

Article

# Modelling Seasonal GWR of Daily PM<sub>2.5</sub> with Proper Auxiliary Variables for the Yangtze River Delta

Man Jiang <sup>1</sup>, Weiwei Sun <sup>1,2,\*</sup>, Gang Yang <sup>1</sup> and Dianfa Zhang <sup>1</sup>

<sup>1</sup> Department of Geography and Spatial Information Techniques, Ningbo University, 818 Fenghua Road, Ningbo 315211, China; jiangman126@126.com (M.J.); yanggang@nbu.edu.cn (G.Y.); zhangdianfa@nbu.edu.cn (D.Z.)

<sup>2</sup> State Key Lab of Information Engineering on Survey, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

\* Correspondence: sunweiwei@nbu.edu.cn; Tel.: +86-182-5879-6120

Academic Editors: Yang Liu, Jun Wang, Omar Torres, Richard Müller and Prasad S. Thenkabail

Received: 9 December 2016; Accepted: 1 April 2017; Published: 5 April 2017

**Abstract:** Over the past decades, regional haze episodes have frequently occurred in eastern China, especially in the Yangtze River Delta (YRD). Satellite derived Aerosol Optical Depth (AOD) has been used to retrieve the spatial coverage of PM<sub>2.5</sub> concentrations. To improve the retrieval accuracy of the daily AOD-PM<sub>2.5</sub> model, various auxiliary variables like meteorological or geographical factors have been adopted into the Geographically Weighted Regression (GWR) model. However, these variables are always arbitrarily selected without deep consideration of their potentially varying temporal or spatial contributions in the model performance. In this manuscript, we put forward an automatic procedure to select proper auxiliary variables from meteorological and geographical factors and obtain their optimal combinations to construct four seasonal GWR models. We employ two different schemes to comprehensively test the performance of our proposed GWR models: (1) comparison with other regular GWR models by varying the number of auxiliary variables; and (2) comparison with observed ground-level PM<sub>2.5</sub> concentrations. The result shows that our GWR models of “AOD + 3” with three common meteorological variables generally perform better than all the other GWR models involved. Our models also show powerful prediction capabilities in PM<sub>2.5</sub> concentrations with only slight overfitting. The determination coefficients R<sup>2</sup> of our seasonal models are 0.8259 in spring, 0.7818 in summer, 0.8407 in autumn, and 0.7689 in winter. Also, the seasonal models in summer and autumn behave better than those in spring and winter. The comparison between seasonal and yearly models further validates the specific seasonal pattern of auxiliary variables of the GWR model in the YRD. We also stress the importance of key variables and propose a selection process in the AOD-PM<sub>2.5</sub> model. Our work validates the significance of proper auxiliary variables in modelling the AOD-PM<sub>2.5</sub> relationships and provides a good alternative in retrieving daily PM<sub>2.5</sub> concentrations from remote sensing images in the YRD.

**Keywords:** seasonal GWR models; auxiliary variable selection; geographically weighted model; MODIS AOD; PM<sub>2.5</sub> concentrations; Yangtze River Delta

## 1. Introduction

Widespread air pollution has become a severe problem in China, with increasing population and pollution emissions. The Yangtze River Delta (YRD), as one of the most developed regions in eastern China, has been suffering deterioration of air quality and even more frequent haze episodes, severely threatening both life and health of its people. Particulate matter with an aerodynamic diameter less than 2.5 μm (PM<sub>2.5</sub>) is one of most harmful components of pollution haze and it has severely toxic effects on climate, environment and human health [1,2]. Numerous epidemiological studies have

validated the direct relation between high PM<sub>2.5</sub> concentrations and rising human health problems like asthma, tumors, and lung cancer [3–7]. Therefore, PM<sub>2.5</sub> concentration monitoring is a significant and pressing issue for both assessing human health exposure and making effective air pollution control measures in the YRD region.

Ground-based monitoring networks could provide accurate and real-time PM<sub>2.5</sub> concentrations. However, the discrete monitoring sites only measure PM<sub>2.5</sub> concentrations around a certain distance of the sites and cannot provide a spatial coverage of PM<sub>2.5</sub> concentrations. Moreover, major monitoring stations are scattered in urban environments and particularly in the metropolis, leaving most rural areas uncovered. Even though the number of monitoring stations in China has been clearly increasing in recent years, the sites are still insufficient to fill all space gaps of the YRD region [8,9].

In contrast, satellite remote sensing has distinct advantages in long-term monitoring and large-scale spatial coverage. Many satellite sensors like MODIS, MISR, and SeaWiFS collect the aerosol information in the atmosphere including aerosol scattering and absorption. They are widely used in estimating and monitoring PM<sub>2.5</sub> concentrations with aerosol optical depth (AOD). AOD measures the light extinction by aerosol scattering and absorption and it reflects the particle number and property of PM<sub>2.5</sub> in the total atmosphere. The satellite sensors then estimate the spatial coverage of daily PM<sub>2.5</sub> concentrations via the retrieval relations between ground-level PM<sub>2.5</sub> concentrations and satellite-based AOD [10–18].

The retrieval models of PM<sub>2.5</sub> concentrations coverage from satellite-based AOD can be divided into three main types: the scaling factor models [19,20], the physical analysis models [21,22], and the empirical statistical models [23]. Scaling factor models mainly originate from the chemical transport model (CTM), and they determine the scale factor between satellite-based AOD and ground-level PM<sub>2.5</sub> concentrations to estimate large-scale spatial distributions of satellite PM<sub>2.5</sub> concentrations. The models were designed for atmospheric regions without ground PM<sub>2.5</sub> monitoring data and the retrieval accuracy of PM<sub>2.5</sub> concentrations is relatively low [23]. Moreover, complicated parameters are mandatorily requiring to initialize and optimize the CTM. Different from scale factor models, physical analysis models analyze the AOD-PM<sub>2.5</sub> relationships and incorporate accountable physical parameters to construct quantitative functions of satellite PM<sub>2.5</sub> concentrations [24]. Unfortunately, it is a big challenge to collect these physical parameters in realistic applications. Furthermore, the physical mechanisms in reality are far more complicated than these ever-proposed formulas. Empirical statistical models bring about more accurate distribution retrievals of PM<sub>2.5</sub> concentrations when compared with the physical analysis models or scaling factor models [21]. Empirical statistical models [25] construct statistical regression functions between satellite-based AOD and in situ PM<sub>2.5</sub> concentration measurements, and they can be grouped into two classes including early-stage statistical models and advanced statistical models.

Early-stage statistical models are mainly referred to as simple or multiple linear regression models, whereas advanced statistical models develop features in delineating spatial and temporal variations in the relationships between AOD and PM<sub>2.5</sub> concentrations. Typical examples of advanced statistical models are the general additive line model (GAM) [26], the geographical weighted regression model (GWR) [27], the linear mixed effects (LME) [28], the geographically and temporarily weighted regression model (GTWR) [29], the two step models [10,14], and the three step models [30,31]. Amongst all of them, the GWR has a simple mathematical theory, low computational complexity, and relatively stable performance in considering unstable relationships between ground-level PM<sub>2.5</sub> concentrations and remote sensed AOD [8,9,15,16,26,32]. Moreover, the GWR shows good compatibility and it is always combined with other schemes to construct complicated statistical models such as the two steps models [14] and three step models [31]. The behaviors of GWR correlate closely with the above formulated models and therefore we focus our study on the GWR model and aim to promote its PM<sub>2.5</sub> retrieval performance in realistic applications of the YRD region.

The GWR model observes that the relation between PM<sub>2.5</sub> concentrations and AOD varies across different spatial locations, and additional factors in geography or meteorology are usually

incorporated into the model to help explain the generation and dilution of PM<sub>2.5</sub> concentrations in the atmosphere [10]. Meteorological factors are mainly derived from physical models, and typical variables adopted in GWR include boundary layer height, relative humidity, temperature, and wind speed etc. Geographical factors, mainly referring to the land use the regression model, to explain spatial variations of air pollution in outdoor environments. Geographical variables mainly include demography, land use type, elevation, and vegetation coverage ratio etc. Table 1 lists representatives of auxiliary variables from meteorological or geographical factors in the GWR model for retrieving daily PM<sub>2.5</sub> concentrations. These auxiliary variables have been proven to enhance the stability of GWR in PM<sub>2.5</sub> concentration retrieval. However, unfortunately, two big problems still exist in how to properly select meteorological or geographical factors, and that severely hinders the performance improvement of the daily GWR model in realistic applications of the YRD region.

**Table 1.** Representatives of auxiliary variables in geographical weighted regression model (GWR) model for PM<sub>2.5</sub> concentrations.

Study Area	Meteorological Factors	Geographical Factors	References
China	relatively humidity, air temperature, wind speed, horizontal visibility	—	[16]
Global	GEOS-Chem chemical transport model (CTM)	urban land cover, elevation	[32]
China	boundary layer height, temperature, wind speed, relative humidity, air pressure	population density, monthly mean normalized difference vegetation index (NDVI)	[9]
Pearl River Delta region	temperature, wind speed, relative humidity	—	[15]
North American Regional	boundary layer height, relative humidity, air temperature, wind speed	percentage of forest cover	[27]

(1) The subjective scheme in selecting meteorological or geographical factors might result in unrepresentative or redundancy among different variables and that would reduce the retrieval performance of the daily GWR model. Different meteorological or geographical factors do have divergent contributions in the GWR model, but the contributions from these variables have never been carefully analyzed. Meanwhile, strong intra-correlations might exist among different meteorological or geographical factors. For example, air temperature has a negative correlation with air pressure, and the elevation correlates closely with demography on the same sites. However, current literatures have never carefully investigated the procedure in selecting proper meteorological or geographical variables for daily GWR modelling. Arbitrary selection of these factors would adversely degrade the accuracy of GWR in realistic applications.

(2) The subjective selection scheme has neglected metabolic contributions of these meteorological or geographical factors to the retrieval performance of the daily GWR model across four different seasons. Working mechanism and contributions of these meteorological or geographical variables vary across different seasons, causing the wide range of daily model performance [27]. For example, in eastern China, the wind dilutes air pollution in summer whereas it might show opposite influences in spatial distributions of PM<sub>2.5</sub> concentrations in winter. The reason is that the winter monsoon brings the articles from the north of China. Accordingly, it is of great necessity to consider the particularity of variable contributions of different factors in different seasons in order to guarantee good performance of the GWR model.

