



Article

Spatiotemporal Weighted for Improving the Satellite-Based High-Resolution Ground PM_{2.5} Estimation Using the Light Gradient Boosting Machine

Xinyu Yu ^{1,2,3,†}, Mengzhu Xi ^{1,2,3,†}, Liyang Wu ^{1,2,3,*}  and Hui Zheng ^{1,2,3}

¹ College of Geography and Environmental Science, Henan University, Kaifeng 475004, China; yuxinyu@henu.edu.cn (X.Y.); xmzz@henu.edu.cn (M.X.); zhenghui@vip.henu.edu.cn (H.Z.)

² Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Henan University), Ministry of Education, Kaifeng 475004, China

³ Henan Key Laboratory of Integrated Air Pollution Control and Ecological Security, Kaifeng 475004, China

* Correspondence: wuly@henu.edu.cn

† These authors contributed equally to this work.

Abstract: Surface fine particulate matter (PM) with a diameter of less than 2.5 microns (PM_{2.5}) negatively impacts human health and the economy. However, due to data and model limitations, obtaining high-quality, high-spatial-resolution surface PM_{2.5} concentration data is a challenging task, and it is difficult to accurately assess the temporal and spatial changes in PM_{2.5} levels at a small regional scale. Here, we combined multi-angle implementation of atmospheric correction (MAIAC) aerosol products, ERA5 reanalysis data, etc., to construct an STW-LightGBM model that considers the spatiotemporal characteristics of air pollution and estimate the PM_{2.5} concentration of China's surface at 1 km resolution from 2015 to 2020. Our model performed well, and the fitting accuracy of the 10-fold cross-validation between years was 0.877–0.917. The fitting accuracy of the model was >0.85 at different time scales (month, season, and year). The average slope of the regression prediction was 0.9 annually. The results showed that PM_{2.5} pollution improved from 2015 to 2020. The average PM_{2.5} concentration decreased by 4.55 µg/m³, and the maximum decrease reached 90.51 µg/m³. The areas with high PM_{2.5} concentrations were predominantly in the North China Plain, Sichuan Basin, and Xinjiang in the west, and the levels in areas with elevated PM_{2.5} levels were consistent across most study years. The standard deviation ellipse for PM_{2.5} in China showed a 'northeast–southwest' spatial distribution. From an interannual perspective, the average values of the four seasonal stations in the country showed a downward trend from 2015 to 2020, with the most obvious decline in winter, from 70.67 µg/m³ in 2015 to 46.75 µg/m³ in 2020. Compared to earlier inversion studies, this work provides a more stable and accurate method for obtaining high-resolution PM_{2.5} data, which is necessary for local air governance and environmental ecological construction at a fine scale.



Citation: Yu, X.; Xi, M.; Wu, L.; Zheng, H. Spatiotemporal Weighted for Improving the Satellite-Based High-Resolution Ground PM_{2.5} Estimation Using the Light Gradient Boosting Machine. *Remote Sens.* **2023**, *15*, 4104. <https://doi.org/10.3390/rs15164104>

Academic Editor: Simone Lolli

Received: 25 June 2023

Revised: 15 August 2023

Accepted: 19 August 2023

Published: 21 August 2023

Keywords: aerosol optical depth; STW-LightGBM; surface PM_{2.5} estimation; air pollution; spatiotemporal variation

1. Introduction

Globally, 3.3 million people die annually from diseases caused by outdoor particulate matter (PM) with a diameter of less than 2.5 microns (PM_{2.5}), and more than a quarter of these premature deaths occur in China [1]. With the in-depth implementation of the “Air Pollution Prevention Plan” and the “3-year Action Plan to Win the Blue Sky Defense War”, ambient air quality in China has improved annually, and the average PM_{2.5} concentration in the country has decreased significantly [2]. However, there are still some areas where the average annual PM_{2.5} concentration exceeds the World Health Organization's medium-term target of 35 µg/m³ [3]. Long-term exposure to air pollution not only increases the risk of cardiovascular and cerebrovascular diseases [4–7] but also creates financial pressure due



Copyright: © 2023 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/>).

to the consequent medical and health expenditures [8–10]. From 2016 to 2017, the additional direct medical expenses related to urban workers in China due to short-term exposure to atmospheric PM_{2.5} were as high as CNY 3.68 billion [11]. In addition to direct medical costs, the reduction in labor productivity due to PM_{2.5} exposure also has a significant indirect economic impact [12]. Therefore, the timely and accurate prediction of PM_{2.5} mass concentrations could help people to effectively avoid the harm of air pollutants, and this is one of the most important strategies for caring for the aging population of China and facilitating high-quality economic development in the country [13].

Currently, research on the simulation of surface PM_{2.5} can be roughly divided into two categories: temporal and spatial variation analyses based on ground stations and temporal and spatial variation analyses based on various satellite remote-sensing datasets. Environmental quality monitoring sites are the most direct sites for analyzing PM_{2.5} [14,15]. However, due to the small number of ground monitoring stations in China and their weak spatial representation, the use of ground monitoring stations alone is not sufficient to fully reflect the spatial characteristics of PM_{2.5} [16]. Compared with ground monitoring, remote-sensing monitoring is characterized by high spatial resolution, wide monitoring range, fast speed, low cost, and all-weather real-time monitoring, which can thus overcome the shortcomings that occur due to the lack of or uneven spatial and temporal distributions of ground monitoring data, and provide new ideas for the study of PM_{2.5} [17–19]. Remote-sensing monitoring of PM_{2.5} mainly uses aerosol optical depth (AOD) and meteorological data as independent variables to establish a regression model to estimate PM_{2.5}. He et al. used the improved geographical and time-weighted regression (GTWR) model to estimate the daily PM_{2.5} concentration data at a 3 km resolution in mainland China [19]. Wang et al. estimated hourly PM_{2.5} concentrations in the Beijing–Tianjin–Hebei (BTH) region from Himawari-8 AOD products using a linear mixing effect (LME) model [20]. Xue et al. proposed an improved geographical and time-weighted regression (IGTWR) model using Himawari-8 AOD products to produce hourly PM_{2.5} datasets for central and eastern China [21]. However, limited by the uneven optical properties and significant nonlinear characteristics of AOD [22,23] and PM_{2.5} monitoring datasets, these regression models may not fully capture the complex spatial and temporal relationships that exist between PM_{2.5} and AOD when evaluated at a large scale. The prediction accuracy is thus not currently ideal, and the data products produced are insufficient to support specific regional research.

Machine learning and deep learning algorithms are providing new ways to solve such problems. Deep learning, which is a branch of machine learning, is widely used in ground object classification. However, compared with machine learning, deep learning is more dependent on computer performance and has a lower computational efficiency [24]; therefore, it is not widely used for PM_{2.5}. Conversely, machine learning algorithms such as Random Forest (RF), Gradient Boosting Tree (GBDT), and Extreme Gradient Boosting Tree (XGBoost) have been widely used to estimate local and global ground PM_{2.5} concentrations [25–27]. RF has a good training speed and prediction accuracy, although it is prone to overfitting with noisy classifications or regression problems. With GBDT, data processing flexibility is improved, and the parameter adjustment time is reduced [28]; however, this decision-tree-based learner does not support the parallel training of data. XGBoost reduces the occurrence of overfitting and shortens the operation time; however, many model parameters exist, and the adjustment process is complicated [29]. Furthermore, the latter model requires a large amount of computer memory, which affects the efficiency of fine spatiotemporal inversion research for PM_{2.5}. In contrast to the above models, the Light Gradient Boosting Machine (LightGBM) model not only has low memory requirements and high computational efficiency but also has high levels of accuracy. It also supports parallel processing and can be run on a GPU, which effectively reduces the burden of computer memory and improves the efficiency and accuracy of large-scale data processing [30–32].

To reduce the influence of the spatial and temporal heterogeneity of ground PM_{2.5} concentrations, this study employs a combination of spatial and temporal parameters and an improved GBDT algorithm proposed by He et al. [33] to construct an improved

LightGBM model (STW-LightGBM) using a spatiotemporal-weighted method. The fitting accuracy of STW-LightGBM to surface PM_{2.5} was verified by combining the PM_{2.5} data from ground environmental monitoring stations, satellite optical depth products (AOD), meteorological data, and land-use data. The stability at different time scales was also analyzed. Finally, a high spatial resolution (1 km) PM_{2.5} concentration map of China from 2015 to 2020 was created to provide a reference for the implementation of carbon emissions reduction programs in China.

2. Datasets and Processing

2.1. In Situ PM_{2.5} Measurements

Hourly data for the ground PM_{2.5} concentrations from 2015 to 2020, were obtained from the China National Environmental Monitoring Centre (CNEMC), <http://www.cnemc.cn/>, accessed on 26 June 2022). The conical element oscillation microbalance method or attenuation monitor was used for on-site concentration measurements and was strictly controlled according to the China Environmental Quality Standard (CNAQS, GB3905-2012) [34,35]. The PM_{2.5} data used were the daily averages from 1 January to 31 December for each year. If there was a missing value for the concentration on a certain day, it was ignored, and the daily average was calculated. The spatial distribution of PM_{2.5} ground monitoring stations in mainland China is detailed in Supplementary Materials Figure S1.

2.2. MAIAC AOD

The 1 km resolution multi-angle atmospheric correction algorithm MAIAC AOD used in this study was derived from the MODIS L1 B data product (<https://ladsweb.nascom.nasa.gov/>, accessed on 26 June 2022). The MAIAC algorithm uses a time-series method to dynamically obtain the surface reflection relationship between the blue and shortwave infrared bands of MODIS in dark and bright regions [36] and can improve detection in cloudy and snowy weather [37]. Compared with the commonly used MODIS DB and DT algorithms, MAIAC AOD inversion has a wider range, higher resolution, and higher inversion accuracy [38,39]. The AOD products of MAIAC also have higher accuracy and smaller statistical errors than those of aerosol monitoring stations in China [40].

Due to the different observation periods of Terra and Aqua satellites equipped with MODIS sensors, we referred to earlier studies [41] and used the linear regression method for matching grid cells between Terra and Aqua AODs, which effectively reduces the loss of AOD data.

2.3. Auxiliary Data

Meteorological data were included in the model (<https://cds.climate.copernicus.eu/cdsapp#!/home>, accessed on 26 June 2022). Seven meteorological variables from the ERA5 reanalysis data were used: atmospheric boundary layer height (BLH), ground pressure (SP), surface temperature above 2 m (T2M), relative humidity (RH), normalized difference vegetation index (NDVI), total precipitation (Pst), and wind speed (Ws). The spatial resolution of the data was $0.25^\circ \times 0.25^\circ$ and the temporal resolution was 1 h. The monthly NDVI had a 1 km resolution and annual land-use data with a 500 m spatial resolution were obtained from MODIS L3 global products. The specific data and sources used are shown in Supplementary Materials Table S1.

2.4. Data Processing

First, the nearest neighbor method was used to resample the raster image data to 1 km. ArcGIS10.2 was then used to create a national 1 km \times 1 km resolution grid and the grid image data were matched. If multiple ground air quality monitoring stations were present in a grid, the average value for all stations was calculated as the PM_{2.5} concentration value of the unit grid. The matching results were sorted, and the missing values and outliers were deleted. In traditional machine learning models, to avoid the influence of the dimensions of variables on the research process, it is usually necessary to normalize or discretize the data

to make different variables comparable. Our model is mainly based on the information gain ratio of the dataset about the feature when the node is split and is not sensitive to the feature value, hence, there is no need to normalize and discretize the data.

