Reviewer: 1

This study built a new space-time extremely randomized trees model (STET), which integrates information from satellite-based aerosol optical depth (AOD) measurements, ground-based PM2.5 observations, and other auxiliary data (e.g., meteorological data), to retrieve daily surface PM2.5 concentrations over China. The newly-developed model outperforms most of the previously reported models in capturing the spatiotemporal variations in surface PM2.5 concentrations and in finer spatial resolution. Overall, this manuscript is well organized with extensive evaluations on the model performance.
Response: We appreciate the time and effort you spent on this manuscript, and we have carefully revised our manuscript. The responses to the questions raised in your report are as follows.

There are some minor concerns that should be addressed before publication.
1. Eq. 1. It is not clear to me how the authors apply these equations. Did the authors apply the relationships between Terra- and Aqua-based AOD measurements to fill the missing AOD value for one sensor while another sensor has a valid measurement on the same day? Please clarify the usage of Eq. 1.
Response: We have replaced the regression method with the average approach according to Reviewer#2's suggestion, and we have clarified this in Section 3.1 of the revised manuscript as follows:
"Terra and Aqua MAIAC AOD retrievals are thus averaged for each pixel on each day to form a new dataset and enlarge the spatial coverage."

2. L201-202. It is possible that the limited impact of precipitation on PM2.5 estimates can be attributed to the fact that there's a high probability of missing AOD measurements on rainy days?
Response: Yes, that's the reason for the limited impact of precipitation on $PM_{2.5}$ estimates. We have added this as "This can be attributed to the high probability of missing AOD retrievals on rainy days." in Section 3.3 of the revised manuscript.

3. It is unclear to me how the authors compare monthly, seasonal, and annual mean PM2.5 retrievals with observed PM2.5 data. For example, for one grid with 100 days of valid daily PM2.5 retrieval, to compare annual mean PM2.5 retrieval with observation, did the authors calculate the corresponding 100-day mean PM2.5 observation or the 365-day mean PM2.5 observation for comparison?
Response: We compared the monthly, seasonal, and annual mean $PM_{2.5}$ retrievals with $PM_{2.5}$ observations using the same number of valid days. We have clarified this in the revised manuscript as follows:
"Synthetized $PM_{2.5}$ retrievals are validated against $PM_{2.5}$ surface observations by calculating the effective values from the same number of valid days at monthly, seasonal, and annual time scales (Figure 10)."

4. L247-248. What's the reason for the overall underestimation of PM2.5 concentration in high polluted days by the STET model?

Response: We have discussed potential reasons in Section 5.1 in the revised manuscript as follows:

"Potential causes are: 1) There are large estimation errors in AOD retrievals under severe pollution conditions in China (Wei et al., 2019c). This is further rooted to the fundamental limitations of satellite-based AOD retrievals, i.e., the non-linear to reflectance and the high sensitivity of the single-scattering albedo (Z. Li et al., 2009); 2) High AOD does not correspond to high $PM_{2.5}$ concentrations because their ratio is highly variable over space and time, affected by both natural and human factors; 3) The number of samples for high-pollution cases is small, hindering the ability to train the model."

5. L310-316. What's the possible impact of variations in the valid sample number of AOD measurement across seasons on the differences in model performance at the seasonal level?

Response: We have discussed the potential causes for the differences in the number of data samples and model performance at the seasonal level in Section 4.2.2 of the revised manuscript as follows:

"Results suggest that there are clear differences in the number of valid data samples because of the long-term snow/ice cover in winter and more frequent clouds in summer, resulting in an overall smaller number of samples than in the other two seasons. … The differences in model performance among the seasons are mainly attributed to seasonal variations in natural conditions and human activities. Meteorological conditions in summer favor the diffusion of pollutants but complicate the $PM_{2.5}$-AOD relationship (Su et al., 2018, 2020), whereas direct emissions of pollutants are greater in winter, resulting in severe air pollution."

6. L361-363. Results in this study cannot support the conclusion here (i.e., air quality improvement from clean air policies) as only one-year PM2.5 concentration data was developed. Please rephrase this sentence.

Response: We have removed this sentence from the manuscript.

7. The caption for Fig.9 is incorrect.

Response: We have corrected the caption in the revised manuscript.

8. L36. "cross-validation coefficient" is unclear here, please clarify whether it means correlation coefficient (R) or coefficient of determination (R2).

Response: We have clarified this in the revised manuscript.

9. Would suggest spelling out all statistical metrics (e.g., R2, RMSE, MAE, MRE) when you first mention them.

Response: Done.

10. Would suggest thoroughly checking the manuscript to avoid grammar errors and make the manuscript more readable.

Response: The manuscript has been more carefully edited by a native speaker.

Reviewer: 2

Using the newly-developed space-time extremely randomized trees (STET) model, this study is aimed at estimating the 1-km-resolution PM2.5 surface concentrations across China. Besides meteorology, land surface conditions and population, a space term and a time term representing the spatial autocorrelation and temporal variation of PM2.5, respectively are also included to derive the PM2.5-AOD relationship. Overall this manuscript is well written, and potentially improves our understanding regarding how to retrieve the PM2.5 concentrations from AOD products and other auxiliary data. However, before I recommend this manuscript to be published, the authors should carefully address and clarify my several comments.
Response: We appreciate the time and effort the reviewer spent on this manuscript and the insightful comments and constructive suggestions. In light of your opinion, we have carefully revised our manuscript. The responses to the questions raised in your report are as follows.

General comments:
1. The relationship between (surface layer) PM2.5 and AOD might largely depend on the compositions (including aerosol water, as Reddington et al. (2019) indicated that aerosol water uptake and hygroscopic growth would also impact the AOD), vertical profile and size distribution of PM2.5. Thus, I find that some results in Figure 2 are confusing, and needs further analysis and clarification: 1) In Section 3.2, it is unclear that how the importance scores of all selected independent variables and spatiotemporal information to PM2.5 estimates for the STET model are calculated.
Response: We agree with you and we have mentioned this in the manuscript and cited the references. In addition, the importance score is described in more detail in the revised manuscript. The importance score of each independent variable used to estimate $PM_{2.5}$ is calculated based on the Gini index (GI). We have added a more detailed description in Section 3.3 of the revised manuscript as follows:

" … the GI index is selected to calculate the importance score of each independent variable on $PM_{2.5}$ estimates because of its higher accuracy and stability as a variable importance measure, especially for continuous variables with low signal-to-noise ratios (Jiang et al., 2009; Calle and Urrea, 2011), expressed as

$$GI(\omega) = \sum_{n=1}^{N} \omega_n (1-\omega_n) = 1 - \sum_{n=1}^{N} \omega_n^2 \ , \qquad (2)$$

where n represents the number of the categories (N = 1, …, n), and ω_n represents the sample weight of each category. The importance of one feature (X_j) on node m is that the GI changes before and after node m branching:

$$\Delta GI_{jm} = GI_m - GI_l - GI_r \ , (3)$$

where GI_l and GI_r represent the GI of two new nodes after branching. The importance score for one feature (IS_j) in then the extra-trees with k trees (i = 1, …, k), calculated as

$$IS_j = \sum_{i=1}^{k} \Delta GI_{ij} = \sum_{i=1}^{k} \sum\nolimits_{m \in M} \Delta GI_{jm} \ , (4)$$

where ΔGI_{ij} represents the importance of X_i in the i^{th} tree when the node of feature X_i in decision tree j belongs to set M. Finally, an additional normalization approach is performed to all obtained importance scores for each feature."

2) Why RH turns out to be a much less important parameter, and it has an importance score that is only slightly higher than those negligible parameters do. RH is an important factor determining the aerosol compositions and water uptake, and recent air quality studies (e.g., Sun et al., 2014; Zheng et al., 2015) showed that high RH conditions facilitate rapid production of secondary PM.

Response: We agree with you that RH should have a large influence on the production of $PM_{2.5}$. However, a potential reason why RH turns out to be less important is that high RH conditions are potentially highly related to cloudy/rainy days, especially in summer, when there is a high probability of missing AOD retrievals. In addition, this importance score only represents the importance of features in splitting during the extra-tree construction, not the contribution of features to $PM_{2.5}$ in physical mechanisms. We have clarified these in in Section 3.3 of the revised manuscript as follows:

"The $PM_{2.5}$-AOD relationship might largely depend on the compositions (e.g., aerosol water, Reddington et al., 2019; Jin et al., 2020). High RH conditions and precipitation should have large influences on the production and removal of $PM_{2.5}$ (Sun et al., 2014; Zheng et al., 2015). However, RH and PRE turn to be less important with overall low importance scores in the STET model, which may be attributed to the fact that aerosol retrieval algorithms only work under cloud-free conditions when RH is relatively low. More importantly, the calculated importance score only represents the importance of features in splitting during the extra-tree construction, not the contribution of features to $PM_{2.5}$ in physical mechanisms."

3) Furthermore, the parameter of precipitation could significantly impact the removal of PM, but is negligible in the STET model. Both RH and precipitation are associated with cloud, and what is the uncertainty for the predicted $PM_{2.5}$-AOD relationship caused by the treatment of AOD data on cloudy dates?

Response: We agree with you that the precipitation should have a large influence on the removal of $PM_{2.5}$. However, it shows the lowest important score and is negligible because remote sensing aerosol retrieval algorithms cannot work when clouds are present, so there are no AOD retrievals on rainy days. Similarly, the importance score only refers to the importance of features in splitting during the extra-tree construction and not the contribution of features to $PM_{2.5}$ in physical mechanisms. We have added this description to Section 3.3 of the revised manuscript (See above comment):

2. The authors declared that STET model exhibited a strong predictive power and could be used to predict the historical PM2.5 records in the Abstract Section (in Line

39). This conclusion could be inappropriate as the authors only tested the year of 2017. Emissions were not expected to change greatly between 2017 and 2018. Actually, I doubt the applicability for the STET model. The space and time terms seem confusing to me, and the former term is represented by the geographical difference between two pixels, while the latter term is represented by the difference for a given pixel on different days in a year. I think they might be "residual terms" to implicitly resolve the "unknown parts" unexplained by other independent parameters. I mean, the authors need more independent parameters that could explicitly explain the PM2.5 compositions, vertical profile and size distribution. Why not emissions for different precursors (e.g., SO2, NOx and VOCs) as well as fine size dust are included as independent parameters?

Response: $PM_{2.5}$ changes dramatically in space, and varies over time, showing significant spatiotemporal heterogeneities and patterns. Thus, introducing the spatial and temporal terms account for the spatiotemporal autocorrelations of $PM_{2.5}$ between different points for each day and between consecutive time series at the same place. In addition, per your suggestion, we have included emissions for main precursors and fine-sized dust as independent parameters to enhance the STET model and improve the estimation of $PM_{2.5}$ in Section 3 of the revised manuscript as follows:

"Different with our previous study (Wei et al., 2019b), pollutant emissions for different precursors (including SO_2, NO_x, CO, and volatile organic compounds) and fine-sized dust are also employed to help explicitly explain the $PM_{2.5}$ composition, collected from a multi-resolution emission inventory for China (Zhang et al., 2007)."

In addition, we have updated and re-described in detail all the results in Sections 3 and 4. Results show that the model performance is overall improved.

3. Equation 1 is confusing. What is the R2 for each linear regression? Are these two linear regressions consistent with each other? Why not to average the Terra and Aqua data directly?

Response: We have replaced the regression method with the average approach per your suggestion and clarified this in the revised manuscript as follows:

"Terra and Aqua MAIAC AOD retrievals are thus averaged for each pixel on each day to form a new dataset and enlarge the spatial coverage."

4. The description for the STET method in Section 3 is not readily to understand. Please add clarification (better to include a schematic) so that ACP readers with less experiences in machine learning could generally understand the fundamentals of the STET method.

Response: We have added clarification and a schematic of the STET model in Section 3.4 of the revised manuscript as follows:

"For the enhanced STET model, all the selected independent variables are first input into the ERT model, and the random splits (S, a_i) are established according to the whole of training data samples; then totally different *K* attributes are selected randomly from all attributes according to spatial and temporal differences; then *K* random splits are generated (s_1, …, s_k), and a split (s*) is selected by calculating the score measure function, i.e., Score(s*, S); then split node (S) is completely randomly generated to establish an extra tree; last the extra tree ensemble is built using the similarity method. Detailed information on ERT algorithm can be found in Geurts et al. (2006). Figure 4 illustrates the schematic of the enhanced STET model."

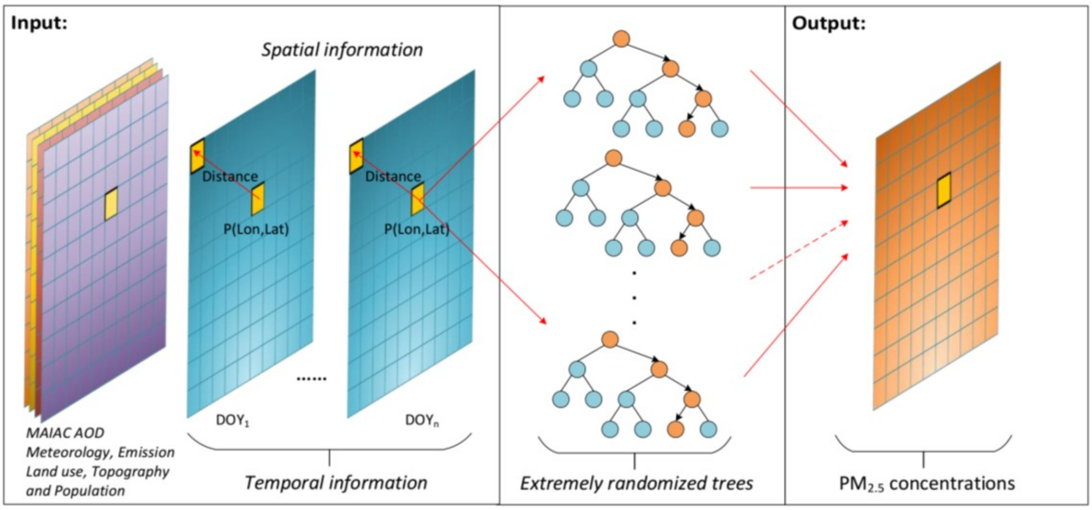


Figure 4. Schematic of the enhanced STET model developed in our study.

5. In Figure 7, what is surprising is that I see a good positive correlation pattern between R and RMSE. Generally, a good model performance is associated with a high R and a low RMSE against observations. Please check and clarify.

Response: We have verified the numbers, which are correct. Mathematically speaking, R^2 and RMSE are two independent measures of a correlation between two variables whose correlation depends on the slope of the regression between the two, higher for a regression slope closer to unity. Since the slope varies from site to site, they may not show the same spatial patterns. We have taken a closer look at the spatial patterns of these quantities and added the following text attempting to give a physical explanation (section 4.2.1 of the revised manuscript):

"In general, high R^2 with overall large RMSE but small MRE values are observed at the beginning and end of the year (in winter). This is because $PM_{2.5}$ concentrations vary more and are always high due to the greater amount of pollutant emissions caused by heating or frequent dust storms. By contrast, lower R^2 with overall small RMSE and large MRE values are observed in the middle of the year (in summer) because air pollution levels are lower."

Specific comments:

1. Line 48, the "evenly dispersed" is confusing, and is conflict with the "PM2.5 shows great spatial and temporal heterogeneities" in Line 80.

Response: Corrected.

2. Line 175, better replace "differences" by variation.

Response: Corrected.

3. Line 227, typos: Figure 2 or Figure 3?
Response: Corrected.

4. Line 247, what is definition for MAE and MRE?
Response: We have provided definitions of these evaluation indicators in the revised manuscript.

