Response to Comments on the Manuscript (acp-2020-1004):

"Estimating daily full-coverage and high-accuracy 5-km ambient particulate matters across China: considering their precursors and chemical compositions"

Response to Comments of Referee #1:

General comment:

Wang et al. present an analysis of particulate matter over China. They use a Light Gradient Boosting Machine regression technique to combine satellite observations and model simulations to estimate surface PM 2.5 and PM 10. While the broad topics addressed in this manuscript (i.e. pollutant estimation, data-model fusion, machine learning, etc.) are important areas of research, I cannot recommend this paper for publication due to serious methodological issues and the fact that very similar and more comprehensive work has already been published.

Response: We would like to take this opportunity to gratefully thank the referee for his/her comments and recommendations for improving the paper. An item-by-item response to the interesting comments raised by the referee follows. Thanks for your time.

Similarity to previous works:

Van Donkelaar et al. (2019) and Hammer et al. (2020) use a far simpler statistical method to achieve similar performance at across larger temporal, spatial, and chemical scales. As a specific example, Hammer et al. (2020) use a relatively simple linear adjustment of satellite observations to achieve similar performance at similar resolution globally for 20 years.

Response: Thank the referee for his/her important comments. The mentioned works (Van Donkelaar et al., 2019; Hammer et al., 2020) provided valuable results and made a significant contribution to the scientific community. However, the proposed study is not similar to these works and many distinctions/highlights can be found in our study compared to them, which are described as follows.

1) The purpose of our study is different from those of them. The proposed study aims at estimating the daily full-coverage PM using the datasets of their precursors & chemical compositions instead of AOD products. In our study, the validation results signify that the estimation model can perform well without the input of AOD. By contrast, the mentioned works adopted numerous AOD products (e.g., MAIAC, DB, and MISR) as major variates. To be specific, see "We use **AOD** retrieved from radiances measured by four satellite instruments..." in Hammer et al. (2020) and "We combined **AOD** from multiple satellite products…" in Van Donkelaar et al. (2019). As for chemical species, these works employed a CTM (GEOS-Chem), while our study utilizes the datasets from two sources: remote sensing (S5P-TROPOMI) and data assimilation (GEOS-FP). These indicate that the intention (or emphasis) of the proposed study entirely differs from these works.

2) The temporal resolution of the estimated PM in the proposed study (daily) is much higher than those in the mentioned works (annual). To be specific, see "The temporal resolution of these globally fused $PM_{2.5}$ estimates focused on **annual** mean values to inform global health assessments…" in Hammer et al. (2020) and " **Annual** $PM_{2.5}$ composition estimates resulting from this effort…" in Van Donkelaar et al. (2019).

3) Only the annual estimated results of the mentioned works were validated at a global scale (Hammer et al., 2020) or over North America (Van Donkelaar et al., 2019). Meanwhile, the ground-based sites over China were not selected for an individual validation in these works. By contrast, the daily estimated PM are validated across China with multiple methods in our study. As a consequence, it cannot be concluded that the mentioned works achieved similar performance.

4) These works exploited an empirical method and the widely used GWR. The specific steps of these works (see their Supplement Information) are complicated and include plenty of data preprocessing procedures. By contrast, the proposed study adopts a convenient end-to-end approach based on an advanced gradient boosting algorithm (i.e., Light-GBM).

5) The motioned works only considered the estimation of ambient $PM_{2.5}$. In our study, ambient $PM_{2.5}$ and PM_{10} are both estimated with high accuracy through the proposed approach.

References:

Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., ... & Brauer, M. (2020). Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998-2018). Environmental Science & Technology.

Van Donkelaar, A., Martin, R. V., Li, C., & Burnett, R. T. (2019). Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. Environmental science & technology, 53(5), 2595-2611.

Methodological Issues:

Section 3.1: This data processing is inappropriate for the input data as described in the supplement. In particular, downscaling wind fields through bicubic interpolation does not preserve mass or energy, nor does it accurately reproduce any higher order variability in the wind fields. More advanced statistical approaches (e.g. Kirchmeier et al., 2014) are necessary to produce physically consistent informations.