3. Models and Methods

3.1. STW-LightGBM Model

LightGBM is a distributed gradient boosting framework based on the decision tree algorithm [30], which improves the slow serial speed of GBDT and its tendency to overfit. LightGBM, provided by Microsoft, is an efficient implementation of GBDT with several advantages over other implementations of GBDT such as XGBoost [42,43]. The negative gradient of the loss function is used as a residual approximation of the current decision tree to fit a new decision tree. LightGBM combines GOSS and EFB algorithms based on GBDT and has the following advantages: faster training efficiency, low memory usage, higher accuracy, parallel learning support, and an ability to handle large-scale data and support the direct use of category features without increasing the complexity of the model while ensuring accuracy [26,27,38]. Specifically, in the estimation of surface PM_{2.5}, LightGBM can easily achieve higher accuracy with fewer sample features, less memory, and faster speed than other gradient enhancement algorithms.

Considering the complex surface environment in China, if only the results of ground stations are considered when constructing and fitting a model for air pollution concentrations, then the spatial representation of the data will be insufficient, and the prediction results will be unreliable. To improve the accuracy and reliability of the data, we referred to earlier research [39] and introduced the spatiotemporal characteristic parameter (Pst), which we used to build the STW-LightGBM model, which can more accurately reflect the spatiotemporal variation laws and future development trends for PM_{2.5}.

For a given sample i , the spatiotemporal characteristic parameters can be represented as follows:

$$P_{St(i)} = \sum_{j=1}^m w_{ij} \text{PM}_{2.5} / \sum_{j=1}^m w_{ij} \quad (1)$$

$$w_{ij} = \exp\left(-\frac{(d_{ij}^s)^2 + \rho * (d_{ij}^t)^2}{h_{st}^2}\right) \quad (2)$$

where w_{ij} represents the weight of sample j relative to sample i and d^s and d^t represent the space and time parameters, respectively, h_{st} represents the temporal and spatial bands, ρ represents time and space control factors, and m is the number of neighbors around sample point i . According to the method adopted by He et al. [33], h_{st} and ρ were set to 2 and 6000, respectively.

The Haversin distance was used to calculate the spatial parameters in the formula. The parameter calculation methods for the time and space distances are shown in Equations (3)–(8):

$$Space_w = f(Lon, Lat) = haversin(\alpha_1 - \alpha_2) + \cos \alpha_1 \times \cos \alpha_2 \times haversin(\beta_1 - \beta_2) \quad (3)$$

$$\alpha = Lon \times \frac{\pi}{180} \quad (4)$$

$$\beta = Lat \times \frac{\pi}{180} \quad (5)$$

$$haversin(\alpha) = \sin^2\left(\frac{\alpha}{2}\right) \quad (6)$$

$$d^s = 2 \times r \times \arcsin\left(\sqrt{Space_w}\right) \quad (7)$$

$$d^t = \cos\left(2\pi \frac{DOY}{Year}\right) \quad (8)$$

where Lon and Lat stand for pixel longitude and pixel latitude, respectively, d^T is the distance between two points determined by the latitude and longitude of the great circle, r is the radius of Earth, DOY indicates the day of a year, and $Year$ indicates the total number of days in the year. Figure 1 shows the structure of the PM2.5 estimation framework constructed in this current study.

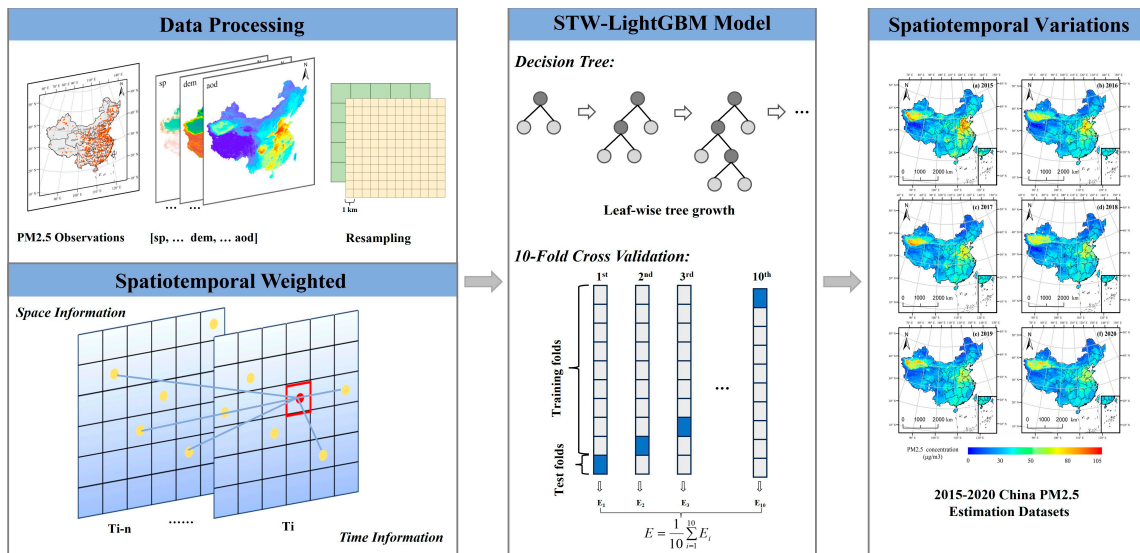


Figure 1. Flowchart of the mapping process of the PM2.5 dataset in our study.

3.2. Parameter Selection

In this study, we adopted manual parameter adjustment and grid optimization to determine the appropriate parameter combination (Supplementary Materials Figure S2). By setting different numerical ranges for three parameters, namely, learning rate (LR), maximum depth (MD), and number of leaf nodes (NL), we performed traversal calculations to determine the optimal model parameter combination. The data from 2016 to 2020 were used as the training set, and the data from 2015 were used as the test set to determine the optimal parameter combination. According to the cycle results, the optimal LR, MD, and NL values were determined to be 0.1, 10, and 375, respectively.

3.3. Standard Deviation Ellipse

The standard deviation ellipse (SDE) can accurately reveal the overall characteristics of the spatial distribution of geographical elements [44,45]. Based on the spatial layout and structure of the research object, the characteristics of the spatial distribution of geographical elements, such as centrality, direction, and spatial form, can be quantitatively explained, and the spatial distribution and spatiotemporal evolution process of geographical elements can be revealed. Based on the site-based annual average PM2.5 concentration, the SDE of the PM2.5 spatial distribution was calculated to show the spatial variation trend for air pollution in China over the past 6 years. The change in the length of the primary and secondary axes and the change in the azimuth angle of the standard deviation ellipse can characterize the increasing or decreasing trend of PM2.5, coverage, and extension direction. The oblateness is the ratio of the long and short half axes and the larger the oblateness, the stronger the sense of data direction, and the more obvious the centrifugal force. At the same time, this also implies that there is a greater degree of data dispersion.

In this study, standard error ellipse images of PM2.5 observations and the predicted values for many years were drawn. Through superposition and comparison, the accuracy of the model prediction could be verified more intuitively from a spatial perspective. The calculation formulas for the main parameters of the standard error ellipse were as follows:

$$P(\bar{X}, \bar{Y}) = \left| \frac{\sum_{i=1}^n \omega_i X_i}{\sum_{i=1}^n \omega_i}, \frac{\sum_{i=1}^n \omega_i Y_i}{\sum_{i=1}^n \omega_i} \right| \quad (9)$$

$$\tan \theta = \frac{A + B}{C} \quad (10)$$

$$A = \left(\sum_{i=1}^n \tilde{X}_i^2 - \sum_{i=1}^n \tilde{Y}_i^2 \right) \quad (11)$$

$$B = \sqrt{\left(\sum_{i=1}^n \tilde{X}_i^2 - \sum_{i=1}^n \tilde{Y}_i^2 \right)^2 + 4 \left(\sum_{i=1}^n \tilde{X}_i \tilde{Y}_i \right)^2} \quad (12)$$

$$C = 2 \sum_{i=1}^n \tilde{X}_i \tilde{Y}_i \quad (13)$$

$$\sigma_x = \sqrt{\sum_{i=1}^n (\omega_i \tilde{X}_i \cos \theta - \omega_i \tilde{Y}_i \sin \theta)^2} / \sum_{i=1}^n \omega_i^2 \quad (14)$$

$$\sigma_y = \sqrt{\sum_{i=1}^n (\omega_i \tilde{X}_i \sin \theta - \omega_i \tilde{Y}_i \cos \theta)^2} / \sum_{i=1}^n \omega_i^2 \quad (15)$$

where $P(\bar{X}, \bar{Y})$ represents the center of gravity of the standard deviation ellipse, ω_i represents weight, (X_i, Y_i) represents the weighted average center, θ represents the azimuth angle of the ellipse, \tilde{X}_i and \tilde{Y}_i represent the coordinate deviations from the location of each research object to the average center, respectively, and σ_x and σ_y represent the standard deviation along the X-axes and Y-axes, respectively.

3.4. Model Verification

A 10-fold cross-validation method was used to perform interannual and overall evaluation training on the model. For all samples, time- and site-based methods were combined to explore the universality of the model in time and space and to verify its robustness and stability [46,47]. The determination coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE) were used to evaluate the fitting results. The formulas used were as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \text{avg}(y))^2} \quad (16)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

4. Results

4.1. Model Fitting Performance and Overall Evaluation

The fitting ability of STW-LightGBM to PM2.5 on an interannual scale is shown in Figure 2. The fitting method uses the 10-fold cross-validation idea commonly utilized in machine learning. The CV- R^2 of the model fitting results for each year from 2015 to 2020 were 0.877, 0.892, 0.915, 0.918, 0.917, and 0.917, respectively. As can be seen, the fitting accuracy of the model steadily increased during the study period. Similarly, the MAE and RMSE of the model decreased over time from 8.796 and 14.376 $\mu\text{g}/\text{m}^3$ in 2015 to 4.446 and 7.833 $\mu\text{g}/\text{m}^3$ in 2020, respectively. The slope of the regression equation for each year ranges from 0.88 to 0.92 (Table 1). As the slope is high and the variation range small, this indicates that the model has good fitting performance and high stability. From the multi-angle evaluation of the numerical variation of the annual fitting results R^2 , MAE, and RMSE, and

the slope stability of the regression equation, the STW-LightGBM was found to perform well for PM2.5 fitting.

