Ultrahigh-Resolution (250 m) Regional Surface $PM_{2.5}$ Concentrations Derived First From MODIS Measurements

Jianjun Liu, Fuzhong Weng, *Member, IEEE*, and Zhanqing Li

***Abstract*— Aerosol optical depth from different satellite sensors are widely used to estimate surface $PM_{2.5}$ concentrations. However, these products generally have coarse resolutions, limiting the ability to evaluate $PM_{2.5}$ concentrations in urban regions where the human activities are relatively high. This study first develops an ensemble machine learning approach to produce $PM_{2.5}$ concentrations with an extremely high spatial resolution of 250 m, based on Moderate Resolution Imaging Spectroradiometer (MODIS) measurements of top-of-atmosphere reflectance and related meteorological variables. The Yangtze River Delta region, with one of the highest levels of $PM_{2.5}$ pollution in China, is the study region chosen. The model shows a very high and stable performance with a coefficient of determination (R^2) of 0.90, a root-mean-square error (RMSE) of 12.0 $\mu g/m^3$, a mean prediction error (MPE) of 7.8 $\mu g/m^3$, and a mean relative prediction error (RPE) 16.9% for sample-based cross validation. The model can accurately capture the distribution patterns and magnitudes of $PM_{2.5}$ concentrations over the study region for seasonal mean, daily variations, and different levels of air pollution. The very high resolution of the model has the advantage of capturing the uneven spatial distribution of $PM_{2.5}$ concentrations at small spatial scales and identifying small areas with very high $PM_{2.5}$ concentrations, offering a possible approach for locating the sources of $PM_{2.5}$ emissions. In general, the model developed here estimates very well $PM_{2.5}$ concentrations at a very high spatial resolution, providing detailed information, useful for air-pollution-related studies, as well as pollution monitoring and evaluation by governments, especially in urban and urban-center areas.**

***Index Terms*— 250-m ultrahigh-resolution, machine learning (ML), Moderate Resolution Imaging Spectroradiometer (MODIS), regional $PM_{2.5}$.**

I. Introduction

PARTICULATE matter with an aerodynamic diameter less than 2.5 μm suspended in the air ($PM_{2.5}$) has become one of the primary air pollutants, especially in rapidly developing megalopolises in developing countries (e.g., China and India). $PM_{2.5}$ is highly associated with human health due to its harmful impact on the cardiovascular system [1], [2] and the respiratory system [3], [4]. It can lead to high morbidity, premature deaths, and ~3.3 million deaths per year around the world [5]. The spatiotemporal distribution of surface $PM_{2.5}$ concentrations is critical for understanding the sources, transportation, and diffusional behavior of regional air pollution episodes and assessing the environmental and health effects as well as pollution-controlling measures. Air quality monitoring sites can accurately measure surface $PM_{2.5}$ concentrations, but their spatial resolution is limited, making it difficult to capture in detail the distribution of $PM_{2.5}$ concentrations, which have large spatial variations.

Contrastingly, satellite remote sensing provides spatially continuous observations and has been widely adopted to estimate surface $PM_{2.5}$ concentrations. The aerosol optical depth (AOD), i.e., the total columnar light attenuation induced by aerosols, is highly correlated with surface $PM_{2.5}$ concentrations. A variety of satellite AOD products have thus been used to derive surface $PM_{2.5}$ concentrations, including retrievals from instruments such as the Moderate Resolution Imaging Spectroradiometer (MODIS) [6], the Multi-angle Imaging SpectroRadiometer (MISR) [7], the Visible Infrared Imaging Radiometer Suite (VIIRS) [8], the Advanced Himawari-8 Imager (AHI) [9], and the Geostationary Ocean Color Imager (GOCI) [10]. The AOD products generated by these instruments have the spatial resolutions of 3 and 10 km for MODIS, 4.4 km for MISR, 6 km for VIIRS (750 m in the VIIRS intermediate product), 5 km for AHI, and 6 km for GOCI. These spatial resolutions are too coarse for use in capturing spatial details about air pollution and in accurately estimating air pollution concentrations at fine scales, such as urban or smaller than urban scales. The high spatial resolution (1 km) MODIS AOD is generated by the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm and has been used to generate high-resolution $PM_{2.5}$ concentration maps [11]–[13]. Zhang *et al.* [14] estimated ultrahigh-resolution $PM_{2.5}$ concentrations in an urban region in China from 160-m Gaofen-1 (GF) AOD retrievals. However, the GF AOD has a coarse temporal resolution of four days. Zhang *et al.* [15] derived $PM_{2.5}$ concentrations from 30-m Landsat 8 Operational Land imager data, but its application is limited due to its coarse temporal resolution of 16 days.

Manuscript received August 31, 2020; revised January 7, 2021 and February 21, 2021; accepted February 23, 2021. Date of publication March 23, 2021; date of current version December 13, 2021. This work was supported by the National Key Research and Development Program of China under Grant 2018YFC1506501 and 2017YFC1501702. *(Corresponding authors: Jianjun Liu; Fuzhong Weng.)*

Jianjun Liu is with the Laboratory of Environmental Model and Data Optima (EMDO), Laurel, MD 20707 USA (e-mail: jianjun.liu@emdous.com).

Fuzhong Weng is with the State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China (e-mail: fweng58@gmail.com).

Zhanqing Li is with the Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740 USA, and also with the Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742 USA (e-mail: zhanqing@umd.edu).

Digital Object Identifier 10.1109/TGRS.2021.3064191

Multiple algorithms have been developed to derive surface $PM_{2.5}$ concentrations from satellite products using ancillary data (e.g., meteorological variables and land use information), including the combination of chemical transform model simulations and satellite AODs [16], physical models [17], empirical statistical models [18]–[20], and machine learning (ML) models [12], [21]–[23]. Of these algorithms, ML, a subset of artificial intelligence, offers the ability to automatically learn and deal with nonlinear complex problems, showing superior performance. Recently, multiple ML algorithms have been used to estimate surface $PM_{2.5}$ concentrations, such as the neural network [15], [24], a back-propagation artificial neural network (BP ANN) [25], deep learning [23], [26], Extreme Gradient Boosting (XGBoost) [27], [28], the Bayesian ensemble model [29], random forest (RF) [21], [30], and space–time extremely randomized trees (STETs) [31]. The RF is an ensemble ML algorithm, providing a nonparametric, nonlinear, and multivariate regression analysis with very high performance when estimating $PM_{2.5}$ concentrations from satellite remote sensing [9], [12], [21], [30].

Previous studies have generally used satellite products with a high frequency (e.g., hourly and twice daily) and coarse spatial resolution (1 km, 3 km, 10 km, and larger) or a high spatial resolution (e.g., 30 m) and coarse frequency (e.g., 16 days). The primary objective of this study is to develop a model for deriving surface $PM_{2.5}$ concentrations with the ultrahigh spatial resolution of 250 m from MODIS measurements and meteorological variables based on the RF algorithm. Fine-resolution and high-frequency $PM_{2.5}$ concentrations would benefit such small-scale environmental and epidemiological studies as in urban areas. The model is validated using the cross-validation (CV) technique with several metrics. $PM_{2.5}$ concentrations are estimated by the proposed model for several clear and high pollution cases over the Yangtze River Delta (YRD) region and metropolises in the YRD (e.g., Nanjing and Shanghai). The advantages of the high-spatial-resolution model are discussed through comparisons with previous studies and the model developed using MODIS 3-km AODs. Section II describes the data and methods, and Section III presents the results. Section IV discusses the advantages of the high-spatial-resolution model and compares it with other models reported in previous studies. Finally, conclusions are given in Section V.

II. Study Region and Data

A. Study Region

The YRD region was chosen as the testbed, covering several cities south of Jiangsu Province, east of Anhui Province, north of Zhejiang Province, and Shanghai city. The YRD is one of the most developed regions in China with intensive human activities and high amounts of complex aerosols in the air. It is also one of the most densely populated regions in the world.

B. Surface $PM_{2.5}$ Concentrations

The China Environmental Monitoring Center (http://www.cnemc.cn/) provided hourly surface $PM_{2.5}$ concentrations from January 1, 2018 to December 31, 2018. A tapered element oscillating microbalance measures $PM_{2.5}$ concentrations with an accuracy of ± 1.5 $\mu g/m^3$ for an hourly average. $PM_{2.5}$ sites in the latitude range of 27.5° N to 35° N and the longitude range of 115.5° E to 122.5° E to eliminate the boundary effect were included to develop the model. Fig. 1 shows the distribution of the 274 sites selected for the study. The colored dots indicate the annual mean surface measured $PM_{2.5}$ concentrations in 2018. Annual mean $PM_{2.5}$ concentrations ranged from ~20 to ~70 $\mu g/m^3$, with larger values occurring over the central and western parts of the YRD region and smaller values occurring over the southern and eastern coastal parts of the YRD region.

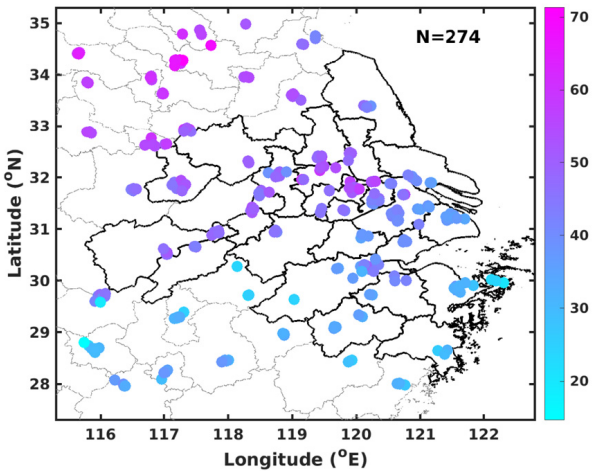

Fig. 1. Distribution of the 274 $PM_{2.5}$ observation sites used in the study. The colored dots indicate the annual mean surface measured $PM_{2.5}$ concentrations (unit: $\mu g/m^3$) in 2018. Black boundaries show the regions comprising the YRD economic zone.

