

# Satellite-based PM<sub>2.5</sub> estimation directly from reflectance at the top of the atmosphere using a machine learning algorithm

Jianjun Liu<sup>a</sup>, Fuzhong Weng<sup>b,\*</sup>, Zhanqing Li<sup>c</sup>



<sup>a</sup> Laboratory of Environmental Model and Data Optima (EMDO), Laurel, MD, USA

<sup>b</sup> State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China

<sup>c</sup> Earth System Science Interdisciplinary Center and Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, USA

## ARTICLE INFO

### Keywords:

PM<sub>2.5</sub> concentration  
TOA reflectances  
Machine learning

## ABSTRACT

Atmospheric particulate matter (PM) that have particle diameter less than 2.5 μm (PM<sub>2.5</sub>) are hazardous to public health whose concentration has been either measured on the ground or inferred from satellite-retrieved aerosol optical depth (AOD). The latter is subject to numerous sources of errors, making the satellite retrievals of PM<sub>2.5</sub> highly uncertain. This study developed an ensemble machine-learning (ML) algorithm for estimating PM<sub>2.5</sub> concentration directly from Advanced Himawari Imager satellite measured top-of-the-atmosphere (TOA) reflectances in 2016 integrated with meteorological parameters. The algorithm is demonstrated to perform well across China with high accuracies at different temporal scales. The model has an overall cross-validation coefficient of determination ( $R^2$ ) of 0.86 and a root-mean-square error (RMSE) of 17.3 μg m<sup>-3</sup> for hourly PM<sub>2.5</sub> concentration estimation. Such accuracies of the estimation on PM<sub>2.5</sub> concentration by using TOA reflectance directly are comparable with those of the common methods on estimating PM<sub>2.5</sub> concentration by using satellite-derived AODs, but the former has a relatively stronger predictive power relating to spatial-temporal coverages than the latter. Annual and seasonal variations of PM<sub>2.5</sub> concentration over three major the developed regions in China are estimated using the model and analyzed. The relatively stronger predictive ability of developed model in this study may help provide information about the diurnal cycle of PM<sub>2.5</sub> concentrations as well as aid in monitoring the processes of regional pollution episodes and the evolution of PM<sub>2.5</sub> concentration.

## 1. Introduction

Atmospheric particulate matter (PM) that have diameters of less than 2.5 μm (PM<sub>2.5</sub>) are air pollutant. Epidemiological studies have revealed significant correlation between the PM<sub>2.5</sub> concentration and premature death, cardiovascular diseases, and respiratory diseases (e.g. Kan and Chen, 2002; Lelieveld et al., 2015). Atmospheric aerosols significantly affect Earth's climate by affecting radiation budgets, cloud properties, precipitation and water cycle processes (e.g. Ramanathan et al., 2001; Rosenfeld et al., 2008; Liu et al., 2012; Fan et al., 2015; Li et al., 2016, 2017; Liu and Li, 2018). PM<sub>2.5</sub> is one of the primary air pollutants in the rapidly growing megacities of developing countries such as China, garnering increasing attention from the government and public alike (e.g. Ma et al., 2016; Yu et al., 2017). The sparse spatial coverage of surface PM<sub>2.5</sub> monitoring sites, especially in rural regions, in China (MEPC, 2015), limits our ability to capture and evaluate the dynamics of air pollution and pursue health- and environment-related studies, as well for making sound policies.

In compensating for the inadequate spatiotemporal coverage of PM<sub>2.5</sub> monitoring sites, many attempts have been made for PM<sub>2.5</sub> estimation from satellite-retrieved aerosol optical depth (AOD) with different approaches, including satellite AOD and atmospheric chemical model simulations joint approaches (e.g. Liu et al., 2004; Geng et al., 2015; Van Donkelaar et al., 2010; Van Donkelaar et al., 2013), statistical approaches (e.g. Engel-Cox et al., 2004; Lee et al., 2011; Ma et al., 2014) and physical methods incorporating aerosol particle size information (e.g., fine mode fraction, or FMF), aerosol vertical distribution and aerosol swelling effect, etc. (Lin et al., 2015; Zhang and Li, 2015). The performance of atmospheric chemical model simulations is strongly affected by the meteorological dynamics that drive chemical transport, emission inventory, and reaction mechanisms (Zhang and Li, 2015). The accuracy of physical models is strongly influenced by uncertainties in the estimations of aerosol properties, such as hygroscopic growth, the mass extinction efficiency, etc. Although statistical approaches have inherent problems (e.g. a model may not be applicable to different locations), they can usually estimate the PM<sub>2.5</sub> concentration

\* Corresponding author.

E-mail address: fweng58@gmail.com (F. Weng).

with an acceptable accuracy (Liu, 2014; Li et al., 2017a). Previous studies have attempted to predict surface PM<sub>2.5</sub> concentration using the linear/non-linear statistical models by directly establishing the PM<sub>2.5</sub>-AOD relationship (e.g. Engel-Cox et al., 2004; Liu et al., 2007; Zhang et al., 2009; Sorek-Hamer et al., 2013) and using the advanced statistical models by incorporating satellite-derived AOD and more related variables, such as meteorological parameters, land use information, emission and so on (e.g. Liu et al., 2009; Lee et al., 2012; Xie et al., 2015; Hu et al., 2013; Song et al., 2014; Ma et al., 2016; You et al., 2016a, 2016b; Fang et al., 2016; Lv et al., 2016, 2017; Yu et al., 2017).

In addition to the conventional statistical models, machine learning, a subdivision of artificial intelligence, is likely one of the best options to address the complex nonlinear relationship among AOD, PM<sub>2.5</sub> and related variables, and generally achieve an outstanding predictive performance (Breiman, 2001; Zhan et al., 2017). Recently, several different machine learning algorithms have been applied on the satellite-based PM<sub>2.5</sub> concentration estimation, including without limitation, the artificial neural network model (e.g. Feng et al., 2015), the back-propagation neural network (e.g. Gupta and Christopher, 2009), the generalized regression neural network (e.g. Li et al., 2017b), the geo-intelligent deep belief network (e.g. Li et al., 2017a), and support vector regression (e.g. Hou et al., 2014). Random forests, a type of ensemble machine learning algorithms, have been widely used for nonparametric regression analysis with competitive accuracy (Hu et al., 2017). Unlike many other machine learning algorithms (e.g., the deep belief network, the gradient boosted machine), the random forest is very user-friendly and has only two parameters to fine-tune to achieve excellent performance (Liaw and Wiener, 2002).

The above studies generally use satellite-retrieved AOD integrated with the other variables, such as meteorological variables, land-cover information, road information, emission, etc. as additional predictors. AOD is retrieved from satellite-measured top-of-the-atmosphere (TOA) reflectances based on different methods such as the Dark Target (Levy et al., 2013). Ristovski et al. (2012) retrieved AOD from MODIS observations using a neural network with a satisfactory accuracy. The results suggest the possibility of estimating PM<sub>2.5</sub> concentrations using TOA reflectances directly rather than AOD retrievals, which possibly circumvent many sources of retrieval uncertainties in the retrievals of AOD (Li et al., 2009).

This study aims to develop a model based on random forest algorithm by incorporating TOA satellite reflectances from a geostationary satellite and meteorological parameters for estimating hourly PM<sub>2.5</sub> concentration. Given the heavy and highly variable air quality condition in China, it is an ideal testbed and we thus choose the region for our study. The model is evaluated using cross validation and several statistical indicators, and the spatial distribution and seasonal variations in PM<sub>2.5</sub> concentration in three developed regions of China are examined. In the following sections, data and methods are described first. The results and discussion follow. Conclusions are given in last.

## 2. Dataset and methodologies

### 2.1. Data

#### 2.1.1. Satellite products

The Advanced Himawari Imager (AHI) onboard the Himawari-8 satellite, the eighth satellite in a series of Himawari geostationary weather satellites of Japan, can measure the TOA reflectances at six visible/near-infrared wavelengths and brightness temperatures at ten infrared wavelengths over the East Asia. Level 1B full-disk calibrated reflectance products and Level-2 AOD products with 5-km spatial and 10-min temporal resolutions were downloaded from the Japan Aerospace Exploration Agency P-Tree system ([ftp://ftp.ptree.jaxa.jp/](http://ftp.ptree.jaxa.jp/)). Compared with the polar-orbiting satellites (e.g. Terra and Aqua), which only supply the “snapshot” observations once or twice per day over a particular location, the high temporal resolution of AHI satellite measurements is useful for capturing

the diurnal cycle and general evolution of PM<sub>2.5</sub>. TOA reflectances at 0.47 μm, 0.64 μm, and 2.3 μm are typically used to retrieve AOD through an atmospheric radiative transfer model based on the Dark Target algorithm (Kaufman et al., 1997). The satellite-measured TOA spectral reflectance at a given wavelength is mainly attributed to the reflection of radiation from the surface and scattering of radiation within the atmosphere without interaction with the surface. The latter is related to the aerosol type and loading (AOD). A lookup table (LUT) which contains pre-computed TOA reflectances is derived using an atmospheric radiative transfer model for a set of aerosol and surface parameters. Surface reflectance in the short-wavelength infrared (e.g., 2.11 μm), the scattering angle, and the “greenness” of the surface in the mid-IR spectrum determines the surface contribution at visible wavelengths. The AOD is derived by finding the best match between the spectral reflectance from the LUT and the satellite-measured reflectance. In this study, TOA reflectances at these three wavelengths and four observation angles (solar zenith angle (SZA), solar azimuth angle (SOA), satellite zenith angle (SAZ), and satellite azimuth angle (SAA)) are used to initially build the random forest model to calculate surface PM<sub>2.5</sub> concentration. For comparison purposes, AODs with the highest confidence (“very good” of confidence levels) from the Himawari-8 level 2 product are also used to build the random forest model. Level-2 AOD products also contain cloud mask information which is used to remove cloud-contaminated TOA reflectances.

