remote sensing

MDPI

Article

Estimation of $PM_{2.5}$ Using Multi-Angle Polarized TOA Reflectance Data from the GF-5B Satellite

Ruijie Zhang [1,2], Hui Chen [2], Ruizhi Chen [1,2], Chunyan Zhou [2], Qing Li [2], Huizhen Xie [2] and Zhongting Wang [2,3,*]

[1] Chinese Research Academy of Environmental Sciences, Beijing 100012, China; zhangruijie222@mails.ucas.ac.cn (R.Z.)
[2] Satellite Application Center for Ecology and Environment, Ministry of Ecology and Environment, Beijing 100094, China
[3] College of Surveying and Geo-Informatics, North China University of Water Resources and Electric Power, Zhengzhou 450046, China
* Correspondence: wzt_07@126.com

Abstract: The use of satellite data to estimate $PM_{2.5}$ is an appropriate approach for long-term, substantial monitoring and assessment. To estimate $PM_{2.5}$, the majority of the algorithms now in use utilize the top-of-atmosphere (TOA) reflectance or aerosol optical depth (AOD) derived from scalar satellite data. However, there is relatively little research on the retrieval of $PM_{2.5}$ using multi-angle polarized data. With its directional polarimetric camera (DPC), the Chinese new-generation satellite Gaofen 5B (henceforth referred to as GF-5B) offers a unique opportunity to close this gap in multi-angle polarized observation data. In this research, we utilized TOA data from the DPC payload and applied the gradient boosting machine method to simulate the impact of the observation angle, wavelength, and polarization information on the accuracy of $PM_{2.5}$ retrieval. We identified the optimal conditions for the effective estimation of $PM_{2.5}$. The quantitative results indicated that, under these optimal conditions, the $PM_{2.5}$ concentrations retrieved by GF-5B showed a strong correlation with the ground-based data, achieving an R^2 of 0.9272 and an RMSE of 7.38 $\mu g \cdot m^{-3}$. By contrast, Himawari-8's retrieval accuracy under similar data conditions consisted of an R^2 of 0.9099 and RMSE of 7.42 $\mu g \cdot m^{-3}$, indicating that GF-5B offers higher accuracy. Furthermore, the retrieval results in this study demonstrated an R^2 of 0.81 when compared to the CHAP dataset, confirming the feasibility and effectiveness of the use of GF-5B for $PM_{2.5}$ retrieval and providing support for $PM_{2.5}$ estimation through multi-angle polarized data.

Keywords: directional polarimetric camera (DPC); Gaofen 5B (GF-5B); $PM_{2.5}$ estimation; machine learning; TOA reflectance

check for updates

Citation: Zhang, R.; Chen, H.; Chen, R.; Zhou, C.; Li, Q.; Xie, H.; Wang, Z. Estimation of $PM_{2.5}$ Using Multi-Angle Polarized TOA Reflectance Data from the GF-5B Satellite. *Remote Sens.* **2024**, *16*, 3944. https://doi.org/10.3390/rs16213944

Academic Editor: Matthew McGill

Received: 1 October 2024
Revised: 18 October 2024
Accepted: 22 October 2024
Published: 23 October 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

Fine particulate matter is the term used to describe atmospheric particles with an aerodynamic diameter of less than 2.5 μm ($PM_{2.5}$). There is a strong link between $PM_{2.5}$ concentrations and respiratory, cardiovascular, and premature death, according to epidemiological research [1–5]. Rapid urbanization and development in developing nations over the past few decades have rendered $PM_{2.5}$ a serious contaminant in regions such as China, creating concern among the public and government alike [6,7]. The accurate monitoring and estimation of $PM_{2.5}$ concentrations have become a global priority.

While the ground-based monitoring of aerosol particles has advanced significantly, it still faces challenges such as spatial discontinuity and difficulties in obtaining comprehensive spatial data on the particulate composition [8,9]. An alternative approach that overcomes the drawbacks of ground monitoring and allows for more extensive temporal and spatial coverage is the estimation of ground-level $PM_{2.5}$ concentrations using satellite remote sensing data [10]. The satellite remote sensing of particulate matter primarily relies on various aerosol monitoring sensors onboard satellites, with $PM_{2.5}$ retrieval

based on the derived AOD [11,12]. The primary method of estimating the ground-level $PM_{2.5}$ from the satellite-derived AOD is to create models that estimate the ground-level $PM_{2.5}$ concentrations by combining meteorological data from monitoring locations with ground-based $PM_{2.5}$ monitoring data that are already available. The main methods include statistical modeling [13–17], the proportionality factor method [18–20], semi-empirical methods based on physical mechanisms [9,21], and increasingly popular machine learning algorithms [22–27]. Machine learning methods are particularly advantageous as they can consider various factors influencing the $PM_{2.5}$ mass concentration, including the local meteorological conditions, land surface types, seasons, and pollution levels. These methods are user-friendly and highly accurate, making them the most popular means of $PM_{2.5}$ remote sensing retrieval [28].

One common limitation in the use of satellite AOD data is that, in certain situations—such as over bright surfaces—AOD retrieval can fail, rendering the data unreliable or unavailable [29,30]. The AOD is derived from TOA reflectance measurements by various methods [30]. Directly using the TOA reflectance to estimate $PM_{2.5}$ concentrations may circumvent many sources of uncertainty in AOD retrieval [31] and demonstrate greater spatial coverage [32,33]. Some researchers have already employed machine learning methods to directly retrieve $PM_{2.5}$ concentrations from TOA data, utilizing satellite payloads such as MODIS [32,34], AHI/Himawari-8 [33,35–37], AGRI/FY-4A [38], and GOCI/COMS [39]. MODIS offers daily global coverage and long-term traceability, although it only provides one observation per day. Himawari-8 and FY-4A, as geostationary satellites, can offer data with an hourly resolution but are limited to monitoring the East Asian region. Despite the advantages of these satellites, they lack the capability for polarized observations.

In recent years, studies utilizing the Gaofen satellite series to investigate aerosol particles have begun to emerge [40,41]. With the launch of China's "GF-5B" satellite on 7 September 2021, the nation's capacity to use remote sensing to track the air quality has advanced significantly. GF-5B offers a unique opportunity for multi-angle polarized $PM_{2.5}$ retrieval by supplying extra remote sensing data sources for extensive aerosol and particulate matter monitoring in China. The GF-5B DPC payload is a key multi-angle polarized observation payload in China [42]. Compared to traditional scalar observation data, multi-angle polarized observation data can provide more information, leading to the more accurate retrieval of the AOD and $PM_{2.5}$. Therefore, it is essential to fully utilize and explore polarized aerosol information for $PM_{2.5}$ retrieval. This study employs a gradient boosting machine method to directly retrieve the $PM_{2.5}$ from GF-5B's multi-angle polarized TOA reflectance data. Data from 1 January 2022 to 31 December 2022 are used in the experiment. This includes measurements of the ground-level $PM_{2.5}$, GF-5B TOA reflectance, meteorological variables, and auxiliary data including NDVI data, night-light data, and DEM elevation models. The performance of this method in directly extracting the ground-level $PM_{2.5}$ from the multi-angle polarized TOA reflectance of GF-5B is assessed, along with its applicability in the Beijing, Tianjin, and Hebei region.

This research is innovative in terms of three aspects: (1) it explores the feasibility of estimating $PM_{2.5}$ using the GF-5B satellite and examines the impact of multi-angle polarized observation data on $PM_{2.5}$ retrieval for the first time; (2) it improves the spatial coverage of ground-level $PM_{2.5}$ estimates by directly utilizing the TOA reflectance instead of the AOD; (3) it directly compares the performance advantages of the GF-5B satellite with those of the Himawari-8 satellite and the CHAP dataset in estimating $PM_{2.5}$.

2. Materials

2.1. Satellite Data

The GF-5B satellite was launched by China on 7 September 2021. Six payloads are carried by the GF-5B satellite, which replaces GF-5: a directional polarization camera (DPC), a greenhouse gas monitoring instrument (GMI), an environmental trace gas monitoring instrument (EMI), an absorbing aerosol sensor (AAS), an advanced hyperspectral imager (AHSI), and a vegetation and ice monitoring imager (VIMI). In the visible to near-infrared

spectrum (wavelengths ranging from 443 nm to 910 nm), the DPC receives signals across eight channels; the channels at 490 nm, 660 nm, and 865 nm are reserved specifically for polarization studies. Compared to the POLDER-3 instrument (with a resolution of 6 km), GF-5B's DPC offers an improved spatial resolution of 3.3 km, and the maximum observation angle has been increased to 17 degrees relative to the DPC/GF-5.

For this study, we utilized Level 1 multi-angle polarized observation data at wavelengths of 490 nm (P), 670 nm (P), 865 nm (P), and 565 nm, along with corresponding solar zenith angles, solar azimuth angles, observation zenith angles, and observation azimuth angles, to retrieve the $PM_{2.5}$ concentrations. This study used data from 1 January 2022 to 31 December 2022, to guarantee a sufficient number of training samples and capture seasonal fluctuations.

2.2. Ground-Based $PM_{2.5}$ Monitoring Data

$PM_{2.5}$ data from roughly 78 monitoring stations in the Beijing–Tianjin–Hebei region (see Figure 1) were taken from the China National Environmental Monitoring Center (CNEMC) between 1 January and 31 December 2022. Data quality calibration and control were implemented, and measurements were collected using the automatic online monitoring methods—the tapered element oscillating microbalance (TEOM) method or the β-attenuation method—specified in the revised ambient air quality standards [43]. The ground-based $PM_{2.5}$ measurements had a temporal precision of one hour. We used the 3 h average values (from 10:00 to 12:00) as the dependent variable and omitted values below the detection limit and missing data (labeled as NA).

