Technical Note

Clarifying Relationship between PM$_{2.5}$ Concentrations and Spatiotemporal Predictors Using Multi-Way Partial Dependence Plots

Haoze Shi [1,†] , Naisen Yang [2,†] , Xin Yang [1,*] and Hong Tang [3]

1 College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China
2 Environment Research Institute, Shandong University, Qingdao 266237, China
3 State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
* Correspondence: yangxin@bnu.edu.cn
† These authors contributed equally to this work.

Abstract: Atmospheric fine particles (PM$_{2.5}$) have been found to be harmful to the environment and human health. Recently, remote sensing technology and machine learning models have been used to monitor PM$_{2.5}$ concentrations. Partial dependence plots (PDP) were used to explore the meteorology mechanisms between predictor variables and PM$_{2.5}$ concentration in the "black box" models. However, there are two key shortcomings in the original PDP. (1) it calculates the marginal effect of feature(s) on the predicted outcome of a machine learning model, therefore some local effects might be hidden. (2) it requires that the feature(s) for which the partial dependence is computed are not correlated with other features, otherwise the estimated feature effect has a great bias. In this study, the original PDP's shortcomings were analyzed. Results show the contradictory correlation between the temperature and the PM$_{2.5}$ concentration that can be given by the original PDP. Furthermore, the spatiotemporal heterogeneity of PM$_{2.5}$-AOD relationship cannot be displayed well by the original PDP. The drawbacks of the original PDP make it unsuitable for exploring large-area feature effects. To resolve the above issue, multi-way PDP is recommended, which can characterize how the PM$_{2.5}$ concentrations changed with the temporal and spatial variations of major meteorological factors in China.

Keywords: PM$_{2.5}$ concentration; machine learning; partial dependence plots (PDP); geospatial visualization

1. Introduction

In recent years, due to the continuous increase in aerosols in developing countries, air pollution has become more serious [1,2]. PM$_{2.5}$ (the particulate matter with aerodynamic equivalent diameter $\leq$ 2.5 μm) has become the main pollutant in the urban environment [3,4]. it contains a variety of toxic and harmful substances and can travel deeply into the lungs and increase the risk of multiple diseases [5]. Since PM$_{2.5}$ is a serious threat to health, it should be monitored [6,7]. Understanding of causal and influential relation of PM$_{2.5}$ concentration and the diverse variables helps to control fine particles and improve air quality. Many studies concern the PM$_{2.5}$-variable relationships and investigate these relations through their spatial distribution patterns [8,9]. Some previous researches analyze PM$_{2.5}$-meteorology relation via the temporal correlation [10,11]. However, the complicated interaction between PM$_{2.5}$ concentrations and the influencing factors cannot be described well by linear correlation.

With the rapid development of remote sensing technology and machine learning methods, satellite-based aerosol optical depth and machine learning model has been widely

used to estimate large areas and high-resolution $PM_{2.5}$ concentrations [12]. These models can handle the complex nonlinear relationship between all predictor variables and $PM_{2.5}$ concentration, so they can generate more accurate and detailed $PM_{2.5}$ concentration distribution [13–16]. Based on such "black box" model with good performance, the previous studies try to explore the underlying mechanisms between predictor variables and $PM_{2.5}$ concentration.

Among interpretable machine learning methods, partial dependence plot (PDP) [17] gets much attention for its superior performance. Previous studies use PDP to investigate the effects of individual features, including temperature, aerosol optical depth (AOD) and winds, on the predicted $PM_{2.5}$ concentrations [18–21]. It can demonstrate the quantitative nonlinear relationships among variables. However, there are two key shortcomings to PDP: (1) it calculates the marginal effect of feature(s) have on the predicted outcome of a machine learning model, therefore some local effects might be hidden. A typical case is that surface winds substantially affect $PM_{2.5}$ concentration, but the PDP shape of winds is almost a horizontal line [14]. The horizontal line means $PM_{2.5}$ concentration does not vary with changes in winds, which is not the truth. However, because the increase [22] and reduce [23] effects of $PM_{2.5}$ concentration could cancel each other out among regions in China. (2) it requires that the feature(s) for which the partial dependence is computed are not correlated with other features, otherwise the estimated feature effect has a great bias. When the independent assumptions are not satisfied, the PDP will create data points that do not exist in reality. This allows the PDP to give opposite interpretation results for the same feature.

In this study, the contradictory relationship between $PM_{2.5}$ and temperature is drawn from the original PDP (hereinafter referred to as one-way PDP) and the spatiotemporal heterogeneity of $PM_{2.5}$-AOD relationship cannot be displayed well by one-way PDP. All results indicate there are uncertainties in the use and interpretation of one-way PDP and it is unsuitable for exploring large-area feature effects. To address the above shortcomings, the one-way PDP is expanded to multi-way PDP, which can reveal the regional and seasonal characteristics in the feature effects on $PM_{2.5}$ concentrations quantitatively. Through embedded geospatial-temporal joint code (GTJC) [14], it can explore the multiple relationships between predictor variables and $PM_{2.5}$ concentrations, i.e., the associations between meteorological factors, spatiotemporal factors and $PM_{2.5}$ concentrations. With the visualization of multi-way PDP, variation details of $PM_{2.5}$ concentration with temperature and AOD are analyzed and discussed. The results of this study suggest that multi-way PDP can give a relatively reliable relation between input features and model predictions.

2. Data

In order to better estimate $PM_{2.5}$ concentrations, four kinds of variables are used as the input data of the model, i.e., satellite aerosol products, Human activity proxies, geographic factors and meteorological re-analysis data.

2.1. Station $PM_{2.5}$ Observation

The hourly $PM_{2.5}$ data are collected from the air quality monitoring network which is administrated by the China National Environment Monitoring Center (CNEMC, http://www.cnemc.cn/, accessed on 1 April 2022). The observation station network is shown in Figure 1a and a total of 1701 monitoring points are located in 367 cities in China. The Beijing-Tianjin-Hebei, Yangtze River Delta, Sichuan Basin, and Pearl River Delta are compared as representatives of urban agglomerations in different geographical contexts. $PM_{2.5}$ observations obtained from air quality monitoring stations from 2015 to 2019 and processed into the daily average. The time series of daily average $PM_{2.5}$ of all sites is shown in Figure 1b. The daily average $PM_{2.5}$ data is matched to the grid cells of other features they fall in.
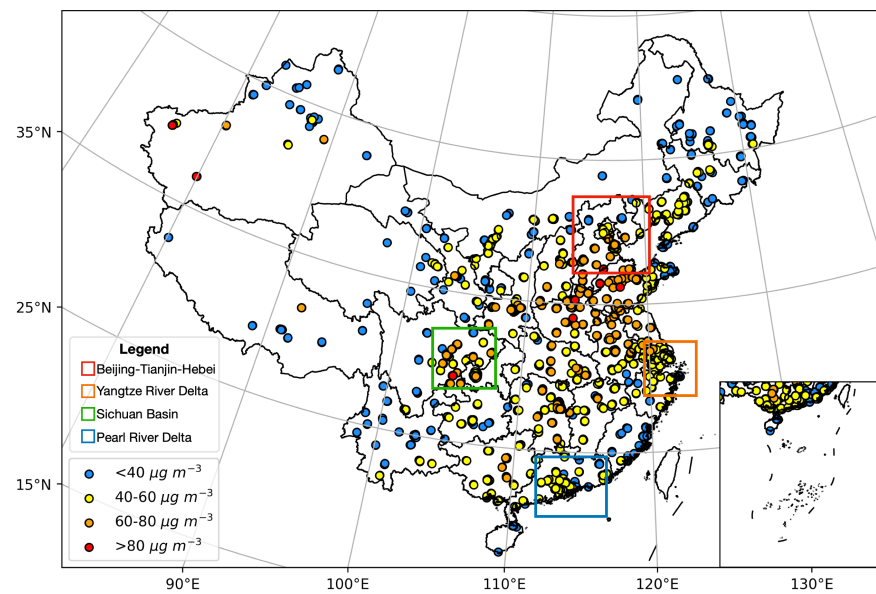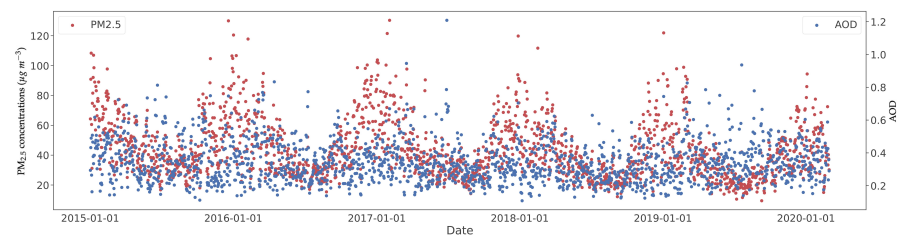
(**a**) The spatial distribution of monitoring sites

(**b**) Scatter plot of the daily average PM$_{2.5}$ concentrations and daily average AOD

Figure 1. (**a**) The spatial distribution of monitoring sites. The monitoring stations are illustrated as points and the color of the scattered point indicates the average of the PM$_{2.5}$ concentration (μg m^{-3}) from 2015 to 2019. (**b**) Scatter plot of the daily average PM$_{2.5}$ concentrations (μg m^{-3}) and daily average aerosol optical depth (AOD) of all observation sites from 2015 to 2019. The PM$_{2.5}$ concentrations have notable seasonal variations, but the fluctuation of AOD is relatively flat.

