Satellite-Derived 1-km-Resolution PM$_1$ Concentrations from 2014 to 2018 across China

Jing Wei,[†,‡] Zhanqing Li,*[,‡] Jianping Guo,[§] Lin Sun,[||] Wei Huang,[⊥] Wenhao Xue,[†] Tianyi Fan,[†] and Maureen Cribb[‡]

[†]State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China

[‡]Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland 20742, United States

[§]State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China

[||]College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China

[⊥]State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100101, China

S Supporting Information

ABSTRACT: Particulate matter with aerodynamic diameters ≤1 μm (PM$_1$) has a greater impact on the human health but has been less studied due to fewer ground observations. This study attempts to improve the retrieval accuracy and spatial resolution of satellite-based PM$_1$ estimates using the new ground-based monitoring network in China. Therefore, a space-time extremely randomized trees (STET) model is first developed to estimate PM$_1$ concentrations at a 1 km spatial resolution from 2014 to 2018 across mainland China. The STET model can derive daily PM$_1$ concentrations with an average across-validation coefficient of determination of 0.77, a low root-mean-square error of 14.6 μg/m^3, and a mean absolute error of 8.9 μg/m^3. PM$_1$ concentrations are generally low in most areas of China, except for the North China Plain and Sichuan Basin where intense human activities and poor natural conditions are prevalent, especially in winter. Moreover, PM$_1$ pollution

has greatly decreased over the past 5 years, benefiting from emission control in China. The STET model, incorporating the spatiotemporal information, shows superior performance in PM$_1$ estimates relative to previous studies. This high-resolution and high-quality PM$_1$ data set in China (i.e., ChinaHighPM$_1$) can be greatly useful for air pollution studies in medium- or small-scale areas.
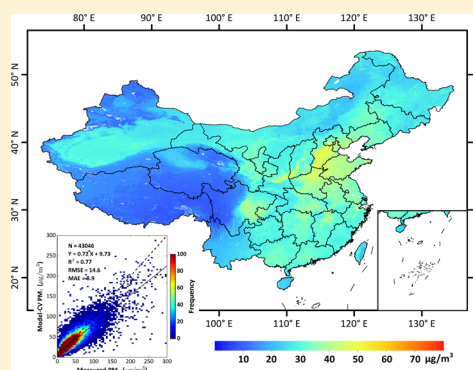
INTRODUCTION

Due to economic development and the acceleration of urbanization over the past few decades in China, intense human activities have emitted abundant pollutants, leading to an increasingly serious air pollution problem. The World Health Organization (WHO) has reported that more than 80% of the world's population experienced air pollution in 2018, especially fine particulate matter with aerodynamic diameters less than 10 μm (PM$_{10}$), 2.5 μm (PM$_{2.5}$), and 1 μm (PM$_1$), and more than 8 million people have lost their lives each year because of pollution.[1,2] Fine particulate matter has become the primary pollutant affecting the air quality of urban areas and has thus attracted worldwide attention.[3]

PM$_{10}$ is mostly emitted by mechanical processes such as construction-generated and wind-blown mineral dust/soil and sea-salt. PM$_{2.5}$ often comes from anthropogenic activities such as power plants, industrial production, and emissions of residues. While some sources are common, PM$_1$ is mainly derived from direct emissions during the combustion

process;[4−6] more importantly, it can incur greater environmental and health effects because these smaller-size particles can be inhaled deeper into the lungs, more easily entering alveoli and contaminating them with toxic and harmful substances.[7,8] In addition, previous studies have illustrated that the spatial distributions and ratios between these fine particles (i.e., PM$_1$/PM$_{2.5}$ and PM$_{2.5}$/PM$_{10}$) vary greatly (range from <0.5 to >0.9) from site to regional scales, in different seasons and polluted levels across China depending on various natural conditions and human activities, indicating more complex relationships.[9,10]

To monitor atmospheric fine particulate matter in real time, many countries around the world have established various ground networks with hundreds to thousands of monitoring

stations. Despite their high accuracy and acquisition frequency, in situ observations cannot realize monitoring on a wide scale. However, satellite remote sensing provides an alternative and more effective way to overcome this limitation by virtue of the relationships between particulate matter and aerosol optical depth (AOD).[11] Based on this, multisource satellite aerosol products derived from different instruments, e.g., the Moderate Resolution Imaging Spectroradiometer (MODIS),[12] the Multi-angle Imaging SpectroRadiometer (MISR),[13] the Visible Infrared Imaging Radiometer Suite (VIIRS),[14] and Himawari-8 Imager,[15] have been used to estimate the concentration of surface particulate matters.

Thanks to the publicly available in situ measurements of $PM_{2.5}$ and PM_{10}, there have been much more studies on these two air pollution indexes including their retrievals from various satellite-derived aerosol products using the physical models,[16] the statistical regression models,[12−14] and the machine or deep learning approaches.[15,17−19] By contrast, studies on PM_1 are much scantier due to much fewer ground monitoring stations, which posed difficulties in developing and validating satellite-derived algorithms. Chen et al. used the generalized additive model (GAM) to study the spatiotemporal variations in PM_1 concentrations at a spatial resolution of 10 km using MODIS AOD products and ancillary data from 2005 to 2014 in China.[20] Zang et al. developed a principal component analysis-general regression neural network (PCA-GRNN) model to estimate hourly PM_1 concentrations at a 5 km resolution using Himawari-8 AOD products from 2013 to 2014 in China[15] and then improved the estimates by integrating $PM_{2.5}$ observations.[21] Wang et al. presented a two-stage model, combining the linear mixed effect (LME) and bagged tree (BT) models, to estimate hourly PM_1 concentrations at a 5 km resolution from 2015 to 2017 over central and eastern China using Himawari-8 AOD products.[22]

These traditional methods still face great challenges in estimating PM_1 concentrations due to weak data mining capability and negligence of spatial-temporal heterogeneities. Besides, the generated PM_1 datasets are at coarse spatial resolutions (5−10 km), which seriously limited their applications over medium- and small-scale areas (e.g., urban). Therefore, in this study, a new space-time extremely randomized trees (STET) model is first proposed to improve the spatial resolution and overall accuracy of PM_1 estimates across mainland China. For this, the newly released MODIS Collection 6 (C6) Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD products at a spatial resolution of 1 km are employed, which is higher than those used in previous studies. Together with meteorological variables, emissions, ancillary data, and the spatiotemporal information are introduced to the STET model. Then, the high-spatial-resolution (1 km) and high-quality PM_1 concentrations across mainland China (i.e., ChinaHighPM$_1$ data set) from 2014 to 2018 are generated for the first time.

MATERIALS AND METHODS

Surface PM$_1$ Measurements. The daily PM_1 in situ measurements from 1 Jan 2014 to 31 Dec 2018 were collected from the China Atmosphere Watch Network (CAWNET) of the China Meteorological Administration. They were measured using the GRIMM Model 1.180 Aerosol Spectrometer, which is an optical particle counter that records particle mass and size distribution every five minutes, covering wavelengths ranging from 1.0 to 10 μm. These measurements inferred from

light scattering were converted into mass concentrations using the GRIMM protocols to be consistent with PM measurements (aerodynamic diameter) according to the China's National Ambient Air Quality Standard.[9,23] Previous studies have reported that the GRIMM-derived PM concentrations are highly consistent with those measured by the tapered element oscillating microbalance instruments in different regions under different background aerosols.[24−26] In addition, two main quality-control procedures, i.e., the limit check and the climatological check, are performed to remove any unreliable, low-quality observations and outliers and invalid values arising from instrument malfunction or improper calibration.[10,11] There are a total of 153 monitoring stations covering most provinces across China (Figure S1). These stations are evenly distributed in central and eastern China where there are intensive human activities and are relatively sparse in western China where there is little human population.