#### 2.1.2. Surface PM<sub>2.5</sub> concentration

PM<sub>2.5</sub> concentration at ~1500 sites over mainland China from January 1, 2016 to December 31, 2016 (Fig. S1) are extracted from the website of China National Environmental Monitoring Center (CNEMC). PM<sub>2.5</sub> mass concentrations were measured with a tapered element oscillating microbalance with an accuracy of ± 1.5 μg m<sup>-3</sup> and ± 0.5 μg m<sup>-3</sup> for hourly and daily averages, respectively. The minimum detectable limit of the hourly averaged PM<sub>2.5</sub> concentration is 0.06 μg m<sup>-3</sup>. Hourly averaged PM<sub>2.5</sub> concentrations are used and missing data (values recorded as NA) and values lower than the detectable limit are removed.

#### 2.1.3. Meteorological variables

The surface atmospheric pressure (P, hPa), total column water (TCW, kg m<sup>-2</sup>), 10-m u-wind (U<sub>10</sub>) and v-wind (V<sub>10</sub>) component, air temperature at an altitude of 2 m (T, K), total column ozone (kg m<sup>-2</sup>), relative humidity (RH, %), and planetary boundary layer height (PBLH, m) were obtained from the ERA-Interim reanalysis. The meteorological variables were selected based on many previous studies (e.g., Ma et al., 2014; Li et al., 2017; Su et al., 2018, etc.) that have revealed those factors having a significant influence on PM<sub>2.5</sub> and its relationship with AOD. Also considered where the physical mechanisms of AOD retrievals from satellite TOA reflectances (e.g., Kaufman et al., 1997; Levy et al., 2013). Most of the variables are routinely observed and easy to acquire. PBLH, P, and T are related to atmospheric stability, which influences the vertical distribution of pollution. Temperature also affects the physical properties and chemical composition of aerosols by influencing the aerosol formation (Price et al., 2016). Wind can influence the horizontal transport of pollution, and RH can affect the particle size and the extinction of radiation in the atmosphere. The total column ozone and total column water are considered because of their potential influence on the AOD retrievals from TOA reflectance (e.g. Levy et al., 2013). Except for the PBLH, which is available at 00:00 and 12:00 UTC, all other data are operationally produced four times daily starting at 00:00 UTC in six-hourly intervals. These data are gridded in the 0.125° × 0.125° spatial resolution.

## 2.2. Methods

We first collocated AHI measurements, surface PM<sub>2.5</sub> concentrations, and meteorological variables to generate the training dataset. Then the random forest model was fitted by applying it to the training

dataset. In model fitting, the modeling dataset was used in both the model fitting and the model validation. The best parameters of the model were determined by adjusting them until the best prediction accuracy was achieved. This procedure does not account for the overfitting problem. Ten-fold cross validation (CV) was then used to further adjust the model to avoid the overfitting problem. The resulting model was finally used to estimate hourly surface PM<sub>2.5</sub> concentrations at the AHI pixel level.

### 2.2.1. Data processing

We collocated AHI reflectance measurements, AOD, cloud mask data, and ECMWF meteorological variables at each surface PM<sub>2.5</sub> site falling within intersecting AHI and ECMWF grids. The reflectance and AOD under cloud-free conditions are averaged over the same hourly observational period at each PM<sub>2.5</sub> site. Daytime samples with SZAs less than 80° and the reflectance at all six wavelengths ∈ (0, 1) are collected.

The surface reflection of radiation significantly contributes to the TOA reflectance. The normalized difference vegetation index (NDVI) can reflect the land surface properties and is heavily affected by aerosol loading. The NDVI<sub>SWIR</sub>, the difference between reflectance at 1.24 and 2.12 μm divided by the sum of reflectance at 1.24 and 2.12 μm, which is less affected by aerosols than the traditional NDVI, is used in the MODIS Dark Target algorithm (Gupta et al., 2016). Since AHI lacks a 1.24-μm channel, the 0.86-μm reflectance is used instead to calculate NDVI<sub>SWIR</sub> as follows:

$$\text{NDVI}_{\text{SWIR}} = \frac{\rho_{0.86} - \rho_{2.26}}{\rho_{0.86} + \rho_{2.26}} \quad (1)$$

where  $\rho_{0.86}$  and  $\rho_{2.26}$  are the AHI-measured reflectances in the 0.86-μm and 2.26-μm channels under cloud-free conditions (Aerosol Product Application, 2012).

### 2.2.2. Model development and validation

Random forest algorithms create multiple decision trees by means of extracting the bootstrap samples from the whole training dataset. At each node of tree, the best split is chosen from among a random subset of all predictors at that node, and then used to construct that tree. Finally, a majority vote or average is applied to the prediction. At each bootstrap iteration, the algorithm uses the predictions of out-of-bag (OOB) samples (i.e., data not in the bootstrap samples) to calculate the error rate (Liaw and Wiener, 2002).

The random forest model used here was developed by incorporating TOA reflectances, observation angles, and meteorological variables to simulate the PM<sub>2.5</sub> concentration (referred to as the Ref-PM<sub>2.5</sub> model). For comparison purposes, a random forest model was also developed based on the AOD and meteorological variables (referred to as the AOD-PM<sub>2.5</sub> model). The input variables for the Ref-PM<sub>2.5</sub> model are PM<sub>2.5</sub> concentrations (as the “true value”), AHI-measured TOA reflectances at three wavelengths (0.47, 0.64, and 2.3 μm), observation angles (SZA, SOA, SAZ, and SAA), the latitudes and longitudes of monitoring sites, all meteorological variables from the ERA-Interim reanalysis, NDVI<sub>SWIR</sub>, and dummy variables (month, day, and hour of observations). Latitudes, longitudes, and dummy variables were included to capture the spatial-temporal variations in the Ref-PM<sub>2.5</sub> and AOD-PM<sub>2.5</sub> relationship due to meteorological conditions (Hu et al., 2017).

Other than other machine learning algorithms (e.g., the deep belief network, the gradient boosted machine), the random forest algorithm has only a few parameters to fine-tune to achieve optimal performance. The most important parameters are the number of trees to grow ( $n_{\text{tree}}$ ), and the number of variables randomly sampled as candidates at each split ( $m_{\text{try}}$ ). The best parameters of the model were determined by changing the values of  $n_{\text{tree}}$  and  $m_{\text{try}}$  until achieving the best prediction accuracy. Comparisons of results generated by the different settings showed that the best performance was achieved when  $n_{\text{tree}}$  and  $m_{\text{try}}$  were assigned values of 1000 and 11, respectively. These values were thus used in the study.

The 10-fold cross-validation (CV) method was then used to check the over-fitting problem and to assess the robustness of the model and its dependence on the subset of training data as used in the many previous studies (e.g. Ma et al., 2016; Li et al., 2017a, etc.). The training dataset was first randomly split into 10 subsets with ~10% of the training data in each subset. Then 9 of the 10 subsets were used for the model training, and the remaining subset was used to validate the model performance. This procedure was repeated 10 times so that every subset was tested. The model performance is then quantitatively evaluated using several statistical indicators, including the coefficient of determination ( $R^2$ ), the root-mean-square error (RMSE,  $\mu\text{g m}^{-3}$ ), and the mean prediction error (MPE,  $\mu\text{g m}^{-3}$ ), between CV estimated and surface-observed PM<sub>2.5</sub> concentrations. The indicators are calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (PM_{2.5}^{\text{obs}}(i) - PM_{2.5}^{\text{mod}}(i))^2}{\sum_{i=1}^n (PM_{2.5}^{\text{obs}}(i) - \bar{PM}_{2.5}^{\text{obs}})^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (PM_{2.5}^{\text{obs}}(i) - PM_{2.5}^{\text{mod}}(i))^2} \quad (3)$$

and

$$\text{MPE} = \frac{1}{n} \sum_{i=1}^n |PM_{2.5}^{\text{obs}}(i) - PM_{2.5}^{\text{mod}}(i)| \quad (4)$$

where  $n$  is the total number of samples;  $PM_{2.5}^{\text{obs}}$  and  $PM_{2.5}^{\text{mod}}$  are the surface-observed and model-estimated PM<sub>2.5</sub> concentrations, respectively.  $\bar{PM}_{2.5}^{\text{obs}}$  is the mean value of the surface-observed PM<sub>2.5</sub> concentration.

