-retrieval algorithms. Most of the current algorithms use the atmospheric radiative transfer model to build a look-up table (LUT) for aerosol retrieval. Although the LUT

bypasses complex mathematical equations, it is impossible to consider all the physical relationships between the satellite radiometric characteristics and atmospheric conditions [15]. For example, the LUT of the dark target (DT) and deep blue (DB) aerosol-retrieval algorithms are built based on limited aerosol models for satellite AOD retrieval [9,10]. In addition, the rich spectral information provided by satellite measurements is only partially utilized in the current LUT-based method. For instance, the MODIS DT land aerosol-retrieval algorithm only uses the information of four of thirty-six wavelength bands to retrieve the AOD (0.47, 0.66, 1.6, and 2.1 μm) [16]. Another issue with respect to LUT-based aerosol retrieval is the construction of a LUT with an appropriate size. It is well known that searching for a large LUT and the interpolation processes for the final AOD output is extremely time-consuming [17,18]. Satellite-based AOD retrievals require a reduction in the calculation time, especially in applications with large spatial scales. Thus, a more efficient method must be established for satellite AOD retrieval. To solve the time-consuming issue, general radiative transfer models have been modified to directly calculate the satellite-based AOD [15,19,20]. They use the single-scattering approximation for Rayleigh, and the computation speed is greatly improved. However, the impact of multiple scattering for both Rayleigh and aerosol reflectance is significant, and the accuracies incorporating them remain limited.

Machine learning methods have recently been investigated with respect to satellite-based AOD retrieval [18,21,22]. In contrast to LUT-based AOD calculations that require iteration or interpolation processes [17], the retrieval of the AOD using a trained machine learning model is instantaneous [22]. Benefiting from this advantage, many classical machine learning models have been applied to describe the nonlinear relationships between satellite measurements and the AOD, such as the fully connected neural network (FCNN), gradient boosting framework (LGBM) [23,24], support vector machine (SVM) [21], and backpropagation neural network (BPNN) [17]. These methods have the advantage of satellite aerosol retrieval but are affected by the training sample. Recently, several modified machine learning models have been developed to obtain better AOD results. The combination of a BPNN and genetic algorithm has been successfully applied for satellite-based AOD retrieval in China [25]. Lanzaco et al. integrated an artificial neural network and SVM to obtain an improved AOD map for South America [26]. The deep learning approach performs better in solving nonlinear problems of atmospheric characteristics [27–29]. A deep learning model with multiple hidden layers can simulate highly varying functions defining nonlinear structures [30,31], which is suitable for the determination of the nonlinear relationships between satellite measurements and the AOD. To date, some novel deep learning models have been advanced, e.g., physics-informed [32,33] methods. In addition, Chen et al. proposed a concept involving a learning model that is “deep” and “wide” [34]. They showed that this wide and deep learning model represents a significant improvement for app recommender systems compared with the deep-only model. This promising result indicates that the wide and deep learning model may be more suitable for AOD calculations than the general deep learning model. However, research on the development of a wide and deep learning model for satellite-based AOD retrievals has not been explored.

In this research, we developed an improved wide and deep learning model that can be used to retrieve the AOD from satellite measurements. Compared to traditional deep learning models, a wide modeling component was embedded. Joint training was applied to the wide and deep components. Comprehensive validation, as well as comparisons with a deep-only model (without the wide component) and extreme gradient boosting (XGBoost), are also presented. Our research presents a new method for satellite-based AOD retrieval using an advanced wide and deep learning model.

2. Materials and Methods

2.1. Training and Test Data

Because of the limited number of AERONET stations in China, we used the multi-angle implementation of atmospheric correction (MAIAC) AOD product (at 550 nm) as

training data in this study to ensure a sufficient amount of AOD data for model training. The advantages of the MAIAC algorithm are that it can retrieve the AOD over both regions with dense vegetation and bright surfaces [35] and has better accuracy than DT and DB over areas with dense vegetation and bright regions, respectively, based on previous studies [36,37]. We resampled the C6.1 MAIAC AOD products to 5 km spatial resolution using the bilinear interpolation and combined them with Himawari-8 measurements for model training. The values for each new cell were determined based on the weighted average of the four nearest cells. To ensure that the training and validation were completely independent, we used MAIAC AOD data from 2016 to 2017 as the training data, 2018 as the validation data (for hyperparameter optimization with Adam and L2 regularization as the optimizer), and AERONET V3 AOD data from 2019 as the test data. We used all available AERONET V3 AOD data (level 2.0 and level 1.5) from China to test the developed model.

2.2. Himawari-8 Measurement Data and Himawari-8 AOD Product

Himawari-8 is a geostationary weather satellite that was launched in October 2014 with a new payload called the advanced Himawari imager (AHI). The AHI is equipped with 16 observation bands (including 3 visible, 3 near-infrared, and 10 infrared bands) with temporal resolutions of 10 min [38]. In this study, the Himawari-8 band 1 to band 6 (top-of-atmosphere reflectance) and band 7 to band 16 (brightness temperature, unit: K) were obtained from the Himawari-8 L1 gridded data during the daytime (0:00–8:00 UTC) from 2016 to 2019. These data were all at 10 min intervals with a spatial resolution of 5 km. We used the Ishida and Nakajima methods to mask the cloud [39]. In addition, the Himawari-8 AOD product (V2.1) at 500 nm with the same spatial resolution of 5 km was used for comparison.

2.3. Wide and Deep Learning Model

The structure of the wide and deep learning model proposed in this study is presented in Figure 1. The wide and deep learning model has two components: a deep learning component and a wide component. The deep learning component is EntityDenseNet [40,41], consisting of entity embedding technology [34] and a batch normalization and robust neural network (BRNN) [30]. Initially, the input data for the deep learning component were divided into two groups: categorical and continuous variables. We used the Himawari-8 band 1 to band 6 (reflectance data), band 7 to 16 (brightness temperature data), latitude, longitude, solar zenith angle, relative azimuth angle, satellite zenith angle, and digital elevation model (DEM) as the continuous variables. An embedded layer [42] was then introduced to the deep learning component to deal with the categorical variables using the entity embedding method. Each continuous variable was normalized with the Z-score method before it was used as the input for the deep learning component [30]. Eventually, the continuous and categorical variables were combined and processed using two hidden layers. Each of the two hidden layers in the deep learning component included one batch normalization (BN) layer [43], one fully connected layer, one dropout layer [44], and one batch normalization (BN) layer [43]. The feed-forward operation in the hidden layer of the deep learning component is described in the following equation:

$$x^{m+1} = BN[f(W^m x^m + b^m)] \quad (1)$$

where $BN[]$ is the BN function [43]; $f()$ is the leaky ReLU activation function; m is the layer number; x^m is the output vector thinned from layer m by the dropout layer [44]; and, W^m and b^m are the weights and biases of layer m , respectively.

The wide component is a generalized linear model based on [34] in the following equation:

$$y = W^T x + b \quad (2)$$

where y is the AOD, x represents the categorical variables of China’s administrative divisions, day type, and month that are subsequently transformed into sparse binary features via one-hot encoding [45], W is the model weight factor, and b is the bias.

To jointly train the deep learning and wide components of the proposed wide and deep learning model, the two components were combined using the weighted sum of their outputs via the following equation:

$$\text{AOD} = W_{\text{wide}}^T x + W_{\text{deep}}^T \alpha^f + b \tag{3}$$

where W_{wide}^T indicates the weight factor of the wide component, W_{deep}^T means the weights for the output α^f of the deep component, and b is the bias term for this joint training process. The AOD is then put into the mean squared error (MSE) loss function for joint training of the wide and deep components. The joint training of these two components is processed by backpropagating the gradients concurrently from the result to both components using Adam (a method for stochastic optimization) [46] with a decoupled weight decay [47]. For the satellite-based AOD retrieval, we used Adam with L2 regularization as the optimizer during the joint training process.