2.2. Satellite Aerosol Products

The aerosol optical depth (AOD) is physically related to PM$_{2.5}$, since it measures the reduction effect of aerosol on the light in the entire atmospheric column [24]. Thus, the daily AOD products (MCD19A2) based on Multiangle Implementation of Atmospheric Correction (MAIAC) [25] are taken as main predictors in this study. It comes from Moderate Resolution Imaging Spectroradiometer (MODIS) and can be accessed at Distributed Active Archive Center (DAAC, https://lpdaac.usgs.gov/products/mcd19a2v006/, accessed on 1 April 2022). A comparison based on AErosol RObotic NETwork (AERONET) AOD shows that the retrieval accuracy of MAIAC AOD is higher than other common retrieval algorithms (MODIS Dark Target/Deep Blue) [26].

2.3. Human Activity Proxies

Human activity proxies include the land use cover (LUC) and nighttime light (NTL) products, which come from MODIS MCD12Q1v006 data [27] and the Suomi NPP VIIRS [28]. The distribution of urban and built-up Lands can be reflected from the land use parameters and the NTL products can provide population density information. So the emission sources caused by human activities can be inferred [29].

2.4. Geographic Factors

Geographic factors are used to describe the surface topography, including the digital elevation model (DEM) product and the normalized difference vegetation index (NDVI) data. The vegetation cover and terrain can be important influencing factors affecting the cross-regional transport of $PM_{2.5}$ [12]. Thus, the Shuttle Radar Topography Mission (SRTM) DEM product [30] and the MODIS MOD13A3data [31] are collected to describe the transmission conditions.

2.5. Meteorological Re-Analysis Data

The variations of meteorological conditions can also affect the increase or decrease of $PM_{2.5}$ concentration [32]. The ERA5 reanalysis products [33,34] provided hourly meteorological data. We collected 8 variables from it, including boundary layer height (BLH), 2 m temperature (TEM), 2 m dew point temperature (T_d), surface pressure (SP), total precipitation (PRE), evaporation (ET), 10 m U-components of wind (U) and 10 m V-components of wind (V). Combined with TEM, T_d and SP, the relative humidity (RH) values were calculated. The wind direction (WD) and wind speed (WS) are converted from the 10m U-components and 10m V-components of wind.

In order to build a more reliable model to explain, a more rigorous data cleaning was performed in this study. If a sample is missing any of the features, it will be removed from the dataset. The observations with an abnormal value of features will also be considered invalid samples. Table 1 summarizes the abnormal criteria and the statistics of collected features. The $PM_{2.5}$ observation lasts for 12 h without change and will be considered an invalid observation. For extending the coverage of data, the each AOD of Terra and Aqua are averaged at every grid point. The hourly data of $PM_{2.5}$ concentrations and meteorological data are averaged into daily values to match the time resolution of AOD data. The spatial resolution of all datasets is unified with AOD products by bilinear interpolation.

Table 1. Statistics of collected features after data cleaning, including unit, range, mean, and standard deviation values.

Variable	Abbreviation	Unit	Range	Mean	STD
$PM_{2.5}$	$PM_{2.5}$	µg/m^3	[1, 500]	48.08	34.30
Aerosol optical dept	AOD	–	[0.01, 2.00]	0.35	0.26
Land use cover	LUC	–	[2, 17]	–	–
Night light	NTL	W cm^{-2}sr^{-1}	[0, 100]	25.37	17.07
Elevation	DEM	m	[0, 4517]	403.23	646.33
Normalized difference vegetation index	NDVI	–	[0, 0.97]	0.28	0.14
Boundary layer height	BLH	m	[18.56, 2782.63]	514.01	265.24
2m air temperature	TEM	K	[239.37, 313.25]	285.33	10.44
Relative humidity	RH	%	[1.74, 99.97]	47.20	18.67
Surface pressure	SP	kpa	[55.71, 104.51]	96.65	7.48
Total precipitation	PRE	mm	[0, 30]	2.28	5.17
Evaporation	ET	mm	[−6.75, 0.45]	−1.32	1.22
Wind direction	WD	degress	[0, 365]	179.84	103.05
Wind speed	WS	m/s	[0, 9.98]	1.93	1.27

3. Methodology

3.1. Overview

Figure 2 shows the workflow for exploring features-$PM_{2.5}$ relation in this study. First, the relationship of $PM_{2.5}$ and predictors is fitted by the gradient boosting decision tree (GBDT) model and geospatial-temporal joint codes (GTJC) model, respectively. Then, the built "black box" model is interpreted by permutation feature importances and partial dependence plots (PDP). The permutation feature importance is used to determine features of interest. The PDP can tell how $PM_{2.5}$ concentrations change quantitatively with a certain amount of variation of selected features based on the built model. Finally, the interpretation

results of the GBDT model and GTJC model are visualized and compared. The detail of each method will be separately described in the rest of this section.
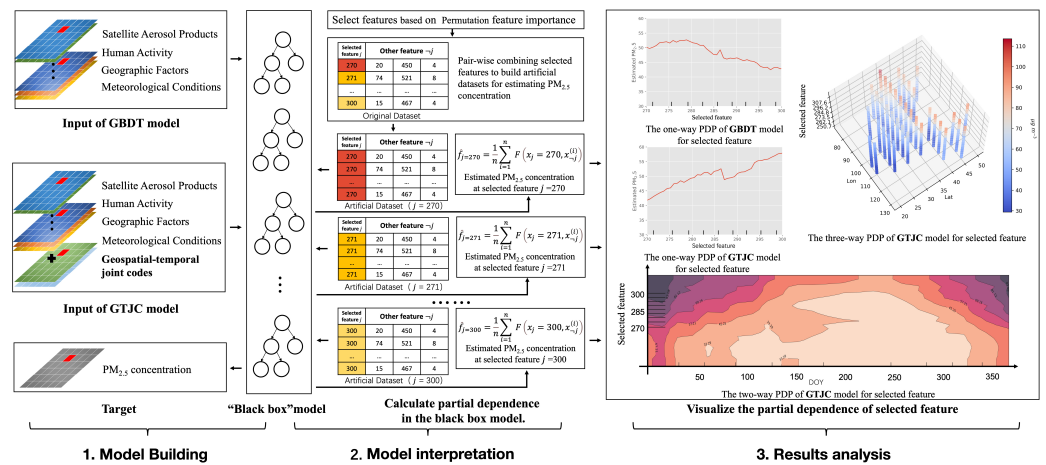

Figure 2. The workflow for exploring features-PM$_{2.5}$ relation, including model building, model interpretation and results analysis. Two models, geospatial-temporal joint codes (GTJC) and gradient boosting decision tree (GBDT) were built and interpreted by partial dependence plots (PDP). The calculation step of one-way PDP is illustrated in model interpretation. The interpretation results are visualized by multi-way PDP.

3.2. Model Building

3.2.1. Gradient Boosting Decision Tree Model

The gradient boosting decision tree (GBDT) [35] is a commonly used machine learning model for PM$_{2.5}$ concentration estimation [36–38]. The results of GBDT are a combination of the output of weak learners. Each added weak learner will fit the residual between the target and the model output in the previous stage. In PM$_{2.5}$ concentrations estimation, the input data of GBDT model **x** include [AOD, LUC, NDVI, DEM, NTL, TEM, SP, WS, WD, BLH, PRE, ET, RH] and the target **y** is PM$_{2.5}$ concentrations.

3.2.2. Geospatial-Temporal Joint Codes Model

The relationship between PM$_{2.5}$ concentrations and input data of GBDT, especially AOD, is spatial and temporal non-stationary [39]. The geospatial-temporal information is required by the model, otherwise the model would assume that all input data occurred at the same time and place. Therefore, geospatial-temporal joint codes (GTJC) [14] is applied to this study to help the model capture the spatiotemporal patterns. The day of the year (DOY), latitude (ϕ) and longitude (θ) of the given sample are converted to temporal codes (TC) and geospatial codes (GC) by a specific coding method, which has been described in Equation (1) and Figure 3.

$$\boldsymbol{TC} = \left[\begin{array}{c} t_x \\ t_y \end{array}\right] = \left[\begin{array}{c} \cos\left(2\pi\frac{DOY}{T}\right) \\ \sin\left(2\pi\frac{DOY}{T}\right) \end{array}\right], GC = \left[\begin{array}{c} g_x \\ g_y \\ g_z \end{array}\right] = \left[\begin{array}{c} R\sin\theta\cos\varphi \\ R\sin\varphi \\ R\cos\theta\cos\varphi \end{array}\right]. \quad (1)$$

Here, T is 365 or 366 and R is the radius of the Earth. R is set to 1 for normalization in the experiment. The day of the year (DOY) and the latitude (ϕ) and longitude (θ) of the given sample are converted into polar coordinates and spherical coordinates, respectively. TC and GC essentially are spatiotemporal information of the given sample in Cartesian coordinates; t_x and t_y and g_x, g_y and g_z are the transformed coordinates.

The GTJC will be coupled with the observed data and this model is named the GTJC model. It should note that the difference between the GBDT model and the GTJC model is the absence or presence of GTJC as input.
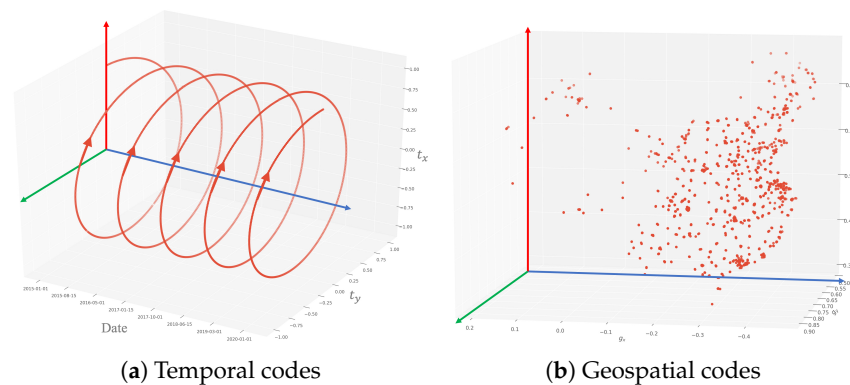
(a) Temporal codes

(b) Geospatial codes

Figure 3. Illustration of the (**a**) temporal codes (TC) and (**b**) geospatial codes (GC), which will be input as variables to the GTJC model.