## 3. Results

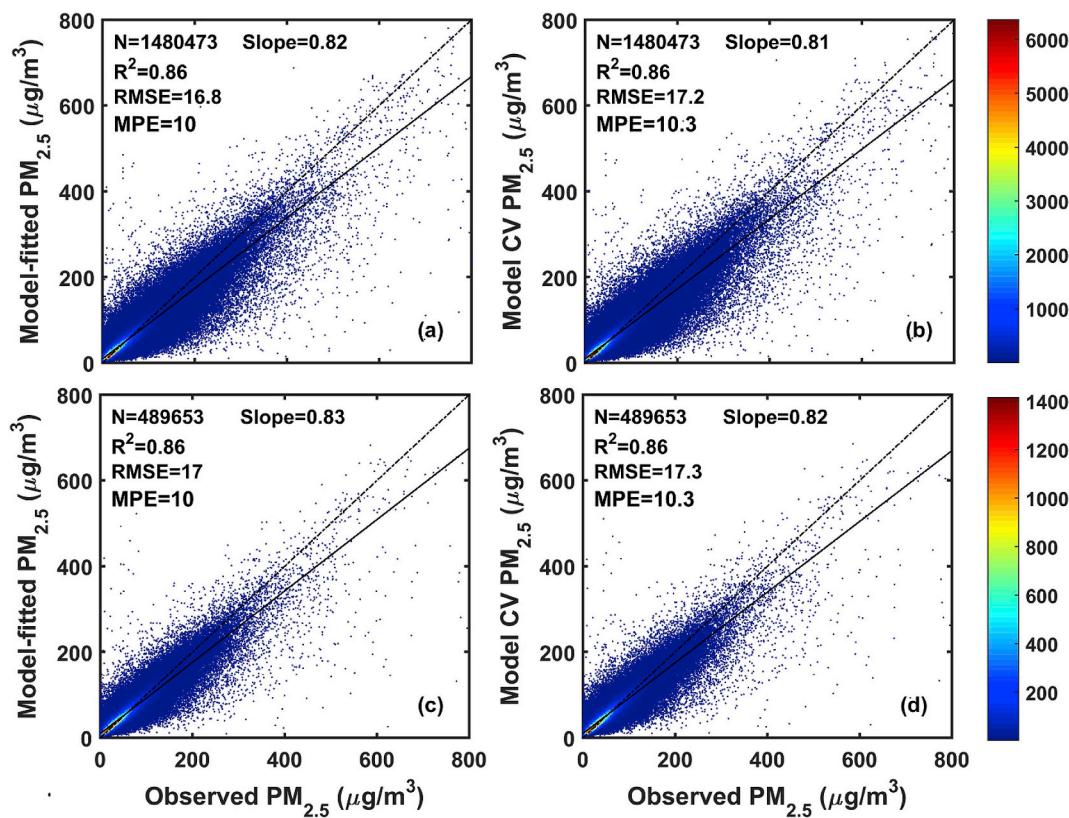
### 3.1. Statistical characteristics

**Table S1** lists the seasonal and annual statistical properties of the Ref-PM<sub>2.5</sub> modeling variables in the training dataset. The PM<sub>2.5</sub> concentration ranges from 1 to 1000  $\mu\text{g m}^{-3}$  with annual mean and standard deviation of 49  $\mu\text{g m}^{-3}$  and 45  $\mu\text{g m}^{-3}$ , respectively. Annual mean (standard deviation) of TOA reflectances at 0.47 μm, 0.64 μm, and 2.3 μm are 0.28 (0.10), 0.20 (0.08), and 0.15 (0.06), respectively. Wintertime mean PM<sub>2.5</sub> concentration and TOA reflectance values are the highest values. Spring and fall mean values are similar, and summertime values are the lowest values. Since the spatial and temporal variabilities of the samples are very large, the range in magnitude of the model variables is also broad.

The Pearson correlation coefficients for all variables are calculated to examine the potential correlation between the predictors and the prediction (Table S3). Results show that the correlation between most of the variables is relatively weak. Note that the random forest algorithm can cope with the problems of complex interactions and is applicable to highly correlated predictor variables (Strobl et al., 2008; Hu et al., 2017).

### 3.2. Model performance

**Fig. 1** gives the model fitted and CV results of the Ref-PM<sub>2.5</sub> (top panels) and AOD-PM<sub>2.5</sub> models (low panels). The observed and Ref-PM<sub>2.5</sub>-model-estimated hourly PM<sub>2.5</sub> concentration is highly correlated ( $p$  value < 0.01), which accounts for 86% of the variability ( $R^2 = 0.86$ ) in the model fitting (Fig. 1a). The RMSE and MPE of the model fitted results are relatively low with the value of 16.8  $\mu\text{g m}^{-3}$  and 10  $\mu\text{g m}^{-3}$ , respectively. The CV results seen in Fig. 1b suggest no substantial overfitting of the Ref-PM<sub>2.5</sub> model ( $R^2 = 0.86$ ). Comparing to the model fitting results, the RMSE and MPE of CV results only increase by 0.4  $\mu\text{g m}^{-3}$  and 0.3  $\mu\text{g m}^{-3}$  with the values of 17.2  $\mu\text{g m}^{-3}$  and 10.3  $\mu\text{g m}^{-3}$ , respectively. The  $R^2$  of the model fitting and CV for



**Fig. 1.** Scatter plots of the results of the model fitting and cross validation (CV) using the Ref-PM<sub>2.5</sub> model (top panels) and the AOD-PM<sub>2.5</sub> model (bottom panels). The color bar shows the counts of data points. The dashed line is the 1:1 line. N: number of samples; R<sup>2</sup>: coefficient of determination; RMSE: root-mean-square error of the predictions ( $\mu\text{g m}^{-3}$ ); MPE: mean prediction error ( $\mu\text{g m}^{-3}$ ). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the AOD-PM<sub>2.5</sub> model (Fig. 1c and d) is also 0.86. The overall RMSE and MPE are  $17 \mu\text{g m}^{-3}$  and  $10 \mu\text{g m}^{-3}$  for model fitting and  $17.3 \mu\text{g m}^{-3}$  and  $10.3 \mu\text{g m}^{-3}$  for model CV, respectively. The performances of Ref-PM<sub>2.5</sub> model is highly consistent with that of AOD-PM<sub>2.5</sub> model, suggesting that PM<sub>2.5</sub> concentrations can be estimated accurately from TOA reflectances in the Ref-PM<sub>2.5</sub> model.

Table 1 gives the CV results of the Ref-PM<sub>2.5</sub> model for each season, which indicated that the model prediction accuracy varies by season. In terms of CV R<sup>2</sup> values, the model has the best performance in winter (December, January, and February, or DJF), followed by fall (September, October, and November, or SON) and spring (March, April, and May, or MAM), and the worst in summer (June, July, and August, or

JJA) with values of 0.86, 0.85, 0.83, and 0.72, respectively. The smallest (largest) RMSE and MPE values are found in summer (winter), partly because PM<sub>2.5</sub> concentrations are at their lowest (highest) then (Li et al., 2015; Ma et al., 2014). The seasonal mean CV-model-simulated PM<sub>2.5</sub> concentration is highly comparable to the seasonal mean surface PM<sub>2.5</sub> concentration.

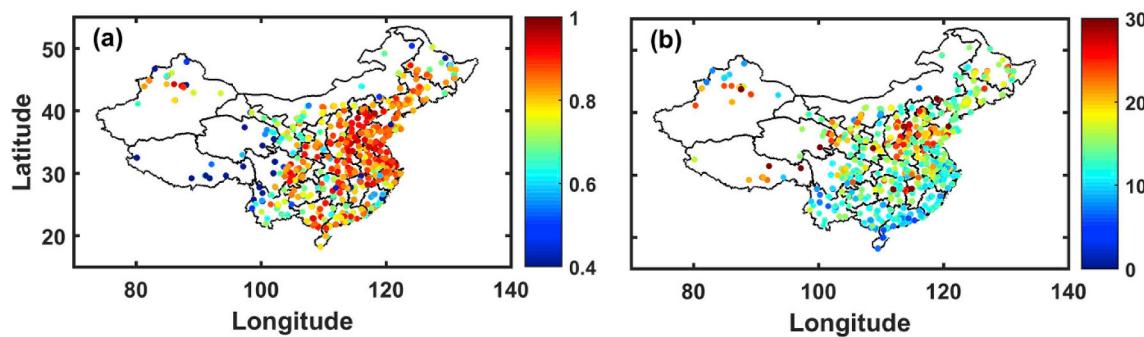
The spatial distributions of the CV R<sup>2</sup> and RMSE of the Ref-PM<sub>2.5</sub> model are shown in Fig. 2. The probability density functions (PDFs) and cumulative density functions (CDFs) of the CV R<sup>2</sup> and RMSE are shown in Fig. 3. Fig. 2a shows that CV site-specific R<sup>2</sup> are generally high with most of the values higher than 0.8 over east/central-east China and are low over western China where surface PM<sub>2.5</sub> sites are scarce, especially

**Table 1**

Results from the ten-fold cross validation of the Ref-PM<sub>2.5</sub> model for different seasons and for different times (local) of the day.

	N	R <sup>2</sup>	RMSE ( $\mu\text{g m}^{-3}$ )	MPE ( $\mu\text{g m}^{-3}$ )	Slope	Estimated PM <sub>2.5</sub> ( $\mu\text{g m}^{-3}$ )	Observed PM <sub>2.5</sub> ( $\mu\text{g m}^{-3}$ )
MAM	400996	0.83	16.0	10.2	0.78	$48.5 \pm 33.3$	$48.0 \pm 38.9$
JJA	367704	0.72	11.4	7.4	0.73	$31.5 \pm 16.2$	$31.2 \pm 21.1$
SON	311912	0.85	16.3	10.1	0.79	$49.0 \pm 35.6$	$48.5 \pm 41.3$
DJF	399861	0.86	22.7	13.1	0.82	$67.9 \pm 54.0$	$67.3 \pm 61.1$
08:00–09:00	133116	0.84	17.2	10.7	0.81	$54.7 \pm 37.7$	$54.0 \pm 42.7$
09:00–10:00	155430	0.85	19.0	11.1	0.81	$58.3 \pm 43.3$	$58.4 \pm 49.4$
10:00–11:00	158646	0.87	19.0	11.3	0.81	$58.8 \pm 44.8$	$58.6 \pm 51.2$
11:00–12:00	155206	0.84	20.4	12.1	0.80	$57.6 \pm 44.4$	$56.8 \pm 51.1$
12:00–13:00	153382	0.85	19.7	11.5	0.79	$53.3 \pm 43.0$	$53.4 \pm 49.8$
13:00–14:00	146520	0.89	16.3	9.8	0.83	$49.6 \pm 42.1$	$49.4 \pm 47.4$
14:00–15:00	146474	0.88	15.7	9.4	0.84	$46.9 \pm 40.5$	$45.8 \pm 45.0$
15:00–16:00	136342	0.86	15.5	9.4	0.82	$43.2 \pm 36.6$	$42.4 \pm 41.3$
16:00–17:00	109711	0.81	14.0	8.6	0.77	$36.5 \pm 27.2$	$35.7 \pm 31.8$
17:00–18:00	60454	0.75	12.1	7.6	0.74	$28.3 \pm 19.8$	$27.7 \pm 23.9$

MAM: March, April, May; JJA: June, July, August; SON: September, October, November; DJF: December, January, February; N: number of samples; R<sup>2</sup>: coefficient of determination; RMSE: root-mean-square error; MPE: mean prediction error.



