Remote Sensing of Environment 231 (2019) 111221

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach

Jing Wei[a], Wei Huang[b], Zhanqing Li[c,*], Wenhao Xue[a], Yiran Peng[d], Lin Sun[e], Maureen Cribb[c]

[a] State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China
[b] State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing, China
[c] Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[d] Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China
[e] College of Geomatics, Shandong University of Science and Technology, Qingdao, China

ARTICLE INFO

Keywords:
$PM_{2.5}$
AOD
MODIS
MAIAC
Space-time random forest model

ABSTRACT

Fine particulate matter ($PM_{2.5}$) is closely related to the atmospheric environment and human life. Satellite-based aerosol optical depth (AOD) products have been widely applied in estimating daily $PM_{2.5}$ concentrations over large areas using statistical regression models. However, they are often given at coarse spatial resolutions which limit their applications in small or medium scales. This study aims to produce $PM_{2.5}$ concentrations at a high spatial resolution (1 km) across China based on the newly released Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6 Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD product using a newly developed space-time random forest (STRF) model. Daily $PM_{2.5}$ concentrations for 2016 were estimated from in-situ surface $PM_{2.5}$ measurements, and meteorological and ancillary variables. The 10-fold cross-validation (CV) approach and three popular models, including the multiple linear regression model, the geographically weighted regression model, and the two-stage model, are employed for validation and cross-comparison. A sample-based CV of the STRF model shows a high and stable accuracy with a coefficient of determination equal to 0.85, a root-mean-square error of 15.57 $\mu g\,m^{-3}$, and a mean prediction error of 9.77 $\mu g\,m^{-3}$. This finding suggests that the STRF model can predict $PM_{2.5}$ daily, monthly, and annual concentrations at an unprecedented spatial resolution and accuracy across China. It also appears to have outperformed the above popular models and the previous related studies. In general, the STRF model is robust and can accurately estimate $PM_{2.5}$ concentrations by taking advantage of the ensemble regression approach, the synergy of space-time information, and the high-resolution, high-quality, and wide-spatial-coverage of the MAIAC AOD product. It may thus also be useful for applications in related air pollution studies, especially those focused on urban areas.

1. Introduction

Particulate matter with aerodynamic diameters ≤ 2.5 μm, or $PM_{2.5}$, also known as pulmonary particles, can significantly damage human health according to numerous epidemiological studies, exacerbating respiratory and cardiovascular diseases (e.g., Bartell et al., 2013; Brauer et al., 2012; Crouse et al., 2012; Pascal et al., 2014; Peng et al., 2009). $PM_{2.5}$ mainly comes from direct emissions from natural processes (e.g., forest fires, tsunamis, and volcanic eruptions), and human activities (e.g., industrial production, transportation, and fuel combustion). In recent decades, the rapid economic development experienced in China has seriously undermined air pollution in this country. Fine particles have been the primary pollutants in urban areas and have become a major public concern (Bi et al., 2014; Huang et al., 2015; Li et al., 2017; Sun et al., 2016). As a result, several $PM_{2.5}$ observation networks (e.g., the Chinese Meteorological Administration Atmosphere Watch Network) have been established in China recently for monitoring real-time fine particles. However, the coverage is still on the sparse side with varying standards of observational quality. High-resolution $PM_{2.5}$ observations with consistent quality and wide-area coverage are thus highly desired.

Thanks to numerous satellite missions and sensors, achieving this objective is possible by means of remote sensing in combination with surface $PM_{2.5}$ ground measurements. As demonstrated in many studies (e.g., Engel-Cox et al., 2004; Guo et al., 2009; Hu et al., 2014), aerosol optical depth (AOD), representing the degree to which aerosols

* Corresponding author.
E-mail address: zli@atmos.umd.edu (Z. Li).

https://doi.org/10.1016/j.rse.2019.111221
Received 29 November 2018; Received in revised form 10 April 2019; Accepted 26 May 2019
Available online 28 June 2019
0034-4257/

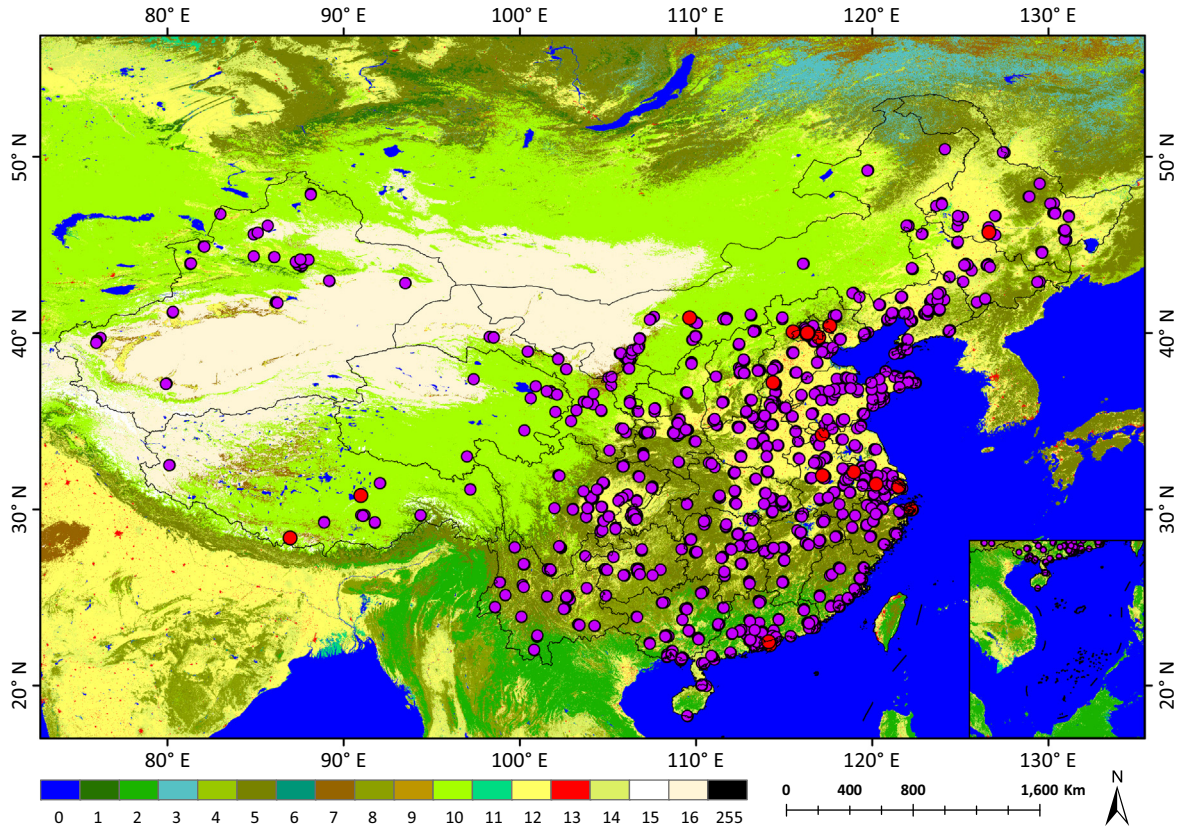

Fig. 1. Locations of $PM_{2.5}$ (purple dots) and Aerosol Robotic Network (red dots) monitoring stations in China. Land use cover types for 2016 (background colored shading) are from the MODIS land use cover product at a 500-m spatial resolution. Descriptions of the land use classes in the legend are given in Table S1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

attenuate light in the atmospheric column, is highly correlated with particulate matter concentrations. With the continuous improvement and maturity of aerosol retrieval algorithms, a variety of aerosol products obtained from different satellite sensors have been widely used in these studies. These sensors include the Multi-angle Imaging SpectroRadiometer (Liu et al., 2004, 2007; You et al., 2015), the Moderate Resolution Imaging Spectroradiometer (MODIS; Fang et al., 2016; Just et al., 2015; Lee et al., 2011; Li et al., 2017; Ma et al., 2014; Sorek-Hamer et al., 2013; Yu et al., 2017), the Visible Infrared Imaging Radiometer Suite (VIIRS; Pang et al., 2018; Yao et al., 2018), the Geostationary Ocean Color Imager (Pang et al., 2018), and the Advanced Himawari-8 Imager (Zang et al., 2018). However, these products generally have coarse resolutions (3–50 km) which limit their applications to atmospheric pollution studies, besides some inherent limitations in the aerosol remote sensing (Li et al., 2009). The coarser resolution renders particular problems for studying air pollution in fine scales like urban areas where air composition is heterogeneous due to complex emission sources.

A new global-coverage high-spatial-resolution (1 km) MODIS Collection 6 (C6) daily AOD dataset (MCD19A2) has been released on 30 May 2018 (Lyapustin et al., 2018). The product is generated based on the new Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm, which is different from the well-known Dark Target (DT; Levy et al., 2013) and Deep Blue (DB, Hsu et al., 2013) algorithms. It uses the semi-empirical Ross-thick/Li-sparse bi-directional reflectance distribution function and the semi-analytical Green's function solution models (Lyapustin et al., 2011a, 2011b, 2012a). The Lambertian equivalent reflector approximation is used to estimate the spectral regression coefficients. For global processing and operation, improvements have been made to numerous key steps such as cloud and snow screening, and aerosol-type selection following a time-series image-based analysis (Lyapustin et al., 2012b).

There are various developed models for examining the relationship between surface ground-based $PM_{2.5}$ and satellite-retrieved AOD, including chemical (Liu et al., 2004), physical (Zhang and Li, 2015), semi-empirical (Lin et al., 2015), and statistical models. Among these, the statistical regression approach has been the most widely used because of its fast and simple characteristics, and sound accuracy. The $PM_{2.5}$-AOD relationships take on many mathematical forms: the linear regression model (Liu et al., 2005), the multiple line regression (MLR) model (Xiao et al., 2017), the geographically weighted regression (GWR) model (Ma et al., 2014; Yu et al., 2017), the linear mixed-effect (LME) model (Xiao et al., 2017; Yu et al., 2017), the two-stage model (Ma et al., 2016a), and the geographically and temporally weighted regression (GTWR) model (He and Huang, 2018). $PM_{2.5}$ concentrations are affected by many factors such as meteorological conditions, land changes, and human activities, making the use of any of these traditional models to estimate $PM_{2.5}$ challenging. Machine learning, commonly known as data mining, has thus become popular when tackling this and many other complex problems because of its superior capability in picking and using a large number of independent factors that can affect the dependent variable to be predicted.

An attempt is made here to use the machine-learning approach to first generate a 1-km daily $PM_{2.5}$ concentration dataset over mainland China where emission sources are exceptionally complex and meteorological conditions highly variable. Based on machine-leaning procedures, the space-time random forest (STRF) model was first developed to cope with the spatial heterogeneity and temporal variations of emissions and meteorological variables. The STRF model developed in this study for application with the MODIS 1-km MAIAC AOD product, together with other pertinent variables related to meteorological conditions, land use, and human activities, was used to generate 1-km daily

$PM_{2.5}$ concentrations. The model performance relative to predicted records is assessed and compared with MLR-, GWR-, and two-stage model- (i.e., stage 1: LME model; stage 2: GWR model) generated $PM_{2.5}$ concentrations. The approaches used to validate the model performance are the sample-, spatial-, and temporal-based 10-fold cross-validation (10-CV) approaches.

2. Datasets used and preprocessing

The datasets used in this study consist mainly of hourly in-situ $PM_{2.5}$ concentrations, 1-km MODIS MAIAC AOD, and auxiliary data related to $PM_{2.5}$ such as meteorological, land- and population-related information.

2.1. $PM_{2.5}$ ground measurements

This study uses calibrated and quality-controlled $PM_{2.5}$ concentration measurements (Guo et al., 2009) from 1480 stations across China acquired by the China National Environmental Monitoring Center (http://www.cnemc.cn) from 2015 to 2016 (Fig. 1). The observation stations are unevenly distributed across China. Eastern China has the greatest density of stations while stations are scarce in the desert and plateau regions of western China. $PM_{2.5}$ measurements at each station that remain the same for more than three continuous hours are removed because of likely instrument malfunction (Rohde and Muller, 2015; Xiao et al., 2017).