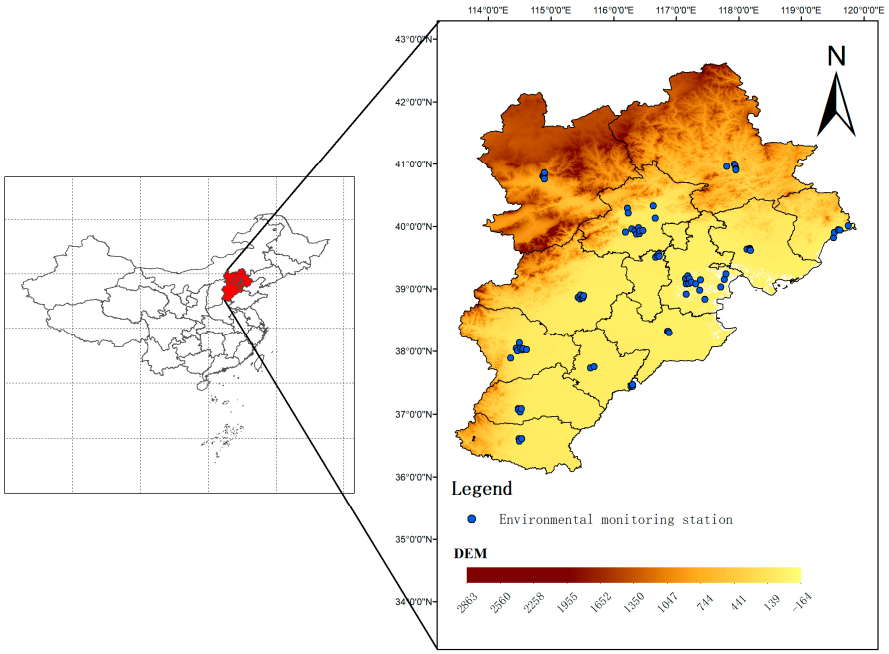

Figure 1. Geographical location and administrative division of the study area.

2.3. Meteorological Data and Auxiliary Data

The European Centre for Medium-Range Weather Forecasts (ECMWF) contributed the ERA-5 reanalysis dataset, which served as the source of the meteorological data used in this study. This dataset features a time resolution of one hour and a geographical resolution of 0.25° × 0.25°. We chose the boundary layer height, ground-level relative humidity, zonal wind (U), and meridional wind (V) from the ERA-5 meteorological data set for 2022. Based on the GF-5B satellite's overhead time of 10:30 am, hourly data for 11:00 am in Beijing time were collected. To ensure that the meteorological data's spatial resolution matched that of the satellite data, preprocessing techniques such as masking, reprojection, and TIFF conversion were used.

Auxiliary data included the digital elevation model (DEM), the normalized difference vegetation index (NDVI), and nighttime light data. Because plant leaves can absorb airborne particles and lower the ground-level $PM_{2.5}$ concentrations, the NDVI provides a measure of the health and coverage of vegetation [44,45]. The MOD13A2 16-Day product provided the NDVI data; they had a temporal resolution of 16 days and a spatial resolution of 1 km. The spatial distribution of human activities, which are strongly correlated with economic conditions and can represent population growth, urbanization, and industrial development, can be studied using data on nighttime light. The VIIRS dataset provided the nighttime light data, which had a 500 m spatial resolution. Google Earth Engine (GEE) was used to download data on nighttime light and the NDVI. The NASA Shuttle Radar Topography Mission (SRTM), which offers a DEM with a resolution of 30 m, was the source of the elevation data. The reprojection, masking, and resampling of these datasets were performed to guarantee that their geographic resolutions and coverage matched those of the satellite observation data.

All data parameters are shown in Table 1.

Table 1. Input variables used to estimate ground-level $PM_{2.5}$ concentrations.

Category	Variable	Abbreviation	Spatial Resolution
TOA reflectance (17 angles; P denotes polarization)	Red channel (670 nm, P)	R	3.3 km
	Blue channel (490 nm, P)	B	
	Green channel (565 nm)	G	
	Near-infrared channel (865 nm, P)	NIR	
Observation angles (17 angles)	Solar azimuth angle	SAA	
	Solar zenith angle	SZA	
	View azimuth angle	VAA	
	View zenith angle	VZA	
Meteorological data	Boundary layer height	blh	0.25°
	Relative humidity	RH	0.25°
	10 m U wind	u	0.25°
	10 m V wind	v	0.25°
Auxiliary data	Digital elevation model	DEM	30 m
	Normalized difference vegetation index	NDVI	1 km
	Nighttime light data	LIGHT	500 m

2.4. Data Processing

In order to match the spatial resolution of the GF-5B data (3.3 km), the multi-source auxiliary data were resampled. There were notable variations in the spatial resolution between the several data sources; hence, different resampling techniques were used [46]. For low-resolution data, bilinear interpolation was used. This method involves taking the four nearest pixel values and performing linear interpolation in both directions to produce a smooth interpolated result. For high-resolution data, the mean value of the grid points within each grid cell was calculated to resample the data to the same resolution as the satellite observations.

The multi-source satellite images and the daily $PM_{2.5}$ readings were matched after the spatial resolution was unified. Following the cropping of the satellite images to the same height and width, the daily $PM_{2.5}$ values were compared to the corresponding pixels from the multi-source satellite data in order to determine the values of various variables. As the dependent variable labels for the study, the matched $PM_{2.5}$ data were used.

3. Methods

Figure 2 shows the research roadmap. This study employed the gradient boosting machine (GBM) method (see Section 3.1) and validated the inversion accuracy through ten-fold cross-validation (see Section 3.2). First, an optimal set of parameters was selected for inversion using a systematic parameter tuning approach (see Section 4.2). Given that the

DPC payload provides up to 17 observation angles in a continuous sequence, we simplified the analysis by selecting five evenly spaced angles: angle 1, angle 5, angle 9, angle 13, and angle 17. We compared the inversion accuracy across different bands (red, green, blue, near-infrared) using the GBM model for each angle to identify the optimal angle and band. Using the best angle and band, we further compared the inversion accuracy when incorporating polarization data (I, Q) versus when they were excluded (I), to determine the impact of including polarization information. Finally, we conducted inversions using the optimal angle, band, and polarization data, resulting in the distribution of $PM_{2.5}$ in the Beijing–Tianjin–Hebei region at annual, seasonal, and monthly scales for 2022. The seasonal and annual results were then compared with data from the Himawari-8 satellite and CHAP datasets.

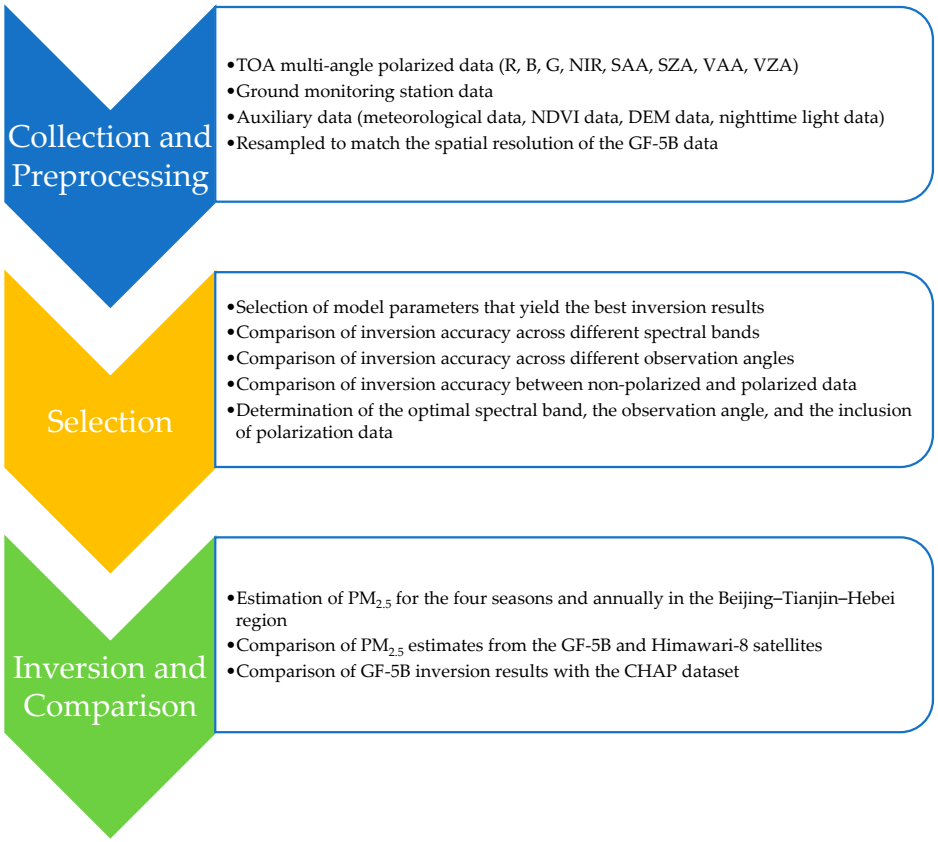

Figure 2. Process flow diagram for estimation of $PM_{2.5}$ concentrations proposed in this study.

3.1. Machine Learning Method

The gradient boosting machine (GBM) is a recently proposed regression analysis prediction model. As a predictive model in ensemble learning, it serves as an additive model using decision trees as base learners, first introduced by Friedman. The creation of new base learners that are maximally correlated with the loss function's negative gradient and linked to the ensemble as a whole is the main notion behind this technique [47,48]. Utilizing Classification and Regression Trees (CART) as its base model, GBM can handle various types of data and offers several advantages, including high predictive accuracy, minimal parameter tuning, robustness, parallel computation, and rapid training. This model effectively overcomes the limitations of linear regression, particularly when dealing with dense datasets and complex relationships among the data.

3.2. Validation Scheme

This study employed a ten-fold cross-validation method for model validation. Ten subgroups of an almost comparable size were created by randomly dividing the dataset.