3.2.3. Hyperparameter Selection

The model is built with scikit-learn [40]. Classification-and-regression-trees (CART) [41] is employed as the weak learner. $PM_{2.5}$ concentration is estimated by CART through the following 5 steps: (1) Randomly sample 66% input data from the training data set for fitting the individual weak learners and randomly select 4 features for determining the best split; (2) The input data was assigned to two branches D_1 and D_2 based on a random split. This split was generated by selecting one arbitrary number a between the maximum and minimum of one feature A in step 1; (3) Calculate the squared error of D_1 and D_2, where c_1, c_2 is average of $PM_{2.5}$ concentrations in D_1 and D_2, respectively. Repeat step 2 to find the best feature A and split point a to minimize squared error; (4) Repeat step 2, 3 on D_1 and D_2 until the depth of the tree reaches 10; (5) The mean value of $PM_{2.5}$ concentrations of input data in the leaf node will be taken as the estimated $PM_{2.5}$ concentrations. The final model was constructed by 22,994 CART. The learning rate was set to 0.05 to shrink the contribution of each CART.

The most commonly used 10-fold cross-validation (10-CV) is adopted to evaluate the reasoning ability of the model [42–44]. For measuring the performance of regression, three quality metrics are used, i.e., determination coefficient (R^2), mean absolute error (MAE), and root-mean-square error (RMSE).

3.3. Model Interpretation

3.3.1. Permutation Feature Importance

The feature importance can be measured by the decrease of the model score when this feature is randomly shuffled, which is also known as Permutation feature importance. In this study, the number of random shuffles is set to 10 times and the model score is measured by the mean squared error.

3.3.2. Partial Dependence Plots

The partial dependence plots (PDP) is a method for interpreting black-box models [45]. It is defined as

$$\hat{f}_j(x_j) = E_{x_{\neg j}}\big[F(x_j, x_{\neg j})\big] = \int F(x_j, x_{\neg j}) dP(x_{\neg j}). \qquad (2)$$

where x is the instances in the dataset and j is the j-th feature that we want to know the effect on the prediction $\hat{f}_j$. $\neg j$ are other features of the black-box model F. E is mathematical expectation and P is probability. The partial function $\hat{f}_j(x_j)$ shows the relationship between j-th feature and the predicted targets.

The illustration of the calculation step of one-way PDP is shown in the model interpretation of Figure 2. First, the selected features are combined with other features one by one to build artificial datasets. Then, all artificial datasets are put into the 'black box' model to estimate $PM_{2.5}$ concentrations. The average of estimated $PM_{2.5}$ concentrations will be

taken as the partial dependence of the corresponding feature. The partial dependence is estimated by marginalizing the predicted targets over the distribution of other features $\neg j$. Therefore, PDP can directly show how the prediction variates change with the input feature. However, as shown in Equation (2), the original PDP, i.e., one-way PDP, only shows the average marginal effects and lost some local effects. In order to show the hidden local effects and the detailed information, we expand PDP from one-way to multi-way. In the calculation of two-way PDP, the selected feature needs to be pair-wise combined with TC first, and then continue to follow the calculation steps of one-way PDP. Finally, to better understand the feature effects of the selected feature, we trace back to corresponding the day of year and visualize the partial dependence results. For three-way PDP, the selected feature is combined with geospatial codes.

4. Results

This section presents the results of both the model performance and interpretation methods. Section 4.1 shows the model performance. The relative importance of the feature is demonstrated in Section 4.2. Section 4.3 investigates TEM-PM$_{2.5}$ concentrations relationship through comparing the one-way PDP of gradient boosting decision tree (GBDT) model and geospatial-temporal joint code (GTJC) model and its annual variation presented in the two-way PDP. Section 4.4 reveals the impacts of water vapor on the AOD-PM$_{2.5}$ concentrations relationship via analysis of seasonal, regional and multi-way PDPs.

4.1. Model Performance

The results of sample-based 10-fold cross-validation are displayed in Table 2. The R^2 value of gradient boosting decision tree without geospatial-temporal joint code (GTJC) is 0.82. The R^2 value is increased to 0.85 after incorporating the spatial information. However, with the aid of the temporal information, R^2 is increased to 0.86. Compared with the general gradient boosting decision tree model, the GTJC improves the prediction performance in terms of both predictive quality and estimated uncertainty. The R^2 of the GTJC model is 0.89. This result indicates that PM$_{2.5}$ concentrations observed by sites across China could be well explained by a combination of satellite-based aerosol products, geographic factors, human activity, surface meteorological conditions, time variables and space variables in the GTJC model.

Table 2. The mean and variance of sample-based 10-fold cross-validation results of gradient boosting decision tree (GBDT), GBDT with geospatial codes, GBDT with temporal codes, and GBDT with geospatial-temporal joint codes.

Method	R^2	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)
Gradient boosting decision tree (GBDT)	0.8226 ± 0.0020	14.4435 ± 0.1551	9.6076 ± 0.0619
GBDT + geospatial codes	0.8532 ± 0.0021	13.1376 ± 0.1444	8.6060 ± 0.0593
GBDT + temporal codes	0.8639 ± 0.0016	12.6532 ± 0.1385	8.3633 ± 0.0596
GBDT + geospatial-temporal joint codes	0.8898 ± 0.0013	11.3814 ± 0.0993	7.4157 ± 0.0368

4.2. Relative Importance of Features

Variables with a larger value of permutation feature importance have a greater influence on the model. Figure 4a,b shows the permutation feature importance of the GTJC model and GBDT model separately.

In the GTJC model, the temporal codes (TC) and geospatial codes (GC) are the two most important features, whose importance values are 36.77% and 17.17%, respectively. This is consistent with previous studies that PM$_{2.5}$ concentrations vary drastically over time and space [46–48], therefore, the estimation of PM$_{2.5}$ concentrations across entire China is highly depending on geospatial location and time period. Generally, in winter, weather conditions are not in favor of the pollutants diffusion that can elevate the PM$_{2.5}$ concentrations [49,50]. The emission amount is larger in northern China than in southern China due to the distribution of industry and indoor heating [51,52]. These synchronously

changing in pollution emission, weather conditions and PM$_{2.5}$ concentrations make TC and GC the most important factors in estimating PM$_{2.5}$ concentrations. Besides the temporal and geospatial codes, the top-4 important features also include aerosol optical depth (AOD), and temperature (TEM). Other features have less contribution to the model, with an importance lower than 10%.
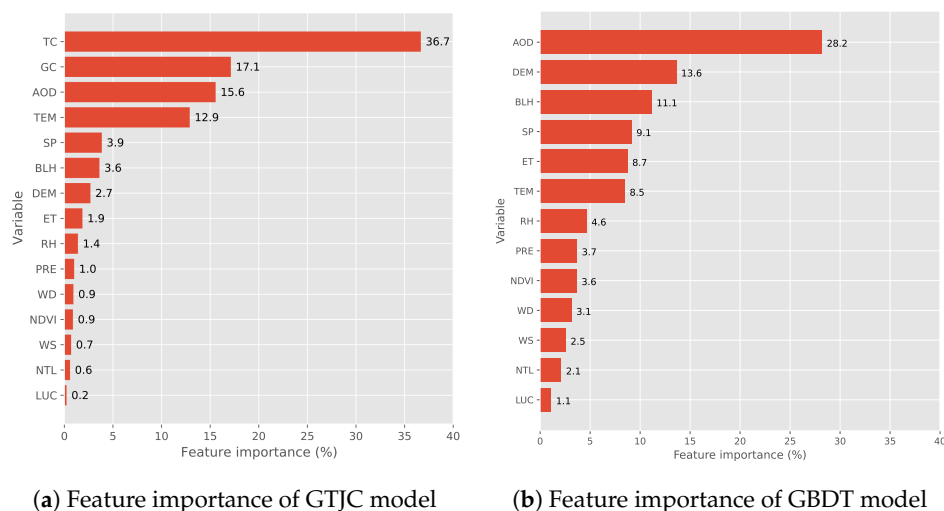

(**a**) Feature importance of GTJC model   (**b**) Feature importance of GBDT model

Figure 4. Permutation feature importances of (**a**) geospatial-temporal joint codes model (GTJC) model and (**b**) gradient boosting decision tree (GBDT) model. The calculated feature importance values are normalized to percentages and sorted in descending order of importance.

In the GBDT model, the importance value of AOD is 28.2%, which is much greater than that of other variables. This is because AOD is the total extinction of particulate matter in the entire atmospheric column, thus physically connected with PM$_{2.5}$ concentration. The other input variables are the influencing factors of PM$_{2.5}$ concentration. There, the contribution of AOD to PM$_{2.5}$ concentration estimation is much greater than other factors. It is worth noting that the values of AOD do not discriminate the extinction for fine and coarse particles. Moreover, AOD is for the whole atmosphere column, but PM$_{2.5}$ concentration is measured at the surface. Therefore, the AOD value does not strictly correspond to changes in PM$_{2.5}$ concentrations. This inconsistency can be shown in Figure 1b, which may partially explain why the feature importance of AOD is falling lower than that of geospatial-temporal joint codes. The other features with an importance value over 10 % are DEM and BLH. In China, the altitude in the eastern plains is low and there will be more human activities here. Thus, the PM$_{2.5}$ concentration there is usually high. On the Tibet Plateau in the west of China, the altitude is highest and the PM$_{2.5}$ concentration is the lowest [53]. The PM$_{2.5}$-meteorology interactions are closely related to the variation in boundary layer height, which determines the vertical structure and diffusion height of PM$_{2.5}$. There is very little space for dispersion of PM$_{2.5}$ if the BLH is small. This results in a higher PM$_{2.5}$ concentration with the same amount of aerosol particles [54–56].