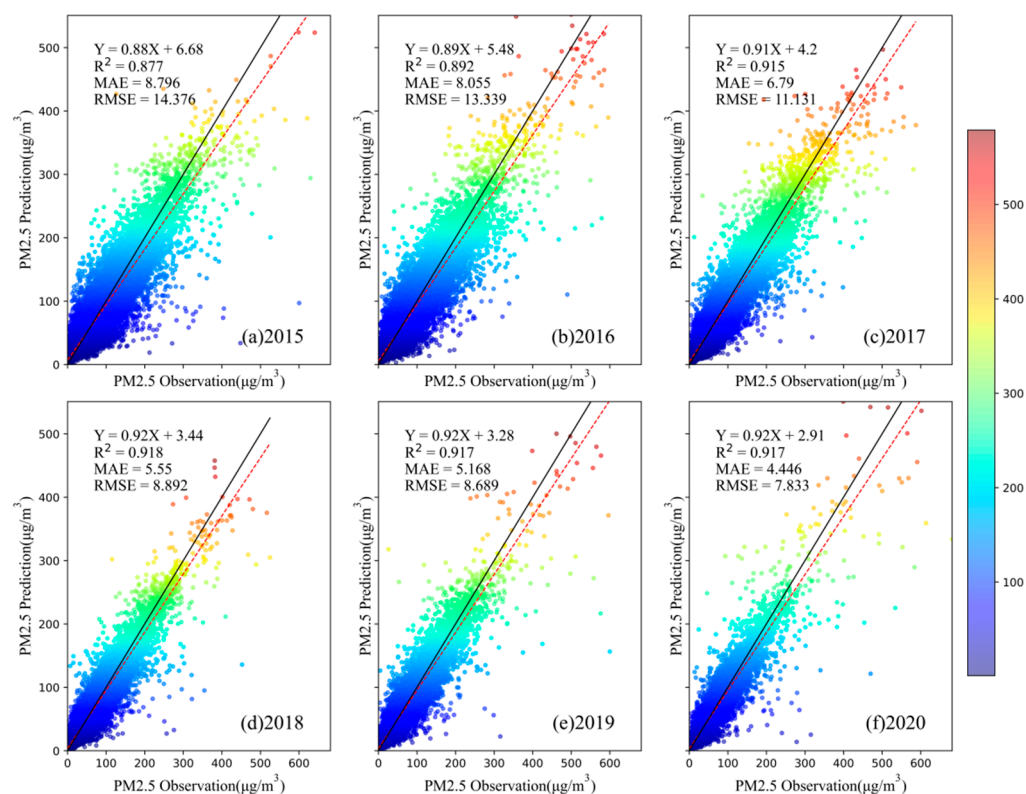


Figure 2. Density scatter plots of the results of the 10-fold CV of the estimated daily PM2.5 concentration from 2015 to 2020 (a–f). The black dotted line denotes the 1:1 line and the red solid line denotes the fitted regression line.

Table 1. Fitting results for the model used in this study at different time scales.

| Time Scale | R^2 | RMSE | MAE | Regression Equation |
|------------|-------|--------|-------|---------------------|
| 2015 | 0.877 | 14.376 | 8.796 | $Y = 0.88X + 6.68$ |
| 2016 | 0.892 | 13.339 | 8.055 | $Y = 0.89X + 5.48$ |
| 2017 | 0.915 | 11.131 | 6.79 | $Y = 0.91X + 4.2$ |
| 2018 | 0.918 | 8.892 | 5.55 | $Y = 0.92X + 3.44$ |
| 2019 | 0.917 | 8.689 | 5.168 | $Y = 0.92X + 3.28$ |
| 2020 | 0.917 | 7.833 | 4.446 | $Y = 0.92X + 2.91$ |
| 2015–2020 | 0.904 | 11.027 | 6.523 | $Y = 0.9X + 4.27$ |
| monthly | 0.906 | 7.295 | 4.838 | $Y = 0.91X + 3.88$ |
| seasonally | 0.907 | 6.342 | 4.838 | $Y = 0.9X + 4.25$ |
| yearly | 0.866 | 5.78 | 3.929 | $Y = 0.85X + 6.53$ |

To further verify the overall fitting accuracy of the model, this study used 983,140 samples from 2015 to 2020 and the trained model was used to predict the surface PM2.5 concentration and for model evaluation. The fitting of the observed and predicted values on a daily scale for all samples is shown in Figure 3. The coefficient of determination (R^2) was 0.904, MAE was $6.523 \mu\text{g}/\text{m}^3$, and RMSE was $11.027 \mu\text{g}/\text{m}^3$.

4.2. Interannual-, Seasonal-, and Monthly Scale Performance

To further verify the model general performance for STW-LightGBM at different time scales, monthly, seasonal-, and annual-scale datasets based on site were fitted. The period March–May is defined as spring, June–August as summer, September–November as

autumn, and December–February as winter. The fitting results of the model at different time scales are shown in Figure 4.

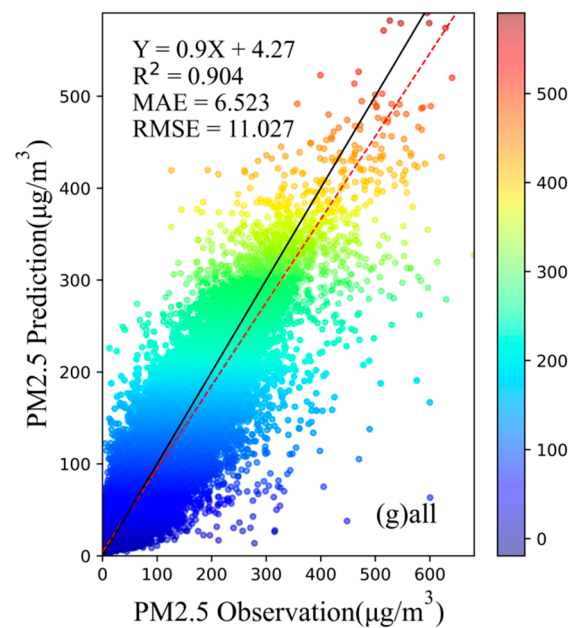


Figure 3. Density scatter plot depicting model fitting for all samples from 2015 to 2020. The black dotted line denotes the 1:1 line and the red solid line denotes the fitted regression line.

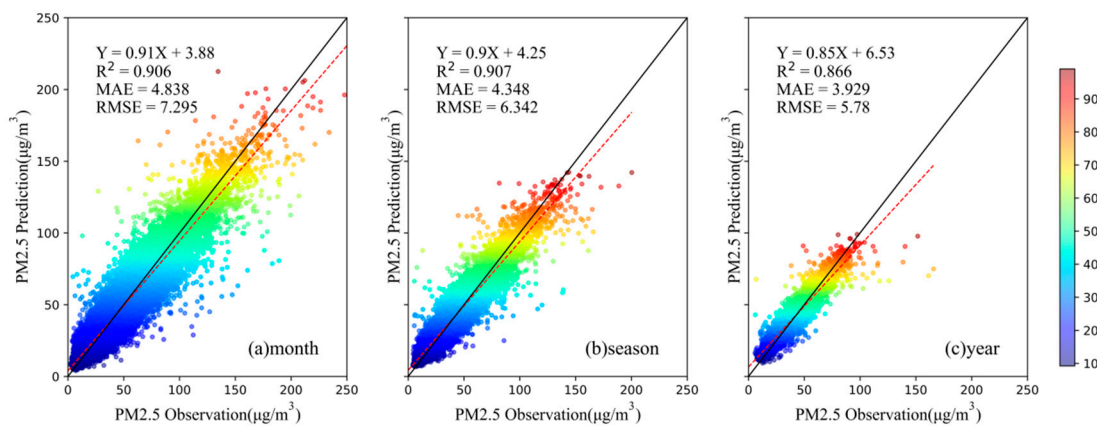


Figure 4. Density scatter plots of the 10-fold CV results for different time scales: (a) month; (b) season; and (c) year. The black dotted line denotes the 1:1 line and the red solid line denotes the fitted regression line.

The R^2 values for the annual, monthly, and seasonal scales were 0.866, 0.907, and 0.906, respectively. The seasonal scale was the highest, followed by annual and monthly scales. Maxima for the MAE and RMSE were no more than $5 \mu\text{g}/\text{m}^3$. By comparing and analyzing the daily (Figure 3) and annual scales (Figure 2), the fitting accuracy of the model used in this study was found to have reached a good level at each time scale, which shows that the parameters adjusted in this study and the established model have good stability.

4.3. Site-Based and Time-Based Authentication

Distribution maps of the PM_{2.5} verification results in China from 2015 to 2020 using a site-based 10-fold cross-validation method are shown in Figure 5. To make the results statistically significant, we removed sites with fewer than 10 samples, calculated the 6-year average value of the sites that met the requirements, and used 10-fold cross-validation. Due to the different service lives of different sites, some differences in the sample sizes were

observed. The site with the best verification result was located in the BTH region; this site was used for the longest time, and its CV-R² reached >0.95. The number of meteorological stations in the southwestern region was small and weakly representative of the sample space, and the site-based CV-R² was generally <0.8. The RMSE distribution results are similar to those for the CV-R² distribution. The RMSE in most areas in the eastern part of the country was <15 µg/m³, which was lower than that in the northwestern region without site data.

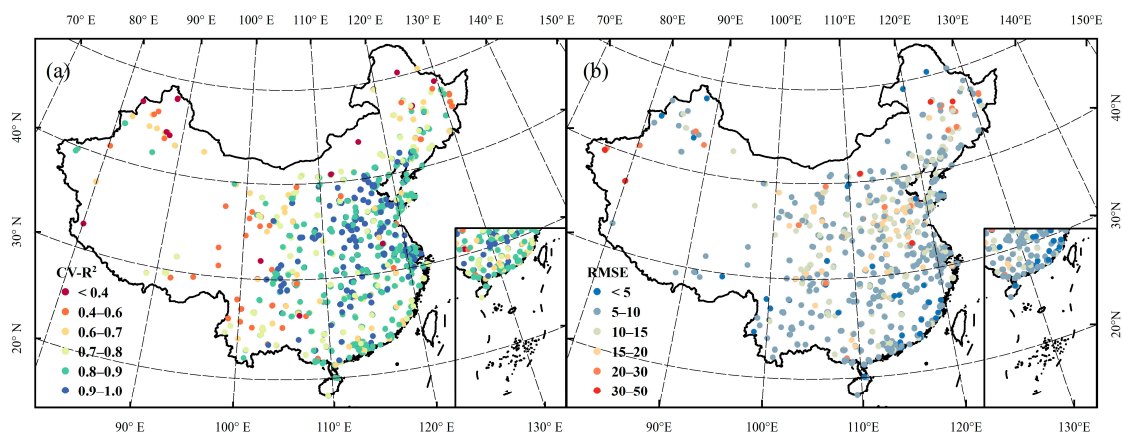


Figure 5. Map showing the site-scale spatial distributions of estimated results based on the STW-LightGBM model: (a) CV-R²; and (b) RMSE.

The distribution map for PM_{2.5} and the verification results from across China for 2015–2020 using a time-based 10-fold cross-validation method are shown in Figure 6. As with the site-based processing operation, daily station data containing fewer than 10 samples per day were removed, the daily mean value of the remaining national daily PM_{2.5} concentration observation data was calculated, and 10-fold cross-validation was applied. As can be seen, the model had the best fitting effect at the beginning and end of each year, i.e., in winter, when the PM_{2.5} concentration was high, and the average CV-R² was greater than 0.82. In the summer and autumn of each year, i.e., when PM_{2.5} was low, the fitting effect of the model decreased, and the mean value of CV-R² was <0.8. The fitting accuracy based on time was comparatively lower than that based on site, which may be related to the fact that the calculation of the mean value weakens the spatial characteristics of PM_{2.5} data.

