



## Estimating PM<sub>2.5</sub> with high-resolution 1-km AOD data and an improved machine learning model over Shenzhen, China

Wenqian Chen <sup>a,b,c</sup>, Haofan Ran <sup>d</sup>, Xiaoyi Cao <sup>d</sup>, Jingzhe Wang <sup>e</sup>, Dexiong Teng <sup>d</sup>, Jing Chen <sup>b</sup>, Xuan Zheng <sup>a,\*</sup>

<sup>a</sup> College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen 518060, China

<sup>b</sup> Geography Department, Hanshan Normal University, Chaozhou 521041, China

<sup>c</sup> College of Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China

<sup>d</sup> College of Resources and Environment Sciences, Xinjiang University, Urumqi 830046, Xinjiang, China

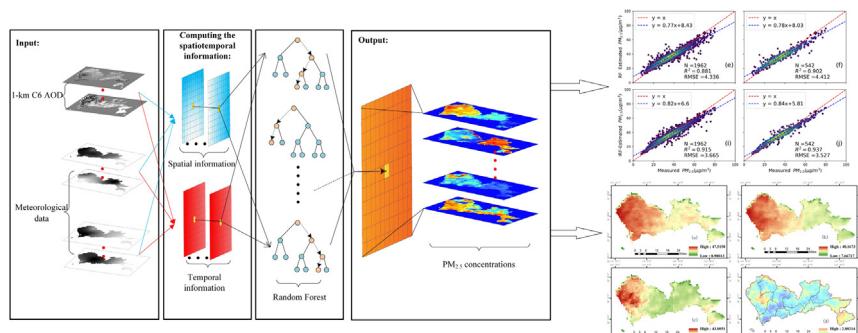
<sup>e</sup> MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area & Guangdong Key Laboratory of Urban Informatics & Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China



### HIGHLIGHTS

- A new prediction method for PM<sub>2.5</sub> based on improved machine learning is proposed.
- IRF model can express varied PM<sub>2.5</sub>-AOD associations in time and space.
- High resolution PM<sub>2.5</sub> concentrations are estimated in coastal urban scale.
- Derived PM<sub>2.5</sub> products have multiple potential applications with high accuracy.

### GRAPHICAL ABSTRACT



### ARTICLE INFO

#### Article history:

Received 15 April 2020

Received in revised form 22 June 2020

Accepted 18 July 2020

Available online 21 July 2020

Editor: Jianmin Chen

#### Keywords:

PM<sub>2.5</sub>

Aerosol optical depth

Machine learning

Improved random forest

Urban area

### ABSTRACT

Studies on fine particulate matter with an aerodynamic diameter of 2.5 μm or smaller (PM<sub>2.5</sub>) are closely related to the atmospheric environment and human activities but are often limited by ground-level in situ observations. Satellite remote sensing techniques have been widely used to estimate the PM<sub>2.5</sub> concentration over large areas where ground-monitoring sites are unavailable. However, satellite-retrieved aerosol optical depth (AOD) products usually feature a coarse resolution, which is insufficient for the estimation of the urban-scale PM<sub>2.5</sub> concentration. We developed a new improved random forest (IRF) model based on machine learning and a newly released AOD product with a high resolution of 1-km, which could more effectively and accurately estimate the PM<sub>2.5</sub> concentration over Shenzhen in the Guangdong-Hong Kong-Macao Greater Bay Area (GBA), China. Daily PM<sub>2.5</sub> concentrations from 2016 to 2018 were estimated from ground-level PM<sub>2.5</sub> and meteorological variable data. The popular linear regression model, geographically and temporally weighted regression (GTWR) model and random forest (RF) model without spatiotemporal information were employed for comparison and validation purposes through the 10-fold cross-validation (CV) approach. The IRF model attained an overall R<sup>2</sup> value of 0.915 and a root mean square error (RMSE) value of 3.66 μg m<sup>-3</sup>. This suggests that the IRF model can estimate the urban PM<sub>2.5</sub> concentration with a high spatial resolution at the daily, seasonal and annual scales, and the improved machine learning method is better than the linear model proposed by previous studies in terms of the estimation accuracy of the PM<sub>2.5</sub> concentration. Generally, the IRF model coupled with AOD data with a 1-km resolution can significantly improve the calculation accuracy of the atmospheric PM<sub>2.5</sub> concentration over coastal urban areas in the future.

© 2020 Elsevier B.V. All rights reserved.

\* Corresponding author.

E-mail address: [zhengxuanqh@163.com](mailto:zhengxuanqh@163.com) (X. Zheng).

## 1. Introduction

The air pollution caused by fine particulate matter with an aerodynamic diameter of  $2.5 \mu\text{m}$  or smaller ( $\text{PM}_{2.5}$ ) has become very serious in China in recent years (Lv et al., 2017; Yang et al., 2019b). It is also known to be closely associated with cardiovascular and respiratory diseases because  $\text{PM}_{2.5}$  can penetrate human lungs and bronchi (Chen et al., 2019). However, in 2013, the National Air Quality Monitoring Network was established to monitor  $\text{PM}_{2.5}$  in China (Wei et al., 2019). However, its monitoring sites do not completely cover the entire area of China. Satellites are a powerful tool for  $\text{PM}_{2.5}$  estimation in areas where ground-monitoring sites are unavailable, and research has adopted satellite-retrieved aerosol optical depth (AOD) data to estimate the  $\text{PM}_{2.5}$  concentration due to their wide coverage (Hutchison et al., 2005; Zang et al., 2017; Liang et al., 2020; Zhang et al., 2018).

Early studies have mainly focused on the AOD- $\text{PM}_{2.5}$  relationship by using statistical models (e.g., linear or multiple regression models), which are faster and have simpler characteristics than chemical, physical, and semiempirical models (Lin et al., 2015; Liu et al., 2007; Yang et al., 2019a; Zhang and Li, 2015). In these studies, the various approaches have primarily focused on the linear AOD and  $\text{PM}_{2.5}$  relationship, and the  $R^2$  value for the predicted  $\text{PM}_{2.5}$  concentration has ranged from 0.42 to 0.62, including the results after vertical and relative humidity (RH) correction of AOD (Rolandone et al., 2003; Wang et al., 2010). More advanced statistical models have been proposed to improve the prediction accuracy by considering meteorological or land use information (Bernardo, 2013; Liu et al., 2009). For example, the linear mixed effects (LME) generalized additive model (GAM), geographically weighted regression (GWR) model, hierarchical model, Bayesian model, and GAM coupled AOD and meteorological parameters with the  $\text{PM}_{2.5}$  concentration (Liu et al., 2017; Song et al., 2015; Zhai et al., 2018). Hu et al. incorporated AOD, meteorological parameters, and land use and cover change (LUCC) in the GWR model, and the model  $R^2$  value reached 0.82 (Hu et al., 2013). These models yielded higher  $R^2$  values than the linear and multiple regression models, but they did not consider the daily variations in certain parameters (e.g., wind and RH) over time, which affects the estimation accuracy when applied in different regions, and the model predictions are unstable (Lee et al., 2011; Yang et al., 2019b).

In general, statistical models have achieved suitable effects in some areas, but they cannot effectively solve the complex nonlinear relationship between the dependent and predictor variables and rarely consider random daily variation effects, resulting in unstable simulation results (Di et al., 2016; Li et al., 2016). Compared to statistical models, machine learning (often called data mining) has become a popular approach to resolve this and many other complex problems because of its superior ability to select and employ many independent factors that may affect the dependent variable to be predicted, thereby providing better air pollution forecasts (Hu et al., 2017; Lv et al., 2016). Xue et al. adopted a geographically weighted gradient boosting machine model to predict the  $\text{PM}_{2.5}$  concentration over China in 2014, which effectively increased the  $R^2$  value (Xue et al., 2019). Chen et al. (2018) and Hu et al. (2017) applied the random forest (RF) model to predict the  $\text{PM}_{2.5}$  exposure level in China and the United States, respectively (Chen et al., 2018; Hu et al., 2017).

The improvement of the AOD resolution is an important factor influencing the model prediction effect of  $\text{PM}_{2.5}$  (van Donkelaar et al., 2011). Previous studies have demonstrated that the Moderate Resolution Imaging Spectroradiometer (MODIS) 3-km AOD product provides more detailed information than the AOD products with a resolution of 10-km or coarser (Munchak et al., 2013; Yang et al., 2019b). According to the work of Ma et al., the annual average  $\text{PM}_{2.5}$  concentration over nearly 81.8% of the areas in Chinese cities exceeds  $35 \text{ mg/m}^3$ , including residential, cultural and commercial areas (Ma et al., 2014). Therefore, the prediction of  $\text{PM}_{2.5}$  at the urban scale is extremely important and requires higher-resolution AOD products. In 2018, a new global-coverage

high-spatial resolution MODIS AOD (1-km) product was released, which has not yet been thoroughly examined for  $\text{PM}_{2.5}$  estimation (Lyapustin et al., 2018).

In the current study, we examine the use of the above MODIS 1-km AOD product in Shenzhen, an important city in the Guangdong-Hong Kong-Macao Greater Bay Area (GBA), where pollutants are greatly affected by meteorological conditions, and daily meteorological fields are adopted as ancillary variables. The GTWR model was applied to calibrate the daily variation in parameters over time, and based on machine learning procedures, an improved random forest (IRF) model was developed to address the spatial heterogeneity and temporal variations in the emissions and meteorological variables. To our knowledge, this is the first study that utilizes MODIS 1-km AOD meteorological parameters with an improved machine learning model for  $\text{PM}_{2.5}$  estimation in Shenzhen.

## 2. Data and methodology

### 2.1. Study area

Shenzhen (22°33'N, 114°06'E) is a coastal city located at the mouth of the GBA near Hong Kong in southern Guangdong Province in China and is home to a population of 13.66 million residents (Fig. 1). Due to its high industrial output, it is the most important special economic zone in China. In recent years, Shenzhen has experienced elevated levels of particulate matter pollution because of its rapid economic development.  $\text{PM}_{2.5}$  has been measured in a suburb of Shenzhen, and a previous study has demonstrated that the mean  $\text{PM}_{2.5}$  concentration is  $101.6 \pm 27.5 \text{ } \mu\text{g m}^{-3}$  in winter and  $32.7 \pm 19.7 \text{ } \mu\text{g m}^{-3}$  in summer, which exceed the 24-h mean ambient air quality standard of the World Health Organization (WHO) of  $25 \text{ } \mu\text{g m}^{-3}$  (Dai et al., 2013). The estimation of the spatial and temporal  $\text{PM}_{2.5}$  concentrations over Shenzhen is of great significance to the atmospheric environment and public health in the GBA.