**Fig. 2.** The cross-validation (a) coefficient of determination ( $R^2$ ) and (b) root-mean-square error (RMSE;  $\mu\text{g m}^{-3}$ ) of the Ref-PM<sub>2.5</sub> model at each site over mainland China.

over the Tibetan region. The CV  $R^2$  of all sites varies from 0.25 to 0.97 with more values located between 0.8 and 0.95. The CDF of  $R^2$  shows that approximately 70% of the sites have CV  $R^2$  values higher than 0.8 (Fig. 3a). Fig. 2b shows that CV RMSE values are less than 20  $\mu\text{g m}^{-3}$  at most sites. Relatively high RMSE values are found over northern China, such as in Hebei, Shanxi, Henan, and Shandong provinces, where PM<sub>2.5</sub> concentrations are generally high. The PDF and CDF of site-specific CV RMSEs of hourly PM<sub>2.5</sub> concentrations (Fig. 3b) show that values range from 4.7 to 51.8  $\mu\text{g m}^{-3}$ . More than 80% of PM<sub>2.5</sub> sites have CV RMSE values lower than 20  $\mu\text{g m}^{-3}$ . The histogram of the bias of model CV estimated PM<sub>2.5</sub> concentration, defined as the model-estimated PM<sub>2.5</sub> concentration minus surface-measured PM<sub>2.5</sub> concentration, shows that the distribution of the bias is centered on  $\sim 0 \mu\text{g m}^{-3}$  (Fig. S2). About 88% and 66% of monitoring sites have an absolute bias of less than 20  $\mu\text{g m}^{-3}$  and 10  $\mu\text{g m}^{-3}$ , respectively. It suggests that the predication accuracy of the Ref-PM<sub>2.5</sub> model is relatively high in the central-eastern of China and low in southwest China, which is consistent with the previous studies from other machine learning models (e.g. Ma et al., 2014; Fang et al., 2016; Li et al., 2017a). Overall, the Ref-PM<sub>2.5</sub> model has high performance on PM<sub>2.5</sub> prediction for the most surface sites.

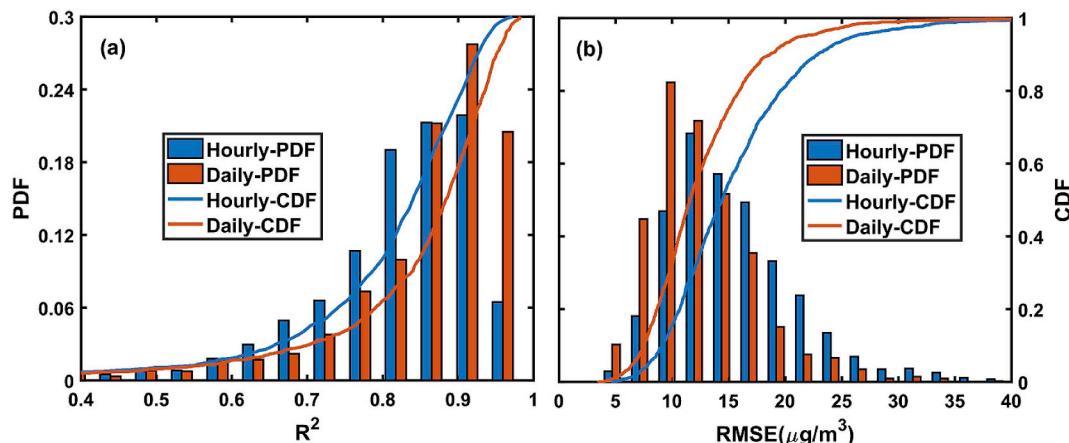
Table 1 gives the CV results of Ref-PM<sub>2.5</sub>-model-simulated and surface-observed PM<sub>2.5</sub> concentration for hours from 8:00–18:00 local time (LT). The CV  $R^2$  ranges from 0.75 to 0.89 for different hours with the highest  $R^2$  occurring from 13:00 to 14:00 LT and the lowest  $R^2$  occurring from 17:00 to 18:00 LT. During the day, CV RMSE values vary from 12.1 to 20.4  $\mu\text{g m}^{-3}$  and MPE values vary from 7.6 to 12.1  $\mu\text{g m}^{-3}$ . The results indicate that mean value of CV Ref-PM<sub>2.5</sub>-model-estimated and surface-observed PM<sub>2.5</sub> concentrations are consistent at all hours. The high  $R^2$  and low RMSE and MPE values for all hours during the day suggest that the Ref-PM<sub>2.5</sub> model can estimate

hourly PM<sub>2.5</sub> concentrations well, offering a way of examining the diurnal cycle of PM<sub>2.5</sub> concentration so that its evolution over time can be better understood.

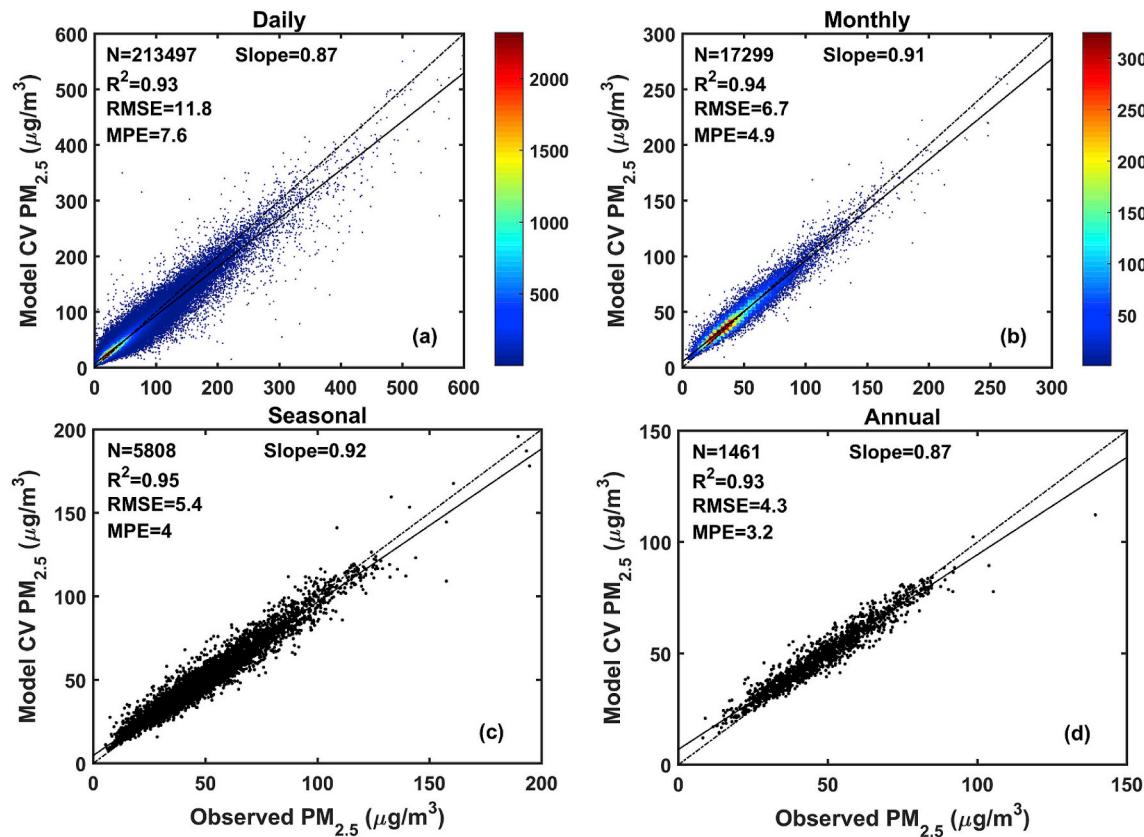
The model performance at different timescales is also evaluated. Fig. 4 shows the cross validation of model-estimated PM<sub>2.5</sub> concentrations as a function of ground-measured PM<sub>2.5</sub> concentration at different timescales. At the daily timescale, the Ref-PM<sub>2.5</sub> model explains 93% ( $R^2 = 0.93$ ) of the variability in PM<sub>2.5</sub> concentration (Fig. 4a). Fig. 3a shows that the CV site-specific  $R^2$  at the daily timescale ranges from 0.40 to 0.98 with nearly 80% of PM<sub>2.5</sub> sites having CV  $R^2$  values higher than 0.8. The overall RMSE and MPE values for daily PM<sub>2.5</sub> concentrations are 11.8  $\mu\text{g m}^{-3}$  and 7.6  $\mu\text{g m}^{-3}$ , respectively. The PDF and CDF of the site-specific RMSE of daily PM<sub>2.5</sub> concentrations (Fig. 3b) show that the RMSE changes from 3.5  $\mu\text{g m}^{-3}$  to 40.0  $\mu\text{g m}^{-3}$  with  $\sim 93\%$  of monitoring sites having RMSE values lower than 20  $\mu\text{g m}^{-3}$  at this timescale. Fig. S2b supports this, showing that  $\sim 93\%$  and 76% of PM<sub>2.5</sub> sites have absolute biases lower than 20  $\mu\text{g m}^{-3}$  and 10  $\mu\text{g m}^{-3}$ , respectively. The CV  $R^2$  (RMSE, MPE) at monthly, seasonal, and annual timescales are 0.94 (6.7  $\mu\text{g m}^{-3}$ , 4.9  $\mu\text{g m}^{-3}$ ), 0.95 (5.4  $\mu\text{g m}^{-3}$ , 4  $\mu\text{g m}^{-3}$ ), and 0.93 (4.3  $\mu\text{g m}^{-3}$ , 3.2  $\mu\text{g m}^{-3}$ ), respectively. This suggests that the Ref-PM<sub>2.5</sub> model showed a strong power in estimating PM<sub>2.5</sub> concentrations at all timescales well.

