

# A spatially structured adaptive two-stage model for retrieving ground-level PM<sub>2.5</sub> concentrations from VIIRS AOD in China

Fei Yao<sup>a,b</sup>, Jiansheng Wu<sup>a,c,\*</sup>, Weifeng Li<sup>d,e,\*</sup>, Jian Peng<sup>c</sup>

<sup>a</sup> Key Laboratory for Urban Habitat Environmental Science and Technology, Shenzhen Graduate School, Peking University, Shenzhen 518055, PR China

<sup>b</sup> School of GeoSciences, University of Edinburgh, Edinburgh, UK

<sup>c</sup> Laboratory for Earth Surface Processes, Ministry of Education, College of Urban and Environmental Sciences, Peking University, Beijing 100871, PR China

<sup>d</sup> Department of Urban Planning and Design, The University of Hong Kong, SAR, PR China

<sup>e</sup> Shenzhen Institute of Research and Innovation, The University of Hong Kong, Shenzhen 518075, PR China

## ARTICLE INFO

### Keywords:

PM<sub>2.5</sub>  
VIIRS AOD  
Spatially structured adaptive  
Two-stage model  
China

## ABSTRACT

While the aerosol optical depth (AOD) product from the Visible Infrared Imaging Suite (VIIRS) instrument has proven effective for estimating regional ground-level particle concentrations with aerodynamic diameters less than 2.5 μm (PM<sub>2.5</sub>), its performance at larger spatial scales remains unclear. Despite the wide application of statistical models in building ground-level PM<sub>2.5</sub> satellite remote sensing retrieval models, a limited number of studies have considered the spatiotemporal heterogeneities for model structures. Taking China as the study area, we used the VIIRS AOD, together with multi-source auxiliary variables, to develop a spatially structured adaptive two-stage model to estimate ground-level PM<sub>2.5</sub> concentrations at a 6-km spatial resolution. To this end, we first defined and calculated a dual distance from the ground-level PM<sub>2.5</sub> monitoring data. We then applied the un-weighted pair-group method with arithmetic means on dual distances and obtained 13 spatial clusters. Subsequently, we combined the time fixed effects regression (TEFR) model and geographically weighted regression (GWR) model to develop the spatially structured adaptive two-stage model. For each spatial cluster, we examined all possible combinations of auxiliary variables and determined the best model structure according to multiple statistical test results. Finally, we obtained the PM<sub>2.5</sub> estimates through regression mapping. At least seven model-fitting data records per day made a good threshold that could best overcome the model overfitting induced by the second-stage GWR model at the minimum price of losing samples. The overall model fitting and ten-fold cross validation (CV) R<sup>2</sup> were 0.82 and 0.60, respectively, under that threshold. Model performances among different spatial clusters differed to a certain extent. High-CV R<sup>2</sup> values always exceeded 0.6 while low-CV R<sup>2</sup> values less than 0.5 also existed. Both the size of the model-fitting data records and the extent of urban-industrial characteristics of spatial clusters accounted for these differences. The PM<sub>2.5</sub> estimates agreed well with the PM<sub>2.5</sub> observations with correlation coefficients all exceeding 0.5 at the monthly, seasonal, and annual scales. East of Hu's line and north of the Yangtze River were characterized by high PM<sub>2.5</sub> concentrations. This study contributes to the understanding of how well VIIRS AOD can retrieve ground-level PM<sub>2.5</sub> concentrations at the national scale and strategies for building ground-level PM<sub>2.5</sub> satellite remote sensing retrieval models.

## 1. Introduction

Epidemiological studies have disclosed various adverse health outcomes (e.g. increasing cardiovascular- and respiratory-related mortality and morbidity) caused by exposure to particles with an aerodynamic diameter of less than 2.5 μm (PM<sub>2.5</sub>) (Burnett et al., 2018; Dominici et al., 2006; Pope et al., 2002). China has become one of the regions facing the most severe PM<sub>2.5</sub> pollution worldwide due to rapid urbanization and industrialization (van Donkelaar et al., 2015; van

Donkelaar et al., 2016). Developing a PM<sub>2.5</sub> dataset with high accuracy, resolution, and coverage can benefit both epidemiological studies and environmental policy making. Using satellite-retrieved aerosol optical depth (AOD) data to estimate ground-level PM<sub>2.5</sub> concentrations is a new technical method that has experienced rapid development in recent years (Hoff and Christopher, 2009). It can effectively extend ground-based air quality monitoring networks due to the broad spatial coverage and relatively high spatial resolution of satellite-retrieved AOD (Della Ceca et al., 2018), thereby allowing for the acquisition of

\* Corresponding authors.

E-mail addresses: [wujs@pkusz.edu.cn](mailto:wujs@pkusz.edu.cn) (J. Wu), [wfli@hku.hk](mailto:wfli@hku.hk) (W. Li).

<https://doi.org/10.1016/j.isprsjprs.2019.03.011>

Received 25 August 2018; Received in revised form 6 January 2019; Accepted 16 March 2019

0924-2716/ © 2019 Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).

the aforementioned PM<sub>2.5</sub> dataset.

Previous studies have employed a series of AOD products to estimate ground-level PM<sub>2.5</sub> concentrations. These mainly include AOD products from geostationary satellite sensors, such as the Geostationary Ocean Color Imager (Xu et al., 2015) and the Advanced Himawari Imager (Wang et al., 2017), and polar-orbiting satellite sensors, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Ma et al., 2016a) and the Visible Infrared Imaging Radiometer Suite (VIIRS) (Wu et al., 2016). The MODIS instrument provides a long time-series of archived AOD products with reliable quality and direct availability that have been most widely used (Chu et al., 2016); however, it is already working beyond its design life. Consequently, the quality of its AOD products is inevitably declining (Lyapustin et al., 2014; Xiao et al., 2016). VIIRS is a follow-on instrument to MODIS with very similar capabilities (Jackson et al., 2013). While the VIIRS AOD data can estimate accurate ground-level PM<sub>2.5</sub> concentrations at regional scales (Wu et al., 2016; Yao et al., 2018), its performance at larger spatial scales (e.g. national scale) remains unclear.

Previous studies have adopted a variety of statistical models to establish quantitative PM<sub>2.5</sub>-AOD relationships. These primarily include the generalized linear regression model (Liu et al., 2007; 2005), generalized additive model (Liu et al., 2009; Zou et al., 2016), geographically weighted regression (GWR) model (Hu et al., 2013; Ma et al., 2014; Song et al., 2014), geographically and temporally weighted regression model (Guo et al., 2017; He and Huang, 2018a,b), linear or nested linear mixed effects model (Lee et al., 2011; Ma et al., 2016b), two-stage model (Hu et al., 2014; Ma et al., 2016a; Wu et al., 2016), and timely structure adaptive model (Fang et al., 2016). The models are relatively easy to calibrate and validate. Additionally, most of them can reach a high level of prediction accuracy due to the contributing strength of model predictors that are capable of changing in time, space, or both. Nevertheless, few of them have considered the spatiotemporal heterogeneities in the model structure. In future studies, both model predictors and their contributing strength should be varied with time and space to produce PM<sub>2.5</sub> estimates with higher accuracy and lower bias (Fang et al., 2016).

The objective of this study was to evaluate the performance of VIIRS AOD in estimating ground-level PM<sub>2.5</sub> concentrations on the national scale and to test a newly developed spatially structured adaptive two-stage model. The model, as a preliminary attempt, captured the spatial heterogeneity of model predictors and the spatiotemporal heterogeneities of their contributing strength. We first divided the geographical area of China into spatial clusters based on the inherent characteristics of ground-level PM<sub>2.5</sub> concentrations. Subsequently, we employed a time fixed effects regression (TFER) model and a GWR model to establish a two-stage model with the most suitable model structure for each spatial cluster. Next, we used a ten-fold cross validation (CV) and residual autocorrelation test to evaluate the model performance. Finally, we estimated the daily PM<sub>2.5</sub> concentrations, and thus the monthly, seasonal, and annual PM<sub>2.5</sub> concentrations.

## 2. Data and methods

Fig. 1 shows the workflow for estimating the ground-level PM<sub>2.5</sub> concentrations in this study. Details of each procedure are described in the following sections.

### 2.1. Ground-level PM<sub>2.5</sub> monitoring data

The Ministry of Ecology and Environment of China has been using the tapered element oscillating microbalance or beta-attenuation method to measure ground-level PM<sub>2.5</sub> concentrations and distributing them on the official website of the China National Environmental Monitoring Center (CNEMC) (<http://www.cnemc.cn/>). Some cities (e.g., Beijing, Tianjin, and Guangzhou) also established additional monitoring sites to measure ground-level PM<sub>2.5</sub> concentrations using

the same method, distributing them online. In this study, we acquired hourly ground-level PM<sub>2.5</sub> concentrations from the official website of CNEMC and the Beijing Municipal Environmental Monitoring Center (<http://zx.bjmemc.com.cn/>) (see Fig. 2). We further calculated the daily averages of PM<sub>2.5</sub> concentrations from the hourly data since the former are more meaningful for epidemiological studies and environmental management.

### 2.2. VIIRS EDR AOD and AERONET AOD data

VIIRS is a key scanning radiometer on board the Suomi-National Polar-orbiting Partnership satellite. It extends and improves upon the Advanced Very High Resolution Radiometer and MODIS (Bian et al., 2018), leading to improved radiometric measurement quality, broad spectral range, and fine spatial resolution (Schueler et al., 2002). The VIIRS aerosol algorithm, which is similar but not identical to the MODIS algorithm, retrieves AOD from reflected solar radiation observed in multiple moderate-resolution (750 m at nadir) VIIRS bands. This generates the VIIRS aerosol Intermediate Product (IP), which is a full set of aerosol parameters, including the AOD at 550 nm. The VIIRS aerosol team further aggregates 8 × 8 750 m IP AOD filtered using quality flags, leading to 6-km AOD, which is included in the VIIRS Environmental Data Record (EDR). In this study, we collected all the VIIRS EDR AOD images covering China from the Comprehensive Large Array-data Stewardship System ([www.class.noaa.gov](http://www.class.noaa.gov)). The VIIRS aerosol team also provides corresponding pixel-level quality assurance (QA) flags for the VIIRS EDR AOD data. These flags are four distinct numbers with zero representing “not produced”, one representing “low-quality”, two representing “medium-quality”, and three representing “high-quality”. Although obtaining the highest spatiotemporal coverage of the VIIRS EDR AOD data by keeping all data regardless of the QA flag would be ideal, the potential negative effects of AOD deviations on PM<sub>2.5</sub> estimates should be prevented. We employed ground monitored AOD from the Aerosol Robotic Network to validate VIIRS EDR AOD with different QA flags. We then decided to use VIIRS EDR AOD with QA flags equaling three or two in this study based on their validation results and spatiotemporal coverage; more details are presented in [Supporting Material: Text S1, Figs. S1–S4](#).

### 2.3. Auxiliary data

Both PM<sub>2.5</sub> emission-related factors and meteorological conditions influence, and thus can adjust, PM<sub>2.5</sub>-AOD relationships. For PM<sub>2.5</sub> emission-related factors, we mainly used the NO<sub>2</sub> and vegetation index data. NO<sub>2</sub> data based on ozone monitoring instrument measurements (Boersma et al., 2011) were from the Tropospheric Emission Monitoring Internet Service ([http://www.temis.nl/airpollution/no2col/no2regioomi\\_v2.php](http://www.temis.nl/airpollution/no2col/no2regioomi_v2.php)). The temporal and spatial resolution of the NO<sub>2</sub> data were 1-d and 0.25° × 0.25°, respectively. We used the NO<sub>2</sub> data of the previous day to adjust the PM<sub>2.5</sub>-AOD relationship of the current day, an effective practice adopted in previous studies (Wu et al., 2016; Yao et al., 2018). The vegetation index data were from the Level-1 and Atmospheric Archive and Distribution System (LAADS) (<https://ladsweb.modaps.eosdis.nasa.gov/search/>). A series of vegetation index products with different temporal and spatial resolutions are available on the LAADS website. We downloaded the Terra and Aqua MODIS vegetation index products with temporal and spatial resolutions of 16-d and 1-km, respectively (code: MOD13A2, MYD13A2). Both included the normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI). The combination of data from the two satellites improved the temporal resolution from 16-d to 8-d. Similar to VIIRS EDR AOD data, pixel reliability data also existed in these products (−1: no data, 0: good data, 1: marginal data, 2: snow or ice, and 3: cloudy). We omitted those data whose pixel reliability was not 0 or 1 to minimize the negative effects of vegetation index deviations on PM<sub>2.5</sub> estimates and simultaneously keep as much data as possible.