MAIAC AOD Product. The National Aeronautics and Space Administration's (NASA) Terra and Aqua MODIS C6 MAIAC Level (L) 2 swath aerosol products (MCD19A2) from 2014 to 2018 across China are employed. MAIAC aerosol products are generated directly from MODIS L1B radiance data at its native resolution of 1 km using the multi-angle implementation of atmospheric correction algorithm. The algorithm is based on the physical atmosphere−surface model with minimal assumption and realized using the look-up-table approach.[27] The MAIAC aerosol algorithm differs from the MODIS Dark Target (DT) and Deep Blue (DB) algorithms, and the MAIAC aerosol products, released on 30 May 2018, are at 3−10 times higher spatial resolution and more accurate than most other widely used DT and DB AOD products, especially over bright urban surfaces.[28−30] Therefore, in this study, the Terra and Aqua MCD19A2 daily AOD products are selected and only those retrievals at 550 nm passing the recommended quality assurance (QA) filtered by the best cloud ($QA_{CloudMask}$ = Clear) and adjacency ($QA_{AdjacencyMask}$ = Clear) masks are used here.

Meteorological Data. Meteorological data used in this study are from the ERA-Interim atmospheric reanalysis products.[31] This reanalysis has provided long-term (since 1979) and real-time (every 3 or 6 h) atmospheric parameters at different spatial resolutions. In this study, eight meteorological variables having potential impacts on PM_1 at the highest spatial resolution of 0.125° are collected, i.e., boundary layer height (BLH, unit: m), evaporation (ET, unit: mm), precipitation (PRE, unit: mm), relative humidity (RH, unit: %), surface pressure (SP, unit: hPa), temperature (TEM, unit: K), wind speed (WS, unit: m/s), and wind direction (WD, unit: °). The mean values of the above meteorological variables between 10 a.m. and 2:00 p.m. local time are calculated to be coincident with MODIS overpass times.

MEIC Emission Data. As aerosol originated from emissions, the multi-resolution emission inventory for China (MEIC) is also used in this study. MEIC provides the inventory of monthly and annual atmospheric pollutants and greenhouse gases (i.e., SO_2, NO_x, CO, PM, BC, etc.) based on more than 700 anthropogenic sources at a spatial resolution of 0.25° in China.[32,33] Four monthly emission variables contributing to fine particulate matters from industry (IDT, unit: Mg/grid), power (POW, unit: Mg/grid), residential (RSD, unit: Mg/grid), and transportation (TST, unit: Mg/grid) sectors are selected. The total amount is first weight-

averaged for each grid and downscaled to a spatial resolution of 0.01° (approximately 1 km).

Other Ancillary Data. Other factors that may affect the PM_1−AOD relationships are also considered such as land cover, topography, traffic, and population. They include the MODIS Level 3 annual Land Use Cover (LUC) product (MCD12Q1) and the monthly normalized difference vegetation index (NDVI) product (MOD13A3) at a 500 m resolution. Data from the Shuttle Radar Topography Mission digital elevation model (DEM) at a 90 m resolution is collected and used to calculate the surface aspect, slope, and relief[34] to reflect surface variations. The spatial distance from roads (road, unit: m) in cities, provinces, and the whole country generated using the multiple ring buffer approach at a spatial resolution of 1 km is selected to represent traffic conditions.[35] The VIIRS monthly day/night band nighttime lights (NTL, $W/cm^2/sr$) product at a 500 m resolution is selected to represent human activities and economic conditions. Table S1 summarizes the data sources used in this study.

Data Integration. Due to the ubiquity of clouds in remote sensing images, there is a large number of missing values in MODIS MAIAC AOD products, especially in southern China (Figure S2). The spatial coverage of Terra MAIAC product ranges from 0 to 71% with an average of 30% from 2014 to 2018 in China, which is larger than the Aqua MAIAC product (i.e., maximum = 64% and average = 26%) due to less cloud cover in the morning. These differences are due to the different imaging times between Terra and Aqua satellites, but the spatial coverage of aerosols can be improved by integrating these two data sets. Therefore, to reduce the system differences, a linear regression approach is used to merge the Terra and Aqua MAIAC AOD products.[19] For a given day, the Terra and Aqua AODs (τ_{Terra} and τ_{Aqua}) are calculated using established linear regression relationships (eq 1), and if both are available, an additional average approach is applied for each pixel (eq 2)

$$\begin{cases} \tau_{Terra} = k_1 \cdot \tau_{Aqua} + b_1 \\ \tau_{Aqua} = k_2 \cdot \tau_{Terra} + b_2 \end{cases} \tag{1}$$

$$\tau = (\tau_{Terra} + \tau_{Aqua})/2 \tag{2}$$

where k_1, k_2 and b_1, b_2 represent the slopes and y-intercepts of the regression lines, respectively. Both Terra and Aqua daily AODs at all monitoring stations in each month are collected and used to obtain monthly relationships. Through data fusion, the spatial coverage of AOD largely increased (i.e., maximum = 80% and average = 40%) over most areas of China, especially northern China (difference > 15%). The percentage increases in the number of daily AOD data samples after data fusion are ~23 and 27% compared to using only Terra or Aqua MAIAC AOD products, respectively. This approach can improve the model training capability by increasing the number of data samples.

In addition, due to different spatial resolutions, the coarser ERA-Interim meteorological and MEIC emission variables and the finer-resolution ancillary data (i.e., LUC, NDVI, topographic conditions, and NTL) are first interpolated to a uniform spatial resolution of 1 km using the bilinear interpolation approach. All independent variables are then matched to the daily PM_1 concentrations at the site scale for each day. After removing all invalid values, there are 12 306,

7631, 7554, 7759, and 7796 matched daily samples covering all days for each year from 2014 to 2018 in China, respectively.

Model Development and Adjustment. The extremely randomized trees (ERT) model is a new tree-based ensemble approach for classification and regression, different from other tree models (e.g., the decision tree and the random forest).[36] The main difference is that the ERT model strongly strengthens the randomization in both attribute and cut-point selection while splitting a tree node. The splitting procedure of the ERT model contains two main parameters, one of which is the number of attributes randomly selected by each node (K), which controls the strength of the attribute randomization. The other main parameter is the minimum sample size for splitting a node (n_{min}), which controls the smoothing degree. The randomized trees are then built, and their structures are independent of the output values using the training samples in the extreme case. This approach has high computational efficiency and good training ability. Each tree in the ERT model is built with the following steps:

1) A training set S is randomly extracted from the whole learning sample rather than the bootstrap approach.
2) If S meets the condition of $|S| < n_{min}$, or all candidate variables or the output variable are constant in S, a leaf labeled by the average output is returned in regression.
3) If not, K attributes $\{a_1, ..., a_K\}$ are randomly selected without replacement among all of the candidate attributes and K splits $\{s_1, ..., s_K\}$ are generated.
4) A split s_* is then determined to meet the condition of $score(s_*, S) = max[Score(s_i, S)]$ (eq 3), and S is split into two subsets (S_l and S_r) based on the s_*.
5) Then, S_l and S_r are used as inputs, and the procedures (2)−(4) are repeated to build t_l and t_r.
6) The resulting tree (t) is built by creating a node of the s_* with t_l and t_r as left and right subtrees, respectively.

$$score(s, S) = \frac{var\{y|S\} - \frac{|S_l|}{S}var\{y|S_l\} - \frac{|S_r|}{S}var\{y|S_r\}}{var\{y|S\}} \tag{3}$$

where $var\{y|S_x\}\}$ represents the variance of the output y in the training set S.