2.2. MODIS AOD products

The newly released Terra and Aqua MODIS C6 daily 1-km MAIAC AOD products at 550 nm from 2015 to 2016 are employed. They are subject to quality assurance (QA), cloud screening ($QA_{CloudMask}$ = Clear), and adjacency ($QA_{AdjacencyMask}$ = Clear) tests (Lyapustin et al., 2018). Three widely used Terra and Aqua MODIS daily DT and DB products at 550 nm at different spatial resolutions for the same period are also collected for comparison purposes. They include the 3-km DT, 10-km DT, and 10-km DB AOD products that have passed the recommended quality assurance tests (Levy et al., 2013).

2.2.1. Validation against AERONET measurements

The above four operational MODIS AOD products at different spatial resolutions are first validated against Aerosol Robotic Network (AERONET) AOD ground measurements at 20 sites (marked as red dots in Fig. 1) across China (Fig. S1). Table S2 provides detailed information about the selected AERONET sites in China. The time- and space-matching approaches developed in our previous studies for matching satellite-derived and ground-based AODs are used in the evaluation (Wei et al., 2019a, 2019b). Results show that the DT algorithm performs the worse with the least number of valid samples and with < 45% of the samples falling within the MODIS expected error envelope (EE, ± (0.05 + 20%), Levy et al., 2013). The largest overestimation uncertainties are also seen (Fig. S1a and b). The reason is that the DT algorithm cannot accurately estimate the surface reflectance over bright (i.e., urban and desert) surfaces (Wei et al., 2018). The 3-km DT retrievals are generally worse than the 10-km retrievals because of the increasing signal-to-noise ratio (Gupta et al., 2018). The DB algorithm can provide about 2.6 to 3.2 times more data samples with ~73% of them falling within the EE envelope. The estimation uncertainties [i.e., the mean absolute error (MAE) and root-mean-square error (RMSE)] are lower than those of the DT algorithm. This mainly arises from the improved surface reflectance modeling in the DB algorithm (Hsu et al., 2013). The MAIAC algorithm generates the most (N = 3111) and best retrievals with ~81% of the samples falling within the EE envelope, an RMSE of 0.150, and an MAE of 0.083. This is mainly attributed to improved cloud screening and an enhanced radiative transfer model used for aerosol retrievals (Lyapustin et al., 2011a, 2012b, 2018). These results illustrate that the MAIAC product is more accurate with improved spatial coverage and higher spatial resolution than the currently widely used DT- and DB-based AOD products in China, and is thus selected for use in this study.

2.2.2. Terra and Aqua combined MAIAC AOD product

Due to differences in observation time and the diurnal variation in cloud cover, the spatial coverage between the Terra and Aqua satellite-derived AOD products is noticeably different. It is thus necessary to integrate the products to enlarge the spatial coverage. An improved linear regression data integration approach by fitting the Terra and Aqua AOD retrievals on a monthly scale is proposed to fill the missing data and improve the spatial continuity. Table S3 presents the monthly linear relationships between Terra and Aqua MAIAC AOD products in China. The estimation has a sound accuracy in terms of regression slope (0.79–0.96), intercept (0.01–0.06), R^2 (0.75–0.88), and p (< 0.01), which is better than other similar approaches on annual or seasonal scales. For each grid, the combined daily MAIAC AOD is estimated, and if both Terra and Aqua AODs are available, additional averaging is done. After the data integration, the number of daily AOD records increases by > 29% (26%) and 34% (36%) in 2015 and 2016, respectively, than if using Terra (Aqua) AOD retrievals only. Overall, the number of effective samples is greater than that from previous studies (Fang et al., 2016; He and Huang, 2018; Li et al., 2017; Zheng et al., 2016). This approach largely minimizes the systematic differences between different sensors and improves the model capability by increasing the number of training samples. A wider spatial coverage of $PM_{2.5}$ concentrations across China will result.

2.3. Meteorological data

Meteorological data used are the ERA-Interim atmospheric reanalysis products (Dee et al., 2011). They are produced based on a sequential data assimilation scheme and have provided continuous real-time atmospheric parameters since 1979 at several spatial resolutions every 3 or 6 h. Eight meteorological quantities, i.e., the 2-m air temperature (TEM; unit: K), total precipitation (PRE; unit: mm), evaporation (ET; unit: mm), boundary layer height (BLH; unit: m), 10-m U/V wind components (unit: m s^{-1}), relative humidity (RH; unit: %), and surface pressure (SP; unit: hPa), at a 0.125° × 0.125° resolution are measured. Wind speed (WS; unit: m s^{-1}) and wind direction (WD; unit: degrees) are calculated from the two wind components using the vector synthesis method. Daily means of the meteorological quantities from 10:00 to 14:00 local time coincident with satellite-derived AODs are used.

2.4. Auxiliary data

Land-related variables include the International Geosphere-Biosphere Program with seventeen classes of MODIS 500-m-resolution annual land use cover (LUC) product and 1000-m-resolution monthly normalized difference vegetation index (NDVI) product. The 90-m-resolution Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) is also used. The population-related variable is the nighttime lights (NTL) product generated from the Suomi National Polar-Orbiting Partnership VIIRS at a 500-m resolution, reflecting human activities and regional economic development. Table S4 summarize the data sources used in this study.

2.5. Data processing

Due to different spatial and temporal resolutions, all quantities used in this study are uniformly resampled to the same spatial size (0.01° × 0.01° ≈ 1 × 1 km) using the bilinear interpolation method and the same temporal interval. The 14 selected independent variables are matched to the daily $PM_{2.5}$ concentrations for each station on each

day. After removing invalid values, there are a total of 149,351 and 153,648 matched daily ground $PM_{2.5}$ concentrations and all independent variables covering all days in 2015 and 2016, respectively.

3. Methodology

$PM_{2.5}$ concentrations vary dramatically in space and time due to vast geographical areas, complex surface structures, and human disturbances in China. A STRF model is thus developed to tackle the spatiotemporal variability and its impact on the AOD-$PM_{2.5}$ relationship. Three popular regression models, i.e., the MLR, GWR, and (LME + GWR), are used to see how their $PM_{2.5}$ concentration estimates compare with STRF-estimated $PM_{2.5}$ concentrations.

3.1. Correlation and collinearity diagnosis

The overall relationship between MAIAC-AOD and $PM_{2.5}$ measurements is not close across China [correlation coefficient (R) = 0.473], suggesting the influence of other variables. Meteorological and other auxiliary quantities are thus introduced to enhance the relationship (Table S5). Variables that have significant positive or negative effects on $PM_{2.5}$ concentrations ($p < 0.01$) are selected. However, the collinearity of independent variables must be taken into account to avoid duplication of information conveyed in the seemingly independent variables. For this, the variance inflation factor (VIF) method is chosen to diagnose collinearity among the selected variables (i.e., AOD, DEM, NDVI, LUC, NTL, TEM, PRE, ET, RH, SP, BLH, WS, and WD). The VIF values of DEM (~17.84) and SP (~21.25) are very high, indicating strong collinearity (Table S6). The ERA-interim SP with a higher daily spatial resolution than DEM remains, along with the other eleven independent variables, for model fitting in this study.

3.2. Space-time random forest model

The traditional Random Forest (RF; Breiman, 2001) is a relatively modern and highly flexible machine-learning approach that integrates multiple decision trees. It can evaluate the importance of each feature during the classification, produce an unbiased estimation, and handle conditions with a large amount of missing data. Moreover, it can efficiently process thousands of input variables without the necessity of dimensionality reduction. There are many classification trees in RFs, and each variable must be input into each tree for classification. Each tree in the forest is independent, and 99.9% of unrelated trees make predictions that cover all situations. These predictions will offset each other. Predictions with a few excellent trees are beyond the "noise" and used for making a good prediction. The basis of RF bagging is to select the results of several weak classifiers and form a strong classifier. Each tree in the RF can be generated by the following steps (Hastie et al., 2008):

(a) For each tree, Z samples are extracted from the training dataset (N) using the bootstrap sample method.
(b) Randomly select m feature subsets from the M-dimension features and select the best one from m features as each tree splits into two daughter nodes.
(c) Every tree grows as far as it can without the pruning process. The characteristics of randomness are crucial to the classification of RFs to avoid over-fitting and noise.
(d) The optimal M is the key parameter in constructing the RF, which can be determined by the out-of-bag (*oob*) error. For each tree (assuming the k_{th} tree), approximately one-third of the training samples do not participate in the generation of the k_{th} tree, which is *oob* samples of the k_{th} tree. For each regression tree [$f(x)$],

$$f(x) = \sum_{z=1}^{Z} c_z I(x \in R_z) \tag{1}$$

$$\widehat{c_z} = mean(y_i \mid x_i \in R_z) \tag{2}$$

$$Z_1(m,n) = \{X \mid X_j \le n\} \& Z_2(m,n) = \{X \mid X_j > n\} \tag{3}$$

$$\min_{m,n}\left[\min_{j,s}\sum_{x_i \in R_1(m,n)}(y_i - c_1)^2 + \min_{m,n}\sum_{x_i \in R_2(m,n)}(y_i - c_2)^2\right] \tag{4}$$

and

$$\widehat{c_1} = mean(y_i \mid x_i \in R_1(m,n)) \& \widehat{c_2} = mean(y_i \mid x_i \in R_2(m,n)) \tag{5}$$

where (x_i, y_i) is the samples for i = 1, 2, …, N for Z regions (R_1, R_2, …, R_Z), c_m is the response to the model, which is a constant, $\widehat{c_z}$ is the best value, m is the splitting variable, and n is the split point.

The RF model is one of the traditional machine learning models, which has been reported previously for applications in different fields. However, there are much fewer applications in remote sensing, more importantly, these applications of the RF model were based on the quantities of input independent variables and did not consider the spatiotemporal information (Hu et al., 2017; Brokamp et al., 2018; Huang et al., 2018). $PM_{2.5}$ is most essentially characterized by its spatial and temporal heterogeneity, and many researchers have tried to solve these problems, leading to the development of well-known models, such as the GWR, the two-stage, and the GTWR models. To further improve the accuracy of the $PM_{2.5}$ estimations, a new space-time random forest (STRF) model is developed. Not only does our model make use of rich information as conveyed in the input variables, it also benefits from the extraction of geospatial information weighted by the distance of adjacent pixels from a center pixel, as well as the time difference from ground-based $PM_{2.5}$ measurements. For a given pixel, its spatial (Ps) and temporal (Pt) characteristics can be expressed as,

$$\mathrm{Ps} = \frac{\sum_{w=1}^{W}\frac{1}{ds_w^2}Ps_w}{\sum_{w=1}^{W}\frac{1}{ds_w^2}} \tag{6}$$

and

$$\mathrm{Pt} = \frac{\sum_{l=1}^{L}\frac{1}{dt_l^2}Pt_l}{\sum_{l=1}^{L}\frac{1}{dt_l^2}} \tag{7}$$

where *ds* and *dt* represent the spatial and temporal distances. The W and L represent the w pixels near the site and the l prior days for the same pixel. The improved AOD-$PM_{2.5}$ relationship can be explored with satellite-derived AOD, meteorological variables, land-related parameters, population-related indices, and spatiotemporal information as inputs using the STRF model [$PM_{2.5} = f$(AOD, NDVI, LUC, NTL, TEM, PRE, ET, RH, SP, BLH, WS, WD, Ps, Pt)]. The output is the estimated 1-km $PM_{2.5}$ concentrations. Fig. 2 illustrates the structure and specific schematics of the STRF model.

3.3. Other statistical regression models

The MLR model (Gupta and Christopher, 2009) used here is expressed as:

[$PM_{2.5} = f$(AOD, NDVI, LUC, NTL, TEM, PRE, ET, RH, SP, BLH, WS, WD)],

$$PM_{2.5} = a_1 AOD + a_2 NDVI + \ldots + a_n WD + b, \tag{8}$$

where a_1, …, a_n are the regression coefficients and b is the intercept.