One subset was used as the validation sample, while the other subsets were utilized for model fitting in each iteration, a process referred to as ten-fold cross-validation [49]. The accuracy evaluation metrics used in this study included the correlation coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE), expressed as follows:

$$R^2 = 1 - \frac{\Sigma_{i=1}^{N}(y_i - \hat{y}_i)^2}{\Sigma_{i=1}^{N}(y_i - \overline{y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (2)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \quad (3)$$

where $\overline{y}$ is the mean value of the ground-based $PM_{2.5}$ measurements, $\hat{y}$ is the value of the ground-based $PM_{2.5}$ measurements, and N is the number of data samples in the dataset.

4. Discussion

4.1. Descriptive Statistical Analysis

Table 2 presents the descriptive statistics for the input variables in the model. The World Health Organization's yearly average air quality standard of 5 $\mu g/m^3$ was greatly exceeded by the $PM_{2.5}$ concentrations, which ranged from 1 to 172 $\mu g/m^3$ on average [50]. This indicates that the $PM_{2.5}$ pollution is quite severe in the study area. Overall, the TOA reflectance observed at angle 9 was the lowest, while angles 1 and 17 recorded higher reflectance values. The average reflectance values in the R band for angles 1, 5, 9, 13, and 17 were 0.1325, 0.1196, 0.1116, 0.1187, and 0.1305, respectively. The patterns for the B, NIR, and G bands exhibited similar trends, with the TOA reflectance initially decreasing and then increasing as the angle varied from 1 to 17. The solar zenith angle (SZA) varied with the seasons, with the maximum values (around 65°) representing winter and the minimum values (around 14°) indicating summer. The solar azimuth angle (SAA) ranged from 120° to 190°, suggesting that the sun was predominantly positioned in the southeast–south direction relative to the observation points, with an average SAA of approximately 160°, coinciding with the GF-5B satellite's overpass time of about 10:30 local time. The viewing zenith angle (VZA) represents the angle between the line connecting the satellite and the observation point and the vertical normal. For angles 1, 5, 9, 13, and 17, the average VZA values were 59.1°, 44.8°, 31.5°, 44.9°, and 58.1°, respectively, indicating a correlation between the VZA magnitude and TOA reflectance values.

Table 2. Descriptive statistical analysis table of TOA variables.

Angle	Index	VZA	VAA	SZA	SAA	R	B	NIR	G
Angle 1	Max	82.630	359.970	65.890	186.710	0.250	0.312	0.701	0.275
	Min	51.450	0.020	14.460	128.120	0.037	0.071	0.030	0.048
	Average	59.098	240.189	37.554	158.891	0.133	0.160	0.185	0.133
	S.Dev	5.136	156.088	13.616	10.014	0.048	0.048	0.060	0.045
Angle 5	Max	79.590	359.950	65.840	187.560	0.250	0.301	0.396	0.283
	Min	28.760	0.010	14.420	128.590	0.034	0.048	0.018	0.036
	Average	44.787	270.772	38.010	159.575	0.120	0.136	0.173	0.120
	S.Dev	10.141	126.301	13.576	9.905	0.047	0.047	0.063	0.044
Angle 9	Max	76.780	358.640	65.780	188.200	0.250	0.306	0.394	0.315
	Min	0.250	2.280	14.400	129.080	0.026	0.042	0.010	0.032
	Average	31.520	273.399	38.136	160.048	0.112	0.124	0.169	0.113
	S.Dev	17.688	58.055	13.605	9.737	0.045	0.045	0.067	0.043

Table 2. *Cont.*

Angle	Index	VZA	VAA	SZA	SAA	R	B	NIR	G
Angle 13	Max	75.300	305.150	65.730	189.050	0.250	0.323	0.382	0.311
	Min	31.200	161.680	14.380	129.620	0.030	0.060	0.013	0.045
	Average	44.924	237.412	37.982	160.511	0.119	0.136	0.185	0.120
	S.Dev	8.643	27.732	13.598	9.697	0.042	0.042	0.065	0.040
Angle 17	Max	75.390	290.510	65.670	189.780	0.250	0.310	0.404	0.269
	Min	51.780	176.260	14.360	130.130	0.037	0.087	0.019	0.065
	Average	58.069	223.383	37.867	160.997	0.131	0.158	0.201	0.132
	S.Dev	7.295	1.836	1.454	0.244	0.003	2.048	2.306	0.051

4.2. Impact of Model Parameters on Inversion

During model training, a consistent random seed was used to shuffle the dataset, minimizing the errors in testing and analysis due to data variability. The optimization process focused on two key parameters, the number of decision trees and the maximum depth of each tree, aimed at enhancing the model's predictive accuracy. The number of trees was varied from 2 to 300 in increments of 2, while the maximum depth was adjusted from 2 to 30, also in increments of 2. As illustrated in the parameter optimization process (Figure 3), the model's predictive accuracy significantly increased as the number of trees and the maximum depth rose, particularly when the number of trees was below 14 and the depth was below 6. The computational time exhibited a roughly linear increase with the number of trees, while the relationship with the tree depth was approximately logarithmic. When the number of decision trees approached 200 and the maximum depth reached around 10, the overall model achieved a coefficient of determination (R^2) of approximately 0.927 and an RMSE of about 7.42 μg/m^3. Further increases in either parameter did not yield significant improvements in the model performance. With this configuration, the computational time was around 17 s. Therefore, in this study, the optimal model parameters were set to a decision tree count of 200 and a maximum depth of 10.

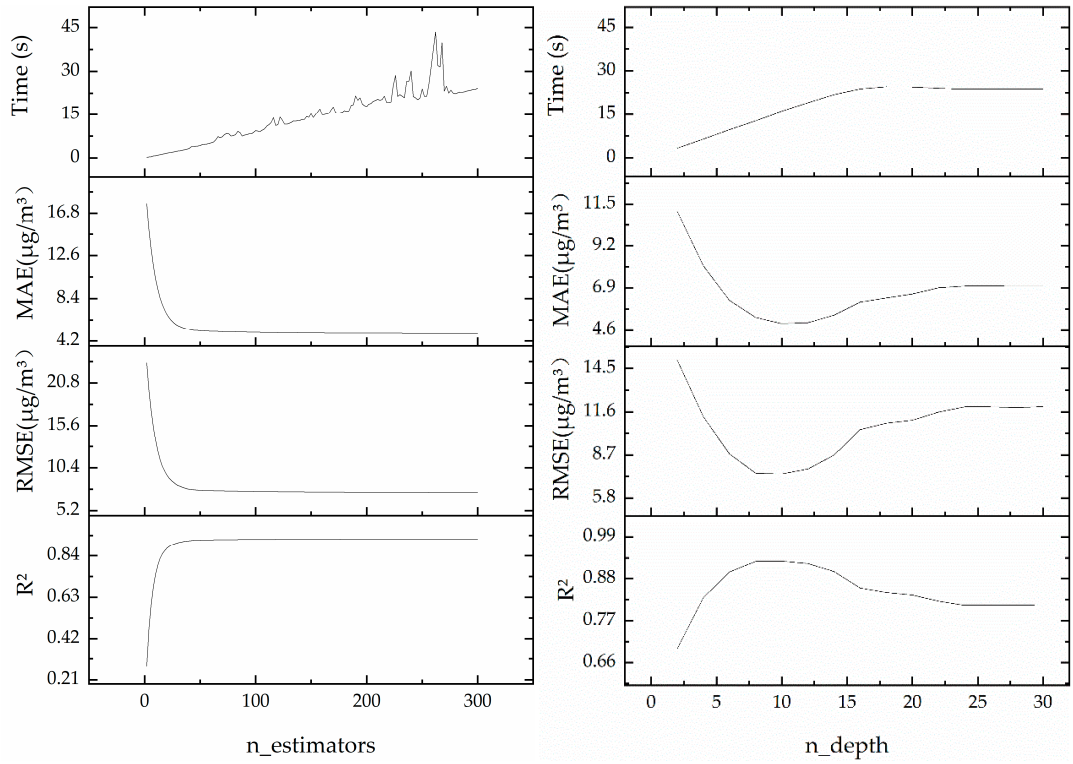

Figure 3. Impact of model parameters on inversion ((**Left**): n_estimators, (**Right**): n_depth).

4.3. Relative Importance Ranking of Model Variables

The model provides an importance ranking of the input variables across various angles (Figure 4). Among all features, their importance is ranked from highest to lowest as follows: TOA (47.9–50.9%), meteorological data (44.9–47.1%), and auxiliary data (4.2–5.3%). Notably, the solar zenith angle (SZA) alone accounts for 24.8–26.7% of the total importance, contributing approximately 50% to the TOA, indicating that the SZA is the most crucial parameter for $PM_{2.5}$ inversion. The overall relative importance of the different spectral bands ranges from 11.8% to 15.2%, with the red band (1.6–5.5%) and blue band (3.1–8.4%) showing substantial contributions, while the near-infrared band (1.4–2.5%) has a comparatively minor impact. Meteorological features have slightly lower overall importance than the TOA, with significant variability among the different meteorological parameters. Specifically, the boundary layer height (BLH) contributes 21.3–21.9%, while the relative humidity (RH) contributes 11.9–14.2%, highlighting their direct influences on the near-surface $PM_{2.5}$ concentrations. The importance of the three auxiliary parameters—DEM, NDVI, and LIGHT—remains relatively low across all seasons. However, removing these features slightly decreases the model's predictive accuracy, which justifies their retention in the predictions. In summary, the top three variables ranked by importance are the SZA, BLH, and RH, underscoring their critical roles in accurately estimating the $PM_{2.5}$ concentrations.