In our previous work of literature [33], we tested different potential influences from meteorological factors and geographical factors in retrieving PM<sub>2.5</sub> concentrations with the regular GWR model. In this manuscript, we design an automatic procedure to select proper variables from meteorological or geographical factors in the YRD region and construct specific GWR models for retrieving daily PM<sub>2.5</sub> concentrations in four different seasons. We validate our seasonal GWR models by comparing with regular GWR models with varying auxiliary variables and by comparing the predicted PM<sub>2.5</sub> concentrations with the observed. As far as we know, few relevant works have carefully explored the situation in current literatures especially for the YRD region. The recently proposed timely

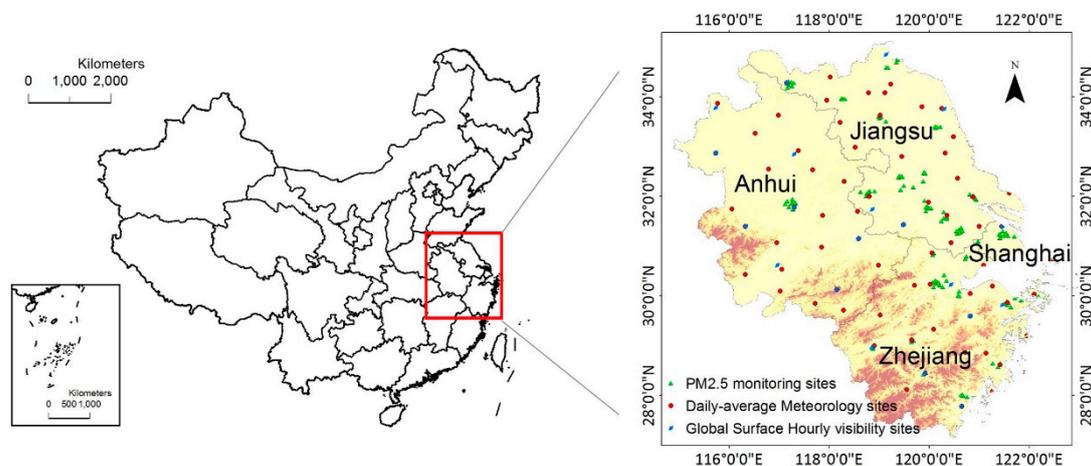
structure adaptive modeling (TSAM) tried to construct a daily AOD-PM<sub>2.5</sub> model by selecting daily varied auxiliary variables [34]. However, the prediction procedure of TSAM is too complicated to implement in the YRD region. Accordingly, the objects of this manuscript are to: (1) make clear different contributions of the main meteorological or geographical factors and their combinations to the GWR model of the YRD region; (2) propose an automatic procedure to find proper auxiliary variables for modelling GWR in different seasons; and (3) provide detailed equations of seasonal GWR models to benefit the retrieval of daily PM<sub>2.5</sub> concentrations in the YRD region.

## 2. Materials and Methods

### 2.1. Data

#### 2.1.1. Ground-Level PM<sub>2.5</sub> Concentration Data

Our study region YRD includes Zhejiang, Jiangsu, Anhui provinces, and Shanghai city. We selected the year of 2013 as our study period because of the increasing public attention to haze episodes from 2013 [35] and the data accessibility of PM<sub>2.5</sub> monitoring sites in 2013. Figure 1 illustrates the 123 PM<sub>2.5</sub> monitoring sites of the YRD region in 2013. Ground-level hourly PM<sub>2.5</sub> concentration data was downloaded from China air quality real-time release system of the Chinese Ministry of Environmental Protection (available at <http://106.37.208.233:20035>). The PM<sub>2.5</sub> concentrations were measured by Tapered Element Oscillating Microbalances (TEOM) or beta attenuation method (BAM or  $\beta$ -gauge). The data has an uncertainty less than 0.75%, with its accuracy reaching up to  $\pm 1.5 \mu\text{g}/\text{m}^3$  for the hourly average, and hence it is accurate enough as ground truth for PM<sub>2.5</sub> concentration measures. From the consideration of simplicity and convenience, the PM<sub>2.5</sub> concentration data at Beijing time 11:00 AM from 1 January to 31 December 2013 was collected to match with the passing time of the MODIS Terra satellite (i.e., approximately 10:30 a.m. at local time).



**Figure 1.** The ground PM<sub>2.5</sub> monitoring sites and meteorology stations in the Yangtze River Delta (YRD) region.

#### 2.1.2. MODIS AOD

The MODIS sensors on the Terra and Aqua satellites provide global information of the Earth-atmosphere system in 36 spectral bands from visible to thermal infrared spectrum range (0.4–14  $\mu\text{m}$ ) with a swath width of 2330 km in 1–2 days. Compared with many other satellite-derived AOD products, the MODIS AOD has the greatest reputation because of its high temporal resolutions, relatively high spatial resolutions, good accuracy, and easy accessibility [36]. The latest MODIS AOD production version Collection 6 is constructed from MODIS imagery via both enhanced DB and DT algorithm and the AOD product is adaptable for both dark and bright surfaces. We in this study

uses the Terra MODIS C6 DT 10-km AOD product and the data was download from Level-2 and Atmosphere Archive & Distribution System of NASA (available at <http://ladsweb.nascom.nasa.gov/>).

### 2.1.3. Meteorological Datasets

Referring to previous studies, we manually choose six factors as preliminarily auxiliary meteorological variables, including temperature (Temp), wind speed (WS), air pressure (Apre), vapor pressure (Vpre), relative humidity (RH), and surface horizontal visibility (VSB). The particle concentrations of PM<sub>2.5</sub> largely depend on the meteorological conditions. High surface temperature or high air pressure accelerates the atmospheric vertical motion to transport ground pollutants into higher places. Wind speed is an effective index of quantifying surface motions of air flow and affects the horizontal transport of ground pollutants. Relative humidity makes a correction of aerosol humidity in the atmosphere in order to better match with ground dry PM<sub>2.5</sub> concentrations. High relative humidity largely enhances the size and light extinction of particles, which comprise the sulfate, nitrate, and ammonium from coal and biomass burning, industrial, and vehicular sources. We also take vapor pressure as a meteorological variable because we regard it is a comprehensive variable and it correlates closely with the generation or aggregation of PM<sub>2.5</sub>. The daily averaged data of above five variables were acquired from the China Daily Surface of Climatic Dataset in Chinese Meteorological Administration (available at <http://data.cma.cn/>). The YRD region has a total of 72 ground-level monitoring sites in Figure 1.

Although BLH has been a common used variable for AOD vertical correction in many previous studies, recent real experiments proved that BLH made unclear contributions in the GWR model for retrieving PM<sub>2.5</sub> concentrations in the YRD [28]. The biomass burning in the YRD greatly influences the aloft aerosol above the BLH, and the vertical correction of AOD might be underestimated by the BLH [37]. In contrast, the visibility directly reflects the relationship between AOD and ground-level extinction coefficient [16] and shows great significance in GWR modelling of PM<sub>2.5</sub> concentrations [8]. Therefore, we implemented VSB as a preliminary meteorological variable rather than BLH. The visibility data was acquired from 23 ground-level monitoring sites in the YRD and the dataset at 11:00 was collected from the National Climatic Data Center (NCDC) Global Surface Hourly database (available at <http://gis.ncdc.noaa.gov/map/viewer/#app=clim&cfg=cdo&theme=hourly&layers=1&node=gis>).

### 2.1.4. Geographical Datasets

We also manually selected three widely used geographical factors as preliminarily auxiliary geographical variables, geomorphy feature (GEOM), elevation, and vegetation coverage. We chose the geomorphy feature because it impacts the spread of air pollutants. The geomorphy feature dataset was obtained from the Institute of Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences (available at <http://www.resdc.cn/data.aspx?DATAID=124>), with spatial resolutions of 10 km equal to the MODIS AOD product. The original geomorphy data was manually recategorized from 26 classes into four new classes in YRD, including plain, platform, hill, and mountain, in order to differentiate their influences in AOD-PM<sub>2.5</sub> relationships. The elevation is supposed to have negative effects on PM<sub>2.5</sub> distributions, because of the gravity sedimentation of air particles. The 90 m Digital Elevation Model (DEM) of SRTMDem3 dataset was acquired from the Geospatial Data Cloud (available at <http://www.gscloud.cn/>). High vegetation coverage reduces the entry of aerosols into the atmosphere and absorbs particles in the atmosphere [38]. The 16-day synthesized normalized difference vegetation index (NDVI) product of MODIS (MOD13A2) was achieved from NASA (available at <http://ladsweb.nascom.nasa.gov/>) to represent the vegetation coverage with spatial resolutions of 1000 m. The time of MOD13\_A2 product was carefully chosen to coincide with those of other daily datasets, including MODIS AOD, PM<sub>2.5</sub> concentrations, and the meteorological dataset prior to the day.

### 2.1.5. Data Pre-Processing and Integration

All the above datasets (Ground-level  $PM_{2.5}$  concentrations, MODIS AOD, meteorological data, and geographical data) were transformed into the WGS84 geographic coordinate system. Meteorological data and ground-level  $PM_{2.5}$  concentrations were acquired at different monitoring stations, and the average distance of two kinds of station was manually measured at 0.144 degrees. We argue that the meteorological condition varies insignificantly within the average distance and meteorological stations are evenly distributed in the study region. So it is reasonable to spatially join each  $PM_{2.5}$  monitoring site with its nearest meteorological station. Accordingly, the ground-level  $PM_{2.5}$  concentration measures and its meteorological data were then registered into the same monitoring site. Meanwhile, the YRD region was digitized into grid cells with a fixed grid size of 0.1 degrees. Using overlay analysis, the averages of MODIS AOD and geographical data (DEM, MODIS NDVI and geomorphy data) within each grid cell were assigned as corresponding values of its grid cell.

The MODIS AOD product has many missing pixels mainly due to cloud coverage, high surface reflectance above bright and urban areas, and model retrieval errors. That greatly reduces the usability of the AOD product in matching it with ground-level  $PM_{2.5}$  concentration measures and also lowers the number of validated records in the GWR modelling. For the daily  $PM_{2.5}$  concentrations retrieval, a proper number of daily validated records of ground-level  $PM_{2.5}$  concentrations and AOD is a key point to ensure the robustness and accuracy of GWR modelling. A too small number of daily records could not reflect the realistic spatial coverage of  $PM_{2.5}$  concentrations on the same day. According to practical experience from our preliminary trials, the daily threshold of 20 records was manually chosen to guarantee a sufficient number of validated daily records of AOD and ground-level  $PM_{2.5}$  concentrations; the final records in our seasonal GWR modelling are 3482 scattering in 66 days of 2013.