4.3. TEM-PM$_{2.5}$ Concentrations Relationship Revealed by Multi-Way PDP

The entire China, the Beijing-Tianjin-Hebei, the Yangtze River Delta region, the Sichuan Basin, and the Pearl River Delta region were selected as study areas to analyze the relationship between PM$_{2.5}$ and temperature (TEM). Figure 5 present the density scatterplots of TEM and observed PM$_{2.5}$ of those regions. In each subfigure, statistical results are attached with the linear regression relation, the number of samples (N), the Pearson correlation coefficient (Pearson's r). In our dataset, the Pearson's r calculated for these five regions are all negative. The slopes of linear regressions are also negative.
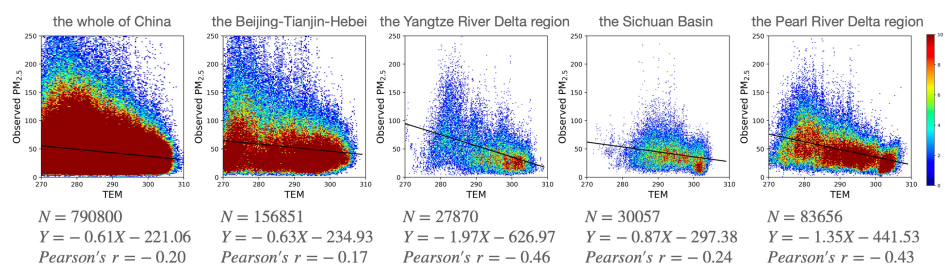
Figure 5. The density scatterplots of temperature (TEM) and observed PM$_{2.5}$ in the whole of China, the Beijing-Tianjin-Hebei, the Yangtze River Delta region, the Sichuan Basin and the Pearl River Delta region. In each subfigure, statistical results are attached with the linear regression relation, the number of samples (N), the Pearson correlation coefficient (Pearson's r).

Figure 6 shows one-way PDP of TEM of the GBDT model and GTJC model across China. The PDP of GTJC model over mainland China is similar to that over the four regions, except for the difference in the range of estimated PM$_{2.5}$ concentrations. All these PDPs of GTJC show that the estimated PM$_{2.5}$ concentrations increased with the rising temperature. On the contrary, one-way PDPs of the GBDT model suggest that there is an inverse relationship between TEM and PM$_{2.5}$ concentrations. The negative TEM-PM$_{2.5}$ relation given by the one-way PDP of the GBDT model is more consistent with the TEM-PM$_{2.5}$ relation demonstrated by the observed dataset in Figure 5.
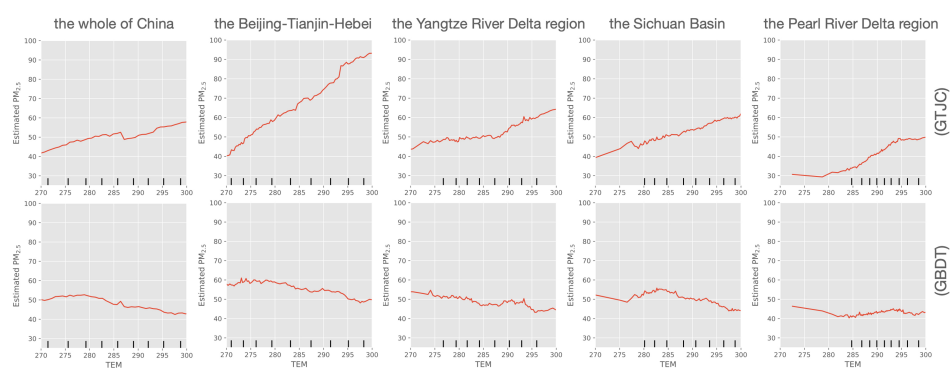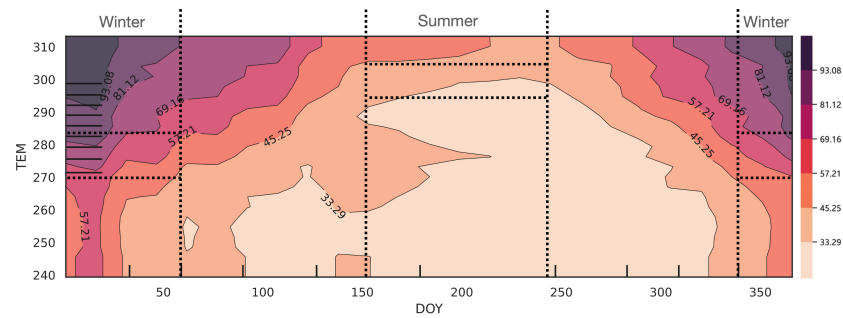

Figure 6. the one-way partial dependence plots (PDP) of temperature (TEM) of GBDT model and GTJC model. The black bars on the axes show the deciles of the feature values in the training data.

As stated in Section 3.2.2 that the difference between the GBDT model and the GTJC model is the absence or presence of GTJC as input. Therefore, the GTJC is speculated to be the reason for this discrepancy. The relation between TEM and PM$_{2.5}$ concentrations in the one-way PDPs is almost the same among the four regions and mainland China for model with and without GTJC, respectively, suggesting the false relation is irrelevant to geospatial codes and induced by temporal codes. The contradictory relationship given by the one-way PDP will be analyzed and discussed in Section 5.1.

To further explore the relationship between TC, TEM, PM$_{2.5}$, we expand PDP to two-way to show how this relation variates with the day of the year. Two-way PDP in Figure 7a displays the joint dependence of estimated PM$_{2.5}$ concentrations on both TC and TEM. Figure 7b,c presents the average of TEM and observed PM$_{2.5}$ concentrations every month, respectively.

As shown in Figure 7b, the temperature is high in summer and low in winter. In the winter (i.e., December, January, and February.), the TEM is mainly distributed from 270 to 285 K. In the summer (i.e., June, July, August), the TEM reached 295 to 305 K. In Figure 7a, when the TEM is between about 270 and 285 K, the estimated PM$_{2.5}$ concentration range is from 45.25 to 81.12 μg m^{-3}. In the summer, when the TEM reached to the range between 295 and 305 K, and the estimated PM$_{2.5}$ concentration changes from 33.29 to 45.25 μg m^{-3}.

The seasonal variation in TEM and estimated $PM_{2.5}$ are diametrically opposed. Therefore, estimated $PM_{2.5}$ concentrations are inversely related with TEM in the annual cycle, i.e., lighter pollution (higher TEM) in summer and heavier pollution (lower TEM) in winter. In Figure 7c, the observed $PM_{2.5}$ ranges from 22 to 100 µg m^{-3} in winter and 10 to 50 µg m^{-3} in summer. Similarly, the variation of observed $PM_{2.5}$ concentration is contrary to the variation of TEM. There is a negative correlation between observed $PM_{2.5}$ and TEM. This result is consistent with that of estimated $PM_{2.5}$ concentrations. It should be noted that the $PM_{2.5}$ in Figure 7a is estimated by marginalizing the predicted targets over the distribution of unselected features. It will be different from the observed $PM_{2.5}$ concentration.

(**a**) The two-way PDP of TEM
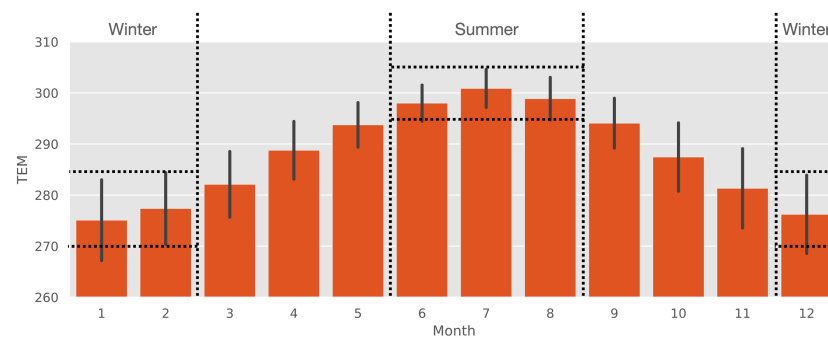
(**b**) The annual average of TEM
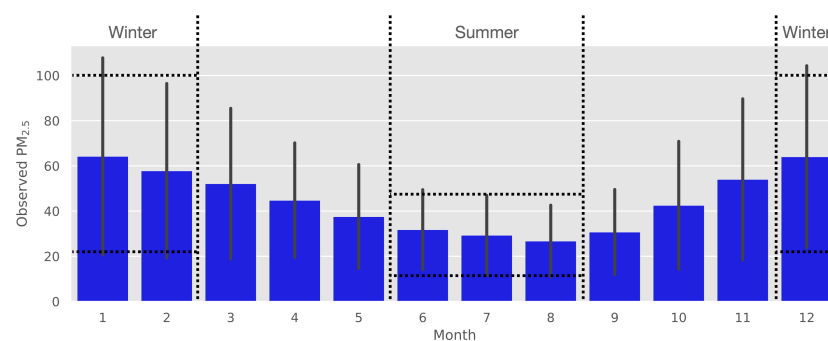

(**c**) The annual average of observed $PM_{2.5}$

Figure 7. (**a**) The two-way partial dependence plots (PDP) of temperature (TEM) for the GTJC model. The contour lines represent the average estimated $PM_{2.5}$ concentrations (µg m^{-3}). The black bars on the axes show the deciles of the feature values in the training data. (**b**) Mean values of the TEM at a monthly scale. (**c**) Mean values of the $PM_{2.5}$ concentrations at a monthly scale. The error bar referred to the standard deviation of the data set.