The GWR model (Hu et al., 2013) is expressed as,

$$PM_{2.5s} = b_0(i,j) + b_1(i,j)AOD_s + b_2(i,j)NDVI_s + \ldots + b_n(i,j)WD_s + \varepsilon_s, \tag{9}$$

where $PM_{2.5s}$ represents the daily $PM_{2.5}$ concentration at location $s(i, j)$,

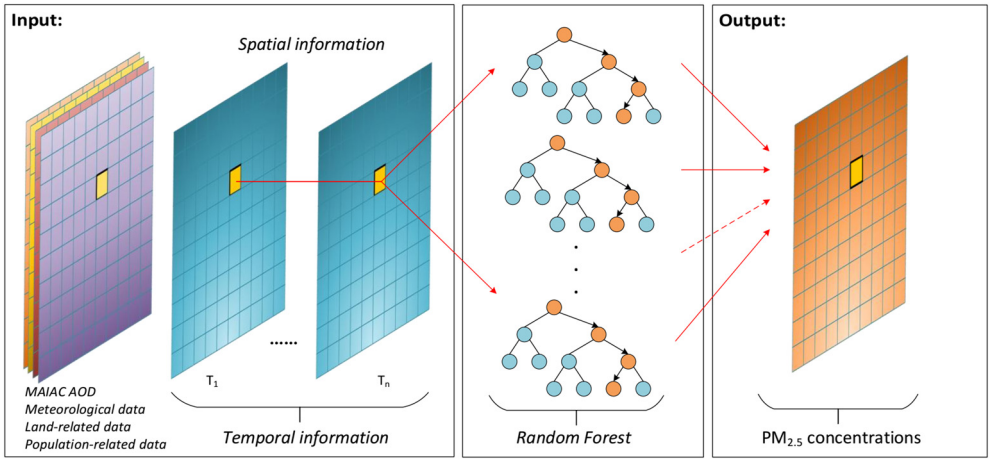

Fig. 2. The structure and specific schematics of the space-time random forest (STRF) model.

b_0 and b_1-b_n denote the intercepts and slopes for each independent variable, and ε stands for the error term.

The two-stage model contains an LME model in stage one and a GWR model in stage two (Ma et al., 2016a). It is expressed as.

$$PM_{2.5st} = (c_0 + c_{0t}) + (c_1 + c_{1t})AOD_{st} + (c_2 + c_{2t})NDVI_{st} + \ldots + (c_n + c_{nt})WD_{st} + \varepsilon_{st}(\beta_{0t}, \beta_{1t}, \ldots, \beta_{nt}) \sim N[(0,\ldots,0), \varphi], \text{ —Stage 1} \quad (10)$$

$$PM_{2.5\_}resi_{st} = d_{0t}(i,j) + d_{1t}(i,j)AOD_{st} + \varepsilon_{st}, \text{ —Stage 2} \quad (11)$$

where $PM_{2.5st}$ and $PM_{2.5\_}resi_{st}$ are $PM_{2.5}$ concentrations and residuals from the first-stage model on day t, c_0 and c_{0t} are the intercepts, c_1-c_n are the slopes, c_{1t}-c_{nt} are the slopes for the independent variables, φ are the variance-covariance matrices, and d_0 and d_{1t} are the intercept and slope for the GWR model.

3.4. Evaluation approaches

For validation purposes, three approaches, i.e., sample-, spatial-, and temporal-based 10-CV, are selected to validate the $PM_{2.5}$ concentrations estimated by the previously mentioned models. The sample-based CV is a widely used validation approach (Rodriguez et al., 2010) that involves randomly selecting 90% of the samples for modeling and using the remaining 10% for validation. This process is repeated 10 times to ensure that all samples are tested. The spatial-based CV and temporal-based CV are like the sample-based CV, which performs the 10-CV process with 10% of ground-based sites (space-CV) and 10% of days (time-CV) randomly dropped. The evaluation includes two parts. The first part is model training and validation based on the three validation approaches. Due to more missing data in 2015 than in 2016, the data from 2016 are used. The second part evaluates the model's predictive power. The models developed for 2015 are used to predict $PM_{2.5}$ concentrations for 2016, which are then validated using surface measurements of $PM_{2.5}$ concentrations from 2016. The following statistical metrics are used to quantitatively evaluate the performance of the models: the linear regression equation (slope and intercept), the coefficient of determination (R^2), the RMSE, and the mean prediction error (MPE).

4. Results and discussion

4.1. Descriptive statistics

Fig. 3 first plots the histograms of all site-based variables used to fit the $PM_{2.5}$-estimation models from 153,648 matched samples for all days in 2016 over mainland China. In general, the surface daily $PM_{2.5}$ concentrations show close similar distributions with AOD. The bivariate correlation analysis illustrates that all independent variables were significantly related to $PM_{2.5}$, especially AOD ($R = 0.473$, $p < 0.01$) and ERA-interim BLH ($R = -0.326$, $p < 0.01$, Table S5). Furthermore, there are significant seasonal variations for several variables. The annual mean observed $PM_{2.5}$ concentration is $50.78 \pm 40.76\ \mu g\ m^{-3}$ averaged at all ground-based stations in China, where the highest $PM_{2.5}$ concentration is $69.50 \pm 52.03\ \mu g\ m^{-3}$ in winter, and the lowest is $29.28 \pm 17.30\ \mu g\ m^{-3}$ in summer. The annual mean MAIAC AOD is 0.39 in China, and seasonally averaged AODs do not vary much: 0.43, 0.37, 0.38, and 0.39 in sping, summer, autumn, and winter, respectively. The discordant seasonal variations are likely caused by varying planetary BLH and hygroscopic growth, both of which have strong seasonal variations (Li et al., 2017; Su et al., 2018). Table S7 gives the seasonal statistics for all selected variables.

4.2. Model fitting and validation

4.2.1. National-scale model performance

Fig. 4 shows the density scatterplots of the fitting and 10-CV results for the traditional RF and newly developed STRF models in 2016 for mainland China. The RF and STRF models perform equally well with equal R^2 values of 0.98, according to model-fitting results. The overall RMSE (MPE) values are 6.40 (4.10) $\mu g\ m^{-3}$ and 5.57 (3.48) $\mu g\ m^{-3}$ for the RF and STRF models, respectively. This suggests that the two RF approaches do well at generating training approximations. For CV on a spatial scale, the two models have relatively poor accuracies because $PM_{2.5}$ concentrations are noticeably spatially heterogeneous. However, the STRF model performs better with a higher space-CV R^2 of 0.63, a lower RMSE of 24.83 $\mu g\ m^{-3}$ and an MPE of 15.53 $\mu g\ m^{-3}$ than the RF model ($R^2 = 0.57$, RMSE = 26.58 $\mu g\ m^{-3}$, and MPE = 17.10 $\mu g\ m^{-3}$). In general, the model accuracies on the temporal scale of the two models are much better than those on the spatial scale. The STRF model is an improvement over the RF model with a better time-CV R^2, RMSE, and MPE of 0.83, 16.63 $\mu g\ m^{-3}$, and 10.55 $\mu g\ m^{-3}$, respectively. For the sample-based validation, the overall R^2, RMSE, and MPE values are 0.85, 15.57 $\mu g\ m^{-3}$, and 9.78 $\mu g\ m^{-3}$ for the STRF model, which are generally better than the values for the RF model ($R^2 = 0.81$, RMSE = 17.90 $\mu g\ m^{-3}$, MPE = 11.50 $\mu g\ m^{-3}$). In general, considering both the geographical correlation and the temporal variation improves the model performance. However, the STRF model tends to underestimate $PM_{2.5}$ concentrations (slope = 0.82, intercept = 9.64 $\mu g\ m^{-3}$), which can lead to poorer estimates on heavily polluted days.

4.2.2. Seasonal model performance

The performances in STRF model fitting and of the sample-based 10-

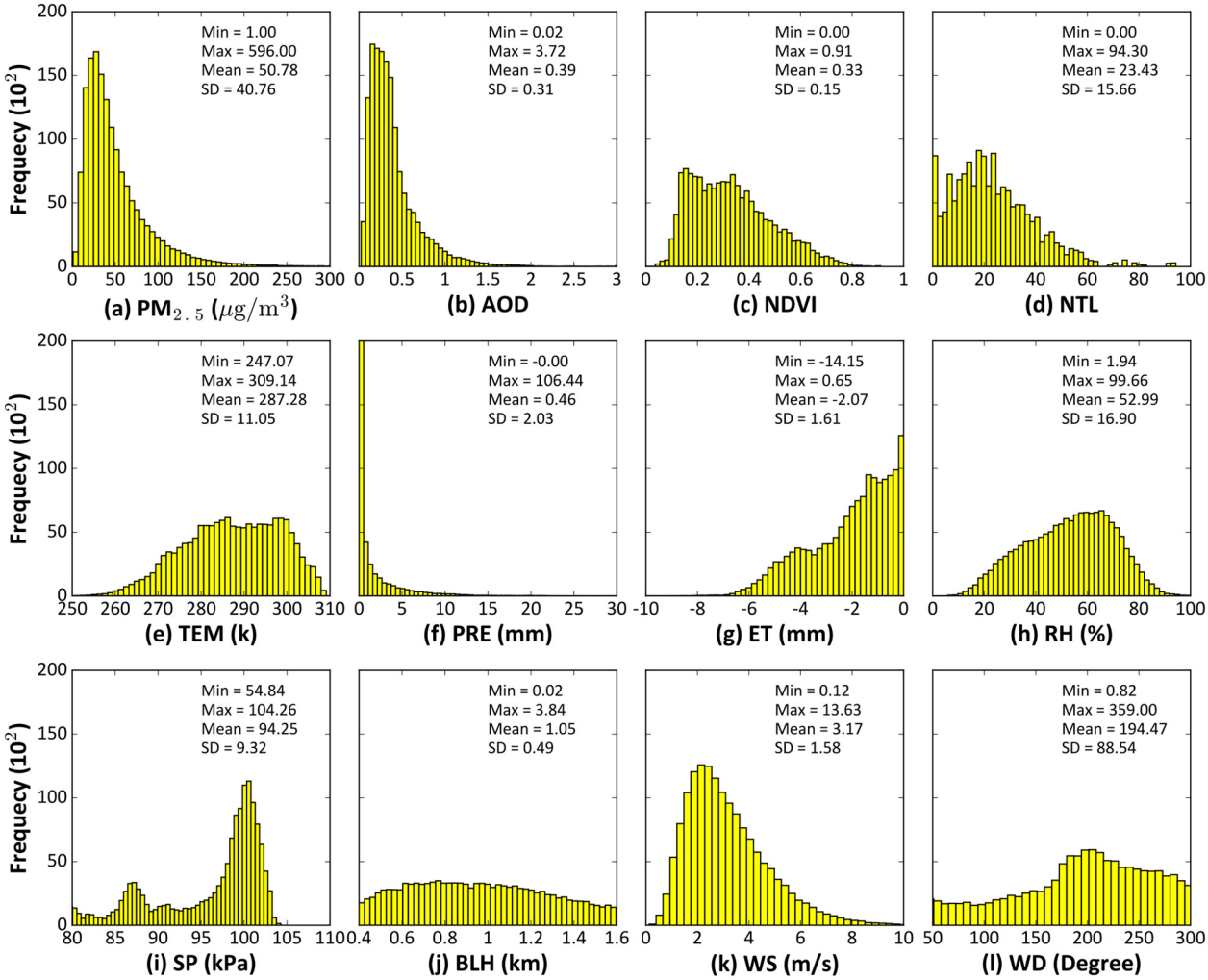

Fig. 3. Histograms and descriptive statistics (minimum, maximum, mean, and standard deviation) for $PM_{2.5}$ and the independent variables used for modeling. Data are from 2016 in China. The number of samples is 153,648.

CV tests according to the season are also evaluated (Fig. 5). The training accuracies of the STRF model-fitting results have R^2 values of 0.98, 0.96, 0.98, and 0.98 in spring, summer, autumn, and winter, respectively. The RMSE and MPE values are generally $< 8\,\mu g\,m^{-3}$ and $5\,\mu g\,m^{-3}$, respectively, for the four seasons. The STRF sample-based validation has similar seasonal characteristics. Winter is the season with the highest R^2 (0.85) but also with the highest RMSE ($20.06\,\mu g\,m^{-3}$) and MPE ($12.69\,\mu g\,m^{-3}$). By contrast, of the four seasons, summer is the season with the lowest R^2 (0.69), RMSE ($9.62\,\mu g\,m^{-3}$), and MPE ($6.43\,\mu g\,m^{-3}$). The large estimation error in winter arises because of the frequent occurrence of high $PM_{2.5}$ concentrations ($> 100\,\mu g\,m^{-3}$) due to anthropogenic pollutants. The small estimation error in summer is mainly due to low fine particulate concentrations caused by high relative humidity and frequent precipitation (Li et al., 2017; Su et al., 2018), complicating the $PM_{2.5}$-AOD relationship. The evaluation metrics in spring and autumn are similar.