5. Figure 9, typos: the year is 2018 or 2017? Also please add the season labels for each plot.
Response: Corrected.

Reddington, C. L., Morgan, W. T., Darbyshire, E., Brito, J., Coe, H., Artaxo, P., Scott, C. E., Marsham, J., and Spracklen, D. V.: Biomass burning aerosol over the Amazon: analysis of aircraft, surface and satellite observations using a global aerosol model, Atmos. Chem. Phys., 19, 9125-9152, 10.5194/acp-19-9125-2019, 2019.

Sun, Y., Jiang, Q., Wang, Z., Fu, P., Li, J., Yang, T., and Yin, Y.: Investigation of the sources and evolution processes of severe haze pollution in Beijing in January 2013, Journal of Geophysical Research: Atmospheres, 119, 4380-4398, 2014.

Zheng, G., Duan, F., Su, H., Ma, Y., Cheng, Y., Zheng, B., Zhang, Q., Huang, T., Kimoto, T., and Chang, D.: Exploring the severe winter haze in Beijing: the impact of synoptic weather, regional transport and heterogeneous reactions, Atmos. Chem. Phys., 15, 2969-2983, 2015.

Reviewer: 3

I noticed that the same authors published a very similar paper in ES&T, https://pubs.acs.org/doi/10.1021/acs.est.9b03258. The only difference is between PM2.5 and PM1.0. However, the ACP paper needs originality.

Response: We would say that the two papers are similar but also differ in many regards that are grossly summarized as follows:

(1) They deal with different pollution quantities: PM_1 and $PM_{2.5}$, whose emission sources, formation and transport mechanisms, and health impact are all different. As such, both the figures and text of the manuscripts differ considerably. Their ratio varies greatly, ranging from less than 0.5 to greater than 0.9 at both spatial and temporal scales, especially in heavily polluted regions due to different influential factors (Wei et al., 2019b). The two papers may thus be regarded as a series of companion studies that do not undermine their respective scientific originality. The reviewer is invited to compare them to see how different they are.

(2) The estimation approaches used to derive PM_1 and $PM_{2.5}$ are similar but also differ in several aspects. While the same kind of machine learning method, namely, the space-time extra-trees (STET) model, is used for retrieving PM_1 and $PM_{2.5}$, there are numerous differences in their applications. For retrieving $PM_{2.5}$, we have 1) used different input parameters by adding the aerosol precursor gases (SO_2, CO, NO_x, VOC, fine-size dust) from pollutant emission inventories; 2) corrected the satellite retrievals of AOD with reference to ground-based measurements; 3) modified the feature selection approach using the Gini index; and 4) improved the determination of spatiotemporal information. We have clearly described these differences in Section 3 as well as in the introduction of the revised manuscript.

Moreover, the manuscript has some fatal defects, (1) It does not work well with high pollution events, which is paid more attention.

Response: Like similar studies, ours suffers from a limitation of having relatively large errors under severely polluted conditions whose causes are further explained, per the reviewer's suggestion. This is a common problem reported in many previous studies. We have added the following text to the revised manuscript (Section 5.1): "We find that all traditional statistical regression models, and machine and deep approaches reported in previous studies underestimated $PM_{2.5}$ concentrations under highly polluted conditions with poor regressions (i.e., slope < 0.9, and intercept > 6 $\mu g/m^3$) between measurements and retrievals of $PM_{2.5}$ in China, a common problem. Potential causes are: 1) There are large estimation errors in AOD retrievals under severe pollution conditions in China (Wei et al., 2019c). This is further rooted to the fundamental limitations of satellite-based AOD retrievals, i.e., the non-linear to reflectance and the high sensitivity of the single-scattering albedo (Z. Li et al., 2009); 2) High AOD does not correspond to high $PM_{2.5}$ concentrations because their ratio is highly variable over space and time, affected by both natural and human factors; 3)

The number of samples for high-pollution cases is small, hindering the ability to train the model."

It appears that all approaches suffer from this inherent limitation, which should thus not be regarded as a "fatal defect" of our study, more importantly, the comparison results suggest that our model can more accurately capture the high pollution events with a larger slope of 0.86 and a smaller intercept of 6.16 μg/m^3 with reference to other models reported from previous studies (Table 2).

(2) Such method seems falling into a dead cycle, the results were compared by the observations which were used to fit the parameters. I do not think it works with another independent database. Some similar comments were pointed by the other two reviewers.

Response: We do not think the method itself is a "dead cycle", but do make more efforts to enhance the validity and effectiveness of the validation approach. Three independent validation methods are applied, ensuring that the training and validation data are independent, as described in Section 3.5, copied below:

"Different from our previous study, three independent validation methods are performed to verify the model's ability to estimate $PM_{2.5}$ concentrations. The first independent validation method, i.e., the out-of-sample cross-validation (CV) approach, is performed by all data samples using the 10-fold CV procedure (Rodriguez et al., 2010). The data samples are divided into ten subsets randomly, and nine (one) of them are used as training (validation) data. This approach is repeated ten times, and error rates are averaged to obtain the final result. This is a common approach to evaluate the overall accuracy of a machine learning model, widely adopted in most satellite-derived PM studies (T. Li et al., 2017a, b; Ma et al., 2014, 2019; Xiao et al., 2017; He and Huang, 2018; Chen et al., 2019; Wei et al., 2019b; Xue et al., 2019; Yao et al., 2019).

The second independent validation method, i.e., out-of-station CV approach, is similar to the first one but performed using data from the monitoring stations to evaluate the spatial performance of the model. Data samples collected from different spatial points make up the training and testing data, and the relationship between spatial predictors and $PM_{2.5}$ built from the training dataset is then estimated for each testing. The third independent validation approach tests the predictive power of the model. It is performed by applying the model built for one year to predict the $PM_{2.5}$ concentrations for other years, then validating the results against the corresponding ground measurements. This approach ensures that the data samples for model training and validation are completely independent on both spatial and temporal scales."

Improved 1-km-resolution $PM_{2.5}$ estimates across China using ~~the~~enhanced space-time extremely randomized trees

Jing Wei[1,2], Zhanqing Li[2*], Maureen Cribb[2], Wei Huang[3], Wenhao Xue[1], Lin Sun[4], Jianping Guo[5], Yiran Peng[6], Jing Li[7], Alexei Lyapustin[8], Lei Liu[9], Hao Wu[1], Yimeng Song[10]

1. State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China
2. Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
3. State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing, China
4. College of Geomatics, Shandong University of Science and Technology, Qingdao, China
5. State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China
6. Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China
7. Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China
8. Laboratory for Atmospheres, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
9. College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, China
10. Department of Urban Planning and Design, Faculty of Architecture, The University of Hong Kong, Hong Kong

Correspondence to: Zhanqing Li (zli@atmos.umd.edu)

Abstract

Fine particulate matter with aerodynamic diameters ≤ 2.5 µm ($PM_{2.5}$) ~~shows~~has adverse effects on human health and the atmospheric environment. ~~Satellite-derived aerosol products have been intensively adopted in estimating~~The estimation of surface $PM_{2.5}$ concentrations~~, but~~ has made intensive use of satellite-derived aerosol products. However, most previous studies failed to monitor air pollution over small-scale areas, limited by the coarse spatial~~-~~ resolution (3–50 km) and ~~low~~the poor data~~-~~ quality of aerosol optical depth (AOD) products. ~~Therefore, a new~~Here, enhanced was the space-time extremely randomized trees (STET) model ~~is developed that integrates~~by integrating updated spatiotemporal information and additional auxiliary data to improve ~~$PM_{2.5}$ estimates at both~~the spatial resolution and overall accuracy of $PM_{2.5}$ estimates across China. To this end, the newly released ~~MODIS MAIAC~~Moderate Resolution Imaging Spectroradiometer Multi-Angle Implementation of Atmospheric Correction AOD product~~,~~, along with meteorological~~ and other auxiliary data are inputs,~~ topographical, land-use data and pollution emissions were input to the STET model~~. Daily~~, and daily 1-km $PM_{2.5}$ maps ~~in~~for 2018 across mainland China ~~are~~were produced. The STET model ~~performs~~performed well with a high out-of-sample (out-of-station) cross-validation coefficient of determination (R^2) of 0.89 (0.88), a low root-mean-square error of 10.~~35~~33 (10.~~97~~93) μg/m^3, a small mean absolute error of 6.~~71~~69 (7.~~17~~15) μg/m^3, and a small mean relative error of 21.~~37~~28 % (23.~~77%), respectively. Particularly, it can well capture the~~69%). In particular, the model captured well $PM_{2.5}$ concentrations at both regional and individual site scales. ~~In addition, it posed a strong predictive power (e.g., monthly R^2 = 0.80) and can be used to predict the historical $PM_{2.5}$ records.~~ The North China Plain, the Sichuan Basin, and Xinjiang Province always ~~are~~ featured ~~with~~ high $PM_{2.5}$ pollution levels, especially in winter. The STET model ~~outperforms~~outperformed most models presented in previous related studies~~.~~, with a strong predictive power (e.g., monthly R^2 = 0.80) which can be used to estimate historical $PM_{2.5}$ records. More importantly, ~~our~~this study provides a new approach ~~to obtain~~toward obtaining high-spatial-resolution and high-quality $PM_{2.5}$ estimates, ~~which is~~ important for air pollution studies ~~over~~focused on urban areas.

1. Introduction

Atmospheric particulate matter is a ~~relatively stable suspension system with~~general term describing all kinds of solid and liquid ~~particulate matter evenly dispersed~~particles in the atmosphere. Fine particles are those particles in ambient air with aerodynamic diameters no more than 2.5 micrometers ($PM_{2.5}$). Compared to coarser particles, $PM_{2.5}$ ~~are~~is rich in toxic and harmful substances and can directly enter the respiratory tract and alveoli of humans. Moreover, they have a long residence time and long transmission distance in the atmosphere (Aggarwal and Jain, 2015). Numerous studies have illustrated that high $PM_{2.5}$ ~~concentration~~concentrations adversely ~~affects~~affect human health (Peng et al., 2009;

Bartell et al., 2013; Chowdhury and Dey, 2016; Crippa et al., 2019; Song et al., 2019), severely impairs the atmospheric environment (Z. Li et al., 2017), and ~~even~~ significantly influences ~~the~~ cloud and precipitation systems ~~by~~through aerosol radiative and microphysical effects (Koren et al., 2014; ~~2016;~~ Seinfeld et al., 2016~~; Ceca et al., 2018~~). Silva et al. (2013) have shown that about 2.1 million people have died each year, resulting from ~~the~~ increasing $PM_{2.5}$ concentrations around the world.

Nowadays, air pollution is becoming more severe due to continuously increasing anthropogenic aerosols in developing countries, especially in China (He et al., 2011; Huang et al., 2014; M. Liu et al., 2017; Zhai et al., 2019). Fine particulate ~~matters have~~matter has become the primary pollutant in urban ~~environment~~environments, garnering much scrutiny from the public (Han et al., 2014; L. Sun et al., 2016; Wu et al., 2018). Therefore, the China Meteorological Administration ~~began to establish~~established in 2004 a ground $PM_{2.5}$ observation network to monitor the urban air quality ~~as early as 2004~~ (Guo et al., 2009), followed by a denser network established by the Chinese Ministry of Environmental Protection ~~since~~in 2013. However, station-based monitoring is largely limited by the instruments and climatic conditions and cannot completely ~~reflect~~characterize air pollution over large areas. Satellite remote sensing technology has led to a variety of operational aerosol optical depth (AOD) products ~~using mature aerosol retrieval algorithms~~ (Levy et al., 2013; Lyapustin et al., 2018), ~~which allows the~~ leading to estimates of $PM_{2.5}$ ~~estimations~~ at large ~~scale~~scales due to ~~their unanimously the~~ positive ~~relationships~~relationship between AOD and $PM_{2.5}$ concentration (Guo et al., 2017; Wei et al., 2019a).

Over the years, numerous approaches have been proposed to improve the $PM_{2.5}$-AOD relationship. Physical models typically construct physical relationships between surface particulate matter concentrations and satellite AOD products through altitude and humidity corrections (Zhang and Li, 2015). Statistical regression models, e.g., the multiple linear regression model, the linear mixed-effect model, the two-stage model, and the geographically weighted regression (GWR) model, have been widely used for applications due to their simplicity and versatility (Gupta ~~&~~and Christopher, 2009; Ma et al., 2014; Xiao et al., 2017; Yao et al., 2019). Artificial intelligence models mainly involve ~~the~~ machine learning and deep learning models, e.g., the random forest (RF; Brokamp et al., 2018; G. Chen et al., 2018; ~~Hu~~Wei et al., ~~2017~~2019a), the extreme gradient boosting model (XGBoost~~,~~; Z. Chen et al., 2019), and the back-propagation and generalized regression neural networks (BRNN and GRNN~~,~~; T. Li et al., 2017a).

~~However,~~ $PM_{2.5}$ is jointly affected by numerous factors, e.g., meteorological conditions, human activities, and topography, showing great spatial and temporal heterogeneities. This makes it difficult for ~~above~~ traditional physical and statistical regression approaches to accurately explain and construct $PM_{2.5}$-AOD relationships, leading to poor $PM_{2.5}$ estimates. Despite their stronger data mining ability, most artificial intelligence approaches have been simplistically adopted in $PM_{2.5}$ predictions, neglecting ~~their crucial~~ the spatiotemporal characteristics ~~(~~of $PM_{2.5}$ (Brokamp et al., 2018; G. Chen et al., 2018~~,~~; Z. Chen et al., 2019; ~~Hu et al., 2017;~~ Li et al., 2017a; ~~Brokamp et al., 2018;~~ Xue et al., 2019). Furthermore, deep learning is highly dependent on the ~~computer~~ performance of a computer and is less computationally efficient. ~~On the other hand~~In addition, most widely used aerosol products are generated ~~with~~at low spatial resolutions (3–50 km), ~~and thus are seriously limited~~a serious limitation for applications over small-scale regions such as urban areas.

~~Focus on these problems, to address~~To account for the spatiotemporal heterogeneity ~~and improve~~of $PM_{2.5}$ ~~estimates, a new~~, the space-time extremely randomized trees (STET) model ~~is~~ developed in our previous study for estimating PM_1 (Wei et al., 2019b) is adopted here with further refinements for improving the estimation of $PM_{2.5}$ using the high-resolution (1 km) Moderate Resolution Imaging Spectroradiometer (MODIS~~-~~) Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD product ~~at 1-km resolution associated with~~. Note that PM_1 and $PM_{2.5}$ emission sources, formation and transport mechanisms, and health impacts differ. Their spatial patterns and distributions also differ, and their particle ratio varies greatly, ranging from less than 0.5 to greater than 0.9 at both spatial and temporal scales, especially in highly polluted regions as in China (Wei et al., 2019b). The STET model has been improved by using corrected AODs, adding pollutant emissions, updating the feature selection, and improving the determination of spatiotemporal information. Based on this, spatially continuous 1-km $PM_{2.5}$ maps covering mainland China in 2018 are generated from the MODIS MAIAC AOD product at a 1-km resolution using meteorological, land-use, topographic, ~~and~~ population, and emission parameters. ~~Then the space continuous 1-km $PM_{2.5}$ maps at different temporal scales covering mainland China in 2018 are generated.~~ Section 2 describes the data sources and integration. Section 3 introduces the ~~space-time extremely randomized trees (~~enhanced STET~~)~~ model in detail, and section 4 presents the validation and comparison of our $PM_{2.5}$ estimates across China. Section 5 compares our model with those models developed in previous related studies, and Section 6 gives a summary and ~~conclusion~~conclusions.