All selected independent variables with potential effects on PM_1 estimations are input to the ERT model. However, atmospheric fine particulate matter (i.e., $PM_{2.5}$, and PM_1) are temporally and spatially heterogeneous.[18,19] Therefore, a new space-time extremely randomized trees (STET) model is first developed by introducing spatiotemporal information to the ERT model. The STET model considers the spatial autocorrelation (space) weighted by the geographical distance (ws) between the center pixel (P) and its adjacent pixels and the temporal difference (time) determined by the different days in a given year from PM_1 observations.[19] The improved PM_1−AOD relationship can then be explored using MAIAC AODs, meteorological data, pollution emissions, land use, topography, and road, population, and spatiotemporal information as input to the STET model.

The potential correlations and effects of 21 selected independent variables on PM_1 estimates are first calculated (Figure S3). Satellite-derived AOD products are not closely correlated to ground PM_1 observations with an average correlation coefficient (R) of 0.378 (Table S2). Therefore, the remaining variables are used to improve the PM_1−AOD relationship developed for China. In general, 10 variables show positive effects on surface PM_1 concentrations, especially RSD,
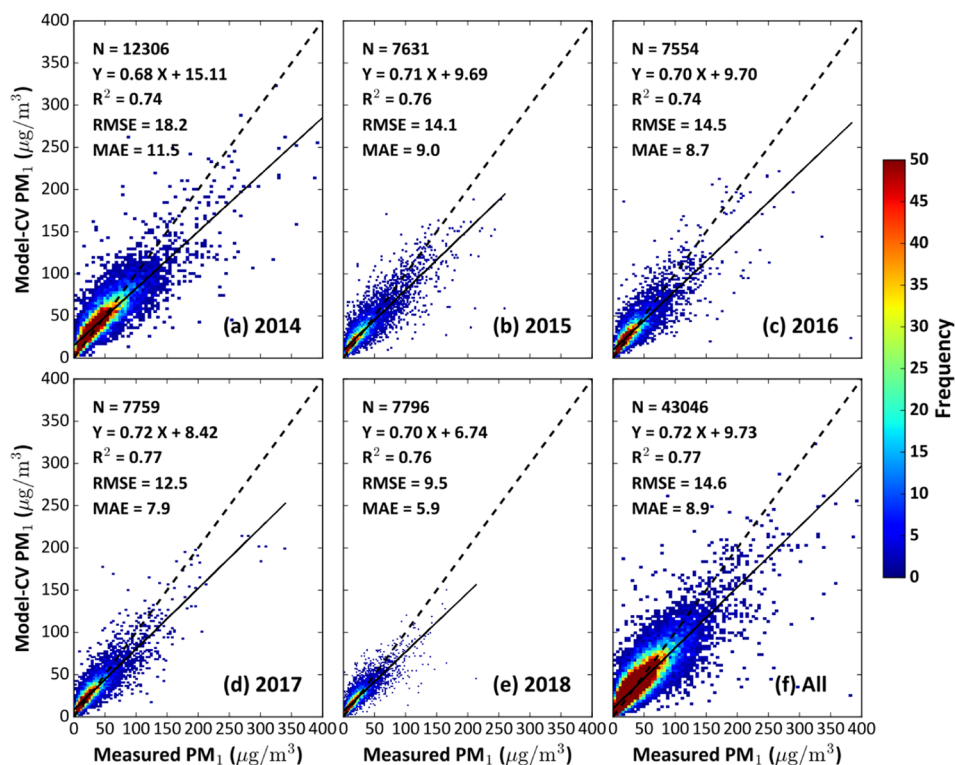
Figure 1. Density scatterplots of 10-fold cross-validation results of 1 km daily PM$_1$ estimates across China for the years (a)−(e) 2014 to 2018 and (f) all years. Statistical metrics are given in each panel: the number of samples (N), the coefficient of determination (R^2), the root-mean-square error (RMSE), and the mean absolute error (MAE). The linear regression relationship is also given in each panel. Dashed lines are the 1:1 lines, and solid lines are the linear best-fit lines through the data points.

SP, and ET with high correlations of 0.372, 0.236, and 0.192, respectively. By contrast, the remaining 10 variables have negative effects, especially BLH, DEM, and TEM with correlations of −0.351, −0.235, and −0.220, respectively. Almost all variables are closely associated with surface PM$_1$ concentrations at the 99% confidence levels (p-value <0.01).

Examined next are the contributions of more important variables rather than all variables to PM$_1$ to establish a stable estimation model for PM$_1$ concentrations and to optimize and adjust the model, improving its overall efficiency and accuracy. All meteorological, MEIC emission, land-use, topographic, traffic, and population variables with potential impacts on PM$_1$ estimations are first input into the STET model for preliminary training, and their importance scores are calculated.

Figure S4 shows the sorted average importance scores of all independent variables on PM$_1$ estimations in China from 2014 to 2018. Different variables influence PM$_1$ estimates differently. AOD is the most critical factor, accounting for ~15% of all variables. Three main meteorological variables, i.e., BLH (~10%), TEM (~6%), and ET (~5%), and residential emissions (~8%) contribute the most to PM$_1$ estimates. More importantly, space and time are two critical variables with an average importance score of ~8%, indicating the importance of spatiotemporal information when estimating PM$_1$ concentrations. The remaining factors show a gradual decline in contributions to PM$_1$ estimates, especially, the last 10 variables show less contributions to PM$_1$ estimation whose scores are below 2.5%. Therefore, these variables are excluded from the STET model and the remaining 13 variables are used for model training and establishment.

Evaluation and Analysis Methods. In this study, the commonly used 10-fold cross-validation (10-CV) approach,[37]

along with several statistical indicators including the root-mean-square error (RMSE) and the mean absolute error (MAE), is selected to validate the model performance. In addition, monthly PM$_1$ maps are generated by averaging adequate daily PM$_1$ values for each grid in a month. Seasonal and annual maps are generated from these monthly PM$_1$ estimates. The spatial coverage is calculated using the area-weighting approach. Linear trends in PM$_1$ pollution are calculated from deseasonalized monthly PM$_1$ anomalies, and the two-side test is selected to assess the statistical significance.[38]