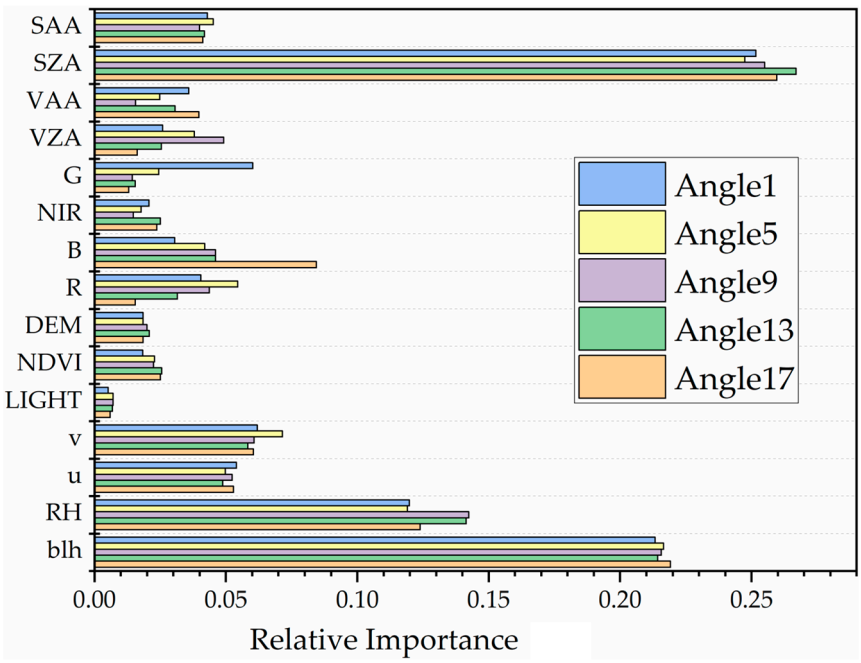

Figure 4. Relative importance ranking of the variables in the model.

4.4. Selection of Angles, Bands, and Polarization

To investigate the influence of different angles and spectral bands on the experimental results, we trained the GBM model using data from angles 1, 5, 9, 13, and 17 across the red, blue, near-infrared, and green bands, as well as a combination of these bands. Figure 5 shows the training results based solely on remote sensing data, without any auxiliary data (including meteorological data, the NDVI, nighttime light data, and digital elevation models). The results indicate that the variation in the R^2 values across different angles is minimal, within a range of 0.03, while the differences in the RMSE and MAE are also relatively small, at 1.02 $\mu g/m^3$ and 0.65 $\mu g/m^3$, respectively. Notably, the blue band yielded the best inversion performance, exhibiting the highest R^2 values (0.8826–0.8898) along with the lowest RMSE (9.03–9.31 $\mu g/m^3$) and MAE (5.63–5.81 $\mu g/m^3$). Conversely, the near-infrared band showed the poorest inversion performance, characterized by the lowest R^2 values (0.8638–0.8713) and the highest RMSE and MAE values (9.72–10.05 $\mu g/m^3$

and 6.12–6.29 μg/m^3, respectively). For the combinations of the four bands, the R^2, RMSE, and MAE values ranged from 0.8701 to 0.8805, 9.41 to 9.83 μg/m^3, and 5.83 to 6.09 μg/m^3, respectively. This suggests that combining bands does not enhance the inversion accuracy; the performance was only better than that of the near-infrared band, while, overall, it remained lower than that of the other three bands. In summary, the effect of different angles on the inversion performance was not significant in this analysis when only using remote sensing data.

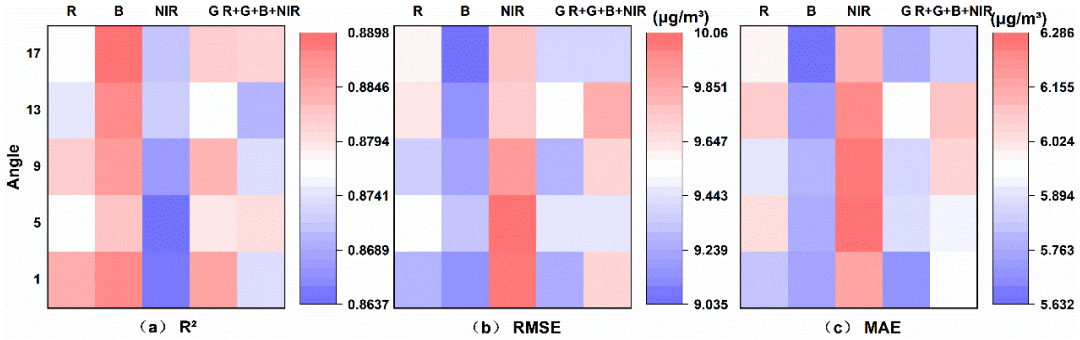

Figure 5. Impact of different angles and bands on inversion accuracy (only TOA remote sensing data).

The inversion accuracy achieved when using only auxiliary data (including meteorological data, DEM, NDVI, and nighttime light) was 0.78, slightly lower than the result obtained with pure remote sensing data (see Figure 6). Figure 7 shows the inversion results for different angles and spectral bands when both remote sensing and auxiliary data were utilized together. The differences in the R^2 values among the various angles were within 0.01, while the differences in the RMSE and MAE were limited to 0.31 μg/m^3 and 0.16 μg/m^3, respectively—much smaller than the variations observed with pure remote sensing data. Consistent with the previous analysis, the blue band demonstrated the best performance in inversion (R^2: 0.9234–0.9272, RMSE: 7.38–7.56 μg/m^3, MAE: 4.76–4.86 μg/m^3), while the near-infrared band yielded the poorest results (R^2: 0.9211–0.9228, RMSE: 7.59–7.67 μg/m^3, MAE: 4.84–4.89 μg/m^3). However, in this study, the impact of different angles on the inversion accuracy was significantly more pronounced. Specifically, angle 13 exhibited the best performance across the remaining four spectral bands, achieving optimal values in all metrics compared to angles 1, 5, 9, and 17, with the R^2, RMSE, and MAE ranging from 0.9222 to 0.9272, 7.36 to 7.61 μg/m^3, and 4.78 to 4.85 μg/m^3, respectively. This improvement may be attributed to the average VZA of approximately 45° at angle 13. Previous research has confirmed that the inversion performance is optimal around this angle when combining remote sensing and auxiliary data [38]. Thus, for subsequent inversions, we selected angle 13 and the blue band for our analysis.

Based on the inclusion of polarization component information, the data were divided into two schemes. Scheme 1 utilized only the I component information from GF-5B at angle 13, and Scheme 2 incorporated both the I and Q component information from GF-5B at angle 13. By comparing the polarization-enhanced Scheme 1 and Scheme 2, we explored the spatial impact of polarization data on $PM_{2.5}$ estimation. The cross-validation (CV) performance of the models for both schemes is illustrated in Figure 8. The sample size for both Scheme 1 and Scheme 2 was 9594. As shown in Figure 8, the CV R^2 and RMSE values for Scheme 1 were 0.9265 and 7.08 μg/m^3, respectively. For Scheme 2, the CV R2 and RMSE values were 0.9268 and 7.10 μg/m^3. The slopes indicated that we slightly underestimated the PM2.5 values, with slopes of 1.0239 for Scheme 1 and 1.0215 for Scheme 2. This suggests that the results from the GF-5B model incorporating polarization data are closer to the 1:1 line compared to those without polarization data. Overall, the inclusion of polarization components resulted in a minor improvement in the model's predictive performance; however, the difference in the R^2 was only in the order of one ten thousandth, and the RMSE for the model including polarization data was higher than that of the model

excluding them. Therefore, to simplify the retrieval, polarization data were not included in subsequent inversions.

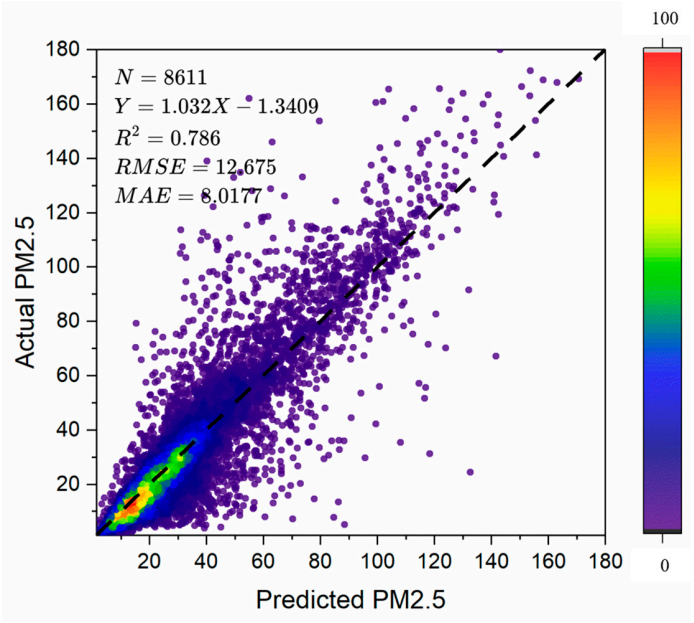

Figure 6. Inversion accuracy of $PM_{2.5}$ under auxiliary data.

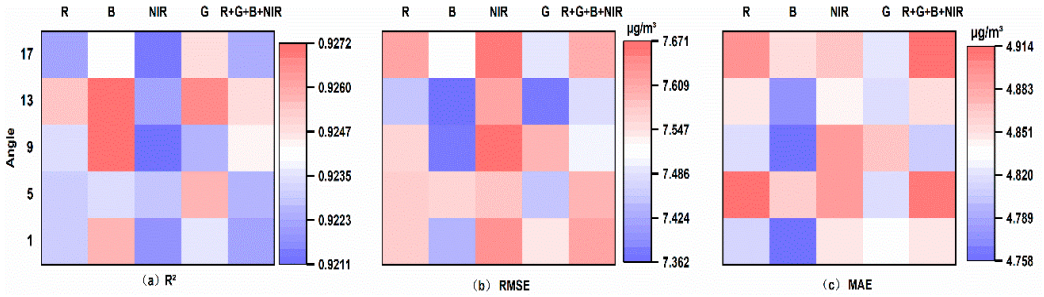

Figure 7. Impact of different angles and bands on inversion accuracy (TOA data combined with meteorological and auxiliary data).