TEM in most regions of mainland China has a distinct annual cycle with hot summer and cold winter due to its location in the northern temperate zone. Simultaneously, $PM_{2.5}$ concentrations in summer stay at a low level because of the rainy weather and less emission

compared with the heating season. This coincidentally seasonal variation in both TEM and PM$_{2.5}$ concentrations makes them have a natural inverse relationship on the yearly scale. The negative TEM-PM$_{2.5}$ concentrations relation is reflected by the one-way PDP of GBDT model in Figure 6.

4.4. AOD-PM$_{2.5}$ Concentration Relationship Revealed by Multi-Way PDP

As shown in Figure 8a,b, the one-way PDPs in different seasons and two-way PDP shows seasonal changes of the AOD-PM$_{2.5}$ relationship. For the same AOD value, the corresponding PM$_{2.5}$ concentration value varies greatly from winter to summer. A more reliable AOD-PM$_{2.5}$ concentration relationship for the individual region is revealed. However, two-way PDP can visualize annual variations in PM$_{2.5}$ concentrations more detailed and intuitively with the changes in AOD across China.
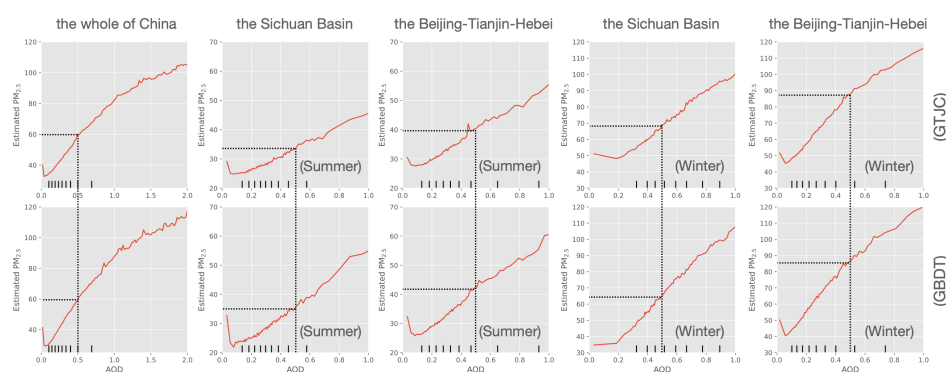
(**a**) The one-way PDP of AOD in summer and winter
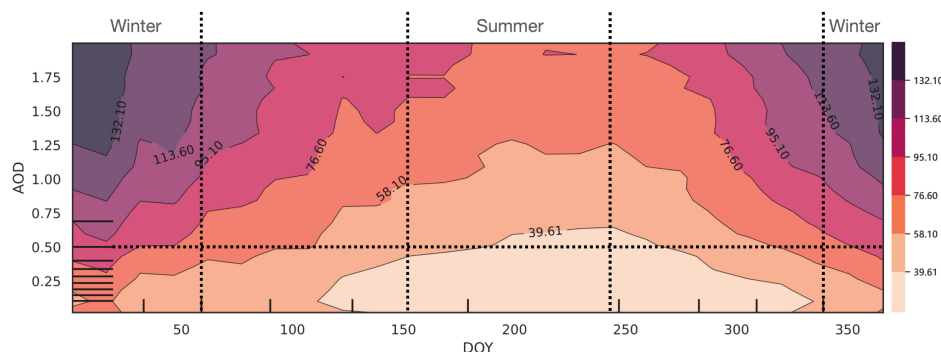
(**b**) The two-way PDP of AOD

Figure 8. (**a**) The one-way partial dependence plots (PDP) of aerosol optical depth (AOD) in the Beijing-Tianjin-Hebei and the Sichuan Basin during summer and winter. The dotted line represents the estimated PM$_{2.5}$ concentration when AOD is 0.5. (**b**) The two-way PDP of AOD. The contour lines on the subfigure (**b**) represent the average estimated PM$_{2.5}$ concentrations (μg m^{-3}). The black bars on the axes show the deciles of the feature values in the training data.

The AOD-PM$_{2.5}$ concentration relation in one-way PDPs are all positive for the whole China, the Beijing-Tianjin-Hebei and the Sichuan Basin in both summer and winter, regardless of whether there is a GTJC in the model. Note that different from the relation of TEM-PM$_{2.5}$ concentrations, PM$_{2.5}$ concentrations are physically relative to AOD values. The positive correlation of TEM-PM$_{2.5}$ concentrations within the season is not the direct relationship, but the regulated expression of the strengthened photochemical reaction and secondary pollutants formation under higher environment temperature. However, PM$_{2.5}$ concentration dose positively related to AOD because they are quantities representing aerosol loading in the atmosphere.

As the columnar integration of the extinction coefficient with altitude, satellite AOD is affected by the atmospheric humidity since the uptake of water vapor can cause a substantial increase of aerosol size and its extinction. Whilst $PM_{2.5}$ is often measured under a dry condition and less affected by the atmospheric humidity. Therefore, atmospheric humidity difference may result in changes in AOD-$PM_{2.5}$ relationship. Figure 9a shows the probability density of relative humidity (RH) in Sichuan Basin and the Beijing-Tianjin-Hebei during winter and summer. As we can see, there is a huge contrast in humidity between the Sichuan Basin and the Beijing-Tianjin-Hebei. In winter, the RH in the Sichuan basin is mainly distributed in 60 to 80% and it is 20 to 50% in the Beijing-Tianjin-Hebei.

With the distinct humidity contrast, the one-way PDP of both the GTJC model and the GBDT model for the Sichuan Basin in winter is markedly different from that of the Beijing-Tianjin-Hebei in winter. The deciles of AOD indicated that the AOD of Sichuan Basin (0.3 to 0.9) is greater than that of the Beijing-Tianjin-Hebei (0.1 to 0.7) in winter but the variation range is smaller for $PM_{2.5}$ concentrations over the Sichuan Basin in winter than that of the Beijing-Tianjin-Hebei in winter. In GTJC model, the $PM_{2.5}$ concentrations vary from $50\ \mu g\ m^{-3}$ to $100\ \mu g\ m^{-3}$ in the Sichuan Basin, which is smaller than that of Beijing-Tianjin-Hebei (45 to $115\ \mu g\ m^{-3}$). In GBDT model, the variation range for $PM_{2.5}$ concentrations of the Sichuan Basin is 35 to $110\ \mu g\ m^{-3}$ and it is 40 to $120\ \mu g\ m^{-3}$. The flatter of one-way PDP means the variation of AOD results in weaker changes in the estimated $PM_{2.5}$ concentrations. The larger AOD in winter is not necessarily due to the high concentration of $PM_{2.5}$, but due to the increased extinction effect by the hygroscopic of particles in higher environmental humidity. This is in line with the situation that in the Sichuan Basin during humid winter, changes in AOD may be more closely related to changes in water vapor rather than variations in $PM_{2.5}$ concentrations. Both the GBDT and GTJC models can distinguish the impacts of water vapor on $PM_{2.5}$-AOD relation.
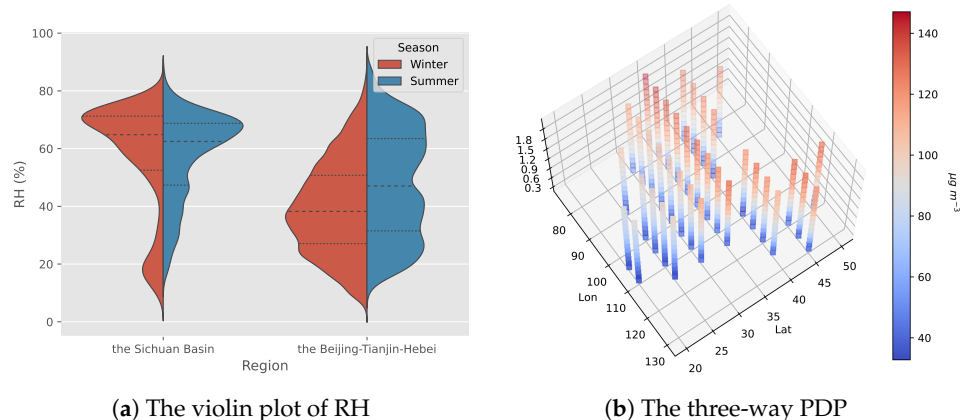

(**a**) The violin plot of RH   (**b**) The three-way PDP

Figure 9. (**a**) The violin plot of relative humidity (RH). The dotted line on (**a**) represents the quartiles of the RH distribution. (**b**) The three-way PDP of geospatial codes (GC) with respect to AOD. The latitude and longitude are inversely transformed from GC. The vertical axis of the three-way PDP is the selected feature, and the color represents the average estimated $PM_{2.5}$ concentration ($\mu g\ m^{-3}$).

The joint dependence of $PM_{2.5}$ concentrations on geographical codes and AOD is presented with three-way PDP in Figure 9b. According to Figure 9b, the sensitivity of estimated $PM_{2.5}$ concentrations to AOD is higher in the north than that in the south of China. This agrees well with the one-way PDPs in Figure 8a that the corresponding increase of $PM_{2.5}$ concentration of a certain variation in AOD is smaller in South China than that in North China. As analyzed above, the prime reason for this South–North contrast is that the higher atmospheric water vapors in South China swell the AOD value by hygroscopic growth of particles. Similarly, the East–West difference of the sensitivity of $PM_{2.5}$ concentrations to AOD in the three-way PDP is also mainly originated from the atmospheric humidity difference between East and West regions in China. In addition, the higher fraction of cloud

cover may introduce a large uncertainty into the AOD retrieval [57], leading to the less sensitivity of estimated PM$_{2.5}$ concentration to AOD variation over southern China.