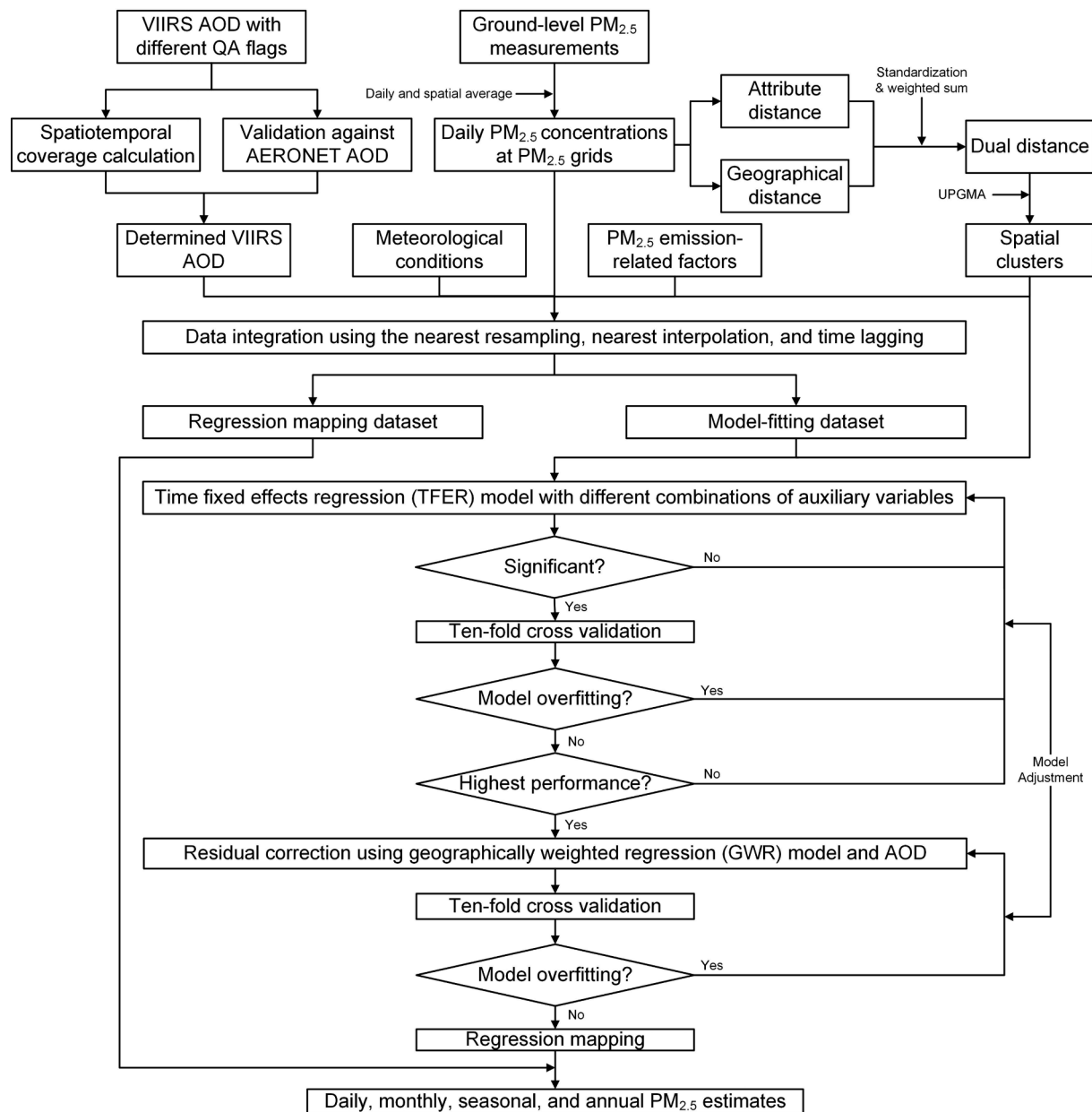


Fig. 1. Workflow for estimating the ground-level  $PM_{2.5}$  concentrations in this study.

For meteorological conditions, we used both the aerological and surface meteorological data. The former mainly contained the planetary boundary layer height (PBLH) and average relative humidity in the nine bottom tropospheric layers (1000 hPa to 800 hPa with a step length of 25-hPa increments) (RH\_PBL). The PBLH and RH\_PBL data, based on the Modern-Era Retrospective Analysis for Research and Application (Rienecker et al., 2011), were from the Goddard Earth Sciences Data and Information Services Center (code: tavgl\_2d\_flux\_Nx, inst3\_3d\_asm\_Cp) (<https://disc.gsfc.nasa.gov/daac-bin/FTPSubset.pl>), and their temporal and spatial resolutions were 1-h,  $0.5^\circ \times 0.67^\circ$  and 3-h,  $1.25^\circ \times 1.25^\circ$ , respectively. We averaged the PBLH and RH\_PBL data from 12:30 to 14:30 and 11:00 to 14:00 local time, respectively, to correspond to the VIIRS overpass time. The surface meteorological data, including the daily averaged temperature (TEMP), daily averaged surface relative humidity (SRH), daily precipitation (PRCP), daily averaged wind speed (WS), and daily maximum wind speed with the corresponding wind direction, came from the China Meteorological Data Sharing System (<http://data.cma.cn/>). We derived four wind vectors

(EWS, SWS, WWS, and NWS) from the daily maximum wind speed and its wind direction (Wu et al., 2016; Yao et al., 2018). Except for TEMP and SRH, we used the remaining variables of both the previous day and the current day to adjust the  $PM_{2.5}$ -AOD relationship of the current day.

#### 2.4. Data integration

We collected the ground-level  $PM_{2.5}$  monitoring, VIIRS EDR AOD, and auxiliary data covering China in 2014 for our modeling demonstration. We created a  $6\text{ km} \times 6\text{ km}$  grid over China to integrate the multi-source data. We averaged those daily averaged  $PM_{2.5}$  concentrations whose monitoring sites fell within a common grid and assigned the average value to that grid. We resampled the VIIRS EDR AOD,  $NO_2$ , NDVI, EVI, PBLH, and RH\_PBL data to the  $6\text{ km} \times 6\text{ km}$  grid and interpolated the surface meteorological data to the  $6\text{ km} \times 6\text{ km}$  grid using the nearest-neighbor method. Both the model-fitting and regression mapping datasets were obtained after data integration.

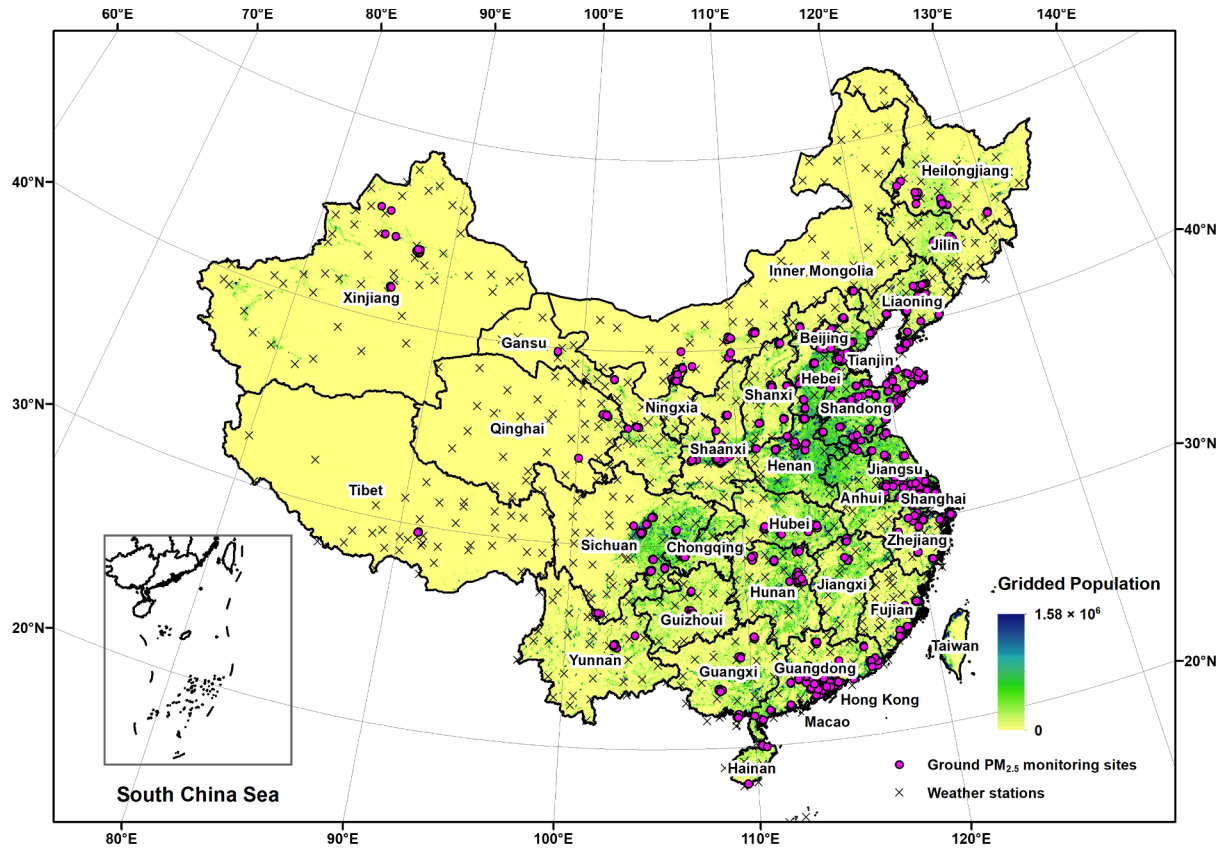


Fig. 2. Spatial distribution of ground PM<sub>2.5</sub> monitoring sites and weather stations. Note that most of the ground PM<sub>2.5</sub> monitoring sites are located in the central and eastern parts of China and many of them in this map are overlaid because of their proximity to each other.

## 2.5. Spatial division

Prior to developing the model, we assumed that those spatially proximate areas with similar PM<sub>2.5</sub> concentrations and PM<sub>2.5</sub> variations should have similar influencing factors on PM<sub>2.5</sub>-AOD relationships and thus could share identical model structures. In addition, we used the Moran's I (a statistic for testing the spatial autocorrelation and its degree (Moran, 1950)) to test the spatial autocorrelation of the PM<sub>2.5</sub> concentrations, and found that significant positive spatial autocorrelations existed in all days and that more than 95% of the days reported a Moran's I of greater than 0.4. This means that spatially proximate areas can always have similar PM<sub>2.5</sub> concentrations. Therefore, we only need to define a dual distance that can reflect the geographical distance and the attribute distance of the PM<sub>2.5</sub> variations. We can then apply a clustering algorithm to that dual distance to divide the whole geographical area of China into multiple spatial clusters, each of which shared an identical model structure when building quantitative PM<sub>2.5</sub>-AOD relationships. The calculation of the geographical distance and the attribute distance of the PM<sub>2.5</sub> variations were among those grids that had daily averaged PM<sub>2.5</sub> concentrations (hereafter referred to as PM<sub>2.5</sub> grids).

$$DG_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \quad (1)$$

where  $DG_{ij}$  is the geographical distance between grid  $i$  and  $j$ , and  $(X_i, Y_i)$  and  $(X_j, Y_j)$  are the coordinates of grid  $i$  and  $j$  under a projected coordinate system.

$$DA_{ij} = 1 - R_{ij} \quad (2)$$

where  $DA_{ij}$  is the attribute distance of PM<sub>2.5</sub> variations between grid  $i$  and  $j$ , and  $R_{ij}$  is the correlation coefficient between the time series of PM<sub>2.5</sub> concentrations of grid  $i$  and  $j$ .

$$DD_{ij} = w_1 \times \widetilde{DA}_{ij} + w_2 \times \widetilde{DG}_{ij} \quad (3)$$

where  $DD_{ij}$  is the dual distance between grid  $i$  and  $j$ ,  $\widetilde{DA}_{ij}$  and  $\widetilde{DG}_{ij}$  are the geographical distance and the attribute distance of the PM<sub>2.5</sub> variations normalized by their ranges, respectively, and  $w_1$  and  $w_2$  are the weightings, and their sum equals one. Their relationship will influence the spatial division result. More specifically, if  $w_1 > w_2$ , the final spatial clusters would likely be more spatially adjacent but share fewer common characteristics from the PM<sub>2.5</sub> variations; the situation would be opposite if  $w_1 < w_2$ . In this study, we assigned 0.5 to both, meaning that the geographical distance and the attribute distance of the PM<sub>2.5</sub> variations were equally important, in order to obtain an intermediate reasonable spatial division result.