RESULTS AND DISCUSSION

Model Fitting and Validation. The STET model shows great but overfitting results with coefficients of determination (R^2) equal to ~1 and with almost no biases (Figure S5). This attests to the strong learning ability of the STEL in data mining because of its high model complexity but is also compromised by the small number of data samples with respective to the large degree of freedom.[18,39] Figure 1 illustrates the 10-CV results from the newly developed STET model from 2014 to 2018 in China. The year 2014 has the largest number of data samples (N = 12 306) and the largest RMSE (18.2 μg/m^3) and MAE (11.5 μg/m^3). The main reason is that the air quality was relatively poor that year, with more than 12% of the data samples exceeding 80 μg/m^3. Another reason may be that the PM$_1$ data records at the beginning of the observation period were of relatively low quality. Approximately one-third of the monitoring stations gradually stopped working after 2014. The number of data samples decreased by ~38% from 2015 to 2018 in China. However, the air quality appears to have
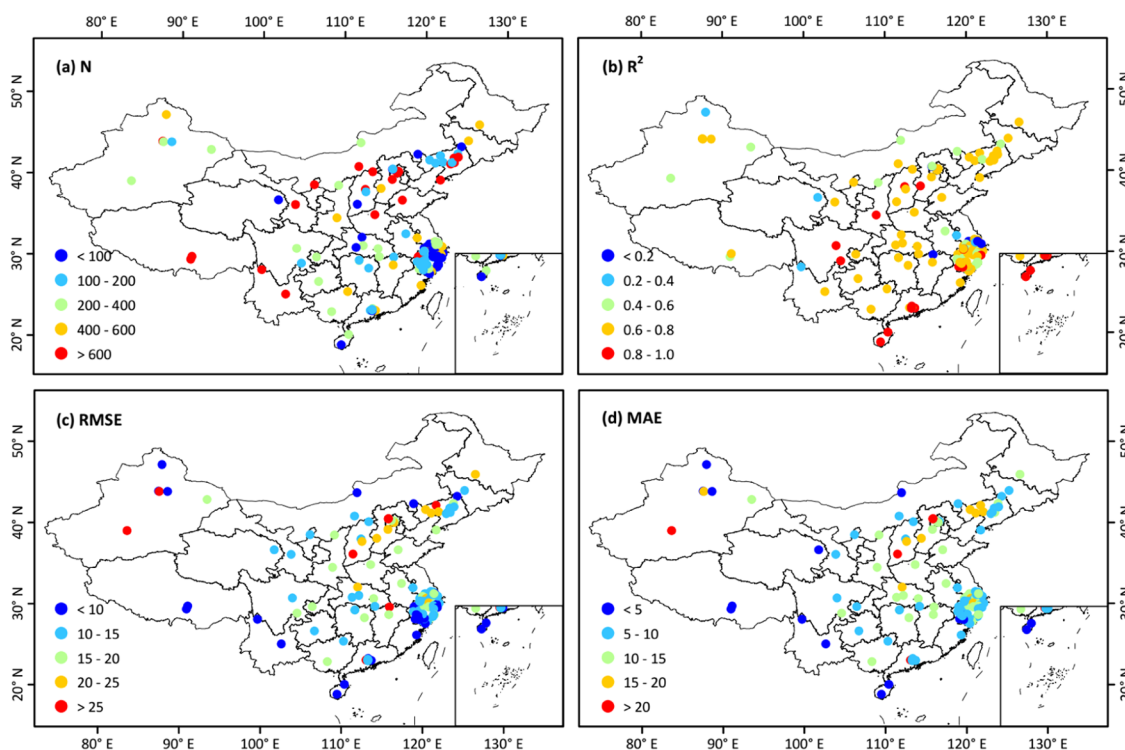
Figure 2. Spatial distributions of model cross-validation statistics based on PM_1 estimates and measurements from 2014 to 2018 across China in terms of (a) number of valid days of data (N), (b) coefficient of determination (R^2), (c) root-mean-square error (RMSE; $\mu g/m^3$), and (d) mean absolute error (MAE; $\mu g/m^3$).

improved, with more than 93% of the data samples below 50 $\mu g/m^3$ in 2018. The estimated PM_1 concentrations are gradually improved with larger cross-validation coefficients (CV-R^2 > 0.74), better linear regression lines, and smaller estimation uncertainties (i.e., RMSE < 14.1 $\mu g/m^3$ and MAE < 9.0 $\mu g/m^3$). There are differences in the model performance among different years, mainly due to different air quality conditions. Nevertheless, overall, PM_1 concentrations on highly polluted days are underestimated (i.e., slope < 0.73). In general, the STET model works well in estimating daily PM_1 concentrations with an average high CV-R^2 value equal to 0.77 and a small RMSE of 14.6 $\mu g/m^3$ and MAE of 8.9 $\mu g/m^3$ during 2014−2018 across mainland China.

The effects of spatiotemporal information on the estimation of PM_1 concentrations in China using the STET model are also investigated (Table S3). The original ERT model, which neglects spatiotemporal variations, has the lowest CV-R^2 and largest RMSE and MAE values. However, when considering either spatial (space) or temporal (time) information, the model performance improves with increasing CV-R^2 values and decreasing RMSE and MAE values. Considering both spatial and temporal information yields the best model performance in terms of the highest CV-R^2 values and the lowest RMSE and MAE values. These results suggest that incorporating spatiotemporal information into models that estimate PM_1 concentrations is crucial because it can significantly improve the model performance.

Figure 2 shows the spatial distributions of 10-fold cross-validation results for the STET model at individual monitoring stations from 2014 to 2018 in China. For statistical significance, only those stations with more than 10 matched data samples are plotted. The number of days with valid PM_1 estimates varies from 13 to 1181 days, with an average of 324

days. The Yangtze River Delta region contains most of those stations with less than 100 valid samples because only 1 year of surface observations is available from this region. Approximately 48% of all stations in China have more than 200 valid samples. Local CV-R^2 values range from 0.10 to 0.89, with an average value of 0.65. Approximately 71% of all stations, more evenly distributed throughout mainland China, have CV-R^2 values greater than 0.6. Local RMSE values range from 3.0 to 45.6 $\mu g/m^3$, with an average value of 14.4 $\mu g/m^3$, and local MAE values range from 2.1 to 29.1 $\mu g/m^3$, with an average value of 9.9 $\mu g/m^3$. In general, their spatial patterns are similar across mainland China. The Beijing−Tianjin−Hebei region has large RMSE and MAE values (i.e., RMSE > 20 $\mu g/m^3$ and MAE > 15 $\mu g/m^3$) due to more severe air pollution in that region. By contrast, the estimation uncertainties of PM_1 are relatively small in most other parts of China, with more than 85 and 88% of the stations showing overall low RMSE and MAE values of less than 20 and 15 $\mu g/m^3$, respectively. These results illustrate that the STET model performs well under different natural and population conditions.