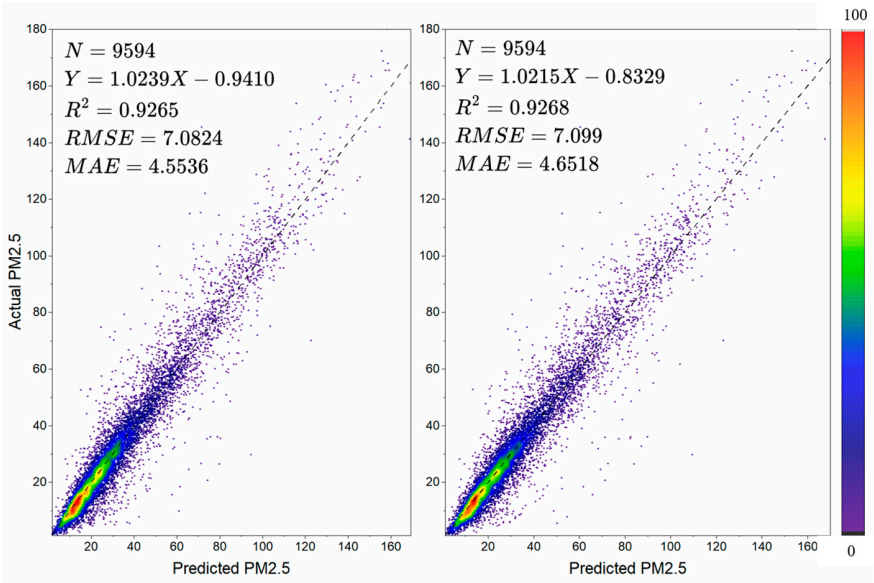

Figure 8. Comparison of inversion accuracy between Scheme 1 (**Left**) and Scheme 2 (**Right**).

5. Results

5.1. Estimated $PM_{2.5}$ Concentrations over Beijing–Tianjin–Hebei Region

Figure 9 shows the estimated monthly average spatial distribution of the $PM_{2.5}$ concentrations in China based on the GF-5B model. It is clear that there are notable seasonal fluctuations in the $PM_{2.5}$ concentrations in the Beijing–Tianjin–Hebei region. According to the rankings of the pollutant levels, January records the highest levels and June the lowest, with winter (January–February, December) > autumn (September–November) > spring (March–May) > summer (June–August). Winter is characterized by the highest $PM_{2.5}$ concentrations and the broadest pollution distribution, with the average simulated concentration reaching 37.45 $\mu g/m^3$. The high pollution levels are primarily concentrated in the central–southern part of Hebei Province. This is largely attributable to the concentration of major pollution-emitting industries such as steel, coke, and cement in this region, coupled with increased coal consumption for heating during the winter months. Additionally, the unique winter meteorological conditions contribute to the formation of an "inversion layer" near the surface, leading to the enhanced accumulation of pollutants and elevated $PM_{2.5}$ levels. The high-value areas in the northern parts of Hebei Province are influenced by the high albedo effect of snow cover. In spring, both the concentration and spatial extent of $PM_{2.5}$ significantly decrease compared to winter, with an average simulated concentration of 27.62 $\mu g/m^3$. Pollution is mainly concentrated in the central–southern part of Hebei Province, where a large population and relatively developed industrial activities contribute to higher $PM_{2.5}$ levels. In autumn, the areas with high $PM_{2.5}$ concentrations primarily extend along the Taihang Mountain range in the central–western part of Hebei Province, with a broader distribution than in spring, resulting in an average simulated concentration of 30.64 $\mu g/m^3$. Summer exhibits the lowest $PM_{2.5}$ concentrations and the smallest pollution extent of the year, with a simulated average concentration of only 21.85 $\mu g/m^3$. This is primarily due to vigorous air convection and active vertical atmospheric movement during the summer, which facilitates the dispersion of pollutants. Furthermore, the increased precipitation in summer promotes the wet deposition of $PM_{2.5}$ and other particulate matter, aiding in their removal from the atmosphere.

The annual average spatial distribution of the $PM_{2.5}$ concentrations in China, estimated using GF-5B, is illustrated in Figure 10a. The $PM_{2.5}$ annual concentration map reveals significant variability across the Beijing–Tianjin–Hebei region, with concentrations ranging from 18.16 to 40.82 $\mu g/m^3$. The spatial distribution of $PM_{2.5}$ pollution generally exhibits a pattern of higher concentrations in the southwest and lower concentrations in the northeast. The highest annual average $PM_{2.5}$ concentrations are primarily located in the urban agglomeration in central and southern Hebei Province, as well as along the Bohai Sea. The substantial population and frequent industrial activities in this area contribute to significant pollutant emissions, while the blocking effect of the Taihang Mountains in the west hinders the dispersion of pollutants. Next, the zone with the second-highest $PM_{2.5}$ concentration is found in Beijing, Tianjin, and their surrounding areas, where the pollution levels are lower compared to the high-value regions. This improvement may be attributed to the successful air quality management measures implemented during the 2022 Beijing Winter Olympic Games, which led to noticeable enhancements in the air quality. The areas with the lowest $PM_{2.5}$ concentrations are located in the northern highland regions of the study area, characterized by a low population density, smaller economic scales, minimal pollution emissions, and rich ecosystems that facilitate pollutant dissipation and deposition. Overall, there is a high degree of consistency between the estimated values and the ground monitoring data.

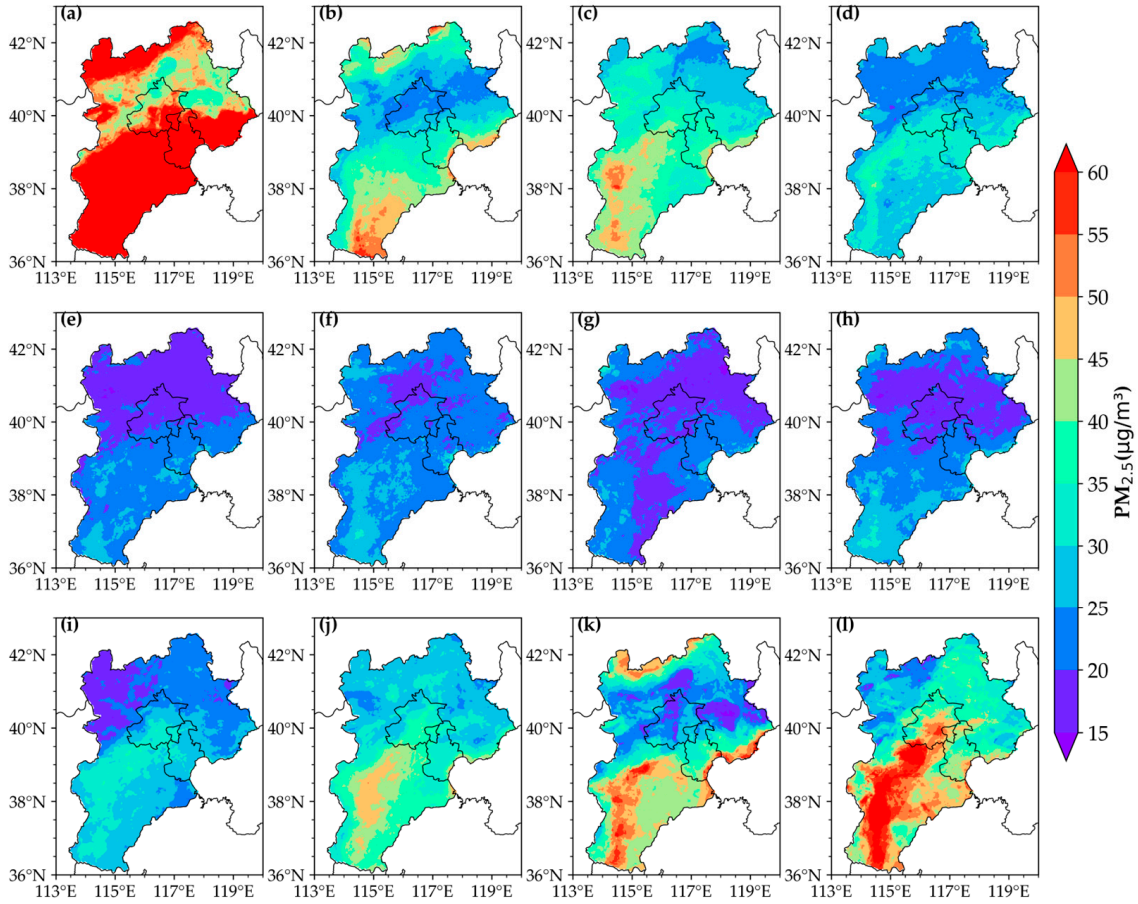

Figure 9. Monthly results of $PM_{2.5}$ estimation by GF-5B ((**a**–**l**) representing January to December).

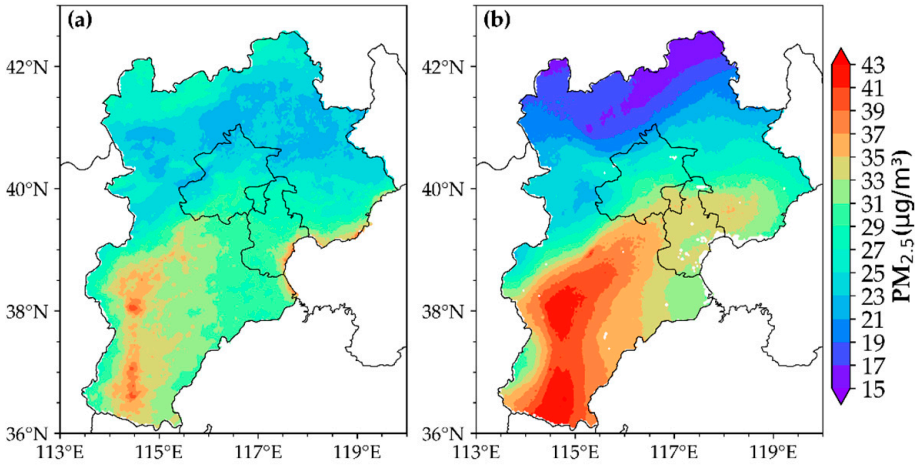

Figure 10. Comparison of annual average values between GF-5B (**a**) results and the CHAP dataset (**b**).