While a series of cluster algorithms are adoptable, we employed the unweighted pair-group method with arithmetic means (UPGMA), taking advantage of its simplicity and reliability (Liu et al., 2018). The clustering process of the UPGMA algorithm was as follows. We first combined those two PM<sub>2.5</sub> grids that had the minimal dual distance. We then updated the dual distances between the combined PM<sub>2.5</sub> grid and the remaining PM<sub>2.5</sub> grids. The new dual distance was the unweighted, averaged dual distance of the original two PM<sub>2.5</sub> grids and the remaining PM<sub>2.5</sub> grids. Next, we repeated the combining and updating process until all of the PM<sub>2.5</sub> grids were combined. Subsequently, we selected a threshold and stipulated that only those PM<sub>2.5</sub> grids whose dual distances or updated dual distances were less than that threshold could be combined. Finally, we assigned the spatial clustering results of the PM<sub>2.5</sub> grids to their corresponding Thiessen polygons (Thiessen, 1911), which were further merged to generate the final spatial clusters. Ideally, all Thiessen polygons falling within the same spatial cluster should be spatially adjacent. Thus, we merged those polygons that were not spatially adjacent to other polygons in the same spatial cluster, or if they were standing alone, as a singleton spatial cluster, to which



adjacent Thiessen polygons had the minimum averaged original dual distance with them. The value of the threshold was pivotal. The number of spatial clusters would be very small if the threshold was set too high. Consequently, their areas would be very large, hiding the spatial heterogeneity of the model structure. The number of spatial clusters would be very large if the threshold was set too low. As a result, their areas and the size of the model-fitting dataset would be very small, decreasing the reliability of the regression analysis.

## 2.6. Model fitting

After the spatial division was completed, we determined the best model structure and conducted the first-stage TFER modeling for each spatial cluster. The TFER model stems from panel data regression models. By incorporating a series of temporal dummy variables, which were set as daily intercepts in this study, into the traditional multiple linear regression model, we simultaneously established the space-invariant temporal impact and the net impact of the predictors on  $PM_{2.5}$  concentrations, and thus, captured the temporal heterogeneity of the relationship between  $PM_{2.5}$  and AOD for each spatial cluster. We first constructed all possible TFER models using the ground-level  $PM_{2.5}$  concentrations as the dependent variable, VIIRS EDR AOD as the main predictor, and different combinations of auxiliary data as the assistant predictors. We then sorted those models that had simultaneously passed the F-test of the regression equation and the t-test of the predictors according to the number of significant daily intercepts in descending order. Because multiple models with an identical number of significant daily intercepts existed, we further selected a threshold for the number of significant daily intercepts, above which  $\geq 10$  models emerged, and they were regarded as candidate models. We finally determined the best model that had the highest adjusted coefficient of determination ( $R^2$ ) from the candidate models. As a result, our best model could have as high an adjusted  $R^2$  and as many significant daily intercepts as possible. An example illustrating the main process has been provided in Table S1. All of the statistical tests were at the significance level  $\alpha = 0.05$ , but we also decreased the significance level in some cases where no model emerged at  $\alpha = 0.05$  (i.e. decreasing  $\alpha$  to 0.4 for AOD t-test for Clusters 4, 9, and 10 and decreasing  $\alpha$  to 0.1 for t-tests of other predictors and daily intercepts for Cluster 10). The first-stage TFER model can be expressed as follows:

$$PM_{2.5gd} = \lambda_d + \beta_{AOD}AOD_{gd} + \beta_{PBLH}PBLH_{gd\_c} + \beta_{RH\_PBL}RH\_PBL_{gd\_c} + \beta_{TEMP}TEMP_{gd\_c} + \beta_{SRH}SRH_{gd\_c} + \beta_{PRCP}PRCP_{gd\_c} + \beta_{PRCP\_Lag}PRCP\_Lag_{gd\_c} + \beta_{WS}WS_{gd\_c} + \beta_{WS\_Lag}WS\_Lag_{gd\_c} + \beta_{EWS}EWS_{gd\_c} + \beta_{SWS}SWS_{gd\_c} + \beta_{WWS}WWS_{gd\_c} + \beta_{NWS}NWS_{gd\_c} + \beta_{EWS\_Lag}EWS\_Lag_{gd\_c} + \beta_{SWS\_Lag}SWS\_Lag_{gd\_c} + \beta_{WWS\_Lag}WWS\_Lag_{gd\_c} + \beta_{NWS\_Lag}NWS\_Lag_{gd\_c} + \beta_{NDVI}NDVI_{gd\_c} + \beta_{EVI}EVI_{gd\_c} + \beta_{NO_2\_Lag}NO_2\_Lag_{gd\_c} + \varepsilon_{gd} \quad (4)$$

where  $PM_{2.5gd}$ ,  $AOD_{gd}$ ,  $PBLH_{gd\_c}$ ,  $RH\_PBL_{gd\_c}$ ,  $TEMP_{gd\_c}$ ,  $SRH_{gd\_c}$ ,  $PRCP_{gd\_c}$ ,  $WS_{gd\_c}$ ,  $EWS_{gd\_c}$ ,  $SWS_{gd\_c}$ ,  $WWS_{gd\_c}$ ,  $NWS_{gd\_c}$ ,  $NDVI_{gd\_c}$ , and  $EVI_{gd\_c}$  are the daily averaged  $PM_{2.5}$  concentrations, AOD, PBLH, RH\_PBL, TEMP, SRH, PRCP, WS, EWS, SWS, WWS, NWS, NDVI, and EVI at grid g during day d, respectively;  $PRCP\_Lag_{gd\_c}$ ,  $WS\_Lag_{gd\_c}$ ,  $EWS\_Lag_{gd\_c}$ ,  $SWS\_Lag_{gd\_c}$ ,  $WWS\_Lag_{gd\_c}$ ,  $NWS\_Lag_{gd\_c}$ , and  $NO_2\_Lag_{gd\_c}$  are the PRCP, WS, EWS, SWS, WWS, NWS, and  $NO_2$  at grid g during day d-1, respectively;  $\beta_{AOD}$ ,  $\beta_{PBLH}$ ,  $\beta_{RH\_PBL}$ ,  $\beta_{TEMP}$ ,  $\beta_{SRH}$ ,  $\beta_{PRCP}$ ,  $\beta_{PRCP\_Lag}$ ,  $\beta_{WS}$ ,  $\beta_{WS\_Lag}$ ,  $\beta_{EWS}$ ,  $\beta_{SWS}$ ,  $\beta_{WWS}$ ,  $\beta_{NWS}$ ,  $\beta_{EWS\_Lag}$ ,  $\beta_{SWS\_Lag}$ ,  $\beta_{WWS\_Lag}$ ,  $\beta_{NWS\_Lag}$ ,  $\beta_{NDVI}$ ,  $\beta_{EVI}$ , and  $\beta_{NO_2\_Lag}$  are coefficients for AOD, PBLH, RH\_PBL, TEMP, SRH, PRCP, PRCP\_Lag, WS, WS\_Lag, EWS, SWS, WWS, NWS, EWS\_Lag, SWS\_Lag, WWS\_Lag, NWS\_Lag, NDVI, EVI, and  $NO_2$ , respectively;  $\lambda_d$  is the intercept of day d;  $\varepsilon_{gd}$  indicates that the

variable was incorporated in spatial cluster c; and  $\varepsilon_{gd}$  is the error term at grid g during day d.

We did not let NDVI and EVI appear in the same model to avoid potential multi-collinearity. Since RH\_PBL has poor spatiotemporal coverage over western China (Fig. S5), we did not consider it when we constructed the TFER models for those spatial clusters located in the west.

We calibrated the first-stage TFER model by applying the model-fitting dataset in Eq. (4). The residuals were obtained accordingly. In the second stage, we built a series of GWR models using those residuals as the dependent variable and VIIRS EDR AOD as the independent variable. By adding the estimated residuals from the second-stage GWR model to the  $PM_{2.5}$  predictions from the first-stage TFER model, we eventually obtained the final  $PM_{2.5}$  predictions with most spatiotemporal heterogeneities captured. The second-stage GWR modeling was conducted daily for each spatial cluster and can be expressed as follows:

$$PM_{2.5\_resi_{gd}} = \beta_{0,gd} + \beta_{AOD,gd}AOD_{gd} + \varepsilon_{gd} \quad (5)$$

where  $PM_{2.5\_resi_{gd}}$  denotes the residual from the first stage at grid g during day d;  $AOD_{gd}$  is VIIRS EDR AOD at grid d during day d;  $\beta_{0,gd}$  and  $\beta_{AOD,gd}$  are the location-dependent intercept and slope, respectively; and  $\varepsilon_{gd}$  is the error term at grid g during day d. For convenience, abbreviations and full names of all the variables appearing in Eqs. (5) and (6) have been provided in Table S2.

We used the weighted least squares method to calibrate the coefficients of the GWR model. The weight was a kernel function value of the distance between regression and data points. According to previous studies, the parameter estimation process is very sensitive to the bandwidth, to an extent where the kernel function plays a role (Tan, 2007). In this study, we used the Gauss kernel function and calculated the best adaptive bandwidth for each regression point by minimizing the CV value since the ground  $PM_{2.5}$  monitoring sites employed in this study were unevenly distributed spatially.

While the minimum number of the model-fitting data records required per day is two to fit an intercept and a slope, previous studies have shown that the practice of directly adding GWR models could cause model overfitting (i.e. the model performs much better in a model-fitting dataset than a non-model-fitting dataset) if that number is too small (Hu et al., 2014; Wu et al., 2016). Therefore, we conducted a sensitivity analysis to identify the best threshold for the minimal number of model-fitting data records per day for all spatial clusters. This threshold should have the characteristic of minimizing the overfitting degree at the cost of losing a minimal number of samples.

## 2.7. Model validation

We performed a regression of the observed and predicted  $PM_{2.5}$  concentrations from the model fitting to obtain the  $R^2$ , mean prediction error (MPE), and root mean squared error (RMSE) and assessed the model performance on the model-fitting dataset. We conducted a ten-fold CV to assess the model performance on the non-model-fitting dataset in three steps. First, we randomly split the model-fitting dataset into ten subsets with approximately 10% of the model-fitting dataset in each subset. Second, we withheld one subset, performed model fitting with the remaining nine subsets, and used the fitted model to predict the withheld subset. We repeated this process ten times to ensure that all the subsets were predicted. Third, we performed a regression of the observed  $PM_{2.5}$  and predicted  $PM_{2.5}$  concentrations from the ten-fold CV and obtained another set of  $R^2$ , MPE, and RMSE. According to the process of the ten-fold CV, the observed  $PM_{2.5}$  concentrations were not from the model-fitting dataset but the non-model-fitting dataset; thus, we were able to assess the model performance on the non-model-fitting dataset. Meanwhile, we could assess the model overfitting degree by comparing the  $R^2$ , MPE, and RMSE from the model fitting and ten-fold CV.

Theoretically, residuals from a reliable statistical model should not correlate sequentially. As for spatial data, residuals should not present spatial autocorrelation. Therefore, we calculated the Moran's I of the residuals from the first-stage and overall model, respectively, to assess the reliability of the model.

## 2.8. $PM_{2.5}$ prediction

While we could readily obtain the daily  $PM_{2.5}$  predictions at locations where  $PM_{2.5}$  observations were unavailable by applying the overall model to the regression mapping dataset, the predictions may be abnormally low or high. To overcome this defect, we reset the abnormally low and high predictions to 60% and 120% of the  $PM_{2.5}$  observations. Previous land use regression (LUR) models and satellite remote sensing retrieval models have proven that this practice is reasonable and effective (Wu et al., 2015, 2016). We obtained the monthly, seasonal, and annual  $PM_{2.5}$  estimates by averaging the daily  $PM_{2.5}$  estimates for each month, season, and the whole year. We then analyzed the spatiotemporal characteristics of the  $PM_{2.5}$  concentrations based on these estimates.