### 2.2. Data

#### 2.2.1. Ground-level $\text{PM}_{2.5}$ monitoring

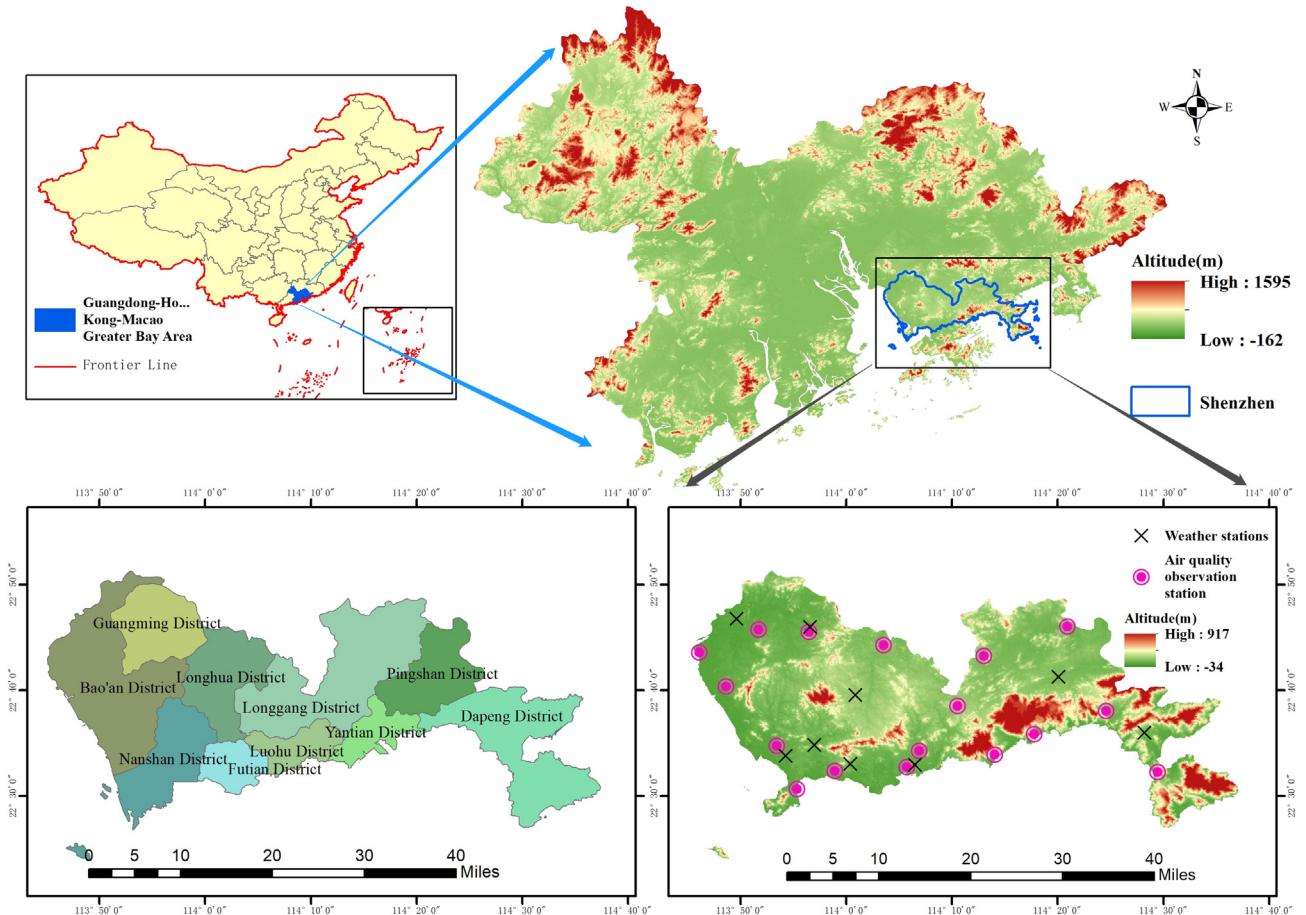
Calibrated and quality-controlled hourly  $\text{PM}_{2.5}$  concentration measurements from 2016 to 2018 were obtained from the China Environmental Monitoring Center (CEMC) (<http://106.37.208.233:20035/>). To match the AOD data obtained from satellites, we used the data collected from 10:00 to 12:00 local time at 19  $\text{PM}_{2.5}$  monitoring sites distributed across the study area (Fig. 1).

#### 2.2.2. MODIS AOD data

The newly released Terra and Aqua MODIS C6 daily 1-km MCD19A2 products at a wavelength of 550 nm from 2016 to 2018 are employed in this study (<https://ladsweb.modaps.eosdis.nasa.gov/>). The data used in this study have already been calibrated by cloud screening constraint and adjacency tests (Lyapustin et al., 2018). These products are based on the new multiangle implementation of atmospheric correction (MAIAC) algorithm, which is different from the Dark Target and Deep Blue algorithms, thereby employing the semiempirical Ross-thick/Li-sparse bidirectional reflectance distribution function and semianalytical Green function solution model, respectively (Lyapustin et al., 2011). After analysis based on time series images, many key steps, such as cloud screening and aerosol type selection, have been improved in the new products over the other products. The satellite observes the Shenzhen area between 10:00 and 12:00 local time, and we obtained the AOD and  $\text{PM}_{2.5}$  measurements simultaneously.

#### 2.2.3. Meteorological parameters

The meteorological data used in this study were retrieved from the China Meteorological Data Service Center (<http://data.cma.cn/en>) and Shenzhen Meteorological Bureau (<http://weather.sz.gov.cn>), and these data were calculated by the kriging interpolation method (stochastic



**Fig. 1.** The study area and the locations of the  $\text{PM}_{2.5}$  monitoring and weather stations.

spatial interpolation process according to the covariance function) in ArcGIS10.3 software to obtain a surface-scale meteorological data set with a resolution of 1-km in Shenzhen from 2016 to 2018.  $\text{PM}_{2.5}$  imposes a strong random effect in space and time, and we selected two parameters with the greatest impact on the  $\text{PM}_{2.5}$  concentration over Shenzhen, including RH (%) and extreme wind speed (EWS, m/s). The daily means of the meteorological parameters from 10:00 to 12:00 local time coincident with the satellite-retrieved AOD and ground-level  $\text{PM}_{2.5}$  monitoring data were used (Fig. 1).

### 2.3. Model development and adjustment

#### 2.3.1. GTWR model

The  $\text{PM}_{2.5}$  concentration varies dramatically in space and time due to the vast geographical areas, complex surface structures, and human disturbances in Shenzhen. We want to develop an as simple as possible prediction model to reduce the complexity of  $\text{PM}_{2.5}$  estimation, improve the generalization ability of the model and increase its estimation accuracy. Therefore, only three parameters with the greatest impact on the  $\text{PM}_{2.5}$  concentration, including AOD, RH and EWS, are selected. To compare the stability of  $\text{PM}_{2.5}$  prediction between linear models and the improved machine learning method, the GTWR model is also run on the R language platform. In contrast to the widely used GWR model, which only considers spatial variations when estimating an empirical relationship, the GTWR model captures the spatiotemporal heterogeneity based on a weighting matrix referencing both the spatial and temporal dimensions and is defined as follows:

$$\text{PM}_{2.5}(x_i, y_i, t_i) = a(x_i, y_i, t_i) + b(x_i, y_i, t_i) \text{AOD}(x_i, y_i, t_i) + \text{eps}(x_i, y_i, t_i) \quad (1)$$

where  $x_i$ ,  $y_i$ , and  $t_i$  are the latitude, longitude and time, respectively, of site  $i$ ;  $a$  and  $b$  are the coefficients of  $n$  explanatory variables; and  $\text{eps}$  is the model residual.

#### 2.3.2. Improved random forest model

The traditional RF model is a relatively flexible machine learning approach with applications in different fields. It is composed of multiple decision trees that employ the tree branch structure to achieve classification and can produce an unbiased result because it evaluates the importance of each feature in the classification process. Moreover, the RF method is easy to implement and can process thousands of input variables efficiently, resulting in relatively low computational costs.

However, the applicability of the RF model depends on the number of input independent variables, and 99.9% of irrelevant trees results in predictions that cover all situations, regardless of the spatiotemporal information. Moreover, as noted in a previous study, there have been few applications of this model to estimate  $\text{PM}_{2.5}$  based on remote sensing data. The precision of  $\text{PM}_{2.5}$  estimation is affected by the spatiotemporal heterogeneity, and many researchers have attempted to solve these spatiotemporal problems, resulting in the development of well-known models such as the GWR and GTWR models. To further improve the accuracy of  $\text{PM}_{2.5}$  estimation, we consider adding information on the temporal and spatial changes in remote sensing products to improve the traditional RF model.

In the RF model, for each tree,  $Z$  samples ( $N$ ) are extracted from the training data using the bootstrap sample method. A feature subset is randomly selected from  $M$ -dimensional features. The optimal value of

M is a key parameter for the construction of the RF model and can be determined by trial and 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 (a sample of the k-th tree) and each regression tree  $[f(x)]$  as follows:

$$f(x) = \sum_{z=1}^Z c_z I(x \in R_z) \quad (2)$$

$$\hat{c}_z = \text{mean}(y_i | x_i \in R_z) \quad (3)$$

$$Z_1(m, n) = \{X | X_j \leq n\} \text{ and } Z_2(m, n) = \{X | X_j > n\} \quad (4)$$

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

and

$$\hat{c}_1 = \text{mean}(y_i | x_i \in R_1(m, n)) \text{ and } \hat{c}_2 = \text{mean}(y_i | x_i \in R_2(m, n)) \quad (5)$$

where  $(x_i, y_i)$  is the sample location in  $i = 1, 2, \dots, N$  out of  $Z$  regions ( $R_1, R_2, \dots, R_Z$ );  $c_m$  is the response to the model, which is a constant;  $\hat{c}_1$  is the optimal value;  $m$  is the splitting variable; and  $n$  is the split point.