Response: Thank the referee for his/her valuable comments. The resampling method is an important part of the data preprocessing steps. The recommended resampling method (Kirchmeier et al., 2014) is advanced and can preserve mass or energy. However, it requires ground-based stations of wind fields to acquire the Probability Density Function (PDF). Since the historical ground measurements of meteorological factors are undisclosed in China, this method is difficult to be employed. For previous related works about the estimation of PM, they usually adopted some simple methods to resample meteorological factors, such as the nearest neighbor method (Hu et al., 2014, 2017; Yao et al., 2019) and bilinear/bicubic/inverse distance weighted interpolation (Chen et al., 2019; Li et al., 2017; Wei et al., 2019; Yang et al., 2020; Shen et al., 2018; Ma et al., 2019). Therefore, the bicubic interpolation is selected as the resampling method in our study and the validation results do not suggest that the adoption of it will introduce large errors. At present, the researches about the influence of the errors from resampling methods on the estimated PM are scarce. This is a topic that is worth exploring and we will study it in our future works.

References:

Chen, Z. Y., Zhang, T. H., Zhang, R., Zhu, Z. M., Yang, J., Chen, P. Y., ... & Guo, Y. (2019). Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China. Atmospheric environment, 202, 180-189.

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., & Liu, Y. (2017). Estimating PM2. 5

concentrations in the conterminous United States using the random forest approach. Environmental science & technology, 51(12), 6936-6944.

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., ... & Liu, Y. (2014). Estimating ground-level PM2. 5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. Remote Sensing of Environment, 140, 220-232.

Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating ground‐level PM2.5 by fusing satellite and station observations: a geo‐intelligent deep learning approach. Geophysical Research Letters, 44(23), 11-985.

Ma, Z., Liu, R., Liu, Y., & Bi, J. (2019). Effects of air pollution control policies on PM 2.5 pollution improvement in China from 2005 to 2017: a satellite-based perspective. Atmospheric Chemistry and Physics, 19(10), 6861-6877.

Shen, H., Li, T., Yuan, Q., & Zhang, L. (2018). Estimating regional ground‐level PM2. 5 directly from satellite top‐of‐atmosphere reflectance using deep belief networks. Journal of Geophysical Research: Atmospheres, 123(24), 13-875.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., & Cribb, M. (2019). Estimating 1-km-resolution PM2. 5 concentrations across China using the space-time random forest approach. Remote Sensing of Environment, 231, 111221.

Yang, Q., Yuan, Q., Yue, L., Li, T., Shen, H., & Zhang, L. (2020). Mapping PM2. 5 concentration at a sub-km level resolution: A dual-scale retrieval approach. ISPRS Journal of Photogrammetry and Remote Sensing, 165, 140-151.

Yao, F., Wu, J., Li, W., & Peng, J. (2019). A spatially structured adaptive two-stage model for retrieving ground-level PM2. 5 concentrations from VIIRS AOD in China. ISPRS Journal of Photogrammetry and Remote Sensing, 151, 263-276.

Section 3.3: The 10x cross validation scheme described in this section does not account for the very large amount of spatial and temporal correlation present in environmental data. This will substantially and inappropriately bias the estimated skill of the machine learning model (e.g. Hastie et al. 2001, Roberts et al., 2017). The authors should perform a more rigorous evaluation of the model skill consistent with the standard in the field (e.g. Barnes et al., 2020). For example, the authors could cross validate through block methods by removing longer time periods of data or removing entire spatial regions of data beyond the autocorrelation length scale.

Response: Thank the referee for his/her significant comments. The 10x cross validation scheme was widely used in previous related works about the estimation of PM over China (Chen et al., 2018a, 2018b; Wei et al., 2019; He et al., 2018; Zhang et al., 2019; Fang et al., 2016; Ma et al., 2014, 2016, 2019). The mentioned works (Van Donkelaar et al., 2019; Hammer et al., 2020) also utilized the 10x cross validation scheme. For instance, see "Performance was evaluated using a **10-fold** cross-validation…" in Van Donkelaar et al. (2019) and "The scatterplot shows **10-fold** out-of-sample 10% cross validation at sites..." in Hammer et al. (2020). In our study, a total of three 10x cross validation schemes (i.e., sample-based, space-based, and time-based) are adopted. Among them, the space-based cross validation scheme is the most commonly used to evaluate the spatial accuracy of the estimated results (Li et al., 2020). Since we need to compare the validation results with the related works, these cross validation schemes are necessary in our study. The recommended methods (region-based or historical validation) are occasionally employed to verify the spatial or temporal predictive ability of the model, such as in Li et al. (2017) and Wei et al. (2019). At present, our study does not focus on the improvement to the predictive ability of the model. However, the historical validation results (removing longer time periods of data) are listed in Table r1 to present the temporal predictive ability of the proposed approach. For comparison, the historical validation results of some related works (Wei et al., 2019; Ma et al., 2016; He et al., 2018) are also provided in Table r2. As can be seen, the temporal predictive ability of the proposed approach is desired compared to these works.