4.2.3. Site-scale model performance

Fig. 6 shows the spatial patterns of the sample-based CV of the STRF model at individual monitoring stations in China. The number of days of valid estimations of $PM_{2.5}$ concentration varies from 2 to 269 days, with an average of 106 days. The number of days increases from southeast to northwest China, mainly due to the decreasing clouds in the remote sensing images. The STRF model performs better but with a noticeable spatial heterogeneity. The mean R^2 is 0.79, and > 90% of the stations show high daily R^2 values > 0.7, especially in densely populated eastern China ($R^2 > 0.8$, $p < 0.01$). By contrast, R^2 is relatively low at a few stations in the northwest and southwest of China, likely because of the sparsity of stations in this part of the country. The average RMSE and MPE in China are $13\,\mu g\,m^{-3}$ and $8\,\mu g\,m^{-3}$, respectively. More than 88% (93%) of the stations have low estimation errors with RMSE (MPE) values $< 20\,\mu g\,m^{-3}$ ($15\,\mu g\,m^{-3}$), respectively. Larger estimation errors are found in northern China where higher $PM_{2.5}$ values are found because of intensive human activities and high pollutant emissions. These results illustrate that the STRF model is robust and can significantly improve the $PM_{2.5}$-AOD relationship through its consideration of the large spatiotemporal heterogeneities.

4.3. Predictive power of the STRF model

This section focuses on evaluating the STRF model's predictability in China. For this purpose, data from 2015 and 2016 are respectively selected for model fitting and sample-based cross-prediction on different time scales. Daily $PM_{2.5}$ estimates for 2016 generated by the 2015-data-fitted STRF model are close to measurements ($R^2 = 0.55$, RMSE = $27.38\,\mu g\,m^{-3}$, MPE = $17.83\,\mu g\,m^{-3}$) (Fig. 7a). When considering the monthly scale, only stations with > 20% valid daily $PM_{2.5}$ concentration measurements in a month are used in the calculations. The STRF model performs better on a monthly scale with a steeper slope (~0.65), increased R^2 (~0.73), decreased RMSE

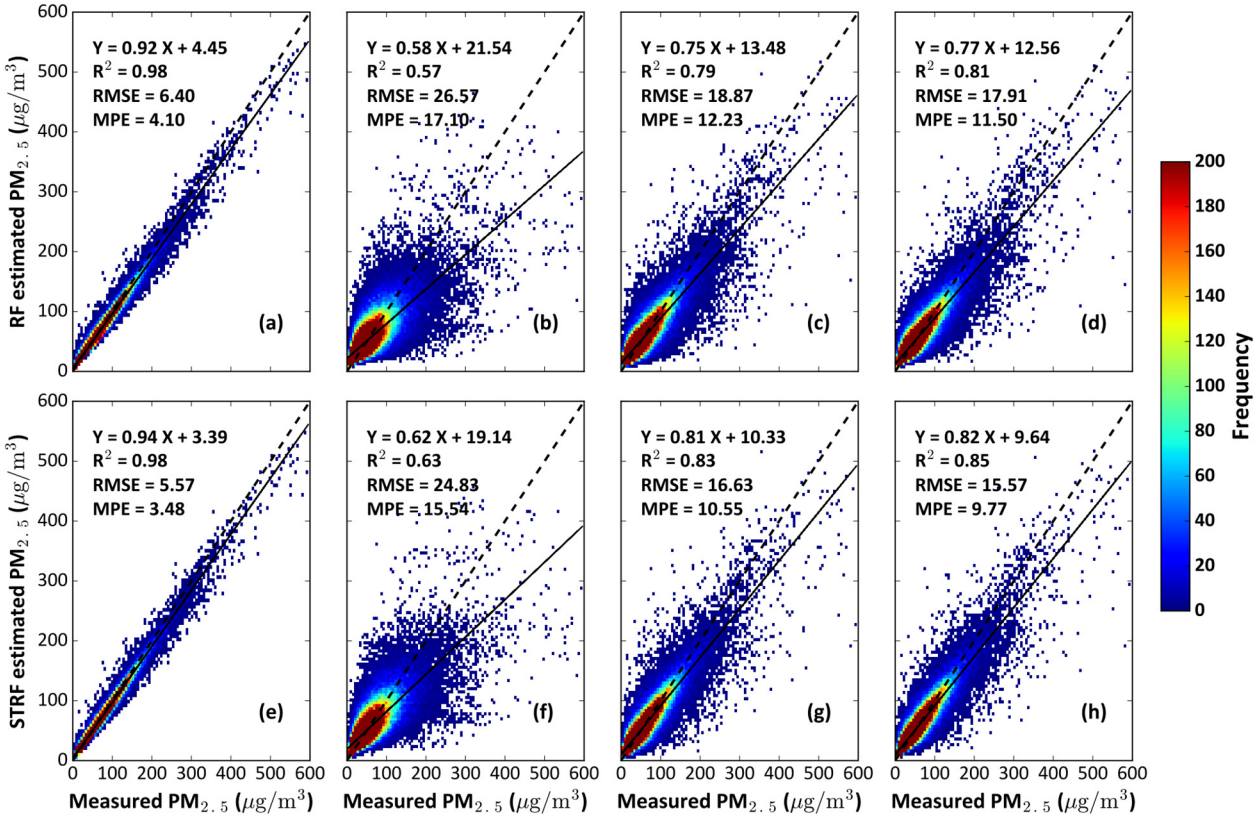

Fig. 4. Density scatterplots of model fitting, and time-based, space-based, and sample-based cross-validation results for the traditional RF (a-d) and newly STRF (e-h) models at the daily scale (N = 153,648) for mainland China. The black dashed and solid lines represent the 1:1 line and the linear regression line, respectively.

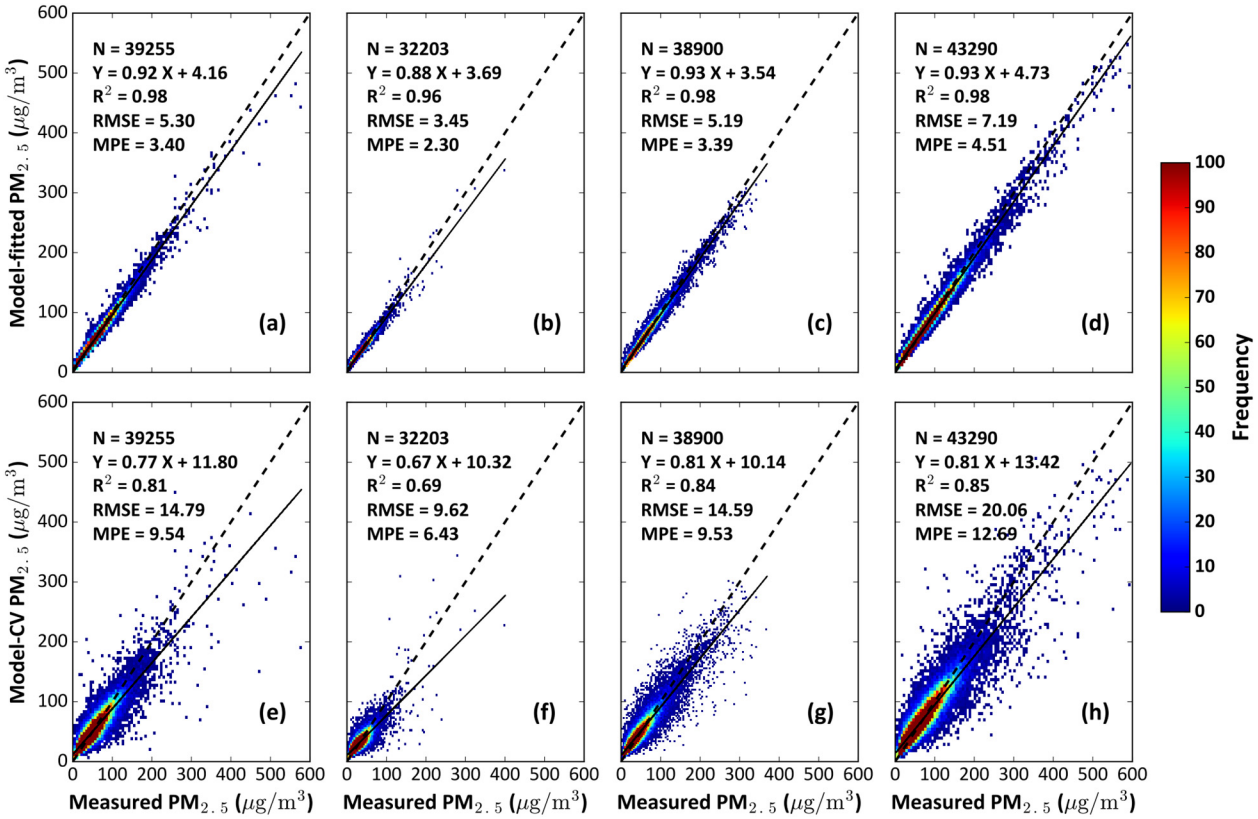

Fig. 5. Density scatterplots of model-fitting results (top row) and station-based cross-validation results (bottom row) from the STRF model for (a, e) spring, (b, f) summer, (c, g) autumn, and (d, h) winter in China.

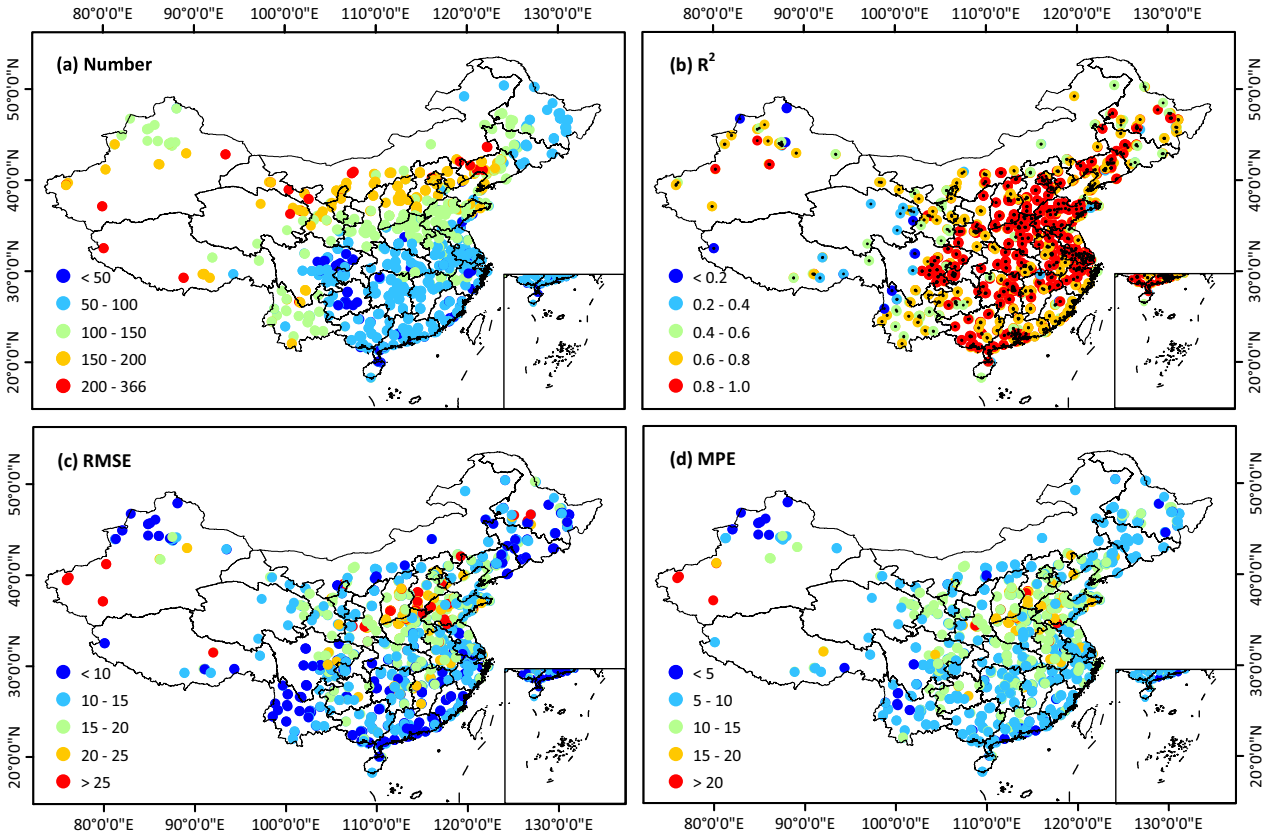

Fig. 6. Spatial distributions of (a) the number of valid samples (N) at each monitoring station, (b) R^2, (c) RMSE ($\mu g\,m^{-3}$), and (d) MPE ($\mu g\,m^{-3}$) between $PM_{2.5}$ estimations and measurements from 2016 in China. Results are from the sample-based 10-cross-validation. The black dots in (b) indicate significance at the 99% confidence level.