### 3.3. Model estimated PM<sub>2.5</sub> concentration over three developed regions

The regions of the Beijing-Tianjin-Hebei (BTH), the Yangtze River Delta (YRD), and the Pearl River Delta (PRD) are three major megalopolises of China, where heavy pollution episodes frequently and periodically occur. Fig. 5 gives the spatial variations of PM<sub>2.5</sub> concentration from Ref-PM<sub>2.5</sub> model-calculated and surface-observed over



**Fig. 3.** Probability density functions (PDFs, bars) and cumulative density functions (CDFs, lines) of the hourly (blue) and daily (orange) cross-validation (a) coefficient of determination ( $R^2$ ) and (b) root-mean-square error (RMSE). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



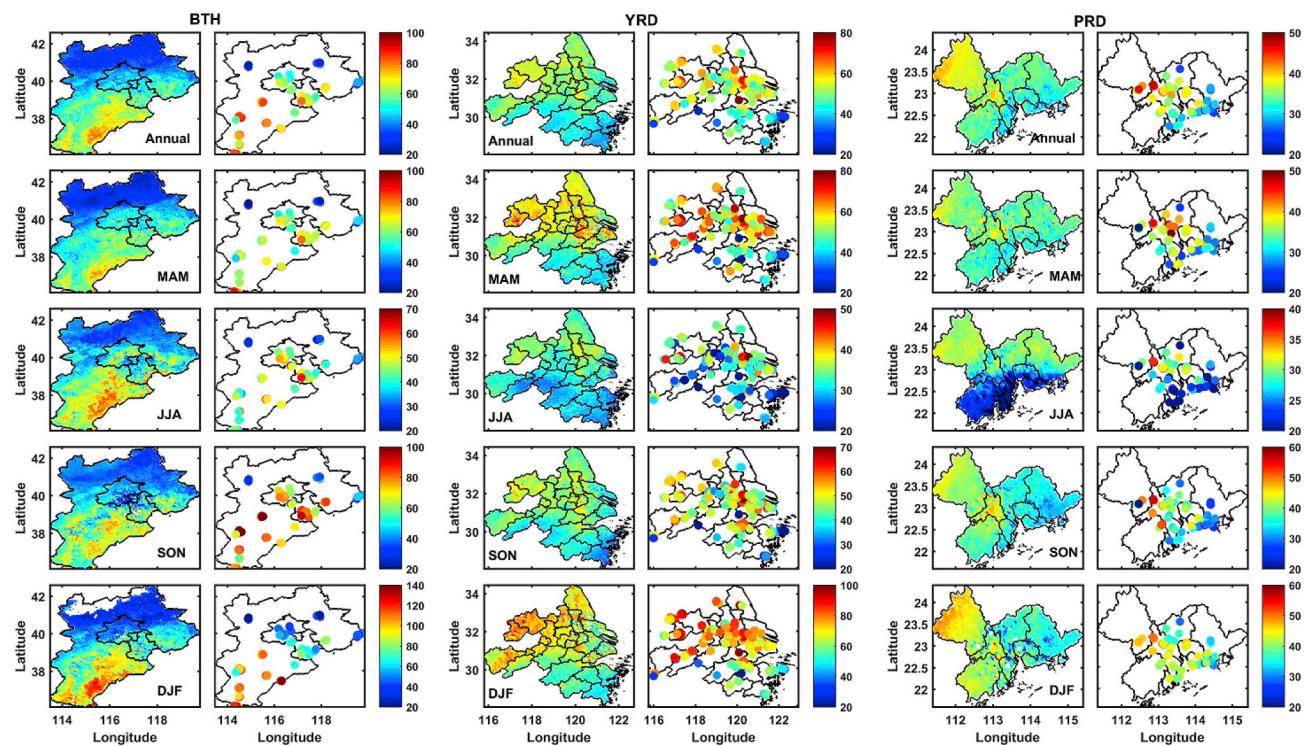
**Fig. 4.** The cross-validation (CV) results of the Ref-PM<sub>2.5</sub> model at (a) daily, (b) monthly, (c) seasonal, and (d) annual timescales. The color bars in (a) and (b) show the counts of data points. The dashed line is the 1:1 line. N: number of samples; R<sup>2</sup>: coefficient of determination; RMSE: root-mean-square error of the predictions ( $\mu\text{g m}^{-3}$ ); MPE: mean prediction error ( $\mu\text{g m}^{-3}$ ). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the three regions for annual and each season. It indicated that annually and seasonally average PM<sub>2.5</sub> concentrations from model-simulated are highly consistent with those from the surface-measured in all three regions. Among the three regions, BTH has the highest PM<sub>2.5</sub> concentration with a mean estimated PM<sub>2.5</sub> concentration of  $50.1 \mu\text{g m}^{-3}$ , followed by the YRD region with an averaged PM<sub>2.5</sub> concentration of  $46.8 \mu\text{g m}^{-3}$ . The PM<sub>2.5</sub> concentrations in the PRD region are relatively lower than those in other two regions with annually average value of  $34.5 \mu\text{g m}^{-3}$ . High PM<sub>2.5</sub> concentrations occur in southern BTH with annual mean values greater than  $60 \mu\text{g m}^{-3}$  and low in northern BTH, especially the mountainous area of northern Beijing City, with annual means less than  $40 \mu\text{g m}^{-3}$ . The highest PM<sub>2.5</sub> concentrations are found in winter with an average of  $67.2 \mu\text{g m}^{-3}$ . The seasonal mean PM<sub>2.5</sub> concentration in autumn and spring is similar with values of  $50.3 \mu\text{g m}^{-3}$  and  $47.8 \mu\text{g m}^{-3}$ , respectively, whereas the lowest value ( $36.5 \mu\text{g m}^{-3}$ ) is seen in summer. In the YRD region, higher PM<sub>2.5</sub> concentrations are seen in the north. Lower concentrations are seen in the hilly southern part of the region where human influences are low. The annual mean value is  $\sim 60 \mu\text{g m}^{-3}$  in the north and is less than  $45 \mu\text{g m}^{-3}$  in the south. PM<sub>2.5</sub> concentration is the largest in winter, followed by spring and autumn, and summer has the lowest PM<sub>2.5</sub> concentration with mean values of  $63.4, 50.8, 41.8$ , and  $31.6 \mu\text{g m}^{-3}$ , respectively. PM<sub>2.5</sub> concentrations over PRD show that the higher concentrations occur over the northwest region, e.g., western Guangzhou, with annually average value of  $\sim 40 \mu\text{g m}^{-3}$  and lower in the eastern coastal region with annual mean values less than  $30 \mu\text{g m}^{-3}$  on average. The seasonal variations over PRD are similar to those over the BTH region, i.e., the highest concentration occurs in winter, followed by autumn and spring, and the lowest mean PM<sub>2.5</sub> concentration occurs in summer. The average values of PM<sub>2.5</sub> concentration from

spring to winter are  $33.5, 25.7, 37.8$ , and  $41.2 \mu\text{g m}^{-3}$ , respectively. The highest (lowest) PM<sub>2.5</sub> concentrations over three regions in winter (summer) can be partly attributed to the higher (lower) atmospheric stability and less (more) precipitation. The emission from the winter heating over BTH region is also one of the reasons for the highest PM<sub>2.5</sub> concentrations in winter. The distributions of variation in predicted PM<sub>2.5</sub> concentrations by our model are in agreement with previous related studies from the satellite-derived AOD with different models over these three megalopolises (Ma et al., 2016; Zheng et al., 2016; Xiao et al., 2017; Lv et al., 2017; He and Huang, 2018).