## 3. Results

### 3.1. Spatial division results

Fig. 3 illustrates the relationship between the threshold for the dual distance and the number of spatial clusters. As mentioned before, the threshold should not be set too high or too low to avoid having too many or too few spatial clusters. We thus decided to find a threshold between 0.2 and 0.3. We found 0.25 was a good threshold, below and above which the number of spatial clusters changed quickly and slowly, respectively. Although 0.25 may not be a perfect value, we think it is good enough, and our model fitting and validation results also proved it. We obtained 13 spatial clusters (Fig. 4) by using a threshold of 0.25 for dual distances and merging the isolated spatial clusters. Among the 13 spatial clusters, some of them (e.g. Cluster 3) had boundaries similar to those of natural basins, while others (e.g. Clusters 6 and 11) had boundaries similar to those of urban agglomerations or provinces. Therefore, both natural and anthropogenic factors influence  $PM_{2.5}$  concentrations. The practice of spatial division based on the inherent characteristics of  $PM_{2.5}$  concentrations could simultaneously reflect these two categories of influencing factors.

### 3.2. Model fitting results

Table 1 shows the coefficient estimates of the first-stage TFER model. The sensitivity analysis determined that the minimum number of model-fitting data records should be greater than six (see Section 3.3). We did not use this threshold for Clusters 4, 9, 10, 13, and 14 because the total number of model-fitting data records of these spatial clusters was not abundant. Although at least three model-fitting data records ( $N > 2$ ) were set to enable the ten-fold CV, the total number of model-fitting data records for Clusters 13 and 14 was still insufficient. Therefore, we decided not to predict the ground-level  $PM_{2.5}$  concentrations in Clusters 13 and 14, but in Clusters 4, 9, and 10, using only the first-stage TFER model.

Fig. 4 and Table 1 illustrate how  $PM_{2.5}$  emission-related factors and meteorological conditions spatially influenced the  $PM_{2.5}$ -AOD relationship. PBLH mainly appeared in the Middle-Upper Yangtze River plain and coastal regions with values approaching zero, indicating a weak role in the  $PM_{2.5}$ -AOD relationships. RH\_PBL mainly appeared in coastal regions and Inner Mongolia with positive or negative signs. Theoretically, high RH\_PBL increases the size and light extinction efficiencies of particles, such as ammonium sulfate and ammonium nitrate, leading to high  $PM_{2.5}$  concentrations; however,  $PM_{2.5}$  measurements take only dry particle mass into account under controlled relative humidity conditions ( $\approx 40\%$ ) (Hu et al., 2013; Liu et al., 2005). Therefore, in Clusters 1 and 7, where the general relative humidity level is always low, the same AOD value under a higher RH\_PBL level means a lower  $PM_{2.5}$  concentration, whereas the situation is the opposite in Cluster 2, where the general relative humidity level is always high. TEMP appeared in most spatial clusters with either positive or negative signs. This could be attributed to the varying relationship between temperature and the generation of secondary particles near the surface, such as sulfate and nitrate. These secondary particles are the precursors of  $PM_{2.5}$ , and their concentrations can either increase or decrease when the temperature becomes higher depending on regions and seasons (Liu et al., 2007; Megaritis et al., 2013; Tai et al., 2010; Yang et al., 2017). SRH appeared in three clusters located in Central China. It was positively correlated with  $PM_{2.5}$  concentrations because high SRH increases water-insoluble organic carbon to elemental carbon ratios, presumably suggesting the formation of more secondary organic aerosols (Cheng et al., 2015). PRCP played a positive role in reducing  $PM_{2.5}$  concentrations due to its washing role. Nevertheless, it only appeared in Clusters 7 and 10, indicating that precipitation is an effective variable

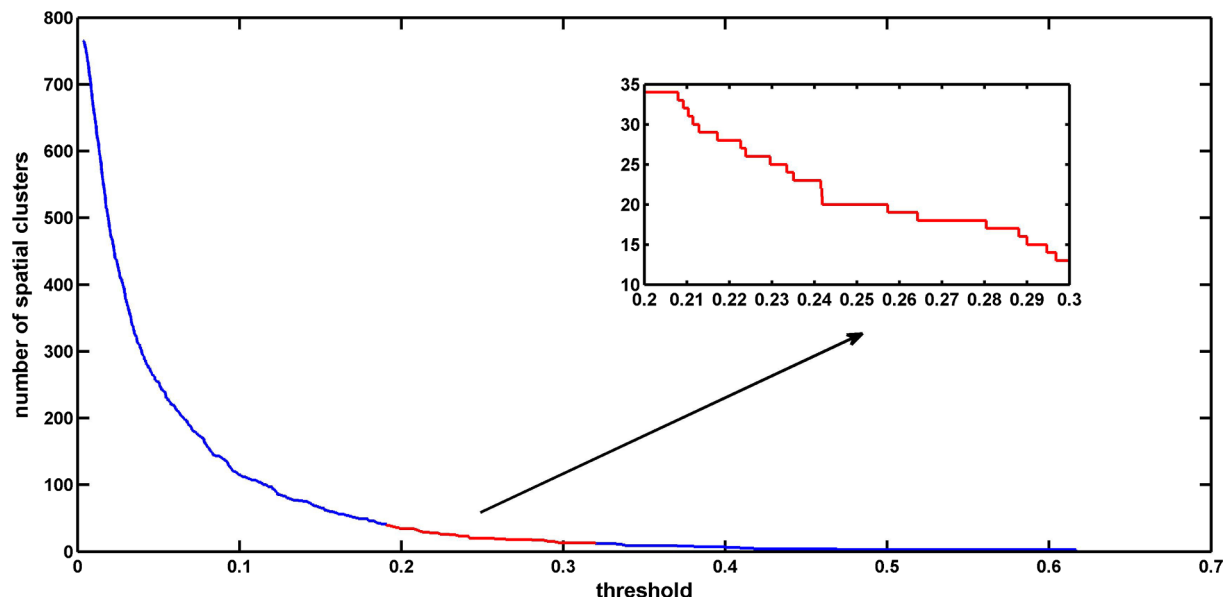


Fig. 3. The relationship between the threshold for the dual distance and the number of spatial clusters.

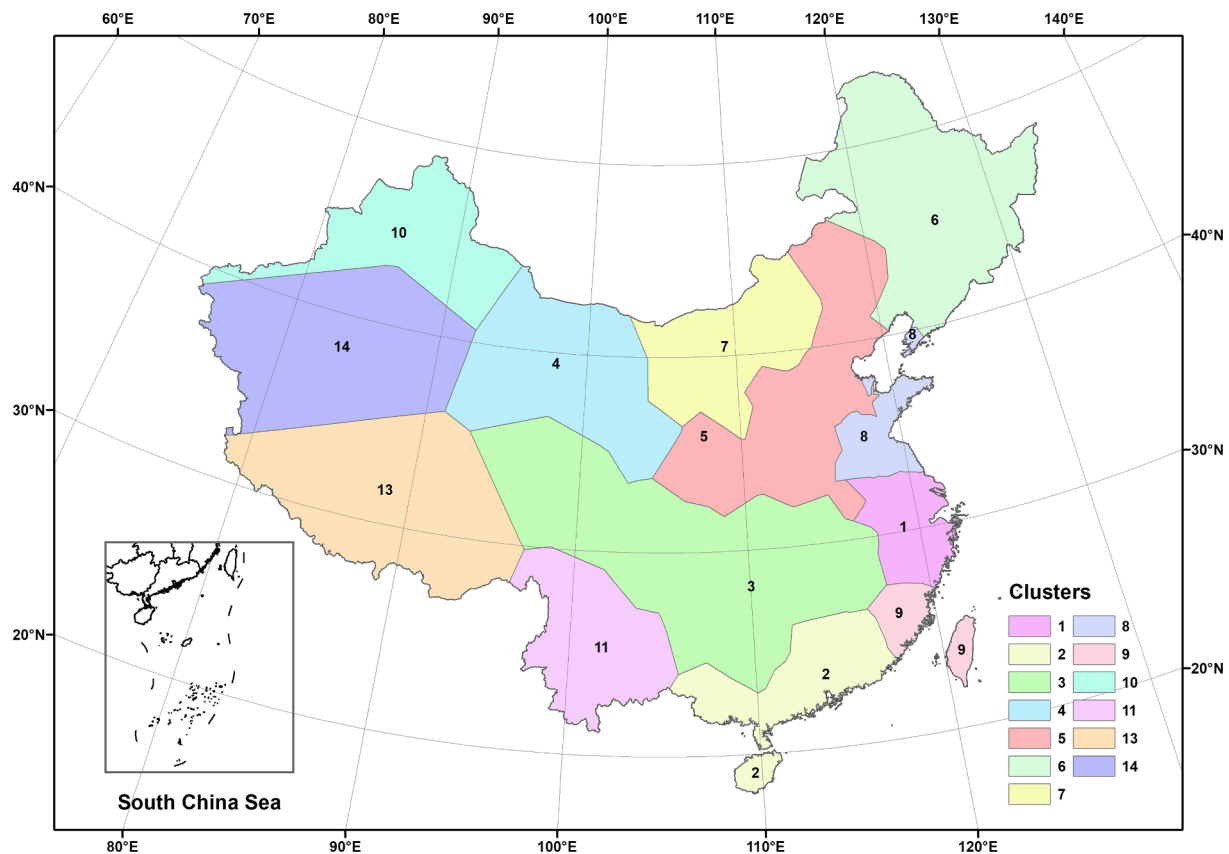


Fig. 4. Spatial division results for developing  $PM_{2.5}$  satellite remote sensing retrieval models.

where it is limited. In addition, it has no significant time lag because PRCP\_Lag did not appear in any spatial clusters. WS and WS\_Lag appeared with both positive and negative signs, indicating different impacts of wind speed on  $PM_{2.5}$ -AOD relationships in different spatial clusters. This can be attributed to the fact that the ground-level  $PM_{2.5}$  concentrations are more sensitive to wind speed than AOD (Zheng et al., 2017), which means the ground  $PM_{2.5}$  concentrations will be higher or lower if more air pollutants are transported in or out by wind speed under the same AOD level. We further analyzed this using the wind direction data. EWS and EWS\_Lag appeared in multiple spatial clusters with most of them having negative signs, indicating that easterly winds could always reduce  $PM_{2.5}$  concentrations. SWS and SWS\_Lag also appeared in multiple spatial clusters with non-lagged forms having negative signs while lagged forms always having positive signs, indicating that the effects of southerly winds on  $PM_{2.5}$  concentrations were subject to the time lag. The signs of WWS and WWS\_Lag were negative in the western spatial clusters, such as Clusters 3, 4, and 11, whereas the signs of WWS and WWS\_Lag were positive in the central and eastern spatial clusters, such as Clusters 5 and 6. This implies that westerly winds may bring pollutants from the west to the east. NWS and NWS\_Lag appeared in several spatial clusters, but their regularity was not evident. NDVI mainly appeared in southern China with both positive and negative signs. Since vegetation generates over 90% of volatile organic compounds (VOC) emissions on the global scale (Guenther et al., 1995), and these emissions can further form secondary aerosols, high NDVI would mean high  $PM_{2.5}$  concentrations. High NDVI could also mean a higher rate of particle deposition on leaves, thus reducing  $PM_{2.5}$  concentrations. In this study, most NDVI signs were reported as negative, indicating that the leaf deposition role has by far the greater impact on  $PM_{2.5}$ . EVI appeared only in Cluster 6, indicating that it is more effective than NDVI in northeastern China.  $NO_2$ \_Lag appeared in Clusters 1, 5, and 7 with positive signs, indicating that its lagged positive role in increasing the  $PM_{2.5}$  concentration occurred in

the North China Plain and coastal regions. Varying coefficients among different spatial clusters suggest that the first-stage TFER model captured the spatial heterogeneity of the  $PM_{2.5}$ -AOD relationship.

Fig. 5 illustrates the distribution of the daily intercepts of the first-stage TFER models in different clusters. The daily intercepts in winter and spring were always larger than those in summer and autumn. This indicates that the  $PM_{2.5}$  concentrations in winter and spring would be higher than the corresponding concentrations in summer and autumn under the same conditions, since the first-stage TFER model only had varying intercepts. These varying daily intercepts in the same spatial cluster also suggest that the first-stage TFER model captured the temporal heterogeneity of the  $PM_{2.5}$ -AOD relationship.

Fig. 6 illustrates the seasonally averaged AOD coefficients from the second-stage GWR model. Winter always had positive coefficients, correcting the underestimation of the first-stage TFER model. Spring had more negative coefficients than positive coefficients, correcting the overestimation of the first-stage TFER model. Summer and autumn had similar numbers of positive and negative coefficients. Spatially, high AOD coefficients were mainly distributed in the North China Plain and coastal areas. High  $PM_{2.5}$  concentrations in these areas may be underestimated in the first-stage TFER model, but the second-stage GWR model amended it. The seasonal and spatial distributions of the AOD coefficients indicate that the second-stage GWR model further captured the temporal and spatial heterogeneities of the  $PM_{2.5}$ -AOD relationship.