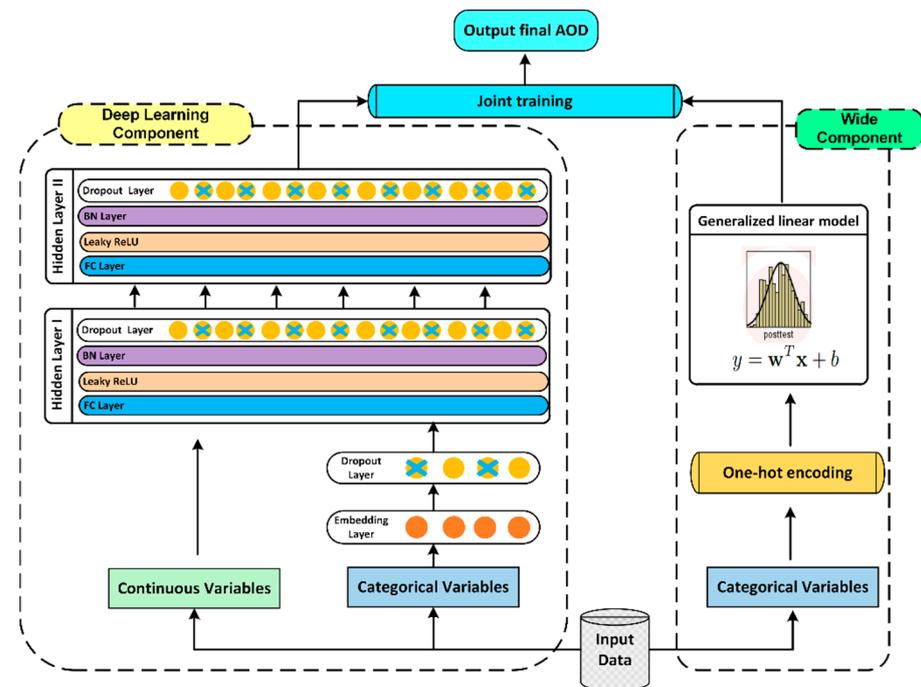


Figure 1. Schematic diagram of the wide and deep learning model proposed in this study.

3. Results

3.1. Validation and Comparison of the Wide and Deep Learning Model

Figure 2 shows the four methods that can be used to derive the AODs as a function of the AERONET AOD. High kernel density values indicate the locations of most of the data. To compare the data accuracies, a test dataset with the same data volume ($N = 8366$) was generated for validation. As shown in Figure 2a, ~63% of the wide and deep AOD retrievals were plotted in the estimated error (EE) envelope. The linear regression relation between the estimated AOD and AERONET AOD had a slope of 0.84, a y-intercept of 0.08, an R^2 of 0.81, and a root-mean-square error (RMSE) of 0.19. Approximately 58% of the AODs derived from the deep-only model (without the wide component) were plot in the EE envelope, whereas 20% of them were below the lower limit of the EE envelope. However, the underestimation at the low AOD level significantly improved based on the wide and deep method, with only 12% of AODs plotting below the EE (Figure 2a). In

In addition, the RMSE of the wide and deep AOD results was lower (RMSE = 0.19) than that of the deep-only model (RMSE = 0.21). Based on the comparison of the wide and deep AODs with the Himawari-8 V2.1 and XGBoost AOD data, the wide and deep model exhibited better performance. A significant underestimation phenomenon was observed for the Himawari-8 V2.1 AOD retrievals, that is, >40% of the results were plotted below the EE envelope. In contrast, 39% of the XGBoost AOD records were plotted above the EE envelope, which indicates an overestimation. Figure 2 shows that the wide and deep learning model outperformed the deep-only model, Himawari-8 V2.1, and XGBoost in terms of R^2 , RMSE, and the fraction of the retrievals plotted in the EE envelope.

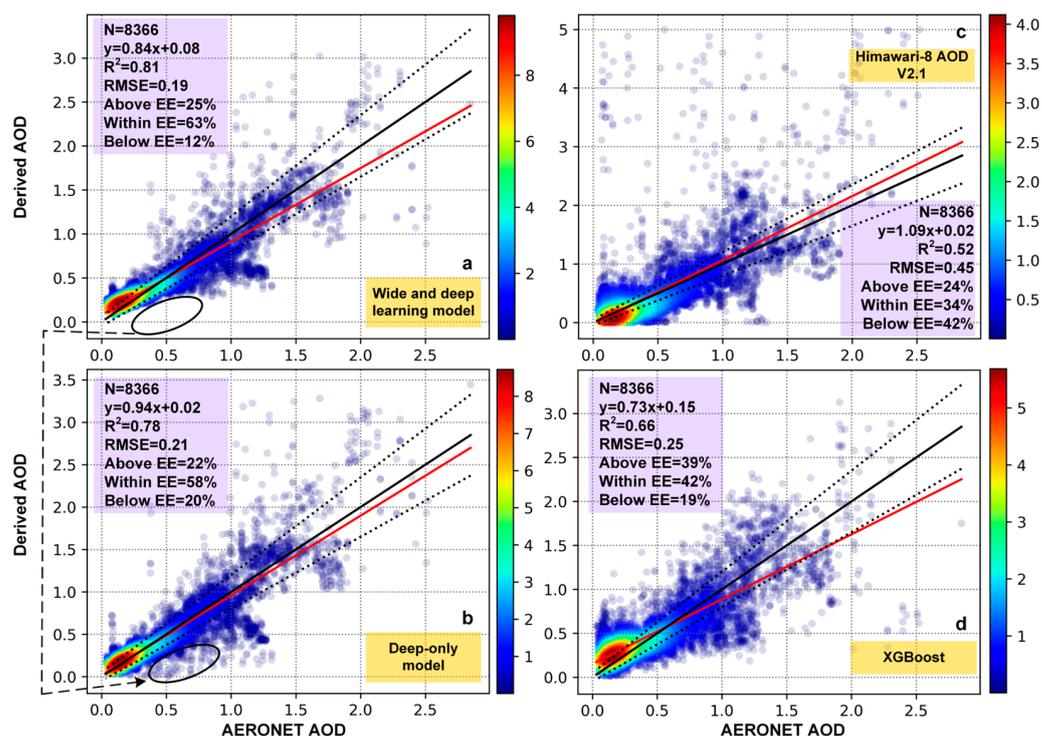


Figure 2. AOD derived as a function of the AERONET AOD based on four methods ((a): wide and deep learning model AOD at 550 nm; (b): deep-only model AOD at 550 nm; (c): Himawari-8 V2.1 AOD at 500 nm; (d): XGBoost AOD at 550 nm). The red lines are the best-fit lines based on linear regression, and the black solid lines are the 1:1 lines. The two dashed error lines are $y = 1.15x + 0.05$ (upper line) and $y = 0.85x - 0.05$ (bottom line), corresponding to the EE envelope $\pm (0.05 + 0.15 \times \text{AERONET AOD})$. The regression equations, coefficients of determination (R^2), number of data points (N), and RMSE are displayed. The colored areas indicate the density of the data points.

Figure 3 shows the AOD retrieval errors for the wide and deep learning model (a), Himawari-8 (b), deep-only model (c), and XGBoost (d). In general, all AOD retrievals worsened with increasing AOD loading. For $\text{AOD} < 0.2$, all retrieval methods overestimated the AOD, especially XGBoost (a mean bias of up to 0.2), and the results were plotted outside the EE envelope, indicating the poor performance of XGBoost at a low AOD level. For AODs between 0.2 and 0.6, the wide and deep learning and deep-only models performed better than the other two methods; most of the biases were plotted in the EE envelope, and the mean/median biases were close to 0. At this AOD level, the Himawari-8 V2.1 AOD was slightly underestimated, whereas the XGBoost AOD was slightly overestimated. For high AOD values ($\text{AOD} > 1.8$), severe overestimation and underestimation were observed in the Himawari-8 (mean bias > 0.6) and XGBoost (mean bias = -0.5) AOD retrievals, respectively. In contrast, good performance was achieved with both the wide and deep learning and deep-only models at this high AOD load ($\text{AOD} > 1.8$). However, at an AOD level between 1.2 and 1.8, the deep-only model was not as accurate as the wide and deep learning model.