## 2.2. Method

### 2.2.1. The Regular GWR Model

The GWR model is a spatial regression model that generates spatially continuous coefficients of all variables across the study area. It mainly contributes in analyzing the unstationary status of the spatially varied relationship between independent variables and dependent variables [39]. The GWR model assumes that the AOD- $PM_{2.5}$  relationship varies greatly with spatial locations in the study area, and it has been adapted to describe the unstable relations between  $PM_{2.5}$  concentrations and AOD, as well as other geographical or meteorological factors. In this study, considering the nine auxiliary variables in Table 2 to constitute the preliminary variable set  $AV = \{DNVI, Geom, Elev, Temp, RH, WS, Apre, Vpre, VSB\}$ , the regular GWR model in retrieving daily  $PM_{2.5}$  concentrations can be formulated as

$$PM_{2.5(i,j)} = \beta_{0(i,j)} + \beta_{AOD(i,j)}AOD_{(i,j)} + \sum_{k=1}^c \beta_{k(i,j)}AV_{SUB(i,j)}^k \quad (1)$$

where  $AOD_{(i,j)}$  and  $PM_{2.5(i,j)}$  represent main variables of the daily GWR model at the position  $i$  on day  $j$ , with the coefficient of AOD as  $\beta_{AOD(i,j)}$ ;  $\beta_{0(i,j)}$  is a constant coefficient denoting the location-specific intercept at the position  $i$  on day  $j$ ;  $AV_{SUB}^k$  denotes the  $k$ -th element of a subset selected from the set of auxiliary variables  $AV$ , with the subset size  $c$  no less than 9, and  $\beta_{k(i,j)}$  is the location-specific slope or coefficient of its corresponding auxiliary variable  $AV_{SUB}^k$ .

**Table 2.** Main parameters of all involved variables in GWR modelling.

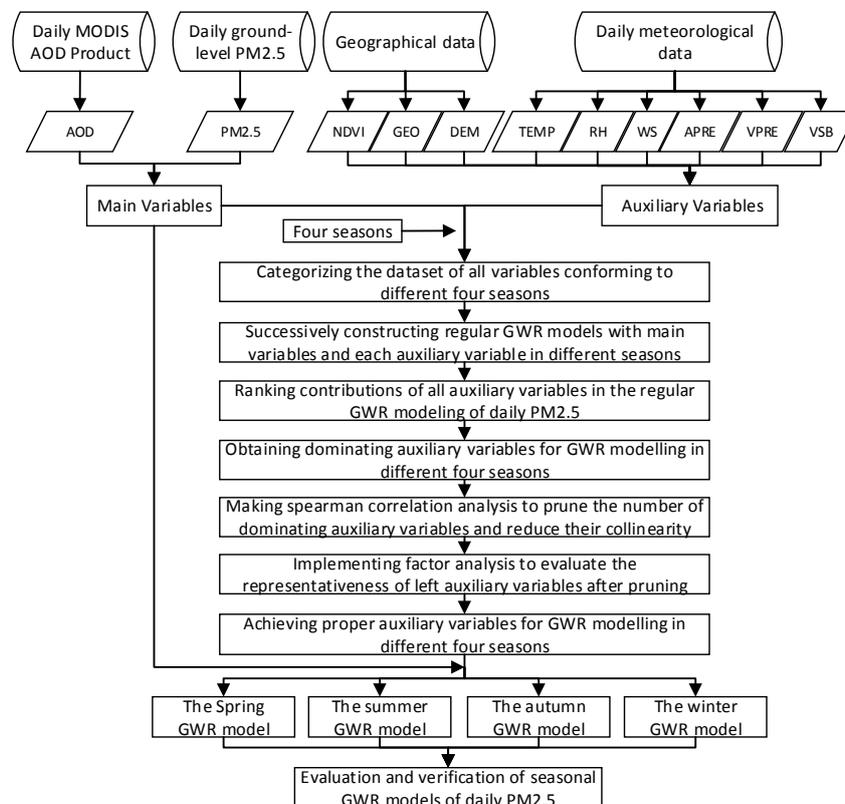
	Data	Variables (Abbreviation)	Unit	Time Frequency	Spatial Parameters
Main Variables	PM <sub>2.5</sub> concentration	PM <sub>2.5</sub>	µg/m <sup>3</sup>	Hourly	121 stations
	MODIS AOD	AOD	—	Daily	10 km
Geographical data		NDVI	—	16 days	1 km
		Geomorphology (Geom)	—	—	10 km
		DEM(Elev)	m	—	90 m
Preliminary Auxiliary Variables	Meteorological data	Temperature (Temp)	°C	Daily	72 stations
		Relative humidity (RH)	%		
		Wind speed (WS)	m/s		
		Air pressure (Apre)	Pa		
		Vapor pressure (Vpre)	Pa		
surface horizontal visibility (VSB)	km	Hourly	23 stations		

### 2.2.2. Seasonal GWR Modelling with Proper Auxiliary Variables

Selecting proper auxiliary variables is of great significance to guarantee the performance of the GWR model as well as to maximize the contribution of each selected auxiliary variable. Previous works validated the seasonal variability of different auxiliary variables in affecting the AOD-PM<sub>2.5</sub> relations in the GWR model. Therefore, we would like to propose an automatic procedure to select the proper auxiliary variables and construct specific GWR models for retrieving daily PM<sub>2.5</sub> concentrations in the four different seasons. With main variables of daily PM<sub>2.5</sub> and daily AOD product, with preliminary auxiliary variables of NDVI, Geom, Elev, Temp, RH, WS, Apre, Vpre, and VSB, the main procedure of constructing seasonal GWR models includes the following steps shown in Figure 2.

1. The datasets of main variables and preliminary auxiliary variables are categorized into four different seasons, spring, summer, autumn, and winter.
2. Different regular GWR models are constructed with main variables and auxiliary variables in different seasons. For each model in each season, we take AOD and PM<sub>2.5</sub> as main variables and separately add each element of nine single auxiliary variables one at a time into the GWR model. The performance of obtained regular GWR models is quantified via the Determination Coefficient ( $R^2$ ). By comparing with the simple seasonal GWR model without auxiliary variables, we rank the contributions of each auxiliary variable in the regular GWR modelling of daily PM<sub>2.5</sub> in descending order. Dominating auxiliary variables for GWR modelling in different seasons are then obtained.
3. Spearman correlation coefficient analysis is implemented into each pair of dominating auxiliary variables in different seasons. The operation is to reduce the collinearity and redundancy among dominating auxiliary variables. The spearman correlation coefficient is a nonparametric rank correlation coefficient, and it is a distribution-free version of the classical Pearson's product-moment correlation coefficient [40]. A higher coefficient means stronger relationships among different auxiliary variables and the coefficient at 0.3 is regarded as the threshold of weak correlations in our study. Once two dominating auxiliary variables have the spearman correlation coefficient over 0.3, and only one of them is chosen for further GWR modelling. The pruned auxiliary variables are obtained after the Spearman correlation coefficient analysis.
4. Factor analysis is carried out to verify the representativeness of pruned auxiliary variables. The idea of factor analysis is to group the variables having high correlations or close connections into the same class, where each class represents a basic structure called the common factor. The main common factors are able to reflect the major information of the original variables. In this study, the average of four season accumulated variance of the first four common factors is 70.97%. Moreover, the factor rotation in factor analysis provides actual physical meaning to explain working mechanisms of each pruned auxiliary variables. In the manuscript, we do not use uniform seasonal load matrix to construct new daily common variables and replace original variables because of the big probability of exaggerated errors.

5. The proper auxiliary variables are achieved for four different seasonal GWR models. The seasonal GWR models for daily  $PM_{2.5}$  are finally obtained in the YRD region.



**Figure 2.** The procedure of constructing seasonal Geographically Weighted Regression (GWR) models with proper auxiliary variables.

### 2.2.3. Model Evaluation and Verification

In the manuscript, we employ two different schemes to comprehensively testify the performance of our seasonal GWR models: (1) comparison with other regular GWR models by varying the number of auxiliary variables; and (2) comparison with observed ground-level  $PM_{2.5}$  concentrations.

We change the number of auxiliary variables to construct different GWR models and compare their performance with our obtained seasonal GWR models in the same season. To assess the performance of model fitting and evaluation, four popular measures are adopted, including  $R^2$ , corrected Akaike Information Criterion (AICc), Root Mean Squared Prediction Error (RMSE), and Mean Absolute Percentage Error (MAPE).  $R^2$  is a common indicator for model fitting, and AICc is used to make comparisons between GWR models with different auxiliary variables. RMSE and MAPE describe the residuals between predicted  $PM_{2.5}$  concentrations and the observed. MAPE weights the residual in terms of measured  $PM_{2.5}$  concentrations, and smaller MAPE indicates higher  $PM_{2.5}$  concentrations in the same residual condition. Compared with MAPE, RMSE is more sensitive to higher residual because it places more punishment on a higher residual than a lower one.

On the other hand, we compare the predicted daily  $PM_{2.5}$  concentrations with the observed  $PM_{2.5}$  measures in the testing sites to further verify the performance of our seasonal GWR models. From the above mentioned, some grids have no daily AOD from MODIS remote sensing to correspond with their  $PM_{2.5}$  measures on the same ground-level monitoring sites because of cloud coverage and other reasons. The  $PM_{2.5}$  concentrations in these grids are chosen to constitute the testing samples and their corresponding AOD measures are filled with the average AOD values of their nearest neighbors using

buffer analysis via 0.3 degrees. The reason for filling absent AOD with buffer analysis is that many experiments from previous works validated that the AOD would not diverge greatly within a small distance. Moreover, with the proper auxiliary variable combination, the 10-fold cross validation is implemented to testify the predication performance of our seasonal GWR models.