5.2. Comparison of Estimation Results with Himawari-8

As a comparative benchmark, we employed the Himawari-8 satellite to compare the performance of the GF-5B model. Himawari-8's Advanced Himawari Imager (AHI) monitors the brightness temperatures in ten infrared wavelengths and the top-of-atmosphere (TOA) reflectance in six visible/near-infrared wavelengths across East Asia [51]. We used the TOA reflectance from the Himawari-8 blue channel (0.46 μm) for the comparison, along with the accompanying solar zenith angle, solar azimuth angle, view zenith angle, and view

azimuth angle data. The data selection timeframe for Himawari-8 was daily at 10:30 BJT, from 1 January to 31 December 2022.

The GF-5B model dataset (N = 9594) had a slightly smaller sample size than the Himawari-8 model dataset (N = 12,109) due to differences in the spatial resolution and cloud masking products. As indicated in Figure 11, the quantitative evaluation based on sample cross-validation (CV) showed that GF-5B outperformed Himawari-8, with an R^2 of 0.93 and an RMSE of 7.08 μg/m^3. In the site-based CV, the GF-5B model was superior to the Himawari-8 model (0.91). Additionally, GF-5B's RMSE and mean absolute error (MAE) were 7.08 and 4.55 μg/m^3, respectively, both lower than the Himawari-8 values of 7.4 and 4.88 μg/m^3. Overall, the GF-5B model demonstrated superior predictive capabilities for $PM_{2.5}$ estimation.

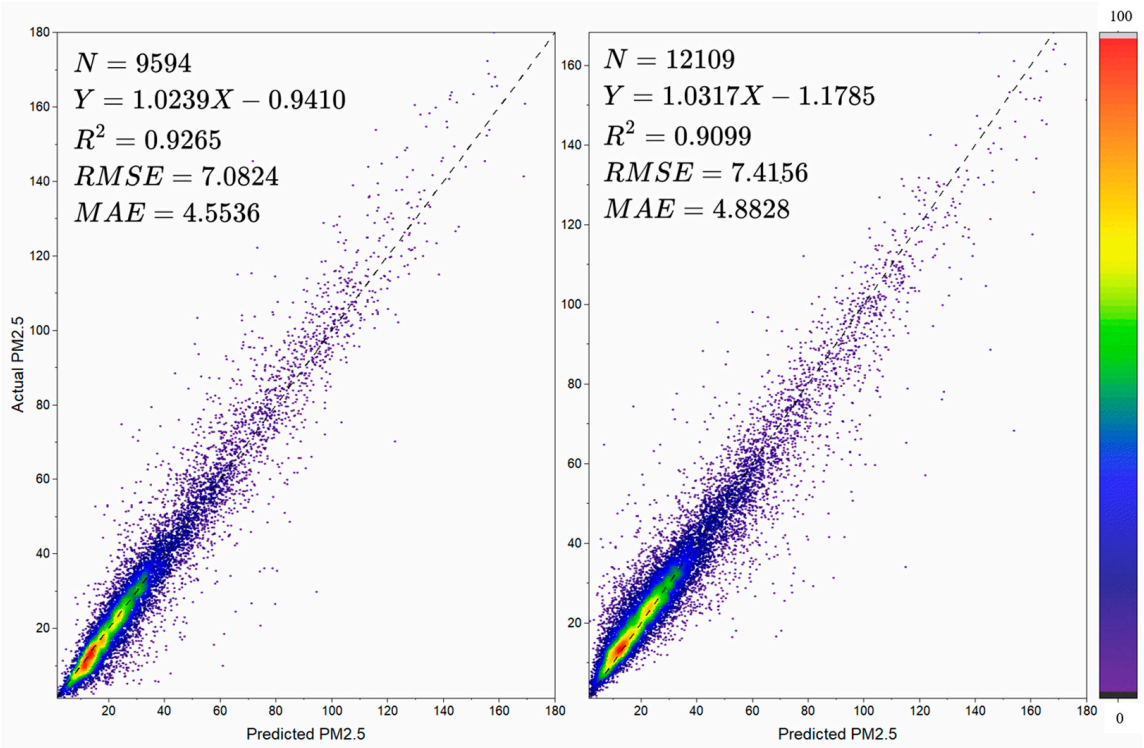

Figure 11. Comparison of estimation results between GF-5B (**left**) and Himawari-8 (**right**).

The spatial distribution of the CV R^2, RMSE, and MAE for GF-5B and Himawari-8 is illustrated in Figure 12. Overall, the use of GF-5B data for $PM_{2.5}$ estimation yielded satisfactory performance, with 80% of the grid cells reporting R^2 values greater than 0.90 and 94% of the grid cells showing RMSE values below 8 μg/m^3. In contrast, only 63% of the grid cells using Himawari-8 data achieved R^2 values above 0.90, while 89% had RMSE values below 8 μg/m^3. These statistics indicate that the GF-5B data have a distinct advantage over Himawari-8 in terms of the spatial prediction of $PM_{2.5}$. Figure 12 also reveals subtle spatial differences in the $PM_{2.5}$ estimations between the two satellites. For instance, Himawari-8 exhibits certain advantages in the Southern Hebei region compared to GF-5B, while GF-5B's performance is slightly better in the Beijing–Tianjin area. The more complex atmospheric pollution sources in the densely populated Beijing–Tianjin region contribute to this trend, suggesting that GF-5B has an advantage in predicting the $PM_{2.5}$ concentrations in areas with intricate pollution dynamics. Notably, the sites with relatively large prediction errors are situated in the northern part of the study area. This is primarily due to the sparse station distribution in the northern highland regions, which hampers the machine learning model's ability to fully capture the relationship between the $PM_{2.5}$ concentrations and their influencing factors.

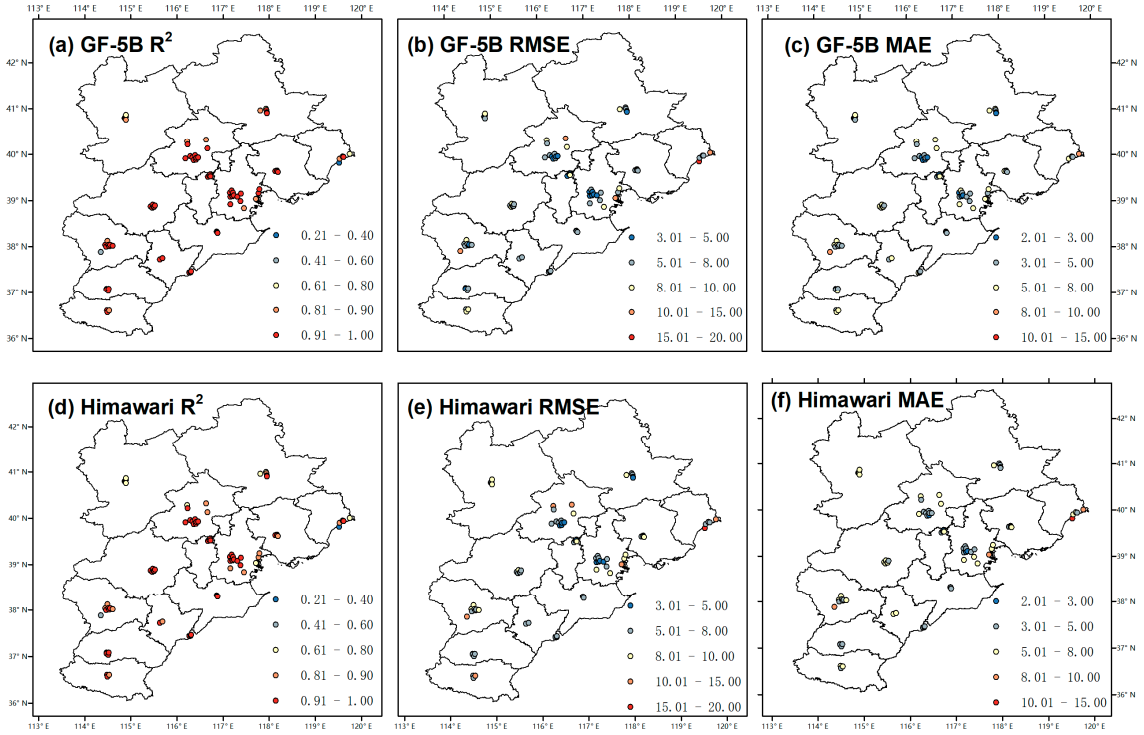

Figure 12. The maps of the R^2, RMSE, and MAE for various stations in the Beijing–Tianjin–Hebei region derived from GF-5B (**a**–**c**) and Himawari-8 (**d**–**f**).

5.3. Comparison of Inversion Results with the CHAP Dataset

The China High Air Pollutants (CHAP) dataset is currently the highest-resolution $PM_{2.5}$ dataset available in China [25,52]. Figure 10 illustrates the annual average $PM_{2.5}$ concentration distributions estimated using the GF-5B satellite data and the CHAP dataset. The overall estimates for the Beijing–Tianjin–Hebei region are similar, with the $PM_{2.5}$ annual average estimated by GF-5B at 27.90 $\mu g/m^3$ and by CHAP at 30.77 $\mu g/m^3$. The ground monitoring stations report an annual average concentration of 33.12 $\mu g/m^3$, indicating that the CHAP estimates closely align with the observed values. Notably, the CHAP dataset predicts higher $PM_{2.5}$ concentrations in most areas of the Beijing–Tianjin–Hebei region compared to the GF-5B data. In terms of the spatial source distribution, the results obtained from GF-5B inversion demonstrate a higher resolution, whereas the CHAP dataset reveals more clustered high-value areas. The high-value regions identified by GF-5B are relatively dispersed, appearing predominantly in the central and southern parts of Hebei Province, allowing for some observation of the pollution source distribution.