Remote sensing data have rich and continuous spatiotemporal features that can be added to the model, and we also considered the spatiotemporal information of meteorological data. For a given pixel of the AOD and meteorological data, its spatial (s) and temporal (t) information can be expressed as:

$$s = \frac{\sum_{i=1}^n \frac{1}{ds_i^2} s_i}{\sum_{i=1}^l \frac{1}{ds_i^2}} \quad (6)$$

$$t = \frac{\sum_{j=1}^m \frac{1}{dt_j^2} t_j}{\sum_{j=1}^k \frac{1}{dt_j^2}} \quad (7)$$

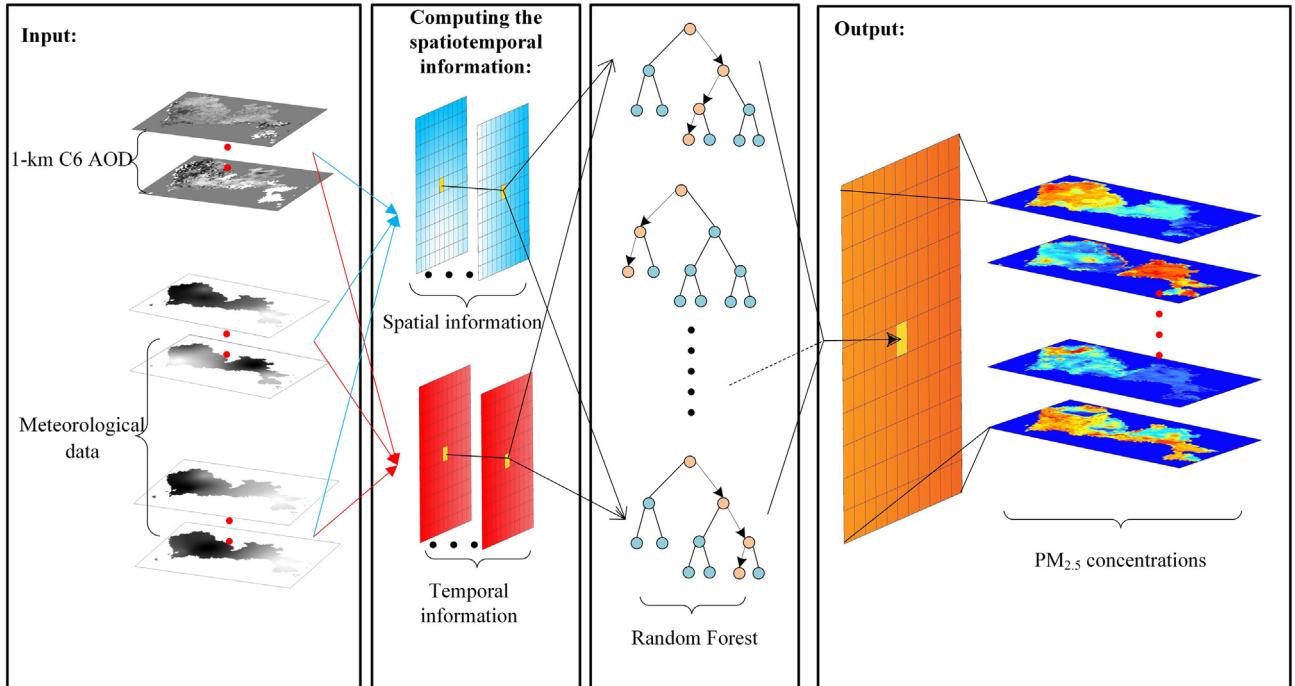
where  $ds$  and  $dt$  are the spatiotemporal distances of adjacent pixels from a central pixel. The term  $n$  represents the  $i$  pixels near the ground site, and  $m$  represents the  $j$  prior days for the same pixel in the AOD image. The spatiotemporal information associated with the satellite-retrieved AOD and meteorological data are used as inputs to develop the improved model of the AOD-PM<sub>2.5</sub> relationship.

The structure and specific schematics of the IRF model are shown in Fig. 2. In the IRF model, decisions are made based on data characteristics. According to the data characteristics, the most suitable splitting method is implemented. Each sample has an associated optimal splitting method, which is the red chain in Fig. 2 (e.g., the first node represents the AOD values larger than 500, and the second node represents the EWS values higher than 90, as shown in Fig. 2). According to these nodes, the predicted PM<sub>2.5</sub> concentration of the other branches is compared to the true concentration at the same time. The red branch with the smallest error is defined as the optimal cutting branch, and the blue branch exhibits the largest error. The above steps are repeated to determine the best decision chain, and the predicted result of the best decision chain is the predicted value of the IRF model.

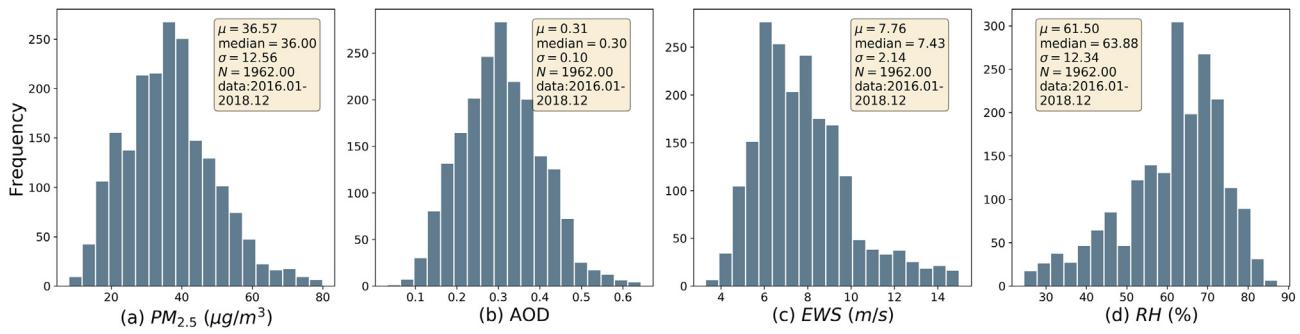
### 2.3.3. Evaluation approaches

In this study, the common 10-fold cross-validation (CV) method and several statistical indicators, including the root mean square error (RMSE), are used to validate the performance of the GTWR and IRF models in this study, and the two-sided test is applied to validate the statistical significance.

The CV method can be described as follows: the data are randomly divided into 10 subsets, 1 of which is selected as the validation subset, while the remaining 9 subsets are the training subsets. Finally, the coefficient of determination ( $R^2$ ) and RMSE between the estimated and measured values are determined to evaluate the model performance. In addition, sufficient average daily PM<sub>2.5</sub> values for each grid over a year are generated to produce yearly PM<sub>2.5</sub> maps. Seasonal maps are generated from the monthly PM<sub>2.5</sub> estimates. The trend changes in PM<sub>2.5</sub> are calculated from the monthly PM<sub>2.5</sub> data after the elimination of seasonal variations and anomalies.



**Fig. 2.** The structure and schematics of the model.



**Fig. 3.** Histograms and descriptive statistics of the independent model variables (mean, median and standard deviation).

### 3. Results and discussion

#### 3.1. Descriptive statistics

When there were fewer than 3 AOD-PM<sub>2.5</sub> matches for a given day, we excluded all the AOD and PM<sub>2.5</sub> observations groups. Finally, during the study period, Shenzhen had a total of 118 valid days, and the descriptive statistics of the data for model fitting are shown in Fig. 3. The annual PM<sub>2.5</sub> concentration was  $39.5 \mu\text{g m}^{-3}$ , and the average AOD was 0.31. The average RH and EWS were 61.5% and 7.7 m/s, respectively, reflecting the monsoon and humid climate characteristics prevailing in the study area. Due to the climate characteristics of Shenzhen, the seasonal changes are not distinct, but there are two precipitation-based seasons, the dry season (October to March) and the wet season (April to September). We analyzed the various parameters and PM<sub>2.5</sub> estimates for the dry and wet seasons.