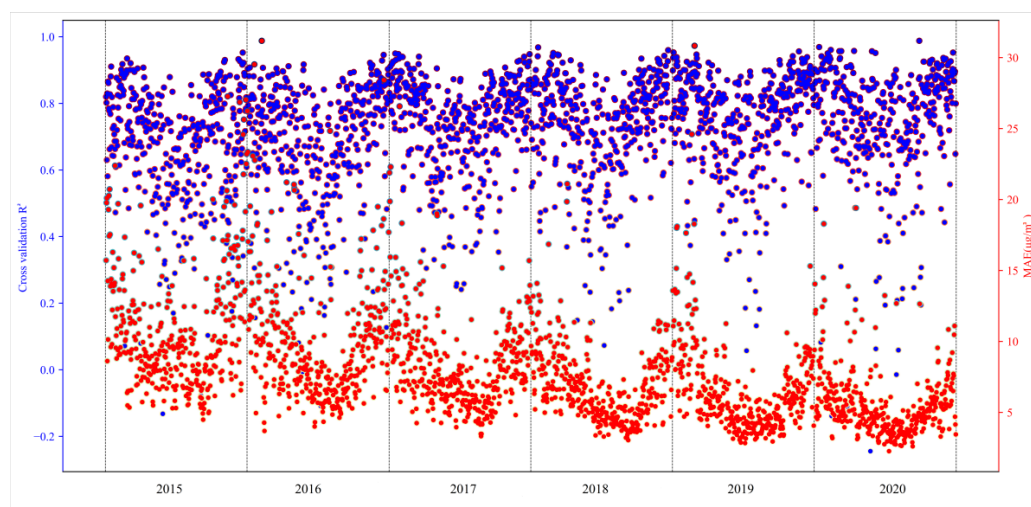


Figure 6. Scatter plot illustrating the time series of evaluation results of the daily averaged PM_{2.5} concentrations from 2015 to 2020. Purple and red points represent CV-R² and MAE, respectively.

4.4. Standard Error Ellipse

In this study, the standard error ellipse is used to study the spatial and temporal evolution of PM_{2.5} concentrations in China from 2015 to 2020, and the spatial characteristics of the annual average PM_{2.5} at all stations for 6 years are quantitatively drawn based on the stations (Supplementary Materials Figure S3). The main body of the ellipse was located in the central and eastern regions of China, and a ‘northeast–southwest’ spatial distribution pattern is generally presented (Figure 7). This is directly related to the dense distribution of stations in the east and sparse distribution in the west and is indirectly related to the population distribution and economic level of the different regions of China.

Specific data relating to the standard deviation ellipse between the observed and predicted values of PM_{2.5} in this study period are shown in Table 2. As can be seen, the length of the short axis of the ellipse decreased from 8.74 in 2015 to 8.70 in 2020 and the ratio of the long axis to the short axis increased from 1.42 in 2015 to 1.44 in 2020. This indicates that the shortening trend of the short axis is stronger than that of the long axis. The standard deviation ellipse shrinks significantly in the east–west direction and does not change significantly in the north–south direction. The directional trend of the standard deviation ellipse of PM_{2.5} concentrations became increasingly obvious. Elliptical oblateness increased from 0.295 to 0.305 during the 2015–2017 period, decreased rapidly in 2018, and reached 0.289 in 2020. This was related to the promotion of the atmospheric governance policies of China in different regions. The rate of fall of PM_{2.5} levels in the eastern region is faster than that in the central and western regions, while the decreasing trend for the PM_{2.5} concentration in the north and south is close, thus forming the aforementioned northeast–southwest spatial distribution trend in the surface PM_{2.5} standard deviation ellipse for China. The increase in PM_{2.5} in the east–west direction in recent years, enhanced the oblateness of the ellipse, making the northeast–southwest tilt trend of the ellipse more obvious, a situation that was verified by the subsequent PM_{2.5} concentration inversion image. Deviations in the latitude and longitude of the center of gravity between the observed and predicted ellipses were $<0.01^\circ$. The observed and predicted ellipses have a high degree of coincidence between the major and minor axes and the center of gravity, which reflects the prediction accuracy of the model on the spatial scale.

Table 2. Standard deviation ellipse information for the observed and predicted PM_{2.5} levels from 2015 to 2020.

| Data | Year | Short Half-Axis | Long Half-Axis | Flattening | Areal Coordinates |
|------------------|------|-----------------|----------------|------------|-------------------------|
| Observation data | 2015 | 8.741986 | 12.400678 | 0.295 | (113.2846, 33.078364) |
| | 2016 | 8.743532 | 12.571961 | 0.305 | (113.303646, 33.141384) |
| | 2017 | 8.743532 | 12.571961 | 0.305 | (113.303646, 33.141384) |
| | 2018 | 8.704039 | 12.379003 | 0.297 | (113.393208, 33.188897) |
| | 2019 | 8.6856 | 12.345031 | 0.296 | (113.53484, 33.127374) |
| | 2020 | 8.711217 | 12.253378 | 0.289 | (113.517829, 33.161107) |
| Prediction data | 2015 | 8.797358 | 12.469937 | 0.295 | (113.293791, 33.102142) |
| | 2016 | 8.691407 | 12.51929 | 0.306 | (113.335734, 33.115923) |
| | 2017 | 8.691407 | 12.51929 | 0.306 | (113.335734, 33.115923) |
| | 2018 | 8.698212 | 12.361702 | 0.296 | (113.381568, 33.166527) |
| | 2019 | 8.696882 | 12.344421 | 0.295 | (113.537326, 33.145294) |
| | 2020 | 8.65032 | 12.172791 | 0.289 | (113.558399, 33.152506) |

To evaluate the ability of the model to predict PM_{2.5} pollution levels in different regions of China, four typical regions in China were selected, BTH, Jiangsu–Zhejiang–Shanghai, Xinjiang, and Sichuan–Chongqing, to analyze the trend in predicted PM_{2.5} levels and draw the standard deviation ellipse (Figure 8). For the BTH region, the standard error ellipse area began to shrink in 2015 and moved northeastward. The short half-axis of the ellipse increased, the long half-axis decreased, the elliptical oblateness increased, and the centrifugal force and directionality increased. The change in ellipse area in the Sichuan–Chongqing and

Jiangsu–Zhejiang–Shanghai regions is not significant. The ellipse in the Jiangsu–Zhejiang–Shanghai region shows a slight trend toward contraction to the northwest, while that in the Sichuan–Chongqing region shows a gradual trend toward expansion to the northeast. The difference in the size of the standard deviation ellipse in Xinjiang was the most obvious. As the ellipse expands to the northeast, the long half-axis is shortened, the short half-axis is increased, the elliptical flatness is reduced, and the centrifugal force and direction are weakened.

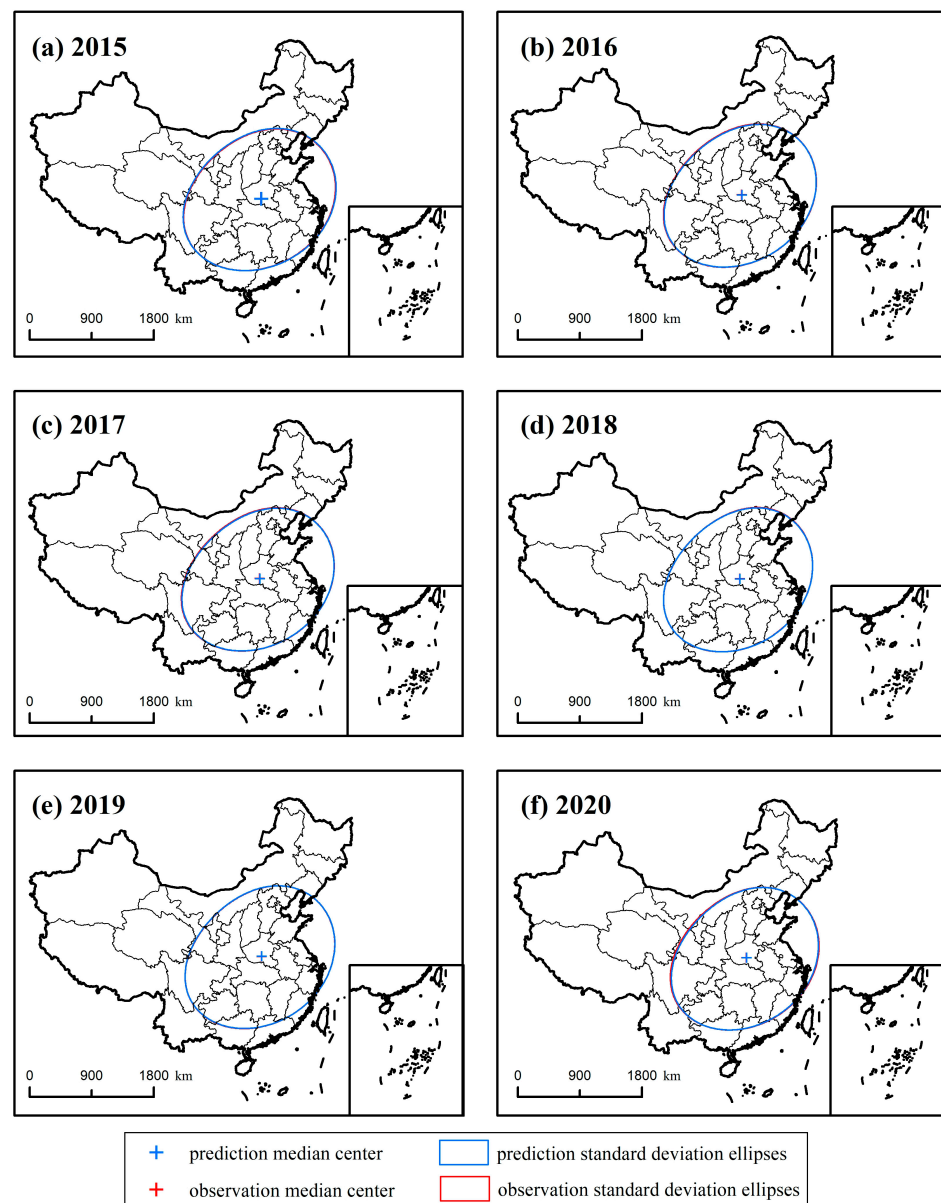


Figure 7. Maps showing the standard error ellipses for observed and predicted PM_{2.5} levels in China from 2015 to 2020. The blue and red ellipses represent the predicted and observed values, respectively. The blue and red crosses represent the center of gravity of the standard deviation ellipse of the predicted value and the observed values, respectively.