We compared the $PM_{2.5}$ retrieval results from GF-5B for the entirety of 2022 with those from the CHAP dataset. The absolute discrepancies between the measurements were, in order, −1.79, 54.65, −70.82, and 10.14 for the average, maximum, minimum, and standard deviation, respectively. The relative differences had the following values: −5%, 166%, −215%, and 31% for the average, maximum, minimum, and standard deviation, respectively. The correlation coefficient between the DPC $PM_{2.5}$ retrieval findings and the CHAP dataset was 0.81, as indicated in Figure 13, with relative accuracy of 5% and an absolute error of 1.79. Our results were underestimated compared to the CHAP dataset, with a scaling factor of 0.9437. Although the average $PM_{2.5}$ difference between GF-5B DPC and the CHAP dataset was −1.79, the high R^2 value indicates that our results were close to those of the CHAP dataset.

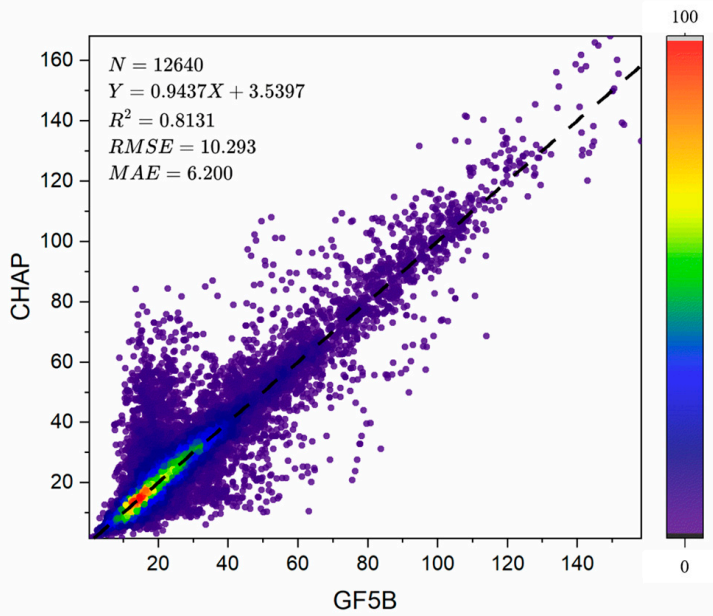

Figure 13. Scatter plot comparing GF-5B results with CHAP data..

6. Conclusions

The aim of this study was to estimate the surface $PM_{2.5}$ concentrations in the Beijing–Tianjin–Hebei region using TOA reflectance data from the GF-5B multi-angle polarized data, conducted within the framework of machine learning using GBM. This research confirms the feasibility of retrieving $PM_{2.5}$ concentrations from GF-5B's multi-angle polarized data. We accomplished two main tasks: first, we compared the $PM_{2.5}$ retrieval performance at different angles (Angle 1, Angle 5, Angle 9, Angle 13, and Angle 17) and across various spectral bands (blue, green, red, near-infrared), analyzing the influence of polarization on the retrieval performance and identifying the optimal parameters for $PM_{2.5}$ retrieval. Second, we examined the differences in the retrieval performance between GF-5B and the Himawari-8 and CHAP datasets under the optimal parameter conditions, highlighting the advantages and disadvantages of GF-5B in retrieving $PM_{2.5}$. The model results under the optimal conditions indicated that the $PM_{2.5}$ concentrations retrieved by GF-5B correlated well with the ground-based data, achieving an R^2 of 0.9272 and a root mean square error (RMSE) of 7.38 $\mu g/m^3$. GF-5B's retrieval accuracy was close to that of Himawari-8, and its results better illustrated the distribution of the pollution sources compared to the CHAP dataset, although the CHAP dataset was closer to the actual ground measurements.

The advantage of GF-5B lies in its ability to control the most crucial parameter in the $PM_{2.5}$ inversion model—the viewing zenith angle (VZA)—by selecting different angles. By choosing the optimal angle, the best results in the inversion process can be achieved. For future research, we propose three considerations: first, while analyzing the optimal observation angles, we selected five angles at equal intervals between Angle 1 and Angle 17; however, it is worth discussing whether the remaining angles might yield better retrieval results. Second, our study found that $PM_{2.5}$ retrieval was not obviously improved by the introduction of polarization observations, and we did not incorporate polarization data in our analysis. It would be valuable to explore how to use polarization data to enhance the $PM_{2.5}$ retrieval accuracy. Finally, we should fully utilize the strengths of different payloads for $PM_{2.5}$ retrieval. MODIS offers a spatial resolution of 1 km, Himawari-8 provides a temporal resolution of 1 h, and GF-5B's key advantage lies in its ability to obtain more aerosol physical information through multi-angle polarization. A crucial focus for future research will be how to effectively combine scalar data and multi-angle polarized data to improve the $PM_{2.5}$ retrieval accuracy.

7. Proclamation of AI-Assisted Generative Writing and AI-Supported Technologies

The writers employed ChatGPT during the writing process to enhance the readability and language. Following their use of this tool/service, the writers examined and made any necessary edits to the text, and they assume full responsibility for the publication's content.

Author Contributions: Conceptualization, Z.W. and R.Z.; methodology, Z.W. and R.Z.; validation, R.Z., R.C. and Z.W.; formal analysis, R.Z., C.Z. and H.C.; data curation, H.X.; writing—original draft preparation, R.Z.; writing—review and editing, R.C.; supervision, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by National Key Research and Development Programme of China (Grant 2021YFE0117100) and the National Natural Science Foundation of China (Grant No. 41971324).

Data Availability Statement: Data are contained within the article.

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

References

1. Dominici, F.; Peng, R.D.; Bell, M.L.; Pham, L.; McDermott, A.; Zeger, S.L.; Samet, J.M. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA J. Am. Med. Assoc.* **2006**, *295*, 1127–1134. [CrossRef]
2. Pope, C.A.; Dockery, D.W. Health effects of fine particulate air pollution: Lines that connect. *J. Air Waste Manag. Assoc.* **2006**, *56*, 709–742.
3. Pope Iii, C.A.; Burnett, R.T.; Thun, M.J.; Calle, E.E.; Krewski, D.; Ito, K.; Thurston, G.D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc.* **2002**, *287*, 1132–1141. [CrossRef]
4. Kan, H.D.; Chen, B.H. Meta-analysis of exposure–response functions of air particulate matter and adverse health outcomes in China. *J. Environ. Health* **2002**, *19*, 422–424.
5. Lelieveld, J.; Evans, J.; Fnais, M.; Giannadaki, D.; Pozzer, A. The contribution of out-door air pollution sources to premature mortality on a global scale. *Nature* **2015**, *525*, 367–371. [CrossRef]
6. Ma, Z.; Hu, X.; Sayer, A.M.; Levy, R.; Zhang, Q.; Xue, Y.; Tong, S.; Bi, J.; Huang, L.; Liu, Y. Satellite-based spatiotemporal trends in $PM_{2.5}$ concentrations: China, 2004–2013. Environ. *Health Perspect.* **2016**, *124*, 184–192. [CrossRef]
7. Yu, W.; Liu, Y.; Ma, Z.; Bi, J. Improving satellite-based $PM_{2.5}$ estimates in China using Gaussian processes modeling in a Bayesian hierarchical setting. *Sci. Rep.* **2017**, *7*, 7048. [CrossRef]
8. Huang, K.; Zhuang, G.; Li, J.; Wang, Q.; Sun, Y.; Lin, Y.; Fu, J.S. Mixing of Asian dust with pollution aerosol and the transformation of aerosol components during the dust storm over China in spring 2007. *J. Geophys. Res. Atmos.* **2010**, *115*. [CrossRef]
9. Zhang, X.Y.; Wang, J.Z.; Wang, Y.Q.; Liu, H.L.; Sun, J.Y.; Zhang, Y.M. Changes in chemical components of aerosol particles in different haze regions in China from 2006 to 2013 and contribution of meteorological factors. *Atmos. Chem. Phys.* **2015**, *15*, 12935–12952. [CrossRef]
10. Semlali, B.E.B.; El Amrani, C.; Ortiz, G.; Boubeta-Puig, J.; Garcia-de Prado, A. SAT-CEP-monitor: An air quality monitoring software architecture combining complex event processing with satellite remote sensing. *Comput. Electr. Eng.* **2021**, *93*, 107257.
11. Geng, G. Estimating long-term PM2.5 concentrations in China using satellite-based aerosol optical depth and a chemical transport model. *Remote Sens. Environ.* **2015**, *166*, 262–270. [CrossRef]
12. Lee, H.J.; Coull, B.A.; Bell, M.L.; Koutrakis, P. Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM 2.5 concentrations. *Remote Sens. Environ.* **2012**, *118*, 8–15.
13. Strawa, A.W.; Chatfield, R.B.; Legg, M.; Scarnato, B.; Esswein, R. Improving $PM_{2.5}$ Retrievals in the San Joaquin Valley Using A-Train Multi-Satellite Observations. *Atmos. Chem. Phys.* **2011**, *11*, 30563–30598.
14. Lee, H.J.; Liu, Y.; Coull, B.A.; Schwartz, J.; Koutrakis, P. A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. *Atmos. Chem. Phys.* **2011**, *11*, 7991–8002. [CrossRef]
15. Benas, N.; Beloconi, A.; Chrysoulakis, N. Estimation of urban PM_{10} concentration, based on MODIS and MERIS/AATSR synergistic observations. *Atmos. Environ.* **2013**, *79*, 448–454. [CrossRef]
16. Chu, D.A.; Kaufman, Y.J.; Zibordi, G.; Chern, J.D.; Mao, J.; Li, C.; Holben, B.N. Global monitoring of air pollution over land from the Earth Observing System-Terra Moderate Resolution Imaging Spectroradiometer (MODIS). *J. Geophys. Res. Atmos.* **2003**, *108*. [CrossRef]
17. Beckerman, B.S.; Jerrett, M.; Serre, M.; Martin, R.V.; Lee, S.-J.; van Donkelaar, A.; Ross, Z.; Su, J.; Burnett, R.T. A hybrid approach to estimating national scale spatiotemporal variability of $PM_{2.5}$ in the contiguous United States. *Environ. Sci. Technol.* **2013**, *47*, 7233–7241. [CrossRef]
18. Liu, Y.; Park, R.J.; Jacob, D.J.; Li, Q.; Kilaru, V.; Sarnat, J.A. Mapping annual mean ground-level $PM_{2.5}$ concentrations using Multiangle Imaging Spectroradiometer aerosol optical depth over the contiguous United States. *J. Geophys. Res. Atmos.* **2004**, *109*, 206.