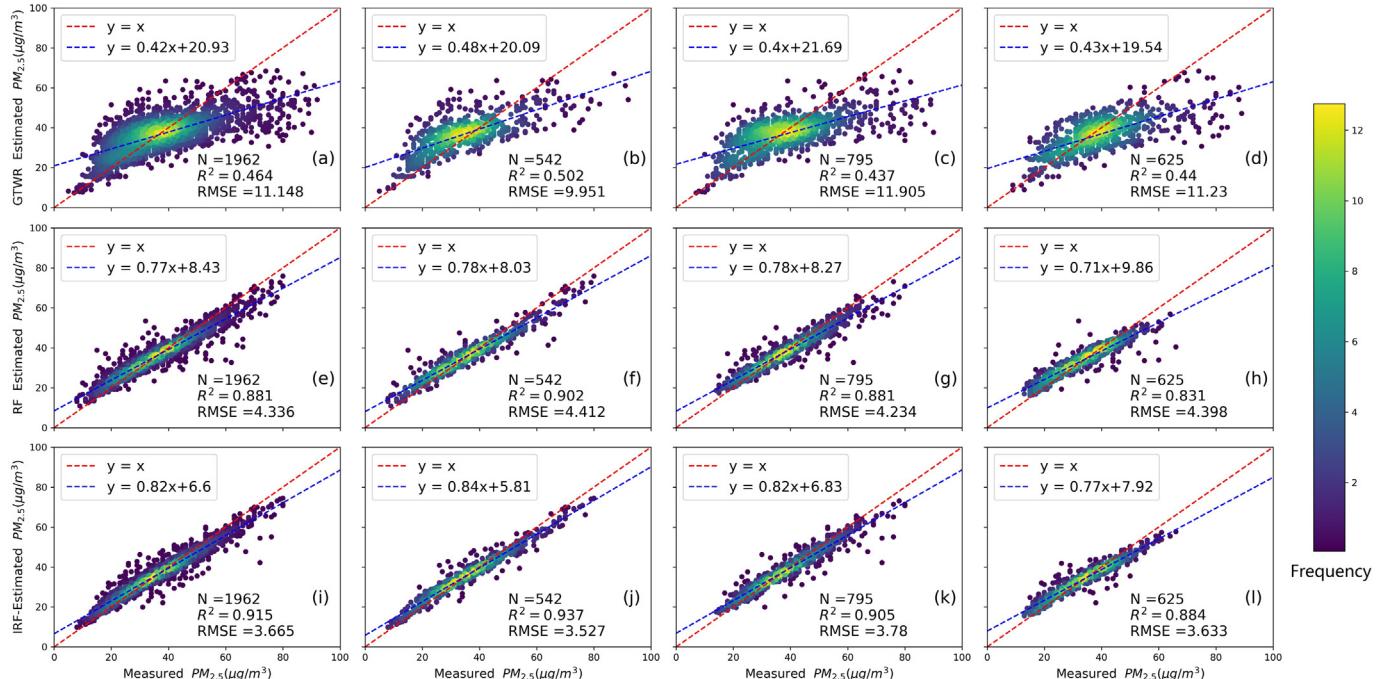
#### 3.2. Model fitting and validation

##### 3.2.1. Annual-scale model performance

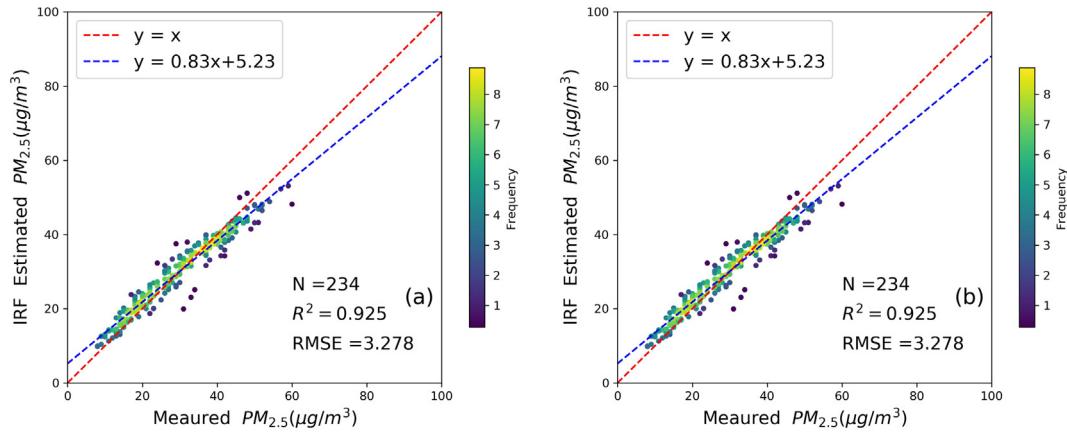
The main objective of this work is to validate the PM<sub>2.5</sub> estimation performance of the IRF model over the Shenzhen. Fig. 4 shows

frequency scatter plots of the fit and 10-fold CV results of the GTWR, RF and improved RF models for Shenzhen from 2016 to 2018. The GTWR model, which is a widely applied linear model in PM<sub>2.5</sub> estimation, is adopted as a benchmark to assess the accuracy improvement between the RF and IRF models with a high capacity of the nonlinear method. The overall R<sup>2</sup> and RMSE values of the two models are shown in Fig. 4. As shown in the figure, for the linear GTWR model, the total R<sup>2</sup> value between the estimated and actual PM<sub>2.5</sub> values is 0.46, with an RMSE value of  $11.14 \mu\text{g m}^{-3}$ . For the RF model without spatial and temporal data, the R<sup>2</sup> value is 0.88, and the RMSE value is  $4.33 \mu\text{g m}^{-3}$ . The consistency between the estimated and measured values of the IRF model is high, with R<sup>2</sup> values of 0.93, 0.90 and 0.88 for 2016, 2017, and 2018, respectively, which are higher than the values of the GTWR model. The RMSE value decreases from 11.14 to  $3.66 \mu\text{g m}^{-3}$ , which indicates that the prediction ability of the model has improved by adding spatial and temporal information to the data as auxiliary parameters.

According to the model fitting results, the RF and IRF models perform well, with an R<sup>2</sup> value of 0.98, when the models are suitably trained. The CV method evaluates the results from a spatial perspective, and the accuracy of these two models is relatively lower than that when the models are trained because the PM<sub>2.5</sub> concentration is unevenly



**Fig. 4.** Estimation and evaluation of the predicted PM<sub>2.5</sub> concentrations based on the cross-validation results for the GTWR (a-d), traditional RF (e-h) and IRF (i-l) models from 2016 to 2018.



**Fig. 5.** Prediction of the IRF model during the dry and wet seasons from 2016 to 2018 over Shenzhen.

distributed in space. However, we find that the RF model is still not as good as the IRF model at the given spatial scale from 2016 to 2018. In general, the temporal accuracy of the two models is much higher than the spatial accuracy, and considering both the geographic spatial relevance and temporal variation will improve the model performance. Overall, the linear AOD-PM<sub>2.5</sub> relationship cannot explain the prediction results well. Compared to the machine learning method, the GTWR model attains the worst effect. Based on the model fitting results, the IRF model also underestimates the PM<sub>2.5</sub> concentration, which may cause inaccurate estimates of days with severe pollution.

### 3.2.2. Seasonal model performance

Fig. 5 shows the spatial pattern over Shenzhen in the dry and wet seasons based on the IRF model. The average number of days available to effectively estimate the PM<sub>2.5</sub> concentration was 119 days due to the occurrence of clouds in the remote sensing images. The IRF model achieves a good performance but exhibits a distinct spatial heterogeneity in the dry and wet seasons, and the space-time RF (STRF) model yields R<sup>2</sup> values of 0.91 and 0.93 in the dry and wet seasons, respectively. The RMSE values are generally 3.71 and 3.27 µg m<sup>-3</sup>, respectively. Between the two seasons, in the dry season, there is less rain, and most forecast results match the PM<sub>2.5</sub> observations from the monitoring stations. The estimation error is caused by the frequent occurrence of high PM<sub>2.5</sub> concentrations (>60 µg m<sup>-3</sup>) caused by man-made pollutants. In contrast, in the wet season, the average R<sup>2</sup> value is high. Although the model accuracy is high, there are fewer sample points in the wet season than in the dry season, which may result in poor model universality. The small error is mainly due to the high RH and low fine-particle concentration caused by frequent precipitation, which complicates the AOD-PM<sub>2.5</sub> relationship.

### 3.3. Comparison to previous studies

#### 3.3.1. Model performance

In this study, we developed a new model for the estimation of the PM<sub>2.5</sub> concentration with the MODIS 1-km AOD product and meteorological parameters based on an improved machine learning approach. This section evaluates the performance of the IRF model and current popular regression models applied in different areas (Table 1), including the LME, GWR, GTWR and other combined models (Ma et al., 2016; Yang et al., 2019b; You et al., 2016). As evidenced by the higher R<sup>2</sup> and lower RMSE values reported for the models based on machine learning, such as the Bayesian maximum entropy (BME)-GWR and stack machine learning models, in PM<sub>2.5</sub> estimation in recent years, these models outperformed the other statistical models (Chen et al., 2019). The combined models, e.g., the LME+support vector regression (SVR) and GTWR+network train (NT) models, also exhibit an excellent prediction ability in PM<sub>2.5</sub> estimation. However, the simple linear

models without other algorithms, such as the generalized additive model (GAM), GLM, and GWR models, do not attain a high accuracy. These models (including the GLM, GAM, and GWR models) do not consider the daily variation characteristics of either the spatial or temporal information of the parameters over time. In these studies, the associated models are not applied in coastal areas with high RH levels and large changes in the wind field, such as Shenzhen near Hong Kong. Therefore, we have attempted to include temporal information to convert the GWR model into the GTWR model, which couples the AOD and meteorological data with the PM<sub>2.5</sub> concentration, because a linear relationship cannot easily capture the variability in AOD and PM<sub>2.5</sub>. Via the investigation of the linear models, we find that the AOD-PM<sub>2.5</sub> relationship is not a simple linear relationship. Hence, a linear relationship cannot accurately predict the PM<sub>2.5</sub> concentration.

Table 1 indicates that from the perspective of the total RMSE and R<sup>2</sup> values obtained from the various statistical models, the LME model (R<sup>2</sup> = 0.45; RMSE = 12.26 µg m<sup>-3</sup>; Sorek-Hamer et al. (2015)) with the 10-km AOD is much less accurate than the other models. The IRF model can explain the daily variations in the sample data effectively (R<sup>2</sup> = 0.91; RMSE = 3.66 µg m<sup>-3</sup>) and attains the best performance among the models applied when predicting the PM<sub>2.5</sub> concentration at the >3-km and 10-km resolutions over different regions, e.g., the GWR and GAM models (R<sup>2</sup> = 0.71 and RMSE = 25.98 µg m<sup>-3</sup>; R<sup>2</sup> = 0.81 and RMSE = 17.2 µg m<sup>-3</sup>; and R<sup>2</sup> = 0.79 and RMSE = 3.6 µg m<sup>-3</sup>) (Liu et al., 2009; Ma et al., 2016; and You et al., 2016; respectively), the two-stage model (R<sup>2</sup> = 0.79 and RMSE = 27.42 µg m<sup>-3</sup>; R<sup>2</sup> = 0.81 and RMSE = 8.83 µg m<sup>-3</sup>; and R<sup>2</sup> = 0.80 and RMSE = 18.58 µg m<sup>-3</sup>) (He and Huang, 2018; Ma et al., 2016; and Yang et al., 2019b; respectively), and the stacking and combined models (R<sup>2</sup> = 0.88 and RMSE = 11.39 µg m<sup>-3</sup>; and R<sup>2</sup> = 0.85 and RMSE = 17.3 µg m<sup>-3</sup>; respectively) (Chen et al., 2019). Moreover, the RMSE values of the GWR model are much higher than those of the machine learning models, even when the R<sup>2</sup> values are similar, because of the