2. Data sources

2.1 $PM_{2.5}$ ground measurements

~~In this study, the hourly~~Hourly in~~-~~ situ $PM_{2.5}$ observations at 1583 monitoring stations (Figure 1) across mainland China from 1~~,~~ January 2017 to 31~~,~~ December 2018 ~~are~~were collected, ~~and they are~~ then averaged to obtain ~~the~~ daily mean $PM_{2.5}$ measurements. ~~The~~ $PM_{2.5}$ observations are measured using the tapered element oscillating microbalance approach~~ method~~ or β-attenuation monitors that have undergone further calibration and strict quality control procedures (Guo et al., 2009).

2.2 MAIAC AOD product

The MAIAC algorithm was developed ~~and applied~~ to generate MODIS aerosol products from the darkest to the brightest surfaces at a 1-km spatial resolution over land (Lyapustin et al., 2011). On 30 May 2018, official 1-km-resolution MAIAC aerosol products were released and made freely available to all users. This dataset is produced using the revised MAIAC algorithm with continuous improvements in scale transition using spectral regression coefficients, cloud detection, determination of aerosol models, over-water processing, and general optimization in the global aerosol retrieval process (Lyapustin et al., 2018). MAIAC daily aerosol products from the Terra and Aqua satellites ~~are~~were collected ~~in~~from 2017 to 2018 across China, and~~ the~~ 550-nm AOD retrievals with high quality assurance ($QA_{CloudMask}$ = Clear and $QA_{AdjacencyMask}$ = Clear) ~~are~~were used.

Here, the MAIAC AOD retrievals were first evaluated against surface observations at 18 AERONET monitoring stations in China (Figure 1) using the spatiotemporal matching approach (Wei et al., 2019c). MAIAC AOD retrievals are highly accurate with small estimation errors across mainland China. More than 84% of the matchups satisfy the MODIS expected error (Levy et al., 2013) at the national scale (Figure 2a). Besides vegetated surfaces, e.g., cropland and grassland, the MAIAC algorithm shows considerable accuracy over heterogeneous urban surfaces (Figure 2b). MAIAC AOD products are more accurate and less biased than the widely used Dark Target (DT) and Deep Blue products at coarse spatial resolutions (N. Liu et al., 2019; Wei et al., 2018, 2019d; Tao et al., 2019; Z. Zhang et al., 2019). More importantly, the DT algorithm generates a large number of missing values over bright surfaces, and aerosol loadings are significantly overestimated over heterogeneous urban surfaces (Levy et al., 2013; Wei et al., 2018, 2019d). Therefore, higher data-quality and spatial-resolution MAIAC products, which can generate more accurate and detailed $PM_{2.5}$ estimates, are selected.

2.3 Auxiliary data

~~The auxiliary~~Auxiliary data ~~mainly includes~~include meteorological, land-cover, surface topographic, and population data. The meteorological variables are collected from ERA-Interim atmospheric reanalysis products, including the boundary layer height (BLH), evaporation (EP), temperature (TEM), precipitation (PRE), relative humidity (RH), surface pressure (SP), wind speed (WS), and wind direction (WD). ~~For~~Observations of meteorological variables~~, the observations~~ made between 1000 to 1400 local time are averaged to be consistent with satellite overpass times. ~~The land~~Land-cover data include the MODIS land use cover and normalized difference vegetation index (NDVI) products. ~~The topographic~~Topographic data~~ include~~, i.e., the surface elevation, slope, aspect, and relief (Wei et al., ~~2019a~~2019e), are calculated from the ~~SRTM~~ Shuttle Radar Topography Mission Digital Elevation Model (DEM) product, and the population ~~derived~~data are from ~~VIIRS~~Visible Infrared Imaging Radiometer Suite nighttime lights ~~data.~~(NTL) data. Different with our previous study (Wei et al., 2019b), pollutant emissions for different precursors (including SO_2, NO_x, CO, and volatile organic compounds) and fine-sized dust are also employed to help explicitly explain the $PM_{2.5}$ composition, collected from a multi-resolution emission inventory for China (Zhang et al., 2007). Table 1 provides detailed information about the data sources.

3. Methodology

Here, a tree-based ensemble learning approach, called the extremely randomized trees (ERT; Geurts et al., 2006), is selected to deal with complex supervised regression issues and to construct robust $PM_{2.5}$-

AOD relationships. This model splits nodes by randomly selecting cut-points and uses all training samples to grow trees instead of the bootstrap approach. The model efficiently solves variance problems and mines more valuable information compared to other widely used tree-based approaches, e.g., the decision tree and RF.

Unlike the STET model used in our previous study for retrieving PM_1 (Wei et al., 2019b), the current algorithm for retrieving $PM_{2.5}$ is partly based on the STET model that is enhanced by a series of refinements to further optimize and strengthen the model capacity to improve the estimation accuracy, including 1) using aerosol precursor gases (SO_2, CO, NO_x, VOC, fine-sized dust) from pollutant emission inventories as additional input; 2) correcting satellite retrievals of AOD with reference to ground-based measurements; 3) modifying the feature selection approach using the Gini index (GI); and 4) improving the determination of spatiotemporal information.

~~2.4~~3.1 Data correction and integration

Although the MAIAC algorithm performs generally well in China with a mean absolute error (MAE) of 0.06 and a root-mean-square error (RMSE) of 0.121 (Figure 2), a systematic error in the AOD retrievals (τ_s) can be corrected by linear regression between in situ AOD measurements collected at all AERONET sites in China matched with the MAIAC retrievals as follows:

$$\tau = 0.911 \cdot \tau_s + 0.018;\ R = 0.963. \quad (1)$$

Due to the difference in cloud distributions at their respective imaging times, the spatial coverages of Terra and Aqua MAIAC AOD products ~~have different spatial coverages due to frequent clouds and difference in their respective imaging times. Therefore, both~~differ. Terra and Aqua MAIAC ~~datasets~~AOD retrievals are ~~combined and merged through the linear regression approach (Eq. 1) to reduce the systematic differences~~thus averaged for each pixel on each day to form a new dataset and enlarge the spatial coverage. By integrating the two datasets, the spatial coverage ~~is greatly~~ increased by more than 15% over most areas ~~across~~in China, ~~which can lead~~leading to $PM_{2.5}$ maps with wider spatial ~~coverage $PM_{2.5}$ maps. More importantly, the~~ coverages. The number of valid data samples ~~has~~also significantly increased by approximately 25–32~~% after combination than just using Terra or Aqua MAIAC products, which can improve the~~ %, improving the model training ability.

$$\cancel{\begin{cases} \tau_T = k_1 \cdot \tau_A + b_1 \\ \tau_A = k_2 \cdot \tau_T + b_2 \\ \tau_C = \text{mean}(\tau_T, \tau_A) \end{cases} \quad (1)}$$

~~where τ_T, τ_A, and τ_C denote the Terra, Aqua, and combined AODs.~~

~~In addition, due~~Due to different spatial resolutions, all ~~the 16~~ auxiliary variables ~~are~~were uniformly aggregated to a 1-km ~~(≈ 0.01° × 0.01°)~~ spatial resolution using the bilinear interpolation approach. After removing invalid or unrealistic values, there are 167,716 matched $PM_{2.5}$-AOD samples and independent variables ~~are~~ collected for 2018 in China.

3.2 Potential effects of variables on $PM_{2.5}$

The potential relationships between all selected independent variables and $PM_{2.5}$ measurements are first investigated (Figure 3). AOD is highly positively related to $PM_{2.5}$ measurements (R = 0.54), and all pollutant emissions, nighttime lights, and land use cover show positive effects on $PM_{2.5}$. By contrast, all topographical variables and NDVI are negatively related to $PM_{2.5}$. Moreover, except for ET (R = 0.24) and SP (R = 0.16), the other meteorological variables show opposite negative effects on $PM_{2.5}$, especially for BLH (R = -0.22) and TEM (R = -0.17). In general, all the selected variables are significantly correlated to $PM_{2.5}$ measurements at the confidence level of 0.01 or 0.05 (two sides), so they are used as inputs to the STET model for preliminary training.

3.3 Updated feature selection

Due to the large number of independent variables considered, ~~this will lead to the unavoidable~~ over-fitting ~~issue~~will occur during the model training process. ~~Therefore, the~~The model ~~need be~~thus needs further ~~adjusted~~adjustment by selecting ~~more~~the most important variables rather than all variables to overcome this issue and improve the model efficiency. ~~For this purpose,~~Instead of using the default out-of-bag error rate (Wei et al., 2019b), the GI index is selected to calculate the importance ~~scores~~score of ~~all selected~~each independent ~~variables and spatiotemporal information to~~ variable on $PM_{2.5}$ estimates because of its higher accuracy and stability as a variable importance measure, especially for ~~the STET model are~~continuous variables with low signal-to-noise ratios (Jiang et al., 2009; Calle and Urrea, 2011), expressed as

$$GI(\omega) = \sum_{n=1}^{N} \omega_n(1-\omega_n) = 1 - \sum_{n=1}^{N}\omega_n^2 \ , \qquad (2)$$

where n represents the number of the categories (N = 1, …, n), and ω_n represents the sample weight of each category. The importance of one feature (X_j) on node m is that the GI changes before and after node m branching:

$$\Delta GI_{jm} = GI_m - GI_l - GI_r \ , (3)$$

where GI_l and GI_r represent the GI of two new nodes after branching. The importance score for one feature (IS_j) in then the extra-trees with k trees (i = 1, …, k), calculated ~~in China (Figure 2).~~ as

$$IS_j = \sum_{i=1}^{k} \Delta GI_{ij} = \sum_{i=1}^{k}\sum\nolimits_{m\in M} \Delta GI_{jm} \ , (4)$$

where ΔGI_{ij} represents the importance of X_i in the i^{th} tree when the node of feature X_i in decision tree j belongs to set M. Finally, an additional normalization approach is performed to all obtained importance scores for each feature.

The results suggest that AOD is the most influential variable, contributing ∼~~31~~32.5% toward daily $PM_{2.5}$ estimates~~. Because there is little precipitation on most days throughout the year, PRE contributes little to PM estimates, by contrast, most other~~ (Figure 3). Most meteorological variables contribute more to $PM_{2.5}$ estimates, especially BLH, EP, and TEM, with an average importance score of 9~~%, 8%, and~~.6%, 7.7%, and 7.3%, respectively. The $PM_{2.5}$-AOD relationship might largely depend on the compositions (e.g., aerosol water, Reddington et al., 2019; Jin et al., 2020). High RH conditions and precipitation should have large influences on the production and removal of $PM_{2.5}$ (Sun et al., 2014; Zheng et al., 2015). However, RH and PRE turn to be less important with overall low importance scores in the STET model, which may be attributed to the fact that aerosol retrieval algorithms only work under cloud-free conditions when RH is relatively low. More importantly, the calculated importance score only represents the importance of features in splitting during the extra-tree construction, not the contribution of features to $PM_{2.5}$ in physical mechanisms. Two main land-use variables, i.e., NDVI and DEM, are also important to $PM_{2.5}$ estimates, while the pollutant emissions show different effects on $PM_{2.5}$ with varying importance scores, especially for NH_3, CO, SO_2, and fine-sized dust. The eight least important variables with low important scores of < 2% are excluded from the STET model, and the remaining 14 more important variables are selected as inputs to build the $PM_{2.5}$-AOD relationship.

3.4 Improved spatiotemporal information

Spatiotemporal heterogeneities, i.e., strong spatial autocorrelations and clear temporal variations, are the key characteristics of $PM_{2.5}$, presenting great challenges and usually neglected in most regression and artificial intelligence models. Therefore, in this study, the STET model is further enhanced to solve this problem by more accurately determining the spatial and temporal information. For this purpose, the Haversine approach is selected to calculate the great-circle distance between two points on a sphere specified by their latitudes and longitudes (Eqs. 5–6). This approach can avoid the problem of insufficient effective numbers due to the short distance between two points by using sines, used to represent the space term (P_S). In addition, instead of using the day of the year (DOY), the time radian difference for each point on different days in a year is calculated (Eq.8) to minimize the impact of the seasonal cycle and is selected to represent the time term (P_T). These two improved space-time terms can account for the spatiotemporal autocorrelations of $PM_{2.5}$ between different points for each day and between consecutive time series at the same place.

$$h = f(Lon_{i,j,t}, Lat_{i,j,t}) = haversin(\alpha_1 - \alpha_2) + \cos(\alpha_1)\cos(\alpha_2)\, haversin(\beta_1 - \beta_2), \quad (5)$$

$$haversin(\theta) = sin^2(\theta/2) = [1 - \cos(\theta)]/2, \quad (6)$$

$$P_{S(i,j,t)} = 2 * r * \text{asin}\,(sqrt(h)), \quad (7)$$

$$P_{T(i,j,t)} = \cos\,(2\pi \frac{d_{i,j,t}}{T}), \quad (8)$$

where α_1 and α_2 denote the latitudes of two points, β_1 and β_2 denote the longitudes of two points in space, *r* denotes the radius (in km) of the earth, *d* represents the DOY, and *T* represents the total number of days in the year in question.

For the enhanced STET model, all the selected independent variables are first input into the ERT model, and the random splits (S, a_i) are established according to the whole of training data samples; then totally different *K* attributes are selected randomly from all attributes according to spatial and temporal differences; then *K* random splits are generated (s_1, …, s_k), and a split (s*) is selected by calculating the score measure function, i.e., Score(s*, S); then split node (S) is completely randomly generated to establish an extra tree; last the extra tree ensemble is built using the similarity method. Detailed information on ERT algorithm can be found in Geurts et al. (2006). Figure 4 illustrates the schematic of the enhanced STET model. Figure 4 illustrates the schematic of the enhanced STET model.

~~2.5~~3.5 Model validation approach

~~In this study, the widely used 10-fold~~ Different from our previous study, three independent validation methods are performed to verify the model's ability to estimate $PM_{2.5}$ concentrations. The first independent validation method, i.e., the out-of-sample cross-validation (~~10-CV)~~ CV) approach, is performed by all data samples using the 10-fold CV procedure (Rodriguez et al., 2010~~) is selected for model validation, where all~~). The data samples are divided into ten subsets randomly, and nine (one) of them are used as ~~the training data and the remaining is the testing data, indicating that the~~ training ~~and testing~~(validation) data ~~are totally independent.~~. This approach is repeated ~~in turn for~~ ten times~~. Then the~~, and error ~~rate of each test is calculated, and the mean error rate from ten tests determines~~rates are averaged to obtain the final result. ~~Here, the out-of-sample and out-of-station 10-CV procedures are involved, which the former one is performed based on the observations and used~~This is a common approach to evaluate the overall accuracy of a machine learning model, widely adopted in most satellite-derived PM studies (T. Li et al., 2017a, b; Ma et al., 2014, 2019; Xiao et al., 2017; He and Huang, 2018; Chen et al., 2019; Wei et al., 2019b; Xue et al., 2019; Yao et al., 2019).

The second independent validation method, i.e., out-of-station CV approach, is similar to the first one ~~is~~but performed ~~based on~~using data from the monitoring stations ~~and used~~ to evaluate the ~~model~~ spatial performance~~. This means that~~ of the model. Data samples collected from different spatial points make up the training and testing ~~are made of different spatial points,~~ data, and the relationship between spatial predictors and $PM_{2.5}$ ~~concentrations estimated in the training dataset is then predicted on the testing.~~built from the training dataset is then estimated for each testing. The third independent validation approach tests the predictive power of the model. It is performed by applying the model built for one year to predict the $PM_{2.5}$ concentrations for other years, then validating the results against the corresponding ground measurements. This approach ensures that the data samples for model training and validation are completely independent on both spatial and temporal scales. Several traditional statistical metrics are selected to describe the model performance, including the correlation coefficient (R), R^2, RMSE, MAE, and the mean relative error (MRE).