The temporal performance of the STET model in China is also evaluated (Figure S6). There are 2892, 1277, and 373 matched data samples from 2014 to 2018 across China on the monthly, seasonal, and annual scales, respectively. Results suggest that the STET model more accurately captures monthly PM_1 concentrations. The data samples are evenly distributed on both sides of the 1:1 line, and the linear best-fit line through the data samples has a slope of 0.88 and a y-intercept of 4.38 $\mu g/m^3$. The average R^2 value is 0.96, and the RMSE and MAE values are 4.8 and 3.2 $\mu g/m^3$, respectively. Annual and seasonal PM_1 estimates are more consistent with ground measurements with stronger linear fits (i.e., slope > 0.89 and y-intercept < 4.24 $\mu g/m^3$), higher correlation
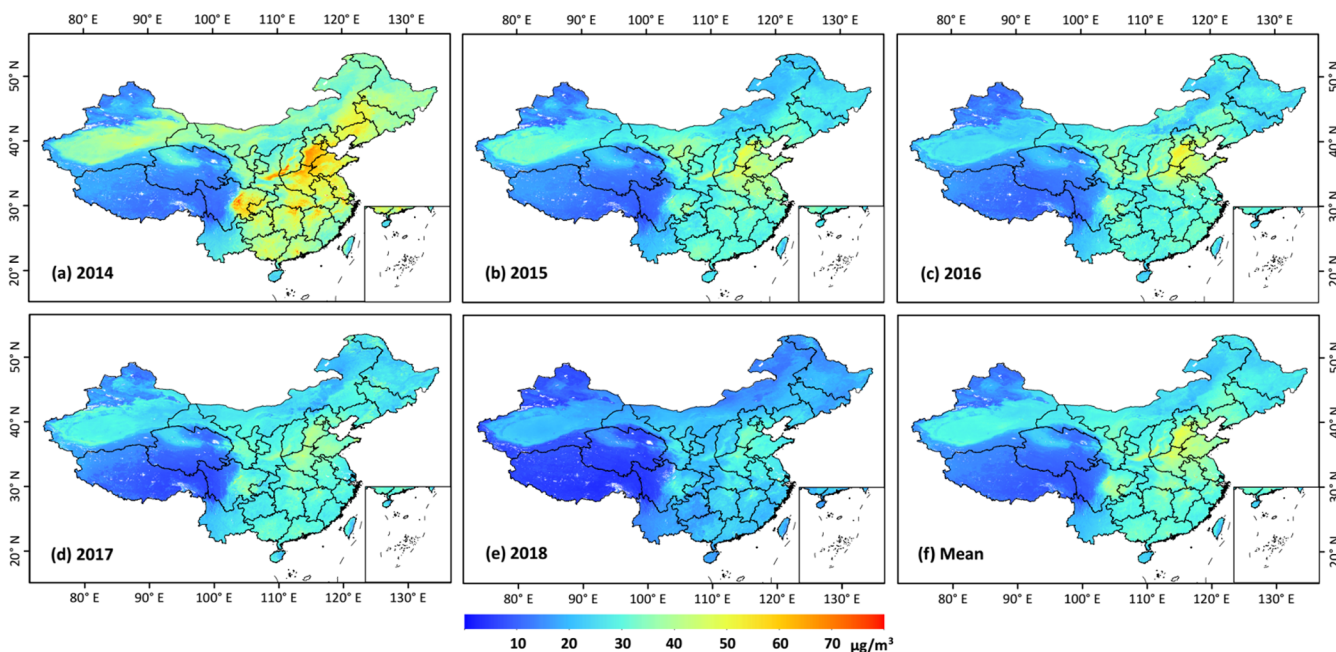
Figure 3. Satellite-derived 1-km-resolution annual mean PM_1 concentrations across China from (a)−(e) 2014 to 2018 and (f) for all years.
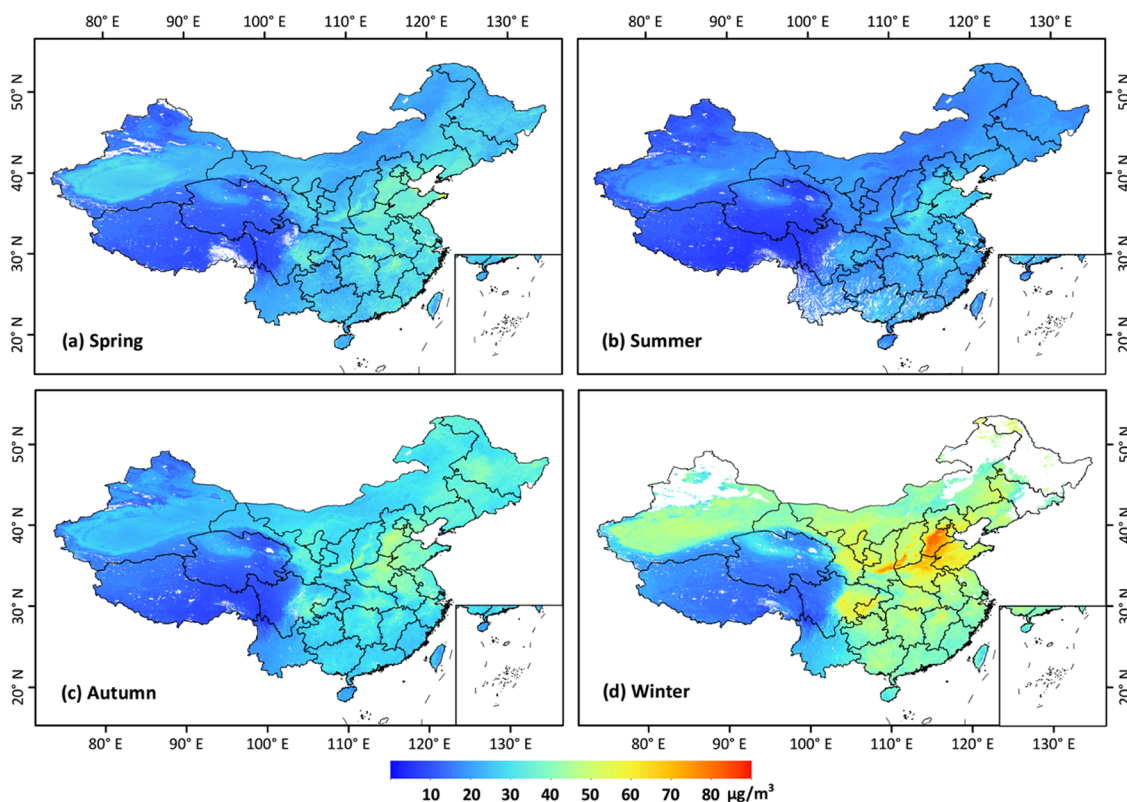

Figure 4. Seasonal mean 1-km-resolution PM_1 concentrations in (a) spring, (b) summer, (c) autumn, and (d) winter during 2014−2018 across China, respectively.

coefficients (0.97), and smaller estimation uncertainties (i.e., RMSE < 4.1 $\mu g/m^3$ and MAE < 2.8 $\mu g/m^3$). These results suggest that the PM_1 data set generated by the STET model can more reliably capture the spatial distributions and temporal variations of PM_1 pollution across China.

Spatial Coverage and Distribution. In this study, daily 1-km-resolution (∼0.01° grids) surface PM_1 concentrations are estimated using the STET model from 2014 to 2018 across

China and then averaged to produce monthly, seasonal, and annual PM_1 maps. This belongs to one of the series of high-resolution and high-quality PM_x data sets across China, i.e., ChinaHighPM$_x$ (x = 1). The spatial coverage of monthly PM_1 maps ranges from 57 to 95% with an average of 79% across China (Figure S7). The spatial coverage of PM_1 estimates is the smallest in June (average = 69%) and February (average = 74%) mainly due to frequent clouds in summer and long-term
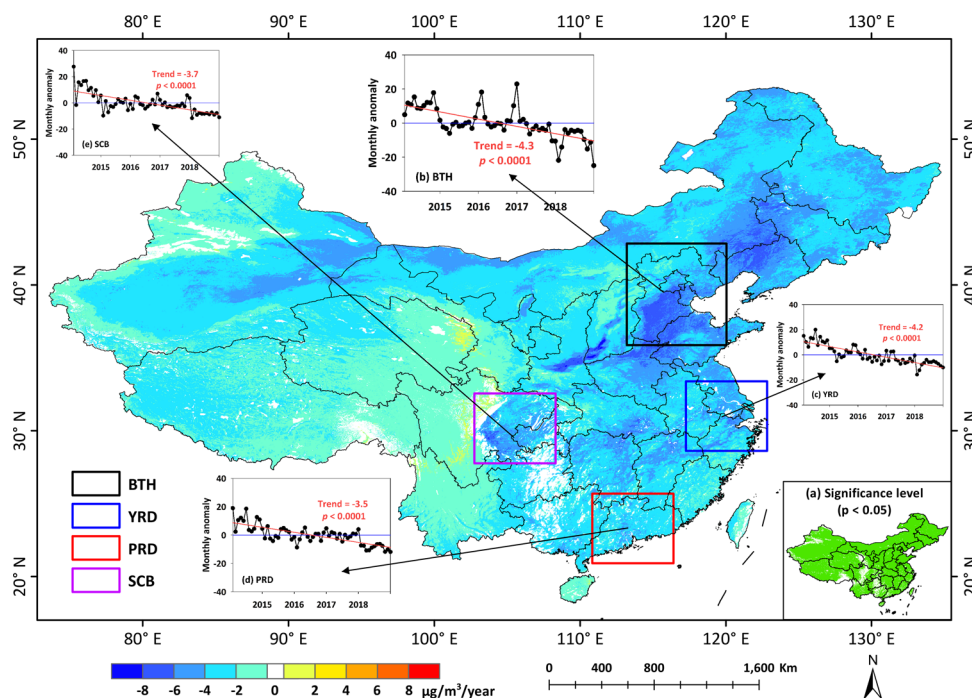
Figure 5. Linear trends ($\mu g/m^3$/year) derived from deseasonalized monthly PM_1 concentration anomalies from 2014 to 2018 in China, where green areas in the (a) inserted figure show areas where the trends are significant at 95% confidence level (*p*-value <0.05). The inserted figures show time series of monthly PM_1 anomalies over (b) Beijing−Tianjin−Hebei (BTH) region, (c) Yangtze River Delta (YRD), (d) Pearl River Delta (PRD), and (e) Sichuan Basin (SCB), where the red and blue lines represent the linear regression lines and *y* = 0 lines.