**Table 1**  
Comparison statistics on the performance of the different models.

Model	R <sup>2</sup>	RMSE	Spatial resolution	Author
GAM	0.79	3.60	4 km	Liu et al., 2009
LME	0.45	12.26	10 km	Sorek-Hamer
GWR	0.71	25.98	10 km	Ma et al.
GWR	0.81	17.2	3 km	You et al.
GTWR+NT	0.80	18.58	10 km	He and Huang, 2018
LME+SVR	0.81	8.83	3 km	Yang et al., 2019b
LME+CAR	0.79	27.42	3 km	Ma et al., 2016
BME-GWR	0.88	11.39	3 km	Xiao et al., 2018
Stack machine learning	0.85	17.3	>1 km	Chen et al., 2019
RF	0.88	4.3	1 km	Our study
IRF	0.91	3.66	1 km	Our study

lack of temporal information and nonlinear processing. However, despite the accuracy improvement by the combined models as a whole, both the spatial and temporal heterogeneities still cannot be captured (Xiao et al., 2017). Overall, the comparison results illustrate that the IRF model consistently exhibits a good model fitting effect and can accurately estimate the PM<sub>2.5</sub> concentration through its ability to consider both spatial and temporal variabilities. The high resolution of the MODIS AOD and meteorological data, which imposes the greatest impact on PM<sub>2.5</sub>, is the reason for the relatively low RMSE values in Shenzhen. Additionally, most of the previous studies employed meteorological parameter data with a coarse resolution of 0.5° \* 0.5°, which lowers the estimation accuracy (Ghotbi et al., 2016).

In these models, due to the limitation of the satellite-retrieved AOD product source, the PM<sub>2.5</sub> concentration can only be provided at spatial resolutions of 3 and 10-km (He and Huang, 2018; Ma et al., 2014). However, with the rapid development of the economy, high concentrations of fine particles in the air are becoming common and very serious in densely populated urban areas (Yang et al., 2019b). Therefore, this coarse-resolution PM<sub>2.5</sub> product is not applicable in urban areas. The National Aeronautics and Space Administration (NASA) recently released a 1-km AOD product that can be used to generate PM<sub>2.5</sub> maps with a high resolution. The latest AOD product with a resolution of 1-km has a larger local-scale coverage than the products with a resolution of 3-km, 10-km and coarser; thus, the latest product has been demonstrated to be a better proxy for PM<sub>2.5</sub> estimation in urban-scale areas (Munchak et al., 2013). The results of this study describe the sensitivity of the MODIS 1-km AOD products in capturing aerosol information in coastal areas.

### 3.3.2. Uncertainty analysis

Although the prediction results are better than the majority of previous studies in urban areas (Ma et al., 2016; Yang et al., 2019b; You et al.,

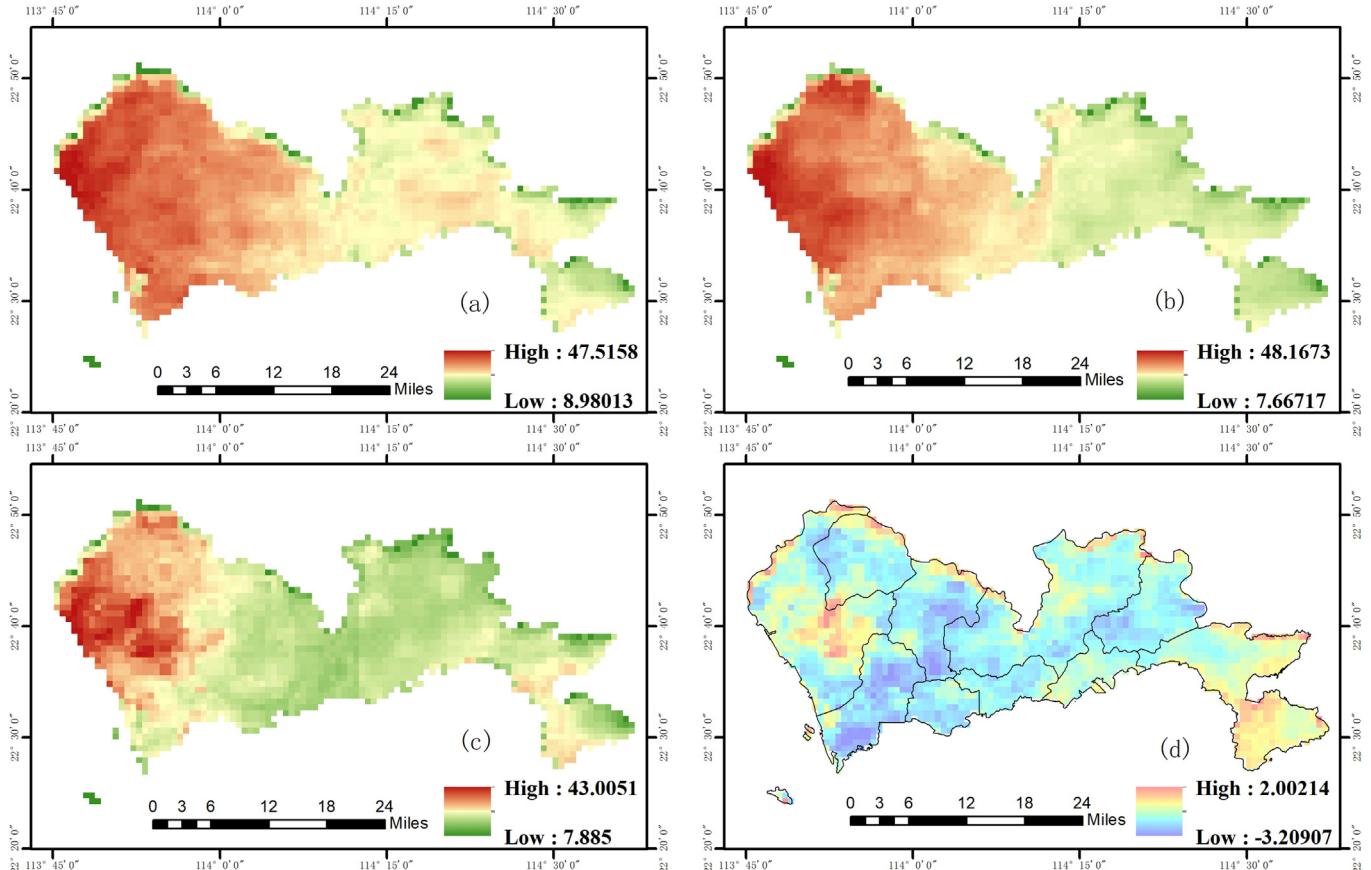
2016), the improved model still has uncertainties in terms of the establishment of the model. Except for AOD, the model only considers the two factors with the greatest impact on PM<sub>2.5</sub> in the study area; thus, additional factors (e.g., land use, population density) should be incorporated to further improve the performance of the model. Certain studies have applied multifactor models to obtain good prediction results (Zang et al., 2017). In addition, higher-resolution AOD data can reflect the distribution information of PM<sub>2.5</sub> on a finer scale (Liu et al., 2007). High-resolution aerosol data (e.g., Landsat/30 m) should be adopted to obtain a more accurate and detailed spatial distribution of the PM<sub>2.5</sub> concentration at the urban scale. Furthermore, due to the specific geographical position of the study area, research on the influence mechanism of the model pollutants in the highly polluted GBA region on the air quality of Shenzhen is currently underway. The performance of the improved model in China is worthy of further research.

Although the IRF model includes few parameters, according to our field survey, compared to other factors, the meteorology is the main factor impacting the atmospheric PM<sub>2.5</sub> over coastal urban areas, such as Shenzhen. Overall, the model is still superior in performance and estimation accuracy than the previous models, and the PM<sub>2.5</sub> products of this model have a spatial resolution that is three to ten times higher than that of the products of the other models. The derived PM<sub>2.5</sub> products have multiple potential high-accuracy applications.

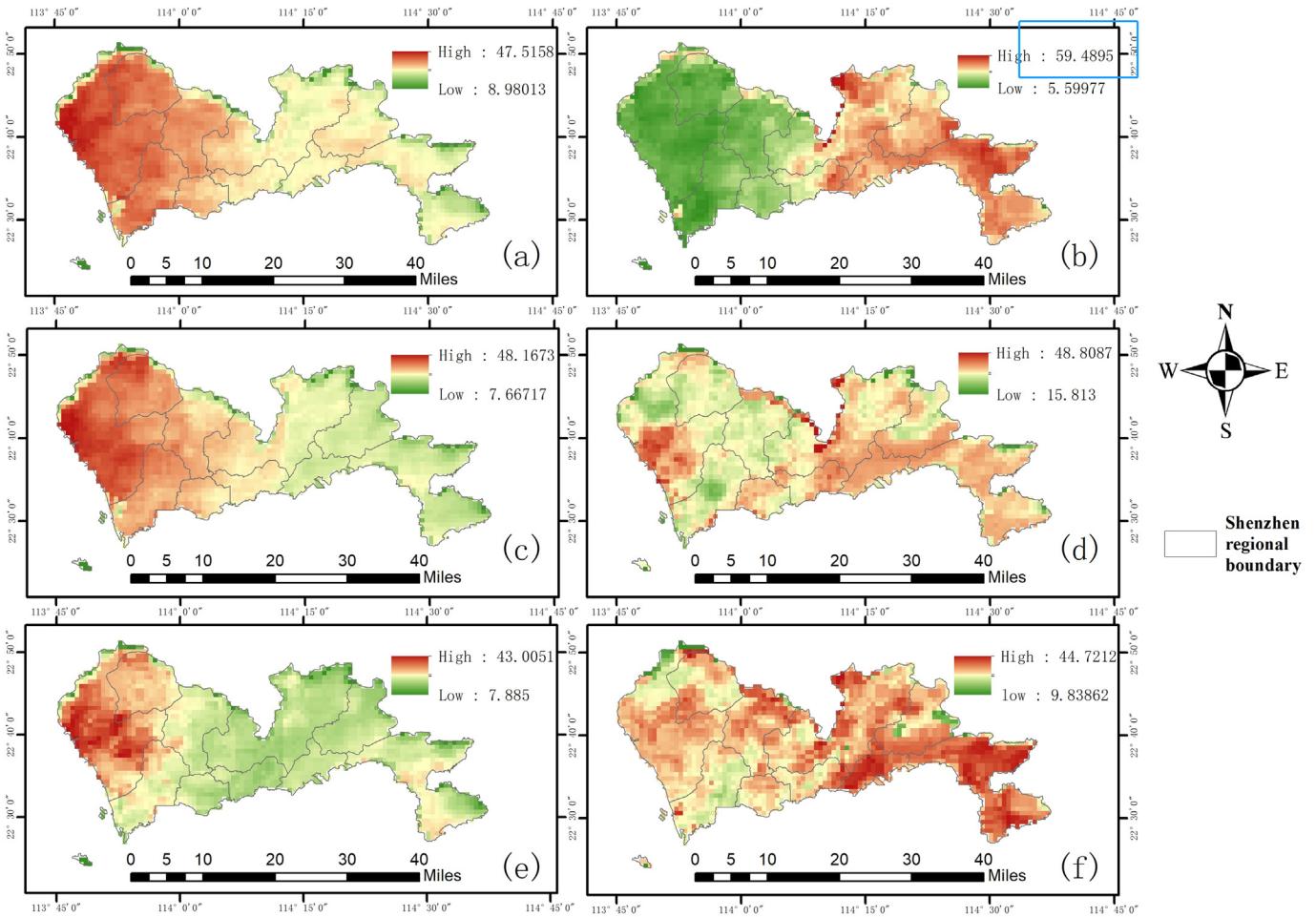
## 4. PM<sub>2.5</sub> spatial distribution in Shenzhen