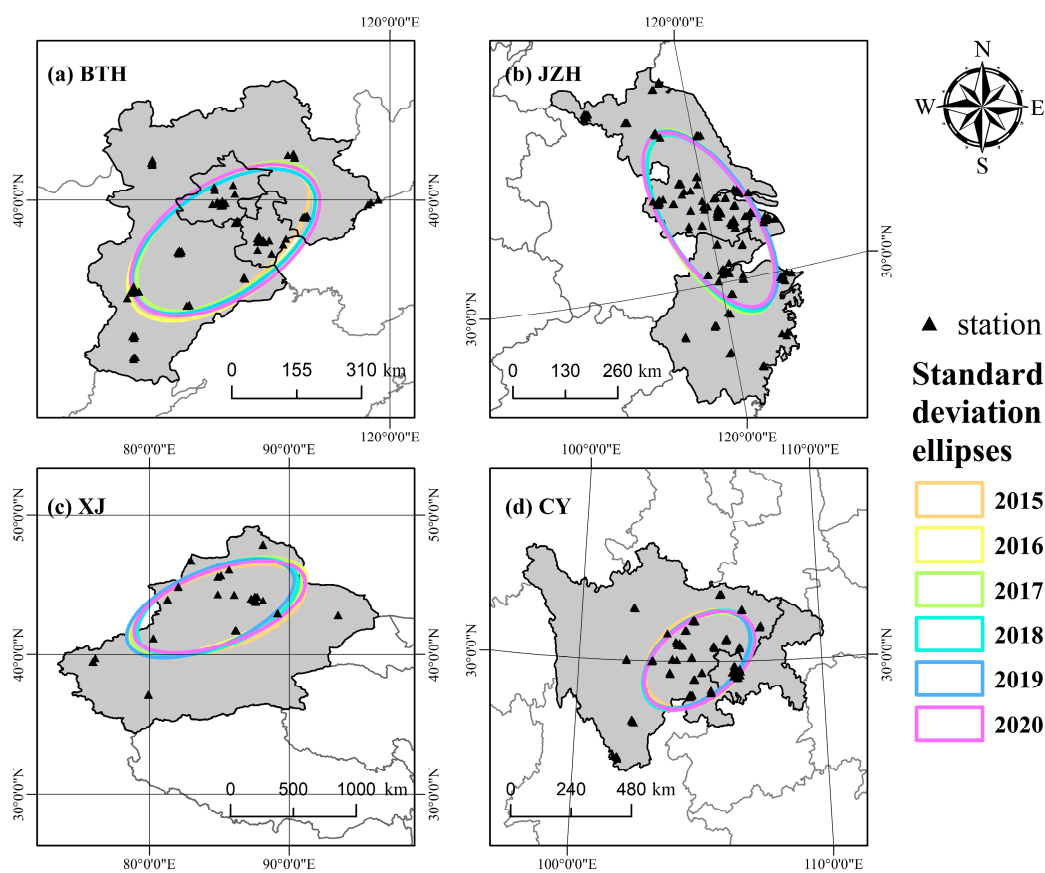


Figure 8. Maps depicting the standard error ellipses of the PM_{2.5} predicted values in some regions of China from 2015 to 2020 (a) Beijing–Tianjin–Hebei (BTH); (b) Jiangsu–Zhejiang–Shanghai; (c) Xinjiang Province; and (d) Sichuan–Chongqing. The black triangles represent the locations of the ground monitoring sites.

As the standard error ellipse is limited by the incomplete distribution of stations and the weakness of the data due to errors in the annual average calculations, it cannot accurately reflect the trend in surface PM_{2.5} concentrations. It needs to be combined with more detailed remote sensing inversion mapping data to jointly characterize the spatial and temporal variation characteristics and trends in near-surface PM_{2.5} concentration in China from 2015 to 2020.

4.5. Mapping of PM_{2.5} Concentrations in China from 2015 to 2020

The mapping results for the PM_{2.5} concentration at a 1 km spatial resolution in China from 2015 to 2020 based on the STW-LightGBM fitting are shown in Figure 9. The fitting results show the full-coverage predictions for the surface concentration and have similar spatial and temporal variation rules to other studies into surface PM_{2.5} concentrations [29,30]. In terms of spatial distribution, the areas with high PM_{2.5} values, from 2015 to 2020, were mainly distributed in the eastern coastal areas, North China Plain, Sichuan Basin, and western Xinjiang. Among them, the high value aggregation phenomenon in the Tarim Basin in Xinjiang is the most significant, and the annual average concentration of PM_{2.5} for this 6-year period is more than 110 $\mu\text{g}/\text{m}^3$. Considering the low intensity of human activities in the region, the presence of PM_{2.5} in the Tarim region may be predominantly the result of natural weather conditions, such as dust. For the BTH, Pearl River Delta, Yangtze River Delta, and other regions with high urbanization rates and population densities, automobile exhausts and factory emissions were the two main causes of high PM_{2.5} concentrations. Terrain is another important factor causing high PM_{2.5}. For example, in the Sichuan Basin, bowl-like terrain gives rise to local air pollutants such as

PM_{2.5} as such terrain does not have the airflow experienced by flat plains, resulting in the accumulation of local air pollutant concentrations and high PM_{2.5} concentrations.

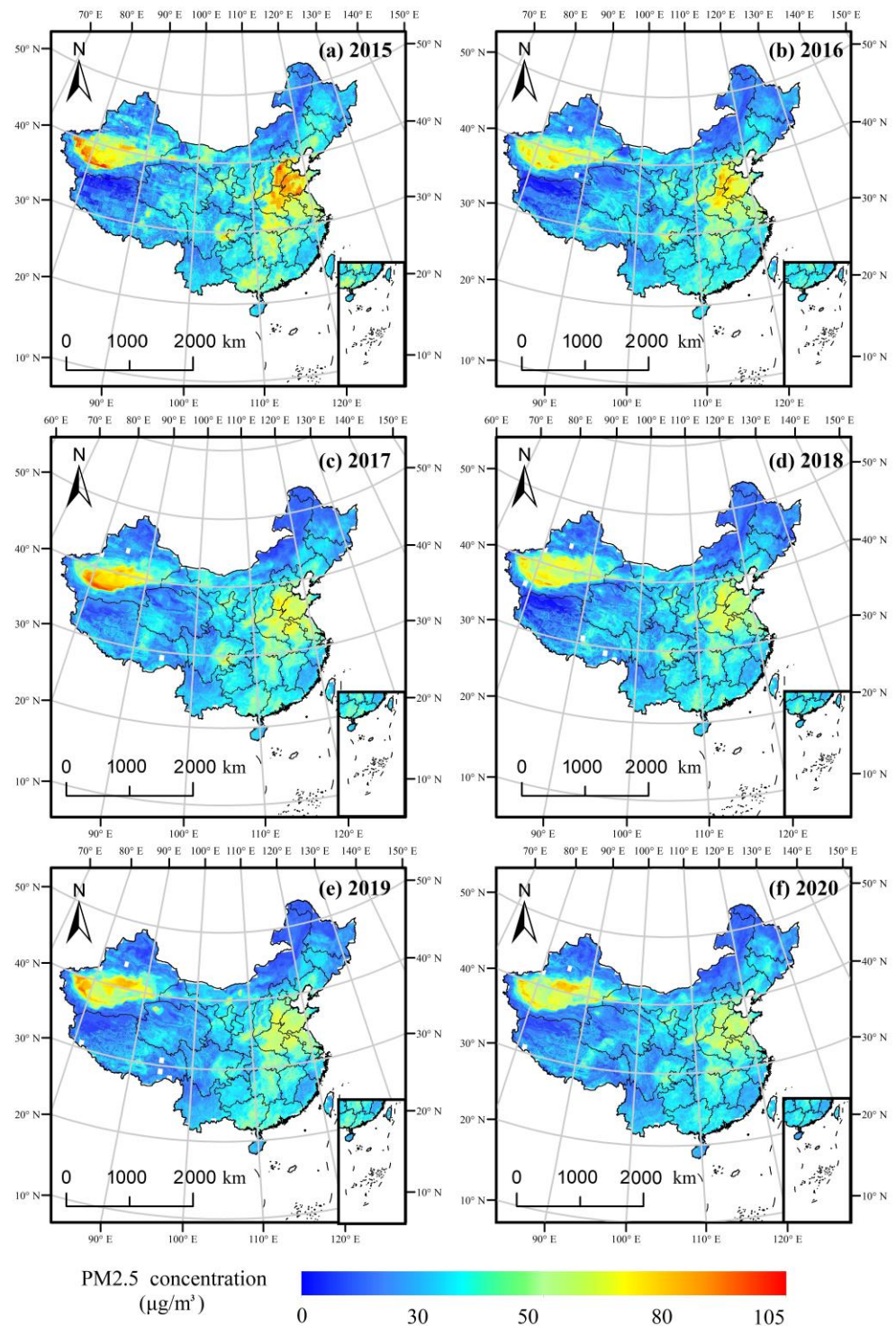


Figure 9. Maps showing the spatial distribution of annual mean estimated PM_{2.5} levels across China at a 1 km resolution from 2015 to 2020 (a–f).

The surface PM_{2.5} concentrations in earlier PM_{2.5} hot spots such as the BTH region, North China, the Pearl River Delta region, and the Jiangsu–Zhejiang–Shanghai region showed a decreasing trend due to the implementation of pollution control measures in China in 2013. Furthermore, the clean air plan [48] and the 3-year Blue Sky action plan

were implemented in 2018 [49] and have had significant impacts. As shown in Supplementary Materials Figure S4, the average PM_{2.5} concentration in the Jiangsu, Zhejiang, and Shanghai areas decreased by 10.82 $\mu\text{g}/\text{m}^3$ in this 6-year period, and the average decrease in Sichuan and Chongqing was 5.34 $\mu\text{g}/\text{m}^3$. The average decrease in Beijing, Tianjin, and Hebei was 11.88 $\mu\text{g}/\text{m}^3$, and the maximum decrease was 72.88 $\mu\text{g}/\text{m}^3$, which is second only to the maximum decrease of 90.51 $\mu\text{g}/\text{m}^3$ seen in Xinjiang. While PM_{2.5} levels in the country have generally fallen, we found that local high PM_{2.5} emissions remain. For example, the PM_{2.5} concentration in a mountainous area in the western part of Hebei Province increased by 30.52 $\mu\text{g}/\text{m}^3$ in this 6-year period. The northeastern part of the Tianshan Mountains in Xinjiang and the central part of the Sichuan Basin experienced the largest growth in pollution in the region, at 65.71 and 27.98 $\mu\text{g}/\text{m}^3$, respectively.

However, with the improvement in the air-quality-related policies of China, industries with high energy consumption and pollution are gradually being replaced by more environmentally friendly and energy-saving technologies. Combined with vehicle restrictions, the emissions of air pollutants have been effectively reduced. Additionally, the combined effects of measures such as shutdowns, specifically those during the outbreak of the new coronavirus epidemic in 2020, have also caused significant reductions in PM_{2.5} levels in China compared with those seen in 2015.

To explore the interannual and seasonal variation characteristics of PM_{2.5} levels in China, box plots can be used, as they are not affected by abnormal values, and thus, can intuitively, accurately, and stably describe the discrete distribution of the data. In this study, a box plot was used to analyze the trends in PM_{2.5} concentrations at different scales. Figures 10 and 11 respectively show the discrete distributions of the annual mean and means of different quarters in each year for the stations of China from 2015 to 2020. The interannual scale shows the annual average value for PM_{2.5} concentrations at stations across China and a first increasing and then decreasing trend was observed, although overall, the trend was downward (Figure 10). In 2016, the median of the box-type diagram was approximately 50 $\mu\text{g}/\text{m}^3$, while in 2020 it was 40 $\mu\text{g}/\text{m}^3$. Thus, a downward trend was observed from 2016 to 2020. Outliers were also present in each box plot, which may be due to the fact that the decrease in background PM_{2.5} concentration in most parts of the country was not obvious or in some places, such as the Tarim Basin in Xinjiang, slightly increased, resulting in abnormal values in this area relative to that of the other sites. From the perspective of the box-plot changes shown in Figure 10, the distance between the upper and lower quartiles represented by the box plot from 2015 to 2020 is shrinking, indicating that the numerical fluctuations of different stations in China are getting smaller, and the regional PM_{2.5} concentration differences are also getting smaller. Additionally, from the perspective of changes in box shape, the upper and lower limit ranges of the boxes also show a narrowing trend, indicating that the concentration of PM_{2.5}, as monitored by most stations, is increasing, and the regional differences are gradually decreasing.