snow cover in winter. By contrast, October always has the most extensive spatial coverage (average = 88%). Moreover, PM_1 concentrations have regular interannual variations, with maximum and minimum concentrations seen in January (~40.0 $\mu g/m^3$) and August (~16.0 $\mu g/m^3$), respectively.

Figure 3 shows annual mean PM_1 concentration maps from 2014 to 2018 across China. The model provides a nearly complete spatial coverage (>98%) of annual PM_1 estimates across mainland China. In general, annual mean PM_1 concentrations vary greatly from essentially nil to 126 $\mu g/m^3$ with averages of 32.3 ± 12.8, 24.5 ± 8.7, 24.7 ± 8.7, 23.8 ± 8.0, and 16.8 ± 7.3 $\mu g/m^3$ for each year from 2014 to 2018 in China, respectively (Table S4). Estimated PM_1 concentrations are generally higher in 2014 than in the other 3 years, especially over eastern China where PM_1 values are always greater than 50 $\mu g/m^3$; by contrast, the PM_1 pollution appears to be the lightest in 2018.

The spatial distributions of PM_1 concentration are similar in different years on the local scale. High PM_1 concentrations are mainly distributed throughout the North China Plain and the Sichuan Basin, mainly due to severe pollutant emissions due to intense human activities and poor topographical or meteorological conditions. By contrast, southwestern, northeastern, and southern China, areas where there are less anthropogenic activities and more favorable climatic conditions, have much lower PM_1 concentrations. Note that the estimated PM_1 concentrations may differ from observations in some parts of China such as Qinghai and Gansu Provinces and Tibet in western China due to the limited number of monitoring stations.

In general, the PM_1 concentrations vary greatly on the seasonal scale (Figure 4). The spatial coverage of estimated PM_1 concentrations is the lowest in winter (~86%) due to snow/ice cover in northeastern and northwestern China

(Table S4). PM_1 pollution is also severe in winter with the highest mean PM_1 concentration (36.3 $\mu g/m^3$) and the largest standard deviation (14.5 $\mu g/m^3$). More importantly, more than 42% of mainland China has high PM_1 concentrations exceeding 40 $\mu g/m^3$, especially the North China Plain. Coal and fossil fuel combustion caused by anthropogenic activities explains this. By contrast, summer has the least PM_1 pollution with an average concentration of 16.4 ± 5.8 $\mu g/m^3$ (spatial coverage = 96%). PM_1 estimates are less than 30 $\mu g/m^3$ over more than 95% of the areas in China. Frequent rainfall and high relative humidity, which promote the diffusion of pollutants and reduce air pollution, are the main explanations. The PM_1 estimations are more consistent in spatial coverage in spring (~97%) and autumn (~99%) and also have similar PM_1 pollution levels with average values of 22.4 ± 7.5 and 24.4 ± 8.8 $\mu g/m^3$, respectively. Moreover, in spring, PM_1 concentrations are much lower than $PM_{2.5}$ concentrations at the Tarim Basin, Xinjiang Province, where coarser dust particles dominate.

Temporal Variation and Trend. Focusing on the numerical differences in PM_1 concentrations among different years, the temporal variations in PM_1 pollution are investigated from 2014 to 2018 in China. PM_1 concentrations have changed greatly in China over this time period (Figure 5), showing a significant decreasing trend of −3.0 $\mu g/m^3$/year (*p*-value <0.0001). In general, PM_1 concentrations have decreased over 96% of mainland China, especially in central and eastern China where the decrease is more than 6 $\mu g/m^3$ per year (*p*-value <0.05). Parts of southwest China where fewer people live show increasing trends in PM_1 concentrations. These trends, however, may not be reliable because few monitoring stations are located in this part of China. Four typical urban agglomerations, where air pollution is a public concern, show significant decreasing trends in PM_1 concentrations, i.e., the

Beijing−Tianjin−Hebei region (-4.3 $\mu g/m^3$/year, p-value <0.0001), the Yangtze River Delta (-4.2 $\mu g/m^3$/year, p-value <0.0001), the Pearl River Delta (-3.5 $\mu g/m^3$/year, p-value <0.0001), and the Sichuan Basin (-3.7 $\mu g/m^3$/year, p-value <0.0001).

Figure S8 shows the number of highly polluted days (daily mean $PM_1 > 50$ $\mu g/m^3$) in each year from 2014 to 2018 in China and four typical regions. Over China as a whole, average PM_1 concentrations are less than 50 $\mu g/m^3$ on most days. However, the Beijing−Tianjin−Hebei and Yangtze River Delta regions have more severe air pollution with a large number of days having concentrations exceeding 50 $\mu g/m^3$. The overall pollution level is the highest in 2014, with high-PM_1 days accounting for approximately 8, 40, 23, 12, and 14% of the total year over the whole of China and the four typical regions, respectively. The number of highly polluted days has decreased significantly over the years in China. In 2018, there were few high-PM_1 days in most areas (<5%). These results suggest that air quality has greatly improved in recent years in China. This is mainly attributed to the implementation by the Chinese government in 2013 of effective measures to reduce pollution emissions.[40,41]

Comparison with Related PM Studies. In general, compared with satellite-derived daily $PM_{2.5}$ concentrations, which can have CV-R^2 values of greater than 0.8 in China, satellite-derived daily PM_1 concentrations do not agree as well with ground measurements (CV-R^2 = 0.7−0.8). There are four possible reasons for this: (1) the larger size range compromises more aerosols as measured by AOD, which is a total measure of aerosol loading; (2) the weaker correlation between PM_1 and AOD indicates a more complex and implicit relationship; (3) given the sparse distribution of less monitoring stations, all possible surface types and atmospheric conditions across mainland China are likely not covered; and (4) fewer data samples likely affect the model training ability and overall accuracy.

Unlike $PM_{2.5}$ studies, few PM_1 studies focused on China have been conducted due to the scarcity of PM_1 data sources. PM_1 concentrations have been estimated by combining MODIS C6 DT and DB AOD products at a 10 km resolution across China (Table S5).[20] Himawari-8 hourly AOD products have also been used to estimate PM_1 concentrations in most parts of China due to lack of satellite observations over western China with an improved spatial resolution of 5 km.[15,21,22] PM_1 concentrations are mainly associated with human activities and are more severe in local urban regions. Estimations of PM_1 concentrations at coarse resolutions are thus not useful. Therefore, in this paper, 1-km-resolution PM_1 concentrations are estimated for the first time across China using the newly released MAIAC AOD products. Our new products have a much higher spatial resolution than those from previous studies, which is important when studying PM_1 in medium- and small-scale areas like urban areas.