### 3. Results

#### 3.1. Descriptive Statistic of Datasets

Table 3 lists the statistic information of two major variables PM<sub>2.5</sub> and AOD for model fitting and evaluation in the experiments. The average PM<sub>2.5</sub> is highest in winter (97.76 µg/m<sup>3</sup>), then followed by spring (68.02 µg/m<sup>3</sup>), autumn (57.95 µg/m<sup>3</sup>), and summer (39.50 µg/m<sup>3</sup>). The averages of AOD reach the highest value in spring (0.82), then followed by summer, winter, and autumn. We argue that the inconsistent seasonal patterns of PM<sub>2.5</sub> and AOD are caused by different seasonal impacts of meteorological or geographical factors, rendering that the PM<sub>2.5</sub> concentrations and AOD are scattered to different extents even in opposite directions. Compared with model fitting in four seasons, the averages of PM<sub>2.5</sub> concentrations implemented in model evaluation are slightly higher. The reason for this is because some grids with higher PM<sub>2.5</sub> concentrations but without AOD caused by cloud coverage are grouped as testing samples for model evaluation.

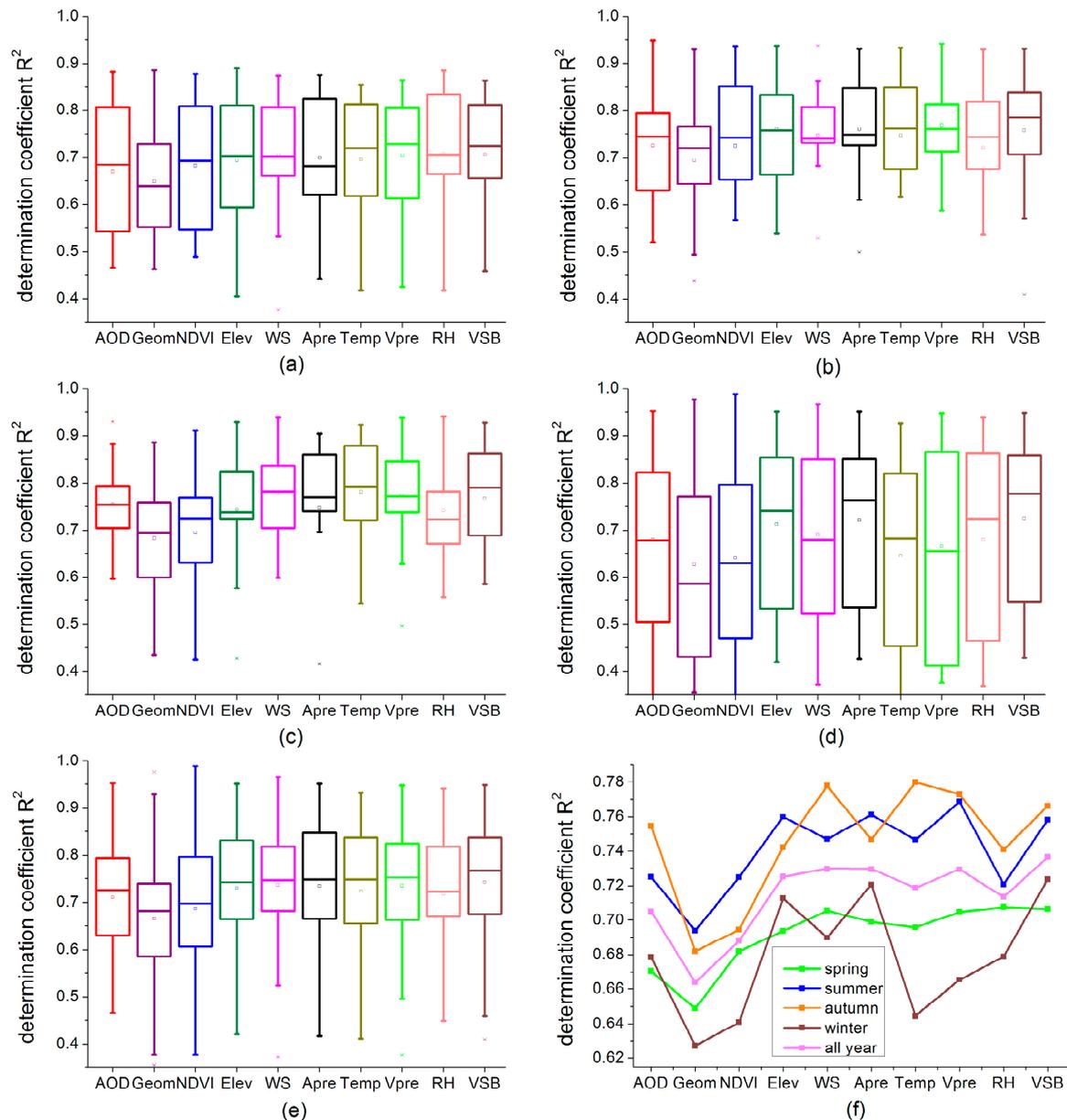
**Table 3.** Description statistics of PM<sub>2.5</sub> and Aerosol Optical Depth (AOD) for GWR model fitting and evaluation.

Whole Year	Variable	Model Fitting (N = 3482, day = 66)				Model Evaluation (N = 715, day = 66)			
		Mean	Min	Max	SD	Mean	Min	Max	SD
	PM <sub>2.5</sub> (µg/m <sup>3</sup> )	61.75	3	400	40.43	67	21	267	32.
	AOD (Unit less)	0.69	0.03	3.51	0.41	0.62	0.04	2.98	0.35
Spring	Variable	Model Fitting (N = 1237, day = 21)				Day-Site Evaluation (N = 198, day = 21)			
		Mean	Min	Max	SD	Mean	Min	Max	SD
	PM <sub>2.5</sub> (µg/m <sup>3</sup> )	68.02	3	279	38.43	68.72	12	257	35.89
	AOD (Unit less)	0.82	0.08	3.51	0.41	0.69	0.11	3.21	0.39
Summer	Variable	Model Fitting (N = 809, day = 16)				Day-Site Evaluation (N = 182, day = 16)			
		Mean	Min	Max	SD	Mean	Min	Max	SD
	PM <sub>2.5</sub> (µg/m <sup>3</sup> )	39.50	3	400	23.25	41.50	14	400	26.25
	AOD (Unit less)	0.67	0.04	2.33	0.34	0.67	0.06	2.33	0.35
Autumn	Variable	Model Fitting (N = 1014, day = 18)				Day-Site Evaluation (N = 181, day = 18)			
		Mean	Min	Max	SD	Mean	Min	Max	SD
	PM <sub>2.5</sub> (µg/m <sup>3</sup> )	57.95	5	205	35.62	62.5	9	235	35.62
	AOD (Unit less)	0.58	0.035	2.50	0.42	0.59	0.04	2.70	0.51
Winter	Variable	Model Fitting (N = 422, day = 11)				Day-Site Evaluation (N = 154, day = 11)			
		Mean	Min	Max	SD	Mean	Min	Max	SD
	PM <sub>2.5</sub> (µg/m <sup>3</sup> )	96.76	5	284	50.10	102.3	25	267	45
	AOD (Unit less)	0.62	0.037	2.94	0.41	0.59	0.04	3.02	0.41

#### 3.2. Proper Auxiliary Variables Analysis

We first respectively add each of the nine auxiliary variables into the simple GWR model. Figure 3 illustrates the determined coefficients R<sup>2</sup> of each auxiliary variable in its seasonal models. In four seasons, VSB contributes greatly to improving the GWR performance, whereas Georm and NDVI have little contributions to the GWR model in fitting the AOD-PM<sub>2.5</sub> relationships. The contributions from Vpre, Temp, WS, Elev, and Apre in GWR modelling diverge greatly across different seasons. In spring, the mean and median of WS, Vpre, VSB, and Temp variables show better contributions than other variables. The wider variations of R<sup>2</sup> from NDVI and Apre indicate more variations of their contributions in spring GWR modelling. In summer, Vpre, and VSB show remarkable performance, followed by Elev and Apre, while the WS variable shows less impact compared with its contribution

in spring GWR modelling. In autumn, WS, Temp, Vpre, and VSB have remarkable contributions compared to other variables, regardless that Apre is less stable than the four variables above. In winter, all the variables have relatively unstable contributions in GWR modelling with wider ranges of  $R^2$ , and the reason for this might be the relatively smaller fitting samples and the fluctuating meteorological conditions in winter. From the above contributions of all the single auxiliary variables, we preliminarily selected the Elev, WS, Apre, Vpre, VSB, and Temp for further analysis.



**Figure 3.** Contributions from nine auxiliary variables in GWR modelling of four seasons and the whole year. (a) spring, (b) summer, (c) autumn, (d) winter, (e) whole year, (f) curve line of contributions from all auxiliary variables in four seasons and the whole year. The box gives the 25%–75% percentile and the line in the box denotes the median. The whisker is the maximum and minimum of  $R^2$ , the points outside the box are outliers, inside the box an average of  $R^2$ .

With Spearman correlation coefficient analysis, the number of auxiliary variables is pruned to reduce the collinearity and redundancy among different variables. For the spring model, the Temp is removed because of its serious collinearity with Vpre. The Apre variable is discarded in the summer

model because of its clear redundancy with Elev. The Vpre variable is dropped in the autumn model because of its collinearity with VSB. The variables in the winter model do not show clear collinearity. Three variables are left for four seasonal models and the yearly model respectively, WS, Vpre, and VSB for spring, Elev, Vpre, VSB for summer, WS, Temp, VSB for autumn, Apre, RH, VSB for winter and WS, Vpre and VSB for the whole year.

We implement factor analysis to verify the representatives of pruned auxiliary variables. We use the factor analysis method to extract the common factor and calculate the factor load matrix. The result shows the first four factors separately reach their 70.52% accumulated variance in spring, 71.31% in summer, 74.64% in autumn, and 67.38% in winter. The first common factor includes AOD and VSB, sometimes with RH in. The second common factor consistently includes Vpre and Temp. The third common factor usually includes WS and Elev. The last common factor includes NDVI and Geom. Apre is possible to be involved in both the first and third common factors. We respectively name these four common factor components comprehensive index, vertical diffusion effect, horizontal diffusion effect, and geographic effect according to their physical mechanism with the ground-level PM<sub>2.5</sub> concentrations. In this case, we find the pruned auxiliary variables exactly represent three common meteorology factors, and therefore factor analysis provides theoretical supports for our selected auxiliary variables. However, the geographic effect does not involve variables in our model, we will explain the reason latter.