The discrete distributions of the seasonal means of the stations in different years and seasons are shown in Figure 11. Overall, the surface PM_{2.5} concentrations in the four seasons showed a decreasing trend from 2015 to 2020. From the perspective of the data distribution, the mean distribution of the PM_{2.5} at stations in the summer was relatively concentrated, and the mean value was the smallest, indicating that in the summer, the PM_{2.5} concentration was the lowest in any given year, while the differences between regions were larger than in some other seasons. The box height of the box-type map in winter was the highest, indicating that in winter, the average distribution of PM_{2.5} at different stations was more dispersed and the regional differences were larger. This may be the result of the weather and heating conditions in northern and southern China. The box heights in spring and autumn were between the two; however, generally, the box height in spring was lower than that in autumn, indicating that the PM_{2.5} concentration at each site in spring was higher than that in autumn, and the regional differences were smaller in the spring than those in autumn.

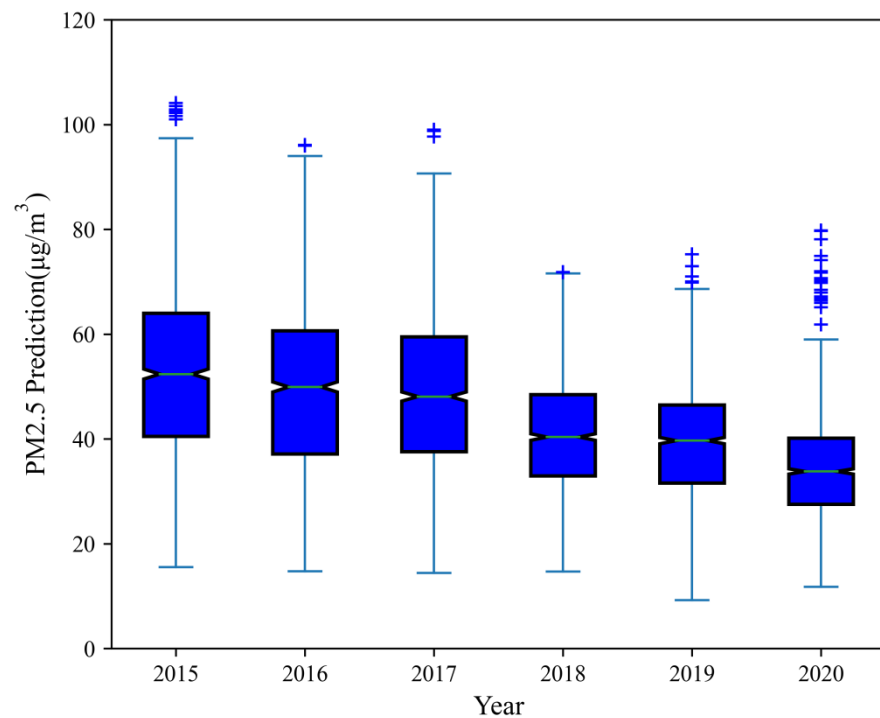


Figure 10. Box plot of the annual mean PM_{2.5} predicted values at the station from 2015 to 2020. In each box, the blue plus signs represent outliers, and the middle, lower, and upper horizontal lines represent the median bias, 25th percentile, and 75th percentile, respectively.

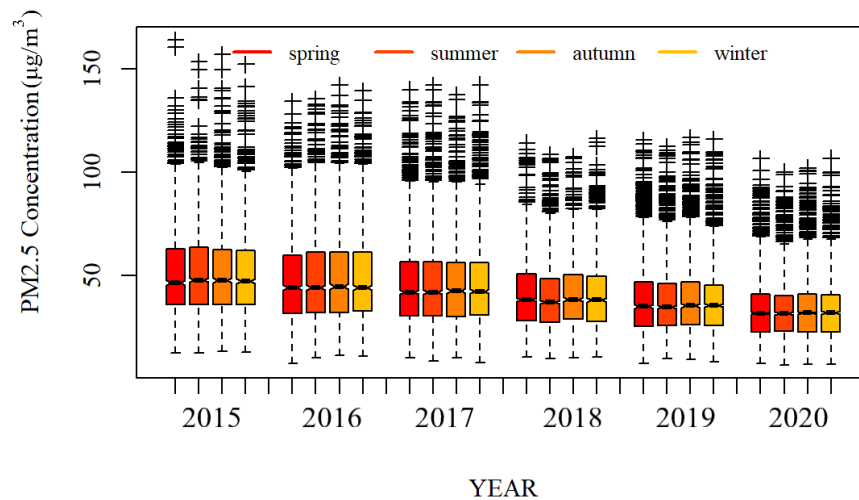


Figure 11. Box chart showing the PM_{2.5} prediction value means by quarter for stations from 2015 to 2020. In each box, the black plus signs represent outliers, and the middle, lower, and upper horizontal lines represent the median bias, 25th percentile, and 75th percentile, respectively.

According to the distribution of data during the year, the PM_{2.5} concentration in the summer was the lowest in any given year, followed by the spring and autumn, and the highest in winter, with the same change rules for different years. For the four seasons across all years, the consistency in the average concentration of PM_{2.5} was highest in the summer, lowest in the winter, and between the two in the spring and autumn, indicating that the average values of the national stations in summer were relatively close when the surface PM_{2.5} concentration in China was the lowest. PM_{2.5} levels were the highest in winter, although the value varied greatly across the country. Interannually, the average value of the four seasonal stations in the country from 2015 to 2020 showed a downward trend, with the most obvious decline in winter. As shown in Figure 11, the surface PM_{2.5} concentration in

winter decreased from $70.67 \mu\text{g}/\text{m}^3$ in 2015 to $46.75 \mu\text{g}/\text{m}^3$ in 2020. In autumn, it decreased from $49.09 \mu\text{g}/\text{m}^3$ in 2015 to $32.28 \mu\text{g}/\text{m}^3$ in 2020, in spring, it decreased from $45.75 \mu\text{g}/\text{m}^3$ in 2015 to $32.66 \mu\text{g}/\text{m}^3$ in 2020, and in summer, it decreased from $33.30 \mu\text{g}/\text{m}^3$ in 2015 to $25.76 \mu\text{g}/\text{m}^3$ in 2020. The heights of the boxes for the four seasons showed a decreasing trend, indicating that the average values for each station were gradually converging, and differences in the surface PM_{2.5} concentration in different regions decreased. The results point to the effective implementation of energy conservation and air quality management policies.

5. Discussion

5.1. Model Fitting Performance

We combined MAIAC AOD ERA5 reanalysis data, etc., to construct an STW-LightGBM model that considers the temporal and spatial characteristics of air pollution and estimate the PM_{2.5} concentration of the surface at 1 km resolution in China from 2015 to 2020. The model proposed in this study performs well and can maintain high prediction accuracy at different timescales (the mean R^2 is 0.902). Compared with the results of earlier studies, the fitting slope of the model is the highest (ranging from 0.88 to 0.92, while the average slope of the 10-fold cross-validation is 0.9), which indicates a good fit for the real surface PM_{2.5} concentration. At different time scales, the change in the range of the predicted slope was <0.2 , and the R^2 change range was <0.45 , demonstrating the robustness and stability of the model. This implies that our model is more sensitive to changes in independent variables and can obtain more detailed PM_{2.5} concentration data under certain conditions. This provides the possibility of obtaining pollution information using remote-sensing data instead of ground monitoring data and could thus reduce the need for ground monitoring stations, which are expensive and difficult to place.

To improve the prediction accuracy of the model, this study adopted a variety of measures, including finding variables related to PM_{2.5}, paying attention to data availability, selecting model algorithms with high accuracy and low complexity for fitting, focusing on overfitting and underfitting problems, selecting appropriate parameter combinations, and providing large amounts of training sample data. The STW-LightGBM model developed in this study is an optimized version of the GBDT machine learning algorithm that can significantly improve computational efficiency and reduce time complexity without affecting accuracy. Simultaneously, this study incorporated the unique spatiotemporal relationship of geography into the machine learning model, which effectively improved the prediction accuracy of the model.

5.2. Comparison with Traditional Models

The model fitting results were compared with the verification results of five widely used traditional statistical models (Supplementary Materials Table S2). The results show that the prediction model proposed in this study is superior to the geographically weighted model (GWR) [50], the spatiotemporal geographically weighted model (GTWR) [19], the combination model for LME, and the generalized additive model (GAM) [27] in terms of temporal and spatial resolution of PM_{2.5} levels. The spatial resolutions of the statistical model research results in the literature were mostly 3 km, 5 km, or 10 km, and it was rarely possible to achieve ≤ 1 km resolution. The verification results for R^2 for these models are generally lower than 0.9 and the fitting slope fluctuated around 0.7. Using statistical models for predictions generally underestimates the surface PM_{2.5} concentration, and cannot fully reflect the changes in regional pollutants. In the long run, these deficiencies will affect the efficacy of the air pollution control measures adopted by decision-makers. Even with the Geo-DBN model [51], which performs best in Table S2 and has an R^2 that reaches 0.88, the spatial resolution is only 10 km, the research time limit is only 2015, and there is a lack of temporal and spatial resolution performance. The STW-LightGBM model used in this study not only has a higher R^2 than that of the above traditional models but also has a fitting slope closer to 1, which can better reflect the real information for ground PM_{2.5}

concentrations to provide a reference for formulating various relevant policies, maintaining air quality, and improving the ecological environment.

5.3. Comparison with Relevant Studies

In terms of using machine learning to estimate PM_{2.5} research in China, the STW-LightGBM used in this study had faster calculation speeds, uses less memory, and has a more accurate performance than RF, XGBoost, GBDT, and other models. Compared to traditional statistical models, the spatial resolution of PM_{2.5} levels in mainland China estimated by machine learning can be improved to a 1 km resolution, the prediction accuracy can be improved accordingly, and R^2 can be maintained at approximately 0.8. At a spatial resolution of 3 km, Zhang et al. [26] and Zhang et al. [52] used GBDT to estimate surface PM_{2.5} concentrations in 2017 and 2013–2017, respectively. The predicted R^2 and RMSE in these two studies were 0.81 and 11.57 $\mu\text{g}/\text{m}^3$, and 0.81 and 19.76 $\mu\text{g}/\text{m}^3$, respectively. Wei et al. [25] and He et al. [19] estimated the surface PM_{2.5} concentration in 2016 and 2013–2018 at a spatial resolution of 1 km, and achieved predicted R^2 and RMSE values of 0.85 and 15.57 $\mu\text{g}/\text{m}^3$ and 0.59 and 27.18 $\mu\text{g}/\text{m}^3$, respectively. However, their approaches had problems similar to those in the traditional statistical models. Their machine learning models underestimate PM_{2.5} levels, which is manifested in the large gap between the values for the slopes of their model predictions (between 0.6 and 0.8) and the standard value of 1. Some studies also overestimate PM_{2.5} concentrations. For example, when Chen et al. [17] predicted the PM_{2.5} concentration at a 10 km resolution in China from 2014 to 2016 based on the RF model, the regression prediction equation was $Y = 1.07X - 4.64$, which was higher than the real PM_{2.5} value.