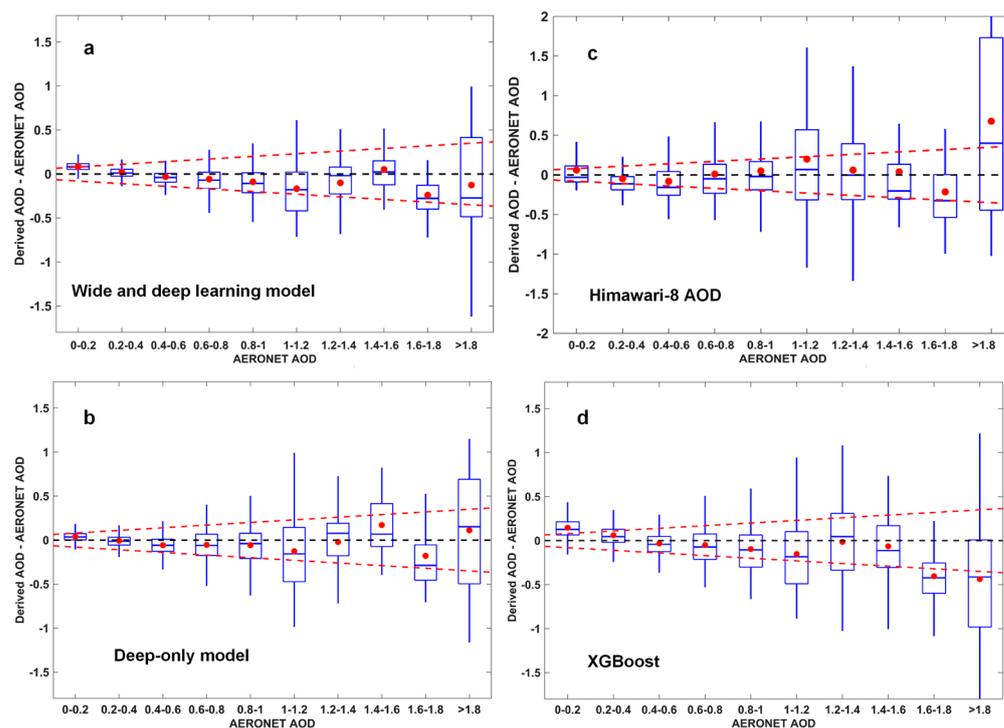


Figure 3. Box plots of the AOD errors (derived—AERONET) as a function of the AERONET AOD. (a) Wide and deep learning model proposed in this study; (b) without the wide component; (c) Himawari-8 AOD; (d) XGBoost. The mean, median, and 66% ($1 - \sigma$) intervals of the differences are shown as the red dots, horizontal lines within the boxes, and boxes themselves, respectively. The blue whiskers are the 96% ($2 - \sigma$) intervals. The black dashed line is 0, and the red dashed lines represent the EE envelope for the total AOD $\pm (0.05 + 15\%)$.

To further investigate the performances of the AOD retrieval methods, the sub-daily variations of the AOD retrievals and AERONET AOD were compared from 8:00 to 16:00 Beijing time (BT). Figure 4 shows the comparisons between the hourly averaged AERONET AOD and wide and deep learning model AOD (Figure 4a), the deep-only model AOD (Figure 4b), and the Himawari-8 V2.1 AOD (Figure 4c), respectively. Generally, in the morning, the AERONET AOD increased from 8:00 to 10:00 BT and then decreased until 12:00 BT; in the afternoon, the AERONET AOD increased from 12:00 to 16:00 BT. Aerosol particles accumulate during the night due to the stable atmosphere, reaching a peak in the morning, then dilute because of the increasing vertical flow, and are affected by increasing anthropogenic emissions in the afternoon [48,49]. Both the AOD data of the wide and deep learning and deep-only models captured the morning trend, but the Himawari-8 V2.1 AOD was lower than the AERONET AOD from 9:00 to 11:00 BT. The sub-daily AOD obtained using the deep-only model was similar to the AERONET AOD, but the data failed to depict the upward trend in the afternoon. From 13:00 to 16:00 BT, the Himawari-8 V2.1 AOD was significantly higher than the AERONET AOD, although the Himawari-8 V2.1 AOD increased. Figure 4 shows that the hourly AOD performance based on the wide and deep learning model was better than that of the deep-only model and much better than that of Himawari-8 V2.1.