C. Satellite Measurements and Retrievals

The MODIS instruments on board the Terra and Aqua satellites provide measurements of the atmosphere, land, and ocean in 36 spectral bands in the visible, near-infrared, and infrared from 0.4 to 14.4 μm. Of the 36 spectral bands, two visible bands have nominal resolutions of 250 m at nadir, five bands have nominal resolutions of 500 m at nadir, and the rest of the bands have nominal resolutions of 1 km at nadir. The instrument collects the data once or twice a day at a given location and views the entire globe every one to two days. MODIS Terra and Aqua L1B products (MOD02QKM and MYD02QKM) were used and downloaded from the Atmosphere Archive and Distribution System website (https://ladsweb.modaps.eosdis.nasa.gov). The products include top-of-atmosphere (TOA) reflectances at 0.65 and 0.86 μm with a 250-m spatial resolution, primarily used to derive land, cloud, and aerosol properties. MODIS Level 2 cloud mask products (MOD35_L2 and MYD35_L2) with a 1-km spatial resolution were used to remove cloud contamination. Here, removed are pixels in the MOD02QKM and MYD02QKM products located within cloud-masked pixels in the MOD35_L2 and MYD35_L2 products. The MODIS cloud mask products have four confidence levels, i.e., "cloudy,"

"uncertain," probably clear," and "confident clear." Only data with the level of "confident clear" are used. L2 AOD products with a 3-km spatial resolution (MOD04_3K and MYD04_3K) and the newly released Terra and Aqua MODIS C6 MAIAC 1-km AOD products were also used to estimate surface PM2.5 concentrations for comparison purposes. The MODIS AOD retrievals have four quality flags, i.e., "bad," "marginal," "good," and "very good." Only data flagged "very good" were used. The quality assurance (QA) flags were provided in MAIAC to indicate the retrieval quality. Only data flagged "best quality" were used. MODIS L3 normalized difference vegetation index (NDVI) products with a temporal resolution of 16 days and a 250-m spatial resolution (MOD13Q1 and MYD13Q1) were downloaded from the website https://search.earthdata.nasa.gov/search, providing surface cover and reflectance information. The NDVI values flagged "good quality" were used.

D. Meteorological Variables

Meteorological variables were obtained from the ERA5 reanalysis [32] and include surface atmospheric pressure (Pa), 2-m air temperature (K), total columnar water ($kg \cdot m^{-2}$), relative humidity (%), the 10-m u-component of wind (ms^{-1}), the 10-m v-component of wind (ms^{-1}), and the planetary boundary layer height (PBLH, km). These meteorological variables have known influences on the relationship between AOD and surface $PM_{2.5}$ concentrations and have been widely used in other studies [12], [22]. The ERA5 is the latest climate reanalysis produced by the European Centre for Medium-Range Weather Forecasts, providing hourly simulations of many atmospheric, terrestrial, and oceanic meteorological variables with a $0.25^{\circ} \times 0.25^{\circ}$ resolution and at 37 pressure levels from the surface to 80 km.

E. Elevation Data

A digital elevation model (DEM), derived from the Shuttle Radar Topography Mission (SRTM) with a 90-m spatial resolution, was used.

III. Methodology

A. Data Integration

Due to the different spatial resolutions of the ERA5 reanalysis, NDVI, DEM, and MODIS measurements, all meteorological parameters, NDVI and DEM, were interpolated onto the grids of the MODIS 3-km AOD, MAIAC 1-km AOD, and 250-m L1B products using a kriging method. The meteorological variables and measured $PM_{2.5}$ concentrations have the same time stamps of hour as the MODIS Terra and Aqua overpass over the surface $PM_{2.5}$ sites were used. MODIS-measured 250-m TOA reflectances and four observation angles [solar zenith angle (SOZ), solar azimuth angle (SOA), satellite zenith angle (SAZ), and satellite azimuth angle (SAA)] of pixels without clouds, NDVI, DEM, and meteorological variables were averaged over a 1-km buffer zone centered on each $PM_{2.5}$ site.

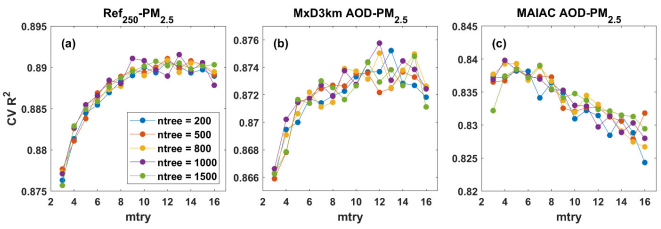

Fig. 2. Tuning parameters for (a) Ref_{250}-$PM_{2.5}$, (b) MxD3km AOD-$PM_{2.5}$, and (c) MAIAC AOD-$PM_{2.5}$ model.

For comparison purposes, MODIS-retrieved 3-km AODs (MxD3km) and MAIAC 1-km AODs and associated observation angles, NDVI, DEM, and meteorological variables were averaged over a 5- and 1-km buffer zone centered on each $PM_{2.5}$ site, generating the data set for developing the models to estimate $PM_{2.5}$ concentrations from AODs, referred to as the MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ model, respectively.

B. Model Development and Validation

A model to estimate surface $PM_{2.5}$ concentrations based on an RF algorithm was developed, which has been used in several related studies and has shown high performance in comparison with surface $PM_{2.5}$ observations [21], [30]. Mathematical details, the structure of the RF algorithm, and how it works have been discussed in previous papers [21], [30], so a brief review of the RF algorithm is given here. The RF is a type of supervised ensemble ML technique using multiple decision trees and the bootstrap aggregation technique, providing a nonparametric, multivariable, and nonlinear regression. The "k" features are first randomly selected from the total features and used to calculate the root node via the best split approach. Then, the tree is constructed with a root node. Multiple randomly constructed trees generated from the above process are used to build multiple decision trees. Finally, each prediction from the multiple trees is merged to obtain the final prediction [33].

Here, surface-observed $PM_{2.5}$ concentrations, TOA reflectances at 0.65 and 0.86 μm, SOA, SOZ, SAA, and SAZ, the locations of the $PM_{2.5}$ sites (longitude and latitude), observation times (month, day, and hour), NDVI, and all meteorological variables mentioned in Section II are used as input variables to develop the RF model for estimating surface $PM_{2.5}$ concentrations with a 250-m spatial resolution (referred to as the Ref_{250}-$PM_{2.5}$ model). The locations of the sites and observation times are included so that spatiotemporal variations in $PM_{2.5}$ concentration are considered. A grid search on hyperparameters with tenfold sample-based CV was used to find the best model performance based on R^2 metrics for different settings of ntree (200, 500, 800, 1000, and 1500) and mtry (3–16 with step 1). The optimum values of ntree and mtree are set to 1000 and 13 for Ref_{250}-$PM_{2.5}$ model, 1000 and 12 for MxD3km AOD-$PM_{2.5}$ model, and 1000 and 4 for MAIAC AOD-$PM_{2.5}$ model (as shown in Fig. 2).

The tenfold CV technique is commonly used to check the model robustness and the overfitting problem. In this study,

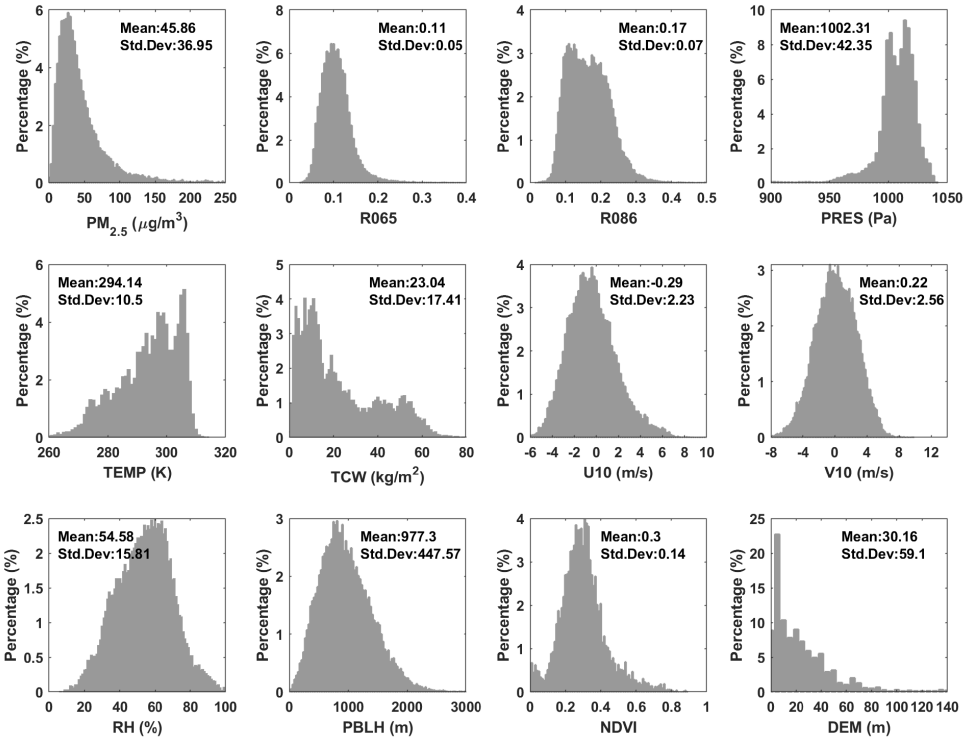

Fig. 3. Probability distributions and means and standard deviations of the variables of the training data set. R065: reflectance at 0.65 μm. R086: reflectance at 0.86 μm. PRES: surface pressure. TEMP: surface temperature. TCW: total column water. U10: 10-m u-component of wind. V10: 10-m v-component of wind. RH: relative humidity. PBLH: planetary boundary layer height. NDVI: normalized difference vegetation index. DEM: digital elevation model.

sample- and site-based CVs are both used. For the sample-based CV, training data were randomly split into ten equal subsets based on all data samples, and for the site-based CV, training data were randomly split into ten equal subsets based on the surface observation stations. Nine of the subsets were used for the model fitting, and the remaining one was used for model validation. This process was repeated ten times until all the subsets were tested. The sample-based CV is a common approach to evaluate the overall accuracy of the model, and the site-based CV is used to evaluate the spatial performance of the model. The coefficient of determination (R^2), the root-mean-square error (RMSE, μg/m3), the mean prediction error (MPE, μg/m^3), and the mean relative prediction error (RPE, %) between CV-model-estimated and surface-measured $PM_{2.5}$ concentrations were used to quantitatively evaluate the model performance.