With the rapid development of satellite big data, the construction of a spatiotemporal prediction system for atmospheric PM_{2.5} concentration is expected to become a new development direction for atmospheric fine PM concentration monitoring by combining constructed time series prediction models and spatial models [53–55].

5.4. Model Prediction Results

This study uses the standard deviation ellipse to measure the accuracy of the model prediction, which can more intuitively compare the difference between the predicted and true values and distinguish the spatial and temporal prediction ability of the model. By comparing the superposition of the standard deviation ellipse of the predicted and observed values year by year and the detailed information of the ellipse, we can find that the deviation of the longitude and latitude of the center of gravity of the two circles in each year does not exceed 0.01° , and the distance between the long and short axes is within 0.01 km. The observed and predicted ellipses have a high degree of coincidence in the long and short axes and the center of gravity, and thus, the model fitting results are visually verified in space. During the study period, the standard deviation ellipse of PM_{2.5} in China generally showed a spatial distribution pattern of ‘northeast–southwest’. The decreasing trend in the east–west short axis of the ellipse is greater than that of the north–south long axis, and the ratio of the long and short axes increases from 1.42 in 2015 to 1.44 in 2020. The north–south directional trend in the PM_{2.5} spatial distribution is enhanced, and the east–west directional trend is weakened. The results show that in recent years, the decreasing trend in the difference in pollution east-to-west in China is greater than that north-to-south. Therefore, different pollution reduction and emission reduction policies can be implemented according to the pollution characteristics of different regions.

Due to the large population density, the level of economic development and the degree of industrialization east of the Mohe–Tengchong line in China are greater than those in the central and western regions. Thus, the air pollution base in the latter is large and the rate of decline is low. This is one of the reasons why the concentration of PM_{2.5} in the eastern region has decreased significantly, and the pollution center is now located in the central and western regions of China. In the future, targeted policies may be formulated to improve the air quality level of high-PM_{2.5}-level clusters. However, PM_{2.5} pollution

in some parts of China is still rebounding, which shows that PM_{2.5} levels in the air over China have not been fundamentally controlled. With the aggravation of ozone pollution in China and the continuous study of the highly nonlinear relationship between ozone and PM_{2.5}, controlling the continued generation of PM_{2.5} pollution has become a major challenge for China to realize the coordinated prevention and control of ozone and PM_{2.5} in the new era. In the future, our team plans to explore the phenomenon of combined PM_{2.5} and ozone pollution at multiple scales and spatial dimensions and provide data support for formulating more accurate, quantitative, and efficient control measures.

Generally, the temporal and spatial variations in surface PM_{2.5} concentration revealed in this study correspond to those found in earlier studies, which provides a reference for a more accurate estimation of PM_{2.5} concentration. However, some limitations persist in this study. For example, the small number of stations in the western region may limit the accuracy of model training. The spatial resolution of the fitted PM_{2.5} concentration is 1 km, and refined data could further improve the prediction ability of the model. Adding more variable adjustment parameter combinations may further improve the accuracy of the model.

6. Conclusions

In this study, a surface PM_{2.5} prediction method based on AOD and site pollution information was proposed. Spatiotemporal-weighted methods are incorporated into the modeling of the enhanced light gradient machine learning model, which effectively improves the prediction efficiency and accuracy of the model. We applied the model to the Chinese mainland for the years 2015 to 2020 and obtained PM_{2.5} data with a resolution of 1 km. The overall validation CV-R² of the model is 0.904 $\mu\text{g}/\text{m}^3$, the predicted slope of all samples is 0.9 $\mu\text{g}/\text{m}^3$, and the interannual predicted slope fluctuation is generally ≤ 0.2 . The 10-fold cross-validation fitting accuracy of the model used in this study is between 0.877 and 0.917 $\mu\text{g}/\text{m}^3$. At different time scales of month, season, and year, the fitting accuracy of the model is more than 0.85. Compared with approaches used in earlier studies, the STW-LightGBM model can better fit PM_{2.5} concentration and also shows higher stability at different time scales, which is suitable for long-term prediction of PM_{2.5}.

We use the standard error ellipse to characterize the spatial distribution of PM_{2.5} pollution in China from 2015 to 2020. The ellipse generally presents a spatial distribution pattern of 'northeast–southwest'. The ratio of the long and short axes of the ellipse increased from 1.42 in 2015 to 1.44 in 2020. The north–south directional trend in PM_{2.5} spatial distribution is enhanced, and the east–west directionality is weakened. Next, based on the model prediction results, we mapped the surface PM_{2.5} concentration map of China from 2015 to 2020. The results showed that during this 6-year period, the national PM_{2.5} pollution situation improved. The national average concentration decreased by 4.55 $\mu\text{g}/\text{m}^3$, and the largest decline—in the Beijing–Tianjin–Hebei region—reached 90.51 $\mu\text{g}/\text{m}^3$. However, PM_{2.5} levels continue to rise in Xinjiang, the Sichuan Basin, and other places, which brings challenges to the life and health of the residents of these regions. Therefore, our research results make an important contribution to the analysis of the spatial and temporal differentiation of PM_{2.5} and the national carbon emission reduction policy.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs15164104/s1>, Figure S1: Spatial distribution of PM_{2.5} monitoring stations included in this study; Figure S2: The visualization result of STW-LightGBM parameter adjustment; Figure S3: The PM_{2.5} observed and predicted standard error ellipses for China from 2015 to 2020; Figure S4: Changes in PM_{2.5} levels in China from 2015 to 2020; Table S1: Summary of the data sources used in this study; Table S2: Comparison of performances of the STW-LightGBM model and other AOD-based models applied in previous studies at the national scale in China [56–60].

Author Contributions: Conceptualization, L.W., H.Z. and X.Y.; methodology, X.Y. and M.X.; software, X.Y. and M.X.; validation, X.Y., M.X. and H.Z.; formal analysis, H.Z.; investigation, X.Y.; resources, L.W.; data curation, X.Y.; writing—original draft preparation, X.Y., and M.X.; writing—review and editing, X.Y., M.X. and L.W.; visualization, X.Y. and H.Z.; supervision, H.Z. and L.W.; project administration, L.W.; funding acquisition, L.W. 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 No. 42005031) and the Science and Technology Development Project of Henan Province China (Grant No. 222102110419).

Data Availability Statement: The datasets generated during and/or analyzed during the current study are open-access or available from the corresponding author on reasonable request.

Acknowledgments: We thank the reviewers who provided valuable comments to improve the paper.

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

References

- Liu, Y.; Tong, D.; Cheng, J.; Davis, S.J.; Yu, S.; Yarlagadda, B.; Clarke, L.E.; Brauer, M.; Cohen, A.J.; Kan, H.; et al. Role of climate goals and clean-air policies on reducing future air pollution deaths in China: A modelling study. *Lancet Planet. Health* **2022**, *6*, e92–e99. [[CrossRef](#)] [[PubMed](#)]
- China Environmental Monitoring Station. 2013–2016 China Environmental Status Bulletin, 2017–2020 China Eco-Environmental Status Bulletin [EB/OL]. Available online: <http://www.cnemc.cn/jcbg/zghjzkgb/> (accessed on 12 August 2022).
- Peng, Y.; Wang, H.; Zhang, X.; Wang, P.; Li, S.; Liu, Z.; Zhang, W.; Che, H. Combined effect of surface PM_{2.5} assimilation and aerosol-radiation interaction on winter severe haze prediction in central and eastern China. *Atmos. Pollut. Res.* **2023**, *14*, 101802. [[CrossRef](#)]
- Zhang, Y.; Yin, Z.; Li, S.; Zhang, J.; Sun, H.Z.; Liu, K.; Shirai, K.; Hu, K.; Qiu, C.; Liu, X.; et al. Ambient PM_{2.5}, ozone and mortality in Chinese older adults: A nationwide cohort analysis (2005–2018). *J. Hazard. Mater.* **2023**, *454*, 131539. [[CrossRef](#)] [[PubMed](#)]
- Cui, F.; Sun, Y.; Xie, J.; Li, D.; Wu, M.; Song, L.; Hu, Y.; Tian, Y. Air pollutants, genetic susceptibility and risk of incident idiopathic pulmonary fibrosis. *Eur. Respir. J.* **2023**, *61*, 2200777. [[CrossRef](#)] [[PubMed](#)]
- Mo, S.; Wang, Y.; Peng, M.; Wang, Q.; Zheng, H.; Zhan, Y.; Ma, Z.; Yang, Z.; Liu, L.; Hu, K.; et al. Sex disparity in cognitive aging related to later-life exposure to ambient air pollution. *Sci. Total Environ.* **2023**, *886*, 163980. [[CrossRef](#)]
- Huang, W.; Zhou, Y.; Chen, X.; Zeng, X.; Knibbs, L.D.; Zhang, Y.; Jalaludin, B.; Dharmage, S.C.; Morawska, L.; Guo, Y.; et al. Individual and joint associations of long-term exposure to air pollutants and cardiopulmonary mortality: A 22-year cohort study in Northern China. *Lancet Reg. Health* **2023**, *36*, 100776. [[CrossRef](#)]
- Xia, F.; Xing, J.; Xu, J.; Pan, X. The short-term impact of air pollution on medical expenditures: Evidence from Beijing. *J. Environ. Econ. Manag.* **2022**, *114*, 102680. [[CrossRef](#)]
- Xie, Y.; Li, Z.; Zhong, H.; Feng, X.L.; Lu, P.; Xu, Z.; Guo, T.; Si, Y.; Wang, J.; Chen, L.; et al. Short-Term Ambient Particulate Air Pollution and Hospitalization Expenditures of Cause-Specific Cardiorespiratory Diseases in China: A Multicity Analysis. *Lancet Reg. Health* **2021**, *15*, 100232. [[CrossRef](#)]
- Rajagopalan, S.; Landrigan, P.J. The Inflation Reduction Act—Implications for climate change, air pollution, and health. *Lancet Reg. Health* **2023**, *23*, 100522. [[CrossRef](#)]
- Xu, M.; Qin, Z.; Zhang, S.; Xie, Y. Health and economic benefits of clean air policies in China: A case study for Beijing-Tianjin-Hebei region. *Environ. Pollut.* **2021**, *285*, 117525. [[CrossRef](#)]
- Cui, L.; Duan, H.; Mo, J.; Song, M. Ecological compensation in air pollution governance: China’s efforts, challenges, and potential solutions. *Int. Rev. Financ. Anal.* **2021**, *74*, 101701. [[CrossRef](#)]
- Pi, T.; Wu, H.; Li, X. Does Air Pollution Affect Health and Medical Insurance Cost in the Elderly: An Empirical Evidence from China. *Sustainability* **2019**, *11*, 1526. [[CrossRef](#)]
- Zhang, Y.-L.; Cao, F. Fine particulate matter (PM_{2.5}) in China at a city level. *Sci. Rep.* **2015**, *5*, 14884. [[CrossRef](#)] [[PubMed](#)]
- Engel-Cox, J.; Nguyen Thi Kim, O.; van Donkelaar, A.; Martin, R.V.; Zell, E. Toward the next generation of air quality monitoring: Particulate Matter. *Atmos. Environ.* **2013**, *80*, 584–590. [[CrossRef](#)]
- Wang, Z.; Chen, L.; Tao, J.; Zhang, Y.; Su, L. Satellite-based estimation of regional particulate matter (PM) in Beijing using vertical-and-RH correcting method. *Remote Sens. Environ.* **2010**, *114*, 50–63. [[CrossRef](#)]
- Chen, G.; Li, S.; Knibbs, L.D.; Hamm, N.A.S.; Cao, W.; Li, T.; Guo, J.; Ren, H.; Abramson, M.J.; Guo, Y. A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* **2018**, *636*, 52–60. [[CrossRef](#)] [[PubMed](#)]
- Zhang, D.; Du, L.; Wang, W.; Zhu, Q.; Bi, J.; Scovronick, N.; Naidoo, M.; Garland, R.M.; Liu, Y. A machine learning model to estimate ambient PM_{2.5} concentrations in industrialized highveld region of South Africa. *Remote. Sens. Environ.* **2021**, *266*, 112713. [[CrossRef](#)]