Table r1: Metrics of the historical validation results in our study.

Type	Training period	Validation period	Approach	R^2	RMSE ($\mu g/m^3$)
$PM_{2.5}$	2018.06.01-2019.05.31 (365)	2019.06.01-2020.03.31 (305)	Proposed	0.59	21.28
			AOD-based	0.54	22.37
PM_{10}			Proposed	0.5	36.82
			AOD-based	0.42	48.12

Table r2: Metrics of the historical validation results in previous related works over China.

Type	Reference	R^2	RMSE ($\mu g/m^3$)	Validation period	Full-coverage
$PM_{2.5}$	Wei et al., 2019	0.55	27.38	2016 (366)	False
	He et al., 2018	0.47	37.57	2014 (365)	
	Ma et al., 2016	0.41	-	2014.01.01-2014.06.30 (181)	

References:

Chen, G., Li, S., Knibbs, L. D., Hamm, N. A., Cao, W., Li, T., ... & Guo, Y. (2018b). A machine learning method to estimate PM2. 5 concentrations across China with remote sensing, meteorological and land use information. Science of the Total Environment, 636, 52-60.

Chen, G., Wang, Y., Li, S., Cao, W., Ren, H., Knibbs, L. D., ... & Guo, Y. (2018a). Spatiotemporal patterns of PM10 concentrations over China during 2005–2016: A satellite-based estimation using the random forests approach. Environmental Pollution, 242, 605-613.

Fang, X., Zou, B., Liu, X., Sternberg, T., & Zhai, L. (2016). Satellite-based ground PM2.5 estimation using timely structure adaptive modeling. Remote Sensing of Environment, 186, 152-163.

Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., ... & Brauer, M. (2020). Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998-2018). Environmental Science & Technology.

He, Q., & Huang, B. (2018). Satellite-based mapping of daily high-resolution ground PM2. 5 in China via space-time regression modeling. Remote Sensing of Environment, 206, 72-83.

Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating ground‐level PM2. 5 by fusing satellite and station observations: a geo‐intelligent deep learning approach. Geophysical Research Letters, 44(23), 11-985.

Li, T., Shen, H., Zeng, C., & Yuan, Q. (2020). A Validation Approach Considering the Uneven Distribution of Ground Stations for Satellite-Based PM 2.5 Estimation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 1312-1321.

Ma, Z., Hu, X., Huang, L., Bi, J., & Liu, Y. (2014). Estimating ground-level PM2. 5 in China using satellite remote sensing. Environmental science & technology, 48(13), 7436-7444.

Ma, Z., Hu, X., Sayer, A. M., Levy, R., Zhang, Q., Xue, Y., ... & Liu, Y. (2016). Satellite-based spatiotemporal trends in PM2. 5 concentrations: China, 2004–2013. Environmental health perspectives, 124(2), 184-192.

Ma, Z., Liu, R., Liu, Y., & Bi, J. (2019). Effects of air pollution control policies on PM 2.5 pollution improvement in China from 2005 to 2017: a satellite-based perspective. Atmospheric Chemistry and Physics, 19(10), 6861-6877.

Van Donkelaar, A., Martin, R. V., Li, C., & Burnett, R. T. (2019). Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. Environmental science & technology, 53(5), 2595-2611.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., & Cribb, M. (2019). Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. Remote Sensing of Environment, 231, 111221.

Zhang, T., Zang, L., Wan, Y., Wang, W., & Zhang, Y. (2019). Ground-level PM2. 5 estimation over urban agglomerations in China with high spatiotemporal resolution based on Himawari-8. Science of the total environment, 676, 535-544.

Section 4.4: The feature importance results in this section call in question the validity of the modelling approach here. The top ranked variables include those that were unphysically downscaled: wind speeds (U, V, and Ustar) and turbulent evaporation, as well as total column ozone. As the vast majority of the ozone within a total column is present in the stratosphere, any surface predictive capacity associated with that variable is dubious. Additionally, the high importance ranking of NO_2 potentially indicates that the machine learning model is predicting based simply off the proximity to combustion sources. This explains why the model performance is so poor over more remote regions in Western China and limits the applicability of the method developed.