IV. Results

A. Descriptive Statistics of the Training Data

Fig. 3 shows the probability distributions of the variables in the training data set. The mean and standard deviation of the $PM_{2.5}$ concentration are 46 and 37 μg/m^3, respectively, with ~70% of the values smaller than 50 μg/m^3 and peak values ranging from 15 to 35 μg/m^3. The reflectance at 0.65 μm ranges from 0.03 to 0.3, with a peak value of ~0.1 and a mean and standard deviation of 0.11 and 0.05, respectively. The mean and standard deviation of reflectance at 0.86 μm is 0.17 and 0.07, respectively, with overall values ranging from 0.05 to 0.3. The distributions of the model variables are broad, indicating that the model training data cover different levels of pollution and different meteorological conditions.

B. Evaluation of the Model Performance

Fig. 4 shows the density scatterplots of the sample- and site-based CV results of the Ref_{250}-$PM_{2.5}$ model. For the sample-based CV [see Fig. 4(a)], the overall R^2, RMSE, MPE, and RPE are 0.90, 12.0 μg/m^3, 7.8 μg/m^3, and 16.9%, respectively. By comparison, R^2 of the site-based CV [see Fig. 4(b)] slightly decreased by 0.04–0.86, and the RMSE, MPE, and RPE slightly increased by 1.7 μg/m^3, 1.1 μg/m^3, and 2.5% to 13.7 μg/m^3, 8.9 μg/m^3, and 19.5%, respectively. Overall, the Ref_{250}-$PM_{2.5}$ model performs well. The slopes of the sample- and site-based CVs are 0.86 and 0.83, respectively, indicating that the model tends to underestimate surface $PM_{2.5}$ concentrations, especially for extremely heavy pollution episodes. Such an underestimation is consistent with most previous studies reporting that $PM_{2.5}$ concentrations were underestimated by ~10%–30% based on different ML algorithms [12], [34].

Fig. 5 shows the spatial distributions of sample-based CV: R^2, RMSE, MPE, and RPE. The number of samples from each surface site varies from 42 to 360, with an average of 252. R^2 [see Fig. 5(a)] varies from 0.48 to 0.98, with a mean and standard deviation of 0.87 and 0.09, respectively, and is generally greater than 0.75 (~82% of the sites have R^2 greater than 0.8). The RMSE [see Fig. 5(b)] at each site ranges from 5.5 to 25.2 μg/m^3, with an average value and standard deviation of 11.4 and 3.1 μg/m^3, respectively. About 90% of

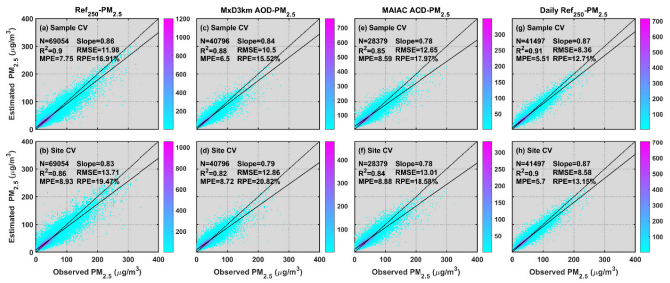

Fig. 4. Density scatterplots of (a), (c), (e), and (g) sample-based and (b), (d), (f), and (h) site-based CV results for (a) and (b) Ref_{250}-$PM_{2.5}$ model, (c) and (d) MxD3km AOD-$PM_{2.5}$ model, (e) and (f) MAIAC AOD-$PM_{2.5}$ model, and (g) and (h) daily Ref_{250}-$PM_{2.5}$ model. Dashed lines are the 1:1 lines. Solid lines are the best-fit lines from linear regression. N: number of samples. R^2: coefficient of determination. RMSE: root-mean-square error ($\mu g/m^3$). MPE: mean prediction error ($\mu g/m^3$). RPE: mean relative prediction error (%).

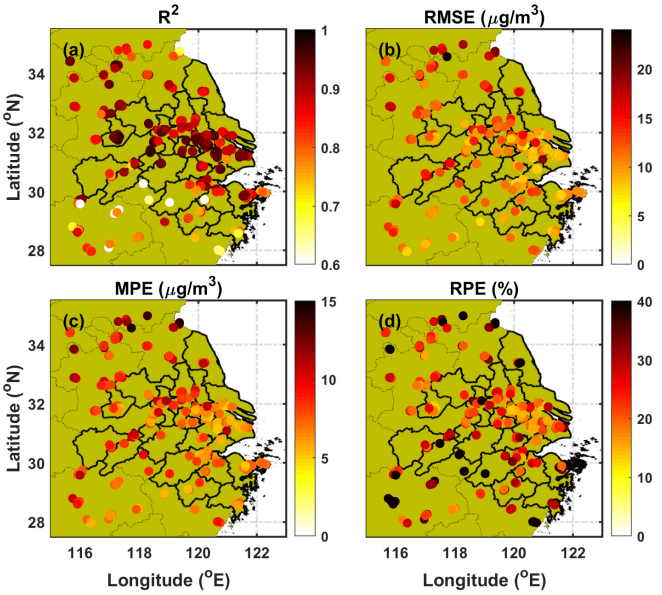

Fig. 5. Spatial distributions of sample-based CV. (a) R^2. (b) RMSE. (c) MPE. (d) RPE. R^2: coefficient of determination. RMSE: root-mean-square error ($\mu g/m^3$). MPE: mean prediction error ($\mu g/m^3$). RPE: mean relative prediction error (%).

the sites have RMSE values smaller than 15 $\mu g/m^3$. Fig. 5(c) shows that ~90% of the stations have MPE values smaller than 10 $\mu g/m^3$, with values ranging from 3.8 to 14.3 $\mu g/m^3$. The mean and standard deviation of MPE for all sites are 7.6 and 1.9 $\mu g/m^3$, respectively. At the majority of the sites, the RPE is smaller than 30%, with a mean and standard deviation of 25% and 16%, respectively [see Fig. 5(d)]. Overall, the model estimates $PM_{2.5}$ concentrations well, although a few of the sites have relatively low R^2 and high RMSE, MPE, and RPE.

C. Spatial Distribution of the Seasonal Mean Estimation of $PM_{2.5}$ Concentrations

Fig. 6 shows the spatial distribution of the seasonal mean $PM_{2.5}$ concentration from Ref_{250}-$PM_{2.5}$ model-estimated and surface-measured over the YDR region in 2018. It shows that the spatial pattern of the seasonal mean $PM_{2.5}$ concentration from model-estimated is highly consistent with surface measured. $PM_{2.5}$ concentration varies seasonally, with the

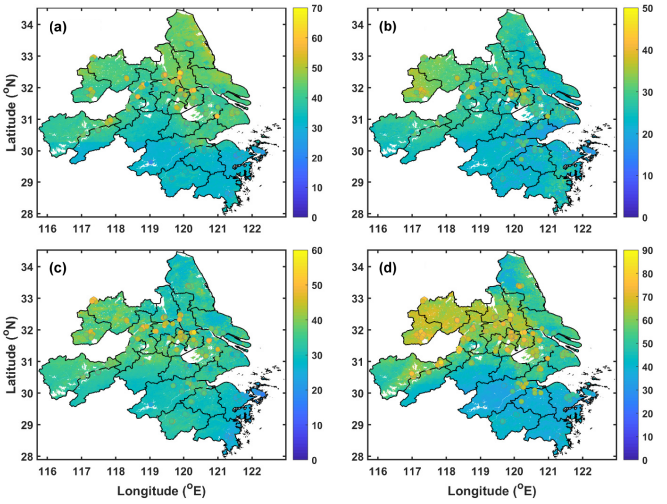

Fig. 6. Spatial distribution of the seasonal mean $PM_{2.5}$ concentration from Ref_{250}-$PM_{2.5}$ model estimated (background shading) and surface measured (color dots) for (a) spring (MAM: March–April–May), (b) summer (JJA: June–July–August), (c) autumn (SON: September–October–November), and (d) winter (DJF: December–January–February) in 2018 over YRD region.

highest concentration in winter, followed by spring and autumn, and the lowest concentration in summer. The mean and standard deviation of model-estimated (surface-measured) $PM_{2.5}$ concentration is 38.8 ± 7.3 (40.4 ± 10.2) $\mu g/m^3$ for spring, 26.1 ± 4.9 (24.2 ± 7.1) $\mu g/m^3$ for summer, 32.8 ± 7.8 (22.3 ± 11.2) $\mu g/m^3$ for autumn, and 49.8 ± 10.3 (51.6 ± 14.4) $\mu g/m^3$ for winter. In winter and spring, more than 97% and 70% of YRD are exposed to high seasonal mean $PM_{2.5}$ levels > 35 $\mu g/m^3$. In contrast, about 98% and 73% of YRD are exposed to seasonal mean $PM_{2.5}$ levels < 35 $\mu g/m^3$ in summer and autumn. Meanwhile, for all seasons, higher $PM_{2.5}$ concentrations are seen in the central and the northern region, and lower concentrations are seen in the southern region.