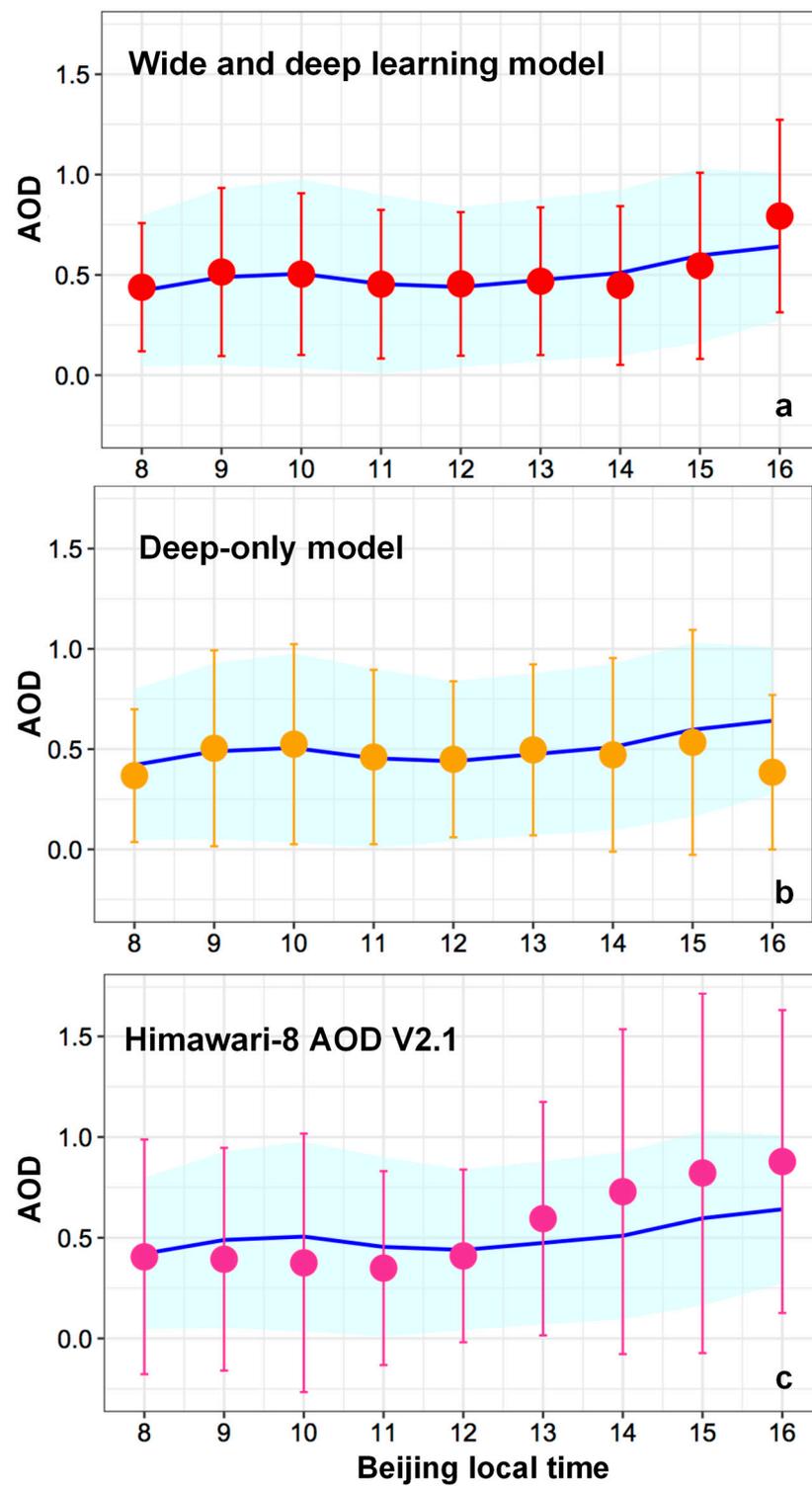


Figure 4. Comparison of the hourly averaged AOD performance from 8:00 to 16:00 Beijing time in 2019. (a) Wide and deep learning model AOD. (b) Deep-only model AOD. (c) Himawari-8 V2.1 AOD. The solid circle is the hourly mean AOD. The whiskers represent the hourly mean AOD \pm median-retrieved AOD. The blue line is the hourly mean AOD obtained using the AERONET and the cyan shades represent the hourly mean AERONET AOD \pm median AERONET AOD.

The retrievals of the wide and deep learning model were also compared with the results of previous studies (Table 1). The selected studies were based on the same satellite data (Himawari-8 for the 10 min-scale AOD retrieval) for China. The AOD results suggested that the wide and deep learning model performs well and has better accuracy than previously used methods. For example, compared with the analytical equation-based methods MAARM and SFART, the RMSE could be reduced to 0.19 in our study, and the AOD results were improved (63% in the EE).

Table 1. Comparison of the wide and deep learning model with other methods for Himawari-8 AOD data for China.

Type	Method	Test Area	N	R ² /R	RMSE	Within EE	Literature
Deep learning	Wide and deep learning	China	8366	R ² = 0.81	0.19	63%	This study
Analytical equations	MAARM	Jing-Jin-Ji (China)	468	R ² ~ 0.83	~0.23	~56.7%	[50]
Analytical equations	SFART	Jing-Jin-Ji (China)	339	R ² = 0.86	0.22	59%	[15]
Analytical equations	Improved time series algorithm	China	9049	R ~ 0.8	~0.19	~45.7%	[51]
LUT	MAIAC	Southeast Asia and southern China	16,532	R = 0.77	0.16	54.95%	[52]
LUT	New DT AHI algorithm	Full coverage of Himawari-8	1982 (Beijing = 82)	R = 0.84 (R = 0.86)	0.2 (0.3)	54.85% (15.85%)	[53]
LUT	DT	Asia	21,666	R = 0.706	0.21	49%	[54]

Notes: Jing-Jin-Ji is in Beijing, Tianjin, and the Hebei Province of China; “~” indicates the average of the validation results.

3.2. Application and Comparison with Other Satellite-Based Aerosol Products

Figure 5 shows an example of the wide and deep learning model AOD retrieval, depicting the variation of the AOD at a fine temporal resolution (10 min) on 25 March 2019. The left panels show the hourly distribution of the AOD over China at 11:00 and 12:00 BT. On the hourly scale, the AOD in the dashed elliptical region (middle panels) significantly increased. Figure 5 shows that the wide and deep learning model AOD in the dashed elliptic region slowly increased in the first half hour (11:00–11:30 BT), suddenly increased from 0.3–0.4 to 0.4–0.65 within 10 min (11:30–11:40 BT), and then insignificantly changed until 12:00 BT. In the solid circle, the AOD suddenly decreased at 11:40–11:50 BT. This indicates that the high temporal resolution of the wide and deep learning model AOD depicts the general AOD variation as well as sudden changes in specific regions.

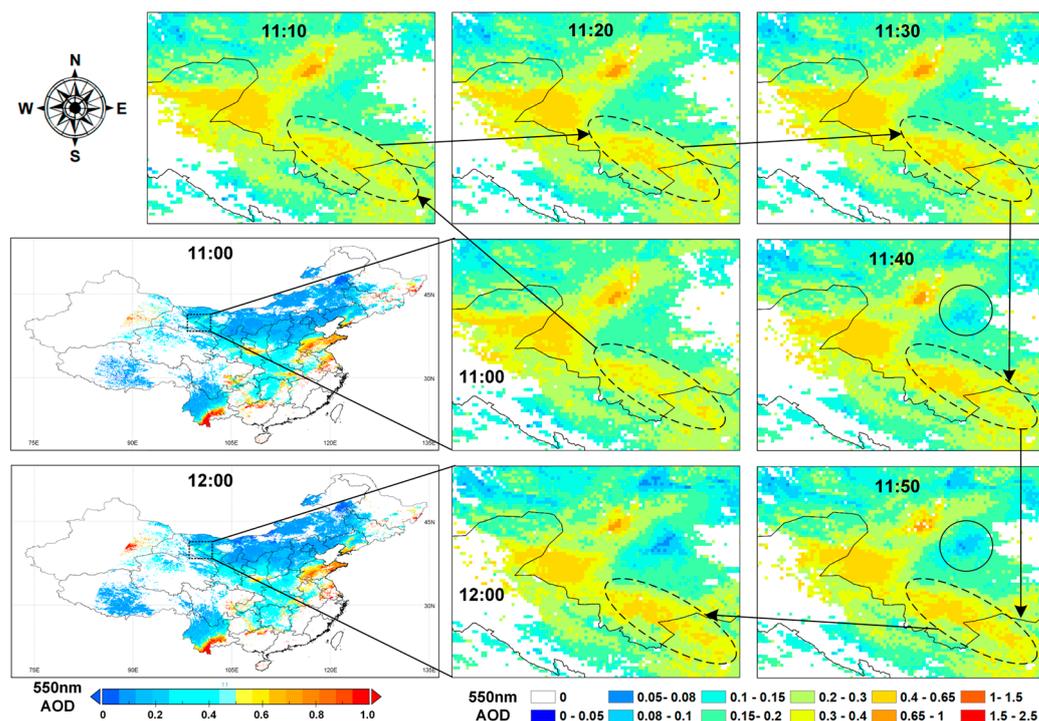


Figure 5. Application of the wide and deep learning model for AOD retrieval on 25 March 2019 from 11:00 to 12:00 BT over China.

Figure 6 compares the wide and deep learning model AOD (Figure 6b) with the Himawari-8 V2.1 AOD (Figure 6c), MODIS DT 3 km AOD (Figure 6d), MODIS DB AOD (Figure 6e), and MODIS DT 10 km AOD (Figure 6f). The true-color map (Figure 6a) shows the satellite image with cloud cover and haze over China on 16 April 2019. The majority of southern China was covered by clouds, and northeastern China suffered from haze. Compared with other AOD retrievals, the wide and deep learning model captured the haze pollution and had good spatial coverage. In the North China Plain, the wide and deep learning model, MODIS DT 3 km, MODIS DB, and MODIS DT 10 km AOD, yielded high AOD levels (>1) due to the haze on this day. However, the Himawari-8 V2.1 AOD was significantly underestimated (~ 0.6). In sparsely populated and urbanized northwestern China, the Himawari-8 V2.1 yielded an abnormally high AOD (>1) compared with the low AOD (<0.2) obtained with the wide and deep learning model and MODIS DB. The results of previous studies indicated a similar underestimation for the Himawari-8 V2.1 AOD in the North China Plain as well as very high values in northwestern China, which were attributed to the improper estimation of surface reflectance and the aerosol model in the aerosol retrieval algorithm of Himawari-8 V2.1 [55,56]. Several values over the Qinhai-Tibetan Plateau were missing in the MODIS DB AOD data (Figure 6e), which might have been due to different cloud mask methods and the pixel aggregation of the DB retrieval method [57]. Although the MODIS DT 3 km AOD had a higher spatial resolution than the MODIS DT 10 km AOD (Figure 6d,f), its accuracy was lower than that of the MODIS DT 10 km AOD, especially in urban areas [58]. Figure 6d,f shows that the spatial coverage was much smaller than that of other methods because the MODIS DT retrieval algorithm is not applicable to bright surfaces [59]. The bright surfaces in northwestern China with deserts, such as the Taklimakan and Gobi deserts, are also unfavorable for the MODIS DT retrieval algorithm [60]. However, the wide and deep learning model (Figure 6b) has no limitations with respect to “bright” or “dark” surface AOD retrievals.