19. van Donkelaar, A.; Martin, R.V.; Park, R.J. Estimating ground-level $PM_{2.5}$ using aerosol optical depth determined from satellite remote sensing. *J. Geophys. Res.-Atmos.* **2006**, *111*, D21. [CrossRef]
20. Van Donkelaar, A.; Martin, R.V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P.J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, 847–855. [CrossRef]
21. 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 $PM_{2.5}$. *Remote Sens. Environ.* **2015**, *156*, 117–128. [CrossRef]
22. Xue, W.; Wei, J.; Zhang, J.; Sun, L.; Che, Y.; Yuan, M.; Hu, X. Inferring Near-Surface $PM_{2.5}$ Concentrations from the VIIRS Deep Blue Aerosol Product in China: A Spatiotemporally Weighted Random Forest Model. *Remote Sens.* **2021**, *13*, 505. [CrossRef]
23. Wei, J.; Huang, W.; Li, Z.; Xue, W.; Peng, Y.; Sun, L.; Cribb, M. Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach. *Remote Sens. Environ.* **2019**, *231*, 111221. [CrossRef]
24. Wei, J.; Li, Z.; Pinker, R.T.; Wang, J.; Sun, L.; Xue, W.; Cribb, M. Himawari-8-derived diurnal variations in ground-level $PM_{2.5}$ pollution across China using the fast space-time Light Gradient Boosting Machine (LightGBM). *Atmos Chem Phys.* **2021**, *21*, 7863–7880. [CrossRef]
25. Wei, J.; Li, Z.; Lyapustin, A.; Sun, L.; Peng, Y.; Xue, W.; Su, T.; Cribb, M. Reconstructing 1-km-resolution high-quality $PM_{2.5}$ data records from 2000 to 2018 in China: Spatiotemporal variations and policy implications. *Remote Sens. Environ.* **2021**, *252*, 112136. [CrossRef]
26. Gündogdu, S.; Tuygun, G.T.; Li, Z.Q.; Wei, J.; Elbir, T. Estimating daily $PM_{2.5}$ concentrations using an extreme gradient boosting model based on VIIRS aerosol products over southeastern Europe. *Air Qual. Atmos. Health* **2022**, *15*, 2185–2198. [CrossRef]
27. Chen, D.; Gu, X.; Guo, H.; Cheng, T.; Yang, J.; Zhan, Y.; Fu, Q. Spatiotemporally continuous $PM_{2.5}$ dataset in the Mekong River Basin from 2015 to 2022 using a stacking model. *Sci. Total Environ.* **2024**, *914*, 169801. [CrossRef]
28. Tian, Z.Y.; Wei, J.; Li, Z.Q. How Important Is Satellite-Retrieved Aerosol Optical Depth in Deriving Surface $PM_{2.5}$ Using Machine Learning? *Remote Sens.* **2023**, *15*, 3780. [CrossRef]
29. Li, C.; Lau, A.K.H.; Mao, J.; Chu, D.A. Retrieval, validation, and application of the 1-km aerosol optical depth from MODIS measurements over Hong Kong. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2650–2658. [CrossRef]
30. Levy, R.C.; Mattoo, S.; Munchak, L.A.; Remer, L.A.; Sayer, A.M.; Patadia, F.; Hsu, N.C. The Collection 6 MODIS aerosol products over land and ocean. *Atmos. Meas. Tech.* **2013**, *6*, 2989–3034. [CrossRef]
31. Li, Z.; Zhao, X.; Kahn, R.; Mishchenko, M.; Remer, L.; Lee, K.-H.; Wang, M.; Laszlo, I.; Nakajima, T.; Maring, H. Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: A review and perspective. *Ann. Geophys.* **2009**, *27*, 2755–2770. [CrossRef]
32. Shen, H.; Li, T.; Yuan, Q.; Zhang, L. Estimating regional ground-level $PM_{2.5}$ directly from satellite Top-Of-Atmosphere reflectance using deep belief networks. *J. Geophys. Res.* **2018**, *123*, 13875–13886. [CrossRef]
33. Liu, J.; Weng, F.; Li, Z. Satellite-based $PM_{2.5}$ estimation directly from reflectance at the top of the atmosphere using a machine learning algorithm. *Atmos. Environ.* **2019**, *208*, 113–122. [CrossRef]
34. Yang, L.; Xu, H.; Yu, S. Estimating $PM_{2.5}$ concentrations in Yangtze River Delta region of China using random forest model and the Top-of-Atmosphere reflectance. *J. Environ. Manag.* **2020**, *272*, 111061. [CrossRef] [PubMed]
35. Wang, L.; Yu, C.; Cai, K.; Zheng, F.; Li, S. Retrieval of Aerosol Optical Depth from the Himawari-8 Advanced Himawari Imager data: Application over Beijing in the summer of 2016. *Atmos. Environ.* **2020**, *241*, 117788. [CrossRef]
36. Bai, H.; Zheng, Z.; Zhang, Y.; Huang, H.; Wang, L. Comparison of Satellite-based $PM_{2.5}$ Estimation from Aerosol Optical Depth and Top-of-atmosphere Reflectance. *Aerosol. Air Qual. Res.* **2021**, *21*, 200257. [CrossRef]
37. Wang, J.; Gao, K.; Hu, X.; Zhang, X.; Wang, H.; Hu, Z.; Yang, Z.; Zhang, P. $PM_{2.5}$ Estimation in Day/Night-Time from Himawari-8 Infrared Bands via a Deep Learning Neural Network. *Remote Sens.* **2023**, *15*, 4905. [CrossRef]
38. Mao, F.; Hong, J.; Min, Q.; Gong, W.; Zang, L.; Yin, J. Estimating hourly full-coverage $PM_{2.5}$ over China based on TOA reflectance data from the Fengyun-4A satellite. *Environ. Pollut.* **2021**, *270*, 116119. [CrossRef]
39. Choi, H.; Park, S.; Kang, Y.; Im, J.; Song, S. Retrieval of hourly $PM_{2.5}$ using top-of-atmosphere reflectance from geostationary ocean color imagers I and II. *Environ. Pollut.* **2023**, *323*, 121169. [CrossRef]
40. Jin, S.; Ma, Y.; Huang, Z.; Huang, J.; Gong, W.; Liu, B.; Wang, W.; Fan, R.; Li, H. A comprehensive reappraisal of long-term aerosol characteristics, trends, and variability in Asia, Atmos. *Chem. Phys.* **2023**, *23*, 8187–8210. [CrossRef]
41. Jin, S.; Ma, Y.; Chen, C.; Dubovik, O.; Hong, J.; Liu, B.; Gong, W. Performance evaluation for retrieving aerosol optical depth from the Directional Polarimetric Camera (DPC) based on the GRASP algorithm, Atmos. *Meas. Tech.* **2022**, *15*, 4323–4337. [CrossRef]
42. Zhang, R.; Zhou, W.; Chen, H.; Zhang, L.; Zhang, L.; Ma, P.; Zhao, S.; Wang, Z. Aerosol Information Retrieval from GF-5B DPC Data over North China Using the Dark Dense Vegetation Algorithm. *Atmosphere* **2023**, *14*, 241. [CrossRef]
43. HJ618-2011, 2011. Determination of Atmospheric Articles PM_{10} and $PM_{2.5}$ in Ambient Air by Gravimetric Method. Available online: https://english.mee.gov.cn/Resources/standards/Air_Environment/air_method/201111/t20111101_219390.shtml (accessed on 8 September 2011).
44. Nowak, D.J.; Crane, D.E.; Stevens, J.C. Air pollution removal by urban trees and shrubs in the United States. *Urban For. Urban Green.* **2006**, *4*, 115–123. [CrossRef]
45. Pugh, T.A.M.; MacKenzie, A.R.; Whyatt, J.D.; Hewitt, C.N. Effectiveness of green infrastructure for improvement of air quality in urban street canyons. *Environ. Sci. Technol.* **2012**, *46*, 7692–7699. [CrossRef]

46. Ngo, T.X.; Pham, H.V.; Phan, H.D.; Nguyen, A.T.; To, H.T.; Nguyen, T.T. A daily and complete $PM_{2.5}$ dataset derived from space observations for Vietnam from 2012 to 2020. *Sci. Total Environ.* **2023**, *857 Pt 3*, 159537. [CrossRef]
47. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
48. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorob.* **2013**, *7*, 21. [CrossRef]
49. Rodriguez, J.D.; Perez, A.; Lozano, J.A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 569–575. [CrossRef]
50. *Global Air Quality Guidelines*; World Health Organization (WHO): Geneva, Switzerland, 2021.
51. Zou, R.; Huang, H.; Lu, X.; Zeng, F.; Ren, C.; Wang, W.; Zhou, L.; Dai, X. PD-LL-Transformer: An Hourly $PM_{2.5}$ Forecasting Methodover the Yangtze River Delta Urban Agglomeration, China. *Remote Sens.* **2024**, *16*, 1915. [CrossRef]
52. Wei, J.; Li, Z.; Cribb, M.; Huang, W.; Xue, W.; Sun, L.; Guo, J.; Peng, Y.; Li, J.; Lyapustin, A.; et al. Improved 1 kmresolution $PM_{2.5}$ estimates across China using enhancedspace-time extremely randomized trees. *Atmos. Chem. Phys.* **2020**, *20*, 3273–3289. [CrossRef]

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