In terms of model performance, the GAM model has an R^2 of 0.59 and an RMSE of 22.5 $\mu g/m^3$ in estimating PM_1 using MODIS AOD products in China.[20] The PM_1 concentrations are estimated from Himawari-8 AOD products in China using a PCA-GRNN model with an R^2 of 0.61, an RMSE of 22.0 $\mu g/m^3$, and an MAE of 13.8 $\mu g/m^3$, as well as some other traditional models, e.g., the multiple linear regression model (R^2 = 0.23, RMSE = 32.8 $\mu g/m^3$, and MAE = 22.5 $\mu g/m^3$), the LME model (R^2 = 0.41, RMSE = 28.7 $\mu g/m^3$, and MAE = 19.4 $\mu g/m^3$), the back propagation neural network model (R^2 =

0.52, RMSE = 26.1 $\mu g/m^3$, and MAE = 17.8 $\mu g/m^3$), and the GRNN model (R^2 = 0.61, RMSE = 24.1 $\mu g/m^3$, and MAE = 14.3 $\mu g/m^3$).[15] The PCA-GRNN model was subsequently improved by including surface $PM_{2.5}$ observations, increasing the CV-R^2 to 0.74 and decreasing the RMSE to 19.0 $\mu g/m^3$ and the MAE to 11.4 $\mu g/m^3$.[21] The two-stage model estimates PM_1 better (R^2 = 0.80, RMSE = 15.4 $\mu g/m^3$, and MAE = 9.3 $\mu g/m^3$) using Himawari-8 AOD products over central and eastern China.[22] Note that our STET model developed for China can outperform most previous models with almost all better evaluation metrics, chiefly because of (1) accounting for spatiotemporal heterogeneity by our model that has not been taken into account in other studies; (2) the use of high-quality MAIAC AOD products at 1 km resolution; and (3) overall low accuracy with large estimation uncertainty for those coarse-resolution aerosol products, especially over urban areas.[42−45] The ChinaHighPM$_1$ data set generated in this study may be most valuable for air pollution studies in urban and megacity areas where pollution is severe and more inhomogeneous in response to intensive human activities.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.9b03258.

Tables S1−S5, and Figures S1−S8 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: zli@atmos.umd.edu. Tel: (301) 405-6699.

ORCID Ⓞ

Jing Wei: 0000-0002-8803-7056
Jianping Guo: 0000-0001-8530-8976

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Dockery, D. W.; Stone, P. Cardiovascular risks from fine particulate air pollution. *N. Engl. J. Med.* **2007**, *356*, 511−513.

(2) Chen, Z.; Wang, J.; Ma, G.; Zhang, Y. China tackles the health effects of air pollution. *Lancet* **2013**, *382*, 1959−1960.

(3) Sun, L.; Wei, J.; Duan, D.; Guo, Y.; Yang, D.; Jia, C.; Mi, X. Impact of land-use and land-cover change on urban air quality in representative cities of China. *J. Atmos. Sol.-Terr. Phys.* **2016**, *142*, 43−54.

(4) Morawska, L.; Keogh, D.; Thomas, S.; Mengersen, K. Modality in ambient particle size distributions and its potential as a basis for

developing air quality regulation. *Atmos. Environ.* **2008**, *42*, 1617–1628.

(5) Meng, X.; Ma, Y.; Chen, R.; Zhou, Z.; Chen, B.; Kan, H. Size-fractionated particle number concentrations and daily mortality in a Chinese city. *Environ. Health Perspect.* **2013**, *121*, 1174–1178.

(6) Jaiprakash; Singhai, A.; Habib, G.; Raman, R.; Gupta, T. Chemical characterization of PM$_1$ aerosol in Delhi and source apportionment using positive matrix factorization. *Environ. Sci. Pollut. Res.* **2017**, *24*, 445–462.

(7) Chen, G.; Li, S.; Zhang, Y.; Zhang, W.; Guo, Y.; et al. Effects of ambient PM$_1$ air pollution on daily emergency hospital visits in China: an epidemiological study. *Lancet Planet. Health* **2017**, *1*, e221–e229.

(8) Yang, B. Y.; Guo, Y.; Morawska, L.; Bloom, M. S.; Markevych, I.; Heinrich, J.; Dharmage, S. C.; Knibbs, L. D.; Lin, S.; Yim, S.; Chen, G.; Li, S.; Zeng, X. W.; Liu, K. K.; Hu, L. W.; Dong, G. H. Ambient PM$_1$ air pollution and cardiovascular disease prevalence: insights from the 33 Communities Chinese Health Study. *Environ. Int.* **2019**, *123*, 310–317.

(9) Wang, Y.; Zhang, X.; Sun, J.; Zhang, X.; Che, H.; Li, Y. Spatial and temporal variations of the concentrations of PM$_{10}$, PM$_{2.5}$ and PM$_1$ in China. *Atmos. Chem. Phys.* **2015**, *15*, 13585–13598.

(10) Chen, G.; Morawska, L.; Zhang, W.; Li, S.; Cao, W.; Ren, H.; Wang, B.; Wang, H.; Knibbs, L.; Williams, G.; Guo, J.; Guo, Y. Spatiotemporal variation of PM$_1$ pollution in China. *Atmos. Environ.* **2018**, *178*, 198–205.

(11) Guo, J.; Zhang, X.; Che, H.; Gong, S.; An, X.; Cao, C.; Guang, J.; Zhang, H.; Wang, Y.; Zhang, X.; et al. Correlation between PM concentrations and aerosol optical depth in eastern China. *Atmos. Environ.* **2009**, *43*, 5876–5886.

(12) Ma, Z.; Hu, X.; Huang, L.; Bi, J.; Liu, Y. Estimating ground-level PM$_{2.5}$ in China using satellite remote sensing. *Environ. Sci. Technol.* **2014**, *48*, 7436–7444.

(13) Liu, Y.; Sarnat, J. A.; Kilaru, V.; Jacob, D. J.; Koutrakis, P. Estimating ground-level PM$_{2.5}$ in the eastern United States using satellite remote sensing. *Environ. Sci. Technol.* **2005**, *39*, 3269–3278.

(14) Yao, F.; Wu, J.; Li, W.; Peng, J. A spatially structured adaptive two-stage model for retrieving ground-level PM$_{2.5}$ concentrations from VIIRS AOD in China. *ISPRS J. Photogramm. Remote Sens.* **2019**, *151*, 263–276.

(15) Zang, L.; Mao, F.; Guo, J.; Gong, W.; Wang, W.; Pan, Z. Estimating hourly PM$_1$ concentrations from Himawari-8 aerosol optical depth in China. *Environ. Pollut.* **2018**, *241*, 654–663.

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

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

(18) Li, T.; Shen, H.; Yuan, Q.; Zhang, X.; Zhang, L. Estimating ground-level PM$_{2.5}$ by fusing satellite and station observations: a geo-intelligent deep learning approach. *Geophys. Res. Lett.* **2017**, *44*, 11985–11993.

(19) Wei, J.; Huang, W.; Li, Z.; Xue, W.; Peng, Y.; Sun, L.; Cribb, M. Estimating 1-km-resolution PM$_{2.5}$ concentrations across China using the space-time random forest approach. *Remote Sens. Environ.* **2019**, *231*, No. 111221.