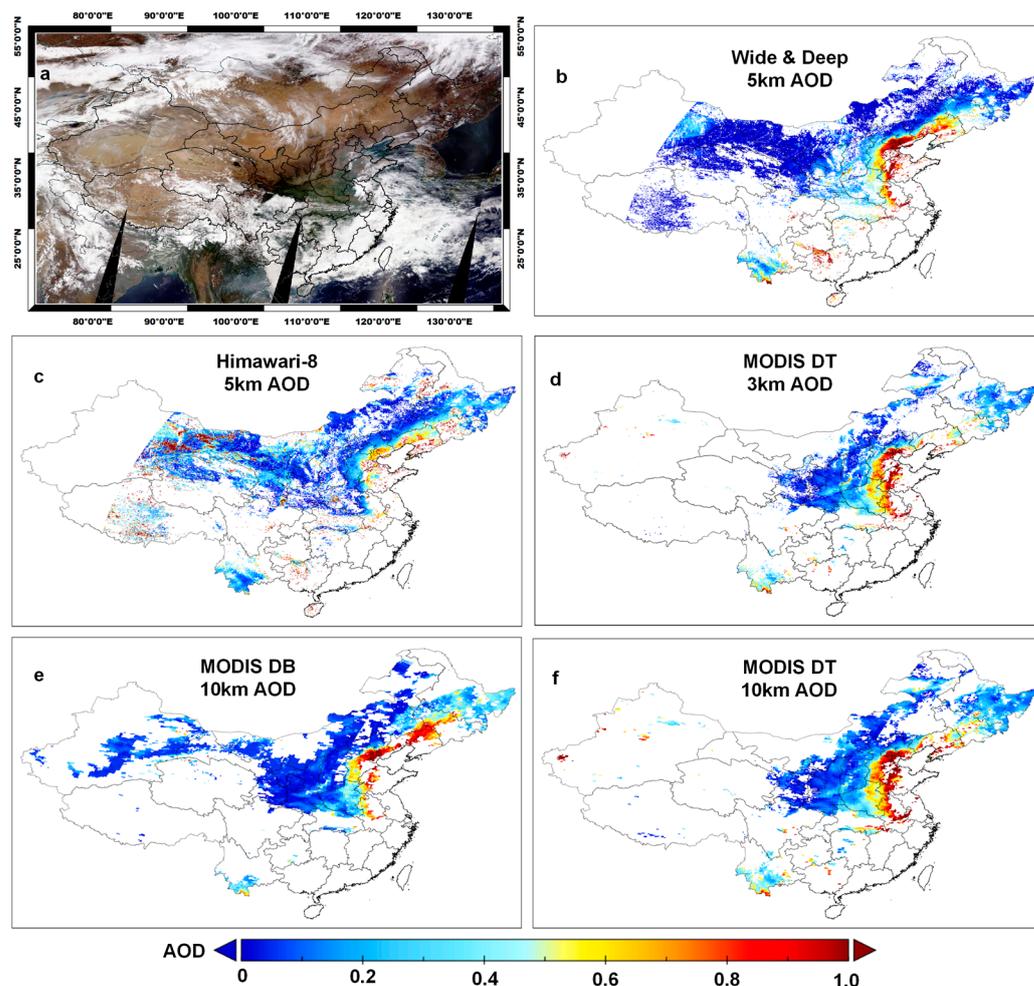


Figure 6. Comparison of the AOD retrievals obtained on 16 April 2019: wide and deep learning model AOD (b), Himawari-8 V2.1 AOD (c), MODIS DT 3 km AOD (d), MODIS DB AOD (e), and MODIS DT 10 km AOD (f). The true-color satellite image is shown in (a).

3.3. Interpretability of the Wide and Deep Learning Model

Traditional deep learning methods are frequently considered to be “black boxes”, which cannot be used to explain the results. In this study, we added a wide component to the common deep learning model and used a joint training method to improve the interpretability of the deep learning model. The wide component is a linear regression model, which predicts the target as a weighted sum of the variable input. The linearity of the relationship learned from the training data makes the interpretation easy. Figure 7 presents the AOD weights (linear regression coefficients) for different months via the wide component. In general, the spring AOD weights (March–May) are obviously higher than those of the other months. In March, April, and May, higher AOD values were obtained across mainland China, which is due to the dust transport in spring [61]. Based on almost 40 years of AOD data (1980 to 2017), Qin et al. [60] showed that the AOD values are higher in spring (0.309) than in summer (0.271), autumn (0.222), and winter (0.202), which they attributed to the frequent sandstorms in North China and straw combustion in South China. Thus, the months of March, April, and May may have higher contributions to AOD levels compared with November and December. Figure 7 shows that the highest and lowest weights are observed in April (0.55) and December (0.25), respectively. Both Sun et al. [61] and Qin et al. [60] reported that they obtained the largest and smallest monthly mean AOD values in April and December, respectively, which is consistent with the weights determined using the wide and deep learning model for these two months.

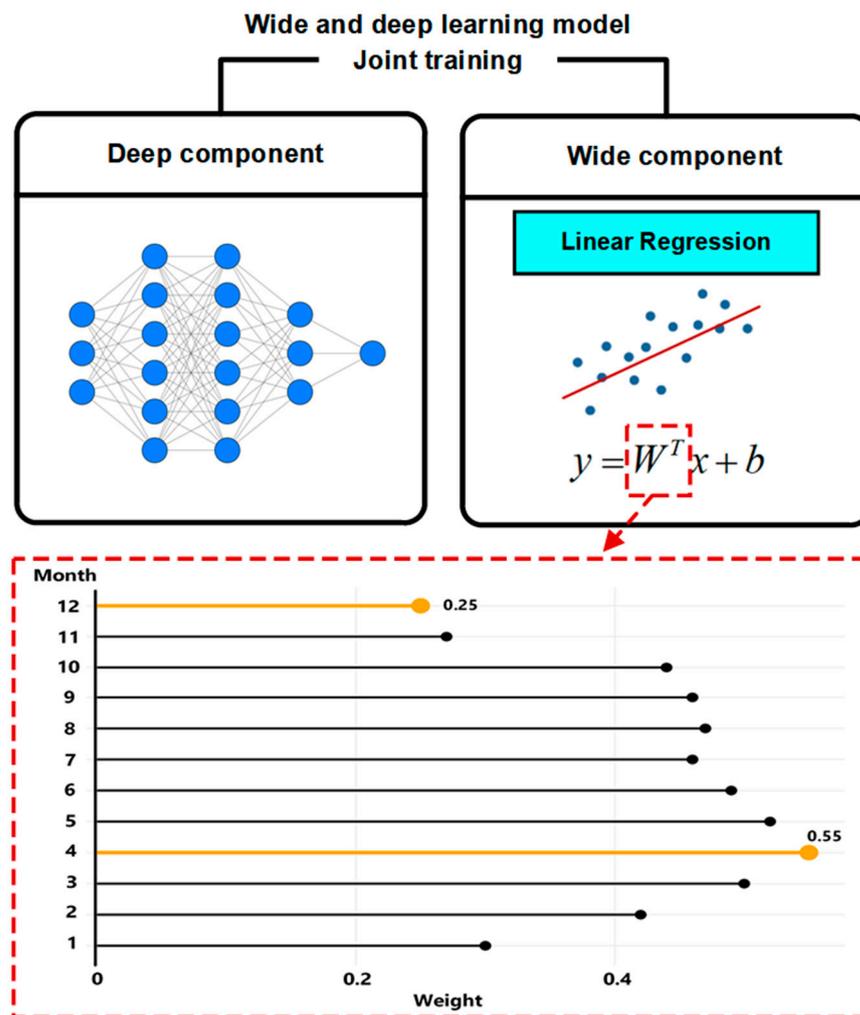


Figure 7. Monthly weights determined using the wide and deep learning model, indicating the contributions of each month to the AOD.