4. Results

~~2.6~~4.1 Validation at the ~~national~~spatial scale

4.1.1 National-scale validation

Figure ~~4~~5 shows the out-of-sample~~-based~~ sample and out-of-station~~-based~~ 10-CV results of daily $PM_{2.5}$ estimates for the traditional ~~ET~~ERT model and our ~~new developed~~enhanced STET model at the national scale in 2018. The ~~results suggest that the~~ original ~~ET~~ERT model works well in estimating $PM_{2.5}$ concentrations with an average out-of-sample CV-R^2~~,~~ of 0.84 and overall small estimation uncertainties. However, when ~~consider the~~considering spatiotemporal information, the model performance ~~has been~~ significantly ~~improved~~improves with ~~an increasing~~a sample-based CV-R^2 ~~equal to~~of 0.89, a stronger regression line ~~(e.g., slope = 0.86),~~, and a decreasing RMSE ~~(~12.46~~of 10.33 μg/m^3~~)~~, MAE ~~(~8.26~~of 6.69 μg/m^3~~)~~, and MRE ~~(~~~of 21.28%. Regarding the spatial performance, compared to the original ET model, the enhanced STET model shows a stronger spatial predictive power with a higher out-of-station CV-R^2 of 0.88, a lower RMSE of 10.~~97~~93 μg/m^3, MAE of 7.~~17~~15 μg/m^3, and MRE of 23.~~77%. These~~69%. In addition, compared to the sample-based validation, the out-of-station accuracy changes little, suggesting that the enhanced STET model can well estimate daily $PM_{2.5}$ concentrations. Moreover, these results illustrate that spatiotemporal information ~~are~~is crucial in improving ~~the~~ $PM_{2.5}$-AOD relationships and should be carefully considered when introducing statistical regression models using remote sensing techniques.

4.1.2 Regional-scale validation

Figure 6 shows the sample-based 10-CV results of the enhanced STET model in $PM_{2.5}$ daily estimates over eastern and western China (according to the widely used Heihe-Tengchong line), and four typical ~~local~~ regions (Figure 1). The enhanced STET model performs differently over eastern and western China, mainly due to significant differences in land cover and climate conditions. There are 1289 uniformly distributed $PM_{2.5}$ stations in eastern China, and 127,241 daily samples were collected. The ~~STET~~ model performs well in eastern China with a high sample-based CV-R^2 equal to 0.90 and low estimation uncertainties, i.e., RMSE = 9.~~77~~72 $\mu g/m^3$, MAE = 6.~~44~~41 $\mu g/m^3$, and MRE = 19.~~24~~16%. By contrast, there are 294 unevenly and sparsely distributed $PM_{2.5}$ stations in western China, ~~thus~~with about three times fewer daily $PM_{2.5}$ estimates ~~were~~ collected. The model performance is overall poorer (e.g., CV-R^2 = 0.~~86, and~~85, RMSE = ~~11.99~~12.04 $\mu g/m^3$, MAE = 7.56 $\mu g/m^3$) than over eastern China. This is mainly ~~contributed~~attributed to brighter surfaces (e.g., desert and bare land) with little vegetation ~~coverage~~ and harsh meteorological conditions over western China.

There were 33,733, 15,199, 6,209, and 6,470 daily samples collected from 233, 184, 95, and 107 uniformly distributed $PM_{2.5}$ monitoring stations in the North China Plain (NCP), the Yangtze River Delta (YRD), the Pearl River Delta (PRD~~)~~), and the Sichuan Basin (SCB), respectively. ~~For former three~~Estimated $PM_{2.5}$ concentrations in the typical urban agglomerations ~~where people closely concerned, the estimated $PM_{2.5}$ concentrations~~ of the NCP, YRD, and PRD are highly consistent with surface measurements (CV-R^2 = 0.~~89~~86–0.92~~)~~), with overall low estimation uncertainties (i.e., RMSE = ~~7~~8–12 $\mu g/m^3$, MAE = 5–8 $\mu g/m^3$, and MRE = 15–19%). ~~In addition, the STET~~The new model also performs well over the Sichuan Basin with an average CV-R^2 value equal to 0.87 and comparable estimation uncertainties to ~~North China Plain. In general~~those from the NCP. Overall, despite some differences in model performance, the enhanced STET model shows an overall good ability in estimating $PM_{2.5}$ ~~estimates~~concentrations at the regional scale.

4.1.3 Site-scale validation

National- and regional-scale aggregated evaluations mainly illustrate the overall performance of the ~~STET~~ model in estimating $PM_{2.5}$ ~~estimates, however~~concentrations. However, due to the inhomogeneity of $PM_{2.5}$ monitoring stations, an additional validation for each monitoring station in China is performed (Figure ~~6~~7). For statistical significance, plotted are only these monitoring stations with more than ten data samples ~~are plotted. The daily~~. Daily $PM_{2.5}$ ~~estimations are~~estimates relate well ~~related~~ to surface measurements at most individual stations across China. The average sample-based CV-R^2 is 0.84, and ~~the~~ CV-R^2 values are ~~higher~~greater than 0.8 at more than 73% of the monitoring stations, especially

~~for~~in eastern China. However, observed are relatively poorer performances (CV-R^2 < 0.6) ~~are observed~~ at some scattered sites located in ~~southwestern~~southwest and ~~southeastern~~southeast China. In general, the ~~STET~~new model shows overall low estimation uncertainties at most sites with average RMSE and MAE values of 9.~~3~~2 and 6.5 μg/m^3, especially ~~for~~in southern China. Moreover, ~~the average RMSE and MAE values are < 10 μg/m^3 at more than 68% and 93~~~94% of the monitoring stations ~~across China.~~in China have mean RMSE and MAE values less than 15 μg/m^3 and 10 μg/m^3, respectively. Note that these stations ~~show~~have larger RMSE values (> 10 μg/m^3) in central China, mainly due to the high ~~polluted~~pollution levels. ~~In addition, the~~The average MRE value in China is 20.~~88~~8%, and most stations (> 86~~%)~~% of them) have ~~low~~ MRE values ~~<~~less than 30~~% in $PM_{2.5}$ estimations in China,~~%, especially ~~for those~~at sites located in eastern and southern China.

4.2 Performance at the temporal scale

4.2.1 Daily-scale validation

Figure ~~7~~8 shows the ~~STET~~ model performance from all available monitoring stations in China as a function of the ~~day of year~~DOY. The number of data samples in one day ranges from 54 to 1155, with an average of 466 in 2018. In general, the ~~STET~~new model ~~shows great performance~~performs well (average CV-R^2 = 0.~~76) at~~77) on most days in the year, and more than ~~76~~77% of ~~the~~these days have CV-R^2 values greater than 0.7. Two main uncertainty metrics, i.e., RMSE and MAE, show similar temporal variations during the year, first decreasing until around day 250, then gradually increasing. ~~In general, approximately equal~~ Approximately 91% and 92% of the days have low RMSE and MAE values of less than 15 and 10 μg/m^3, respectively, over the year. ~~Large estimation uncertainties always occur at the beginning and end of the year mainly due to intense human activities and harsh natural environment. Furthermore,~~ MRE is relatively stable, ranging from 13% to ~~52~~49% with an average value of 23.~~29~~2%, and more than 87% of the days ~~yield low~~have MRE values of less than 30% in China. ~~These results illustrate that the STET model show great performance in capturing~~In general, high R^2 with overall large RMSE but small MRE values are observed at the beginning and end of the year (in winter). This is because $PM_{2.5}$ concentrations ~~on most days of the year.~~vary more and are always high due to the greater amount of pollutant emissions caused by heating or frequent dust storms. By contrast, lower R^2 with overall small RMSE and large MRE values are observed in the middle of the year (in summer) because air pollution levels are lower. Nevertheless, these results illustrate that the enhanced STET model captures well $PM_{2.5}$ concentrations on most days of the year.

4.2.2 Seasonal-scale validation

Figure 9 shows sample-based ~~cross-validation~~CV results for $PM_{2.5}$ daily estimates ~~divided by four seasons~~according to the season in 2018 ~~across~~in China. ~~The results~~Results suggest that there are ~~obvious~~clear differences in ~~model performance at the seasonal level. The~~the number of valid data samples because of the long-term snow/ice cover in winter and more frequent clouds in summer, resulting in an overall smaller number of samples than in the other two seasons. The enhanced STET model performs best in autumn with the highest CV-R^2 value of 0.90 and the strongest regression line (i.e., slope = 0.88, and intercept = 4.~~88~~85 μg/m^3). ~~The average~~Mean RMSE, MAE, and MRE values in autumn are ~~9.01~~8.97 μg/m^3, 5.~~87~~84 μg/m^3, and 21.~~10~~ 02%, respectively. By contrast, the ~~STET~~new model performs the worst in summer with the lowest CV-R^2 of 0.~~76~~79 and ~~smallest~~a less steep slope of ~~0.74~~7.37, indicating ~~obvious~~clear underestimations. However, summer ~~shows~~experiences the least amount of air pollution with most daily $PM_{2.5}$ values < ~~80~~50 μg/m^3, leading to ~~smallest estimation uncertainties. The main reason is that the meteorological conditions in place in summer accelerated the diffusion of pollutants but complicated the $PM_{2.5}$-AOD relationships. The air~~the smallest RMSE and MAE values but the largest MRE values. Air quality is about two or three times worse in spring and winter ~~than in winter~~ with wider $PM_{2.5}$ ranges and larger standard deviations. ~~Moreover, the STET~~The model ~~shows~~performance in these seasons is similar ~~performances in these two seasonal~~, with almost equal CV-R^2 and slope values, ~~as well as~~and close estimation uncertainties. The differences in model performance among the seasons are mainly attributed to seasonal variations in natural conditions and human activities. Meteorological conditions in summer favor the diffusion of pollutants but complicate the $PM_{2.5}$-AOD relationship (Su et al., 2018, 2020), whereas direct emissions of pollutants are greater in winter, resulting in severe air pollution.

4.2.3 Synthetic-scale validation

Synthetized $PM_{2.5}$ retrievals are validated against $PM_{2.5}$ surface observations by calculating the effective values from the same number of valid days at monthly, seasonal, and annual time scales (Figure 10). Monthly $PM_{2.5}$ estimates and ground measurements (N = 12,410) are highly correlated (R^2 = 0.93), with a steep slope of 0.91. Mean RMSE, MAE, and MRE values are 5.63 μg/m^3, 4.08 μg/m^3, and 11.59%, respectively. Seasonal mean $PM_{2.5}$ estimates (N = 5,231) have a good accuracy (i.e., R^2 = 0.93, RMSE = 5.00 μg/m^3, MAE = 3.69 μg/m^3, and MRE = 10.31%). Annual mean $PM_{2.5}$ estimates (N = 1,462) agree well with ground measurements (R = 0.91), with small uncertainties (i.e., RMSE = 4.11 μg/m^3, MAE = 3.12 μg/m^3, and MPE = 8.58%). This illustrates that the synthetic dataset can more accurately reflect the spatiotemporal $PM_{2.5}$ loadings and variations across China.

~~2.7~~4.3 Predicted $PM_{2.5}$ maps across China

~~The monthly~~Monthly $PM_{2.5}$ maps are thus synthesized and averaged from at least 20% of available daily $PM_{2.5}$ estimates for each grid in a month~~ in 2018 ~~, and annual $PM_{2.5}$ maps are generated from monthly $PM_{2.5}$ maps if there are more than eight available values for each grid across China (Hsu et al., 2012; Wei et al., 2019f). The spatial coverage of monthly $PM_{2.5}$ maps varies from 73% to 92%, with an average of 83% across mainland China. The ~~highest (lowest) spatial~~maximum coverage ~~occcurs around October (~~occurs in April, and the minimum coverage occurs in January~~) of the year. Similarly, the~~. The monthly mean $PM_{2.5}$ values vary conversely from ~~21.2~~24.4 μg/m^3 to ~~45.1~~42.9 μg/m^3~~ with~~, where the highest (lowest) $PM_{2.5}$ concentration ~~occurring around March~~is observed in December (August) of the year.

The satellite-derived 1-km-resolution $PM_{2.5}$ map in 2018 covers almost the full scene (spatial coverage = 99%) across mainland China (Figure 11a) and is highly consistent in spatial ~~patterns are similar between the STET-derived 1-km $PM_{2.5}$ map and calculated in~~ pattern with the corresponding in situ measurements (Figure 11b). The average $PM_{2.5}$ concentration is 32.7±13.6 μg/m^3 in 2018 across mainland China. In general, the most severe $PM_{2.5}$ pollution occurs in the Taklamakan Deseret, where most areas ~~expose~~ are exposed to high $PM_{2.5}$ concentrations of > 80 μg/m^3. There are also high~~-polluted~~ pollution levels over the ~~North China Plain, Sichuan Basin, and Yangtze River Delta~~NCP, the SB, and the YRD, with annual mean $PM_{2.5}$ values of 46.~~8±11.8, 38.3~~7±10.~~3~~5, 39.8±9.9, and ~~37.6±9~~38.4±8.3

μg/m^3, respectively~~. These mainly contributed to~~, arising from intensive human activities, and special topographic and meteorological conditions. By contrast, the annual mean $PM_{2.5}$ ~~loadings are~~loading is overall low ~~in~~over the rest ~~areas~~ of China, e.g., the PRD (33.4±3.9 μg/m^3). However, there may be poor representativeness for ~~these~~ areas ~~over~~in western China with few ground monitoring stations. ~~In general, we have to say that the $PM_{2.5}$ pollution has been significantly reduced in 2018 across China due to the effective emission control measures implemented by the Chinese government (Fang et al., 2019; Ma et al., 2019). However, more~~More than ~~30~~34% of mainland China ~~still~~ experienced high $PM_{2.5}$ levels in 2018 exceeding the international and national recommended air quality level ($PM_{2.5}$ > 35 μg/m^3).

Figure 12 shows seasonal mean $PM_{2.5}$ maps, ~~which are~~ averaged from ~~the~~ available monthly values for each grid, in 2018 across China. The average $PM_{2.5}$ concentration (spatial coverage) is 37.2±20.7 μg/m^3 (~ 96%), 25.5±12.1 μg/m^3 (~ 92%), 29.5±11.5 μg/m^3 (~ 97%), and 41.3±15.4 μg/m^3 (~ 88%) for spring, summer, autumn, and winter, respectively. There are noticeable spatial differences in $PM_{2.5}$ distributions on the seasonal scale. In winter and spring, more than ~~77~~49% and ~~66~~42% of mainland China ~~exposing the~~were exposed to high $PM_{2.5}$ levels > of 30 μg/m^3, ~~yielding poorer air~~resulting in poor quality. By contrast, $PM_{2.5}$ pollution is ~~slighter~~lower in summer and autumn, with more than ~~91~~90% and ~~81~~74% of mainland China, respectively, experiencing ~~low~~ $PM_{2.5}$ levels below the acceptable air quality level. Note that in spring, $PM_{2.5}$ concentrations are particularly high in Xinjiang province due to frequent sand and dust episodes in 2018.

5. Discussion

5.1 Model accuracy

There is an increasing number of studies on estimating $PM_{2.5}$ using satellite AOD products from local to national scales across China. However, limited by the operational satellite aerosol products, $PM_{2.5}$ can only be estimated at coarse spatial resolutions of approximately 6–10 km (Fang et al., 2016; T. Li et al., 2017b; Yu et al., 2017; Chen et al., 2018; Ma et al., 2019; Yao et al., 2019). Recently, with the release of MODIS 3-km DT aerosol products, ~~the~~ $PM_{2.5}$ estimates can be improved to a 3-km spatial resolution across China (You et al., 2016; T. Li et al., 2017a; He ~~&~~and Huang, 2018; Chen et al., 2019; Xue et al., 2019). ~~Therefore, in our~~This study~~,~~ improves the spatial resolution of $PM_{2.5}$ estimates ~~has been~~

~~significantly improved by 3–10 times~~across mainland China to 1 km based on the newly released high-quality MAIAC products~~ across mainland China~~.