($\sim$14.88 $\mu g\,m^{-3}$), and decreased MPE ($\sim$10.56 $\mu g\,m^{-3}$) (Fig. 7b). Seasonal $PM_{2.5}$ estimates are averaged from at least two months of mean $PM_{2.5}$ concentrations in each season at each station (Fig. 7c). Annual $PM_{2.5}$ estimates are averaged from at least eight months of monthly mean $PM_{2.5}$ concentrations in a year at each station (Fig. 7d). They are consistent with ground measurements with improved regression lines, higher R^2 values of 0.78 and 0.79, lower RMSEs of 11.42 $\mu g\,m^{-3}$ and 8.08 $\mu g\,m^{-3}$, and MPEs of 8.59 $\mu g\,m^{-3}$ and 6.25 $\mu g\,m^{-3}$, respectively. These results illustrate that the newly developed STRF model can be well applied for predicting the $PM_{2.5}$ concentrations, which is of great significance for studying its long-term variations and their causes in China.

One hundred randomized trials on the model fitting, sample-based CV, and prediction power of $PM_{2.5}$ concentrations were also conducted. The model-fitting results are stable with almost constantly high R^2 values of 0.98, RMSE values of 5.64 $\mu g\,m^{-3}$, and MPE values of 3.51 $\mu g\,m^{-3}$ with small standard deviations of $\sim$0.000, 0.009, and 0.003, respectively. The sample-based accuracy is 0.85 $\pm$ 0.001 in R^2 with small and stable estimation errors, i.e., RMSE = 15.57 $\pm$ 0.027 $\mu g\,m^{-3}$ and MPE = 9.76 $\pm$ 0.011 $\mu g\,m^{-3}$. Moreover, concerning the predictive power, the R^2, RMSE, and MPE values are 0.55 $\pm$ 0.001, 27.38 $\pm$ 0.029 $\mu g\,m^{-3}$, and 17.83 $\pm$ 0.015 $\mu g\,m^{-3}$, respectively. The overall low standard deviations close to zero further affirm that the STRF model is reasonably stable for estimating $PM_{2.5}$ concentrations.

4.4. Comparison with other regression models

This section compares the performance of the STRF model with those of currently popular regression models including the MLR, GWR, and two-stage models using the same samples from 2016 in China (Table 1). The MLR model has the least steep slopes and the largest intercepts and significantly underestimates daily $PM_{2.5}$ values. The GWR model is overall better than the MLR model with higher correlations (R^2 = 0.60, CV-R^2 = 0.53) and reduced estimation errors. However, the distribution of monitoring stations and the potential scarcity of observations can limit the use of the GWR model. This model also does not consider time information and cannot capture temporal variations in the time series samples. The first-stage LME model of the two-stage model considers daily variability and shows a notable improvement over the MLR model in terms of modeling (R^2 = 0.67) and validation (CV-R^2 = 0.65). The $PM_{2.5}$-AOD relationship improves with the same R^2 values of 0.71 in modeling and validation when the GWR model, which considers spatial variability, is added as the second stage. However, despite the overall increased accuracy of the two-stage model, the model still cannot simultaneously capture both spatial and temporal heterogeneities. By contrast, the RF model is more robust than these three models in terms of training (R^2 = 0.98) and CV (CV-R^2 = 0.81). By making use of time and space information, the STRF model performs the best among these models according to all evaluation metrics. Note that most models have poor predictive abilities (CV-R^2 < 0.45), while the STRF model shows the best performance in predicting daily $PM_{2.5}$ concentrations (CV-R^2 = 0.55). The comparison results illustrate that the STRF model is robust in modeling, estimating, and predicting $PM_{2.5}$ concentrations through its ability to take both spatial and temporal variabilities into account.

4.5. Comparison with recent studies

Due to the serious air pollution problem in China, especially fine particulate particles, many studies have tried to improve the spatial coverage and resolution of $PM_{2.5}$ concentrations in China. In earlier

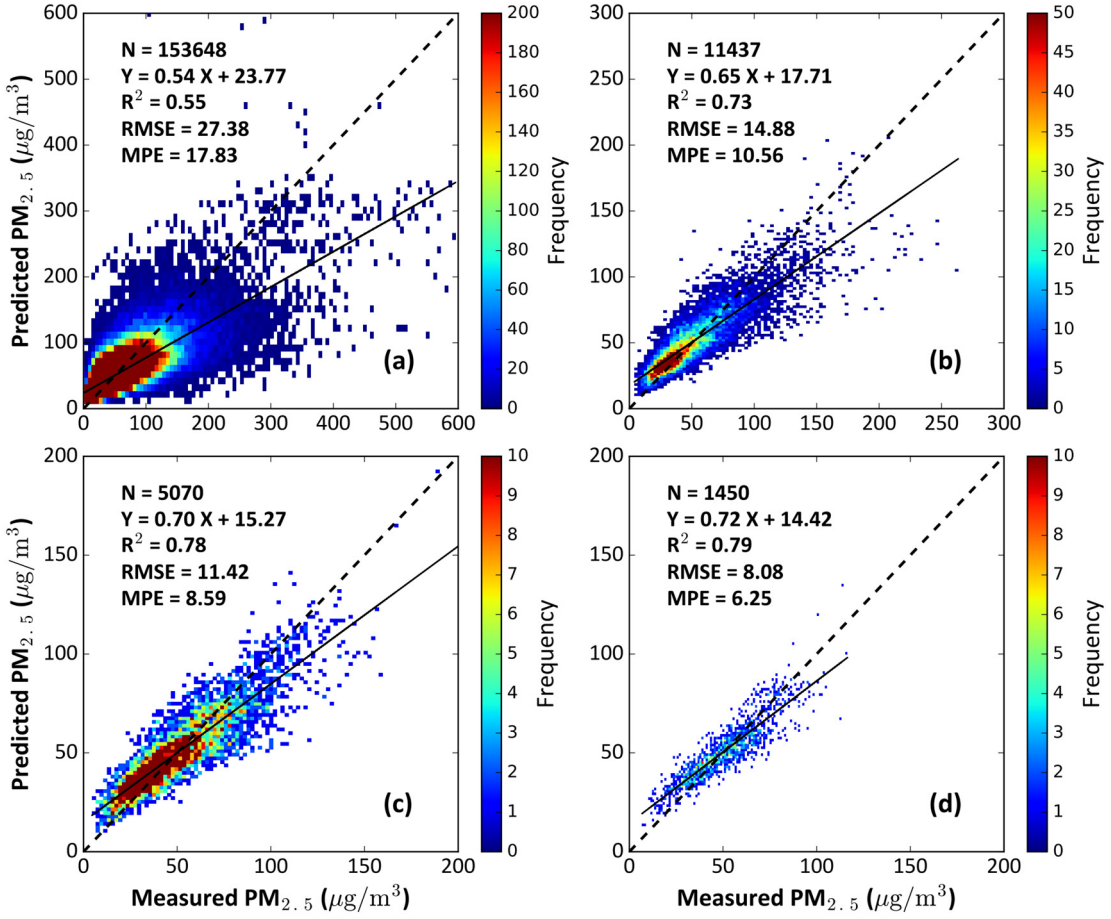

Fig. 7. Estimation and evaluation of predicted $PM_{2.5}$ concentrations on (a) daily, (b) monthly, (c) seasonal, and (d) annual scales in 2016 in China.

studies, due to the limitation of aerosol sources, $PM_{2.5}$ maps of China could only be provided at a spatial resolution of 10 km (Fang et al., 2016; Ma et al., 2014, 2016b; Yu et al., 2017). However, fine particulate particles always occur and are severe in local urban areas with dense populations. Such coarse-resolution $PM_{2.5}$ products are thus not satisfactory. The National Aeronautics and Space Administration (NASA) has recently released a new 3-km AOD product which allows for the creation of $PM_{2.5}$ maps at a higher resolution of 3 km (He and Huang, 2018; Li et al., 2017; You et al., 2016). The latest released MAIAC 1-km aerosol product is used in the current study to improve the spatial coverage of and generate more accurate 1-km $PM_{2.5}$ concentrations.

To be comparable with our study, only those studies on satellite-aerosol-based $PM_{2.5}$ estimations over the whole of China are selected (Table 2). The STRF model can capture 85% of the daily variations in the sample-based CV and can outperform most previous models used for generating 10-km-resolution $PM_{2.5}$ maps of China, e.g., the GWR model (CV-R^2 = 0.64; Ma et al., 2014), the two-stage model (CV-R^2 = 0.78 and 0.79 for the stage-1 LME and stage-2 GWR models; Ma et al., 2016a), the timely structure adaptive modeling (TSAN) model (CV-R^2 = 0.80; Fang et al., 2016), and the Gaussian process model (CV-R^2 = 0.82; Yu et al., 2017). Moreover, the STRF model also surpasses most of the high-spatial-resolution (3 km) models, including the GWR model (CV-R^2 = 0.79; You et al., 2016), the GTWR model (CV-R^2 = 0.80; He and Huang, 2018), and the generalized regression neural network (GRNN) model (CV-R^2 = 0.67; Li et al., 2017). In addition,

Table 1
Model performances and predictive powers of different models.

Model	Model fitting			Model validation			Predictive power		
	R^2	RMSE	MPE	R^2	RMSE	MPE	R^2	RMSE	MPE
MLR	0.41	20.04 y = 0.41 x + 30.02	20.85	0.41	20.04 y = 0.41 x + 30.03	20.85	0.38	21.97 y = 0.41 x + 30.70	22.20
GWR	0.60	22.83 y = 0.62 x + 20.25	18.96	0.53	23.28 y = 0.61 x + 20.93	19.26	0.44	26.47 y = 0.55 x + 23.35	22.23
Two-stage (1)	0.67	19.04 y = 0.66 x + 17.63	15.41	0.65	19.50 y = 0.65 x + 17.63	15.72	0.31	27.73 y = 0.45 x + 26.76	24.03
Two-stage (2)	0.71	18.51 y = 0.71 x + 15.00	14.17	0.71	18.59 y = 0.71 x + 15.10	14.54	0.35	27.65 y = 0.49 x + 25.46	23.30
RF	0.98	6.40 y = 0.92 x + 4.45	4.10	0.81	17.91 y = 0.77 x + 12.56	11.50	0.53	28.09 y = 0.52 x + 24.95	18.43
STRF	0.98	5.57 y = 0.94 x + 3.39	3.48	0.85	15.57 y = 0.82 x + 9.64	9.77	0.55	27.38 y = 0.54 x + 23.77	17.83

Table 2
Statistics for the comparison in performances of different regression models in China.