4. Discussion

The focus of previous machine learning models was placed on extracting accurate nonlinear relationships, and the results showed that multi-layer neural networks (deep learning models) have better performance than traditional BPNN models with only one single hidden layer [27]. Recently, the combination of a linear regression model (wide component) with deep learning neural networks was proposed to improve the model's accuracy [34]. In addition to the better accuracy of the wide and deep learning model proposed in this study (Figure 2), the interpretability of the model was enhanced due to the wide component. In addition to the AOD retrieval, the wide and deep learning model can be used to automatically extract temporal AOD characteristics. The linear regression coefficients determined using the wide component and the joint training method can be used to extract the contributions of each month to the overall AOD (Figure 7).

With respect to satellite-based AOD retrievals, most AOD retrieval methods use limited satellite band information. For instance, the DT algorithm used for MODIS and Kikuchi et al.'s [12] algorithm used for Himawari-8 is only based on the information of four wavelength bands. However, in addition to the visible bands, short-wave and thermal infrared bands also contain unique AOD information [62]. To fully apply the spectral information, including the TOA reflectance and brightness temperature, measured using the satellite for the AOD retrieval, a machine learning model (or deep learning) can be used to integrate the information from these bands, taking advantage of their intricate

relationships with the AOD instead of complex physical models. In addition, the wide and deep learning model proposed in this study does not require surface reflectance and aerosol model information for the AOD calculation. The surface reflectance and aerosol model are considered to be two key factors affecting the uncertainties in satellite-based AOD calculations [63]. However, the wide and deep learning model has several limitations. First, highly accurate AOD products are required as training data. In this study, we used the MAIAC AOD product for the training of the model. Although the accuracy of the wide and deep learning model AOD results was better than that of the Himawari-8 V2.1 AOD, overestimation was still observed, which is mainly due to the training data. The MAIAC AOD was also overestimated over China [64]. Second, the deep learning model relies heavily on learned information and will perform worse when feedback is provided for information that has not been learned. If sudden changes in the AOD occur due to policy interventions, which are not included in the training datasets, the predictions of the wide and deep learning model proposed in this study will fail, resulting in high uncertainties in the AOD values. In the future, more efforts will be made to improve these issues, including the use of more accurate AOD data as training datasets and the improvement of the model with respect to unknown knowledge prediction.

5. Conclusions

In this study, a wide and deep learning model was developed for satellite-based AOD retrievals using Himawari-8 imagery. The model was trained with MAIAC AOD data from 2016 to 2018 and tested with AERONET AOD data from 2019. One strength of the model is that the test data are completely independent of the training dataset with reasonable accuracy. We compared the AOD retrievals from the deep-only model (the wide and deep learning model without the wide component), Himawari-8 V2.1 (a LUT-based method), and XGBoost (a traditional machine learning model). Based on the validation with the ground-based AERONET AOD, our model AOD retrieval ($R^2 = 0.81$, RMSE = 0.19, 63% in the EE) outperformed the deep-only model ($R^2 = 0.78$, RMSE = 0.21, 58% in EE), XGBoost ($R^2 = 0.66$, RMSE = 0.25, 42% in the EE), and Himawari-8 V2.1 AOD ($R^2 = 0.52$, RMSE = 0.45, 34% in the EE). Another novelty of this study is that, by introducing the wide component, the underestimation was significantly reduced (12% below the EE) compared with the deep-only model (20% below the EE). Furthermore, a key feature of our model is its interpretability, which can be used to explore the temporal AOD characteristics, revealing the higher contribution of spring months to the AOD levels in China.

Our work shows that the proposed wide and deep learning model can capture the near-real-time variation of the AOD with an ultrafine temporal resolution in China with high accuracy. The wide and deep learning model also has potential with respect to the application of interpretation capabilities to explain the predictions, which is particularly valuable for the automatic extraction of the spatiotemporal characteristics from Earth observation datasets.

Author Contributions: Conceptualization, X.Y.; methodology, X.Y.; validation, N.L., J.Z. and Z.Z.; formal analysis, N.L. and Z.Z.; investigation, N.L. and Z.Z.; data curation, Z.Z.; writing—original draft preparation, N.L.; writing—review and editing, X.Y. and T.C.; visualization, N.L. and Z.Z.; supervision, X.Y.; project administration, X.Y.; funding acquisition, N.L. and X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant numbers: 42301395, 42030606, and 91837204), the Natural Science Foundation of Beijing (grant no. 8222058), R and D Program of Beijing Municipal Education Commission (KM202310016004), and the Pyramid Talent Training Project of Beijing University of Civil Engineering and Architecture (grant no. JDYC20220823).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, Xing Yan, upon reasonable request. The data are not publicly available due to privacy.

Acknowledgments: The authors gratefully acknowledge the MODIS and AERONET teams for their effort in making the data available. We would like to thank the Meteorological Satellite Center (MSC) of the Japan Meteorological Agency (JMA) for providing the Himawari-8 data.