D. Daily Variations in Estimation of $PM_{2.5}$ Concentrations

Fig. 7 shows the daily variations of the Ref_{250}-$PM_{2.5}$ model-estimated ($PPM_{2.5}$) and surface observed ($OPM_{2.5}$) $PM_{2.5}$ concentration and the differences between $PPM_{2.5}$ and $OPM_{2.5}$ in 2018 over the YRD region. It indicates that the daily variation of the model estimated $PM_{2.5}$ concentration is highly consistent with surface observed. R^2 and RMSE between model-estimated and surface-observed daily $PM_{2.5}$ concentration are 0.83 and 14.3 $\mu g/m^3$, respectively. The model-estimated (surface-observed) daily $PM_{2.5}$ concentration varies from 8.2 (4.0) to 225.5 (232.0) $\mu g/m^3$ with the mean and standard deviation of 45.3 ± 30.8 (44.3 ± 35.1) $\mu g/m^3$, respectively. The higher (lower) $PM_{2.5}$ concentrations are usually seen on days in winter (summer). The difference between model-estimated and surface-measured $PM_{2.5}$ concentration changes from −45.2 to 71.5 $\mu g/m^3$ with ~87% and ~67% of the values ranging from −20 to 20 and −10 to 10 $\mu g/m^3$, respectively. This indicates that the Ref_{250}-$PM_{2.5}$ model can accurately capture the temporal variations in the surface $PM_{2.5}$ concentration.

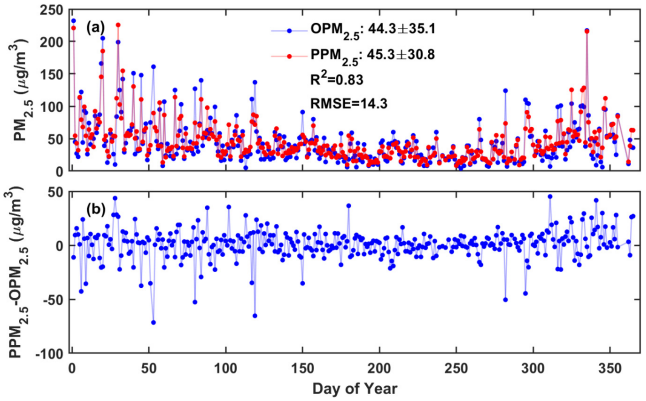

Fig. 7. Daily variations of (a) model estimated ($PPM_{2.5}$) and surface observed ($OPM_{2.5}$) $PM_{2.5}$ concentration and (b) differences between $PPM_{2.5}$ and $OPM_{2.5}$ in 2018 over the YRD region.

E. Estimation of $PM_{2.5}$ Concentrations for Days With Different Pollution Levels

To verify the performance and the power of the Ref_{250}-$PM_{2.5}$ model, $PM_{2.5}$ concentrations using the model for four MODIS overpass (mainly cloudless) cases in each season for different pollution levels were estimated [see Fig. 8(a)–(d); background shading]. The corresponding surface-observed $PM_{2.5}$ concentrations for the same four cases are also shown [see Fig. 8(a)–(d); color dots]. Relative to the whole region, higher model-estimated $PM_{2.5}$ concentrations were seen in the center of the YRD region for case (a), the north and northwest for case (b), the northwest for case (c), and the center and north for case (d), and relatively low model-estimated $PM_{2.5}$ concentrations were generally found in the south and eastern coastal areas of the YRD region, consistent with the distribution patterns of surface-observed $PM_{2.5}$ concentrations for all cases. The means and standard deviations of model-estimated (surface-observed) $PM_{2.5}$ concentrations over the whole YRD region are 38.0 ± 7.3 (39.9 ± 14.5) $\mu g/m^3$ for case (a), 16.7 ± 4.8 (15.8 ± 7.4) $\mu g/m^3$ for case (b), 30.4 ± 6.8 (30.6 ± 11.9) $\mu g/m^3$ for case (c), and 72.7 ± 23.4 (80.0 ± 33.0) $\mu g/m^3$ for case (d). These results suggest that the model can accurately capture the spatial distributions of $PM_{2.5}$ concentrations and the mean status of $PM_{2.5}$ concentrations for different air pollution conditions. The near-complete spatial coverage and extremely high spatial resolution of the model reveal locations where the surface-observed $PM_{2.5}$ concentration is high.

V. Discussion

A. Comparison With the Performance of the AOD-$PM_{2.5}$ Models

Many pollution studies focused on China have estimated surface $PM_{2.5}$ concentrations from satellite AOD retrievals based on different models due to the heavy $PM_{2.5}$ pollution levels in that country. AOD, a measure of columnar aerosol loading, is widely used as the primary predictor in these models. Thus, the AOD-$PM_{2.5}$ models were also developed using the MODIS L2 3-km AOD (MxD3km AOD-$PM_{2.5}$) and MAIAC 1-km AOD (MAIAC AOD-$PM_{2.5}$) products as the primary predictor based on the same RF algorithm as the

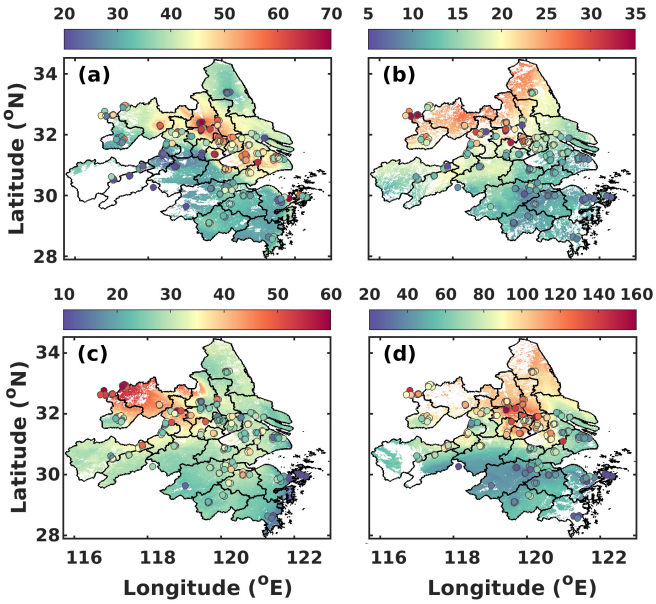

Fig. 8. Ref_{250}-$PM_{2.5}$-model-estimated (background shading) and surface-observed (color dots) $PM_{2.5}$ concentrations (unit: $\mu g/m^3$) for four MODIS overpass (mainly cloudless) cases in each season. Low (b), moderate (a) and (c), and high (d) pollution level cases are shown. The dates of the four cases are (a) May 4, 2018, 03:00 UTC, (b) July 14, 2018, 03:05 UTC, (c) October 27, 2018, 03:00 UTC, and (d) January 15, 2018, 05:05 UTC.

Ref_{250}-$PM_{2.5}$ model. The predictors of the MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ model are the same as the predictors of the Ref_{250}-$PM_{2.5}$ model except that AOD was used to replace TOA reflectances at 0.65 and 0.86 μm.

Fig. 4 shows the density scatterplots of the sample- and site-based CV results of the MxD3km and MAIAC AOD-$PM_{2.5}$ models. The results indicate that the sample- and site-based CV-R^2 of the MxD3km and MAIAC AOD-$PM_{2.5}$ models are both lower than the sample- and site-based CV-R^2 of the Ref_{250}-$PM_{2.5}$ model [as shown in Fig. 4(a) and (b), respectively]. In addition, the number of samples in the training data set of the Ref_{250}-$PM_{2.5}$ model is more than that of the MxD3km and MAIAC AOD-$PM_{2.5}$ models, indicating the higher performance on spatial–temporal coverage of Ref_{250}-$PM_{2.5}$ model than that of MxD3km and MAIAC AOD-$PM_{2.5}$ models.

To better comparisons in the $PM_{2.5}$ estimations from the TOA reflectance and the AODs, the Ref_{250}-$PM_{2.5}$ model was also developed with uniform conditions of MAIAC AOD-$PM_{2.5}$ model, except the reflectance and AOD. The sample-based CV result indicates that the performance of the Ref_{250}-$PM_{2.5}$ model is comparable to that of the MAIAC AOD-$PM_{2.5}$ model (as shown in Fig. 9). Overall, results here suggest that the performance of the Ref_{250}-$PM_{2.5}$ model shows a comparable performance to the AOD-$PM_{2.5}$ method. The Ref_{250}-$PM_{2.5}$ model has a relatively stronger predictive power with greater spatiotemporal coverage than the AOD-$PM_{2.5}$ model.

B. Variable Importance Assessment

Fig. 10(a)–(c) shows the contribution of each predictor to the model performance for Ref_{250}-$PM_{2.5}$, MxD3km

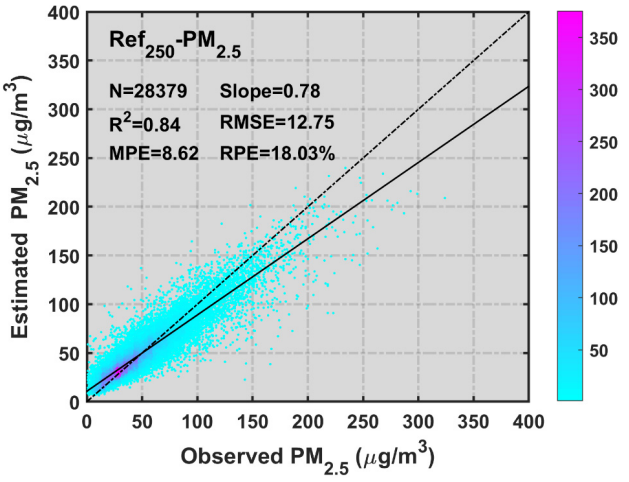

Fig. 9. Sample-based CV result of Ref_{250}-$PM_{2.5}$ model developed with both TOA reflectance and MAIAC AOD available.