Related study	Spatial resolution	Model	Model fitting			Model validation			Predictive power		
			R^2	RMSE	MPE	R^2	RMSE	MPE	R^2	RMSE	MPE
Ma et al. (2014)	10 km	GWR	0.71	29.58	19.04	0.64	32.98	21.25	–	–	–
Ma et al. (2016a)	10 km	Stage-1	0.81	26.31	17.59	0.78	27.99	18.67	–	–	–
		Stage-2	0.82	25.74	17.32	0.79	27.42	18.37	0.41	–	–
Fang et al. (2016)	10 km	TSAM	0.82	18.13	13.21	0.80	22.75	15.99	–	–	–
Yu et al. (2017)	10 km	Gaussian	1.00	00.01	–	0.81	21.87	–	–	–	–
You et al. (2016)	3 km	GWR	0.81	17.20	18.50	0.79	18.60	19.60	–	–	–
Li et al. (2017)	3 km	GRNN	0.79	16.51	11.01	0.67	20.93	13.90	–	–	–
He and Huang (2018)	3 km	GTWR	0.85	15.28	10.16	0.80	18.00	12.03	0.47	37.57	24.51
Our study	1 km	RF	0.98	06.40	04.10	0.81	17.91	11.50	0.52	28.80	18.99
Our study	1 km	STRF	0.98	05.57	03.48	0.85	15.57	09.77	0.55	27.38	17.83

only a few studies have examined the model ability in predicting $PM_{2.5}$ records, e.g., the two-stage model (R^2 = 0.41; Ma et al., 2016a) and the GTWR model (R^2 = 0.47; He et al., 2018). The STRF model developed in this study has a stronger predictive power than the other models, having a CV-R^2 of 0.55. The comparison results affirm that the STRF model performs better than most models in estimating and predicting $PM_{2.5}$ concentrations in China.

5. $PM_{2.5}$ spatial distribution across China

5.1. Annual and seasonal maps

Fig. 8 shows the spatial distributions of STRF-modeled, i.e., satellite-estimated on 0.01° × 0.01° grids, and ground-based $PM_{2.5}$ measurements for 2016 in China. The model provides a nearly complete spatial coverage of $PM_{2.5}$ concentrations except for a few areas. The spatial distribution patterns of the 1-km $PM_{2.5}$ estimations from this study and those reported in previous studies are similar except for the wider spatial coverage of the 1-km $PM_{2.5}$ estimations, especially in northwest and southwest China (Fang et al., 2016; He and Huang, 2018; Li et al., 2017; Yu et al., 2017). The satellite-derived $PM_{2.5}$ estimations agree spatially well with surface observations. In general, the annual mean $PM_{2.5}$ concentration is 38.5 ± 12.4 μg m^{-3} with high $PM_{2.5}$ values (> 70 μg m^{-3}) mainly found in the North China Plain, Xinjiang, and the Sichuan Basin. The North China Plain is heavily industrial and densely populated, having undergone rapid economic development which has led to severe emissions of pollutants. Xinjiang Province which includes the Taklimakan Desert experiences frequent sand dust outbreaks. Intensive human activities are the likely cause of pollutants in the Sichuan Basin which are hard to remove because of the special terrain there. By contrast, $PM_{2.5}$ concentrations in southwest and northeast China are generally low and are < 30 μg m^{-3}. This is most likely due to the low anthropogenic aerosol loading there and the favorable meteorological conditions that keep the loading low. Nevertheless, > 55% of China's expanse exceed the standards for the acceptable amount of fine particulate pollution of both the World Health Organization and China's

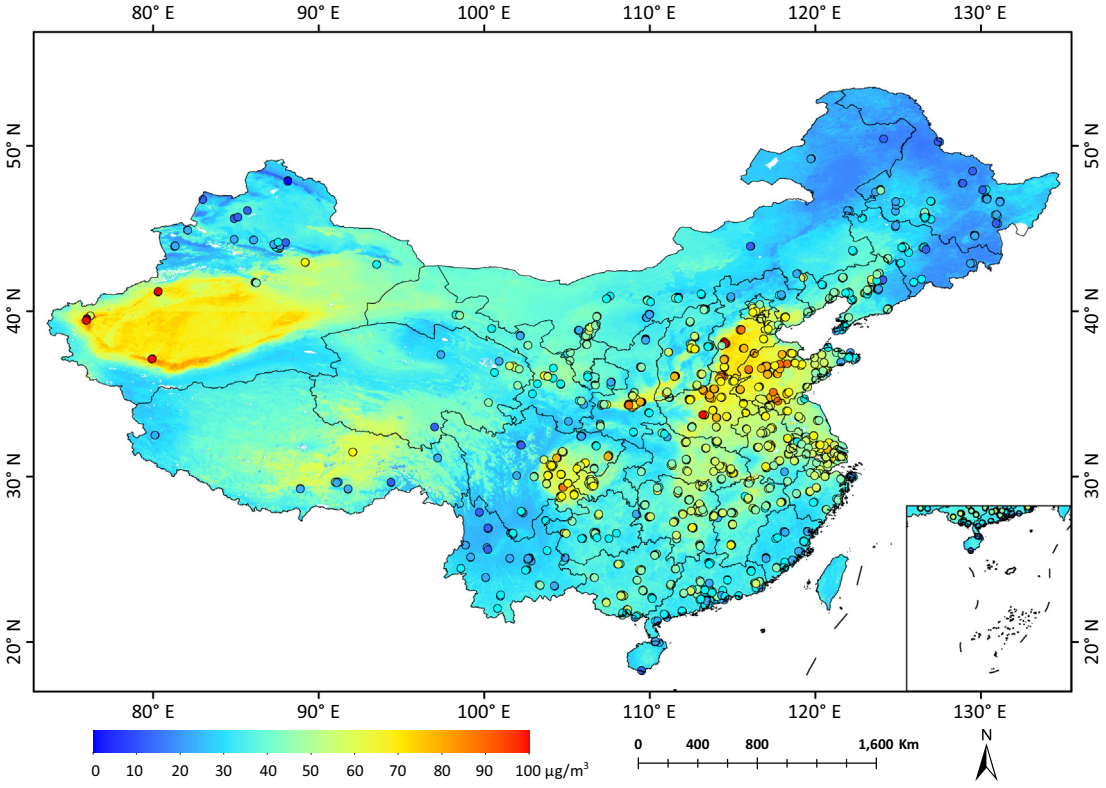

Fig. 8. Spatial distributions of satellite-estimated (0.01° × 0.01° grid, background shading) and ground-measured (colored dots) $PM_{2.5}$ concentrations in 2016 across China.

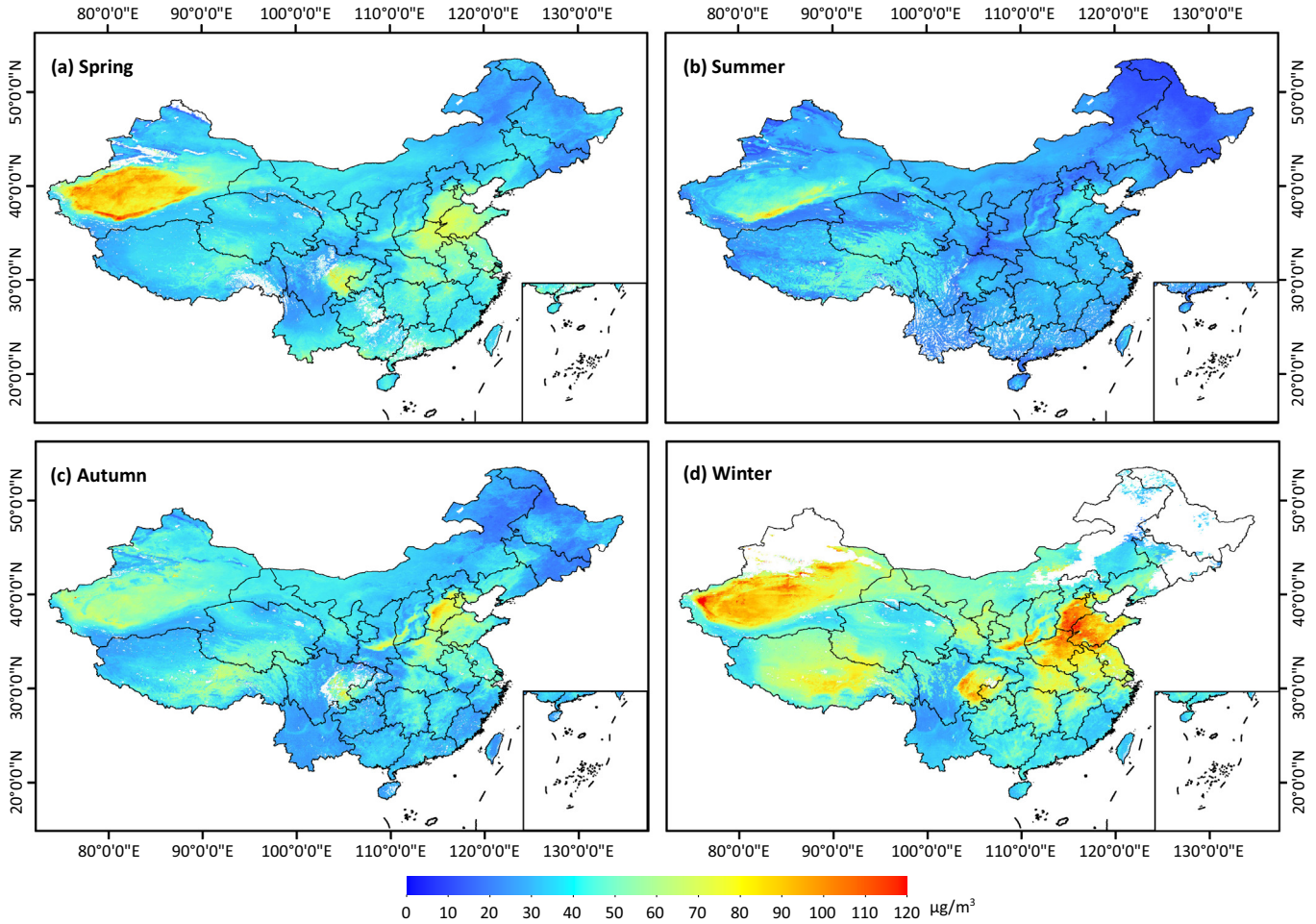

Fig. 9. Seasonal mean distributions of satellite-estimated (0.01° × 0.01° grid, background shading) $PM_{2.5}$ concentrations for (a) spring, (b) summer, (c) autumn, and (d) winter in 2016 across China.

Ministry of Ecology and Environment (i.e., $PM_{2.5} > 35\,\mu g\,m^{-3}$).

Fig. 9 shows seasonal $PM_{2.5}$ spatial distributions across China in 2016. On a seasonal scale, the STRF model can provide an almost complete $PM_{2.5}$ spatial coverage in most areas. However, missing $PM_{2.5}$ values are observed in south-southwest China due to frequent cloud contamination (especially in summer) and in the high-latitude areas of northwest-northeast China due to the long-term snow cover or inadequate observations made in the winter. Spatial differences in fine particulate pollution are noticeable on a seasonal scale. Winter experiences the most severe fine-particulate-matter pollution ($\sim 54.3 \pm 18.6\,\mu g\,m^{-3}$) with > 80% of the country exceeding the national air quality standard. This is mainly due to coal burning and unfavorable weather for the dispersion of pollutants in eastern China and frequent dusty weather in the western deserts. By contrast, the lightest air pollution is found in summer ($\sim 28.1 \pm 8.9\,\mu g\,m^{-3}$) with only ~17% of the country exceeding the air quality standard. The main reason is that the favorable weather conditions, e.g., abundant precipitation, high relative humidity, and abundant water vapor, diffuse atmospheric fine particles. The $PM_{2.5}$ concentration levels in spring and autumn are similar and close to the annual mean with average $PM_{2.5}$ concentrations of $40.1 \pm 15.3\,\mu g\,m^{-3}$ and $36.9 \pm 11.2\,\mu g\,m^{-3}$, respectively.

5.2. Regional $PM_{2.5}$ hotspots

Several heavily polluted areas are selected which are of great concern to the public, including the North China Plain, Yangtze River Delta region, Pearl River Delta region, and Sichuan Basin. Our 1-km higher spatial-resolution $PM_{2.5}$ product enables the study of air pollution on small and medium scales, especially for urban areas. Compared with other studies in these hotspot regions (Ma et al., 2016a, 2016b; Zheng et al., 2016; Guo et al., 2017; He and Huang, 2018; Yao et al., 2018), our product shows similar spatial distributions and variations but can provide more detailed and clearer $PM_{2.5}$ information because the spatial resolution is increased by three to ten times.