#### 4. Discussion

Different from the most of published studies for estimating surface PM<sub>2.5</sub> concentrations based on statistical models that use different satellite-retrieved AODs as the primary predictor, this study presents a model for estimating surface PM<sub>2.5</sub> concentrations using a machine learning algorithm, i.e., the random forest, that directly uses satellite-measured TOA reflectances as the primary predictor (Ref-PM<sub>2.5</sub> model). We also built a model for PM<sub>2.5</sub> estimation from AOD (AOD-PM<sub>2.5</sub> model) using the same algorithm behind the Ref-PM<sub>2.5</sub> model. The overall cross validation shows that the R<sup>2</sup> (RMSE, MPE) of PM<sub>2.5</sub> concentration from the Ref-PM<sub>2.5</sub> model is 0.86 ( $17.2 \mu\text{g m}^{-3}$ ,  $10.3 \mu\text{g m}^{-3}$ ) at hourly scale and that the model performance is comparable to the capability of the AOD-PM<sub>2.5</sub> model which has a cross validation R<sup>2</sup> (RMSE, MPE) of 0.86 ( $17.3 \mu\text{g m}^{-3}$ ,  $10.3 \mu\text{g m}^{-3}$ ). We have also summarized some previous relative studies using satellite-derived AODs based on different statistical models applied over the national scale of China (Table 2). In terms of CV R<sup>2</sup> and RMSE, our Ref-PM<sub>2.5</sub> model better predicts PM<sub>2.5</sub> concentrations than most of the



**Fig. 5.** Spatial variation of annually and seasonally average  $\text{PM}_{2.5}$  concentrations from the Ref- $\text{PM}_{2.5}$  model-estimated and surface-measured in the Beijing-Tianjin-Hebei region (BTH, two left columns), the Yangtze River Delta region (YRD, two middle columns), and the Pearl River Delta region (PRD, two right columns). Model estimates (observations) are shown in the left (right) column of each set of two columns. Units:  $\mu\text{g m}^{-3}$ . MAM: March, April, May; JJA: June, July, August; SON: September, October, November; DJF: December, January, February.

**Table 2**

Summary of estimates of  $\text{PM}_{2.5}$  concentration from satellite-retrieved AOD based on statistical models applied on the national scale of China.

Reference	Model	$R^2$ , RMSE ( $\mu\text{g m}^{-3}$ ), and Slope (sample-based CV)	Time Scale	Study Period	Source of AOD
Ma et al. (2014)	GWR	0.64, 32.98, 0.67	daily	2000–2013	MODIS and MISR
Fang et al. (2016)	TSAM	0.80, 22.75, 0.79	daily	Jun. 2013–May. 2014	MODIS
Ma et al. (2016)	LME	0.78, 27.99, 0.77	daily	Jan. 2004–Jun. 2014	MODIS
	LME + GAM	0.79, 27.42, 0.79			
Li et al. (2016)	GRNN	0.67, 20.93, 0.62	daily	Feb. 2013–Dec. 2014	MODIS
You et al. (2016a)	GWR	0.79, 18.6, 0.83	daily	2014	MODIS
You et al. (2016b)	GWR	MODIS: 0.79, 20.85, 0.82 MSIR: 0.85, 24.86, 0.87	daily	2014	MODIS MISR
Li et al. (2017a)	DBN	0.88, 13.03, 0.88	daily	2015	MODIS
Yu et al. (2017)	Gaussian process	0.81, 21.87, 0.73	daily	2013	MODIS
Zhan et al. (2017)	GW-GBM	0.76, 23.0, 0.77	daily	2014	MODIS
This study	RF	0.86, 17.2, 0.81	hourly	2016	AHI

CV: cross validation.

MODIS: Moderate Resolution Imaging Spectroradiometer.

MISR: Multi-angle Imaging Spectro Radiometer.

AHI: Advanced Himawari Imager.

GWR: geographically weighted regression model; LME: linear mixed effects model.

GAM: generalized additive model; TSAM: Timely structure adaptive modeling.

GRNN: generalized regression neural network (GRNN) mode; DBN: deep belief network.

GW-GBM: geographically weighted gradient boosting machine. RF: random forest.

previous studies summarized in Table 2 and is comparable to some machine-learning approaches (e.g. Li et al., 2017a). AOD retrievals from satellite measurements have some inherent challenges such as the quantification of the contribution of surface reflectances to TOA reflectances over bright surfaces and the accurate assignment of aerosol type, etc. (Li et al., 2015). This study proposes a short-cut solution for  $\text{PM}_{2.5}$  concentration estimations from satellite observations by directly using TOA reflectance which can circumvent the numerous sources of errors in the retrievals of AOD. Another advantage of the Ref- $\text{PM}_{2.5}$  model is that the sample size of the TOA reflectance is much higher than that of the retrieved AOD. The Ref- $\text{PM}_{2.5}$  modeling dataset has a

total of 1,480,473 samples (Fig. 1) which is more than three times the number of samples from the AOD- $\text{PM}_{2.5}$  modeling dataset ( $N = 489,653$ ). This is mainly because satellite AOD retrieval algorithms cannot accurately retrieve AODs over bright surfaces (Levy et al., 2007). Fig. S3 shows the cross-validation results of the Ref- $\text{PM}_{2.5}$  model for samples (a) with TOA reflectance but without AOD available and (b) with both AOD and TOA reflectance available. Pixels without AOD available are associated with high surface reflectance or a low retrieval confidence level or both. Fig. S3a shows that the model estimates  $\text{PM}_{2.5}$  concentrations well for pixels where AOD cannot be reliably retrieved. Fig. S3b shows that the performance of the Ref- $\text{PM}_{2.5}$

model is comparable or slightly better than that of the AOD-PM<sub>2.5</sub> model (as shown in Fig. 1d). In general, estimates of PM<sub>2.5</sub> concentration based on TOA reflectances and based on AOD retrievals have comparable accuracies. However, the Ref-PM<sub>2.5</sub> model has a relatively stronger predictive power than the AOD-PM<sub>2.5</sub> model when it comes to the spatiotemporal coverage of PM<sub>2.5</sub> concentration estimates.

Underestimations of PM<sub>2.5</sub> concentrations at national and regional of China were reported with magnitude of ~10%–30% in most previous studies used satellite-retrieved AOD based on different algorithms (e.g. Lin et al., 2015; You et al., 2016a; Zheng et al., 2016; Ma et al., 2016; Li et al., 2017a; Yu et al., 2017) and are seen in the results from AOD-PM<sub>2.5</sub> models in the current study. Such an underestimation is also seen in the results from the Ref-PM<sub>2.5</sub> model whose hourly CV slope ranges from 0.74 to 0.84 at different local hours (the overall CV slope is 0.81). This underestimation was mitigated for averaging PM<sub>2.5</sub> concentration estimations at daily, monthly, seasonal, and annual timescales with CV slope values of 0.87, 0.91, 0.92, and 0.87, respectively. Possible reasons were discussed in previous works (e.g. Li et al., 2017a), e.g., the model training using site-based PM<sub>2.5</sub> measurements possibly can not completely represent the whole satellite pixel. Due to the similar spectral reflectance characteristics of haze and low clouds, distinguishing between the two using passive sensors is very challenging, especially in China where hazy weather and high pollution cases are common. Studies have indicated that the current cloud mask algorithms tend to mistake haze for clouds under heavily polluted conditions (Shang et al., 2017; Tan et al., 2018). This misclassification leads to lack the severe polluted cases in the model training dataset can possibly partly contribute to the underestimation of PM<sub>2.5</sub> concentration under heavy pollution conditions.

In the current study, reanalysis meteorological field data are used to improve the model performance. Reanalysis meteorological data have coarse temporal resolutions (two times or four times per day), and their accuracy still needs more validation. The coarse temporal resolution of meteorological variables can result in a temporal mismatch among meteorological data, satellite measurements, and PM<sub>2.5</sub> concentrations, which may incur uncertainties in estimating the PM<sub>2.5</sub>. The proposed model performance may possibly be improved if high temporal ground meteorological observations over the sites were used. For example, Su et al. (2017) showed that lidar-derived PBLHs were more effective for PM<sub>2.5</sub> remote sensing than climatologies of radiosonde- and reanalysis-derived PBLHs.

More predictor variables, including without limitation land-use variables (e.g. coverage of agriculture, urban, rural areas, etc.), population data, and elevation data have been shown to contribute toward the performance of PM<sub>2.5</sub> prediction models (Ma et al., 2016). The column amounts of nitrogen dioxide (NO<sub>2</sub>) can represent the strength of anthropogenic emissions such as fossil fuel consumption (Zhang et al., 2012), so they may reflect the strong relationship between ambient PM<sub>2.5</sub> concentrations and local emitting sources. Moreover, since NO<sub>2</sub> concentrations are commonly used to correct atmospheric gas absorption and scattering in AOD retrieval algorithms, the NO<sub>2</sub> concentration may also contribute to the magnitude of the TOA reflectance. It is noted that even without the inclusion of those variables, our Ref-PM<sub>2.5</sub> model predicted PM<sub>2.5</sub> concentration well.

Understanding the results is still a major challenge when determining the effectiveness of machine learning algorithms. The prediction performance of machine learning algorithms is acceptable and relatively static over the observation time period of the training dataset but deteriorates after application beyond the realm of the training data. Machine learning algorithms, therefore, have some limitations when it comes to extrapolation (Reichstein et al., 2019). The model can be updated simultaneously by intaking real-time satellite measurements, surface PM<sub>2.5</sub> observations, and related meteorological variables to predict PM<sub>2.5</sub> concentrations over regions without surface PM<sub>2.5</sub> measurements.

There are many machine learning algorithms that can be used for

such studies. Unlike many other machine learning algorithms (e.g., the deep belief network, the gradient boosted machine), the random forest is user-friendly and has only two parameters to fine-tune to achieve optimal performance. The random forest algorithm is also one of the most effective machine learning algorithms for a wide variety of regression tasks. It can cope with the problem of complex interactions and is applicable to highly correlated predictor variables (Strobl et al., 2008; Hu et al., 2017). Systematic comparisons of the applicability and performance of different machine learning algorithms on PM<sub>2.5</sub> estimations need to be done. Although not beyond the scope of this study, this will be considered in future work.