From the above, we finally obtain three proper auxiliary variables for four seasonal GWR models and the year GWR model (listed in the supplementary file). The proper auxiliary variables are WS, Vpre, and VSB for spring GWR model, Elev, Vpre, VSB for summer model, WS, Temp, VSB for autumn model and Apre, RH, VSB for winter model. Integrating the proper auxiliary variables with the main variable, four seasonal GWR models and the year model for further comparison are finally obtained.

### 3.3. Evaluation and Verification of Seasonal GWR Models

In order to validate our four seasonal models, we implement two groups of experiments to validate our four seasonal models. The first experiment in comparison with other GWR models by changing different auxiliary variable combination is to verify the performance of our proper auxiliary variable combination in four seasonal models. The second experiment is to evaluate the behaviors of four seasonal GWR models in predicting daily PM<sub>2.5</sub> concentrations.

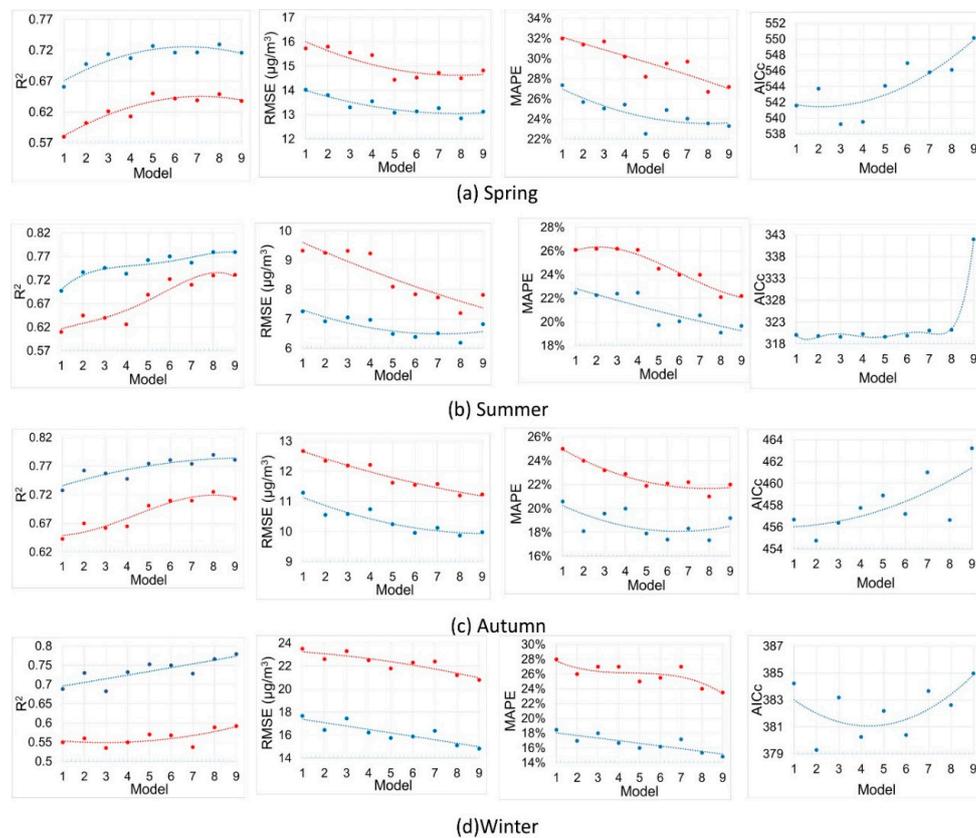
#### 3.3.1. Comparison of Regular GWR Models with Varied Auxiliary Variables

In order to testify the performance of three variable combinations, we respectively change the number of auxiliary variables from 0–4 to construct eight comparison GWR models in each season. Table 4 lists auxiliary variables of 36 GWR models of four seasons in the comparison, and the nine models can be grouped into five types: “AOD + 0”, “AOD + 1”, “AOD + 2”, “AOD + 3” (our model with three proper auxiliary variables), and “AOD + 4”. The fourth variable in the comparison compared against our seasonal models is manually selected according to its contribution of a single variable in regular GWR modelling.

**Table 4.** The list of all GWR models with different variable combinations.

Model Groups	Models	Spring	Summer	Autumn	Winter	Year
AOD + 0	1			AOD		
AOD + 1	2	AOD, WS	AOD, Elev	AOD, WS	AOD, Apre	AOD, WS
	3	AOD, Vpre	AOD, Vpre	AOD, Temp	AOD, RH	AOD, Vpre
	4	AOD, VSB	AOD, VSB	AOD, VSB	AOD, VSB	AOD, VSB
AOD + 2	5	AOD, WS, Vpre	AOD, Elev, Vpre	AOD, WS, Temp	AOD, Apre RH	AOD, WS, Vpre
	6	AOD, WS, VSB	AOD, Elev, VSB	AOD, WS, VSB	AOD, Apre, VSB	AOD, WS, VSB
	7	AOD, Vpre, VSB	AOD, Vpre, VSB	AOD, Temp, VSB	AOD, RH, VSB	AOD, Vpre, VSB
AOD + 3 (Ours)	8	AOD, WS, Vpre, VSB	AOD, Elev, Vpre, VSB	AOD, WS, Temp, VSB	AOD, Apre, RH, VSB	AOD, WS, Vpre, VSB
AOD + 4	9	AOD, WS, Vpre, VSB, Elev	AOD, Elev Temp, Vpre, VSB,	AOD, WS, Temp, Vpre, VSB	AOD, WS, Apre, RH, VSB	AOD, WS, Temp, Vpre, VSB

Figure 4 demonstrates all the model fitting and evaluation results from all the models in four seasons. In spring, the  $R^2$  of model fitting and model evaluation rises when the auxiliary variables gradually increase from 1–3, consistent with the decreasing RMSE and MAPE from  $14.0 \mu\text{g}/\text{m}^3$  to  $12.8 \mu\text{g}/\text{m}^3$  and from 27% to 22% in model fitting respectively. The optimal  $R^2$ , MAPE, and AICc reach the optimal result in model 5, although the RMSE has the minimum value in model 8 of “AOD + 3”. With the variable number varying from 1–4, the overfitting degree of all models decreases from “AOD + 0”, achieves the bottom at “AOD + 3” and then increases from “AOD + 3” to “AOD + 4”. Specifically, our spring GWR model of “AOD + 3” has the slightest overfitting degree whereas its “AOD + 2” model encounters the most serious overfitting.



**Figure 4.** (a–d) The comparison of all seasonal GWR models with different auxiliary variables. Blue scattering points are quantitative measures ( $R^2$ , RMSE, MAPE and AICc) of model fitting and red points are those of model evaluation. The dash lines are fitted curves of all scattered points in different seasons.

In summer, an overall gradual improvement of model performance exists from “AOD + 0” to “AOD + 3”, with a slight descending in “AOD + 4”. For the GWR model fitting process,  $R^2$  increases from 0.69 to 0.78, RMSE decreases from  $7.25 \mu\text{g}/\text{m}^3$  to  $6.18 \mu\text{g}/\text{m}^3$ , and MAPE decreases from 22.5% to 19.0%. The model performance of GWR slightly descends from “AOD + 3” to “AOD + 4”, and the model 8 of “AOD + 3” behaves best of all the models in summer. From model 1 to model 8, AICc values vary within a smaller range from 320.1 to 321.7, and the overall tendency of AICc in model evaluation is consistent with that of model fitting.

Similar to that of summer, the  $R^2$ , RMSE, and MAPE in autumn models show gradual improvement in performance from model 1 to model 8, with  $R^2$  increasing from 0.72 to 0.75, RMSE decreasing from  $11.35 \mu\text{g}/\text{m}^3$  to  $10.17 \mu\text{g}/\text{m}^3$  and MAPE decreasing from 22.5% to 19.0%. Model 8 of “AOD + 3” has the highest  $R^2$ , lowest RMSE and MAPE and relatively less overfitting among all the models and it performs best of all.

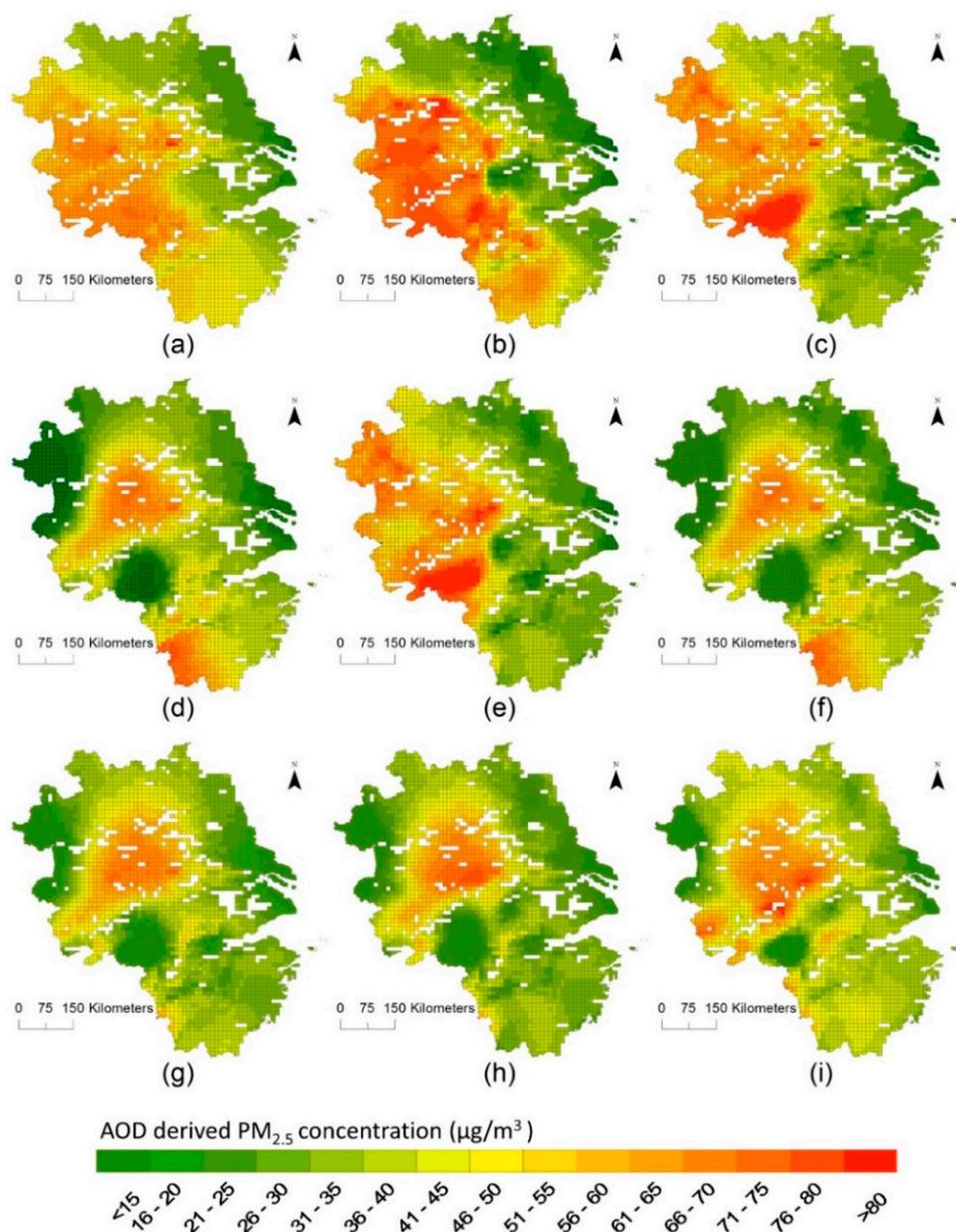
In winter, the variation of  $R^2$ , RMSE, and MAPE diverges more than those of the other three seasons. The explanation for this is partly because of fewer samples in both model fitting and evaluation. In terms of  $R^2$ , Figure 4 shows a more apparent increasing tendency and does not have descending tendency at “AOD + 4” model. From model 1 to model 8,  $R^2$  increases from 0.68 to 0.77, RMSE decreases from  $17.67 \mu\text{g}/\text{m}^3$  to  $15.11 \mu\text{g}/\text{m}^3$ , MAPE decreases from 18.5% to 15.3%. More variation and severer overfitting also occur in the winter models, with the decreasing  $R^2$  averaged at 0.173 from model fitting to model evaluation.