~~For~~Regarding model performance, our newly developed STET model ~~shows much higher accuracy~~is more accurate with higher CV-R^2 values, and smaller RMSE and MAE values than ~~the~~those from statistical regression models (Table 2), e.g., the timely structure adaptive model (TSAM~~,~~; Fang et al., 2016~~) model, the Gaussian model (Yu et al., 2017~~), the Generalized Additive Model (GAM~~,~~; Chen et al., 2018) model, ~~and~~ the GWR model (Ma et al., 2014; You et al., 2016), and the geographically and temporally weighted regression model (GTWR~~ model (~~; He and Huang, 2018). The enhanced STET model can also outperform most machine learning (ML) and deep learning approaches including the ~~RF~~Gaussian model (Yu et al., 2017), the Random Forest model (Chen et al., 2018; Wei et al., 2019e), the XGBoost model (Chen et al., 2019), the ~~Geo- BPNN,~~ GRNN and deep brief network (DBN) models (T. Li et al., 2017a, ~~2017b~~b), and some optical combined models, e.g., the Daily-GWR model (D-GWR~~)~~ ~~model (~~; He and Huang, 2018), the two-stage model (He and Huang, 2018; Ma et al., 2019; Yao et al., 2019), and the ML + GAM model (Xue et al., 2019).

We find that all traditional statistical regression models, and machine and deep approaches reported in previous studies underestimated $PM_{2.5}$ concentrations under highly polluted conditions with poor regressions (i.e., slope < 0.9 and intercept > 6 μg/m^3) between measurements and retrievals of $PM_{2.5}$ in China, a common problem. Potential causes are: 1) There are large estimation errors in AOD retrievals under severe pollution conditions in China (Wei et al., 2019c). This is further rooted to the fundamental limitations of satellite-based AOD retrievals, i.e., the non-linear to reflectance and the high sensitivity of the single-scattering albedo (Z. Li et al., 2009); 2) High AOD does not correspond to high $PM_{2.5}$ concentrations because their ratio is highly variable over space and time, affected by both natural and human factors; 3) The number of samples for high-pollution cases is small, hindering the ability to train the model. Therefore, our model also tends to underestimate $PM_{2.5}$ concentrations on highly polluted days ($PM_{2.5}$ > 150 μg/m^3), however, it can more accurately capture the high pollution events with a stronger slope of 0.86 and a smaller intercept of 6.16 μg/m^3 with reference to other models reported from previous studies (Table 2).

Furthermore, compared with daily PM_1 estimates using the STET model in our previous study (CV-R^2 = 0.76 and slope = 0.70; Wei et al., 2019b), the overall accuracy of daily $PM_{2.5}$ estimates using the enhanced STET model has improved significantly with a much higher CV-R^2 of 0.89 and a steeper slope of 0.86, based on data from 2018 in China. Continuous improvements of the model can further improve the determination of the relationship between fine particulate matter and AOD so as to improve the model performance. More data samples may also help improve the training ability of the model.

5.2 Predictive power

To test the predictive power ~~in $PM_{2.5}$ concentrations~~ of the enhanced STET model, the model built for the year of 2018 was used to predict daily $PM_{2.5}$ concentrations in 2017, validated against the ground measurements from 2017. Results suggest that our new model can correctly capture more than 65% of the historical daily $PM_{2.5}$ concentrations (N = 177,616). Monthly (N = 12,408), seasonal (N = 5,227), and annual (N = 1,461) mean $PM_{2.5}$ predictions across China~~. The comparison results~~ are highly correlated with surface observations with R^2 values of 0.80, 0.81, and 0.82, respectively, having overall small estimation uncertainties (i.e., RMSE < 12 μg/m^3, MAE < 9 μg/m^3, and MRE < 26 μg/m^3). There are only a handful of studies examining the predictive powers of models estimating $PM_{2.5}$ concentrations in China. Comparisons show that ~~our~~the enhanced STET model is superior to those ~~results~~ reported ~~by~~in previous studies, i.e., the two-stage model (Ma et al., 2019), the GTWR model (He and Huang, 2018), the ML + GAM model (Xue et al., 2019), and the ~~STRF~~space-time RF model (Wei et al., 2019e). The enhanced STET model has a strong predictive power and can be used to estimate historical $PM_{2.5}$ concentrations in China.

~~3.~~6. Summary and conclusions

With the increase in air pollution over recent years, abundant studies on estimating $PM_{2.5}$ have been performed using satellite remote sensing. However, most of the $PM_{2.5}$ estimates are reported at spatial resolutions of 3–10 km, which is inadequate for monitoring air quality ~~at~~in urban areas. ~~The~~ Traditional models also limit the accuracy of $PM_{2.5}$ estimates ~~is also limited by traditional models. Therefore~~. Here, we ~~try to generate~~present spatially continuous high-quality $PM_{2.5}$ maps at a 1-km ~~higher~~ spatial resolution across China. For this, ~~a new space-time extremely randomized trees (~~an enhanced STET~~)~~ ~~approach is~~ model was developed to minimize ~~the~~ spatiotemporal heterogeneities ~~in $PM_{2.5}$~~ and improve the overall estimate accuracy of ground-level $PM_{2.5}$ concentrations.

Our results suggest that the enhanced STET model ~~shows great performance in estimating~~estimates well daily $PM_{2.5}$ concentrations at the national scale with a relatively high sample-based cross-validation coefficient of 0.89, low RMSE of 10.35 μg/m^3, MAE of 6.71 μg/m^3, and MRE of 21.37~~% at the national scale.~~%. Comparisons illustrate that spatiotemporal information is ~~of great importance~~important and should be carefully considered during model development. The enhanced STET model ~~shows better performance~~estimates $PM_{2.5}$ concentrations well at most monitoring stations and individual days in the year. The North China Plain and the Sichuan Basin regions, under the influence of intense human activities and poor dispersion conditions, have high $PM_{2.5}$ loadings. ~~Moreover, the~~The enhanced STET model can outperform most models presented in previous related studies in terms of spatial resolution, model accuracy, and predictive power. This study suggests that the 1-km-resolution $PM_{2.5}$ dataset will be ~~of great importance~~useful in future atmospheric pollution studies focused on medium- or small-scale areas. ~~In addition, the~~ The enhanced STET model ~~will~~may be applied in the future to produce ~~the~~ historical $PM_{2.5}$ ~~dataset across~~datasets for China ~~in our future studies since~~ because the MODIS ~~can cover global observations nearly over the past~~data record extends back 20 years.

Data availability

Data are available by contacting the first author (weijing_rs@163.com).

Author contributions

ZL designed the research, and JW carried out the research and wrote the initial draft of this manuscript. All authors made substantial contributions to this work.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The in-situ $PM_{2.5}$ measurements are available from the China National Environmental Monitoring Center (http://www.cnemc.cn). The MODIS series products are available at https://search.earthdata.nasa.gov/, and the ERA-Interim reanalysis products are available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim. The AERONET measurements are available at https://aeronet.gsfc.nasa.gov/. We would like to thank Dr. Qiang Zhang at Tsinghua University for providing MEIC pollution emission data in China.

Financial support

This research has been supported by the National Key R&D Program of China (2017YFC1501702), the National Natural Science Foundation of China (91544217), the U.S. National Science Foundation (AGS1534670), and the BNU Interdisciplinary Research Foundation for ~~the~~ First-Year Doctoral Candidates (BNUXKJC1808).

References

Aggarwal, P., and Jain, S.: Impact of air pollutants from surface transport sources on human health: a modeling and epidemiological approach, Environ. Int., 83, 146–157, 2015.

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

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

Calle, M., and Urrea, V.: Letter to the editor: satiability of random forest importance measures, Briefings Bioinform., 12(1), 86–89, 2011.

Chen, G., Li, S., Knibbs, L., Hamm, N., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M., and Guo, Y.: A machine learning method to estimate $PM_{2.5}$ concentrations across China with remote sensing, meteorological and land use information, Sci. Total Environ., 636, 52–60, 2018.

Chen, Z., Zhang, T., Zhang, R., Zhu, Z., Yang, J., Chen, P., Ou, C., and Guo, Y.: Extreme gradient boosting model to estimate $PM_{2.5}$ concentrations with missing-filled satellite data in China, Atmos. Environ., 202, 180–189, 2019.

Chowdhury, S., and Dey, S.: Cause-specific premature death from ambient $PM_{2.5}$ exposure in India: estimate adjusted for baseline mortality, Environ. Int., 91, 283–290, 2016.

Crippa, M., Janssens-Maenhout, G., Guizzardi, D., Van Dingenen, R., and Dentener, F.: Contribution and uncertainty of sectorial and regional emissions to regional and global $PM_{2.5}$ health impacts, Atmos. Chem. Phys., 19, 5165–5186, 2019.

Fang, X., Zou, B., Liu, X., Sternberg, T., and Zhai, L.: Satellite-based ground $PM_{2.5}$ estimation using timely structure adaptive modeling. Remote Sens. Environ., 186, 152–163, 2016.

Geurts, P., Ernst, D., and Wehenkel, L.: Extremely randomized trees, Mach. Learn., 63(1), 3–42, 2006.

Guo, J., Xia, F., Zhang, Y., Liu, H., Li, J., Lou, M, He, J., Yan, Y., Wang, F., Min, M., and Zhai, P.: Impact of diurnal variability and meteorological factors on the $PM_{2.5}$-AOD relationship: implications for $PM_{2.5}$ remote sensing, Environ. Pollut., 221(94), 94, 2017.

Guo, J., Zhang, X., Che, H., Gong, S., An, X., Cao, C., Guang, J., Zhang, H., Wang, Y., Zhang, X., Xue, M., and Li, X.: Correlation between PM concentrations and aerosol optical depth in eastern China, Atmos. Environ., 43(37), 5876–5886, 2009.

Gupta, P., and Christopher, S.: Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: multiple regression approach, J. Geophys. Res. Atmos., 114(D14205), https://doi.org/10.1029/2008JD011496, 2009.

Han, L., Zhou, W., Li, W., and Li, L.: Impact of urbanization level on urban air quality: a case of fine particles ($PM_{2.5}$) in Chinese cities, Environ. Pollut., 194, 163–170, 2014.

He, K., Hong Huo, A., and Zhang, Q.: Urban air pollution in China: current status, characteristics, and progress, Annu. Rev. Energ. Env., 27(1), 397–431, 2011.

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

Hsu, N., Gautam, R., Sayer, A., Bettenhausen, C., Li, C., Jeong, M. J., Tsay, S., and Holben, B.: Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010, Atmos. Chem. Phys., 12, 8037–8053, 2012.

Huang, R., Zhang, Y., Bozzetti, C., Ho, K., Cao, J., Han, Y., Daellenbach, K., Slowik, J., Platt, S., Canonaco, F., Zotter, P., Wolf, R., Pieber, S., Bruns, E., Cripa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., An, Z., Szidat., S., Baltensperger, U., Haddad, I., and Prévôt, A.: High secondary aerosol contribution to particulate pollution during haze events in China, Nature, 514(7521), 218–222, 2014.

Jiang, R., Tang, W., Wu, X., and Fu, W.: A random forest approach to the detection of epistatic interactions in case-control studies, BMC Bioinformatics, 10(2), 135–135, 2009.

Jin, X., Wang, Y., Li, Z., Zhang, F., Xu, W., Sun, Y., Fan, X., Chen, G., Wu, H., Ren, J., Wang, Q., and Cribb, M.: Significant contribution of organics to aerosol liquid water content in winter in Beijing, China, Atmos. Chem. Phys., 20, 901–914, https://doi.org/10.5194/acp-20-901-2020, 2020.

Koren, I., Dagan, G., and Altaratz, O.: From aerosol-limited to invigoration of warm convective clouds, Science, 344, 1143–1146, 2014.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmos. Meas. Tech., 6, 2989–3034, 2013.

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

Li, T., Shen, H., Yuan, Q., Zhang, X., and Zhang, L.: Estimating ground-level $PM_{2.5}$ by fusing satellite and station observations: a geo-intelligent deep learning approach, Geophys. Res. Lett., 44(23), 11,985–11,993, 2017b.

Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., Sun, Y., Wang, T., Xue, H., Zhang, H., and Zhu, B.: Aerosol and boundary-layer interactions and impact on air quality, Natl. Sci. Rev., 4(6), 810–833, 2017.

Liu, M., Huang, Y., Ma, Z., Jin, Z., Liu, X., Wang, H., Liu, Y., Wang, J., Jantunen, M., Bi, J., Kinney, P. L.: Spatial and temporal trends in the mortality burden of air pollution in China: 2004-2012, Environ. Intl., 98, 75–81, 2017.

Liu, N., Zou, B., Feng, H., Wang, W., Tang, Y., and 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., 19, 8243–8268, 2019.

Lyapustin, A., Wang, Y., Korkin, S., and Huang, D.: MODIS Collection 6 MAIAC algorithm, Atmos. Meas. Tech., 11, 5741–5765, 2018.

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R., and Reid, J.: Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm, J. Geophys. Res. Atmos., 116, https://doi.org/10.1029/2010JD014985, 2011.

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

Ma, Z., Liu, R., Liu, Y., and Bi, J.: Effects of air pollution control policies on $PM_{2.5}$ pollution improvement in China from 2005 to 2017: a satellite-based perspective, Atmos. Chem. Phys., 19, 6861–6877, 2019.

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

Reddington, C. L., Morgan, W. T., Darbyshire, E., Brito, J., Coe, H., Artaxo, P., Scott, C. E., Marsham, J., and Spracklen, D. V.: Biomass burning aerosol over the Amazon: analysis of aircraft, surface and satellite observations using a global aerosol model, Atmos. Chem. Phys., 19, 9125-9152, 10.5194/acp-19-9125-2019, 2019.

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

Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., Demott, P. J., Dunlea, E. J., Feingold, G., Ghan, S., Chan, S., Guenther, A., Kahn, R., Kredenweis, S., Molina, M., Nenes, A., Penner, J., Prather, K., Ramanathan, V., Ramaswamy, V., Rashch, P., and Ravishankara, A.: Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system, P. Natl. Acad. Sci. USA, 113(21), 5781–5790, 2016.

Silva, R., West, J., Zhang, Y., Anenberg, S., Lamarque, J., Shindell, D., Collins, W., Dalsøren, S., Faluvegi, G., Folberth, G., Horowitz, L., Nagashima, T., Naik, V., Rumbold, S., Skeie, R., Sudo, K., Takemura, T., Bergmann, D., Cameron-Smith, P., Cionni, I., Doherty, R., Eyring, V., Josse, B., MacKenzie, I., Plummer, D., Righi, M., Stevenson, D., Strode, S., Szopa, S., and Zeng, G.: Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change, Environ. Res. Lett., 8(3), 034005, 2013.

Song, Y., Huang, B., He, Q., Chen, B., Wei, J., and Mahmood, R.: Dynamic assessment of $PM_{2.5}$ exposure and health risk using remote sensing and geo-spatial big data, Environ. Pollut., 253, 288–296, 2019.

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

Su, T., Li, Z., and Kahn, R.: A new method to retrieve the diurnal variability of planetary boundary layer height from lidar under different thermodynamic stability conditions. Remote Sens. Environ., 237, 111519, 2020.

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

Sun, Y., Jiang, Q., Wang, Z., Fu, P., Li, J., Yang, T., and Yin, Y.: Investigation of the sources and evolution processes of severe haze pollution in Beijing in January 2013, J. Geophys. Res. Atmos., 119, 4380–4398, 2014.