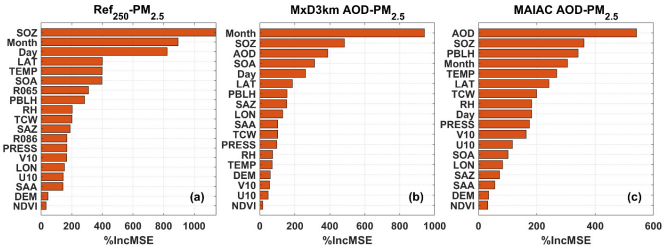

Fig. 10. Importance assessment for predictors of (a) Ref_{250}-$PM_{2.5}$, (b) MxD3km AOD-$PM_{2.5}$, and (c) MAIAC AOD-$PM_{2.5}$ model.

AOD-$PM_{2.5}$, and MAIAC AOD-$PM_{2.5}$ models, respectively. The %IncMSE indicates the increase in mean square error of the prediction if that predictor is not involved in training data. The higher the %IncMSE for a predictor, the more important is that predictor. The results showed that the spatial–temporal information (e.g., SOZ and month) is important to all models. For the Ref_{250}-$PM_{2.5}$ model, the reflectance @ 0.65 μm shows large influences on model performance and is more important than the reflectance @ 0.86 μm, which is reasonable since the reflectance @ 0.65 μm is more sensitive to aerosol particles than that of reflectance @ 0.86 μm. However, the reflectance @ 0.86 μm is also important. For the MxD3km and MAIAC AOD-$PM_{2.5}$ model, AOD is among the top three most important predictor variables. Contrastingly, DEM and NDVI are relatively less important than other predictors for all models.

C. Performance of the Daily Ref_{250}-$PM_{2.5}$ Model

The Ref_{250}-$PM_{2.5}$ model was developed based on $PM_{2.5}$ concentrations measured the same to the hour of the MODIS Terra and Aqua overpass times over the surface $PM_{2.5}$ sites. Many previous studies developed models to estimate the daily mean surface $PM_{2.5}$ concentration by matching measured daily mean surface $PM_{2.5}$ concentrations with the AODs retrieved when the MODIS Terra or Aqua platforms overpassed the surface stations. For comparisons with previous studies, a daily Ref_{250}-$PM_{2.5}$ model was also developed to estimate daily mean $PM_{2.5}$ concentrations based on the same algorithms using a combination of daily mean surface-observed $PM_{2.5}$ concentrations, daily mean meteorological variables, and MODIS Terra measurements of reflectance. Fig. 4 shows the density scatterplots of the sample- and site-based CVs of the daily Ref_{250}-$PM_{2.5}$ model. The model performs well in estimating daily mean $PM_{2.5}$ concentrations with R^2 values of 0.91 and 0.90 for sample- and site-based CVs, respectively, and RMSE (MPE and RPE) values of 8.4 μg/m^3 (5.5 μg/m^3 and 12.7%) for the sample-based CV and 8.6 μg/m^3 (5.7 μg/m^3 and 13.2%) for the site-based CV. These results suggest that the Ref_{250}-$PM_{2.5}$ model can be used to accurately estimate surface $PM_{2.5}$ concentrations at the MODIS Terra and Aqua overpass times and on a daily mean basis as well.

D. Comparison With Recent Studies

Table I summarizes the model performances of different algorithms reported in previous studies on $PM_{2.5}$ concentration estimations from satellite retrieved AOD over the YRD region. The spatial resolutions of the different models range from 1 to 10 km, with most of the studies having coarse spatial resolutions of greater than 3 km. The CV-R^2 of the different models varies from 0.67 to 0.88, with most R^2 values less than 0.85. The Ref_{250}-$PM_{2.5}$ model captures 90% of the variability in $PM_{2.5}$ concentrations in the sample-based CV, which is larger than the sample-based CV-R^2 of the other models. The RMSE, MPE, and RPE values of the Ref_{250}-$PM_{2.5}$ model are generally lower than those of the other models. Overall, the Ref_{250}-$PM_{2.5}$ model has a robust and superior performance in estimating $PM_{2.5}$ concentrations with an extremely high spatial resolution of 250 m.

E. Advantages of the Ref_{250}-$PM_{2.5}$ Model

The greatest advantage of the proposed model developed in this study is its high spatial resolution of 250 m. The annual mean $PM_{2.5}$ concentrations derived from the Ref_{250}-$PM_{2.5}$ model in Nanjing and its central district and Shanghai and its central district are, respectively, shown in Fig. 11(a-1)–(a-4). Also, the annual mean $PM_{2.5}$ estimation in these two cities and their central district based on the MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ models are shown in Fig. 11(b-1)–(b-4) and (c-1)–(c4), respectively. The annual mean $PM_{2.5}$ concentration derived from the three models possessed approximately similar spatial distributions. The MAIAC AOD-$PM_{2.5}$ model tends to overestimate the $PM_{2.5}$ concentration along the riverbank and over the coastal region. Compared with the $PM_{2.5}$ concentration estimated from the MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ models, the $PM_{2.5}$ concentration estimations from the Ref_{250}-$PM_{2.5}$ model have an obvious advantage in the spatial resolution, especially in the urban central district of Nanjing and Shanghai city with more population (as shown in Fig. 12). For instance, the Ref_{250}-$PM_{2.5}$ model estimations have a better illustration of the gradual variation in the spatial distribution of $PM_{2.5}$ concentration than the MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ model estimations, especially in the small areas. In addition, the Ref_{250}-$PM_{2.5}$ model could capture the highest $PM_{2.5}$ concentrations over the relatively

TABLE I

MODEL PERFORMANCES REPORTED IN SOME PREVIOUS STUDIES ON $PM_{2.5}$ ESTIMATIONS FROM SATELLITE RETRIEVALS OVER THE YRD REGION

Reference	Model	R^2	RMSE ($\mu g/m^3$)	MPE ($\mu g/m^3$)	RPE	Spatial Resolution	Satellite Retrieval
Ma et al.[18]	Nested LME	0.73	18.30	12.53	-	10 km	MODIS AOD
Ma et al.[18]	Nested LME	0.67	15.82	11.44	-	3 km	MODIS AOD
Zheng et al.[35]	LME	0.8	17.9	-	17.8%	10 km	MODIS AOD
Xiao et al.[36]	LME-GAM	0.81	25.0	-	36%	1 km	MAIAC AOD
Jiang et al.[37]	Seasonal GRW	0.75/0.74/ 0.88/0.78	16.1/8.3/ 12.3/21.2	-	28.4%, 25.3%, 21.0%, 19.1%	10 km	MODIS AOD
Tang et al.[38]	Two-stage RF	0.86/0.86	12.4/12.4	-	28.5%/ 28.4%	6/5 km	GOCI/AHI AOD
She et al.[39]	LME-GAM	0.73	23	-	-	6 km	GOCI AOD
This study	RF	0.90	12.0	7.8	17.1%	0.25 km	MODIS Reflectance

LME: linear mixed-effects model; GAM: generalized additive model; GRW: geographically weighted regression model; GTWR: geographically and temporally weighted regression model; RF: random forest model; AOD: aerosol optical depth; MODIS: Moderate Resolution Imaging Spectroradiometer; MAIAC: Multi-Angle Implementation of Atmospheric Correction; GOCI: Geostationary Ocean Color Imager; AHI: Advanced Himawari Imager

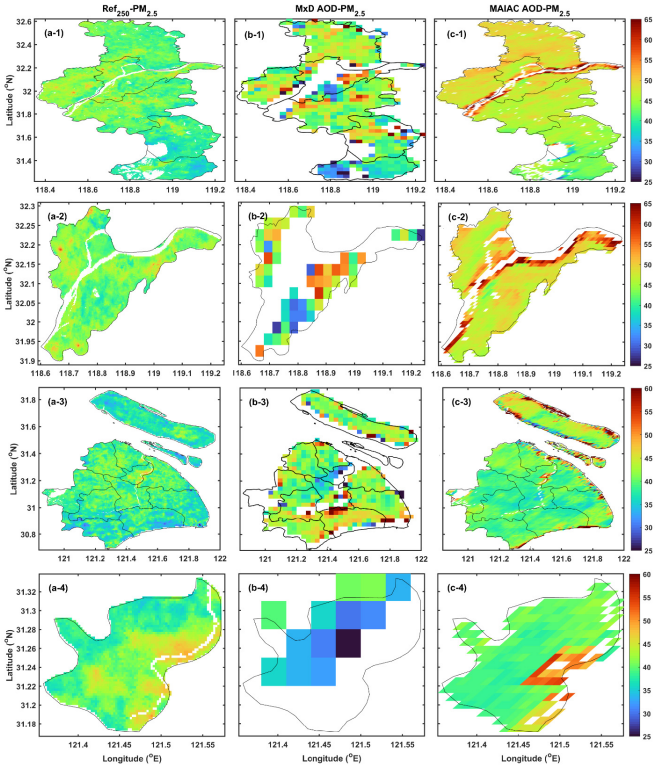

Fig. 11. $PM_{2.5}$ concentration estimated from (a-1)–(a-4) Ref_{250}-$PM_{2.5}$ model, (b-1)–(b-4) MxD3km AOD-$PM_{2.5}$ model, and (c-1)–(c-4) MAIAC AOD-$PM_{2.5}$ model in Nanjing (a-1), (b-1), and (c-1), urban central district of Nanjing (a-2), (b-2), and (c-2), Shanghai (a-3), (b-3), and (c-3), and urban central district of Shanghai (a-4), (b-4), and (d-4).

small areas, whereas MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ models could not detect these small areas. These small areas with high $PM_{2.5}$ concentration could be considered possible sources of $PM_{2.5}$ point emissions. Therefore, the Ref_{250}-$PM_{2.5}$ model with higher spatial resolution can provide more details in spatial variations of $PM_{2.5}$ than MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ models.