From the above, the comparison with changing numbers of auxiliary variables explains the best performance of our “AOD + 3” model, and verifies the effectiveness of our selected proper auxiliary variables in modelling seasonal GWR models.

Moreover, we listed key coefficients of auxiliary variables involved in nine GWR models from 1–9 on 16 September 2013 to further explain the special performance of our “AOD + 3” GWR model. We selected the day because of its high AOD coverage rate. Table 5 shows the coefficients of all involved auxiliary variables in the nine models. The results show that the auxiliary variables have clear effects in model fitting and evaluation. From model 1 to model 9, the changing combinations of the auxiliary variables explain their divergent contributions in improving the performance of the GWR model. Among all the nine models, the “AOD + 3” of model 8 performs best, with lowest overfitting in model fitting and evaluation. Figure 5 illustrates spatial distributions of  $\text{PM}_{2.5}$  retrieved from these nine models on 16 September 2013.

**Table 5.** Parameter estimations and Mean Absolute Percentage Errors (MAPEs) for GWR models 1–9 on 16 September 2013.

Model	Parameter Estimate					MAPE (%)	
	$\beta_{\text{AOD}}$	$\beta_{\text{WS}}$	$\beta_{\text{Temp}}$	$\beta_{\text{VSB}}$	$\beta_{\text{Vpre}}$	Fitting	Evaluation
1	6.87	—	—	—	—	22.17	35.91
2	7.56	1.28	—	—	—	21.89	35.29
3	6.67	—	4.31	—	—	21.85	36.34
4	0.81	—	—	−6.65	—	22.64	33.18
5	6.38	0.55	2.24	—	—	22.24	34.34
6	2.77	1.33	—	−9.49	—	20.92	33.05
7	3.22	—	6.38	−5.82	—	20.99	34.04
8	3.09	0.61	3.89	−5.00	—	20.81	32.25
9	3.84	0.81	5.98	−4.81	1.45	20.92	32.79



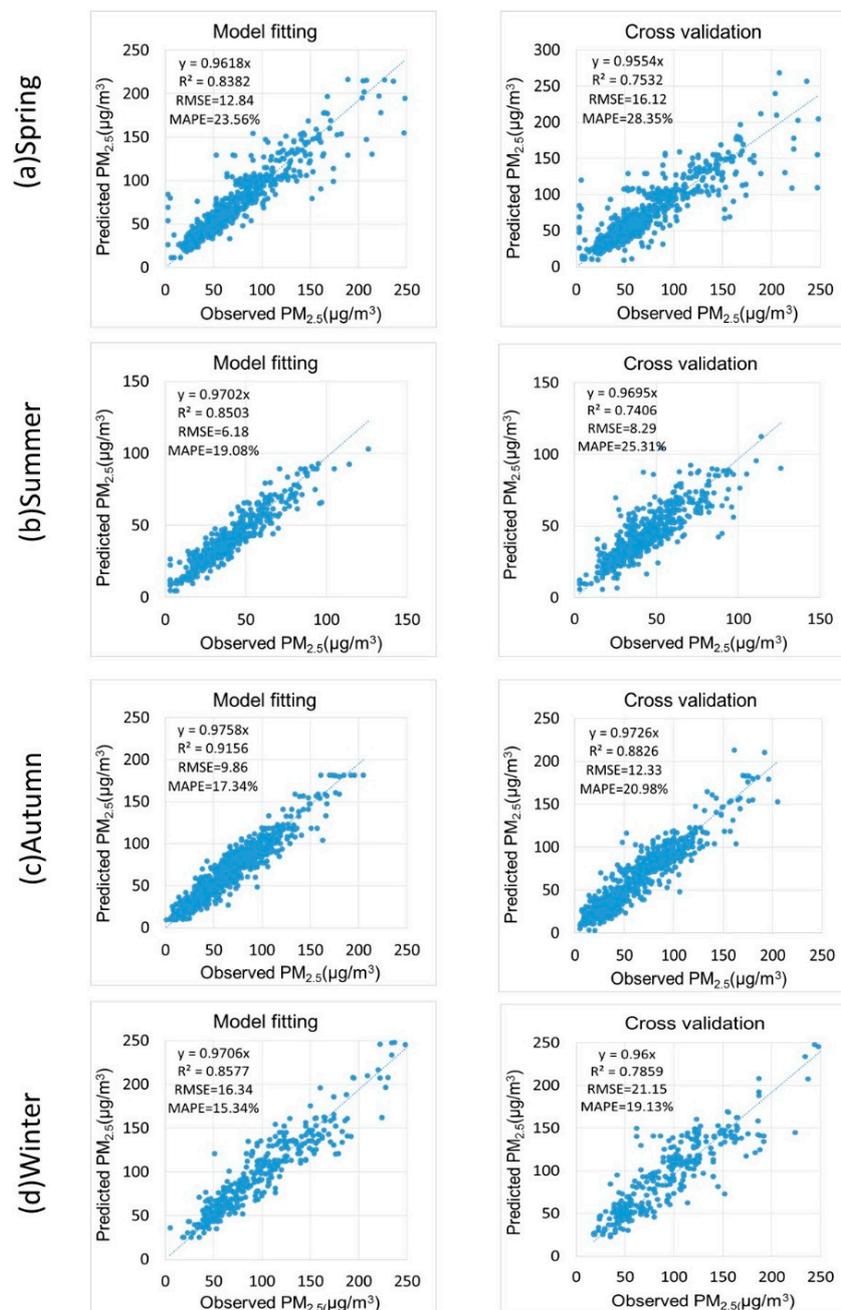
**Figure 5.** Spatial distribution maps of retrieved  $PM_{2.5}$  concentrations of all 9 models on 16 September 2013. The radial basis interpolation (RBF) method was utilized to interpolate meteorological variables and grid them into cells with spatial resolution of 0.1 degree. (a–i) correspond to retrieved maps of  $PM_{2.5}$  concentrations from models 1–9 in Table 4.

### 3.3.2. Comparison with the Observed $PM_{2.5}$ Concentrations

In this section, we validate the behaviors of four seasonal “AOD + 3” GWR models in predicting daily  $PM_{2.5}$  concentrations, using 10-fold cross validation scheme.

Figure 6 depicts the regression results with zero intercept between our predicted  $PM_{2.5}$  against observed  $PM_{2.5}$  measures. All the four models behave as slightly overfitting when compared against their cross-validation results. The result also shows that our seasonal GWR models in summer and autumn perform better than those in spring and winter. In spring, the slope and  $R^2$  for model fitting are 0.9618 and 0.8328. In summer, the slope and  $R^2$  for model fitting are 0.9702 and 0.8503. In autumn, the slope and  $R^2$  for model fitting are 0.9758 and 0.9156. In winter, the slope and  $R^2$  for model fitting

are 0.9706 and 0.8577. The slopes of four seasons are all less than 1, indicating that the estimated model generally underestimates the actual observed data. All the four models behave as slightly overfitting from model fitting to evaluation, with the slightest increase of RMSE in the summer and the most severe increase in the winter. The main reason for the worse performance in winter is partly because of the relatively small sample size. We argue that the frequent eruption of heavy pollution episodes in the winter of 2013 [35] also aggravated the possibility of unexpected situations in  $PM_{2.5}$  estimation, causing higher RMSE values in the winter GWR model.



**Figure 6.** (a–d) Comparison between observed  $PM_{2.5}$  and predicted  $PM_{2.5}$  concentrations in the four seasonal models. The dashed lines are regression lines.

Moreover, we compare the autumn GWR model with the yearly model to further validate the specific seasonal variables pattern of our GWR models. The yearly model has the same auxiliary

variables with our autumn model, and the reason for choosing the autumn model is because of its relatively larger sample size for GWR modelling. The result shows the autumn GWR model performs better than the yearly model, with  $1.2 \mu\text{g}/\text{m}^3$  lower RMSE and 2.79% lower MAPE in model fitting (Figure S1 in the supplementary file). The cases are similar in model evaluation. In addition, the overfitting is more severe in the yearly model, with a bigger gap between fitting and evaluation. The above also indicates that it is of great necessity to consider the seasonal auxiliary variables and establish seasonal models rather than the yearly model to guarantee the performance of the GWR model.