### 4.1. Annual and seasonal maps

Fig. 6(a–c) shows that the predicted PM<sub>2.5</sub> concentrations obtained by satellite retrieval in 2016, 2017 and 2018 are spatially consistent with the surface observations. The annual PM<sub>2.5</sub> concentration is



**Fig. 6.** The annual PM<sub>2.5</sub> concentration distributions in Shenzhen in 2016, 2017 and 2018 (a–c) and the rate of change from 2016 to 2018 (d) by satellite retrieval.



**Fig. 7.** The average predicted PM<sub>2.5</sub> concentrations from 2016 to 2018 in the wet and dry seasons based on satellite data.

37.98  $\mu\text{g m}^{-3}$ , and higher values primarily occur in western Shenzhen, e.g., the Baoan, Longhua, Nanshan, Guangming and Dapeng districts.

The estimated mean PM<sub>2.5</sub> concentrations in Shenzhen in 2016, 2017, and 2018 were 38.86, 39.42, and 37.1  $\mu\text{g m}^{-3}$ , respectively, and the average PM<sub>2.5</sub> concentration over the period from 2016 to 2018 was 37.98  $\mu\text{g m}^{-3}$ . In Fig. 6(d), the values larger than 0 represent increase, and the values smaller than 0 represent decrease. The rate of change of the PM<sub>2.5</sub> concentration from 2016 to 2018 reveals an increase trend of PM<sub>2.5</sub> over these three years, as nearly 90% of the land area exhibits a difference larger than 0. The estimated PM<sub>2.5</sub> concentration decreases in most districts because of the air pollution control implemented in Shenzhen in recent years and increases only in the Baoan District and Dapeng New District. The estimated mean PM<sub>2.5</sub> values of the 10 districts in Shenzhen were calculated, and the results showed that the average PM<sub>2.5</sub> concentration in each district ranged from 35.56 to 39.1  $\mu\text{g m}^{-3}$  (Fig. 6).

Fig. 7 shows the spatial distribution of the seasonal average PM<sub>2.5</sub> concentration from 2016 to 2018 in Shenzhen as estimated by the IRF model, where a, c, and e are the forecast maps of the dry season. Among the three years, the PM<sub>2.5</sub> distribution was wider in 2016, and the situation improved in 2018. The estimated seasonal mean PM<sub>2.5</sub> concentrations in the dry and wet seasons were 39.56 and 30.22  $\mu\text{g m}^{-3}$ , respectively. The average PM<sub>2.5</sub> concentration in the dry season was much higher than that in the wet season.

The maximum value in the dry season can reach 48  $\mu\text{g m}^{-3}$ , and almost 20% of the daily average PM<sub>2.5</sub> concentrations exceeds 25  $\mu\text{g m}^{-3}$ . Additionally, we find that the distribution of the PM<sub>2.5</sub> concentration in the dry season remains relatively stable, basically revealing that the

PM<sub>2.5</sub> concentration in the western part of the urban area is higher than that in the eastern part. Based on the spatial distribution in the different seasons, we also examine the reasons for the seasonal PM<sub>2.5</sub> changes. These seasonal changes are closely related to the climatic characteristics of Shenzhen. Shenzhen is characterized by wet and humid conditions in the wet season, and the temperature and boundary layer height are relatively high, which promotes the dispersion of PM<sub>2.5</sub>. In addition, typhoons from the tropical ocean occur from July to September and are accompanied by high wind speeds and heavy rainfall, which can also reduce the pollutant concentration. The dry season mainly occurs in spring and winter, with little precipitation, and the vertical transport of air pollutants can lead to the accumulation of particulate matter, resulting in high concentrations.

#### 4.2. PM<sub>2.5</sub> concentration center of gravity analysis

There are many factors that influence PM<sub>2.5</sub>, including the surface elevation, industrial pollution sources and urban ventilation conditions, and these factors can be adopted to examine the spatial distribution characteristics of the PM<sub>2.5</sub> concentration. The k-means clustering analysis method was used to analyze the center of gravity shift of the PM<sub>2.5</sub> concentration in Shenzhen based on the prediction maps (Fig. 8). These centers of gravity represent the centers of gravity of the concentration in the prediction maps. Hence, these centers of gravity were clustered into two groups, indicating that the pollution sources are located in two main areas. Fig. 8 reveals that the pollution source remains stable, with 75% of the points falling within the range of the clusters from 2016 to 2018. There are two high-value clusters: one is located at the

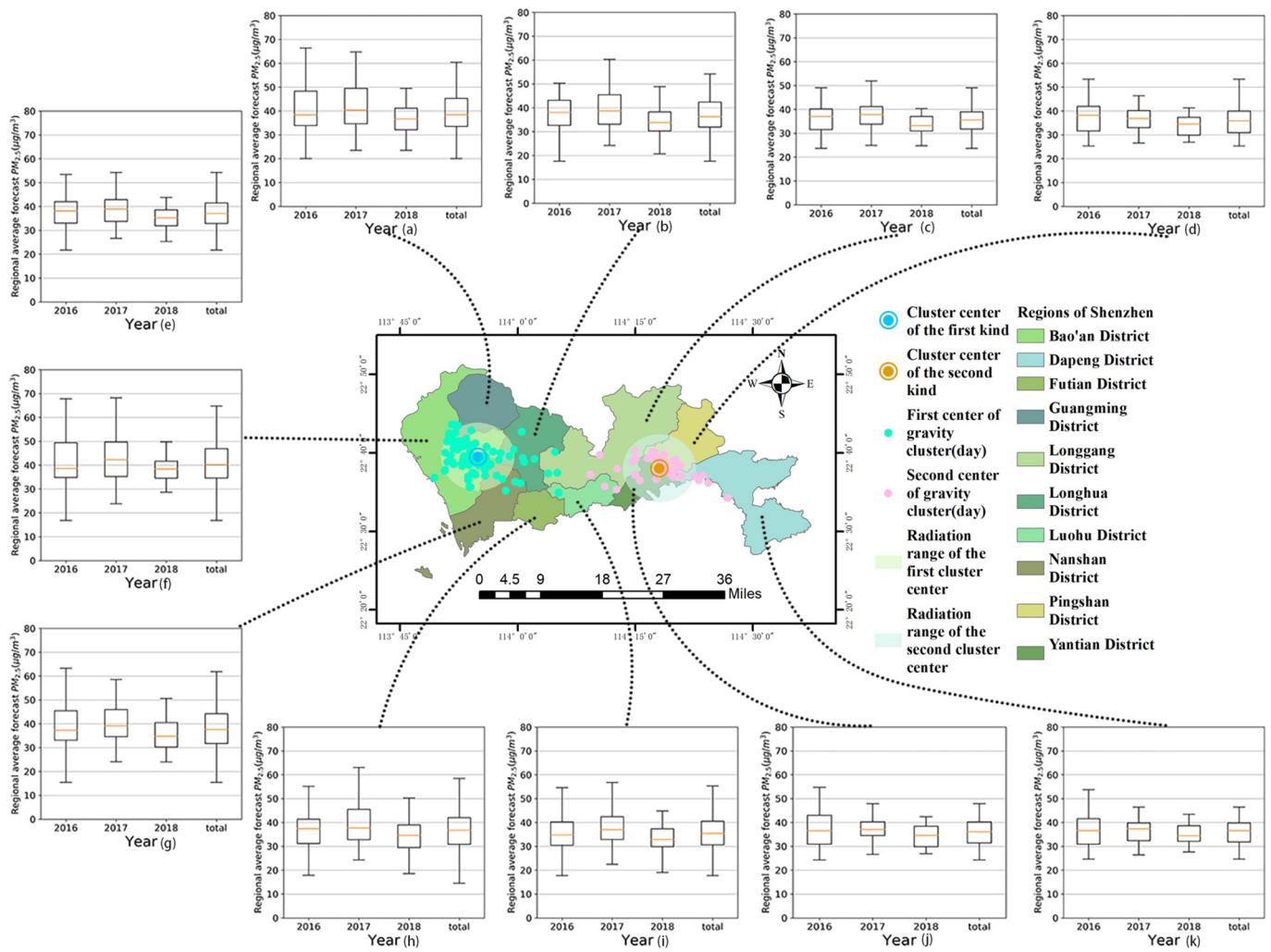


Fig. 8. Estimated migration of the center of gravity of the PM<sub>2.5</sub> concentration from 2016 to 2018 in Shenzhen.

junction of the Baoan, Longhua, Guangming, and Nanshan districts, with a high-value range of 63–69 µg m<sup>-3</sup>, and the other is located at the junction of the Yantian and Pingshan districts, with a range of 52–55 µg m<sup>-3</sup>. The PM<sub>2.5</sub> pollution sources in Shenzhen are mainly located in the Baoan, Guangming and Pingshan districts, and these areas exhibit high annual and interannual values. In addition, the accumulation of pollution sources is also closely related to the prevailing wind direction and wind speed in coastal areas. During the dry season, Shenzhen is dominated by a northeast wind direction, and the pollutants are mainly concentrated in areas near Yantian and Pingshan. In the wet season, the pollutants are mainly concentrated in the Baoan and Longhua areas because a southeast wind direction prevails in Shenzhen.