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

References

1. Boucher, O.; Randall, D.; Artaxo, P.; Bretherton, C.; Feingold, G.; Forster, P.; Kerminen, V.-M.; Kondo, Y.; Liao, H.; Lohmann, U. Clouds and aerosols. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2013; pp. 571–657.
2. Luo, N.; Zhang, Y.; Jiang, Y.; Zuo, C.; Chen, J.; Zhao, W.; Shi, W.; Yan, X. Unveiling global land fine-and coarse-mode aerosol dynamics from 2005 to 2020 using enhanced satellite-based inversion data. *Environ. Pollut.* **2024**, *348*, 123838. [[CrossRef](#)] [[PubMed](#)]
3. Zang, Z.; Zhang, Y.; Zuo, C.; Chen, J.; He, B.; Luo, N.; Zou, J.; Zhao, W.; Shi, W.; Yan, X. Exploring Global Land Coarse-Mode Aerosol Changes from 2001–2021 Using a New Spatiotemporal Coaction Deep-Learning Model. *Environ. Sci. Technol.* **2023**, *57*, 19881–19890. [[CrossRef](#)] [[PubMed](#)]
4. Al-Saadi, J.; Szykman, J.; Pierce, R.B.; Kittaka, C.; Neil, D.; Chu, D.A.; Remer, L.; Gumley, L.; Prins, E.; Weinstock, L. Improving national air quality forecasts with satellite aerosol observations. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 1249–1262. [[CrossRef](#)]
5. Grell, G.; Baklanov, A. Integrated modeling for forecasting weather and air quality: A call for fully coupled approaches. *Atmos. Environ.* **2011**, *45*, 6845–6851. [[CrossRef](#)]
6. Zuo, C.; Chen, J.; Zhang, Y.; Jiang, Y.; Liu, M.; Liu, H.; Zhao, W.; Yan, X. Evaluation of four meteorological reanalysis datasets for satellite-based PM_{2.5} retrieval over China. *Atmos. Environ.* **2023**, *305*, 119795. [[CrossRef](#)]
7. Yan, X.; Shi, W.; Li, Z.; Li, Z.; Luo, N.; Zhao, W.; Wang, H.; Yu, X. Satellite-based PM_{2.5} estimation using fine-mode aerosol optical thickness over China. *Atmos. Environ.* **2017**, *170*, 290–302. [[CrossRef](#)]
8. Kaufman, Y.J.; Tanré, D.; Boucher, O. A satellite view of aerosols in the climate system. *Nature* **2002**, *419*, 215–223. [[CrossRef](#)] [[PubMed](#)]
9. Levy, R.C.; Remer, L.A.; Mattoo, S.; Vermote, E.F.; Kaufman, Y.J. Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *J. Geophys. Res. Atmos.* **2007**, *112*, D13211. [[CrossRef](#)]
10. Sayer, A.M.; Hsu, N.; Bettenhausen, C.; Jeong, M.J. Validation and uncertainty estimates for MODIS Collection 6 “Deep Blue” aerosol data. *J. Geophys. Res. Atmos.* **2013**, *118*, 7864–7872. [[CrossRef](#)]
11. Yan, X.; Zang, Z.; Li, Z.; Luo, N.; Zuo, C.; Jiang, Y.; Li, D.; Guo, Y.; Zhao, W.; Shi, W. A global land aerosol fine-mode fraction dataset (2001–2020) retrieved from MODIS using hybrid physical and deep learning approaches. *Earth Syst. Sci. Data Discuss.* **2021**, *2021*, 1193–1213. [[CrossRef](#)]
12. Kikuchi, M.; Murakami, H.; Suzuki, K.; Nagao, T.M.; Higurashi, A. Improved hourly estimates of aerosol optical thickness using spatiotemporal variability derived from Himawari-8 geostationary satellite. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3442–3455. [[CrossRef](#)]
13. Kahn, R.A.; Gaitley, B.J.; Martonchik, J.V.; Diner, D.J.; Crean, K.A.; Holben, B. Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on 2 years of coincident Aerosol Robotic Network (AERONET) observations. *J. Geophys. Res. Atmos.* **2005**, *110*, D10S04. [[CrossRef](#)]
14. Shao, Y.; Liu, R.; Li, W.; Bi, J.; Ma, Z. A MISR-Based Method for the Estimation of Particle Size Distribution: Comparison with AERONET over China. *Remote Sens.* **2023**, *3*, 0032. [[CrossRef](#)]
15. Yan, X.; Luo, N.; Liang, C.; Zang, Z.; Zhao, W.; Shi, W. Simplified and Fast Atmospheric Radiative Transfer model for satellite-based aerosol optical depth retrieval. *Atmos. Environ.* **2020**, *224*, 117362. [[CrossRef](#)]
16. Kokhanovsky, A. Remote sensing of atmospheric aerosol using spaceborne optical observations. *Earth-Sci. Rev.* **2013**, *116*, 95–108. [[CrossRef](#)]
17. Okada, Y.; Mukai, S.; Sano, I. Neural network approach for aerosol retrieval. In Proceedings of the IGARSS 2001, Scanning the Present and Resolving the Future, Proceedings, IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217), Sydney, NSW, Australia, 9–13 July 2001; pp. 1716–1718.
18. Sun, Y.; Hang, R.; Liu, Q.; Zhu, F.; Pei, H. Graph-regularized low-rank representation for aerosol optical depth retrieval. *Int. J. Remote Sens.* **2016**, *37*, 5749–5762. [[CrossRef](#)]
19. Seidel, F.; Kokhanovsky, A.; Schaepman, M.E. Fast retrieval of aerosol optical depth and its sensitivity to surface albedo using remote sensing data. *Atmos. Res.* **2012**, *116*, 22–32. [[CrossRef](#)]
20. Mei, L.; Rozanov, V.; Burrows, J.P. A fast and accurate radiative transfer model for aerosol remote sensing. *J. Quant. Spectrosc. Transf.* **2020**, *256*, 107270. [[CrossRef](#)] [[PubMed](#)]

21. Nguyen, T.N.T.; Mantovani, S.; Campalani, P.; Cavicchi, M.; Bottoni, M. Aerosol optical thickness retrieval from satellite observation using support vector regression. In Proceedings of the Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 15th Iberoamerican Congress on Pattern Recognition, CIARP 2010, Sao Paulo, Brazil, 8–11 November 2010; Proceedings 15; pp. 492–499.
22. Taylor, M.; Kazadzis, S.; Tsekeri, A.; Gkikas, A.; Amiridis, V. Satellite retrieval of aerosol microphysical and optical parameters using neural networks: A new methodology applied to the Sahara desert dust peak. *Atmos. Meas. Tech.* **2014**, *7*, 3151–3175. [[CrossRef](#)]
23. Fan, Y.; Sun, L. Satellite Aerosol Optical Depth Retrieval Based on Fully Connected Neural Network (FCNN) and a Combine Algorithm of Simplified Aerosol Retrieval Algorithm and Simplified and Robust Surface Reflectance Estimation (SREMARA). *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4947–4962. [[CrossRef](#)]
24. Kang, E.; Park, S.; Kim, M.; Yoo, C.; Im, J.; Song, C.-K. Direct aerosol optical depth retrievals using MODIS reflectance data and machine learning over East Asia. *Atmos. Environ.* **2023**, *309*, 119951. [[CrossRef](#)]
25. Qin, W.; Wang, L.; Lin, A.; Zhang, M.; Bilal, M. Improving the estimation of daily aerosol optical depth and aerosol radiative effect using an optimized artificial neural network. *Remote Sens.* **2018**, *10*, 1022. [[CrossRef](#)]
26. Lanzaco, B.L.; Olcese, L.E.; Palancar, G.G.; Toselli, B.M. An improved aerosol optical depth map based on machine-learning and MODIS data: Development and application in South America. *Aerosol Air Qual. Res.* **2017**, *17*, 1623–1636. [[CrossRef](#)]
27. Brahma, P.P.; Wu, D.; She, Y. Why deep learning works: A manifold disentanglement perspective. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *27*, 1997–2008. [[CrossRef](#)] [[PubMed](#)]
28. Shikhovtsev, A.Y.; Kovadlo, P.G.; Kiselev, A.V.; Eselevich, M.V.; Lukin, V.P. Application of Neural Networks to Estimation and Prediction of Seeing at the Large Solar Telescope Site. *Publ. Astron. Soc. Pac.* **2023**, *135*, 014503. [[CrossRef](#)]
29. Ali, Y.; Aly, H.H. Short term wind speed forecasting using artificial and wavelet neural networks with and without wavelet filtered data based on feature selections technique. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108201. [[CrossRef](#)]
30. 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*, 8427–8437. [[CrossRef](#)]
31. 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*, 102516. [[CrossRef](#)]
32. Jiang, J.; Tao, M.; Xu, X.; Jiang, Z.; Man, W.; Wang, J.; Wang, L.; Wang, Y.; Zheng, Y.; Tao, J. A Generalized Aerosol Algorithm for Multi-Spectral Satellite Measurement With Physics-Informed Deep Learning Method. *Geophys. Res. Lett.* **2023**, *50*, e2023GL106806. [[CrossRef](#)]
33. Tao, M.; Chen, J.; Xu, X.; Man, W.; Xu, L.; Wang, L.; Wang, Y.; Wang, J.; Fan, M.; Shahzad, M.I. A robust and flexible satellite aerosol retrieval algorithm for multi-angle polarimetric measurements with physics-informed deep learning method. *Remote Sens. Environ.* **2023**, *297*, 113763. [[CrossRef](#)]
34. Cheng, H.-T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anderson, G.; Corrado, G.; Chai, W.; Ispir, M. Wide & deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, Boston, MA, USA, 15 September 2016; pp. 7–10.
35. Lyapustin, A.; Wang, Y.; Korokin, S.; Huang, D. MODIS collection 6 MAIAC algorithm. *Atmos. Meas. Tech.* **2018**, *11*, 5741–5765. [[CrossRef](#)]
36. Martins, V.S.; Lyapustin, A.; de Carvalho, L.A.; Barbosa, C.C.F.; Novo, E.M.L.d.M. Validation of high-resolution MAIAC aerosol product over South America. *J. Geophys. Res. Atmos.* **2017**, *122*, 7537–7559. [[CrossRef](#)]
37. Superczynski, S.D.; Kondragunta, S.; Lyapustin, A.I. Evaluation of the multi-angle implementation of atmospheric correction (MAIAC) aerosol algorithm through intercomparison with VIIRS aerosol products and AERONET. *J. Geophys. Res. Atmos.* **2017**, *122*, 3005–3022. [[CrossRef](#)] [[PubMed](#)]
38. Yumimoto, K.; Nagao, T.; Kikuchi, M.; Sekiyama, T.; Murakami, H.; Tanaka, T.; Ogi, A.; Irie, H.; Khatri, P.; Okumura, H.; et al. Aerosol data assimilation using data from Himawari-8, a next-generation geostationary meteorological satellite. *Geophys. Res. Lett.* **2016**, *43*, 5886–5894. [[CrossRef](#)]
39. Ishida, H.; Nakajima, T.Y. Development of an unbiased cloud detection algorithm for a spaceborne multispectral imager. *J. Geophys. Res. Atmos.* **2009**, *114*, D07206. [[CrossRef](#)]
40. Yan, X.; Zang, Z.; Luo, N.; Jiang, Y.; Li, Z. New interpretable deep learning model to monitor real-time PM_{2.5} concentrations from satellite data. *Environ. Int.* **2020**, *144*, 106060. [[CrossRef](#)] [[PubMed](#)]
41. Yan, X.; Guo, Y.; Zhang, Y.; Chen, J.; Jiang, Y.; Zuo, C.; Zhao, W.; Shi, W. Combining physical mechanisms and deep learning models for hourly surface ozone retrieval in China. *J. Environ. Manag.* **2024**, *351*, 119942. [[CrossRef](#)] [[PubMed](#)]
42. Guo, C.; Berkhahn, F. Entity embeddings of categorical variables. *arXiv* **2016**, arXiv:1604.06737.
43. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
44. 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*, 1929–1958.
45. Qu, Y.; Cai, H.; Ren, K.; Zhang, W.; Yu, Y.; Wen, Y.; Wang, J. Product-based neural networks for user response prediction. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016; pp. 1149–1154.