19. He, Q.; Huang, B. Satellite-based mapping of daily high-resolution ground PM_{2.5} in China via space-time regression modeling. *Remote Sens. Environ.* **2018**, *206*, 72–83. [[CrossRef](#)]
20. Wang, Y.; Yuan, Q.; Li, T.; Tan, S.; Zhang, L. Full-coverage spatiotemporal mapping of ambient PM_{2.5} and PM₁₀ over China from Sentinel-5P and assimilated datasets: Considering the precursors and chemical compositions. *Sci. Total Environ.* **2021**, *793*, 148535. [[CrossRef](#)]
21. Xue, Y.; Li, Y.; Guang, J.; Tugui, A.; She, L.; Qin, K.; Fan, C.; Che, Y.; Xie, Y.; Wen, Y.; et al. Hourly PM_{2.5} Estimation over Central and Eastern China Based on Himawari-8 Data. *Remote Sens.* **2020**, *12*, 855. [[CrossRef](#)]
22. Sun, J.; Wang, Z.; Zhou, W.; Xie, C.; Wu, C.; Chen, C.; Han, T.; Wang, Q.; Li, Z.; Li, J.; et al. Measurement report: Long-term changes in black carbon and aerosol optical properties from 2012 to 2020 in Beijing, China. *Atmos. Chem. Phys.* **2022**, *22*, 561–575. [[CrossRef](#)]
23. Sun, Y.; He, Y.; Kuang, Y.; Xu, W.; Song, S.; Ma, N.; Tao, J.; Cheng, P.; Wu, C.; Su, H.; et al. Chemical Differences between PM₁ and PM_{2.5} in Highly Polluted Environment and Implications in Air Pollution Studies. *Geophys. Res. Lett.* **2020**, *47*, e2019GL086288. [[CrossRef](#)]
24. Zhang, J.; Wang, H.; Guo, Y.; Hu, X. Review of deep learning. *Appl. Res. Comput.* **2018**, *35*, 1921–1928+1936.
25. 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](#)]
26. Zhang, T.; He, W.; Zheng, H.; Cui, Y.; Song, H.; Fu, S. Satellite-based ground PM_{2.5} estimation using a gradient boosting decision tree. *Chemosphere* **2021**, *268*, 128801. [[CrossRef](#)]
27. Yang, N.; Shi, H.; Tang, H.; Yang, X. Geographical and temporal encoding for improving the estimation of PM_{2.5} concentrations in China using end-to-end gradient boosting. *Remote Sens. Environ.* **2022**, *269*, 112828. [[CrossRef](#)]
28. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
29. Chen, T.; Guestrin, C.; Assoc Comp, M. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
30. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
31. Zhong, J.; Zhang, X.; Gui, K.; Wang, Y.; Che, H.; Shen, X.; Zhang, L.; Zhang, Y.; Sun, J.; Zhang, W. Robust prediction of hourly PM_{2.5} from meteorological data using LightGBM. *Natl. Sci. Rev.* **2021**, *8*, nwaa307. [[CrossRef](#)]
32. Wei, J.; Li, Z.; Pinker, R.T.; Wang, J.; Sun, L.; Xue, W.; Li, R.; 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](#)]
33. He, W.; Zhang, S.; Meng, H.; Han, J.; Zhou, G.; Song, H.; Zhou, S.; Zheng, H. Full-Coverage PM_{2.5} Mapping and Variation Assessment during the Three-Year Blue-Sky Action Plan Based on a Daily Adaptive Modeling Approach. *Remote Sens.* **2022**, *14*, 3571. [[CrossRef](#)]
34. GB 3095-2012; China, M. Ambient Air Quality Standards. China Environmental Science Press: Beijing, China, 2012.
35. Guo, J.-P.; Zhang, X.-Y.; Che, H.-Z.; Gong, S.-L.; An, X.; Cao, C.-X.; Guang, J.; Zhang, H.; Wang, Y.-Q.; Zhang, X.-C.; et al. Correlation between PM concentrations and aerosol optical depth in eastern China. *Atmos. Environ.* **2009**, *43*, 5876–5886. [[CrossRef](#)]
36. Martins, V.S.; Lyapustin, A.; de Carvalho, L.A.S.; Barbosa, C.C.F.; Novo, E.M.L.M. Validation of high-resolution MAIAC aerosol product over South America. *J. Geophys. Res. Atmos.* **2017**, *122*, 7537–7559. [[CrossRef](#)]
37. Chudnovsky, A.A.; Kostinski, A.; Lyapustin, A.; Koutrakis, P. Spatial scales of pollution from variable resolution satellite imaging. *Environ. Pollut.* **2013**, *172*, 131–138. [[CrossRef](#)] [[PubMed](#)]
38. Tao, M.; Wang, J.; Li, R.; Wang, L.; 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](#)]
39. Lyapustin, A.; Wang, Y.; Korkin, S.; Huang, D. MODIS Collection 6 MAIAC algorithm. *Atmos. Meas. Technol.* **2018**, *11*, 5741–5765. [[CrossRef](#)]
40. 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](#)]
41. Fotheringham, A.S.; Crespo, R.; Yao, J. Geographical and Temporal Weighted Regression (GTWR). *Geogr. Anal.* **2015**, *47*, 431–452. [[CrossRef](#)]
42. Fan, J.L.; Ma, X.; Wu, L.F.; Zhang, F.C.; Yu, X.; Zeng, W.Z. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agric. Water Manag.* **2019**, *225*, 105758. [[CrossRef](#)]
43. Bentejac, C.; Csorgo, A.; Martinez-Munoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [[CrossRef](#)]
44. Wang, B.; Shi, W.; Miao, Z. Confidence Analysis of Standard Deviation Ellipse and Its Extension into Higher Dimensional Euclidean Space. *PLoS ONE* **2015**, *10*, e0118537. [[CrossRef](#)]
45. Gong, J. Clarifying the Standard Deviation Ellipse. *Geogr. Anal.* **2002**, *34*, 155–167. [[CrossRef](#)]

46. He, Q.; Gu, Y.; Zhang, M. Spatiotemporal trends of PM_{2.5} concentrations in central China from 2003 to 2018 based on MAIAC-derived high-resolution data. *Environ. Int.* **2020**, *137*, 105536. [[CrossRef](#)]
47. Xue, T.; Zheng, Y.; Tong, D.; Zheng, B.; Li, X.; Zhu, T.; Zhang, Q. Spatiotemporal continuous estimates of PM_{2.5} concentrations in China, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations. *Environ. Int.* **2019**, *123*, 345–357. [[CrossRef](#)] [[PubMed](#)]
48. Chinese State Council, Action Plan on Air Pollution Prevention and Control. 2013. Available online: http://www.gov.cn/zwggk/2013-09/12/content_2486773.htm (accessed on 26 July 2022). (In Chinese)
49. Chinese State Council, Three-Year Action Plan on Defending the Blue Sky (In Chinese). 2018. Available online: http://www.gov.cn/zhengce/content/2018-07/03/content_5303158.htm (accessed on 26 July 2022).
50. You, W.; Zang, Z.; Zhang, L.; Li, Y.; Pan, X.; Wang, W. National-Scale Estimates of Ground-Level PM_{2.5} Concentration in China Using Geographically Weighted Regression Based on 3 km Resolution MODIS AOD. *Remote Sens.* **2016**, *8*, 184. [[CrossRef](#)]
51. Ma, Z.; Hu, X.; Huang, L.; Bi, J.; Liu, Y. Estimating Ground-Level PM_{2.5} in China Using Satellite Remote Sensing. *Rev. Environ. Sci. Biol.* **2014**, *48*, 7436–7444. [[CrossRef](#)] [[PubMed](#)]
52. He, W.; Meng, H.; Han, J.; Zhou, G.; Zheng, H.; Zhang, S. Spatiotemporal PM_{2.5} estimations in China from 2015 to 2020 using an improved gradient boosting decision tree. *Chemosphere* **2022**, *296*, 134003. [[CrossRef](#)]
53. Boudriki Semlali, B.-E.; El Amrani, C. Satellite Big Data Ingestion for Environmentally Sustainable Development. In *Emerging Trends in ICT for Sustainable Development*; Springer: Cham, Switzerland, 2021; pp. 269–284. [[CrossRef](#)]
54. Semlali, B.-E.B.; Amrani, C.E. Amrani, C.E. A Stream Processing Software for Air Quality Satellite Datasets. In *Advanced Intelligent Systems for Sustainable Development (AI2SD'2020)*; Springer: Cham, Switzerland, 2022; pp. 839–853. [[CrossRef](#)]
55. Semlali, B.-E.B.; Amrani, C.E.; 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. [[CrossRef](#)]
56. Li, T.; Shen, H.; Yuan, Q.; Zhang, X.; Zhang, L. Estimating Ground-Level PM_{2.5} by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach. *Geophys. Res. Lett.* **2017**, *44*, 11,985–11,993. [[CrossRef](#)]
57. Li, T.; Shen, H.; Zeng, C.; Yuan, Q.; Zhang, L. Point-surface fusion of station measurements and satellite observations for mapping PM_{2.5} distribution in China: Methods and assessment. *Atmos. Environ.* **2017**, *152*, 477–489. [[CrossRef](#)]
58. Chen, Z.; Zhang, T.; Zhang, R.; Zhu, Z.-M.; Yang, J.; Chen, P.-Y.; Ou, C.-Q.; Guo, Y. Extreme gradient boosting model to estimate PM_{2.5} concentrations with missing-filled satellite data in China. *Atmos. Environ.* **2019**, *202*, 180–189. [[CrossRef](#)]
59. 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](#)]
60. He, Q.; Gao, K.; Zhang, L.; Song, Y.; Zhang, M. Satellite-derived 1-km estimates and long-term trends of PM_{2.5} concentrations in China from 2000 to 2018. *Environ. Int.* **2021**, *156*, 106726. [[CrossRef](#)] [[PubMed](#)]

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.