### 3.3. Model validation results

Fig. 7 illustrates how the minimal number of model-fitting data records per day influenced the model performance. While the total number of the samples decreased as the minimal number of model-fitting data records per day increased,  $R^2$ , MPE, RMSE, slope, and intercept of the first-stage TFER model present a slight change during both model fitting and ten-fold CV. The  $R^2$ , MPE, RMSE, slope, and

**Table 1**  
Coefficient estimates of the first-stage TFER model.

Cluster	N	Minimum number of N per day	Coefficients						
			AOD	PBLH	RH_PBL	TEMP	SRH	PRCP	PRCP_Lag
1	3097	N > 6	16.04	0.01	−27.75	0.06			
2	2354	N > 6	18.07	0.01	27.43				
3	2013	N > 6	23.24	0.01		0.14	0.54		
4	357	N > 2	6.45			−0.15			
5	7142	N > 6	21.81			0.22	0.66		
6	2753	N > 6	16.68			0.07			
7	1307	N > 6	12.13		−30.26	0.22	0.69	−0.06	
8	3421	N > 6	18.00	0.00		0.10			
9	262	N > 2	−2.38**						
10	460	N > 2	−5.91***			−0.44		−0.36	
11	901	N > 6	17.84			0.06			
13	57	N > 2	14.36	−0.10					
14	6	N > 2	−34.80	−0.04					

Cluster	N	Minimum number of N per day	Coefficients												
			WS	WS_Lag	EWS	SWS	WWS	NWS	EWS_Lag	SWS_Lag	WWS_Lag	NWS_Lag	NDVI	EVI	NO <sub>2</sub> _Lag
1	3097	N > 6		−0.18			0.07	0.10		0.13			−7.39		0.15
2	2354	N > 6		−0.17	−0.17			−0.10					−4.95		
3	2013	N > 6	−0.23										−23.17		
4	357	N > 2									−0.14				
5	7142	N > 6			−0.05	−0.08		−0.08		0.09	0.04	0.07			0.54
6	2753	N > 6		0.18	0.05		0.15		−0.09	−0.10	−0.12	−0.28		−8.44	
7	1307	N > 6	0.24							0.04			−12.49		1.13
8	3421	N > 6			−0.06	−0.07							−12.20		
9	262	N > 2						−0.08					−11.37		
10	460	N > 2													
11	901	N > 6					−0.15				−0.13		−41.84		
13	57	N > 2					−0.24						56.08		
14	6	N > 2									−0.54				

\* p = 0.361;

\*\* p = 0.494;

\*\*\* p = 0.315.

intercept of the overall model also exhibit a slight change during model fitting. During ten-fold CV, however, before the value of  $N > 6$ ,  $R^2$  increased rapidly, MPE and RMSE decreased rapidly, and the slope and intercept rapidly approached one and zero, respectively. After the value of  $N > 6$ , the speed at which these changes occurred decreased. This indicates that the difference between the performance of the overall model during model fitting and ten-fold CV decreased as the minimal number of model-fitting data records increased; i.e. the over-fitting degree of the model declined. We chose  $N > 6$  as the threshold because the changing speed before and after that value differed greatly, which means that  $N > 6$  can minimize the overfitting degree at the cost of losing a minimal number of samples.

Fig. 8 illustrates the model validation results when at least seven model-fitting data records existed per day. While the over-fitting degree of the first-stage TFER model was less than that of the overall model (e.g.  $R^2$  decreased 6.45% and 26.83% from model fitting to ten-fold CV for the first-stage TFER model and the overall model, respectively), the overall model performed better than the first-stage TFER model during both model fitting and ten-fold CV. Moreover, the overall model captured the  $PM_{2.5}$  concentrations that were larger than  $250 \mu g/m^3$ , which were not found in the first-stage TFER model. Therefore, the practice of adding the second-stage GWR model was necessary and meaningful. It not only improved the model accuracy but also enhanced the model capability of retrieving high  $PM_{2.5}$  concentrations. The satisfactory model performance also proves the validity of the assumption proposed in Section 2.5.

Fig. 9 illustrates the temporal distribution of the spatial auto-correlation test results for the daily residuals from the first-stage TFER model and the overall model. It demonstrates the necessity of adding the second-stage GWR model. Although the daily residuals on some

days still presented significant, though much weaker, spatial auto-correlation after adding the second-stage GWR model, the proportion of the days that the daily residuals were significantly spatially auto-correlated declined greatly during both model fitting and ten-fold CV. In short, adding the second-stage GWR model improved the model reliability.

Table 2 shows the model performance in each spatial cluster. Clusters 1, 2, 3, and 8 had relatively high CV  $R^2$  values that were larger than 0.6, followed by Clusters 5 and 6 with CV  $R^2$  values greater than or approaching 0.5. The CV  $R^2$  values of Clusters 7 and 11 were slightly lower because these two clusters are not characterized by strong urban-industrial conditions. The satellite linkage with the ground cannot be well guaranteed if the principal contribution to AOD does not come from the particles near the surface. Clusters 4, 9, and 10 only had the first-stage TFER model. Clusters 4 and 10 had a relatively low CV  $R^2$  while Cluster 9 had a relatively high CV  $R^2$ . The fluctuations in the  $R^2$  values may be attributed to insufficient model-fitting data records.

### 3.4. Predictions of $PM_{2.5}$ concentrations

The resulting daily  $PM_{2.5}$  estimates were not temporally or spatially complete due to missing VIIRS EDR AOD data. As a result, uncertainties emerged when deriving the monthly, seasonal, and annual  $PM_{2.5}$  estimates from the daily  $PM_{2.5}$  estimates. We calculated the correlation coefficients between the  $PM_{2.5}$  estimates and measurements at the ground-based  $PM_{2.5}$  monitoring sites and found that all of the correlation coefficients were larger than 0.5 (Fig. S6) at the monthly, seasonal, and annual scales. Therefore, the monthly, seasonal, and annual  $PM_{2.5}$  estimates could be used to reveal the spatiotemporal characteristics of the  $PM_{2.5}$  concentrations.



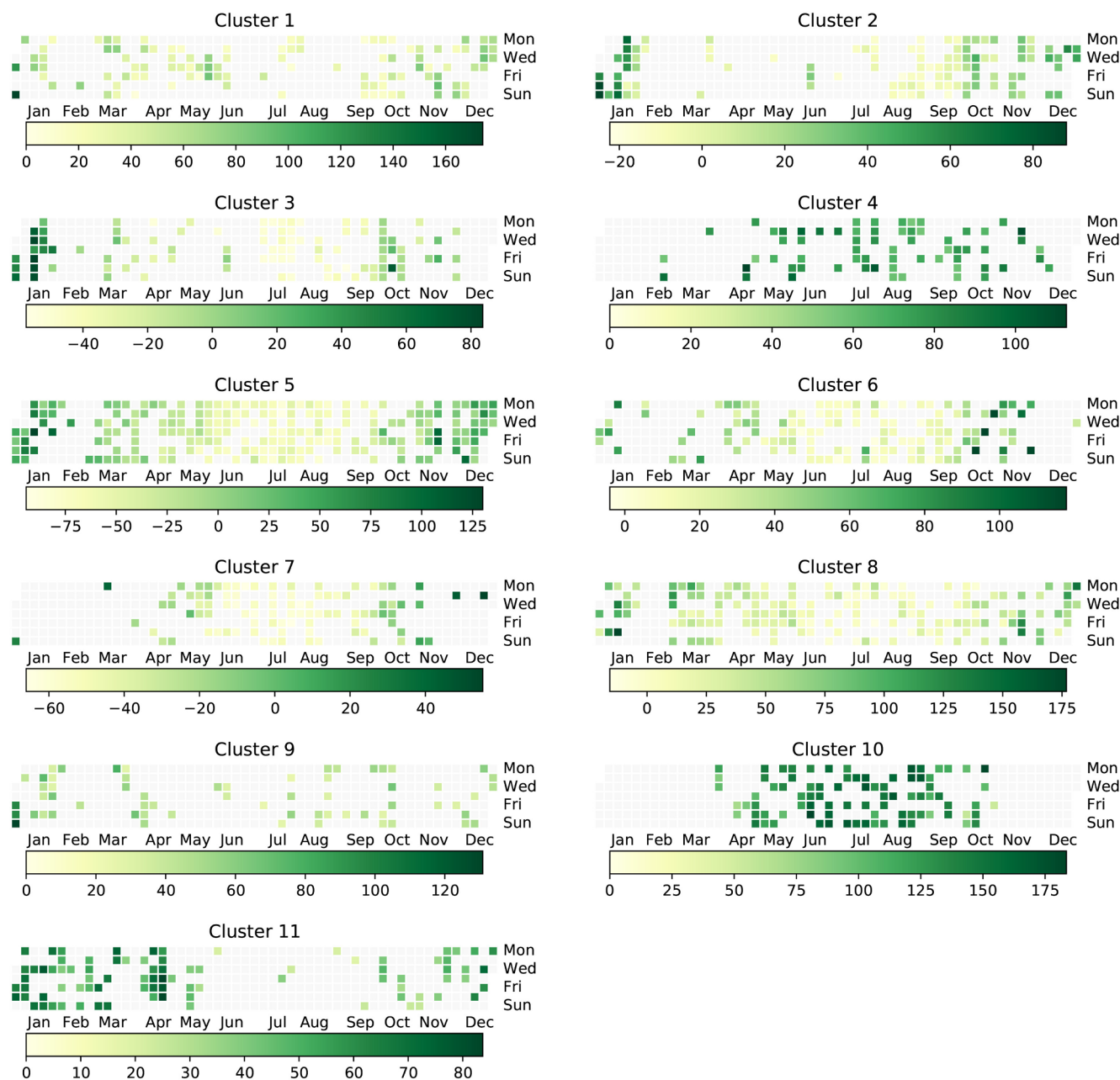


Fig. 5. Distribution of daily intercepts for the first-stage TFER models in different clusters. Note that most intercepts were significant at  $\alpha = 0.05$ , and thus their  $p$ -values were not shown.

Figs. 10 and S7 illustrate the spatial distribution of the seasonal, annual, and monthly  $PM_{2.5}$  estimates. Seasonally, the ground-level  $PM_{2.5}$  concentrations were high in winter, low in summer, and moderate in spring and autumn. Spatially, high  $PM_{2.5}$  concentrations mainly occurred to the east of Hu's line (Hu, 1935) and north of the Yangtze River, while the core polluted area was in the North China Plain. The core polluted area expanded from October to January and contracted from March to September. February observations were not conclusive due to the missing data.