Tao, M., Wang, J., Li, R., Wang, L., Wang, L., Wang, Z., Tao, J., Che, H., and Chen, L.: Performance of MODIS high-resolution MAIAC aerosol algorithm in China: characterization and limitation, Atmos. Environ., 213, 159–169, 2019.

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

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., and Cribb, M.: Estimating 1-km-resolution $PM_{2.5}$ concentrations across China using the space-time random forest approach, Remote Sens. Environ., 231, 111221, https://doi.org/10.1016/j.rse.2019.111221, 2019a.

Wei, J., Li, Z., Guo, J., Sun, L., Huang, W., Xue, W., Fan, T, and Cribb, M.: Satellite-derived 1-km-resolution PM_1 concentrations from 2014 to 2018 across China, Environ. Sci. Tech., 53(22), 13,265–13,274, https://doi.org/10.1021/acs.est.9b03258, 2019b.

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

Wei, J., Li, Z., Peng, Y., Sun, L., and Yan, X.: A regionally robust high-spatial-resolution aerosol retrieval algorithm for MODIS images over Eastern China, IEEE T. Geosci. Remote, 57(7), 4748–4757, 2019d.

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., 202, 315–327, 2019e.

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

Wu, J., Zheng, H., Zhe, F., Xie, W., and Song, J.: Study on the relationship between urbanization and fine particulate matter ($PM_{2.5}$) concentration and its implication in China, J. Clean. Prod., 182, 872–882, 2018.

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

Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., and Zhang, Q.: Spatiotemporal continuous estimates of $PM_{2.5}$ concentrations in China, 2000–2016: a machine learning method with inputs from satellites, chemical transport model, and ground observations, Environ. Intl., 123, 345–357, 2019.

Yao, F., Wu, J., Li, W., and 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., 151, 263–276, 2019.

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

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

Zhai, S., Jacob, D., Wang, X., Shen, L., Li, K., Zhang, Y., Gui, K., Zhang, T., and Liao, H.: Fine particulate matter ($PM_{2.5}$) trends in China, 2013–2018: separating contributions from anthropogenic emissions and meteorology, Atmos. Chem. Phys., 19, 11,031–11,041, 2019.

Zhang, Q., Streets, D., He, K., and Klimont, Z.: Major components of China's anthropogenic primary particulate emissions, Environ. Res. Lett., 2(4), 045027, 2007.

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

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

Zheng, G., Duan, F., Su, H., Ma, Y., Cheng, Y., Zheng, B., Zhang, Q., Huang, T., Kimoto, T., and Chang, D.: Exploring the severe winter haze in Beijing: the impact of synoptic weather, regional transport and heterogeneous reactions, Atmos. Chem. Phys., 15, 2969–2983, 2015.

Table 1. Summary of the data sources used in this study.

Dataset	Variable	Content	Unit	Spatial Resolution	Temporal Resolution	Data source
$PM_{2.5}$	$PM_{2.5}$	~~$PM_{2.5}$~~Particulate matter ≤ 2.5 μm	μg/m^3	~~-~~in situ	Hourly	CNEMC
AOD	AOD	MAIAC AOD	-	1 km ×1 km	Daily	MCD19A2
~~Meteorological data~~Meteorology	BLH	Boundary layer height	m	0.125°×0.125°	3-hour	ERA-Interim ~~reanalysis product~~
	PRE	Total precipitation	mm		3-hour	
	EP	Evaporation	mm		3-hour	
	RH	Relative humidity	%		3-hour	
	TEM	2-m air temperature	K		6-hour	
	SP	Surface pressure	hPa		6-hour	
	WS	10-m wind speed	m/s		6-hour	
	WD	10-m wind direction	m/s		6-hour	
Land ~~cover~~use	NDVI	NDVI	-	500 m × 500 m	Monthly	MOD13A3
	LUC	Land use cover	-		Annually	MCD12Q1
Topography	DEM	DEM	m	90 m × 90 m	-	SRTM
	Relief	Surface relief	m			
	Aspect	Surface aspect	degree			
	Slope	Surface slope	degree			
Emission	SO_2	Sulfur dioxide	Mg/grid	0.25°×0.25°	Monthly	MEIC
	NO_x	Nitrogen oxide				
	CO	Carbon monoxide				
	VOC	Volatile organic compounds				
	Dust	Fine-sized dust				
Population	NTL	Night lights	W/cm^2/sr	500 m × 500 m	Monthly	VIIRS

Table 2. Comparison between model performances of the enhanced STET model and other models from previous related studies focused on China.

Model	Resolution	Model Validation					Predictive power		Literature
		R^2	RMSE	MAE	Slope	Intercept	Daily	Monthly	
GWR	10 km	0.64	32.98	21.25	-0.67	-21.22	-	-	Ma et al~~.~~., (2014)
TSAM	10 km	0.80	22.75	15.99	-0.79	-15.31	-	-	Fang et al. (2016)
Gaussian	10 km	0.81	21.87	-	-0.73	-17.97	-	-	Yu et al. (2017)
RF	10 km	0.83	18.08	-	-	-	-	-	Chen et al. (2018)
GAM		0.55	29.13	-	-	-	-	-	
DBN	10 km	0.54	25.86	18.10	-0.55	-24.56			Li et al. (2017b)
Geo-DBN		0.88	13.03	08.54	-0.86	-6.39	-	-	
Two-stage	10 km	0.77	17.10	11.51	0.76	11.64	0.41	0.73	Ma et al. (2019)
Two-stage	6 km	0.60	21.76	14.41	-0.85	-8.63	-	-	Yao et al. (2019)
GRNN	3 km	0.67	20.93	13.90	-0.62	-22.90	-	-	Li et al. (2017a)
GWR	3 km	0.81	21.87	-	-0.83	-9.44	-	-	You et al. (2016)
D-GWR	3 km	0.72	21.01	14.59	-0.79	-12.92	-	-	He ~~&~~and Huang (2018)
Two-stage		0.71	21.21	13.50	-0.73	-16.67	-	-	
GTWR		0.80	18.00	12.03	0.81	11.69	0.41	-	
XGBoost	3 km	0.86	14.98	-	-	-	-	-	Chen et al. (2019)
ML	3 km	0.53	30.40	19.60	0.53	25.3			Xue et al. (2019)
ML + GAM		0.61	27.80	17.70	0.61	21.2	0.57	0.74	
MLR	1 km	0.41	20.04	30.03	0.41	30.03	0.38	-	Wei et al. (2019e)
GWR		0.53	23.28	19.26	0.61	20.93	0.44	-	
Two-stage		0.71	18.59	14.54	0.71	15.10	0.35	-	
RF		0.81	17.91	11.50	0.77	12.56	0.53	-	
STRF		0.85	15.57	9.77	0.82	9.64	0.55	0.73	
STET	1 km	0.89	10.35	6.71	0.~~60~~86	6.16	0.65	0.80	~~Our~~This study

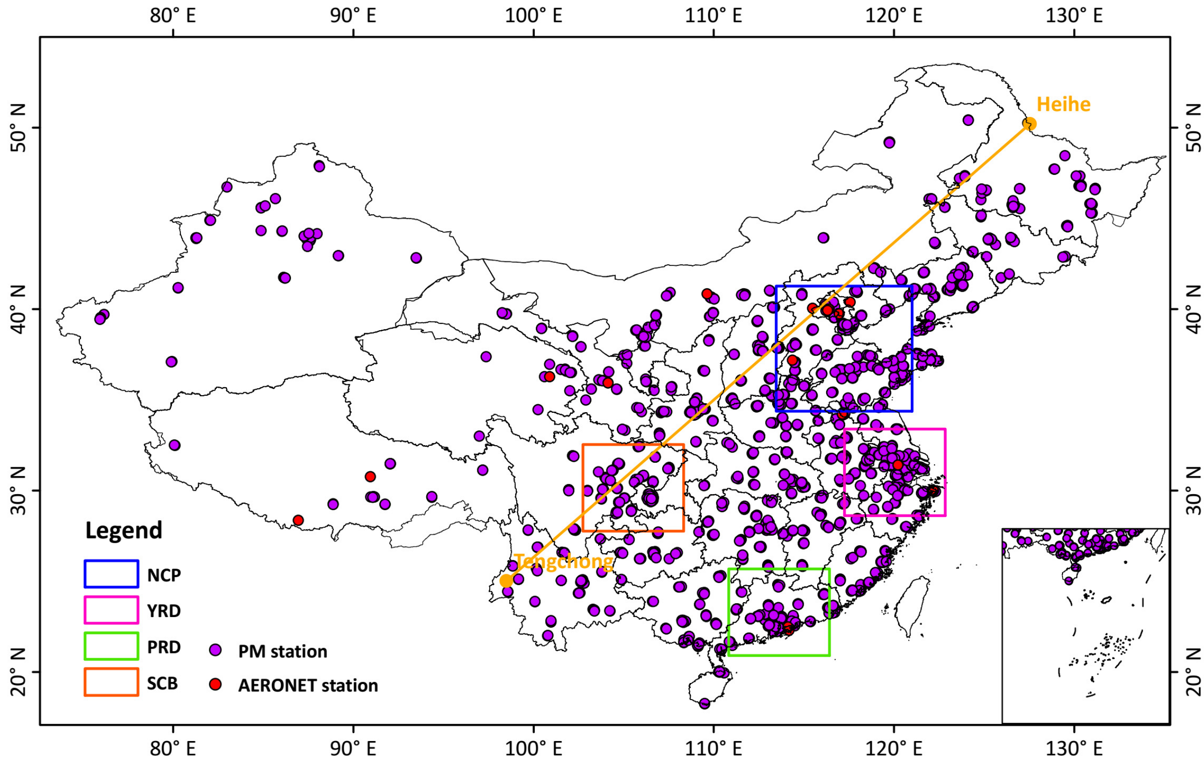


Figure 1. Spatial distributions of $PM_{2.5}$ and AERONET monitoring stations in China. The Heihe-Tengchong line (orange line) shows the boundary between ~~Eastern~~eastern and ~~Western~~western China.

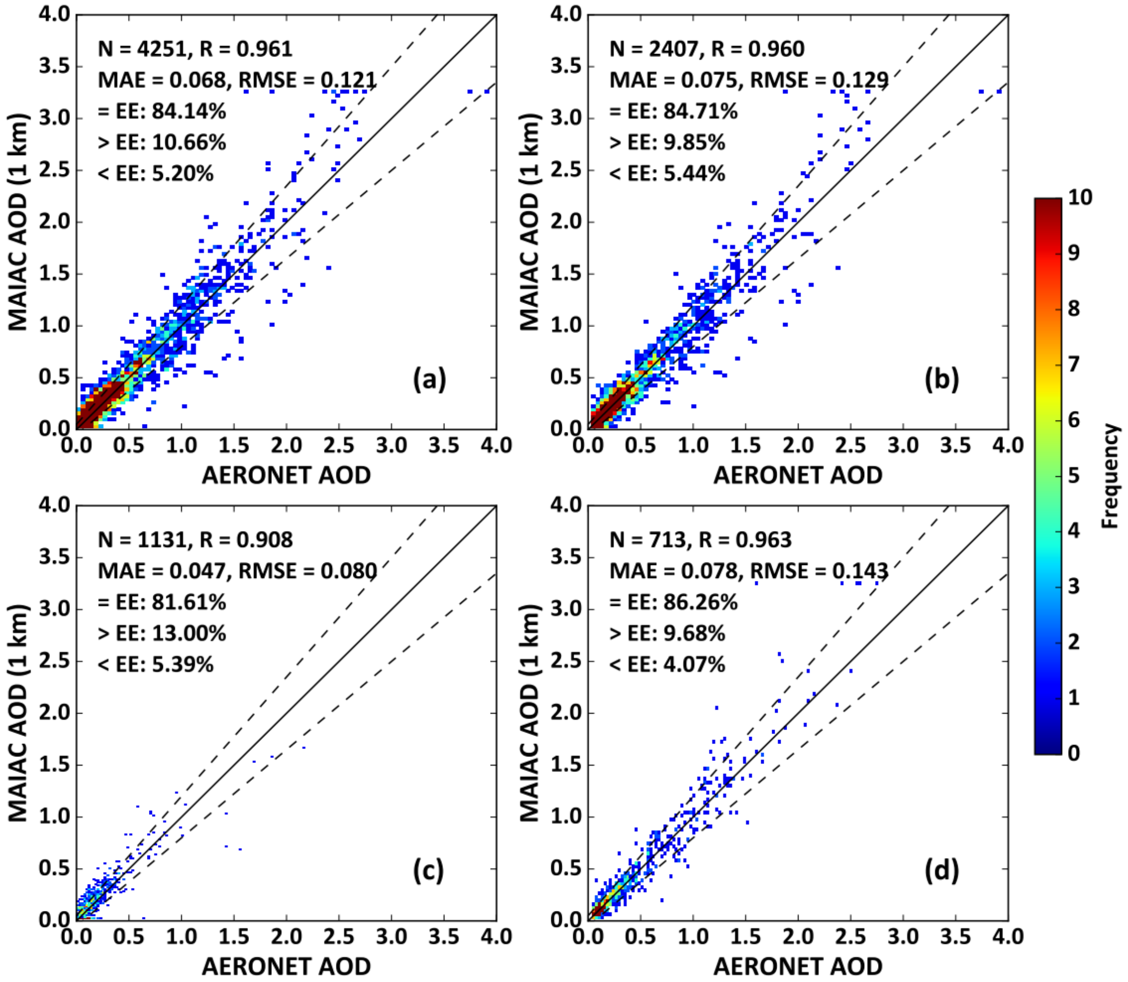
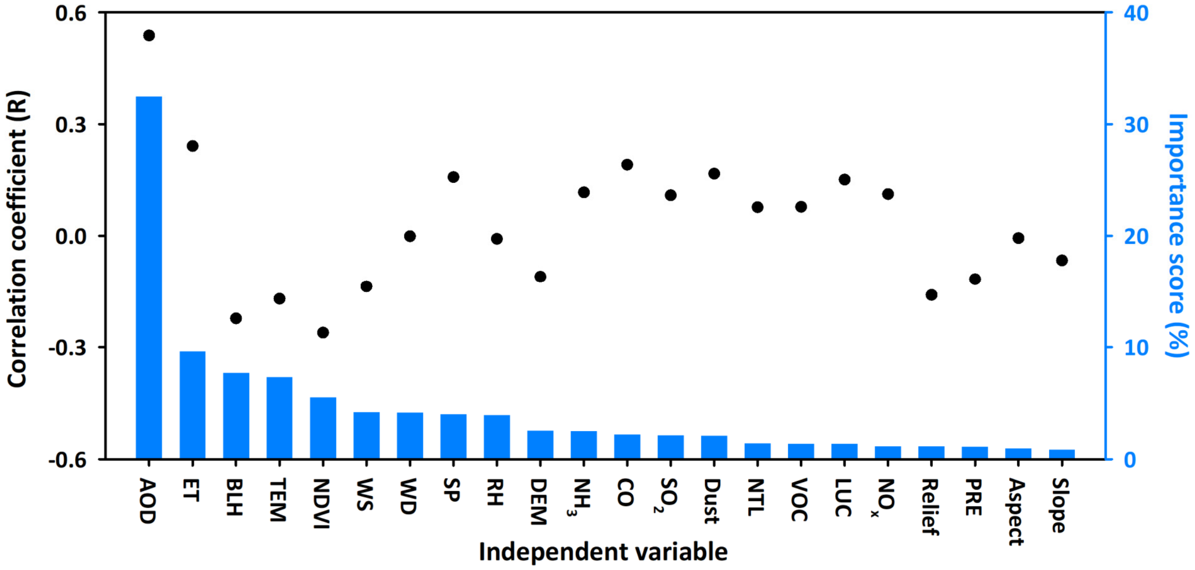
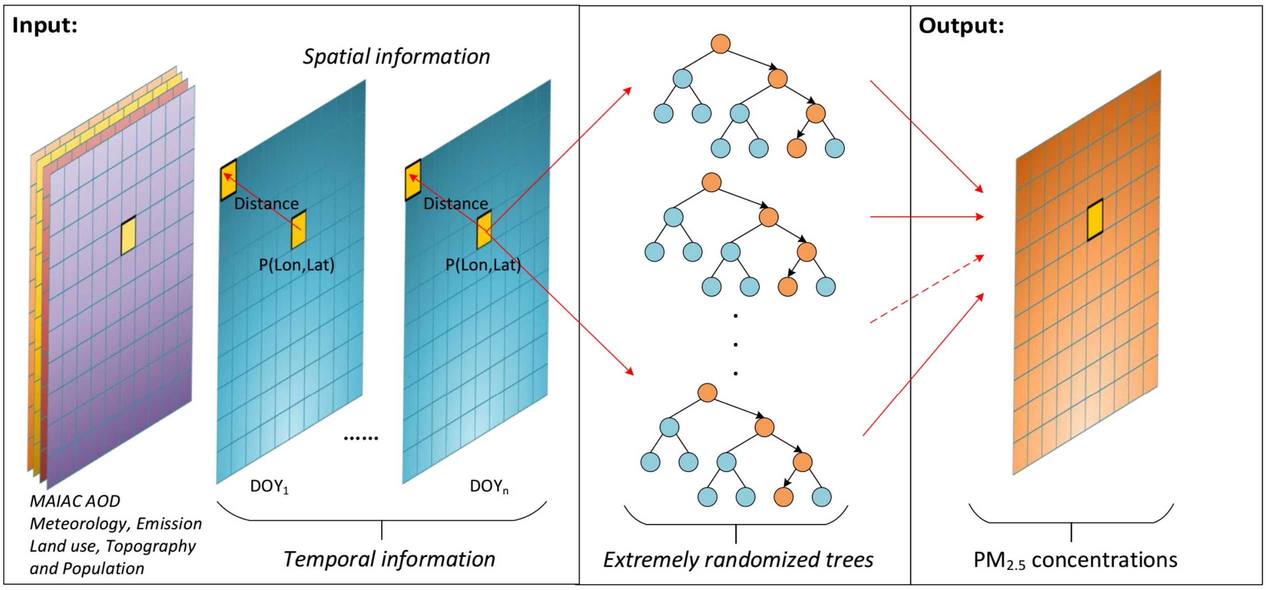