The Pingshan and Baoan districts in Shenzhen are its main industrial areas and have a high population density and rapid economic development, leading to high pollutant emissions. In addition, the areas near the sea exhibit a high PM<sub>2.5</sub> concentration. The atmospheric aerosols along the coast include both marine and terrestrial aerosols, and their physical and chemical components and sources are complex, which may result in higher pollutant predictions in these areas. In contrast, the northern region of Shenzhen usually has a low PM<sub>2.5</sub> concentration (<40 µg m<sup>-3</sup>), which is probably due to the low anthropogenic aerosol loading and favorable meteorological conditions that inhibit pollutant accumulation. However, the level in more than 55% of the Shenzhen area exceeded the fine-particle pollution standards considered

acceptable by the WHO and the Chinese Ministry of Ecology and the Environment.

## 5. Conclusions

At present, most satellite-retrieved AOD-PM<sub>2.5</sub> concentrations constrained by input data with a coarse spatial resolution (>3-km) cannot accurately produce small- and medium-scale air pollution predictions, especially in urban areas. Shenzhen is a coastal city in the GBA with few spatiotemporal PM<sub>2.5</sub> estimations based on remote sensing data. Additionally, because of the high uncertainty in the AOD-PM<sub>2.5</sub> relationship, most traditional linear models typically used to estimate PM<sub>2.5</sub> concentration from satellite data experience problems that cannot be solved, such as the influence of factors with a strong spatiotemporal heterogeneity. An IRF model based on machine learning is developed in this work, and it resolves the complex nonlinear relationship between AOD and PM<sub>2.5</sub> and considers two independent input variables in coastal regions characterized by varying climatic conditions to better account for the temporal and spatial changes.

In the IRF model, the newly released 1-km AOD product was applied to generate a 1-km PM<sub>2.5</sub> product for Shenzhen for the first time, and this product yielded accurate daily PM<sub>2.5</sub> concentration estimates. The results showed that the estimation accuracy of the IRF model is considerably higher than that of the RF and GTWR models ( $R^2 = 0.88$  and

$\text{RMSE} = 4.33 \mu\text{g m}^{-3}$ , and  $R^2 = 0.46$  and  $\text{RMSE} = 11.14 \mu\text{g m}^{-3}$ , respectively). The IRF model can also accurately estimate the  $\text{PM}_{2.5}$  concentration at the seasonal ( $R^2 = 0.92$  and  $\text{RMSE} = 3.44 \mu\text{g m}^{-3}$ ) and annual ( $R^2 = 0.91$  and  $\text{RMSE} = 3.66 \mu\text{g m}^{-3}$ ) scales. This model may be adopted to reconstruct  $\text{PM}_{2.5}$  information on large spatial and temporal scales because long time series of satellite data are available. However, except for AOD, the model only considers the two factors with the greatest impact on  $\text{PM}_{2.5}$  in the study area. Therefore, additional factors should be incorporated to further improve the performance of the model.

Although the IRF model includes few parameters, it is still superior in performance and estimation accuracy to the previous models, and the  $\text{PM}_{2.5}$  products of this model yield a spatial resolution that is three to ten times higher than that yielded by the products of the other models. Thus, this model can produce more detailed  $\text{PM}_{2.5}$  information for small- and medium-sized areas. The results of the study indicate that the 1-km  $\text{PM}_{2.5}$  product we have developed can be effectively applied for air quality research in urban areas, especially in coastal areas.

## CRediT authorship contribution statement

**Wenqian Chen:** Conceptualization, Methodology, Software, Investigation, Writing - original draft. **Haofan Ran:** Resources, Validation, Formal analysis, Visualization, Data curation. **Xiaoyi Cao:** Resources, Writing - review & editing, Supervision, Data curation. **Jingzhe Wang:** Writing - review & editing, Supervision. **Dexiong Teng:** Writing - review & editing. **Jing Chen:** Writing - review & editing. **Xuan Zheng:** Validation, Formal analysis, Visualization, Software.

## Declaration of competing interest

There is no conflict of interest exist in the submission of this manuscript, and manuscript is approved by all authors for publication.

## Acknowledgments

This work was supported by Joint Fund for Basic and Applied Basic Research of Guangdong Province of China (2019A1515110692) and the National Natural Science Foundation of China (51978404). We are especially grateful to the editors, anonymous reviewers for appraising our manuscript and instructive comments, and Meteorological Bureau and Institute of Environmental Sciences of Shenzhen offering data.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2020.141093>.

## References

- Bernardo, S.B., 2013. A hybrid approach to estimating national scale spatiotemporal variability of  $\text{PM}_{2.5}$  in the contiguous United States. *Environ. Sci. Technol.* 13, 7233–7241.
- Chen, G., Li, S., Knibbs, L.D., Hamm, N.A.S., Cao, W., Li, T., et al., 2018. A machine learning method to estimate  $\text{PM}_{2.5}$  concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* 636, 52–60.
- Chen, J., Yin, J., Zang, L., Zhang, T., Zhao, M., 2019. Stacking machine learning model for estimating hourly  $\text{PM}_{2.5}$  in China based on Himawari 8 aerosol optical depth data. *Sci. Total Environ.* 697.
- Dai, W., Gao, J., Cao, G., Ouyang, F., 2013. Chemical composition and source identification of  $\text{PM}_{2.5}$  in the suburb of Shenzhen, China. *Atmos. Res.* 122, 391–400.
- Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing  $\text{PM}_{2.5}$  exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol.* 50, 4712–4721.
- Ghotbi, S., Sotoudehian, S., Arhami, M., 2016. Estimating urban ground-level  $\text{PM}_{10}$  using MODIS 3km AOD product and meteorological parameters from WRF model. *Atmos. Environ.* 141, 333–346.
- He, Q., Huang, B., 2018. Satellite-based mapping of daily high-resolution ground  $\text{PM}_{2.5}$  in China via space-time regression modeling. *Remote Sens. Environ.* 206, 72–83.
- Hu, X., Waller, L.A., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G., Estes, S.M., et al., 2013. Estimating ground-level  $\text{PM}_{2.5}$  concentrations in the southeastern U.S. using geographically weighted regression. *Environ. Res.* 121, 1–10.
- Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., et al., 2017. Estimating  $\text{PM}_{2.5}$  concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol.* 51, 6936–6944.
- Hutchison, K.D., Smith, S., Faruqui, S.J., 2005. Correlating MODIS aerosol optical thickness data with ground-based  $\text{PM}_{2.5}$  observations across Texas for use in a real-time air quality prediction system. *Atmos. Environ.* 39, 7190–7203.
- Lee, H.J., Liu, Y., Coull, B.A., Schwartz, J., Chemistry PKJA, Physics, et al., 2011. A novel calibration approach of MODIS AOD data to predict  $\text{PM}_{2.5}$  concentrations. *Atmos. Chem. Phys.* 11, 7859–7873.
- Li, T., Shen, H., Zeng, C., Yuan, Q., Zhang, L., 2016. Point-surface fusion of station measurements and satellite observations for mapping  $\text{PM}_{2.5}$  distribution in China: methods and assessment. *Atmos. Environ.* 152.
- Liang, X.Y., Zhang, S.J., Wu, X., Guo, X., Han, L., Liu, H., Wu, Y., Hao, J.M., 2020. Air quality and health impacts from using ethanol blended gasoline fuels in China. *Atmos. Environ.* 228, 117396.
- Lin, C., Li, Y., Yuan, Z., Lau, A.K.H., Li, C., Fung, J.C.H., 2015. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level  $\text{PM}_{2.5}$ . *Remote Sens. Environ.* 156, 117–128.
- Liu, Y., Franklin, M., Kahn, R., Koutrakis, P., 2007. Using aerosol optical thickness to predict ground-level  $\text{PM}_{2.5}$  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  $\text{PM}(2.5)$  concentrations using satellite data, meteorology, and land use information. *Environ. Health Perspect.* 117, 886–892.
- Liu, M., Bi, J., Ma, Z., 2017. Supporting Information for Visibility-based  $\text{PM}_{2.5}$  Concentrations in China.
- Lv, B., Hu, Y., Chang, H.H., Russell, A.G., Bai, Y., 2016. Improving the accuracy of daily  $\text{PM}_{2.5}$  distributions derived from the fusion of ground-level measurements with aerosol optical depth observations, a case study in North China. *Environ. Sci. Technol.* 50, 4752–4759.
- Lv, B., Hu, Y., Chang, H.H., Russell, A.G., Cai, J., Xu, B., et al., 2017. Daily estimation of ground-level  $\text{PM}_{2.5}$  concentrations at 4km resolution over Beijing-Tianjin-Hebei by fusing MODIS AOD and ground observations. *Sci. Total Environ.* 580, 235–244.
- Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I., Korkin, S., 2011. Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *J. Geophys. Res.* 116.
- 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  $\text{PM}_{2.5}$  in China using satellite remote sensing. *Environ. Sci. Technol.* 48, 7436–7444.
- Ma, Z., Liu, Y., Zhao, Q., Liu, M., Zhou, Y., Bi, J., 2016. Satellite-derived high resolution  $\text{PM}_{2.5}$  concentrations in Yangtze River Delta Region of China using improved linear mixed effects model. *Atmos. Environ.* 133, 156–164.
- Munchak, L.A., Levy, R.C., Mattoe, S., Remer, L.A., Holben, B.N., Schafer, J.S., et al., 2013. MODIS 3 km aerosol product: applications over land in an urban/suburban region. *Atmos. Meas. Tech.* 6, 1747–1759.
- Rolandone, F., Mareschal, J.-C., Jaupart, C., 2003. Temperatures at the base of the Laurentide Ice Sheet inferred from borehole temperature data. *Geophys. Res. Lett.* 30.
- Song, Y.-Z., Yang, H.-L., Peng, J.-H., Song, Y.-R., Sun, Q., Li, Y., 2015. Estimating  $\text{PM}_{2.5}$  concentrations in Xi'an City using a generalized additive model with multi-source monitoring data. *PLoS One* 10, e0142149.
- Sorek-Hamer, M., Kloog, I., Koutrakis, P., Strawa, A.W., Chatfield, R., Cohen, A., et al., 2015. Assessment of  $\text{PM}_{2.5}$  concentrations over bright surfaces using MODIS satellite observations. *Remote Sens. Environ.* 163, 180–185.
- van Donkelaar, A., Martin, R.V., Levy, R.C., da Silva, A.M., Krzyzanowski, M., Chubarova, N.E., et al., 2011. Satellite-based estimates of ground-level fine particulate matter during extreme events: a case study of the Moscow fires in 2010. *Atmos. Environ.* 45, 6225–6232.
- Wang, Z., Chen, L., Tao, J., Zhang, Y., Su, L., 2010. Satellite-based estimation of regional particulate matter (PM) in Beijing using vertical-and-RH correcting method. *Remote Sens. Environ.* 114, 50–63.
- Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., et al., 2019. Estimating 1-km-resolution  $\text{PM}_{2.5}$  concentrations across China using the space-time random forest approach. *Remote Sens. Environ.* 231.
- Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., et al., 2017. Full-coverage high-resolution daily  $\text{PM}_{2.5}$  estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ.* 199, 437–446.
- Xiao, L., Lang, Y.C., Christakos, G., 2018. High-resolution spatiotemporal mapping of  $\text{PM}_{2.5}$  concentrations at Mainland China using a combined BME-GWR technique. *Atoms. Environ.* 173, 295–305.
- Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., et al., 2019. Spatiotemporal continuous estimates of  $\text{PM}_{2.5}$  concentrations in China, 2000–2016: a machine learning method with inputs from satellites, chemical transport model, and ground observations. *Environ. Int.* 123, 345–357.
- Yang, D.Y., Zhang, S.J., Niu, T.L., Wang, Y.J., Xu, H.L., Zhang, K., Wu, Y., 2019a. High-resolution mapping of vehicle emissions of atmospheric pollutants based on large-scale, real-world traffic datasets. *Atmos. Chem. Phys.* 19, 8831–8843.
- Yang, L., Xu, H., Jin, Z., 2019b. Estimating ground-level  $\text{PM}_{2.5}$  over a coastal region of China using satellite AOD and a combined model. *J. Clean. Prod.* 227, 472–482.
- You, W., Zang, Z., Zhang, L., Li, Y., Wang, W., 2016. Estimating national-scale ground-level  $\text{PM}_{2.5}$  concentration in China using geographically weighted regression based on MODIS and MISR AOD. *Environ. Sci. Pollut. Res.* 23, 8327–8338.