To further examine the performance of the Ref_{250}-$PM_{2.5}$ model at urban scales, Fig. 13 shows the Ref_{250}-$PM_{2.5}$-model-estimated $PM_{2.5}$ concentrations over Nanjing city, Shanghai

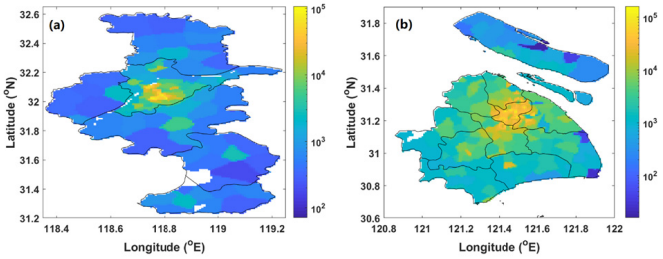

Fig. 12. Distribution of population in (a) Nanjing city and (b) Shanghai city in 2018. Data are obtained from the NASA Socioeconomic Data and Applications Center (Gridded Population of the World, v4; http://sedac.ciesin.columbia.edu/) with approximately 1-km resolution.

city, and their urban central districts for low and high air pollution cases. $PM_{2.5}$ concentrations were also estimated by the MxD3km AOD-$PM_{2.5}$ and the MAIAC AOD-$PM_{2.5}$ model over Nanjing and Shanghai for the same low and high air pollution cases. $PM_{2.5}$ concentrations measured in the central districts of the two cities are also shown. For the case with low $PM_{2.5}$ pollution in Nanjing city [Fig. 13(a-1)–(a-4)], the models estimated $PM_{2.5}$ concentration possessed almost the same spatial distribution, but the Ref_{250}-$PM_{2.5}$ estimations have a more complete spatial coverage than the MxD3km AOD-$PM_{2.5}$ model estimations. The $PM_{2.5}$ concentration generally ranges from ~30 to ~55 $\mu g/m^3$ with a relatively higher (lower) concentration in the north (south) of city. For the case with high $PM_{2.5}$ pollution in Nanjing city [Fig. 13(b-1)–(b-4)], the high (low) $PM_{2.5}$ concentration is distributed in the north (south) of city, generally higher (lower) than ~110 (80) $\mu g/m^3$. The spatial distribution of $PM_{2.5}$ concentration derived from the MxD3km AOD-$PM_{2.5}$ model is consistent with the Ref_{250}-$PM_{2.5}$ model. However, the MxD3km AOD-$PM_{2.5}$ model has no estimations in a large area of the city. Meanwhile, compared with the Ref_{250}-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ model estimations, the MxD3km AOD-$PM_{2.5}$ model obviously overestimated the $PM_{2.5}$ concentrations in the south of the city. The MAIAC AOD-$PM_{2.5}$ model estimated $PM_{2.5}$ concentration in the south of the city is consistent with the Ref_{250}-$PM_{2.5}$ model estimations. However, there are no $PM_{2.5}$ estimations from the MAIAC AOD-$PM_{2.5}$ model

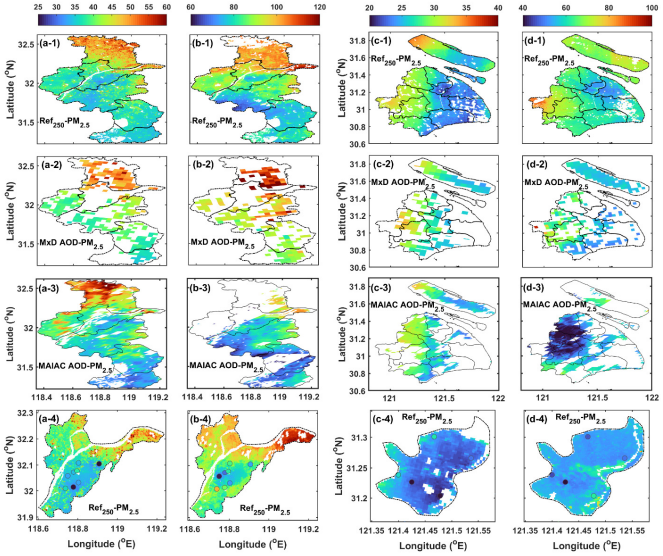

Fig. 13. Ref_{250}-$PM_{2.5}$-model-estimated $PM_{2.5}$ concentrations over (a-1) and (b-1) Nanjing City, (a-4 and b-4) urban central district of Nanjing City, (c-1) and (d-1) Shanghai city, and (c-4) and (d-4) urban central district of Shanghai city, as well as MxD3km and MAIAC AOD-$PM_{2.5}$-model-estimated $PM_{2.5}$ concentrations over (a-2), (b-2), (a-3), and (b-3) Nanjing City and (c-2), (d-2), (c-3), and (d-3) Shanghai city for low and high $PM_{2.5}$ concentration cases. The colored dots in (a-4), (b-4), (c-4), and (d-4) show surface-measured $PM_{2.5}$ concentrations. Units are $\mu g/m^3$. The dates of the four cases are (a) February 13, 2018, 03:00 UTC, (b) January 15, 2018, 05:05 UTC, (c) October 2, 2018, 03:05 UTC, and (d) January 15, 2018, 05:05 UTC.

in the north of city, where the $PM_{2.5}$ concentrations are high. For the case with low air pollution in Shanghai city [Fig. 13(c-1)–(c-4)], the $PM_{2.5}$ concentration varies from ~20 to ~35 $\mu g/m^3$ with relatively higher and lower values that are seen in the western and the eastern part of city, respectively. The spatial distribution of the Ref_{250}-$PM_{2.5}$, MxD3km AOD-$PM_{2.5}$, and MAIAC AOD-$PM_{2.5}$ model estimations are consistent with each other. However, the $PM_{2.5}$ estimations from the Ref_{250}-$PM_{2.5}$ model have a much more complete spatial coverage than those from the MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ models. For the case with high air pollution in Shanghai city [Fig. 13(d-1)–(d-4)], relatively higher and lower values are distributed in the west and the east of the city with values larger and smaller than ~90 and ~50 $\mu g/m^3$, respectively. The $PM_{2.5}$ estimations from the Ref_{250}-$PM_{2.5}$ model cover almost the entire city, but there are no $PM_{2.5}$ estimations from the MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ model in a large area of the city. In addition, the $PM_{2.5}$ estimations from the MAIAC AOD-$PM_{2.5}$ model are obviously lower than those from the Ref_{250}-$PM_{2.5}$ and MxD3km AOD-$PM_{2.5}$ models in the west of the city. Overall, the spatial distribution of the $PM_{2.5}$ concentration estimated from the Ref_{250}-$PM_{2.5}$, MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ models is generally consistent, but the Ref_{250}-$PM_{2.5}$ model estimations have a much more complete spatial coverage than the MxD3km AOD-$PM_{2.5}$ and MAIAC AOD-$PM_{2.5}$ model estimations, which is likely related to the AOD retrieval algorithm [40]. In addition, more pixels were masked as the cloud contamination in the MxD3km AOD retrieval algorithm than the MODIS cloud mask

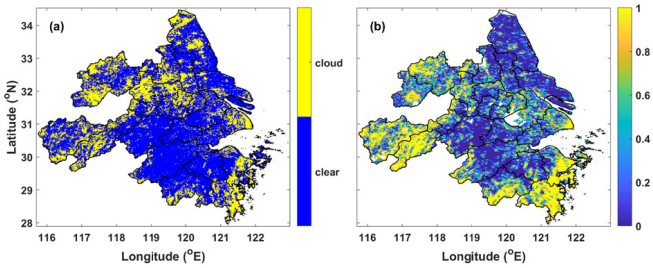

Fig. 14. (a) Cloud mask result in the MODIS cloud mask product and (b) cloud fraction result in MODIS 3-km AOD product over YRD region for one case as shown in Fig. 5 (July 14, 2018, 03:05 UTC).

product (as shown in Fig. 14) is also partially contributed to the large missing of the AOD retrievals. The Ref_{250}-$PM_{2.5}$ model estimates $PM_{2.5}$ concentrations by directly using TOA reflectances, avoiding AOD retrievals.

The results indicate that the Ref_{250}-$PM_{2.5}$ model successfully derives $PM_{2.5}$ concentrations at the city level and smaller scales within the city under low and high air pollution conditions. The Ref_{250}-$PM_{2.5}$ model also captures the uneven spatial distribution of $PM_{2.5}$ concentrations at small spatial scales and identifies locations with the highest $PM_{2.5}$ concentration, considered possible sources of $PM_{2.5}$ point emissions. The AOD-$PM_{2.5}$ model, however, cannot reveal such details, given the large number of missing retrievals and the coarse spatial resolution of the MODIS DT AOD product. The Ref_{250}-$PM_{2.5}$ model, with its spatial resolution of 250 m, can provide more details about the gradual variations in $PM_{2.5}$ concentration at the city level and at smaller scales within the city than models developed using the MxD3km and MAIAC AOD products.

The spatial distribution of annual mean $PM_{2.5}$ concentration derived from the Ref_{250}-$PM_{2.5}$ model in Nanjing city is given in Fig. 15(II) and the satellite image in Nanjing city from Google Earth image is given in Fig. 15(I) as a reference. The results indicate that the high $PM_{2.5}$ concentration derived from the Ref_{250}-$PM_{2.5}$ model is generally seen in the regions with intensive human activities and dense urban constructions, e.g., industrial zones and residential areas. In addition, several typical areas, including an industrial park located in the north of the city [see Fig. 15(a)], a combination of industrial zones and residential areas [see Fig. 15(b)], and the Nanjing Lukou International Airport [see Fig. 15(c)], with high $PM_{2.5}$ concentration are zoomed in upon in order to illustrate the performance of high spatial resolution of Ref_{250}-$PM_{2.5}$ model. The $PM_{2.5}$ concentration is typically higher in the industrial zone and the high-density residential area than that in the surrounding area of nonindustrial and nonresidential areas [as shown in Fig. 15(a) and (b)]. Fig. 15(c) shows that the $PM_{2.5}$ concentration is obviously higher in the entire airport than that in the surround area, with relatively higher values in the parking apron, indicating that our Ref_{250}-$PM_{2.5}$ model has the potential ability to locate $PM_{2.5}$ emission source inside a large $PM_{2.5}$ emissions unit to a certain extent. Overall, the high resolution of the model developed in this study is possibly used to locate main $PM_{2.5}$ emission sources, consisting of industrial zones, residential areas, and transportation hubs.