Response: Thank the referee for his/her meaningful comments. The total comments can be divided into three parts, which are replied as follows.

1) "unphysically downscaled meteorological factors (e.g., wind speed)": The reasons for the adoption of bicubic interpolation in our study have been explained above. Previous related works about the estimation of PM usually applied some simple methods to resample meteorological factors. There is no indication that bicubic interpolation will lead to large errors in our study.

2) "total column ozone": For the stratospheric O_3, a latest study (Chen et al., 2020) has shown that the downward transport of O_3 stemming from the stratosphere-to-troposphere exchange can be a significant contributor to background O_3. Such enhancement of background O_3 will affect ambient PM. In addition, ambient O_3 pollution is rapidly increasing over China in recent years (Liu et al., 2020; Wang et al., 2020) and the proportion of it may also rise in the total O_3 column. It is inferred that more surface information can be extracted from the total O_3 column in China compared to other regions. At present, the total O_3 column has been used to estimate ambient O_3 over China (Liu et al., 2020) and Tibetan Plateau (Li et al., 2020a), suggesting its surface predictive capacity. In China, ambient PM is associated with ambient O_3 (Chen et al., 2019). Therefore, the total O_3 column is introduced as an auxiliary variate in our study and the feature importance of it is ranked 9^{th} and 7^{th} for $PM_{2.5}$ and PM_{10}, respectively.

3) "poor model performance in Western China": The poor performance of the proposed approach in Western China primarily results from the imbalanced matched samples. Since most of the ground-based sites distribute in the populous regions, the matched samples are small in Western China. This makes it difficult to extract useful information over these regions. In our study, the high importance ranking of NO_2 potentially signifies that nitrate is generally the major contribution to ambient PM in China according to the current distribution of ground-based sites. Ambient PM presents strong heterogeneous spatial patterns over China (Li et al., 2017, 2020b); consequently, the model performance is spatially various due to the imbalanced matched samples. In previous related works about the estimation of PM over China, this phenomenon is common. Some examples (Chen et al., 2019; Wei et al., 2019; Li et al., 2020b) are provided in Figure r1-r3. At present, the purpose of our study does not aim at addressing the issue caused by the imbalanced (or small) matched samples. We will study it in our future works.

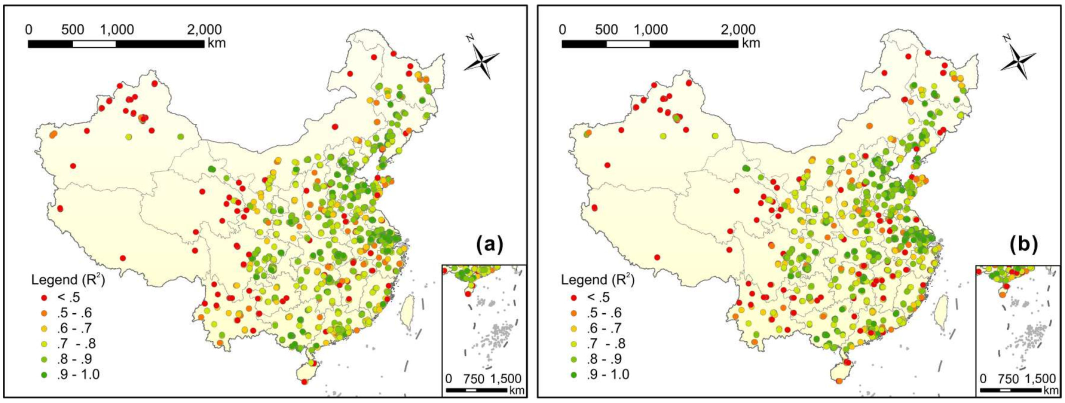

Figure r1: Distribution of the (a) sample-based CV R^2 and (b) site-based CV R^2 of the GTW-GRNN model (Li et al., 2020b).