5. Discussion

5.1. The Contradictory Relationship Shown in the One-Way PDP

As stated in Section 4.3, for the same feature, temperature, the one-way PDP will give two opposed results since temporal codes. Here, Temporal codes are essentially spatiotemporal information of the given sample in Cartesian coordinates and t_x, t_x are the transformed coordinates. The daily average of PM$_{2.5}$ concentrations, TEM, t_x and t_y are shown in Figure 10. The PM$_{2.5}$ concentrations have notable seasonal variations, with the heaviest pollution in winter. The annual cycle is also substantial in TEM, but with the lowest values of TEM in winter. The Pearson correlation coefficient of PM$_{2.5}$ and TEM is -0.17. Clearly, PM$_{2.5}$ concentrations are negatively correlated with TEM from the perspective of annual changes. The variation of TC, represented by t_x and t_y, has a very similar annual cycle to that of PM$_{2.5}$ concentrations. The Pearson correlation coefficients between PM$_{2.5}$ and t_x, t_y are 0.34, 0.08, respectively. For better describe the relationship among TC, TEM, and PM$_{2.5}$, we claim that TC is positively correlated with PM$_{2.5}$ even if TC consists of two elements. The Pearson correlation coefficients between TEM and t_x, t_y are -0.78 -0.15, respectively. It should be noted that TEM is physically determined by the seasonal changed solar radiation. Therefore, the two input variables of TEM and TC are naturally strongly correlated with each other.
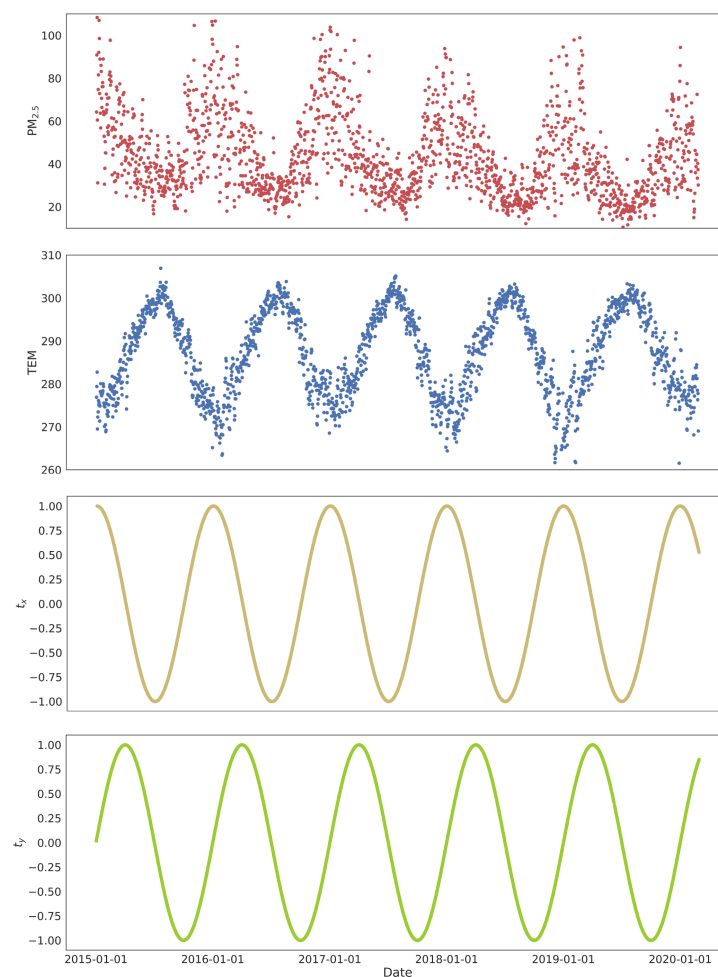

Figure 10. Scatter plot of the daily average PM$_{2.5}$ concentrations (μg m^{-3}), temperature (TEM), t_x and t_y of all observation sites from 2015 to 2019.

As shown in Figure 2, TEM are combined with TC one by one to build artificial datasets for estimating partial dependence. Based on the strong negative correlation of TEM and TC, a paired combination is not suitable for these two features. Therefore, some data points that do not exist in reality will be created, for example, extremely cold temperatures occur in the summer. These data points will be involved in the estimation of the average of $PM_{2.5}$ concentrations.

For the complicated interactions between different factors, the extracted correlation coefficient between two variables could be influenced significantly by other variables in the ecosystem. In addition, correlation analysis between two variables in a complicated ecosystem might lead to mirage correlations [58]. Similarly, it is unreliable to analyze complex patterns learned by a "black box" model from the perspective of a feature, especially when the selected feature has a strong correlation with other predictors.

The coincidentally seasonal variation in TEM, $PM_{2.5}$ and TC cause a strong correlation among them. However, the one-way PDP can only reveal the relationship between $PM_{2.5}$ and a single feature. Furthermore, as shown in Figure 4a, TC has more influence on the model compared to TEM. Thus, the relationship between TEM and $PM_{2.5}$ given by one-way PDP can be very partial. Actually, Christoph Molnar [45] indicates that the correlated features would complicate the partial dependence analysis. The results in this study further suggests the effect of the powerful feature with the output seems to cover up that of less influential feature with the output.

Note that within a certain period, such as a time scale of months, seasons, etc., $PM_{2.5}$ concentrations increase with the increase of TEM, according to the two-way PDP in Figure 7a. The possible explanations are that more precursors of particles and secondary pollutants are generated under high temperatures, since it promotes the photochemical reaction in the atmosphere. Thereby resulting in the rise in $PM_{2.5}$ concentrations within the corresponding period [32].

5.2. The Single Perspective of One-Way PDP

The one-way PDP can only demonstrate the relationship between a single predictor and target. In addition, the partial dependence is estimated by marginalizing the predicted target over the distribution of unselected features, as shown in Equation (2). Therefore some local effects might be hidden and quantifying the relationship from one-way PDP may deviate from the actual local situation.

For example, the dotted line in Figure 8a,b represents the estimated $PM_{2.5}$ concentration when AOD is 0.5. It is 60 µg m^{-3} in the whole of China both in the GTJC model and GBDT model when AOD is 0.5. However, in summer, the estimated $PM_{2.5}$ concentration of the GTJC model and GBDT model both is around 35 µg m^{-3} in Sichuan Basin and 40 µg m^{-3} in the Beijing-Tianjin-Hebei. Those results in summer are lower than the national average $PM_{2.5}$ concentration. Meanwhile, in the winter, the estimated $PM_{2.5}$ concentration of the GTJC model and GBDT model both reached 65 µg m^{-3} in Sichuan Basin and 85 µg m^{-3} in the Beijing-Tianjin-Hebei. The $PM_{2.5}$ concentrations estimated in winter are higher than the national average. In terms of regions, the estimated $PM_{2.5}$ concentration also have slight differences.

These sub-regional and sub-seasonal results indicate that the $PM_{2.5}$-AOD relationship has spatiotemporal heterogeneity. The heterogeneous interactions might be hidden because one-way PDP can only quantify the influence of individual factors in an average state. However, the expanded multi-way partial dependence plots can visualize spatiotemporal variations of $PM_{2.5}$ concentrations more detailed and intuitively with the changes in selected factors across China. As mentioned above, the drawbacks of the original one-way PDP make it unsuitable for exploring large-area feature effect. To reveal the spatiotemporal characteristics of feature effect, it is recommended to visualize the joint dependence in the form of multi-way PDP.

5.3. The Effect of Incorporating Spatial and Temporal Terms

The spatial and temporal variables are widely incorporated into $PM_{2.5}$ concentrations estimation models because they can address the spatiotemporal heterogeneities of the $PM_{2.5}$-AOD relationship [12]. The common forms of spatial and temporal terms include the days of year [59], Unix Epoch time [21,60], Haversine [15], two-dimension smooth term of geographical coordinate [61]. The geospatial-temporal joint codes [14] used in this study is also one of the forms spatiotemporal variables. However, not all spatiotemporal variables will affect the results of one-way PDP.

The time variable, days of the year, is converted into a cycle with a range of $[0, 2\pi]$ in this study to solve the problem of the time variable jumping from 365/366 on the last day of a year to 1 on the first day of the next year. This conversion enables the time variable variates seasonally, resulting in the strong correlation among TC, temperature and $PM_{2.5}$ concentrations. Other $PM_{2.5}$ concentration prediction models with temporal codes but in the form of Unix Epoch time (number of seconds since 1 January 1970) get negative TEM-PMPM$_{2.5}$ concentrations relationship [21,60], verifying the above suppose that it is the strong correlation between TC and temperature rather than TC itself that makes the model get different TEM-$PM_{2.5}$ concentrations relation.

Spatiotemporal variables can also make the one-way PDP more reliable in the case of sparse data points. As shown in Figure 8a, in the summer of the Sichuan basin, AOD is mainly distributed in the range of 0.1 to 0.6. It almost has no data points between 0.6 and 1.0. The estimated $PM_{2.5}$ concentrations of the GBDT model changed by more than $15\ \mu g\ m^{-3}$ when AOD changed from 0.1 to 0.6 and it is $8\ \mu g\ m^{-3}$ in GTJC model. This is where the two models' one-way PDP differ most. When the distribution of data points is sparse, the spatiotemporal variables can offer additional information to the model's decision. This is also why GTJC can improve the model's ability.