#### 4. Discussion

Satellite images provide a reliable method to retrieve spatial coverages of  $\text{PM}_{2.5}$  concentrations. One key problem is how to construct the GWR relations between satellite based AOD and ground-level  $\text{PM}_{2.5}$  measures, with the help of auxiliary meteorological or geographical factors. The auxiliary variables arbitrarily adapted in the GWR model might have unclear contributions to the  $\text{PM}_{2.5}$  concentration retrievals. Meanwhile, the contribution from single auxiliary variables or their combinations to the GWR modelling varies across different seasons. Therefore, we proposed an automatic procedure to select proper meteorological or geographical factors for GWR modelling in different seasons, rendering their proper auxiliary variable combinations to construct optimal daily GWR models in the YRD region. We made careful comparisons between our seasonal GWR models and regular GWR models with different auxiliary variables, between the predicted  $\text{PM}_{2.5}$  concentrations from our models and the observed ones, in order to validate our model performance. The results explain that our seasonal models have better performance for  $\text{PM}_{2.5}$  retrievals than the comparison models.

Theoretically speaking, the relationship between daily AOD and  $\text{PM}_{2.5}$  is widely affected by few auxiliary variables. Also the proper choice of auxiliary variables could then upgrade the model accuracy as well as reduce the model redundancy. The effects from nine preliminary auxiliary variables which we adopted in  $\text{PM}_{2.5}$  concentrations, can be grouped into four different aspects—comprehensive effect, vertical diffusion effect, horizontal diffusion effect, and geographical effect. The contribution of each selected auxiliary variable in our models is not constant, and it varies greatly among different seasons like that of TSAM [34]. VSB is a comprehensive indicator of air quality showing significant influences in all four seasons.  $V_{\text{pre}}$  is the partial pressure of water vapor in the whole atmosphere column, and our experimental results indicate that it is a good auxiliary variable for GWR modelling in spring, summer, and autumn when compared with the widely adopted RH in the current literature. The reason for this we guessed is because  $V_{\text{pre}}$  reflects and corrects the humidity of the whole aerosol column.  $A_{\text{pre}}$  shows greater contribution in the winter GWR model than in those of other seasons. We guess the reason for this is that the high air pressure occurring in winter always comes along with cold air and sand-dust from the north or northwest of China, and that greatly increases the  $\text{PM}_{2.5}$  concentrations in the YRD. Two geographical variables, Geom and NDVI, were not significant in the GWR modelling and accordingly were excluded from the auxiliary variable combinations. The relatively simple geomorphy features of the YRD, with 49% of plain and 7.9% of mountain, render that the Geom shows insignificant contributions in the GWR modelling. Although NDVI is relatively weaker with respect to other contemporary meteorology variables, it shows significant contributions in the spring and winter models. That indicates seasonal contribution patterns of NDVI in the  $\text{PM}_{2.5}$  retrieval model. In general, our experimental results illustrate that the daily AOD- $\text{PM}_{2.5}$  relationships closely correlate with meteorological factors rather than geographical factors. The explanation for this is that the shifting meteorology conditions highly affect the generation or diffusion of daily  $\text{PM}_{2.5}$  concentrations in the short term whereas geographical factors are crucial to long term prediction [30].

The combination of proper auxiliary variables in our seasonal models is carefully investigated in our study by comparing with regular GWR models via changing the number of auxiliary variables. The comparison between our seasonal models and the yearly model illustrates the importance of

proper auxiliary variables. Generally, our seasonal models have the optimal performance with three auxiliary variables, and less or more variables than three would bring about instability or redundancy of the GWR models. Adding extra auxiliary variables shows little improvement in the GWR model but even reduces the model performance and aggravates the overfitting, due to the increasing risk of intra-correlations among variables [41]. We further analyzed the auxiliary variables in four seasonal model “AOD + 3” and guessed that they coincided with three meteorological effects in PM<sub>2.5</sub> concentrations. For the spring, WS determines the horizontal diffusion effect of PM<sub>2.5</sub>, Vpre correlates closely with the vertical diffusion of PM<sub>2.5</sub>, and VSB represents the comprehensive effect of PM<sub>2.5</sub> concentrations. For the summer season, Elev replaces WS in spring and has a close relation with the horizontal diffusion effect. For the autumn model, Temp correlates closely with the air vertical movement. For winter, Apre shows more correlation with the air vertical movement. The seasonal average PM<sub>2.5</sub> prediction of model evaluation R<sup>2</sup> is 0.79, and the result of our seasonal models is comparable to those of the previous works by Fang [34], but with less variables and computation cost. We also applied the specific seasonal GWR models to 2014, and the achieved result is satisfying (Figure S2 in the supplementary file).

Unfortunately, our study simultaneously has the following shortcomings needing improvement in the future. First, the selection of the seasonal variable largely depends on our numerous empirical statistical works and the analysis of common factors. Also, theoretical explanations of specific seasonal variables cannot be fully figured out and need to be further investigated. Second, only natural factors were involved in our study and which might result in neglecting some potential effects from anthropic factors especially social geographical factors. Actually, anthropic factors, especially social geographical variables have great potential in improving the PM<sub>2.5</sub> retrieval model [42]. The supplement work of more anthropic variables into our seasonal GWR models will be completed in a further study. Finally, the filling problem of absent pixels in the AOD product requires careful investigation in order to increase the number of validated records for more robust GWR modelling and to promote the utility of proposed seasonal GWR model in all daily conditions of the YRD region.

## 5. Conclusions

Auxiliary variables like meteorological or geographical factors show great potential in improving the GWR model accuracy for retrieving PM<sub>2.5</sub> concentrations from satellite AOD. However, the selection of proper variables and their different contributions in the four seasons have never been carefully investigated, hampering further applications of the GWR model in the YRD region. In this study, we put forward an automatic procedure of seasonal proper variable selection considering the contribution of each single variable and its inner structure in the GWR modelling. Moreover, we investigated the seasonal pattern of auxiliary variables, and constructed four seasonal GWR models with properly selected auxiliary variables. Our seasonal GWR models with proper variable combinations were tested with two groups of experiments. The seasonal GWR models was compared with regular GWR models by changing auxiliary variables and the predicted PM<sub>2.5</sub> concentrations from the four seasonal models were compared against the observed measures and that of the yearly GWR model. Our selected proper auxiliary variables in four models of “AOD + 3” reduce the redundancy of regular GWR models and simultaneously help to obtain better model accuracy than with other regular GWR models. The prediction performance of the four seasonal models behaves better than the yearly model and the predicted result coincides well with the observed ones, having high R<sup>2</sup> in model evaluation, averaged as 0.79 in 10-fold cross validation. Therefore, seasonal varied contribution of variables should be taken into consideration in PM<sub>2.5</sub> concentration retrieval and our seasonal GWR models could be a good alternative in modelling daily PM<sub>2.5</sub> concentrations from remote sensing in the YRD region.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2072-4292/9/4/346/s1](http://www.mdpi.com/2072-4292/9/4/346/s1).

**Acknowledgments:** This research was funded by National Natural Science Foundation (41401389, 41671342), by the Chinese Postdoctoral Science Foundation (2015M570668, 2016T90732), by the Public Projects of Zhejiang Province (2016C33021), by the Zhejiang University Student Science and Technology Innovation and Xin-Miao Talented Plan Program (2016R405088), and by the K. C. Wong Magna Fund in Ningbo University. The authors thank Qingyue Open Environmental Data Center (<https://data.epmap.org>) for their generous support in hourly ground level PM<sub>2.5</sub> data processing. The authors also would like to thank the editor and referees for their suggestions that improved the manuscript.

**Author Contributions:** All the coauthors made significant contributions to the manuscript. Man Jiang carried on most experiments and wrote the manuscript. Weiwei Sun presented the key idea of seasonal GWR model and designed the experiments. Gang Yang compared the seasonal GWR model with the yearly GWR model and helped to revise the manuscript. Dianfa Zhang provided the background knowledge of atmospheric remote sensing and gave important suggestions for improving the manuscript.

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

## References

1. Kaufman, Y.J.; Tanré, D.; Boucher, O. A satellite view of aerosols in the climate system. *Nature* **2002**, *419*, 215–223. [[CrossRef](#)] [[PubMed](#)]
2. An, X.; Hou, Q.; Li, N.; Zhai, S. Assessment of human exposure level to PM<sub>10</sub> in China. *Atmos. Environ.* **2013**, *70*, 376–386. [[CrossRef](#)]
3. Dockery, D.W.; Pope, C.A.; Xu, X.; Spengler, J.D.; Ware, J.H.; Fay, M.E.; Ferris, B.G., Jr.; Speizer, F.E. An association between air pollution and mortality in six U.S. Cities. *N. Engl. J. Med.* **1993**, *329*, 1753–1759. [[CrossRef](#)] [[PubMed](#)]
4. Wang, Z.; Yang, L.; Mu, H.; Pan, X.; Jing, S.; Feng, C.; He, K.; Koutrakis, P.; Christiani, D.C. Acute health impacts of airborne particles estimated from satellite remote sensing. *Environ. Int.* **2013**, *51*, 150–159. [[CrossRef](#)] [[PubMed](#)]
5. Weichenthal, S.; Villeneuve, P.J.; Burnett, R.T.; van Donkelaar, A.; Martin, R.V.; Jones, R.R.; Dellavalle, C.T.; Sandler, D.P.; Ward, M.H.; Hoppin, J.A. Long-term exposure to fine particulate matter: Association with nonaccidental and cardiovascular mortality in the agricultural health study cohort. *Environ. Health Perspect.* **2014**, *122*, 609–615. [[CrossRef](#)] [[PubMed](#)]
6. Hu, Z. Spatial analysis of MODIS aerosol optical depth, PM<sub>2.5</sub>, and chronic coronary heart disease. *Int. J. Health Geogr.* **2009**, *8*, 1–10. [[CrossRef](#)] [[PubMed](#)]
7. Shi, L.; Zanobetti, A.; Kloog, I.; Coull, B.A.; Koutrakis, P.; Melly, S.J.; Schwartz, J.D. Low-concentration PM<sub>2.5</sub> and mortality: Estimating acute and chronic effects in a population-based study. *Environ. Health Perspect.* **2016**, *124*, 46–52. [[CrossRef](#)] [[PubMed](#)]
8. Wei, Y.; Zang, Z.; Zhang, L.; Yi, L.; Wang, W. Estimating national-scale ground-level PM<sub>25</sub> concentration in china using geographically weighted regression based on MODIS and MISR AOD. *Environ. Sci. Pollut. Res.* **2016**, 1–12. [[CrossRef](#)] [[PubMed](#)]
9. Ma, Z.; Hu, X.; Huang, L.; Bi, J.; Liu, Y. Estimating ground-level PM<sub>2.5</sub> in China using satellite remote sensing. *Environ. Sci. Technol.* **2014**, *48*, 7436–7444. [[CrossRef](#)] [[PubMed](#)]
10. Liu, Y.; Paciorek, C.J.; Koutrakis, P. Estimating regional spatial and temporal variability of PM<sub>2.5</sub> concentrations using satellite data, meteorology, and land use information. *Environ. Health Perspect.* **2009**, *117*, 886–892. [[CrossRef](#)] [[PubMed](#)]
11. Yao, L.; Lu, N. Spatiotemporal distribution and short-term trends of particulate matter concentration over China, 2006–2010. *Environ. Sci. Pollut. Res.* **2014**, *21*, 9665–9675. [[CrossRef](#)] [[PubMed](#)]
12. 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 PM<sub>2.5</sub> mass and species during 2010 in Xi'an, China. *Sci. Total Environ.* **2015**, *508*, 477–487. [[CrossRef](#)] [[PubMed](#)]
13. Strawa, A.W.; Chatfield, R.B.; Legg, M.J.; Scarnato, B.V.; Esswein, R. In improving retrievals of regional PM<sub>2.5</sub> concentrations from MODIS and OMI multi-satellite observations. In Proceedings of the American Geophysical Union 2013 Fall Meeting, San Francisco, CA, USA, 9–13 December 2013.