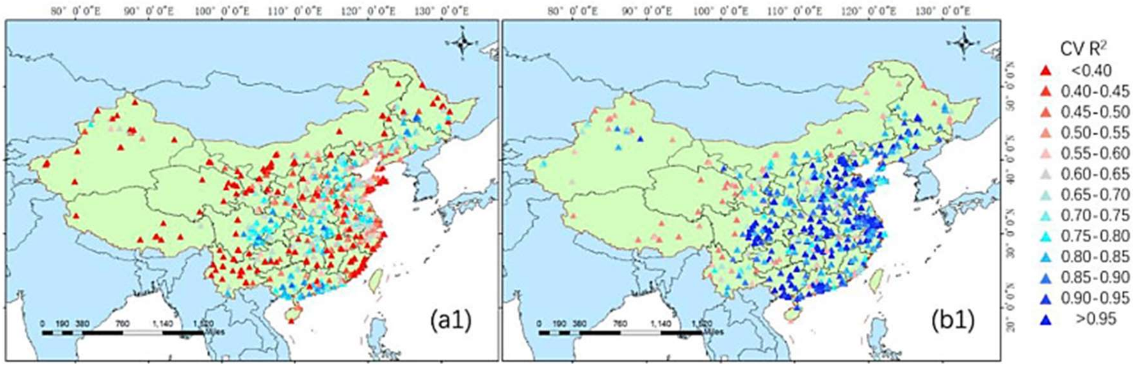

Figure r2: Spatial distribution of the spatial cross-validation result (blue color indicates a better fit) with each grid with $PM_{2.5}$ monitors (Chen et al., 2019).

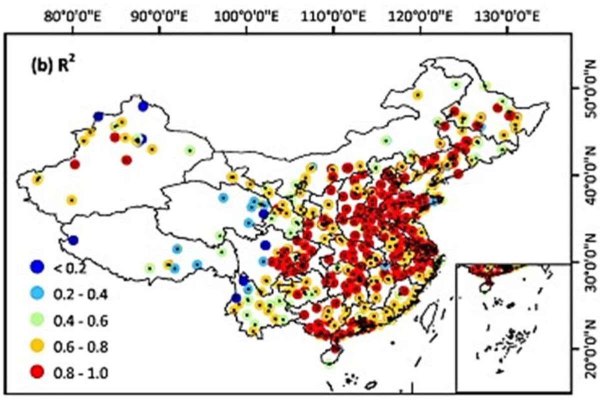

Figure r3: Spatial distributions of R^2 between $PM_{2.5}$ estimations and measurements from 2016 in China. Results are from the sample-based 10-cross-validation (Wei et al., 2019).

References:

Chen, L., Xing, J., Mathur, R., Liu, S., Wang, S., & Hao, J. (2020). Quantification of the enhancement of PM2.5 concentration by the downward transport of ozone from the stratosphere. Chemosphere, 126907.

Chen, J., Shen, H., Li, T., Peng, X., Cheng, H., & Ma, C. (2019). Temporal and Spatial Features of the Correlation between PM2.5 and O3 Concentrations in China. International Journal of Environmental Research and Public Health, 16(23), 4824.

Chen, Z. Y., Zhang, T. H., Zhang, R., Zhu, Z. M., Yang, J., Chen, P. Y., ... & Guo, Y. (2019). Extreme gradient boosting model to estimate PM2. 5 concentrations with missing-filled satellite data in China. Atmospheric environment, 202, 180-189.

Li, T., Shen, H., Yuan, Q., Zhang, X., & Zhang, L. (2017). Estimating ground - level PM2. 5 by fusing satellite and station observations: a geo - intelligent deep learning approach. Geophysical Research Letters, 44(23), 11-985.

Li, R., Zhao, Y., Zhou, W., Meng, Y., Zhang, Z., & Fu, H. (2020a). Developing a novel hybrid model for the estimation of surface 8 h ozone (O3) across the remote Tibetan Plateau during 2005–2018. Atmospheric Chemistry and Physics, 20(10), 6159-6175.

Li, T., Shen, H., Yuan, Q., & Zhang, L. (2020b). Geographically and temporally weighted neural networks for satellite-based mapping of ground-level PM2. 5. ISPRS Journal of Photogrammetry and Remote Sensing, 167, 178-188.

Liu, R., Ma, Z., Liu, Y., Shao, Y., Zhao, W., & Bi, J. (2020). Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach. Environment International, 142, 105823.

Shao, Y., Ma, Z., Wang, J., & Bi, J. (2020). Estimating daily ground-level PM2. 5 in China with random-forest-based spatiotemporal kriging. Science of The Total Environment, 740, 139761.

Wang, Y., Wild, O., Chen, X., Wu, Q., Gao, M., Chen, H., ... & Wang, Z. (2020). Health impacts of long-term ozone exposure in China over 2013–2017. Environment International, 144, 106030.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., & Cribb, M. (2019). Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. Remote Sensing of Environment, 231, 111221.