Fig. 10a shows that the North China Plain with its high levels of fine particulate matter has become one of the most polluted areas in China. The annual mean $PM_{2.5}$ concentration is $44.08 \pm 23.79\,\mu g\,m^{-3}$, ranging from $5\,\mu g\,m^{-3}$ to $166\,\mu g\,m^{-3}$. Higher $PM_{2.5}$ concentrations are observed over southwest Beijing, southern Hebei, and Henan and Shandong Provinces, mainly the result of intensive human activities (especially fossil fuel burning and straw combustion) and adverse pollutant diffusion conditions (Fang et al., 2016; He and Huang, 2018; Wang et al., 2018). By contrast, the northwest part of the North China Plain has lower $PM_{2.5}$ concentrations due to the dense vegetation coverage and lesser human activities. The Yangtze River Delta region has relatively light particulate-matter pollution with an annual mean $PM_{2.5}$ concentration of $44.1 \pm 7.9\,\mu g\,m^{-3}$. The high $PM_{2.5}$ concentrations in this region are in Shanghai, Jiangsu, and northern Anhui Provinces (Fig. 10b). The Pearl River Delta region has the least amount of atmospheric fine particulate pollution with $PM_{2.5}$ concentrations < $30\,\mu g\,m^{-3}$ in most areas (Fig. 10c) among all developed regions in China. The annual mean $PM_{2.5}$ concentration is $34.5 \pm 4.3\,\mu g\,m^{-3}$. Favorable meteorological conditions, i.e., adequate precipitation and

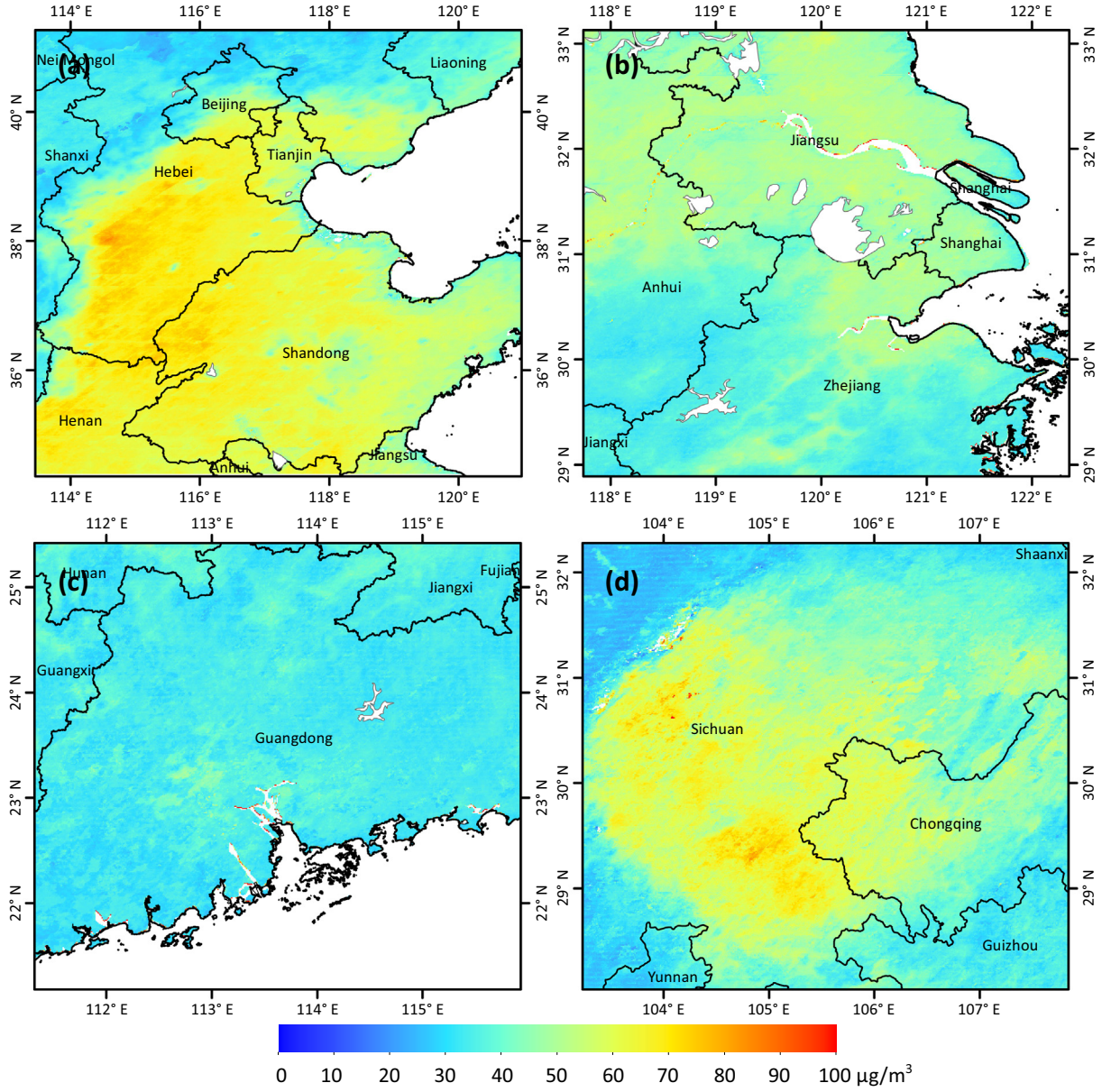

Fig. 10. Zoomed-in map of 1-km annual mean $PM_{2.5}$ concentrations in 2016 for four pollution hotspots: (a) the North China Plain, (b) the Yangtze River Delta region, (c) the Pearl River Delta region, and (d) the Sichuan Basin.

the impact of the monsoon, and less anthropogenic pollutant emissions contribute to this. The Sichuan Basin, however, is a hotpot featuring high levels of fine particulate matter (Fig. 10d) with $PM_{2.5}$ concentrations generally $> 60\,\mu g\,m^{-3}$ with an average mean value of $42.6 \pm 12.0\,\mu g\,m^{-3}$. This is mainly due to the discharge of anthropogenic pollutants and the combination of special topographic (i.e., basin) and meteorological conditions that limit the diffusion of pollutants (Wang et al., 2018; Zhang et al., 2012). These results illustrate that the newly generated 1-km $PM_{2.5}$ product appears to be beneficial for the study of air pollution in small and medium-scale areas.

6. Summary and conclusion

Nowadays, most satellite-derived $PM_{2.5}$ concentrations in China are reported at coarse spatial resolutions (3–50 km) constrained by the input datasets, making it difficult to meet the needs of air pollution monitoring on small- or medium- scales, especially in urban areas. Meanwhile, most traditional models used in estimating $PM_{2.5}$ from satellite face great challenges due to large uncertainties in the AOD-$PM_{2.5}$ relationship as it is influenced by numerous factors that have strong spatial and temporal heterogeneities. To tackle with these challenges, a new space-time random forest (STRF) model is developed that makes use of both the values of the independent input variables and their spatial-temporal variations. The STRF model is able to improve the estimation of $PM_{2.5}$ considerably than the traditional approaches relying on the deterministic relationship between $PM_{2.5}$ observations and satellite-derived AODs. It accounts for the influences of meteorological, land-related, population-related data, and spatio-temporal heterogeneities. The model performance is evaluated using the sample-, spatial-, and temporal-based 10-fold cross-validation (10-CV) approaches, with reference to the widely used regression models (i.e., the MLR, GWR, and two-stage models) for the sake of comparison.

By applying the STRF model to the newly released 1-km MODIS MAIAC AOD products, a satellite-based 1-km $PM_{2.5}$ product was first generated across the mainland China of persistent high accuracy in estimating the daily $PM_{2.5}$ concentrations (model fitting: $R^2 = 0.98$,

RMSE = 5.64 μg m^{-3}, MPE = 3.51 μg m^{-3}, and model validation: R^2 = 0.85, RMSE = 15.57 μg m^{-3}, MPE = 9.76 μg m^{-3}). Meanwhile, the STRF model can more accurately predict the $PM_{2.5}$ concentrations at monthly (R^2 = 0.73), seasonal (R^2 = 0.78), and annual (R^2 = 0.79) scales, which can be used to develop a much longer records of $PM_{2.5}$ than the ground-based $PM_{2.5}$ measurements, as satellite data cover much longer periods. The STRF model is superior in both model performance and predictive power to those of widely used regression models and most models presented in previous studies. The newly generated $PM_{2.5}$ product with its three-to-ten times higher spatial resolution provides a wider spatial coverage than most traditional products. More detailed $PM_{2.5}$ information at small to medium scales across China is thus obtained. These results illustrate that our newly generated 1-km $PM_{2.5}$ product can be useful in air quality studies, especially when the focus in urban areas.

Declaration of Competing Interest

Z. Li designed the research. J. Wei carried out the research and wrote the initial draft. W. Huang performed the model training and validation, and W. Xue compared our results with traditional models. J. Wei and W. Huang made equal contributions to this work. Z. Li, Y. Peng and L. Sun helped with the refinement of this manuscript. All authors contributed to the interpretation of the results.

Acknowledgments

This work was supported by the National Key R&D Program of China (2017YFC1501702), the National Natural Science Foundation of China (91544217), the U.S. National Science Foundation (AGS1534670), and the BNU Interdisciplinary Research Foundation for the First-Year Doctoral Candidates (BNUXKJC1808). The in situ $PM_{2.5}$ ground-measurements used in this study are available from the China National Environmental Monitoring Center (http://www.cnemc.cn). The MODIS series products are available at the Level 1 and Atmosphere Archive and Distribution System (http://ladsweb.nascom.nasa.gov), and the AERONET AOD measurements are available at the Goddard Space Flight Center (https://aeronet.gsfc.nasa). ERA-interim reanalysis data are obtained from the European Centre for Medium-Range Weather Forecasts (http://apps.ecmwf.int/datasets/), and the VIIRS nighttime lights data are available from the National Centers for Environmental Information (https://ngdc.noaa.gov).

Appendix A

Nomenclature

Acronym	Full name	Acronym	Full name
10-CV	10-fold cross-validation	MPE	mean prediction error
AERONET	Aerosol Robotic Network	NDVI	normalized difference vegetation index
AOD	aerosol optical depth	NTL	nighttime lights
BLH	boundary layer height	$PM_{2.5}$	particulate matter with aerodynamic diameters less than 2.5 μm
DB	Deep Blue	PRE	total precipitation
DEM	digital elevation model	R^2	coefficient of determination
DT	Dark Target	RF	random forest
EE	expected error	RH	relative humidity
ET	evaporation	RMSE	root-mean-square error
GRNN	generalized regression neural network	SP	surface pressure
GTWR	geographically and temporally weighted regression	SRTM	Shuttle Radar Topography Mission
GWR	geographically weighted regression	STRF	space-time random forest
LME	linear mixed effect	TEM	air temperature
LUC	land use cover	TSAN	timely structure adaptive modeling
MAE	mean absolute error	VIF	variance inflation factor
MAIAC	Multi-Angle Implementation of Atmospheric Correction	VIIRS	Visible Infrared Imaging Radiometer Suite
MLR	multiple line regression	WD	wind direction
MODIS	Moderate Resolution Imaging Spectroradiometer	WS	wind speed

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2019.111221.

References

Bartell, S.M., Longhurst, J., Tjoa, T., Sioutas, C., Delfino, R.J., 2013. Particulate air pollution, ambulatory heart rate variability, and cardiac arrhythmia in retirement community residents with coronary artery disease. Environ. Health Perspect. 121 (10), 1135–1141. https://doi.org/10.1289/ehp.1205914.

Bi, J., Huang, J., Hu, Z., Holben, B., and Guo, Z. (2014). Investigating the aerosol optical and radiative characteristics of heavy haze episodes in Beijing during January of 2013. Journal of Geophysical Research: Atmospheres, 119(16), 9884–9900. https://doi.org/https://doi.org/10.1002/2014JD021757.

Brauer, M., Amann, M., Burnett, R.T., Cohen, A., Dentener, F., Ezzati, M., Thurston, G.D., 2012. Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. Environ. Sci. Technol. 46 (2), 652–660.

Brokamp, C., Jandarov, R., Hossain, M., Ryan, P., 2018. Predicting daily urban fine particulate matter concentrations using a random forest model. Environ. Sci. Technol. 52 (7), 4173–4179.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Crouse, D.L., Peters, P.A., van Donkelaar, A., Goldberg, M.S., Villeneuve, P.J., Brion, O., Burnett, R.T., 2012. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: a Canadian national-level cohort study. Environ. Health Perspect. 120 (5), 708–714. https://doi.org/10.1289/ehp.1104049.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Vitart, F., 2011. The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597.