6. Conclusions

The success of the machine learning model in estimating $PM_{2.5}$ concentrations from satellite remote sensing data has led to the pursuit of interpretability. Due to the high spatiotemporal heterogeneity, understanding the relationship between predictors and $PM_{2.5}$ concentration remains challenging. To obtain a more reliable relation of temperature (TEM) and AOD with $PM_{2.5}$ concentrations, partial dependence plots (PDP) and their extended forms are used to thoroughly explore the relationship of the prediction and input features in the machine learning model.

Temperature is found to be inversely correlated with $PM_{2.5}$ concentration in the one-way PDP of the model without geospatial-temporal joint codes (GTJC) and in the annual scale of the two-way PDP with GTJC. The comparison of one-way PDPs between the model with and without GTJC suggests that the strong correlations of temporal codes with $PM_{2.5}$ concentration and TEM lead to an artificial positive TEM-$PM_{2.5}$ concentration relation. Future studies need to use one-way PDP with caution because it is susceptible when the input variables have a correlation. As recommended in this study, the extension of PDP from one-way to two-way can clarify to some degree of the TEM-$PM_{2.5}$ concentration relation by visualizing its annual evolution.

For AOD-$PM_{2.5}$ concentration relation, If we analyze the one-way PDP of the whole of China, the distinct regional and seasonal characters would be hidden. From the perspective of obtaining empirical and numerical relationships, one-way PDP is not suitable for exploring AOD-$PM_{2.5}$ concentration relation over large-scale regions of high inhomogeneous. Therefore, we combine GTJC with the model and thus expand PDP to two-way and three-way to dive into the details of seasonal and regional variations.

This study let us dive into the spatial and temporal characters of the relationship of AOD, TEM and $PM_{2.5}$ concentration. The results suggest that multi-way PDP can give a relatively reliable relation between input features and model predictions than one-way PDP. This further illustrates the important role of GTJC-based multi-way PDP in the analysis of individual factor-$PM_{2.5}$ concentration relationship. Therefore, it is recommended to use

multi-way PDP to explain the "black box" model. In the future, the dependence of PM$_{2.5}$ concentrations on interactions between meteorological factors is interesting and needs to be further analyzed. It will contribute to the improvement of climate prediction with better understanding of PM$_{2.5}$ concentrations–multifactor relationship.

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

Data Availability Statement: Publicly available datasets were analyzed in this study. The hourly PM$_{2.5}$ data are collected from the China National Environment Monitoring Center (CNEMC, http://www.cnemc.cn/, accessed on 1 April 2022). The Moderate Resolution Imaging Spectroradiometer (MODIS) data can be accessed at Distributed Active Archive Center (DAAC, https://lpdaac.usgs.gov/products/mcd19a2v006/, accessed on 1 April 2022). The nighttime light (NTL) products come from the Suomi NPP VIIRS (https://ncc.nesdis.noaa.gov/VIIRS/, accessed on 1 April 2022) The Shuttle Radar Topography Mission (SRTM) DEM product are available online (http://srtm.csi.cgiar.org/, accessed on 1 April 2022). The ERA5 were available from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (https://cds.climate.copernicus.eu/, accessed on 1 April 2022)

References

1. Pant, P.; Harrison, R.M. Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: A review. *Atmos. Environ.* **2013**, *77*, 78–97. [CrossRef]
2. Van Donkelaar, A.; Martin, R.V.; Brauer, M.; Hsu, N.C.; Kahn, R.A.; Levy, R.C.; Lyapustin, A.; Sayer, A.M.; Winker, D.M. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environ. Sci. Technol.* **2016**, *50*, 3762–3772.
3. Sun, L.; Wei, J.; Duan, D.; Guo, Y.; Yang, D.; Jia, C.; Mi, X. Impact of Land-Use and Land-Cover Change on urban air quality in representative cities of China. *J. Atmos. Sol.-Terr. Phys.* **2016**, *142*, 43–54.
4. Jandacka, D.; Durcanska, D. Seasonal Variation, Chemical Composition, and PMF-Derived Sources Identification of Traffic-Related PM$_1$, PM$_{2.5}$, and PM$_{2.5–10}$ in the Air Quality Management Region of Žilina, Slovakia. *Int. J. Environ. Res. Public Health* **2021**, *18*, 10191. [CrossRef] [PubMed]
5. Wang, S.; Kaur, M.; Li, T.; Pan, F. Effect of Different Pollution Parameters and Chemical Components of PM$_{2.5}$ on Health of Residents of Xinxiang City, China. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6821. [CrossRef]
6. Wang, Y.; Hu, J.; Huang, L.; Li, T.; Yue, X.; Xie, X.; Liao, H.; Chen, K.; Wang, M. Projecting future health burden associated with exposure to ambient PM$_{2.5}$ and ozone in China under different climate scenarios. *Environ. Int.* **2022**, *169*, 107542. [CrossRef]
7. Chowdhury, S.; Pozzer, A.; Haines, A.; Klingmüller, K.; Münzel, T.; Paasonen, P.; Sharma, A.; Venkataraman, C.; Lelieveld, J. Global health burden of ambient PM$_{2.5}$ and the contribution of anthropogenic black carbon and organic aerosols. *Environ. Int.* **2022**, *159*, 107020. [CrossRef]
8. He, J.; Gong, S.; Yu, Y.; Yu, L.; Wu, L.; Mao, H.; Song, C.; Zhao, S.; Liu, H.; Li, X.; et al. Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major Chinese cities. *Environ. Pollut.* **2017**, *223*, 484–496.
9. Chen, Z.; Xie, X.; Cai, J.; Chen, D.; Gao, B.; He, B.; Cheng, N.; Xu, B. Understanding meteorological influences on PM$_{2.5}$ concentrations across China: A temporal and spatial perspective. *Atmos. Chem. Phys.* **2018**, *18*, 5343–5358.
10. Jin, Q.; Fang, X.; Wen, B.; Shan, A. Spatio-temporal variations of PM2. 5 emission in China from 2005 to 2014. *Chemosphere* **2017**, *183*, 429–436.
11. Lim, C.H.; Ryu, J.; Choi, Y.; Jeon, S.W.; Lee, W.K. Understanding global PM2. 5 concentrations and their drivers in recent decades (1998–2016). *Environ. Int.* **2020**, *144*, 106011.
12. Ma, Z.; Dey, S.; Christopher, S.; Liu, R.; Bi, J.; Balyan, P.; Liu, Y. A review of statistical methods used for developing large-scale and long-term PM2.5 models from satellite data. *Remote Sens. Environ.* **2022**, *269*, 112827.
13. Yang, Q.; Yuan, Q.; Li, T. Ultrahigh-resolution PM2.5 estimation from top-of-atmosphere reflectance with machine learning: Theories, methods, and applications. *Environ. Pollut.* **2022**, *306*, 119347. [CrossRef]
14. Yang, N.; Shi, H.; Tang, H.; Yang, X. Geographical and temporal encoding for improving the estimation of PM2.5 concentrations in China using end-to-end gradient boosting. *Remote Sens. Environ.* **2022**, *269*, 112828.

15. Wei, J.; Li, Z.; Lyapustin, A.; Sun, L.; Peng, Y.; Xue, W.; Su, T.; Cribb, M. Reconstructing 1-km-resolution high-quality PM2.5 data records from 2000 to 2018 in China: Spatiotemporal variations and policy implications. *Remote Sens. Environ.* **2021**, *252*, 112136. [CrossRef]

16. Bai, K.; Li, K.; Wu, C.; Chang, N.B.; Guo, J. A homogenized daily in situ PM$_{2.5}$ concentration dataset from the national air quality monitoring network in China. *Earth Syst. Sci. Data* **2020**, *12*, 3067–3080. [CrossRef]

17. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]

18. Grange, S.K.; Carslaw, D.C. Using meteorological normalisation to detect interventions in air quality time series. *Sci. Total Environ.* **2019**, *653*, 578–588. [CrossRef] [PubMed]

19. Liu, Y.; Cao, G.; Zhao, N. Integrate machine learning and geostatistics for high-resolution mapping of ground-level PM2.5 concentrations. In *Spatiotemporal Analysis of Air Pollution and Its Application in Public Health*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 135–151. [CrossRef]

20. Analitis, A.; Barratt, B.; Green, D.; Beddows, A.; Samoli, E.; Schwartz, J.; Katsouyanni, K. Prediction of PM2.5 concentrations at the locations of monitoring sites measuring PM10 and NOx, using generalized additive models and machine learning methods: A case study in London. *Atmos. Environ.* **2020**, *240*, 117757. [CrossRef]

21. Ly, B.T.; Matsumi, Y.; Vu, T.V.; Sekiguchi, K.; Nguyen, T.T.; Pham, C.T.; Nghiem, T.D.; Ngo, I.H.; Kurotsuchi, Y.; Nguyen, T.H.; et al. The effects of meteorological conditions and long-range transport on PM$_{2.5}$ levels in Hanoi revealed from multi-site measurement using compact sensors and machine learning approach. *J. Aerosol Sci.* **2021**, *152*, 105716. [CrossRef]

22. Ding, A.J.; Fu, C.B.; Yang, X.Q.; Sun, J.N.; Zheng, L.F.; Xie, Y.N.; Herrmann, E.; Nie, W.; Petäjä, T.; Kerminen, V.M.; et al. Ozone and fine particle in the western Yangtze River Delta: An overview of 1 yr data at the SORPES station. *Atmos. Chem. Phys.* **2013**, *13*, 5813–5830. [CrossRef]

23. Yang, X.; Zhao, C.; Guo, J.; Wang, Y. Intensification of aerosol pollution associated with its feedback with surface solar radiation and winds in Beijing. *J. Geophys. Res. Atmos.* **2016**, *121*, 4093–4099.