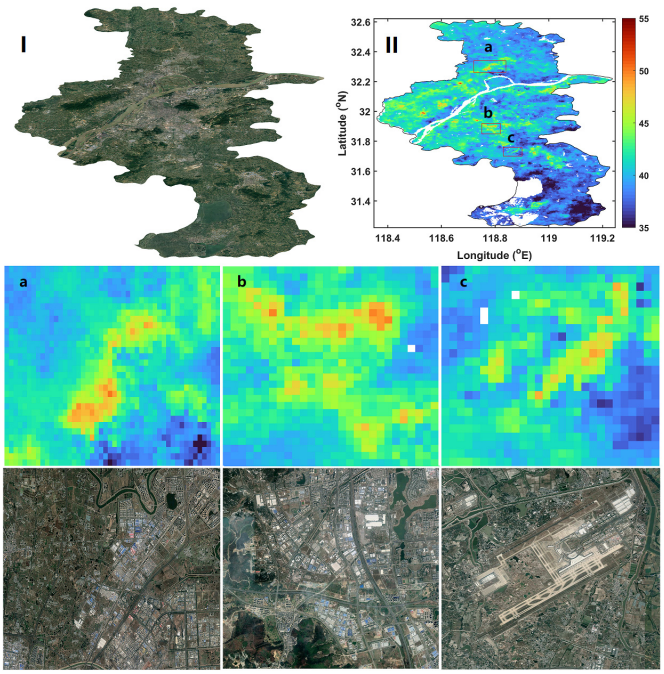

Fig. 15. Illustration of the locations with high annual mean $PM_{2.5}$ concentration derived from Ref_{250}-$PM_{2.5}$ model in Nanjing city, combined with Google Earth images as references.

F. Limitations and Future Work

In this section, we developed a model to estimate surface $PM_{2.5}$ concentrations with a high spatial resolution of 250 m by using the direct measurements of MODIS TOA reflectance in combination with meteorological variables, in lieu of the conventional approach of using the retrieved AOD. Compared with the latter, the proposed model has a much higher spatial resolution at the same temporal frequency. However, there are some limitations and potential room for model improvements.

Reanalysis meteorological variables were used in the model development, having relatively coarse spatial resolutions of $0.25° \times 0.25°$ and whose accuracy still needs further validation. For example, the ECMWF Reanalysis v5 (ERA5) tends to overestimate/underestimate the PBLH in spring/summer over China [41] and underestimate the 10-m wind speed in most regions of China [42]. The coarse spatial resolution of reanalysis data may lead a spatial mismatch between the surface $PM_{2.5}$ concentrations and the meteorological variables. The possible unrepresentativeness of the meteorological conditions over a site may result in errors in $PM_{2.5}$ estimations. More accurate surface measurements of meteorological variables over a site may improve the model performance.

The RF algorithm was used to develop the model in the current study because it has been shown to estimate $PM_{2.5}$ concentrations from satellite remote sensing very well and easy to fine-tune to derive optimal model parameters. Some of the previous studies have developed models based on different ML algorithms, which shows that the performance is comparable to or slightly better than the RF algorithm. Thus, developing high-spatial-resolution models with different ML algorithms and systematically comparing their performances should be done in the future. The proposed approach was initially tested in the YRD region in the current study, and thus, the model developed over other regions should be tested in future.

The red channel is sensitive to aerosol particles and reports relatively low surface reflectance on dark targets. Thus, the TOA reflectances at red and blue channels from MODIS have been widely used for aerosol retrievals [43]. The fact is that much larger sample size of the TOA reflectance than that of the AOD products is possibly contributed to the higher performance of the model with reflectance than that with AOD products. In addition, the ML approach can generally cope with complex statistics, multiple variables, different noise sources, and complicated relationship among variables, and thus, it offers a promising tool to build a new data-driven model for possibly capturing the complicated relationship between the initial predictors and the final predictions [44]. However, interpretability has been identified as a potential weakness of ML approaches. Meanwhile, ML algorithms have acceptable and relatively stable performances when estimating over the observation period of the training data set but deteriorates after application beyond the realm of the training data. ML algorithms, therefore, have some limitations when it comes to temporal extrapolation [44]. Using the sample- and site-based CV, the accuracy of the model and spatial variability can be well verified. To assess the accuracy of model prediction in those days beyond the modeling days, the model is also verified based on the time-based CV with R^2 of 0.56 (figure not shown). The time-based CV is like the sample- and site-based CV and performs the 10-CV process with 10% of days randomly dropped. A few previous studies validated the model with the time-based CV approach with the most of them have reported a low value of CV-R^2, i.e., 0.57 for the RF model, 0.63 for the space–time RF model [12], 0.68 for the XGBoost model [45], and 0.52 for the nested linear mixed effect (LME) model [14]. These results indicate that the model has the limitation on the $PM_{2.5}$ prediction for future time. However, the training data set can be iteratively updated by incorporating with the near real-time surface measured $PM_{2.5}$ concentration, satellite observed reflectances, and the related meteorological variables to update the Ref_{250}-$PM_{2.5}$ model and then estimate the surface $PM_{2.5}$ concentration at satellite pixels at near real-time level.

VI. Conclusion

In this section, a $PM_{2.5}$ estimation model with an extremely high spatial resolution of 250 m was developed based on the RF algorithm by directly using MODIS measurements of TOA reflectance and relevant meteorological variables. To the best of our knowledge, it is the first time to derive such high spatial resolution of $PM_{2.5}$ product from common satellite retrievals, such as MODIS. Compared with most previous models with coarse spatial resolutions, the proposed model estimates $PM_{2.5}$ concentrations on ultrafine spatial resolution (250 m) very well, especially at city and city district scales. The model performs very well with R^2 and RMSE (MPE and RPE) of 0.90 and 12.0 $\mu g/m^3$ (7.8 $\mu g/m^3$ and 16.9%) for the sample-based CV, respectively, and 0.86 and 13.7 $\mu g/m^3$ (8.9 $\mu g/m^3$ and 19.5%) for the site-based CV, respectively.

The model accurately captures the fine-detailed features in the distribution pattern and magnitude of $PM_{2.5}$ concentrations over the YRD region for seasonal mean, daily variations, and different air pollution conditions (low, moderate, and severely high). The proposed model also generates $PM_{2.5}$ concentrations with more complete spatial coverage than the MxD3km and MAIAC AOD-$PM_{2.5}$ models do. It can capture the uneven spatial distribution of $PM_{2.5}$ concentrations at a fine spatial resolution, suggesting a way of locating the main sources of $PM_{2.5}$ emissions. $PM_{2.5}$ concentrations generated by the proposed model have a much finer spatial resolution than most previously developed products, useful for air-pollution-related studies, and pollution monitoring and evaluation by governments, especially in urban areas and their districts where human activities are generally intensive.

ACKNOWLEDGMENT

Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), date of access. https://cds.climate.copernicus.eu/cdsapp#!/home.

REFERENCES

[1] R. D. Brook *et al.*, "Particulate matter air pollution and cardiovascular disease," *Circulation*, vol. 121, no. 21, pp. 2331–2378, 2010.

[2] R. B. Hayes *et al.*, "$PM_{2.5}$ air pollution and cause-specific cardiovascular disease mortality," *Int. J. Epidemiol.*, vol. 49, no. 1, pp. 25–35, Feb. 2020.

[3] X.-D. Yan *et al.*, "Polydatin protects the respiratory system from $PM_{2.5}$ exposure," *Sci. Rep.*, vol. 7, no. 1, Feb. 2017, Art. no. 40030.

[4] J.-Z. Wu, D.-D. Ge, L.-F. Zhou, L.-Y. Hou, Y. Zhou, and Q.-Y. Li, "Effects of particulate matter on allergic respiratory diseases," *Chronic Diseases Transl. Med.*, vol. 4, no. 2, pp. 95–102, Jun. 2018.

[5] J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer, "The contribution of outdoor air pollution sources to premature mortality on a global scale," *Nature*, vol. 525, no. 7569, pp. 367–371, Sep. 2015.

[6] L. Xiao, Y. Lang, and G. Christakos, "High-resolution spatiotemporal mapping of $PM_{2.5}$ concentrations at mainland China using a combined BME-GWR technique," *Atmos. Environ.*, vol. 173, pp. 295–305, Jan. 2018.

[7] W. You, Z. Zang, L. Zhang, Y. Li, and W. Wang, "Estimating national-scale ground-level PM_{25} concentration in China using geographically weighted regression based on MODIS and MISR AOD," *Environ. Sci. Pollut. Res.*, vol. 23, no. 9, pp. 8327–8338, May 2016.

[8] F. Yao, J. Wu, W. Li, and J. Peng, "Estimating daily $PM_{2.5}$ concentrations in Beijing using 750-M VIIRS IP AOD retrievals and a nested spatiotemporal statistical model," *Remote Sens.*, vol. 11, no. 7, p. 841, Apr. 2019.

[9] J. Liu, F. Weng, Z. Li, and M. C. Cribb, "Hourly $PM_{2.5}$ estimates from a geostationary satellite based on an ensemble learning algorithm and their spatiotemporal patterns over central east China," *Remote Sens.*, vol. 11, no. 18, p. 2120, Sep. 2019.

[10] S. Park *et al.*, "Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea," *Atmos. Chem. Phys.*, vol. 19, no. 2, pp. 1097–1113, Jan. 2019.