Engel-Cox, J.A., Holloman, C.H., Coutant, B.W., Hoff, R.M., 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. Atmos. Environ. 38 (16), 2495–2509.

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.

Guo, J.P., Zhang, X.Y., Che, H.Z., Gong, S.L., An, X., Cao, C.X., Li, X.-W., 2009. Correlation between PM concentrations and aerosol optical depth in eastern China. Atmos. Environ. 43 (37), 5876–5886.

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.

Gupta, P., Christopher, S.A., 2009. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: multiple regression approach. Journal of Atmospheric Research: Atmospheres 114, D14205. https://doi.org/10.1029/2008JD011496.

Gupta, P., Remer, L.A., Levy, R.C., Mattoo, S., 2018. Validation of MODIS 3-km land aerosol optical depth from NASA's EOS Terra and Aqua missions. Atmospheric Measurement Techniques 11, 3145–3159. https://doi.org/10.5194/amt-11-3145-2018.

Hastie, T., Tibshirani, R., Friedman, J., 2008. The Elements of Statistical Learning Data Mining, Inference, and Prediction. World Book Inc.

He, Q., Huang, B., 2018. Satellite-based mapping of daily high-resolution ground $PM_{2.5}$ in China via space-time regression modeling. Remote Sens. Environ. 206, 72–83.

Hsu, N.C., Bettenhausen, C., Sayer, A.M., Hansell, R., Seftor, C.S., Huang, J., 2013. Enhanced deep blue aerosol retrieval algorithm: the second generation. Journal of Geophysical Research: Atmospheres 118 (16), 9296–9315.

Hu, X., Waller, L.A., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G., Estes, S.M., Liu, Y., 2013. Estimating ground-level $PM_{2.5}$ concentrations in the southeastern U.S. using geographically weighted regression. Environ. Res. 121, 1–10.

Hu, X., Waller, L.A., Lyapustin, A., Wang, Y., Liu, Y., 2014. Ten-year spatial and temporal trends of $PM_{2.5}$ concentrations in the southeastern US estimated using high-resolution satellite data. Atmospheric Chemistry & Physics 14 (12), 6301–6314.

Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L., Strickland, M., Liu, Y., 2017. Estimating $PM_{2.5}$ concentrations in the conterminous united states using the random forest approach. Environ. Sci. Technol. 51 (12), 6936–6944.

Huang, J., Liu, J., Chen, B., Nasiri, S.L., 2015. Detection of anthropogenic dust using CALIPSO lidar measurements. Atmos. Chem. Phys. 15, 11,653–11,665. https://doi.org/10.5194/acp-15-11653-2015.

Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyyapustin, A., Gu, D., Liu, Y., 2018. Predicting monthly high-resolution $PM_{2.5}$ concentrations with random forest model in the North China Plain. Environ. Pollut. 242, 675–683.

Just, A.C., Wright, R.O., Schwartz, J., Coull, B.A., Baccarelli, A.A., Tellez-Rojo, M.M., Kloog, I., 2015. Using high-resolution satellite aerosol optical depth to estimate daily $PM_{2.5}$ geographical distribution in Mexico City. Environ. Sci. Technol. 49 (14), 8576–8584.

Lee, H. J., Liu, Y., Coull, B. A., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. *Atmospheric Chemistry and Physics*, 11, 7991–8002. https://doi.org/https://doi.org/10.5194/acp-11-7991-2011.

Levy, R.C., Mattoo, S., Munchak, L.A., Remer, L.A., Sayer, A.M., Patadia, F., Hsu, N.C., 2013. The collection 6 MODIS aerosol products over land and ocean. Atmospheric Measurement Techniques 6, 2989–3034. https://doi.org/10.5194/amt-6-2989-2013.

Li, T., Shen, H., Zeng, C., Yuan, Q., Zhang, L., 2017. Point-surface fusion of station measurements and satellite observations for mapping $PM_{2.5}$ distribution in China: methods and assessment. Atmos. Environ. 152, 477–489.

Li, Z., Zhao, X., Kahn, R., Mishchenko, M., Remer, L., Lee, K.-H., Wang, M., Laszlo, I., Nakajima, T., Maring, H., 2009. Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective. Ann. Geophys. 27, 1–16.

Lin, C.Q., Li, Y., Yuan, Z.B., Lau, A., H, K., Li, C.C., Fung, J.C.H., 2015. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level $PM_{2.5}$. Remote Sens. Environ. 156, 117–128.

Liu, Y., Park, R.J., Jacob, D.J., Li, Q., Kilaru, V., Sarnat, J.A., 2004. Mapping annual mean ground-level $PM_{2.5}$ concentrations using multiangle imaging spectroradiometer aerosol optical thickness over the contiguous United States. Journal of Geophysical Research Atmospheres 109, D22206. https://doi.org/10.1029/2004JD005025.

Liu, Y., Sarnat, J.A., Kilaru, V., Jacob, D.J., Koutrakis, P., 2005. Estimating ground-level $PM_{2.5}$ in the eastern United States using satellite remote sensing. Environ. Sci. Technol. 39 (9), 3269–3278. https://doi.org/10.1021/es049352m.

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 (1), 33–44.

Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I., Korkin, S., 2011a. Multi-Angle Implementation of Atmospheric Correction (MAIAC): 1. Radiative transfer basis and look-up tables. Journal of Geophysical Research: Atmospheres 116, D03210. https://doi.org/10.1029/2010JD014985.

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., …. Reid, J. S. (2011b). Multi-Angle Implementation of Atmospheric Correction (MAIAC): 2. Aerosol algorithm. Journal of Geophysical Research: Atmospheres, 116, D03211. https://doi.org/10.1029/2010JD014986.

Lyapustin, A., Wang, Y., Laszlo, I., Hilker, T., Hall, F., Sellers, P., … Korkin, S., 2012a. Multi-Angle Implementation of Atmospheric Correction for MODIS (MAIAC). 3: atmospheric correction. Remote Sens. Environ. 127, 385–393.

Lyapustin, A., Wang, Y., Laszlo, I., Korkin, S., 2012b. Improved cloud screening in MAIAC aerosol retrievals using spectral and spatial analysis. Atmospheric Measurement Techniques 5, 843–850.

Lyapustin, A., Wang, Y., Korkin, S., Huang, D., 2018. MODIS collection 6 MAIAC algorithm. Atmospheric Measurement Techniques 11, 5741–5765.

Ma, Z., Hu, X., Huang, L., Bi, J., Liu, Y., 2014. Estimating ground-level $PM_{2.5}$ in China using satellite remote sensing. Environ. Sci. Technol. 48 (13), 7436–7444. https://doi.org/10.1021/es5009399.

Ma, Z., Hu, X., Sayer, A., Levy, R., Zhang, Q., Liu, Y., 2016a. Satellite-based spatiotemporal trends in $PM_{2.5}$ concentrations: China, 2004–2013. Environ. Health Perspect. 124 (2), 184–192. https://doi.org/10.1289/ehp.1409481.

Ma, Z., Liu, Y., Zhao, Q., Liu, M., Zhou, Y., Bi, J., 2016b. Satellite-derived high resolution $PM_{2.5}$ concentrations in Yangtze river delta region of China using improved linear mixed effects model. Atmos. Environ. 133, 156–164.

Pang, J., Liu, Z., Wang, X., Bresch, J., Ban, J., Chen, D., Kim, J., 2018. Assimilating AOD retrievals from GOCI and VIIRS to forecast surface $PM_{2.5}$ episodes over eastern China. Atmos. Environ. 179, 288–304.

Pascal, M., Falq, G., Wagner, V., 2014. Short-term impacts of particulate matter (PM_{10}, $PM_{10-2.5}$, $PM_{2.5}$) on mortality in nine French cities. Atmos. Environ. 95, 175–184.

Peng, R.D., Bell, M.L., Geyh, A.S., McDermott, A., Zeger, S.L., Samet, J.M., Dominici, F., 2009. Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. Environ. Health Perspect. 117 (6), 957–963.

Rodriguez, J.D., Perez, A., Lozano, J.A., 2010. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. 32 (3), 569–575.

Rohde, R.A., Muller, R.A., 2015. Air pollution in China: mapping of concentrations and sources. PLoS One 10 (8), e0135749.

Sorek-Hamer, M., Strawa, A.W., Chatfield, R.B., Esswein, R., Cohen, A., Broday, D.M., 2013. Improved retrieval of $PM_{2.5}$ from satellite data products using non-linear methods. Environ. Pollut. 182, 417–423.

Su, T., Li, Z., Kahn, R., 2018. Relationships between the planetary boundary layer height and surface pollutants derived from lidar observations over China: regional pattern and influencing factors. Atmos. Chem. Phys. 18 (21), 15,921–15,935.

Sun, L., Wei, J., Duan, D., Guo, Y., Mi, X., 2016. Impact of land-use and land-cover change on urban air quality in representative cities of China. J. Atmos. Sol. Terr. Phys. 142, 43–54.

Wang, X., Dickinson, R., Su, L., Zhou, C., Wang, K., 2018. $PM_{2.5}$ pollution in China and how it has been exacerbated by terrain and meteorological conditions. Bull. Am. Meteorol. Soc. 99 (1), 105–119.

Wei, J., Sun, L., Huang, B., Bilal, M., Zhang, Z., Wang, L., 2018. Verification, improvement and application of aerosol optical depths in China. Part 1: inter-comparison of NPP-VIIRS and Aqua-MODIS. Atmos. Environ. 175, 221–233.

Wei, J., Li, Z., Peng, Y., Sun, L., 2019a. MODIS collection 6.1 aerosol optical depth products over land and ocean: validation and comparison. Atmos. Environ. 201, 428–440.

Wei, J., Li, Z., Sun, L., Peng, Y., Wang, L., 2019b. Improved merge schemes for MODIS collection 6.1 dark target and deep blue combined aerosol products. Atmos. Environ. 202, 315–327.

Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., Liu, Y., 2017. Full-coverage high-resolution daily $PM_{2.5}$ estimation using MAIAC AOD in the Yangtze River Delta of China. Remote Sens. Environ. 199, 437–446.

Yao, F., Si, M., Li, W., Wu, J., 2018. A multidimensional comparison between MODIS and VIIRS AOD in estimating ground-level $PM_{2.5}$ concentrations over a heavily polluted region in China. Sci. Total Environ. 618, 819–828. https://doi.org/10.1016/j.scitotenv.2017.08.209.

You, W., Zang, Z., Pan, X., Zhang, L., Chen, D., 2015. Estimating $PM_{2.5}$ in Xi'an, China using aerosol optical depth: a comparison between the MODIS and MISR retrieval models. Sci. Total Environ. 505, 1156–1165.

You, W., Zang, Z., Zhang, L., Li, Y., Pan, X., Wang, W., 2016. National-scale estimates of ground-level $PM_{2.5}$ concentration in China using geographically weighted regression based on 3-km resolution MODIS AOD. Remote Sens. 8 (3), 184. https://doi.org/10.3390/rs8030184.

Yu, W., Liu, Y., Ma, Z., Bi, J., 2017. Improving satellite-based $PM_{2.5}$ estimates in China using Gaussian processes modeling in a Bayesian hierarchical setting. Sci. Rep. 7 (1). https://doi.org/10.1038/s41598-017-07478-0.

Zang, L., Mao, F., Guo, J., Gong, W., Wang, W., Pan, Z., 2018. Estimating hourly PM_1 concentrations from Himawari-8 aerosol optical depth in China. Environ. Pollut. 241, 654–663.

Zhang, H., Li, J., Ying, Q., Yu, J.Z., Wu, D., Cheng, Y., ... Jiang, J., 2012. Source apportionment of $PM_{2.5}$ nitrate and sulfate in China using a source-oriented chemical transport model. Atmos. Environ. 62 (7), 228–242.

Zhang, Y., Li, Z., 2015. Remote sensing of atmospheric fine particulate matter ($PM_{2.5}$) mass concentration near the ground from satellite observation. Remote Sens. Environ. 160, 252–262.

Zheng, Y., Zhang, Q., Liu, Y., Geng, G., He, K., 2016. Estimating ground-level $PM_{2.5}$ concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. Atmos. Environ. 124, 232–242.