46. Kingma, P.D.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
47. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
48. Yang, F.; Wang, Y.; Tao, J.; Wang, Z.; Fan, M.; De Leeuw, G.; Chen, L. Preliminary investigation of a new AHI aerosol optical depth (AOD) retrieval algorithm and evaluation with multiple source AOD measurements in China. *Remote Sens.* **2018**, *10*, 748. [[CrossRef](#)]
49. Lu, C.; Li, Q.; Xing, C.; Hu, Q.; Tan, W.; Lin, H.; Lin, J.; Zhang, Z.; Chang, B.; Liu, C.S. A Novel Hyperspectral Remote Sensing Technique with Hour-Hectometer Level Horizontal Distribution of Trace Gases: To Accurately Identify Emission Sources. *J. Remote Sens.* **2023**, *3*, 0098. [[CrossRef](#)]
50. Yan, X.; Li, Z.; Luo, N.; Shi, W.; Zhao, W.; Yang, X.; Jin, J. A minimum albedo aerosol retrieval method for the new-generation geostationary meteorological satellite Himawari-8. *Atmos. Res.* **2018**, *207*, 14–27. [[CrossRef](#)]
51. Li, D.; Qin, K.; Wu, L.; Mei, L.; de Leeuw, G.; Xue, Y.; Shi, Y.; Li, Y. Himawari-8-derived aerosol optical depth using an improved time series algorithm over eastern China. *Remote Sens.* **2020**, *12*, 978. [[CrossRef](#)]
52. She, L.; Zhang, H.; Wang, W.; Wang, Y.; Shi, Y. Evaluation of the multi-angle implementation of atmospheric correction (MAIAC) aerosol algorithm for Himawari-8 data. *Remote Sens.* **2019**, *11*, 2771. [[CrossRef](#)]
53. Gupta, P.; Levy, R.C.; Mattoo, S.; Remer, L.A.; Holz, R.E.; Heidinger, A.K. Applying the Dark Target aerosol algorithm with Advanced Himawari Imager observations during the KORUS-AQ field campaign. *Atmos. Meas. Tech.* **2019**, *12*, 6557–6577. [[CrossRef](#)]
54. Lim, H.; Choi, M.; Kim, M.; Kim, J.; Chan, P. Retrieval and validation of aerosol optical properties using Japanese next generation meteorological satellite, Himawari-8. *Korean J. Remote Sens.* **2016**, *32*, 681–691. [[CrossRef](#)]
55. Zhang, Z.; Wu, W.; Fan, M.; Wei, J.; Tan, Y.; Wang, Q. Evaluation of MAIAC aerosol retrievals over China. *Atmos. Environ.* **2019**, *202*, 8–16. [[CrossRef](#)]
56. Yang, X.; Zhao, C.; Luo, N.; Zhao, W.; Shi, W.; Yan, X. Evaluation and Comparison of Himawari-8 L2 V1. 0, V2. 1 and MODIS C6. 1 aerosol products over Asia and the oceania regions. *Atmos. Environ.* **2020**, *220*, 117068. [[CrossRef](#)]
57. Tao, M.; Wang, J.; Li, R.; Wang, L.; Wang, Z.; Tao, J.; Che, H.; Chen, L. Performance of MODIS high-resolution MAIAC aerosol algorithm in China: Characterization and limitation. *Atmos. Environ.* **2019**, *213*, 159–169. [[CrossRef](#)]
58. Gupta, P.; Remer, L.A.; Levy, R.C.; Mattoo, S. Validation of MODIS 3 km land aerosol optical depth from NASA’s EOS Terra and Aqua missions. *Atmos. Meas. Tech.* **2018**, *11*, 3145–3159. [[CrossRef](#)]
59. Liu, N.; Zou, B.; Feng, H.; Wang, W.; Tang, Y.; Liang, Y. Evaluation and comparison of multiangle implementation of the atmospheric correction algorithm, Dark Target, and Deep Blue aerosol products over China. *Atmos. Chem. Phys.* **2019**, *19*, 8243–8268. [[CrossRef](#)]
60. Qin, W.; Liu, Y.; Wang, L.; Lin, A.; Xia, X.; Che, H.; Bilal, M.; Zhang, M. Characteristic and driving factors of aerosol optical depth over mainland China during 1980–2017. *Remote Sens.* **2018**, *10*, 1064. [[CrossRef](#)]
61. Sun, E.; Xu, X.; Che, H.; Tang, Z.; Gui, K.; An, L.; Lu, C.; Shi, G. Variation in MERRA-2 aerosol optical depth and absorption aerosol optical depth over China from 1980 to 2017. *J. Atmos. Sol.-Terr. Phys.* **2019**, *186*, 8–19. [[CrossRef](#)]
62. Lee, S.-S.; Sohn, B. Nighttime AOT retrieval for Asian dusts from MODIS IR measurements: An artificial neural network approach. *J. Meteorol. Soc. Jpn.* **2012**, *90*, 163–177. [[CrossRef](#)]
63. Mhawish, A.; Banerjee, T.; Sorek-Hamer, M.; Lyapustin, A.; Broday, D.M.; Chatfield, R. Comparison and evaluation of MODIS Multi-angle Implementation of Atmospheric Correction (MAIAC) aerosol product over South Asia. *Remote Sens. Environ.* **2019**, *224*, 12–28. [[CrossRef](#)]
64. Wang, Y.; Yuan, Q.; Wang, H.; Li, T.; Shen, H.; Zhang, L. Validation of MODIS 1-Km MAIAC Aerosol Products with AERONET in China During 2008–2016. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 7610–7613.

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