#### 4. Discussion

##### 4.1. Comparison with the spatially structured fixed model

To further demonstrate the advantage of the spatially structured adaptive two-stage model (hereafter referred to as the adaptive model),

we developed a spatially structured fixed two-stage model (hereafter referred to as the fixed model). The fixed model considered all of China as a whole, tested all possible TFER models, and selected the one with the highest adjusted  $R^2$  value that passed the statistical tests. It then developed daily GWR models to correct the daily residuals, using AOD as the independent variable. The only difference between the adaptive model and the fixed model was that the former took advantage of the spatial clusters analyzed from  $PM_{2.5}$  measurements. The sensitivity analysis of the adaptive model showed that each spatial cluster should have  $N > 6$  model-fitting data records per day. Therefore, the minimum and maximum numbers of model-fitting data records per day for all of China should fall within [7, 77]. We used the lower and upper limit of this range to conduct the model fitting and ten-fold CV, respectively. The results are shown in Figs. S8 and S9. Comparisons between Figs. S8, S9 and 8 show that the adaptive model outperformed the fixed model during both model fitting and ten-fold CV for Stage I

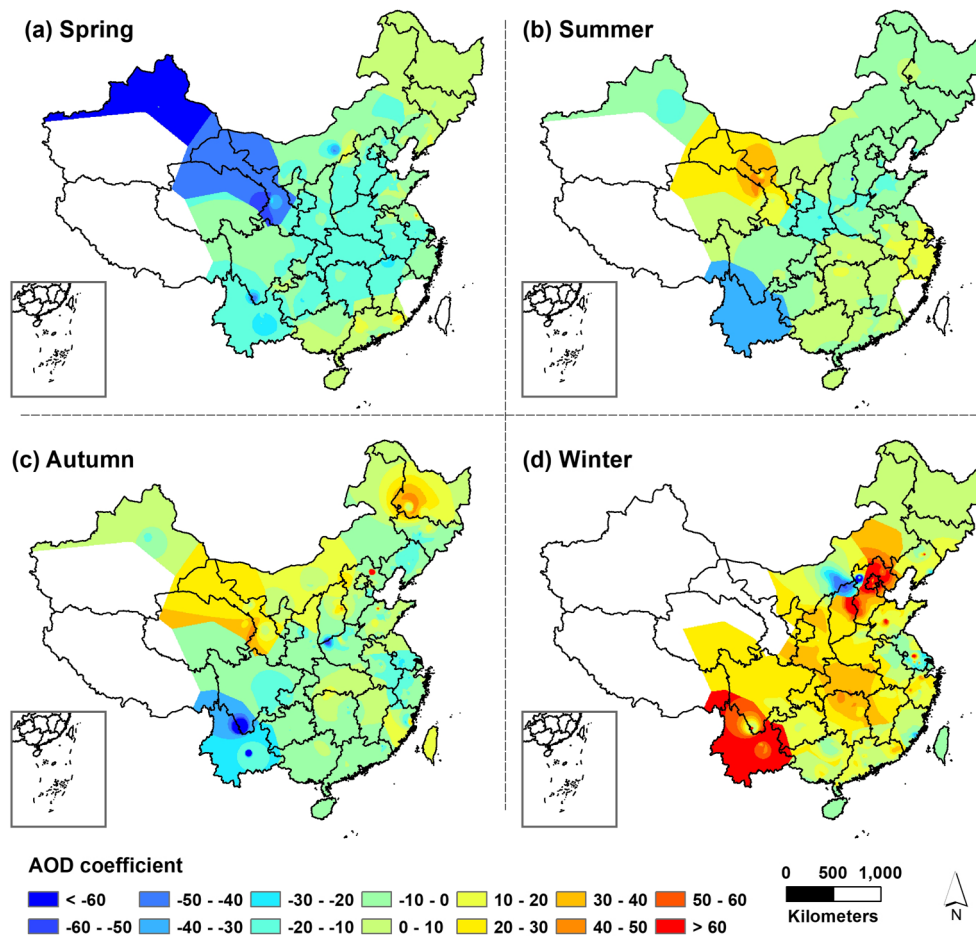


Fig. 6. Seasonally averaged AOD coefficients from the second-stage GWR model.

and model fitting for Stage II. Regarding the ten-fold CV of Stage II, while the  $R^2$  value of the fixed model was similar to that of the adaptive model when the fixed model required at least 77 model-fitting data records per day, the total number of model-fitting data records was smaller, and the MPE and RMSE values were larger, likely because of their different practices for requiring a minimum number of model-fitting data records per day. The number of model-fitting data records per day for China from the adaptive model should fall within [7, 77] because its requirement for the minimum number of model-fitting data records per day did not apply on every spatial cluster simultaneously. If a spatial cluster met the condition, it was included in the model; otherwise, it was excluded. Therefore, the total number of model-fitting data records of the adaptive model outweighed that of the fixed model when it required at least 77 model-fitting data records per day. In addition, the practice of requiring a minimum number of model-fitting data records per day for the adaptive model was more likely to produce spatially evenly distributed model-fitting data records than that of the fixed model. Therefore, smaller MPE and RMSE values of the adaptive model were obtained. In short, the development of the adaptive model was necessary.

The performance improvement from the second-stage GWR model of the adaptive model was worse than that of the fixed model because the model structure varied across the spatial clusters when the first-stage TFER models were constructed; thus, both the spatial and temporal heterogeneities of the  $PM_{2.5}$ -AOD relationships were captured by the adaptive model in Stage I. The GWR coefficient analysis in Section 3.2 shows that the adaptive model developed in this study could further capture the spatiotemporal heterogeneities in Stage II.

#### 4.2. Potential applications of this study

Potential applications of this study are based on the spatially structured adaptive two-stage model and resulting  $PM_{2.5}$  estimates. The model developed in this study can be applied to other satellite-retrieved AOD products and extended to other regions and periods. For example, using other products like MODIS AOD, we could develop the corresponding spatially structured adaptive two-stage models for the year 2014 simply by substituting VIIRS EDR AOD with data from these products. We could develop similar adaptive models for other countries or regions that share similar characteristics with China, such as the United States of America and Southeast Asia, using the model steps described in Section 2. If we assume that  $PM_{2.5}$ -AOD relationships do not change at all or change only slightly over the years, we could directly apply the spatially structured adaptive two-stage model developed in this study to China in other years. While this practice is valid for producing historical  $PM_{2.5}$  data for China, as the national ground-based  $PM_{2.5}$  monitoring network was not built until 2012, the uncertainties associated with it should be considered, especially for the earlier years.

The resulting  $PM_{2.5}$  estimates in this study had a higher spatial resolution ( $6\text{ km} \times 6\text{ km}$ ) than in previous studies that used the 10 km MODIS AOD or the 17.6 km MISR (Multiangle Imaging SpectroRadiometer) AOD. Van Donkelaar et al. (2016) estimated a long time-series of global  $PM_{2.5}$  estimates using the AOD products from MODIS, MISR, and SeaWiFS instruments and a combined geophysical-statistical method. We downloaded their results for 2014 ([http://fizz.phys.dal.ca/~atmos/martin/?page\\_id=140](http://fizz.phys.dal.ca/~atmos/martin/?page_id=140)) and replotted it with our results using the same color scale (Fig. S10). Fig. S10 shows that the  $PM_{2.5}$  values from our estimates are slightly higher than those from van

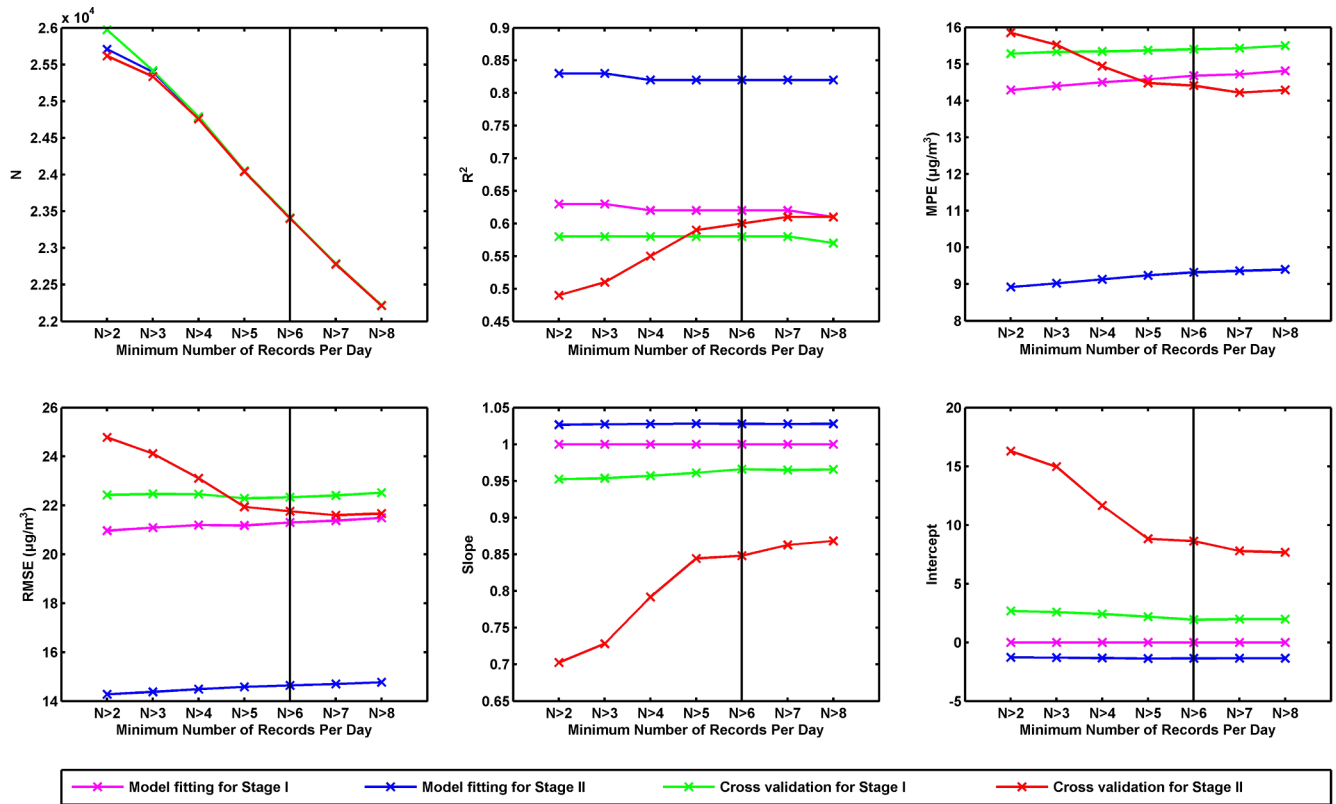


Fig. 7. The influence of the minimal number of model-fitting data records per day on the model performance.

Donkelaar et al. (2016) over the whole study domain. Fig. 10 also shows that the observed  $\text{PM}_{2.5}$  values are even higher than the  $\text{PM}_{2.5}$  estimates in our results. This, to some extent, indicates that our estimates are closer to the real  $\text{PM}_{2.5}$  pollution level. Nevertheless, the spatiotemporal coverage of our results was inferior, and thus more

efforts should be made to improve the spatiotemporal coverage of the original VIIRS EDR AOD. In short, the resulting  $\text{PM}_{2.5}$  estimates in this study will advance the epidemiological studies and pollution control policies in China by revealing more accurate spatial details.

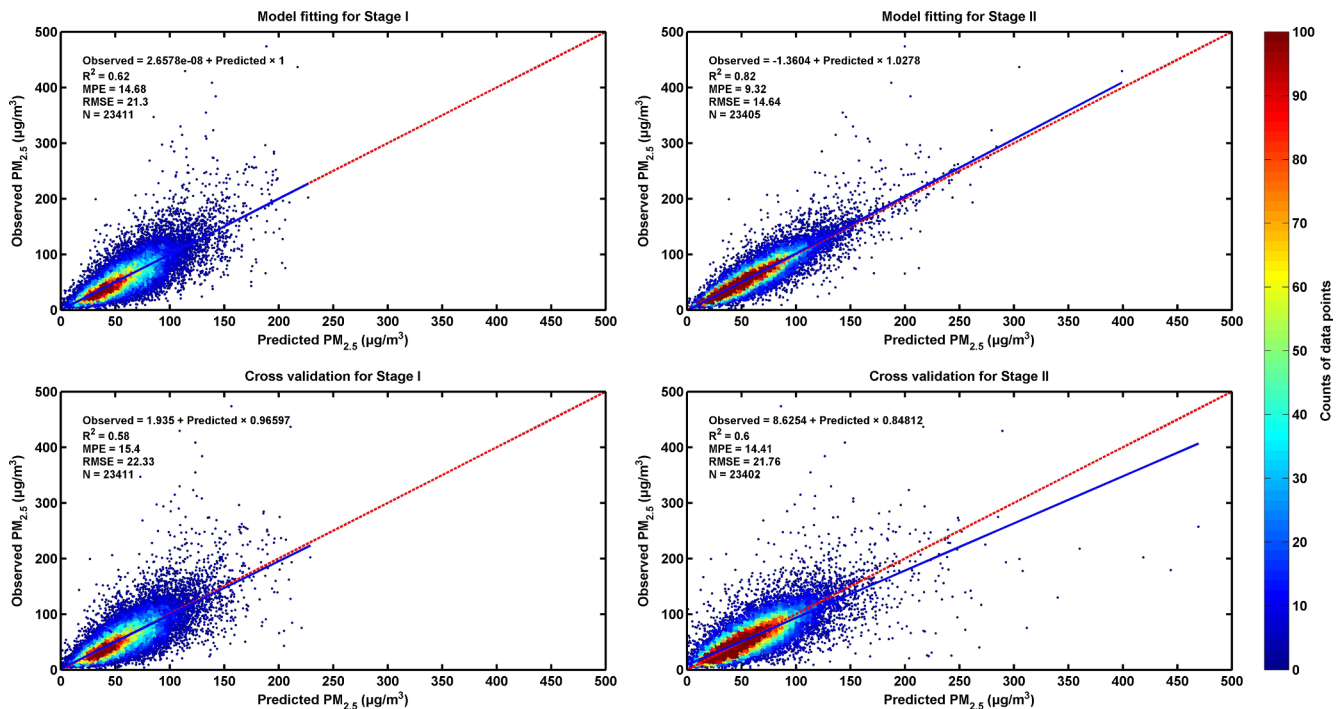


Fig. 8. Model validation results when greater than six model-fitting data records existed per day.