## 5. Conclusions

Different from commonly used methods which predict PM<sub>2.5</sub> concentrations from satellite-retrieved AOD, the current study developed a model based on an assemble learning algorithm for predicting PM<sub>2.5</sub> concentrations using TOA reflectances as inputs instead of AODs, together with meteorological variables. CV results indicated that the Ref-PM<sub>2.5</sub> model explain 86% ( $R^2 = 0.86$ ) of the overall hourly variability in PM<sub>2.5</sub> concentration. The distribution of  $R^2$  showed that approximately 70% of PM<sub>2.5</sub> monitoring sites had CV  $R^2$  values greater than 0.8. The overall CV RMSE was  $17.2 \mu\text{g m}^{-3}$  with more than 80% of the values less than  $20 \mu\text{g m}^{-3}$ . The Ref-PM<sub>2.5</sub> model predicted PM<sub>2.5</sub> well for daily, monthly, seasonal, and annual timescales with  $R^2$  of 0.93, 0.94, 0.95, and 0.93, respectively. The performance of PM<sub>2.5</sub> concentrations estimated directly from satellite TOA reflectances is comparable with the performance of PM<sub>2.5</sub> concentration predication by using the satellite retrieved AODs in the current study and most previous studies. Compared with the AOD-PM<sub>2.5</sub> model, the Ref-PM<sub>2.5</sub> model had a relatively stronger predictive power in terms of spatiotemporal coverage. Using the Ref-PM<sub>2.5</sub> model, PM<sub>2.5</sub> concentrations over three developed regions, BTH, YRD, and PRD, were estimated and analyzed. The Ref-PM<sub>2.5</sub> model can capture spatial variations of PM<sub>2.5</sub> at different timescales. This may provide information about the diurnal cycle of PM<sub>2.5</sub> concentrations as well as aid in monitoring the process of regional pollution episodes and the evolution of PM<sub>2.5</sub> concentrations.

Traditionally, compared with the TOA reflectance, the AOD is regarded as more related with the surface PM<sub>2.5</sub> concentration. This study proposes a short-cut solution for PM<sub>2.5</sub> concentration estimations from satellite observations by using TOA reflectance. The result implies that the machine learning algorithms can possibly capture the complicated relationship between the initial predictors and the final predictions, which possibly simplify and improve the solutions for the questions by avoiding the complicated intermediated processing.

## Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

ERA-Interim reanalysis data were downloaded from <http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/>. The JAXA P-Tree system provides the Himawari Standard Data, which is provided by the Japan Meteorological Agency, Chinese Ministry of Science and Technology, China, project ID:2018YFC1506500. M. Cribb helped edit the manuscript.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.atmosenv.2019.04.002>.