(20) Chen, G.; Knibbs, L.; Zhang, W.; Li, S.; Cao, W.; Guo, J.; Ren, H.; Wang, B.; Wang, H.; Williams, G.; Hamm, N.; Guo, Y. Estimating spatiotemporal distribution of PM$_1$ concentrations in China with satellite remote sensing, meteorology, and land use information. *Environ. Pollut.* **2018**, *233*, 1086–1094.

(21) Zang, L.; Mao, F.; Guo, J.; Wang, W.; Pan, Z.; Shen, H.; Zhu, B.; Wang, Z. Estimation of spatiotemporal PM$_{1.0}$ distributions in China by combining PM$_{2.5}$ observations with satellite aerosol optical depth. *Sci. Total Environ.* **2019**, *658*, 1256–1264.

(22) Wang, W.; Mao, F.; Zou, B.; Guo, J.; Wu, L.; Pan, Z.; Zang, L. Two-stage model for estimating the spatiotemporal distribution of hourly PM$_{1.0}$ concentrations over central and east China. *Sci. Total Environ.* **2019**, *675*, 658–666.

(23) Grimm, H.; Eatough, D. J. Aerosol Measurement: The Use of Optical Light Scattering for the Determination of Particulate Size Distribution, and Particulate Mass, Including the Semi-Volatile Fraction. *J. Air Waste Manage. Assoc.* **2009**, *59*, 101–107.

(24) Hansen, J. C.; Woolwine, W. R.; Bates, B. L.; Clark, J. M.; Kuprov, R. Y.; Mukherjee, P.; Murray, J. A.; Simmons, M. A.; Waite, M. F.; Eatough, N. L.; Eatough, D. J.; Long, R.; Grover, B. D. Semicontinuous PM$_{2.5}$ and PM$_{10}$ mass and composition measurements in Lindon, Utah, during winter 2007. *J. Air Waste Manage. Assoc.* **2010**, *60*, 346–355.

(25) Sciare, J.; Cachier, H.; Sarda-Esteʹve, R.; Yu, T.; Wang, X. Semi-volatile aerosols in Beijing (R.P. China): Characterization and influence on various PM$_{2.5}$ measurements. *J. Geophys. Res.* **2007**, *112*, No. D18202.

(26) Zhao, X.; Zhang, X.; Pu, W.; Meng, W.; Xu, X. Scattering properties of the atmospheric aerosol in Beijing, China. *Atmos. Res.* **2011**, *101*, 799–808.

(27) Lyapustin, A.; Wang, Y.; Korkin, S.; Huang, D. MODIS Collection 6 MAIAC algorithm. *Atmos. Meas. Tech.* **2018**, *11*, 5741–5765.

(28) Wei, J.; Li, Z.; Peng, Y.; Sun, L.; Yan, X. A regionally robust high-spatial-resolution aerosol retrieval algorithm for MODIS images over Eastern China. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4748–4757.

(29) Zhang, Z.; Wu, W.; Fan, M.; Wei, J.; Tan, Y.; Wang, Q. Evaluation of MAIAC aerosol retrievals over China. *Atmos. Environ.* **2019**, *202*, 8–16.

(30) Liu, N.; Zou, B.; Feng, H.; Wang, W.; Tang, Y.; Liang, Y. Evaluation and comparison of multiangle implementation of the atmospheric correction algorithm, Dark Target, and Deep Blue aerosol products over China. *Atmos. Chem. Phys.* **2019**, *19*, 8243–8268.

(31) Dee, D. P.; Uppala, S. M.; Simmons, A. J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.; Balsamo, G.; Bauer, P.; Bechtold, P.; Beljaars, A.; van de Berg, L.; Bidlot, J.; Bormann, N.; Delsol, C.; Dragani, R.; Fuentes, M.; Geer, A.; Haimberger, L.; Healy, S.; Hersbach, H.; Hólm, V.; Isaksen, L.; Kållberg, P.; Köhler, M.; Matricardi, M.; McNally, A.; Monge-Sanz, B.; Morcrette, J.; Park, B.; Peubey, C.; de Rosnay, P.; Tavolato, C.; Thépaut, J.; Vitart, F. The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597.

(32) Zhang, Q.; Streets, D.; He, K.; Klimont, Z. Major components of China's anthropogenic primary particulate emissions. *Environ. Res. Lett.* **2007**, *2*, No. 045027.

(33) Li, M.; Zhang, Q.; Kurokawa, J.-I.; Woo, J.-H.; He, K.; Lu, Z.; Ohara, T.; Song, Y.; Streets, D. G.; Carmichael, G. R.; Cheng, Y.; Hong, C.; Huo, H.; Jiang, X.; Kang, S.; Liu, F.; Su, H.; Zheng, B. MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP. *Atmos. Chem. Phys.* **2017**, *17*, 935–963.

(34) Wei, J.; Li, Z.; Sun, L.; Peng, Y.; Wang, L. Improved merge schemes for MODIS Collection 6.1 Dark Target and Deep Blue combined aerosol products. *Atmos. Environ.* **2019**, *202*, 315–327.

(35) Fang, X.; Zou, B.; Liu, X.; Sternberg, T.; Zhai, L. Satellite-based ground PM$_{2.5}$ estimation using timely structure adaptive modeling. *Remote Sens. Environ.* **2016**, *186*, 152–163.

(36) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42.

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

(38) Wei, J.; Peng, Y.; Mahmood, R.; Sun, L.; Guo, J. Intercomparison in spatial distributions and temporal trends derived from multi-source satellite aerosol products. *Atmos. Chem. Phys.* **2019**, *19*, 7183–7207.

(39) Yu, W.; Liu, Y.; Ma, Z.; Bi, J. Improving satellite-based $PM_{2.5}$ estimates in China using Gaussian processes modeling in a Bayesian hierarchical setting. *Sci. Rep.* **2017**, *7*, No. 7048.

(40) Fang, D.; Chen, B.; Hubacek, K.; Ni, R.; Chen, L.; Feng, K.; Lin, J. Clean air for some: unintended spillover effects of regional air pollution policies. *Sci. Adv.* **2019**, *5*, 1−10.

(41) Liu, M.; Huang, X.; Song, Y.; Tang, J.; Cao, J.; Zhang, X.; Zhang, Q.; Wang, S.; Xu, T.; Kang, L.; Cai, X.; Zhang, H.; Yang, F.; Wang, H.; Yu, J.; Lau, A.; He, L.; Huang, X.; Duan, L.; Ding, A.; Xue, L.; Cao, J.; Liu, B.; Zhu, T. Ammonia emission control in China would mitigate haze pollution and nitrogen deposition, but worsen acid rain. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 7760−7765.

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

(43) Wei, J.; Li, Z.; Peng, Y.; Sun, L. MODIS Collection 6.1 aerosol optical depth products over land and ocean: validation and comparison. *Atmos. Environ.* **2019**, *201*, 428−440.

(44) Zhang, Z.; Wu, W.; Fan, M.; Tao, M.; Wei, J.; Jin, J.; Tan, Y.; Wang, Q. Validation of Himawari-8 aerosol optical depth retrievals over China. *Atmos. Environ.* **2019**, *199*, 32−44.

(45) Wei, J.; Li, Z.; Sun, L.; Peng, Y.; Zhang, Z.; Li, Z.; Su, T.; Feng, L.; Cai, Z.; Wu, H. Evaluation and uncertainty estimate of next-generation geostationary meteorological Himawari-8/AHI aerosol products. *Sci. Total Environ.* **2019**, *692*, 879−891.