- Zang, Z., Wang, W., You, W., Li, Y., Ye, F., Wang, C., 2017. Estimating ground-level PM<sub>2.5</sub> concentrations in Beijing, China using aerosol optical depth and parameters of the temperature inversion layer. *Sci. Total Environ.* 575, 1219–1227.
- Zhai, L., Li, S., Zou, B., Sang, H., Fang, X., Xu, S., 2018. An improved geographically weighted regression model for PM<sub>2.5</sub> concentration estimation in large areas. *Atmos. Environ.* 181, 145–154.
- Zhang, Y., Li, Z., 2015. Remote sensing of atmospheric fine particulate matter (PM<sub>2.5</sub>) mass concentration near the ground from satellite observation. *Remote Sens. Environ.* 160, 252–262.
- Zhang, T., Zhu, Z., Gong, W., Zhu, Z., Sun, K., Wang, L., et al., 2018. Estimation of ultrahigh resolution PM<sub>2.5</sub> concentrations in urban areas using 160 m Gaofen-1 AOD retrievals. *Remote Sens. Environ.* 216, 91–104.

Update

**Science of the Total Environment**

Volume 765, Issue , 15 April 2021, Page

DOI: <https://doi.org/10.1016/j.scitotenv.2021.145325>



## Corrigendum

**Corrigendum to “Estimating PM<sub>2.5</sub> with high-resolution 1-km AOD data and an improved machine learning model over Shenzhen, China” [Sci. Total Environ. 746 (December 1, 2020) 141093]**



Wenqian Chen <sup>a,b,c</sup>, Haofan Ran <sup>d</sup>, Xiaoyi Cao <sup>d</sup>, Jingzhe Wang <sup>e</sup>, Dexiong Teng <sup>d</sup>, Jing Chen <sup>b</sup>, Xuan Zheng <sup>a,\*</sup>

<sup>a</sup> College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen 518060, China

<sup>b</sup> Geography Department, Hanshan Normal University, Chaozhou 521041, China

<sup>c</sup> College of Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China

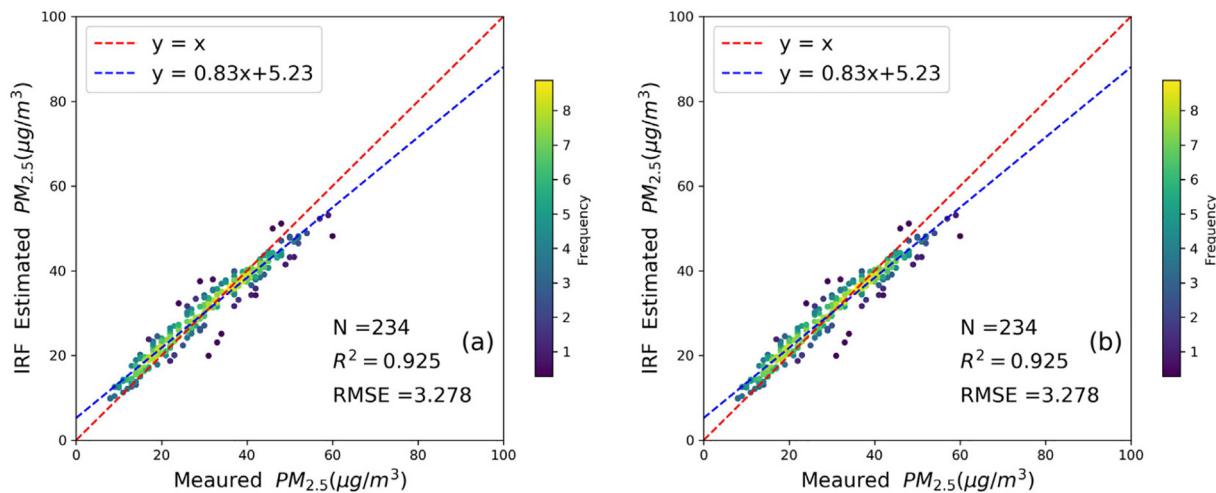
<sup>d</sup> College of Resources and Environment Sciences, Xinjiang University, Urumqi 830046, Xinjiang, China

<sup>e</sup> MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area & Guangdong Key Laboratory of Urban Informatics & Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China

Dear editors,

We found that in our paper which published on Science of the Total Environment, “Estimating PM<sub>2.5</sub> with high-resolution 1-km AOD data and an improved machine learning model over Shenzhen, China”, STOTEN 141093, the two pictures in Fig. 5 are repeat, as following,

W. Chen et al. / Science of the Total Environment 746 (2020) 141093

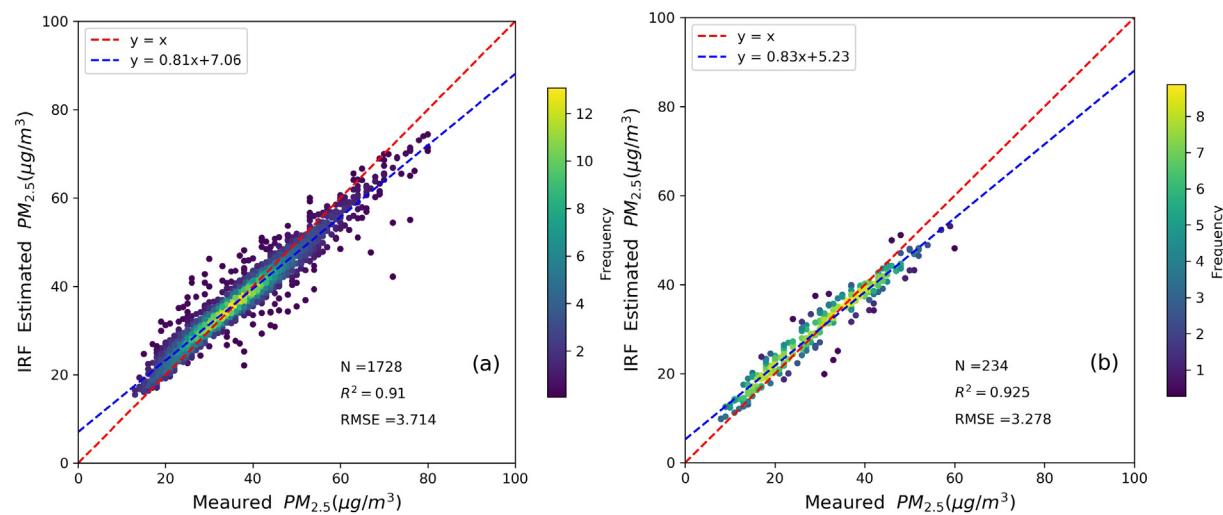


DOI of original article: <https://doi.org/10.1016/j.scitotenv.2020.141093>.

\* Corresponding author.

E-mail address: [zhengxuanqh@163.com](mailto:zhengxuanqh@163.com) (X. Zheng).

The authors regret that the printed version of the above article contained a number of errors. The correct and final version follows. The authors would like to apologise for any inconvenience caused.



**Fig. 5.** Prediction of IRF model during the dry and wet seasons from 2016 to 2018 over Shenzhen.

The correction version of Fig. 5 as following,