## References

- Aerosol Product Application, 2012. Team of the AWG Aerosols/Air Quality/Atmospheric Chemistry Team: GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document for Suspended Matter/Aerosol Optical Depth and Aerosol Size Parameter, NOAA, NESDIS, Center for Satellite Applications and Research, Version 3.0, July 30.
- Breiman, L., 2001. Statistical modeling: the two cultures. *Stat. Sci.* 16, 199–215.
- 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, 2495–2509. <https://doi.org/10.1016/j.atmosenv.2004.01.039>.
- Fan, J., Rosenfeld, D., Yang, Y., Zhao, C., Ruby Leung, L., Li, Z., 2015. Substantial contribution of anthropogenic air pollution to catastrophic floods in Southwest China. *Geophys. Res. Lett.* 42, 6066–6075. <https://doi.org/10.1002/2015GL064479>.
- Fang, X., Zou, B., Liu, X., Sternberg, T., Zhai, L., 2016. Satellite-based ground PM<sub>2.5</sub> estimation using timely structure adaptive modeling. *Remote Sens. Environ.* 186, 152–163. <https://doi.org/10.1016/j.rse.2016.08.027>.
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J., 2015. Artificial neural networks forecasting of PM<sub>2.5</sub> pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos. Environ.* 107, 118–128. <https://doi.org/10.1016/j.atmosenv.2015.02.030>.
- Geng, G., Zhang, Q., Martin, R.V., Van Donkelaar, A., Huo, H., Che, H., Lin, J., He, K., 2015. Estimating long-term PM<sub>2.5</sub> concentrations in China using satellite-based aerosol optical depth and a chemical transport model. *Remote Sens. Environ.* 166, 262–270. <https://doi.org/10.1016/j.rse.2015.05.016>.
- Gupta, P., Christopher, S.A., 2009. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: multiple regression approach. *J. Geophys. Res.* 114, D14205. <https://doi.org/10.1029/2008JD011496>.
- Gupta, P., Levy, R., Mattoo, S., Remer, L., Munchak, L., 2016. A surface reflectance scheme for retrieving aerosol optical depth over urban surfaces in MODIS Dark Target retrieval algorithm. *Atmos. Meas. Tech.* 9 (7), 3293–3308. <https://doi.org/10.5194/amt-9-3293-2016>.
- He, Q., Huang, B., 2018. Satellite-based high-resolution PM<sub>2.5</sub> estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model. *Environ. Pollut.* 236, 1027–1037. <https://doi.org/10.1016/j.envpol.2018.01.053>.
- Hou, W., Li, Z., Zhang, Y., Xu, H., Zhang, Y., Li, K., Li, D., Wei, P., Ma, Y., 2014. Using support vector regression to predict PM<sub>10</sub> and PM<sub>2.5</sub>. IOP Conf. Ser. Earth Environ. Sci. 17 (1), 012268.
- Hu, X., Waller, L.A., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G., Estes, S.M., Quattrochi, D.A., Sarnat, J.A., Liu, Y., 2013. Estimating ground-level PM<sub>2.5</sub> concentrations in the southeastern U.S. using geographically weighted regression. *Environ. Res.* 121, 1–10. <https://doi.org/10.1016/j.enres.2012.11.003>.
- Hu, X., Belle, J., Meng, X., Wildani, A., Waller, L., Strickland, M., Liu, Y., 2017. Estimating PM<sub>2.5</sub> concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol.* 51 (12), 6936–6944. <https://doi.org/10.1021/acs.est.7b01210>.
- Kan, H.D., Chen, B.H., 2002. Meta-analysis of exposure-response functions of air particulate matter and adverse health outcomes in China. *J. Environ. Health* 19 (6), 422–424.
- Kaufman, Y.J., Tanré, D., Remer, L.A., Vermote, E., Chu, A., Holben, B.N., 1997. Operational remote sensing of tropospheric aerosol over land from EOS Moderate Resolution Imaging Spectroradiometer. *J. Geophys. Res.* 102, 17 051–17 067. <https://doi.org/10.1029/96JD03988>.
- Lee, H.J., Coull, B.A., Bell, M.L., Kourakis, P., 2012. Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM<sub>2.5</sub> concentrations. *Environ. Res.* 118, 8–15. <https://doi.org/10.1016/j.enres.2012.06.011>.
- Lee, H.J., Liu, Y., Coull, B.A., Schwartz, J., Kourakis, P., 2011. A novel calibration approach of MODIS AOD data to predict PM<sub>2.5</sub> concentrations. *Atmos. Chem. Phys.* 11, 7991–8002. <https://doi.org/10.5194/acp-11-7991-2011>.
- Lelieveld, J., Evans, J., Fnais, M., Giannadaki, D., Poszer, A., 2015. The contribution of out-door air pollution sources to premature mortality on a global scale. *Nature* 525, 367–371. <https://doi.org/10.1038/nature15371>.
- 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. *Atmos. Meas. Tech.* 6, 2989–3034. <https://doi.org/10.5194/amt-6-2989-2013>.
- Levy, R.C., Remer, L.A., Mattoo, S., Vermote, E.F., Kaufman, Y.J., 2007. Second-generation operational algorithm: retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *J. Geophys. Res. D Atmos.* 112, D13211. <https://doi.org/10.1029/2006JD007811>.
- Li, R., Gong, J., Chen, L., Wang, Z., 2015. Estimating ground-level PM<sub>2.5</sub> using fine-resolution satellite data in the megacity of Beijing, China. *Aerosol Air Qual. Res.* 15, 1347–1356. <https://doi.org/10.4209/aaqr.2015.01.0009>.
- Li, T., Shen, H., Yuan, Q., Zhang, X., Zhang, L., 2017a. Estimating ground level PM<sub>2.5</sub> by fusing satellite and station observations: geo-intelligent deep learning approach. *Geophys. Res. Lett.* 44, 11,985–11,993. <https://doi.org/10.1002/2017GL075710>.
- Li, T., Shen, H., Zeng, C., Yuan, Q., Zhang, L., 2017b. Point-surface fusion of station measurements and satellite observations for mapping PM<sub>2.5</sub> distribution in China: methods and assessment. *Atmos. Environ.* 152, 477–489. <https://doi.org/10.1016/j.atmosenv.2017.01.004>.
- Li, Z., Lau, W.K.M., Ramanathan, V., Wu, G., Ding, Y., Manoj, M.G., Liu, J., Qian, Y., Li, J., Zhou, T., Fan, J., Rosenfeld, D., Ming, Y., Wang, Y., Huang, J., Wang, B., Xu, X., Lee, S.S., Cribb, M., Zhang, F., Yang, X., Zhao, C., Takemura, T., Wang, K., Xia, X., Yin, Y., Zhang, H., Guo, J., Zhai, P.M., Sugimoto, N., Brasseur, G.P., 2016. Aerosol and monsoon climate interactions over Asia. *Rev. Geophys.* 54, 4. <https://doi.org/10.1002/2015RG000500>.
- Li, Z., Rosenfeld, D., Fan, J., 2017. Aerosols and their impact on radiation, clouds, precipitation, and severe weather events. *Oxf. Res. Encyclopedias.* <https://doi.org/10.1093/acrefore/9780199389414.013.126>.
- 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, 2755–2770. <https://doi.org/10.5194/angeo-27-2755-2009>.
- Liaw, A., Wiener, M., 2002. Classification and regression by random forest. *R. News* 2 (3), 18–22.
- 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 PM<sub>2.5</sub>. *Remote Sens. Environ.* 156, 117–128. <https://doi.org/10.1016/j.rse.2014.09.015>.
- Liu, J., Zheng, Y., Li, Z., Flynn, C., Cribb, M., 2012. Seasonal variations of aerosol optical properties, vertical distribution and associated radiative effects in the Yangtze Delta region of China. *J. Geophys. Res.* 117, D00K38. <https://doi.org/10.1029/2011JD016490>.
- Liu, J., Li, Z., 2018. First surface-based estimation of the aerosol indirect effect over a site in southeastern China. *Adv. Atmos. Sci.* 35 (2), 169–181. <https://doi.org/10.1007/s00376-017-7106-2>.
- Liu, Y., Park, R.J., Jacob, D.J., Li, Q., Kilaru, V., Sarnat, J.A., 2004. Mapping annual mean ground-level PM<sub>2.5</sub> concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. *J. Geophys. Res.: Atmospheres* 109, D22206. <http://doi.org/10.1029/2004JD005025>.
- Liu, Y., 2014. Monitoring PM<sub>2.5</sub> from space for health: past, present, and future directions. *Environ. Monit. Assess.* 6, 6–10.
- Liu, Y., Paciorek, C.J., Kourakis, P., 2009. Estimating regional spatial and temporal variability of PM<sub>2.5</sub> concentrations using satellite data, meteorology, and land use information. *Environ. Health Perspect.* 117 (6), 886–892. <https://dx.doi.org/10.1289/ehp.0800123>.
- Liu, Y., Franklin, M., Kahn, R., Kourakis, P., 2007. Using aerosol optical thickness to predict ground-level PM<sub>2.5</sub> concentrations in the St. Louis area: a comparison between MISR and MODIS. *Remote Sens. Environ.* 107, 33–44. <http://doi.org/10.1016/j.rse.2006.05.022>.
- Lv, B., Hu, Y., Chang, H.H., Russell, A.G., Cai, J., Xu, B., Bai, Y., 2017. Daily estimation of ground-level PM<sub>2.5</sub> concentrations at 4km resolution over Beijing-Tianjin-Hebei by fusing MODIS AOD and ground observations. *Sci. Total Environ.* 580, 235–244. [https://doi.org/10.1016/j.scitotenv.2016.12.049](http://doi.org/10.1016/j.scitotenv.2016.12.049).
- Lv, B., Hu, Y., Chang, H.H., Russell, A.G., Bai, Y., 2016. Improving the accuracy of daily PM<sub>2.5</sub> 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. <http://doi.org/10.1021/acs.est.5b05940>.
- Ma, Z.W., Hu, X.F., Huang, L., Bi, J., Liu, Y., 2014. Estimating ground-level PM<sub>2.5</sub> in China using satellite remote sensing. *Environ. Sci. Technol.* 48, 7436–7444. <https://doi.org/10.1021/es5009399>.
- Ma, Z., Hu, X., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L., Liu, Y., 2016. Satellite-based spatiotemporal trends in PM<sub>2.5</sub> concentrations: China, 2004–2013. *Environ. Health Perspect.* 124, 184–192. <https://doi.org/10.1289/ehp.1409481>.
- MEPC, 2015. Air Quality Daily Report: Ministry of Environmental Protection of the People's Republic of China. . <http://datacenter.mep.gov.cn/>.
- Price, D., Kacarab, M., Cocker, D., Purvis-Roberts, K., Silva, P., 2016. Effects of temperature on the formation of secondary organic aerosol from amine precursors. *Aerosol Sci. Technol.* 50, 1216–1226.
- Ramanathan, V., Crutzen, P.J., Kiehl, J.T., Rosenfeld, D., 2001. Aerosols, climate, and the hydrological cycle. *Science* 294 (5549), 2119–2124. <https://doi.org/10.1126/science.1064034>.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204.
- Ristovski, K., Vučetić, S., Obradović, Z., 2012. Uncertainty analysis of neural-network-based aerosol retrieval. *IEEE Trans. Geosci. Remote Sens.* 50, 409–414. <https://doi.org/10.1109/TGRS.2011.2166120>.
- Rosenfeld, D., Lohmann, U., Raga, G.B., O'Dowd, C.D., Kulmala, M., Fuzzi, S., Reissell, A., Andreae, M.O., 2008. Flood or drought: how do aerosols affect precipitation? *Science* 321, 1309–1313. <https://doi.org/10.1126/science.1160606>.
- Shang, H., Chen, L., Letu, H., Zhao, M., Li, S., Bao, S., 2017. Development of a daytime cloud and haze detection algorithm for Himawari-8 satellite measurements over central and eastern China. *J. Geophys. Res.: Atmosphere* 122, 3528–3543. <https://doi.org/10.1002/2016JD025659>.
- Song, W., Jia, H., Huang, J., Zhang, Y., 2014. A satellite-based geographically weighted regression model for regional PM<sub>2.5</sub> estimation over the Pearl River Delta region in China. *Remote Sens. Environ.* 154, 1–7. <https://doi.org/10.1016/j.rse.2014.08.008>.
- Sorek-Hamer, M., Strawa, A.W., Chatfield, R.B., Esswein, R., Cohen, A., Broday, D.M., 2013. Improved retrieval of PM<sub>2.5</sub> from satellite data products using non-linear methods. *Environ. Pollut.* 182, 417–423. <http://doi.org/10.1016/j.envpol.2013.08.002>.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC Bioinf.* 9 (1), 307.
- Su, T., Li, J., Li, C., Lau, A.K.-H., Yang, D., Shen, C., 2017. An intercomparison of AOD-converted PM<sub>2.5</sub> concentrations using different approaches for estimating aerosol vertical distribution. *Atmos. Environ.* 166, 531–542. <http://doi.org/10.1016/j.atmosenv.2017.07.054>.
- 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, 15921–15935. <https://doi.org/10.5194/acp-18-15921-2018>.
- Tan, S.-C., Zhang, X., Wang, H., Chen, B., Shi, G.-Y., Shi, C., 2018. Comparisons of cloud detection among four satellite sensors on severe haze days in eastern China. *Atmos. Ocean Sci. Lett.* 11, 86–93. <https://doi.org/10.1080/16742834.2017.1381547>.
- Van Donkelaar, A., Martin, R.V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., Villeneuve, P.J., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ. Health Perspect.* 118, 847–855. <https://doi.org/10.1289/ehp.0901623>.
- Van Donkelaar, A., Martin, R.V., Spurr, R.J.D., Drury, E., Remer, L.A., Levy, R.C., Wang, J., 2013. Optimal estimation for global ground-level fine particulate matter concentrations. *J. Geophys. Res.: Atmosphere* 118, 5621–5636. <http://doi.org/10.1002/jgrd.50479>.
- Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., Liu, Y., 2017. Full-coverage high-resolution daily PM<sub>2.5</sub> estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ.* 199, 437–446. <https://doi.org/10.1016/j.rse.2017.07.023>.
- Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B., Bai, Y., 2015. Daily estimation of ground-level PM<sub>2.5</sub> concentrations over Beijing using 3 km resolution MODIS AOD. *Environ. Sci. Technol.* 49 (20), 12280–12288. <http://doi.org/10.1021/acs.est.5b01413>.
- You, W., Zang, Z., Zhang, L., Li, Y., Wang, W., 2016a. Estimating national-scale ground-level PM<sub>2.5</sub> concentration in China using geographically weighted regression based on MODIS and MISR AOD. *Environ. Sci. Pollut. Control Ser.* 23, 8327–8338. <https://doi.org/10.1007/s11356-015-6027-9>.
- You, W., Zang, Z., Zhang, L., Li, Y., Pan, X., Wang, W., 2016b. National-scale estimates of ground-level PM<sub>2.5</sub> concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD. *Rem. Sens.* 8, 184. <https://doi.org/10.3390/rs8030184>.
- Yu, W., Liu, Y., Ma, Z., Bi, J., 2017. Improving satellite-based PM<sub>2.5</sub> estimates in China using Gaussian processes modeling in a Bayesian hierarchical setting. *Sci. Rep.* 7, 7048. <https://doi.org/10.1038/s41598-017-07478-0>.
- Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M.L., Shen, X., Zhu, L., Zhang, M., 2017. Spatiotemporal prediction of continuous daily PM<sub>2.5</sub> concentrations across China using a spatially explicit machine learning algorithm. *Atmos. Environ.* 155, 129–139. <https://doi.org/10.1016/j.atmosenv.2017.02.023>.
- Zhang, H., Hoff, R.M., Engel-Cox, J.A., 2009. The relation between Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth and PM<sub>2.5</sub> over the United States: geographical comparison by U.S. Environmental Protection Agency regions. *J. Air Waste Manag. Assoc.* 59 (11), 1358–1369. <https://doi.org/10.3155/1047-3289.59.11.1358>.
- Zhang, Q., Geng, G.N., Wang, S.W., Richter, A., He, K., 2012. Satellite remote sensing of changes in NO<sub>x</sub> emissions over China during 1996–2010. *Chin. Sci. Bull.* <https://doi.org/10.1007/s11434-01205015-4>.
- 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. <http://doi.org/10.1016/j.rse.2015.02.005>.
- Zheng, Y., Zhang, Q., Liu, Y., Geng, G., He, K., 2016. Estimating ground-level PM<sub>2.5</sub> concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements. *Atmos. Environ.* 124, 232–242. <https://doi.org/10.1016/j.atmosenv.2015.06.046>.