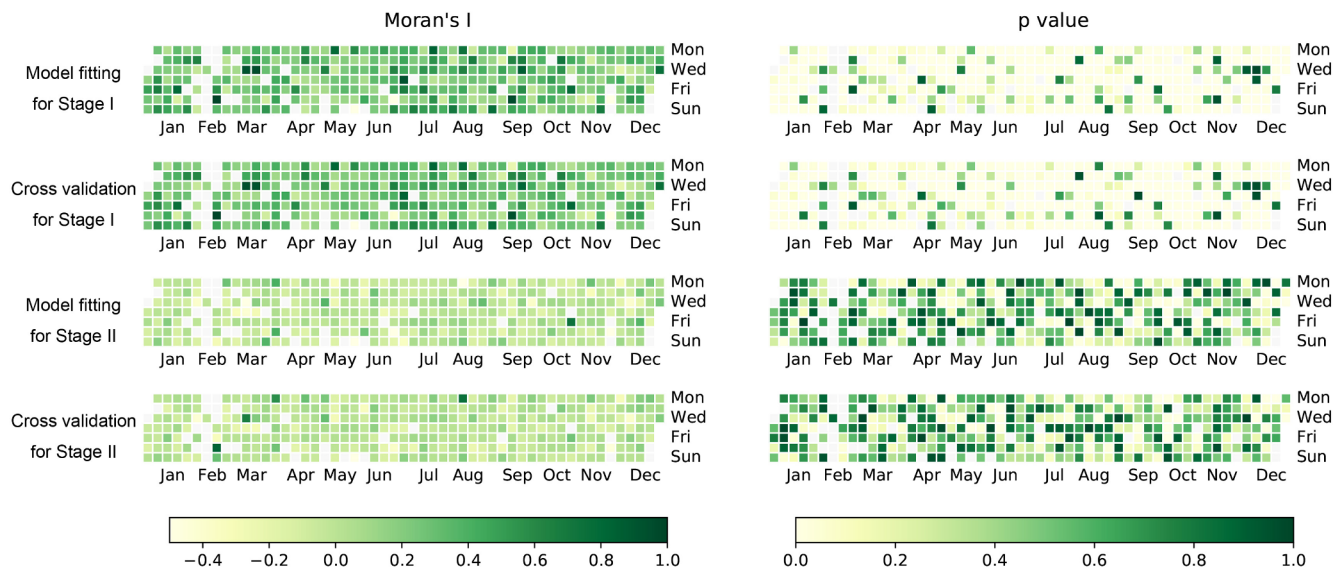


Fig. 9. Temporal distribution of spatial autocorrelation tests for daily residuals from the first-stage TFER model and the overall model.

#### 4.3. Limitations and prospects

Several limitations associated with this study are discussed here and should be addressed in the future. First, we did not retrieve  $PM_{2.5}$  concentrations over Clusters 13 and 14 due to the limited number of model-fitting data records. We could overcome this limitation by applying the spatially structured adaptive two-stage model on other satellite-retrieved AOD products or on an integrated product that has higher spatiotemporal coverage. A larger number of ground-level  $PM_{2.5}$  monitoring data records will also help solve this problem. Second, we noted that the CV  $R^2$  of some clusters were not satisfactory due to insufficient model-fitting data records even with both high- and medium-quality AOD. Another reason is that some spatial clusters, such as Cluster 7, lacked strong urban-industrial conditions, and thus the linear linkages between AOD and ground  $PM_{2.5}$  concentrations were weak. Nonparametric machine learning models usually involve less restrictive assumptions and have demonstrated great potential for estimating accurate  $PM_{2.5}$  concentrations from satellite-retrieved AOD (Hu et al., 2017; Li et al., 2017; Xiao et al., 2018; Zhang et al., 2018). Recognizing the natural match of machine learning models and remotely sensed big data (Mountrakis et al., 2018), we are more likely to improve the accuracy of retrieved  $PM_{2.5}$  concentrations by using machine learning models. Finally, due to the limitations of data and calculation resources,

not all useful auxiliary variables that have been proposed and proven effective in previous studies have been collected and examined, and some of the auxiliary variables are also not in their best form (e.g. it would be much better to decompose daily average wind speed using the main wind direction of the day, but instead, we used the maximum wind speed and corresponding wind direction). We firmly believe that the model structure for each spatial cluster will be more robust when more reasonable variables are included.

#### 5. Conclusion

In this study, we aimed to develop a spatially structured adaptive model to estimate ground-level  $PM_{2.5}$  concentrations at a 6-km spatial resolution in China. To our knowledge, this is one of the earliest studies that examined the performance of VIIRS AOD in estimating ground-level  $PM_{2.5}$  concentrations on the national scale and using the newly developed spatially structured adaptive two-stage model. The results are satisfactory despite some limitations. The main conclusions are as follows.

- i. The inherent characteristics of ground-level  $PM_{2.5}$  concentrations produced 13 spatial clusters using the UPGMA cluster algorithm. The boundaries of these 13 spatial clusters indicate that both

Table 2

Performance of the first-stage TFER model and the overall model in each cluster.

Cluster	N	Minimum number of N per day	The first-stage TFER model					The overall model						
			Model-fitting $R^2$	CV $R^2$	Model-fitting MPE	CV MPE	Model-fitting RMSE	CV RMSE	Model-fitting $R^2$	CV $R^2$	Model-fitting MPE	CV MPE	Model-fitting RMSE	CV RMSE
1	3097	N > 6	0.64	0.60	14.34	14.98	19.51	20.43	0.87	0.69	8.02	12.61	11.81	18.01
2	2354	N > 6	0.67	0.64	11.66	12.24	15.78	16.48	0.85	0.66	7.25	11.58	10.53	16.07
3	2013	N > 6	0.68	0.64	14.20	15.09	19.49	20.70	0.86	0.66	8.85	14.03	13.00	19.91
4	357	N > 2	0.45	0.23	12.67	15.26	18.51	21.89						
5	7142	N > 6	0.53	0.50	17.96	18.54	26.18	27.16	0.78	0.56	11.35	16.33	18.11	25.27
6	2753	N > 6	0.57	0.52	13.21	13.89	20.57	21.61	0.77	0.49	9.24	13.83	14.84	22.34
7	1307	N > 6	0.57	0.49	9.33	10.17	12.51	13.67	0.76	0.31	6.58	11.46	9.25	15.85
8	3421	N > 6	0.63	0.59	15.22	16.02	21.08	22.19	0.82	0.60	9.87	15.38	14.81	21.94
9	262	N > 2	0.86	0.75	5.44	7.32	7.58	10.21						
10	460	N > 2	0.45	0.19	12.77	14.96	17.12	20.72						
11	901	N > 6	0.62	0.54	9.81	10.74	13.20	14.47	0.81	0.41	6.28	11.69	9.27	16.28
13	57	N > 2	0.81	0.58	4.62	6.67	6.59	9.69						
14	6	N > 2	0.72	0.07	8.70	15.51	10.76	19.94						



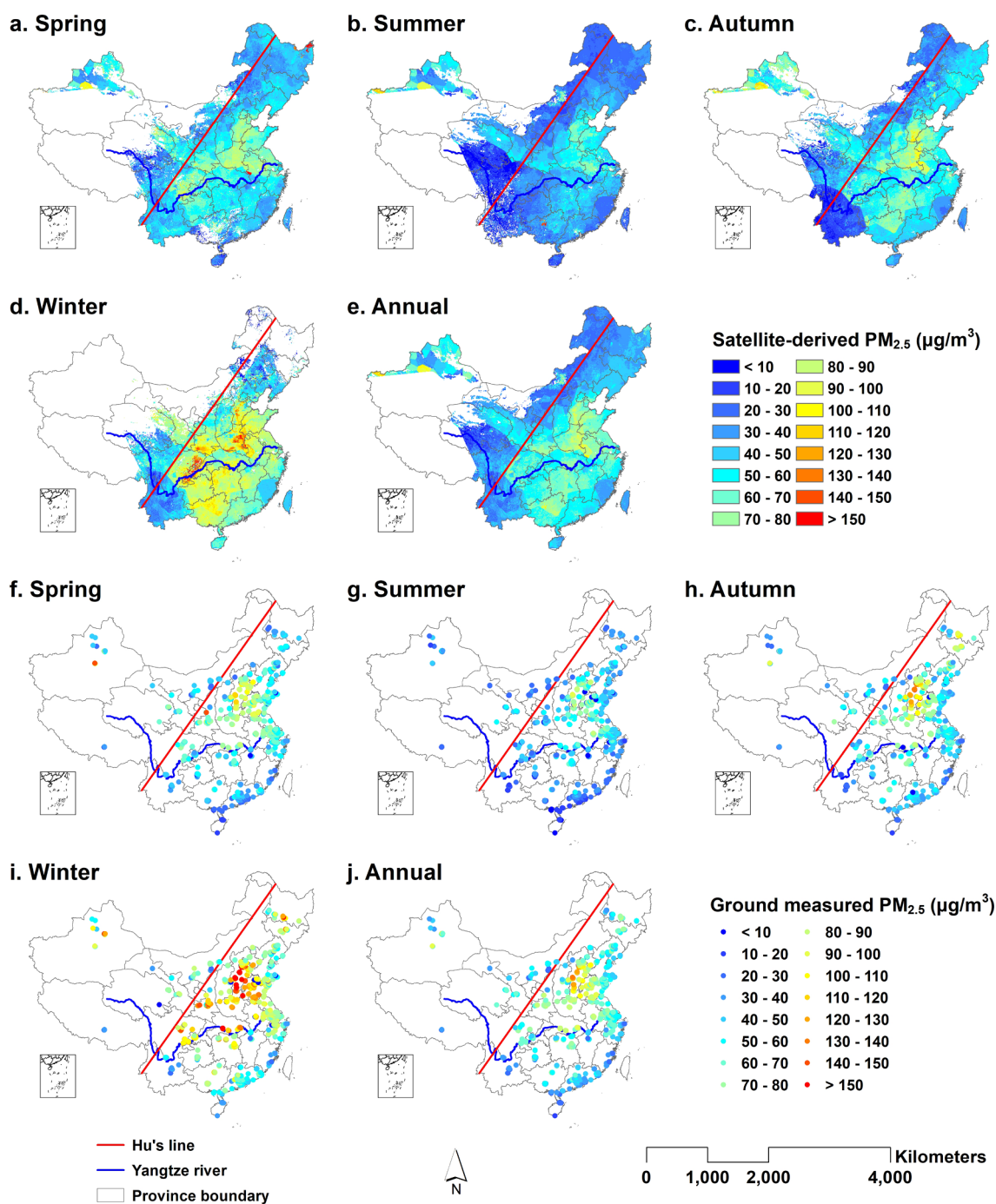


Fig. 10. Spatial distribution of seasonal and annual PM<sub>2.5</sub> estimates.

- natural and anthropogenic factors influence PM<sub>2.5</sub> concentrations.
- ii. "Seven" was a good threshold number of model-fitting data records for constructing the spatially structured adaptive two-stage model. The overall model fitting and ten-fold CV R<sup>2</sup> values were 0.82 and 0.60, respectively, when using this threshold.
  - iii. Correlation coefficients between the PM<sub>2.5</sub> estimates and observations all exceeded 0.5 at the monthly, seasonal, and annual scales. High PM<sub>2.5</sub> concentrations mainly occurred to the east of Hu's line and north of the Yangtze River in winter.

#### Acknowledgements

This study was financially supported by the Shenzhen Science and

Technology Innovation Committee (No. JCYJ20170412150910443), National Natural Science Foundation of China (No. 41471370), and China Scholarship Council (CSC). We thank anonymous reviewers and the editor for their constructive comments.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.isprsjprs.2019.03.011>.