Figure 2. Scatter plots of MAIAC AOD retrievals versus AERONET AODs at 550 nm in (a) China, and (b) urban, (c) cropland, and (d) grassland areas. The dotted lines represent the upper and lower boundaries of the expected error (EE). Statistical metrics are given in each panel: the number of samples (N), the correlation coefficient (R), the mean absolute error (MAE), and the root-mean-square error (RMSE).



760 Figure 3. Potential effects and importance scores (blue bars; unit: %) of independent variables to $PM_{2.5}$ estimates for the STET model.



765 Figure 4. Schematic of the enhanced STET model developed in our study.

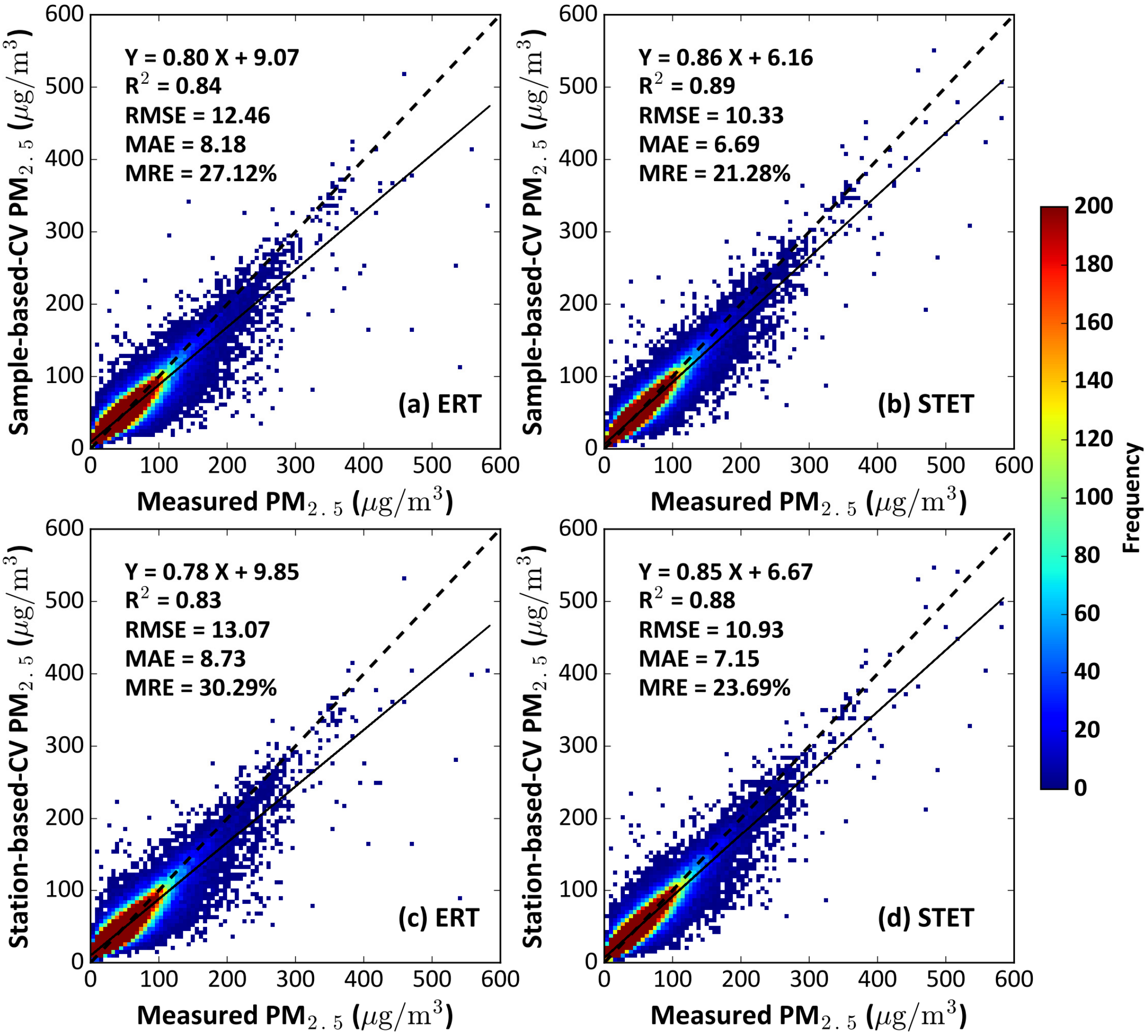


Figure 5. Density scatter plots of out-of-sample~~-based~~ (top row) and out-of-station~~-based~~ (bottom row) 10-CV results for the ~~ET~~ERT (left column) and STET (right column) models at the daily level ~~(N = 167,692)~~ in 2018 ~~across~~for mainland China. Statistical metrics are given in each panel, along with the linear regression relation: the correlation of determination (R^2), the root-mean-square error (RMSE), the mean absolute error (MAE), and the mean relative error (MRE).

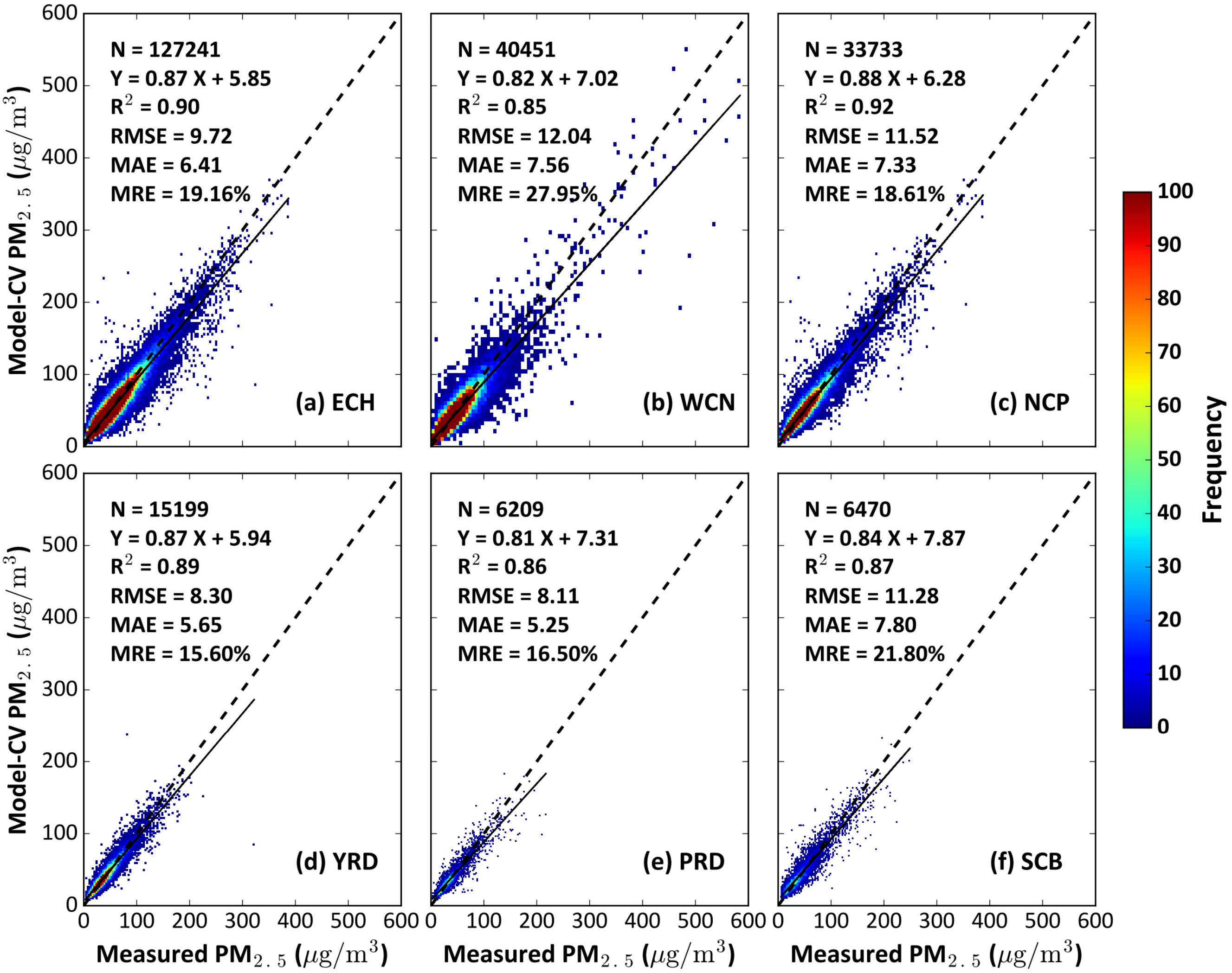


Figure 6. Density scatter plots of out-of-sample~~-based~~ 10-CV results for (a) eastern China (ECH), (b) western China (WCH), (c) the North China Plain (NCP), (d) the Yangtze River Delta (YRD), (e) the Pearl River Delta (PRD), and (f) the Sichuan Basin (SCB) in 2018. Statistical metrics are given in each panel, along with the linear regression relation: the number of samples (N), the correlation of determination (R^2), the root-mean-square error (RMSE), the mean absolute error (MAE), and the mean relative error (MRE).

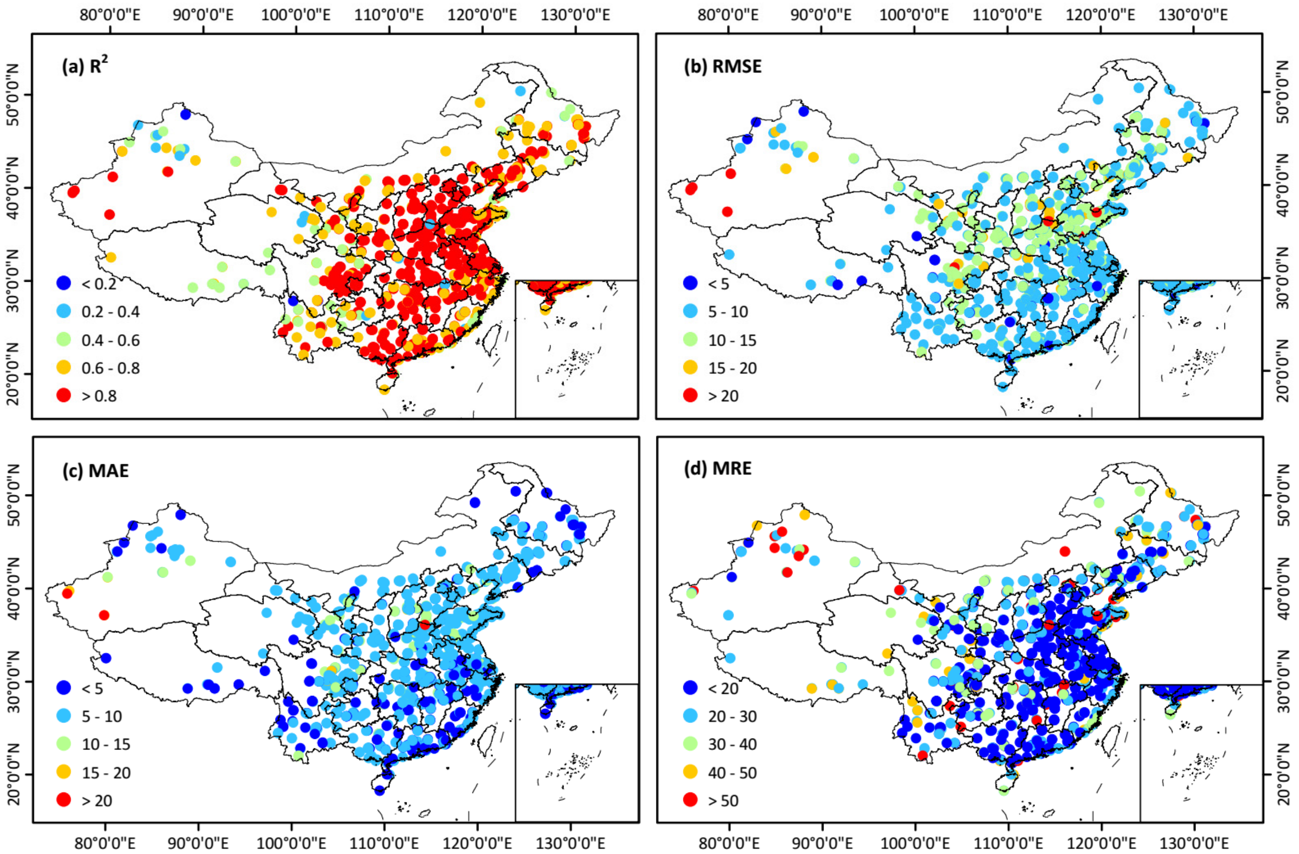


Figure 7. Spatial distributions of the site-scale performance of the STET model for (a) the sample-based ~~CV R^2,~~cross-validation coefficient of determination (R^2), (b) ~~RMSE,~~the root-mean-square error (RMSE), (c) ~~MAE,~~the mean absolute error (MAE), and (d) the mean relative error (MRE) in 2018 across China.