24. Hu, X.; Waller, L.A.; Lyapustin, A.; Wang, Y.; Liu, Y. 10-Year spatial and temporal trends of PM2.5 concentrations in the southeastern US estimated using high-resolution satellite data. *Atmos. Chem. Phys.* **2014**, *14*, 6301–6314.

25. Lyapustin, A.; Wang, Y. *MCD19A2 MODIS/Terra+ Aqua Land Aerosol Optical Depth Daily L2G global 1km SIN Grid V006 [data set]*; NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA, 2018.

26. Mhawish, A.; Banerjee, T.; Sorek-Hamer, M.; Lyapustin, A.; Broday, D.M.; Chatfield, R. Comparison and evaluation of MODIS Multi-angle Implementation of Atmospheric Correction (MAIAC) aerosol product over South Asia. *Remote Sens. Environ.* **2019**, *224*, 12–28. [CrossRef]

27. Sulla-Menashe, D.; Friedl, M. *MCD12Q1 MODIS/Terra+ Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006*; NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA, 2019. [CrossRef]

28. Elvidge, C.D.; Baugh, K.E.; Zhizhin, M.; Hsu, F.C. Why VIIRS data are superior to DMSP for mapping nighttime lights. *Proc. Asia-Pac. Adv. Netw.* **2013**, *35*, 62. [CrossRef]

29. Liu, Y. New Directions: Satellite driven PM2.5 exposure models to support targeted particle pollution health effects research. *Atmos. Environ.* **2013**, *68*, 52–53.

30. Jarvis, A.; Reuter, H.I.; Nelson, A.; Guevara, E.; others. Hole-filled SRTM for the globe Version 4. available from the CGIAR-CSI SRTM 90m Database. 2008, Volume 15. Available online: http://srtm.csi.cgiar.org (accessed on 1 April 2022).

31. Didan, K. *Mod13a3 Modis/Terra Vegetation Indices Monthly L3 Global 1km Sin Grid V006*; NASA EOSDIS Land Processes DAAC: Sioux Falls, SD, USA, 2015; Volume 10. [CrossRef]

32. Chen, Z.; Chen, D.; Zhao, C.; Kwan, M.p.; Cai, J.; Zhuang, Y.; Zhao, B.; Wang, X.; Chen, B.; Yang, J.; et al. Influence of meteorological conditions on PM2. 5 concentrations across China: A review of methodology and mechanism. *Environ. Int.* **2020**, *139*, 105558.

33. Muñoz Sabater, J. *ERA5-Land Hourly Data from 1981 to Present*; Copernicus Climate Change Service (C3S) Climate Data Store (CDS): Bracknell, UK, 2019. [CrossRef]

34. Hersbach, H.; Bell, B.; Berrisford, P.; Biavati, G.; Horányi, A.; Muñoz Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Rozum, I.; et al. *ERA5 Hourly Data on Single Levels from 1979 to Present*; Copernicus Climate Change Service (C3S) Climate Data Store (CDS): Bracknell, UK, 2018; Volume 10. [CrossRef]

35. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

36. Wei, J.; Huang, W.; Li, Z.; Xue, W.; Peng, Y.; Sun, L.; Cribb, M. Estimating 1-Km-Resolution PM2.5 concentrations across China using the space-time random forest approach. *Remote Sens. Environ.* **2019**, *231*, 111221.

37. Zhang, T.; He, W.; Zheng, H.; Cui, Y.; Song, H.; Fu, S. Satellite-based ground PM2.5 estimation using a gradient boosting decision tree. *Chemosphere* **2021**, *268*, 128801. [CrossRef]

38. He, W.; Meng, H.; Han, J.; Zhou, G.; Zheng, H.; Zhang, S. Spatiotemporal PM2.5 estimations in China from 2015 to 2020 using an improved gradient boosting decision tree. *Chemosphere* **2022**, *296*, 134003. [CrossRef]

39. Shin, M.; Kang, Y.; Park, S.; Im, J.; Yoo, C.; Quackenbush, L.J. Estimating ground-level particulate matter concentrations using satellite-based data: A review. *GIScience Remote Sens.* **2020**, *57*, 174–189.

40. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

41. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: London, UK, 2017.

42. Gui, K.; Che, H.; Zeng, Z.; Wang, Y.; Zhai, S.; Wang, Z.; Luo, M.; Zhang, L.; Liao, T.; Zhao, H.; et al. Construction of a virtual PM2. 5 observation network in China based on high-density surface meteorological observations using the Extreme Gradient Boosting model. *Environ. Int.* **2020**, *141*, 105801.

43. Fan, W.; Qin, K.; Cui, Y.; Li, D.; Bilal, M. Estimation of hourly ground-level PM2. 5 concentration based on himawari-8 apparent reflectance. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 76–85.

44. Huang, C.; Hu, J.; Xue, T.; Xu, H.; Wang, M. High-resolution spatiotemporal modeling for ambient PM2.5 exposure assessment in china from 2013 to 2019. *Environ. Sci. Technol.* **2021**, *55*, 2152–2162.

45. Molnar, C. *Interpretable Machine Learning*; Lulu Press: Morrisville, NC, USA, 2020.

46. de Leeuw, G.; van der A, R.; Bai, J.; Xue, Y.; Varotsos, C.; Li, Z.; Fan, C.; Chen, X.; Christodoulakis, I.; Ding, J.; et al. Air Quality over China. *Remote Sens.* **2021**, *13*, 3542.

47. Varotsos, C.A.; Ondov, J.M.; Cracknell, A.P.; Efstathiou, M.N.; Assimakopoulos, M.N. Long-range persistence in global Aerosol Index dynamics. *Int. J. Remote Sens.* **2006**, *27*, 3593–3603.

48. Varotsos, C.; Ondov, J.; Tzanis, C.; Öztürk, F.; Nelson, M.; Ke, H.; Christodoulakis, J. An observational study of the atmospheric ultra-fine particle dynamics. *Atmos. Environ.* **2012**, *59*, 312–319.

49. Wang, P.; Cao, J.j.; Shen, Z.x.; Han, Y.m.; Lee, S.c.; Huang, Y.; Zhu, C.s.; Wang, Q.y.; Xu, H.m.; Huang, R.j. Spatial and seasonal variations of PM2. 5 mass and species during 2010 in Xi'an, China. *Sci. Total Environ.* **2015**, *508*, 477–487.

50. Chen, L.; Zhu, J.; Liao, H.; Yang, Y.; Yue, X. Meteorological influences on PM2. 5 and O3 trends and associated health burden since China's clean air actions. *Sci. Total Environ.* **2020**, *744*, 140837. [CrossRef] [PubMed]

51. Deng, L.; Bi, C.; Jia, J.; Zeng, Y.; Chen, Z. Effects of heating activities in winter on characteristics of PM2.5-Bound Pb, Cd and lead isotopes in cities of China. *J. Clean. Prod.* **2020**, *265*, 121826.

52. Fan, M.; He, G.; Zhou, M. The winter choke: Coal-fired heating, air pollution, and mortality in China. *J. Health Econ.* **2020**, *71*, 102316.

53. Wang, J.; Wang, S.; Li, S. Examining the spatially varying effects of factors on PM2.5 concentrations in Chinese cities using geographically weighted regression modeling. *Environ. Pollut.* **2019**, *248*, 792–803. [CrossRef]

54. Dong, Z.; Li, Z.; Yu, X.; Cribb, M.; Li, X.; Dai, J. Opposite long-term trends in aerosols between low and high altitudes: A testimony to the aerosol–PBL feedback. *Atmos. Chem. Phys.* **2017**, *17*, 7997–8009.

55. Sheng, Z.; Che, H.; Chen, Q.; Liu, D.; Wang, Z.; Zhao, H.; Gui, K.; Zheng, Y.; Sun, T.; Li, X.; et al. Aerosol vertical distribution and optical properties of different pollution events in Beijing in autumn 2017. *Atmos. Res.* **2019**, *215*, 193–207.

56. Wang, F.; Li, Z.; Ren, X.; Jiang, Q.; He, H.; Dickerson, R.R.; Dong, X.; Lv, F. Vertical distributions of aerosol optical properties during the spring 2016 ARIAs airborne campaign in the North China Plain. *Atmos. Chem. Phys.* **2018**, *18*, 8995–9010.

57. Zhang, X.; Wang, H.; Che, H.Z.; Tan, S.C.; Shi, G.Y.; Yao, X.P. The impact of aerosol on MODIS cloud detection and property retrieval in seriously polluted East China. *Sci. Total Environ.* **2020**, *711*, 134634.

58. Sugihara, G.; May, R.; Ye, H.; Hsieh, C.h.; Deyle, E.; Fogarty, M.; Munch, S. Detecting causality in complex ecosystems. *Science* **2012**, *338*, 496–500.

59. Zhan, Y.; Luo, Y.; Deng, X.; Chen, H.; Grieneisen, M.L.; Shen, X.; Zhu, L.; Zhang, M. Spatiotemporal prediction of continuous daily PM2.5 concentrations across China using a spatially explicit machine learning algorithm. *Atmos. Environ.* **2017**, *155*, 129–139. [CrossRef]

60. Wang, M.; Zhang, Z.; Yuan, Q.; Li, X.; Han, S.; Lam, Y.; Cui, L.; Huang, Y.; Cao, J.; Lee, S.c. Slower than expected reduction in annual PM2.5 in Xi'an revealed by machine learning-based meteorological normalization. *Sci. Total Environ.* **2022**, *841*, 156740. [CrossRef] [PubMed]

61. Xiao, Q.; Wang, Y.; Chang, H.H.; Meng, X.; Geng, G.; Lyapustin, A.; Liu, Y. Full-coverage high-resolution daily PM2.5 estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ.* **2017**, *199*, 437–446. [CrossRef]