14. Hu, X.; Waller, L.A.; Lyapustin, A.; Wang, Y.; Al-Hamdan, M.Z.; Crosson, W.L.; Estes, M.G.; Estes, S.M.; Quattrochi, D.A.; Puttaswamy, S.J. Estimating ground-level PM<sub>2.5</sub> concentrations in the southeastern united states using MAIAC AOD retrievals and a two-stage model. *Remote Sens. Environ.* **2014**, *140*, 220–232. [[CrossRef](#)]
15. Song, W.; Jia, H.; Huang, J.; Zhang, Y. A satellite-based geographically weighted regression model for regional PM 2.5 estimation over the Pearl River Delta region in China. *Remote Sens. Environ.* **2014**, *154*, 1–7. [[CrossRef](#)]
16. You, W.; Zang, Z.; Zhang, L.; Li, Y.; Pan, X.; Wang, W. National-scale estimates of ground-level PM<sub>2.5</sub> concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD. *Remote Sens.* **2016**, *8*, 184. [[CrossRef](#)]
17. Guo, H.; Cheng, T.; Gu, X.; Chen, H.; Wang, Y.; Zheng, F.; Xiang, K. Comparison of four ground-level PM<sub>2.5</sub> estimation models using parasol aerosol optical depth data from China. *Int. J. Environ. Res. Public Health* **2016**, *13*, 180. [[CrossRef](#)] [[PubMed](#)]
18. Ma, Z.; Hu, X.; Sayer, A.M.; Levy, R.; Zhang, Q.; Xue, Y.; Tong, S.; Bi, J.; Huang, L.; Liu, Y. Satellite-based spatiotemporal trends in PM<sub>2.5</sub> concentrations: China, 2004–2013. *Environ. Health Perspect.* **2016**, *124*, 184–192. [[CrossRef](#)] [[PubMed](#)]
19. Liu, Y.; Park, R.J.; Jacob, D.J.; Li, Q.; Kilaru, V.; Sarnat, J.A. Mapping annual mean ground-level PM<sub>2.5</sub> concentrations using multiangle imaging spectroradiometer aerosol optical thickness over the contiguous united states. *J. Geophys. Res. Atmos.* **2004**, *109*, 2285–2311.
20. Boys, B.L.; Martin, R.V.; Van, D.A.; Macdonell, R.J.; Hsu, N.C.; Cooper, M.J.; Yantosca, R.M.; Lu, Z.; Streets, D.G.; Zhang, Q. Fifteen-year global time series of satellite-derived fine particulate matter. *Environ. Sci. Technol.* **2014**, *48*, 11109–11118. [[CrossRef](#)] [[PubMed](#)]
21. Lin, C.; Li, Y.; Yuan, Z.; Lau, A.K.; Li, C.; Fung, J.C. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM 2.5. *Remote Sens. Environ.* **2015**, *156*, 117–128. [[CrossRef](#)]
22. Zhang, Y.; Li, Z. Remote sensing of atmospheric fine particulate matter (PM<sub>2.5</sub>) mass concentration near the ground from satellite observation. *Remote Sens. Environ.* **2015**, *160*, 252–262. [[CrossRef](#)]
23. Ma, Z. Study on Spatiotemporal Distribution of PM<sub>2.5</sub> in China Using Satellite Remote Sensing. Ph.D. Thesis, Nanjing University, Nanjing, China, 2015.
24. Li, Z.; Zhang, Y.; Shao, J.; Li, B.; Hong, J.; Liu, D.; Li, D.; Wei, P.; Li, W.; Li, L. Remote sensing of atmospheric particulate mass of dry PM<sub>2.5</sub> near the ground: Method validation using ground-based measurements. *Remote Sens. Environ.* **2016**, *173*, 59–68. [[CrossRef](#)]
25. Tian, J.; Chen, D. A semi-empirical model for predicting hourly ground-level fine particulate matter (PM 2.5) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sens. Environ.* **2010**, *114*, 221–229. [[CrossRef](#)]
26. You, W.; Zang, Z.; Zhang, L.; Li, Z.; Chen, D.; Zhang, G. Estimating ground-level PM<sub>10</sub> concentration in northwestern China using geographically weighted regression based on satellite AOD combined with Calipso and MODIS fire count. *Remote Sens. Environ.* **2015**, *168*, 276–285. [[CrossRef](#)]
27. HuHu, X.; Waller, L.A.; Al-Hamdan, M.Z.; Crosson, W.L.; Estes, M.G.; Estes, S.M.; Quattrochi, D.A.; Sarnat, J.A.; Yang, L. Estimating ground-level PM(2.5) concentrations in the southeastern U.S. Using geographically weighted regression. *Environ. Res.* **2012**, *121*, 1–10. [[CrossRef](#)] [[PubMed](#)]
28. Ma, Z.; Liu, Y.; Zhao, Q.; Liu, M.; Zhou, Y.; Bi, J. Satellite-derived high resolution PM concentrations in Yangtze river delta region of China using improved linear mixed effects model. *Atmos. Environ.* **2016**, *133*, 156–164. [[CrossRef](#)]
29. Bai, Y.; Wu, L.; Qin, K.; Zhang, Y.; Shen, Y.; Zhou, Y. A geographically and temporally weighted regression model for ground-level PM<sub>2.5</sub> estimation from satellite-derived 500 m resolution AOD. *Remote Sens.* **2016**, *8*, 262. [[CrossRef](#)]
30. Kloog, I.; Koutrakis, P.; Coull, B.A.; Lee, H.J.; Schwartz, J. Assessing temporally and spatially resolved PM 2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ.* **2011**, *45*, 6267–6275. [[CrossRef](#)]
31. Hu, X. Estimation of PM<sub>2.5</sub> concentrations in the conterminous U.S. using MODIS data and a three-stage model. In Proceedings of the American Geophysical Union 2015 Fall Meeting, San Francisco, CA, USA, 14–18 December 2015.

32. Van Donkelaar, A.; Martin, R.V.; Spurr, R.J.; Burnett, R.T. High-resolution satellite-derived PM<sub>2.5</sub> from optimal estimation and geographically weighted regression over North America. *Environ. Sci. Technol.* **2015**, *49*, 10482–10491. [[CrossRef](#)] [[PubMed](#)]
33. Jiang, M.; Sun, W. Investigating meteorological and geographical effect in remote sensing retrieval of PM<sub>2.5</sub> concentration in Yangtze River Delta. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium, Beijing, China, 10–18 July 2016; pp. 4108–4111.
34. Fang, X.; Zou, B.; Liu, X.; Sternberg, T.; Zhai, L. Satellite-based ground PM 2.5 estimation using timely structure adaptive modeling. *Remote Sens. Environ.* **2016**, *186*, 152–163. [[CrossRef](#)]
35. Mu, Q.; Zhang, S. Assessment of the Trend of Heavy PM<sub>2.5</sub> Pollution Days and Economic Loss of Health Effects during 2001–2013. *Acta Sci. Nat. Univ. Pekin.* **2015**, *51*, 694–706.
36. Remer, L.A.; Kaufman, Y.J.; Tanré, D.; Mattoo, S.; Chu, D.A.; Martins, J.V.; Li, R.R.; Ichoku, C.; Levy, R.C.; Kleidman, R.G. The MODIS aerosol algorithm, products, and validation. *J. Atmos. Sci.* **2005**, *62*, 947–973. [[CrossRef](#)]
37. Cheng, Z.; Wang, S.X.; Fu, X.; Watson, J.G.; Jiang, J.K.; Fu, Q.Y.; Chen, C.H.; Xu, B.Y.; Yu, J.S.; Chow, J.C. Impact of biomass burning on haze pollution in the Yangtze River Delta, China: A case study of summer in 2011. *Atmos. Chem. Phys.* **2014**, *14*, 4573–4585. [[CrossRef](#)]
38. Xu, J.H.; Jiang, H. Estimation of PM<sub>2.5</sub> concentration over the Yangtze Delta using remote sensing: analysis of spatial and temporal variations. *Environ. Sci.* **2015**, *36*, 3119–3127. (In Chinese).
39. Fotheringham, A.S.; Charlton, M.E.; Brunsdon, C. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environ. Plan. A* **1998**, *30*, 1905–1927. [[CrossRef](#)]
40. Hauke, J.; Kossowski, T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaest. Geogr.* **2015**, *30*, 87–93. [[CrossRef](#)]
41. Tu, Y.K.; Kellett, M.; Clerehugh, V.; Gilthorpe, M.S. Problems of correlations between explanatory variables in multiple regression analyses in the dental literature. *Br. Dent. J.* **2005**, *199*, 457–461. [[CrossRef](#)] [[PubMed](#)]
42. Lin, G.; Fu, J.; Jiang, D.; Hu, W.; Dong, D.; Huang, Y.; Zhao, M. Spatio-temporal variation of PM<sub>2.5</sub> concentrations and their relationship with geographic and socioeconomic factors in China. *Int. J. Environ. Res. Public Health* **2013**, *11*, 173–186. [[CrossRef](#)] [[PubMed](#)]



© 2017 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 (<http://creativecommons.org/licenses/by/4.0/>).