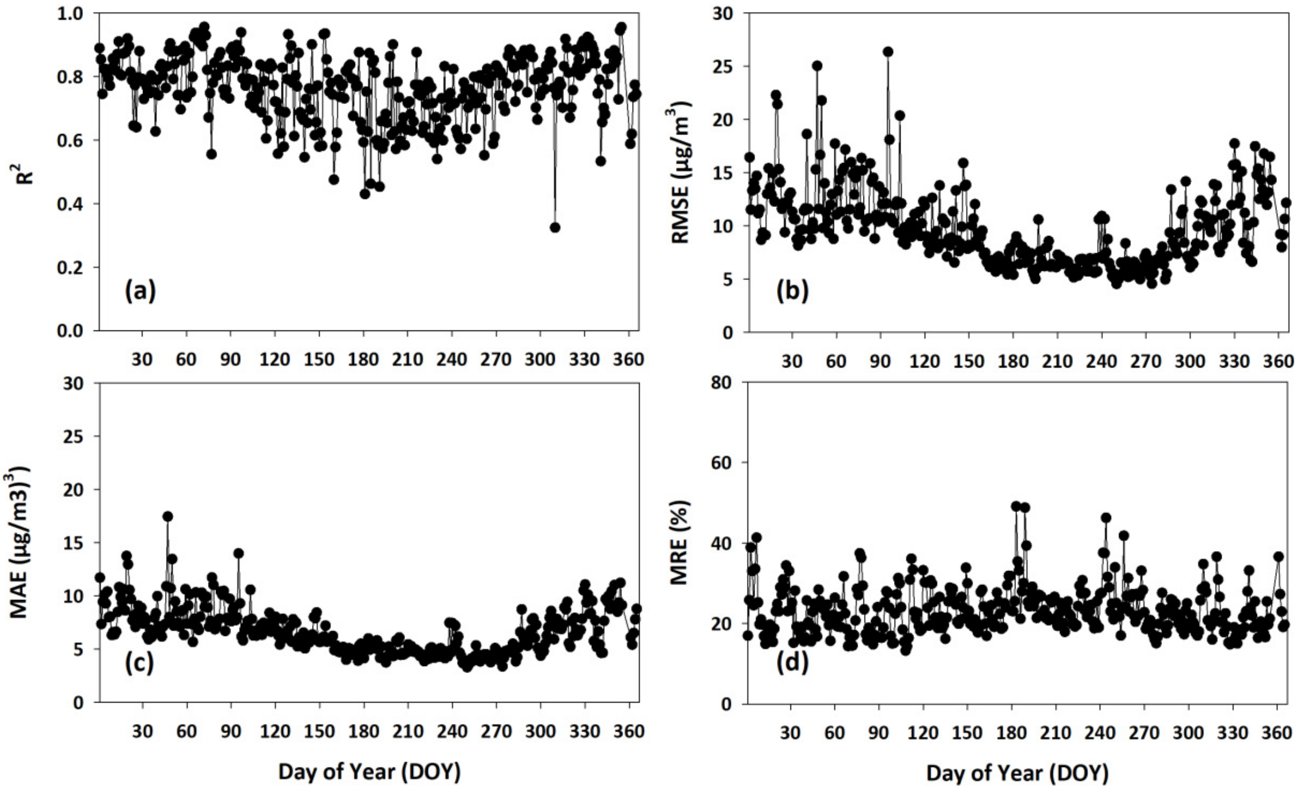


Figure 8. Time series of the daily performance of the STET model in terms of (a) sample-based ~~CV-R^2,~~cross-validation coefficient of determination (R^2), (b) ~~RMSE,~~the root-mean-square error (RMSE), (c) ~~MAE,~~the mean absolute error (MAE), and (d) the mean relative error (MRE) in 2018 across China.

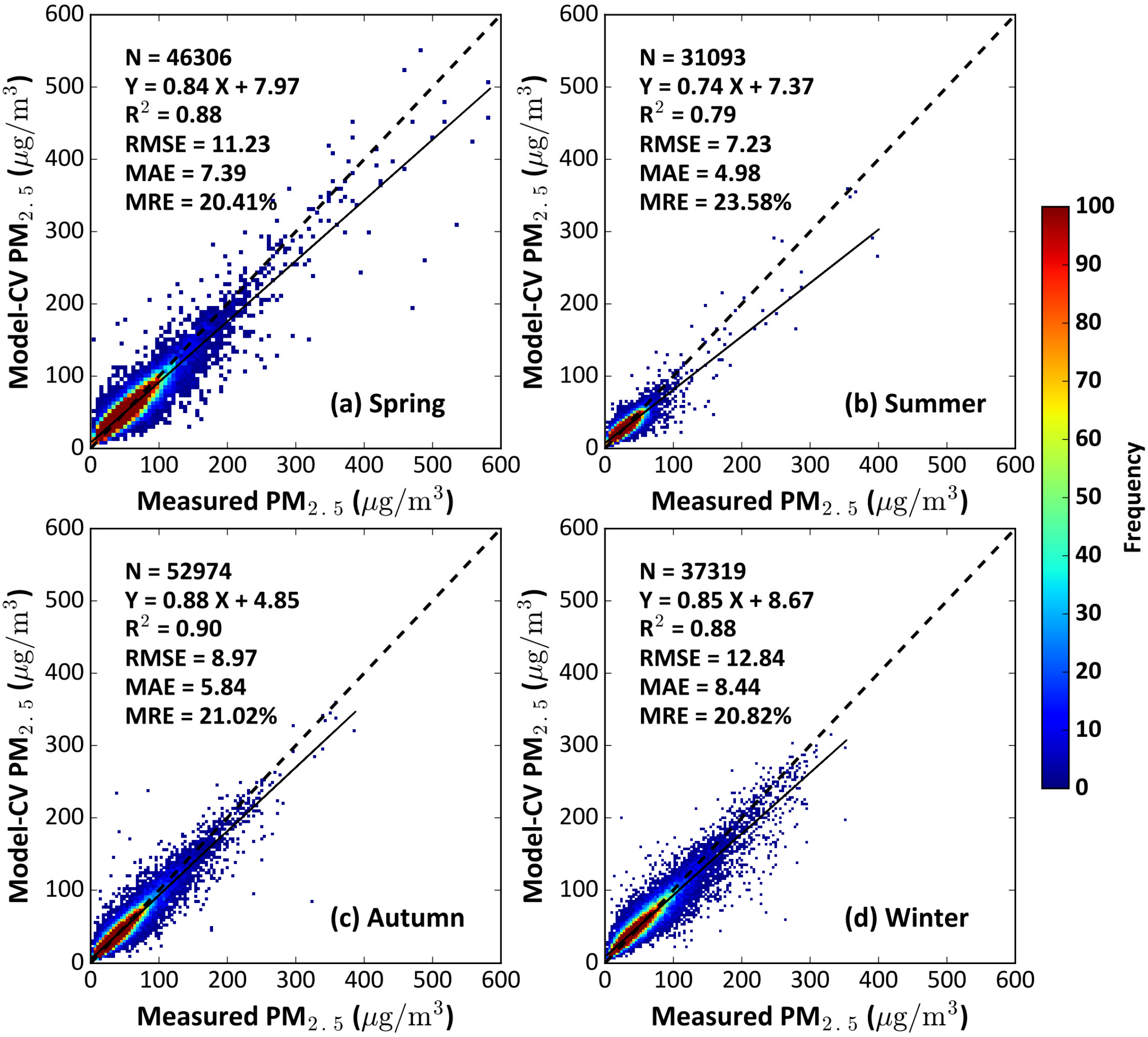


Figure 9. Density scatter plots of sample-based 10-CV results for the STET model for ~~four seasons~~the four seasons in 2018 across China. Statistical metrics are given in each panel, along with the linear regression relation: the number of samples (N), the correlation of determination (R^2), the root-mean-square error (RMSE), the mean absolute error (MAE), and the mean relative error (MRE).

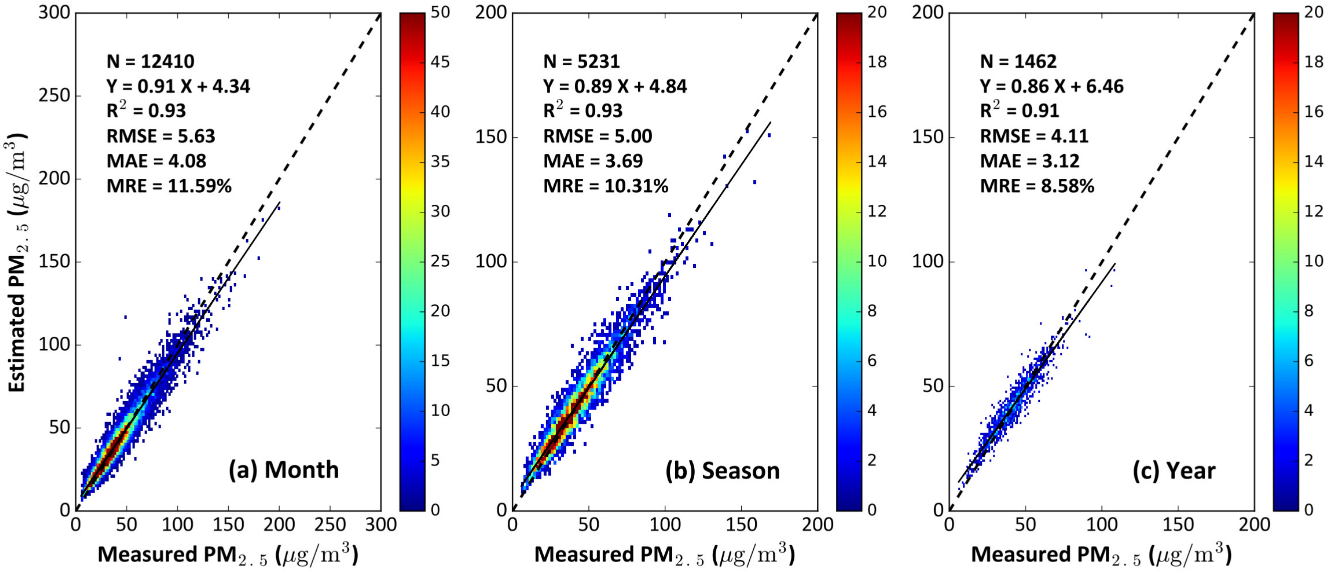


Figure 10. Validation of (a) monthly, (b) seasonal, and (c) annual $PM_{2.5}$ estimates in 2018 ~~across~~in China. Statistical metrics are given in each panel, along with the linear regression relation: the number of samples (N), the correlation of determination (R^2), the root-mean-square error (RMSE), the mean absolute error (MAE), and the mean relative error (MRE).

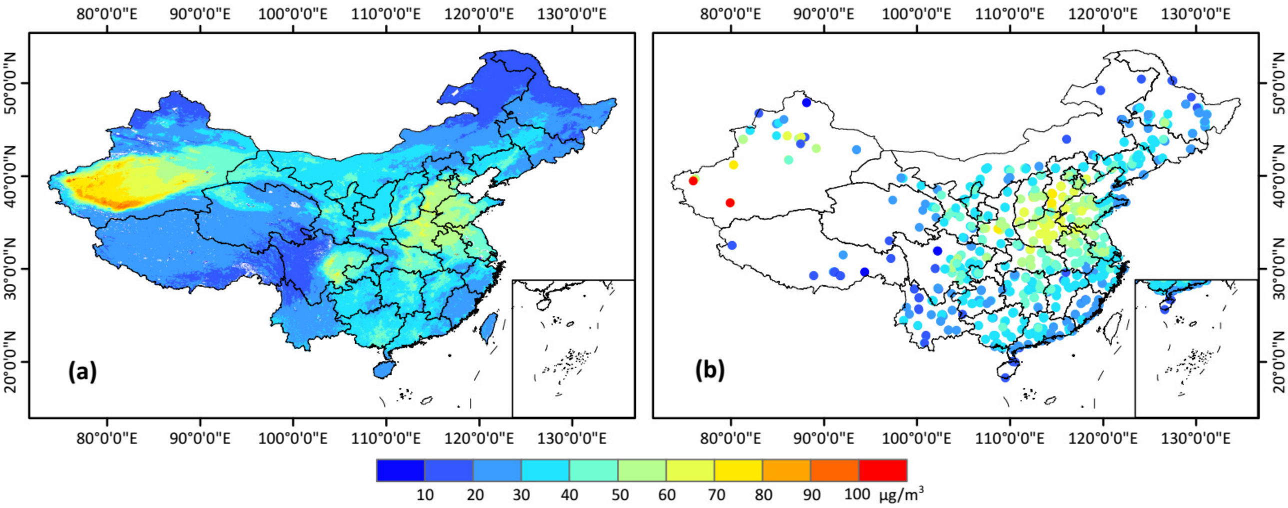
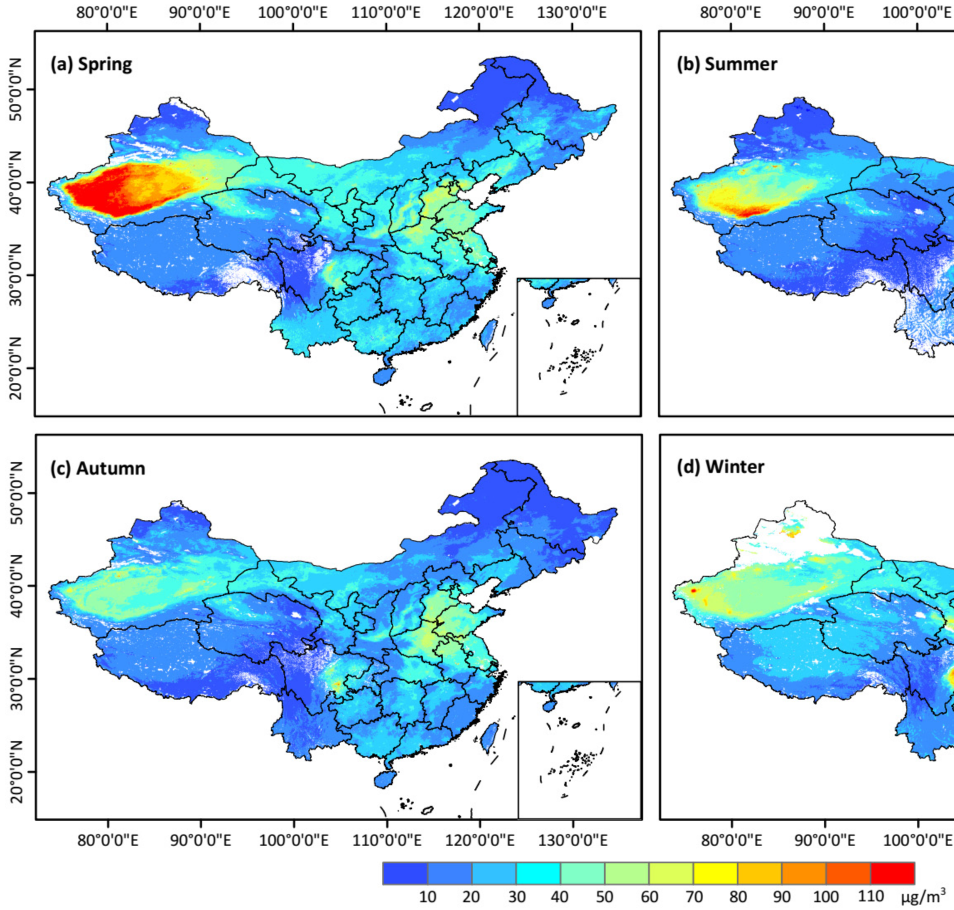


Figure 11. Spatial distributions of annual mean (a) $PM_{2.5}$ estimates and (b) surface observations in 2018 across China.



810 Figure 12. Spatial distributions of seasonal mean 1-km-resolution $PM_{2.5}$ concentrations in 2018 across China.