[11] J. Hu, Y. Wang, Q. Ying, and H. Zhang, "Spatial and temporal variability of $PM_{2.5}$ and PM_{10} over the north China plain and the yangtze river delta, China," *Atmos. Environ.*, vol. 95, pp. 598–609, Oct. 2014.

[12] J. Wei *et al.*, "Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 111221.

[13] A. Mhawish *et al.*, "Estimation of high-resolution $PM_{2.5}$ over the indo-gangetic plain by fusion of satellite data, meteorology, and land use variables," *Environ. Sci. Technol.*, vol. 54, no. 13, pp. 7891–7900, Jul. 2020.

[14] T. Zhang *et al.*, "Estimation of ultrahigh resolution $PM_{2.5}$ concentrations in urban areas using 160 m Gaofen-1 AOD retrievals," *Remote Sens. Environ.*, vol. 216, pp. 91–104, Oct. 2018.

[15] B. Zhang, M. Zhang, J. Kang, D. Hong, J. Xu, and X. Zhu, "Estimation of PMx concentrations from landsat 8 OLI images based on a multilayer perceptron neural network," *Remote Sens.*, vol. 11, no. 6, p. 646, Mar. 2019.

[16] A. van Donkelaar *et al.*, "Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application," *Environ. Health Perspect.*, vol. 118, no. 6, pp. 847–855, Jun. 2010.

[17] Y. Zhang and Z. Li, "Remote sensing of atmospheric fine particulate matter ($PM_{2.5}$) mass concentration near the ground from satellite observation," *Remote Sens. Environ.*, vol. 160, pp. 252–262, Apr. 2015.

[18] Z. Ma, Y. Liu, Q. Zhao, M. Liu, Y. Zhou, and J. Bi, "Satellite-derived high resolution $PM_{2.5}$ concentrations in yangtze river delta region of China using improved linear mixed effects model," *Atmos. Environ.*, vol. 133, pp. 156–164, May 2016.

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

[20] Y. Liu, J. A. Sarnat, V. Kilaru, D. J. Jacob, and P. Koutrakis, "Estimating ground-level $PM_{2.5}$ in the eastern united states using satellite remote sensing," *Environ. Sci. Technol.*, vol. 39, no. 9, pp. 3269–3278, May 2005.

[21] X. Hu *et al.*, "Estimating $PM_{2.5}$ concentrations in the conterminous United States using the random forest approach," *Environ. Sci. Technol.*, vol. 51, no. 12, pp. 6936–6944, Jun. 2017.

[22] G. Chen *et al.*, "A machine learning method to estimate $PM_{2.5}$ concentrations across China with remote sensing, meteorological and land use information," *Sci. Total Environ.*, vol. 636, pp. 52–60, Sep. 2018.

[23] H. Shen, T. Li, Q. Yuan, and L. Zhang, "Estimating regional ground-level $PM_{2.5}$ directly from satellite top-of-atmosphere reflectance using deep belief networks," *J. Geophys. Research: Atmos.*, vol. 123, no. 24, p. 13, Dec. 2018.

[24] P. Gupta and S. A. Christopher, "Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach," *J. Geophys. Res.*, vol. 114, no. D20205, 2009, Art. no. 2008JD011496.

[25] Y. Wu *et al.*, "Synergy of satellite and ground based observations in estimation of particulate matter in eastern China," *Sci. Total Environ.*, vol. 433, pp. 20–30, Sep. 2012.

[26] T. Li *et al.*, "Estimating ground-level $PM_{2.5}$ by fusing satellite and station observations: A geo-intelligent deep learning approach," *Geophys. Res. Lett.*, vol. 44, no. 23, 2017, Art. no. 2017GL075710.

[27] Y. Zhan *et al.*, "Spatiotemporal prediction of continuous daily $PM_{2.5}$ concentrations across China using a spatially explicit machine learning algorithm," *Atmos. Environ.*, vol. 155, pp. 129–139, Apr. 2017.

[28] Z.-Y. Chen *et al.*, "Extreme gradient boosting model to estimate $PM_{2.5}$ concentrations with missing-filled satellite data in China," *Atmos. Environ.*, vol. 202, pp. 180–189, Apr. 2019.

[29] G. Geng *et al.*, "Satellite-based daily $PM_{2.5}$ estimates during fire seasons in colorado," *ISEE Conf. Abstr.*, vol. 2018, no. 1, pp. 8159–8171, Sep. 2018.

[30] J. Liu, F. Weng, and Z. Li, "Satellite-based $PM_{2.5}$ estimation directly from reflectance at the top of the atmosphere using a machine learning algorithm," *Atmos. Environ.*, vol. 208, pp. 113–122, Jul. 2019.

[31] J. Wei *et al.*, "Improved 1 km resolution $PM_{2.5}$ estimates across China using enhanced space–time extremely randomized trees," *Atmos. Chem. Phys.*, vol. 20, no. 6, pp. 3273–3289, 2020.

[32] H. Hersbach *et al.*, "The ERA5 global reanalysis," *Quart. J. Royal Meteorolog. Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020.

[33] A. Liaw and M. Wiener, "Classification and regression by random forest," *R Newslett.*, vol. 2, no. 3, pp. 18–22, 2002.

[34] Z. Ma *et al.*, "Satellite-based spatiotemporal trends in $PM_{2.5}$ concentrations: China, 2004–2013," *Environ. Health Perspect.*, vol. 124, no. 2, pp. 184–192, Feb. 2016.

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

[36] Q. Xiao *et al.*, "Full-coverage high-resolution daily $PM_{2.5}$ estimation using MAIAC AOD in the yangtze river delta of China," *Remote Sens. Environ.*, vol. 199, pp. 437–446, Sep. 2017.

[37] M. Jiang, W. Sun, G. Yang, and D. Zhang, "Modelling seasonal GWR of daily $PM_{2.5}$ with proper auxiliary variables for the yangtze river delta," *Remote Sens.*, vol. 9, no. 4, p. 346, Apr. 2017.
[38] D. Tang, D. Liu, Y. Tang, B. C. Seyler, X. Deng, and Y. Zhan, "Comparison of GOCI and Himawari-8 aerosol optical depth for deriving full-coverage hourly $PM_{2.5}$ across the yangtze river delta," *Atmos. Environ.*, vol. 217, Nov. 2019, Art. no. 116973.
[39] Q. She *et al.*, "Satellite-based estimation of hourly $PM_{2.5}$ levels during heavy winter pollution episodes in the yangtze river delta, China," *Chemosphere*, vol. 239, Jan. 2020, Art. no. 124678.
[40] W. Fan, K. Qin, Y. Cui, D. Li, and M. Bilal, "Estimation of hourly ground-level $PM_{2.5}$ concentration based on Himawari-8 apparent reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 76–85, Jan. 2020.
[41] J. Guo *et al.*, "The climatology of planetary boundary layer height in China derived from radiosonde and reanalysis data," *Atmos. Chem. Phys.*, vol. 16, no. 20, pp. 13309–13319, Oct. 2016.
[42] J. Yu, T. Zhou, Z. Jiang, and L. Zou, "Evaluation of near-surface wind speed changes during 1979 to 2011 over China based on five reanalysis datasets," *Atmosphere*, vol. 10, no. 12, p. 804, Dec. 2019.
[43] R. C. Levy, L. A. Remer, S. Mattoo, E. F. Vermote, and Y. J. Kaufman, "Second–generation operational algorithm: Retrieval of aerosol properties over land from inversion of moderate resolution imaging spectroradiometer spectral reflectance," *J. Geophys. Res., Atmos.*, vol. 112, no. D13, pp. 1–21, Jul. 2007.
[44] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019.
[45] Q. He and B. Huang, "Satellite-based high-resolution $PM_{2.5}$ estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model," *Environ. Pollut.*, vol. 236, pp. 1027–1037, May 2018.

Jianjun Liu received the M.S. and Ph.D. degrees in atmospheric physics and atmospheric environment from the Nanjing University of Information Science and Technology, Nanjing, China, in 2008 and 2012, respectively.

He is a Senior Research Scientist with the Laboratory of Environmental Model and Data Optima (EMDO), Laurel, MD, USA. His research interests include remote sensing of aerosol and cloud properties, aerosol direct effects, and the interaction between aerosol and cloud.

Fuzhong Weng (Member, IEEE) received the Ph.D. degree from the Department of Atmospheric Sciences, Colorado State University, Fort Collins, CO, USA, in 1992.

He was a Physical Scientist at the United States National Oceanic and Atmospheric Administration (NOAA), Washington, DC, USA, from 1992 to 2017. He is a Professor with the Chinese Academy of Meteorological Sciences, Beijing, China. He has published more than 300 peer-reviewed articles and nine book chapters on remote sensing sciences. In 2017, he has also published a book on *Passive Microwave Remote Sensing of the Earth for Meteorological Applications* (Wiley Series in Atmospheric Physics and Remote Sensing). His major research areas include radiative transfer, satellite remote sensing, satellite data assimilation, and instrument calibration.

Dr. Weng is a fellow of the American Meteorological Society. He received many awards, including the NOAA David Johnson Award in 2000, the U.S. Department of Commerce Gold Medal in 2005, and the NOAA Administrator Award on Science and Technology in 2009. He is the Co-Chief Editor of the *Journal of Meteorological Research*.

Zhanqing Li received the B.Sc. and M.Sc. degrees from the Nanjing University of Information Science and Technology, Nanjing, China, in 1983 and 1989, respectively, and the Ph.D. degree from McGill University, Montreal, QC, Canada, in 1991.

He is a Professor with the University of Maryland, College Park, MD, USA. He has authored over 350 articles. His research interests include remote sensing, atmospheric physics, and climate and environment focusing on aerosol, cloud, radiation budget, and precipitation.

Dr. Li is a fellow of the American Meteorological Society (AMS), the American Geophysical Union (AGU), and the American Association for the Advancement of Science (AAAS). He received numerous awards in the USA, Canada, and Germany. He is an Editor of the *Atmospheric Chemistry and Physics*.