#### References

- Bian, J.H., Li, A.N., Huang, C.Q., Zhang, R., Zhan, X.W., 2018. A self-adaptive approach for producing clear-sky composites from VIIRS surface reflectance datasets. *ISPRS J.*

- Photogram. Remote Sensing 144, 189–201.
- Boersma, K.F., Eskes, H.J., Dirksen, R.J., Van der, A.R.J., Veefkind, J.P., Stammes, P., Huijnen, V., Kleipool, Q.L., Snee, M., Claas, J., Leitao, J., Richter, A., Zhou, Y., Brunner, D., 2011. An improved tropospheric NO<sub>2</sub> column retrieval algorithm for the Ozone Monitoring Instrument. *Atmos. Meas. Tech.* 4, 1905–1928.
- Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C.A., Apte, J.S., Brauer, M., Cohen, A., Weichenhal, S., Coggins, J., Di, Q., Brunekreef, B., Frostad, J., Lim, S.S., Kan, H.D., Walker, G.D., Thurston, G.D., Hayes, R.B., Lim, C.C., Turner, M.C., Jerrett, M., Krewski, D., Gapstur, S.M., Diver, W.R., Ostro, B., Goldberg, D., Crouse, D.L., Martin, R.V., Peters, P., Pinault, L., Tjepkema, M., Donkelaar, A., Villeneuve, P.J., Miller, A.B., Yin, P., Zhou, M.G., Wang, L.J., Janssen, N.A.H., Marra, M., Atkinson, R.W., Tsang, H., Thach, Q., Cannon, J.B., Allen, R.T., Hart, J.E., Laden, F., Cesaroni, G., Forastiere, F., Weinmayr, G., Jaensch, A., Nagel, G., Concin, H., Spadaro, J.V., 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *P. Natl. Acad. Sci. USA* 115, 9592–9597.
- Cheng, Y., He, K.-B., Du, Z.-Y., Zheng, M., Duan, F.-K., Ma, Y.-L., 2015. Humidity plays an important role in the PM<sub>2.5</sub> pollution in Beijing. *Environ. Pollut.* 197, 68–75.
- Chu, Y., Liu, Y., Li, X., Liu, Z., Lu, H., Lu, Y., Mao, Z., Chen, X., Li, N., Ren, M., 2016, 129. A review on predicting ground PM<sub>2.5</sub> concentration using satellite aerosol optical depth. *Atmosphere* 7.
- Della Ceca, L.S., Ferreyra, M.F.G., Lyapustin, A., Chudnovsky, A., Otero, L., Carreras, H., Barnaba, F., 2018. Satellite-based view of the aerosol spatial and temporal variability in the Cordoba region (Argentina) using over ten years of high-resolution data. *ISPRS J. Photogram. Remote Sens.* 145, 250–267.
- Dominici, F., Peng, R.D., Bell, M.L., Pham, L., McDermott, A., Zeger, S.L., Samet, J.M., 2006. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA-J. Am. Med. Assoc.* 295, 1127–1134.
- Fang, X., Zou, B., Liu, X., Sternberg, T., Zhai, L., 2016. Satellite-based ground PM 2.5 estimation using timely structure adaptive modeling. *Remote Sens. Environ.* 186, 152–163.
- Guenther, A., Hewitt, C.N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., Mckay, W.A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., Zimmerman, P., 1995. A global model of natural volatile organic compound emissions. *J. Geophys. Res. Atmos.* 100, 8873–8892.
- Guo, Y., Tang, Q., Gong, D.Y., Zhang, Z., 2017. Estimating ground-level PM 2.5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sens. Environ.* 198, 140–149.
- He, Q. and Huang, B., **Satellite-based high-resolution PM<sub>2.5</sub> estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model.** *Environ. Pollut.* 236, 2018a, 1027–1037.
- He, Q., Huang, B., 2018b. Satellite-based mapping of daily high-resolution ground PM<sub>2.5</sub> in China via space-time regression modeling. *Remote Sens. Environ.* 206, 72–83.
- Hoff, R.M., Christopher, S.A., 2009. Remote sensing of particulate pollution from space: have we reached the promised land? *J. Air Waste Manage.* 59, 645–675.
- Hu, H., 1935. The distribution of population in China, with statistics and maps. *Acta Geographica Sinica* 2, 33–74.
- Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L., Strickland, M., Liu, Y., 2017. Estimating PM<sub>2.5</sub> concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol.* 51, 6936.
- Hu, X., Waller, L.A., Al-Hamdan, M.Z., Crosson Jr, W.L.E.M., Estes, S.M., Quattrochi, D.A., Sarnat, J.A., Liu, Y., 2013. Estimating ground-level PM(2.5) concentrations in the southeastern U.S. using geographically weighted regression. *Environ. Res.* 121, 1.
- Hu, X., Waller, L.A., Lyapustin, A., Wang, Y., Al-Hamdan, M.Z., Crosson Jr, W.L.M.G.E., Estes, S.M., Quattrochi, D.A., Puttaswamy, S.J., 2014. Estimating ground-level PM 2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sens. Environ.* 140, 220–232.
- Jackson, J.M., Liu, H., Laszlo, I., Kondragunta, S., Remer, L.A., Huang, J., Huang, H.C., 2013. Suomi-NPP VIIRS aerosol algorithms and data products. *J. Geophys. Res. Atmos.* 118, 673–689.
- Lee, H.J., Liu, Y., Coull, B.A., Schwartz, J., Koutrakis, P., 2011. A novel calibration approach of MODIS AOD data to predict PM<sub>2.5</sub> concentrations. *Atmos. Chem. Phys.* 11, 9769–9795.
- Li, T., Shen, H., Yuan, Q., Zhang, X., Zhang, L., 2017. Estimating ground-level PM<sub>2.5</sub> by fusing satellite and station observations: a geo-intelligent deep learning approach. *Geophys. Res. Lett.* 44, 11,985–991,993.
- Liu, J.Z., Li, W.F., Wu, J.Z., 2018. A framework for delineating the regional boundaries of PM<sub>2.5</sub> pollution: a case study of China. *Environ. Pollut.* 235, 642–651.
- Liu, Y., Franklin, M., Kahn, R., Koutrakis, P., 2007. Using aerosol optical thickness to predict ground-level PM 2.5 concentrations in the St. Louis area: a comparison between MISR and MODIS. *Remote Sens. Environ.* 107, 33–44.
- Liu, Y., Paciorek, C.J., Koutrakis, P., 2009. Estimating regional spatial and temporal variability of PM(2.5) concentrations using satellite data, meteorology, and land use information. *Environ. Health Perspect.* 117, 886–892.
- Liu, Y., Sarnat, J.A., Kilaru, A., Jacob, D.J., Koutrakis, P., 2005. Estimating ground-level PM<sub>2.5</sub> in the eastern United States using satellite remote sensing. *Environ. Sci. Technol.* 39, 3269–3278.
- Lyapustin, A., Wang, Y., Xiong, X., Meister, G., Platnick, S., Levy, R., Franz, B., Korkin, S., Hilker, T., Tucker, J., 2014. Scientific impact of MODIS C5 calibration degradation and C6+ improvements. *Atmos. Meas. Tech. Discuss.* 7, 4353–4365.
- Ma, Z., Hu, X., Huang, L., Bi, J., Liu, Y., 2014. Estimating ground-level PM<sub>2.5</sub> in China using satellite remote sensing. *Environ. Sci. Technol.* 48, 7436–7444.
- Ma, Z., Hu, X., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L., Liu, Y., 2016a. Satellite-based spatiotemporal trends in PM<sub>2.5</sub> concentrations: China, 2004–2013. *Environ. Health Perspect.* 124, 184–192.
- Ma, Z.W., Liu, Y., Zhao, Q.Y., Liu, M.M., Zhou, Y.C., Bi, J., 2016b. Satellite-derived high resolution PM<sub>2.5</sub> concentrations in Yangtze River Delta Region of China using improved linear mixed effects model. *Atmos. Environ.* 133, 156–164.
- Megaritis, A.G., Fountoukis, C., Charalampidis, P.E., Pilinis, C., Pandis, S.N., 2013. Response of fine particulate matter concentrations to changes of emissions and temperature in Europe. *Atmos. Chem. Phys.* 13, 3423–3443.
- Moran, P.A.P., 1950. A test for the serial independence of residuals. *Biometrika* 37, 178–181.
- Mountrakis, G., Li, J., Lu, X.Q., Hellwich, O., 2018. Deep learning for remotely sensed data. *ISPRS J. Photogram. Remote Sens.* 145, 1–2.
- Pope, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA-J. Am. Med. Assoc.* 287, 1132–1141.
- Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.K., Bloom, S., Chen, J.Y., Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F.R., Ruddick, A.G., Sienkiewicz, M., Woollen, J., 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Climate* 24, 3624–3648.
- Schueler, C.F., Ardanuy, P.E., Swenson, H., 2002. NPOESS VIIRS sensor design overview. *Proc. SPIE* 11–23.
- Song, W., Jia, H., Huang, J., Zhang, Y., 2014. A satellite-based geographically weighted regression model for regional PM 2.5 estimation over the Pearl River Delta region in China. *Remote Sens. Environ.* 154, 1–7.
- Tai, A.P.K., Mickley, L.J., Jacob, D.J., 2010. Correlations between fine particulate matter (PM<sub>2.5</sub>) and meteorological variables in the United States: implications for the sensitivity of PM<sub>2.5</sub> to climate change. *Atmos. Environ.* 44, 3976–3984.
- Tan, W.Z., 2007. The Basic Theoretics and Application Research on Geographically Weighted Regression. Tongji University, Shanghai.
- Thiessen, A.H., 1911. Precipitation averages for large areas. *Mon. Weather Rev.* 39, 1082–1089.
- van Donkelaar, A., Martin, R.V., Brauer, M., Boys, B.L., 2015. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ. Health Perspect.* 123, 135–143.
- 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., 2016. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environ. Sci. Technol.* 50, 3762–3772.
- Wang, W., Mao, F.Y., Du, L., Pan, Z.X., Gong, W., Fang, S.H., 2017. Deriving hourly PM<sub>2.5</sub> concentrations from Himawari-8 AODs over Beijing-Tianjin-Hebei in China. *Remote Sensing* 9, 17.
- Wu, J., Li, J., Peng, J., Li, W., Xu, G., Dong, C., 2015. Applying land use regression model to estimate spatial variation of PM<sub>2.5</sub> in Beijing, China. *Environ. Sci. Pollut. Res.* 22, 7045–7061.
- Wu, J., Yao, F., Li, W., Si, M., 2016. VIIRS-based remote sensing estimation of ground-level PM 2.5 concentrations in Beijing-Tianjin-Hebei: a spatiotemporal statistical model. *Remote Sens. Environ.* 184, 316–328.
- Xiao, Q., Chang, H.H., Geng, G., Liu, Y., 2018. An ensemble machine-learning model to predict historical PM<sub>2.5</sub> concentrations in China from satellite data. *Environ. Sci. Technol.* 52, 13260–13269.
- Xiao, Q., Zhang, H., Choi, M., Li, S., Kondragunta, S., Kim, J., Holben, B., Levy, R.C., Liu, Y., 2016. Evaluation of VIIRS, GOCI, and MODIS Collection 6 AOD retrievals against ground sunphotometer observations over East Asia. *Atmos. Chem. Phys.* 16, 20709–20741.
- Xu, J., Martin, R., Kim, J., Choi, M., Zhang, Q., Geng, G., Liu, Y., Ma, Z., Huang, L., Wang, Y., 2015. Estimating ground-level PM<sub>2.5</sub> in China using aerosol optical depth determined from the GOCI satellite instrument. *Atmos. Chem. Phys.* 15, 13133–13144.
- Yang, Q., Yuan, Q., Li, T., Shen, H., Zhang, L., 2017. The relationships between PM<sub>2.5</sub> and meteorological factors in China: seasonal and regional variations. *Int. J. Environ. Res. Public Health* 14, 1510.
- Yao, F., Si, M., Li, W., Wu, J., 2018. A multidimensional comparison between MODIS and VIIRS AOD in estimating ground-level PM<sub>2.5</sub> concentrations over a heavily polluted region in China. *Sci. Total Environ.* 618, 819–828.
- Zhang, R., Di, B., Luo, Y., Deng, X., Grieneisen, M.L., Wang, Z., Yao, G., Zhan, Y., 2018. A nonparametric approach to filling gaps in satellite-retrieved aerosol optical depth for estimating ambient PM<sub>2.5</sub> levels. *Environ. Pollut.* 243, 998–1007.
- Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., Qiu, Y., 2017. Analysis of influential factors for the relationship between PM<sub>2.5</sub> and AOD in Beijing. *Atmos. Chem. Phys.* 17, 13473–13489.
- Zou, B., Chen, J., Zhai, L., Fang, X., Zheng, Z., 2016. Satellite based mapping of ground PM<sub>2.5</sub> concentration using generalized additive modeling. *Remote Sens.* 9, 1.