Figures: All figures with maps in this paper violate the ACP maps policy: https://www.atmospheric-chemistry-and-physics.net/submission.html#mapsaerials. "In order to depoliticize scientific articles, authors should avoid the drawing of borders or use of contested topographical names." The inset of the South China Sea does not aid in the scientific interpretation of the results presented in this manuscript in any way, and only confuses the figures.

Response: Thank the referee for his/her comment. These figures have been redrawn in the manuscript.

An example of revision is as follows:

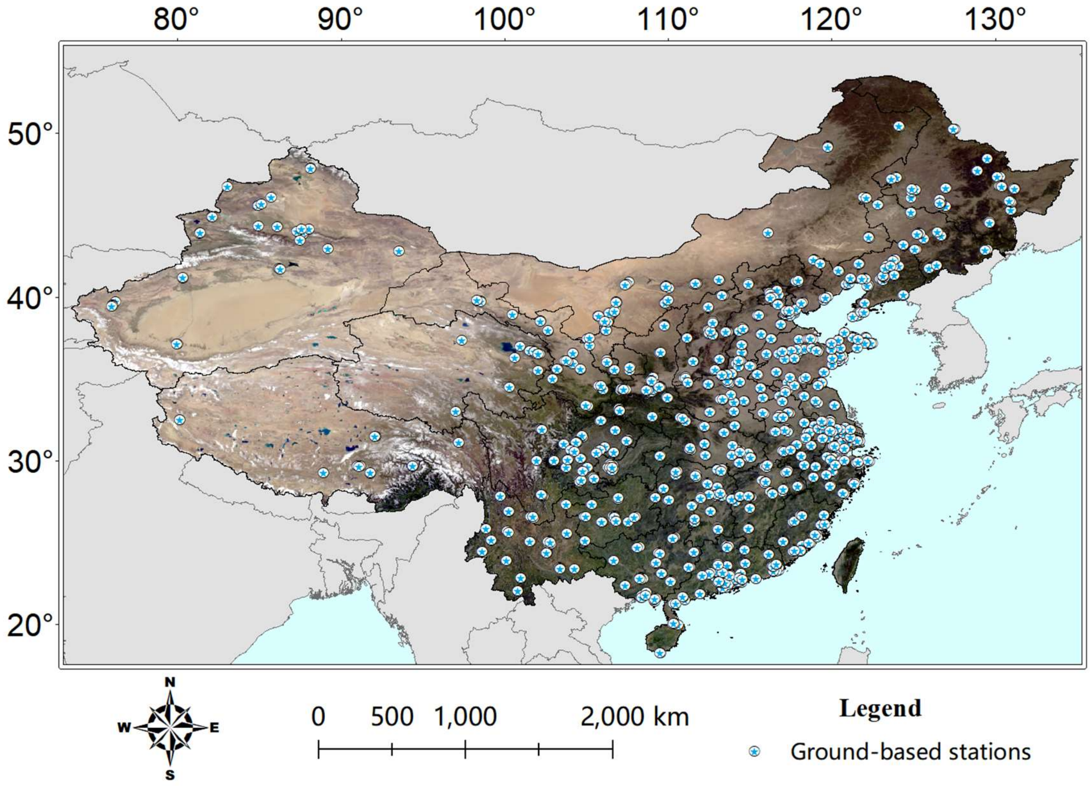

Figure 1. The spatial distribution of the ground-based stations over China. The base-map is the true color image of MODIS.

Specific Comments:

Q1.1: L14 "Most of the existing approaches for the estimation of PM 2.5 and PM 10 employed the remote sensing Aerosol Optical Depth (AOD) products as the main variate." I don't believe this is the case. Most approaches to estimate PM 2.5 and PM 10 come from in situ observations of aerosol mass and size distributions.

Response: Thank the referee for his/her comment. This statement is inaccurate and has been reworded in the manuscript.

The main revision is as follows:

At present, most of remote sensing based approaches for the estimation of $PM_{2.5}$ and PM_{10} employed Aerosol Optical Depth (AOD) products as the main variate.

Q1.2: L29 "conducive to the researches". I'm not sure what this sentence means.

Response: Thank the referee for his/her comment. This statement has been reworded in the manuscript.

The main revision is as follows:

It is concluded that the full-coverage estimated results from our study will be helpful in the field of large-scale $PM_{2.5}$ and PM_{10} monitoring over the regions where the AOD values are missing.

Q1.3: L38 Satellite observations are much more expensive than ground-based monitoring.

Response: Thank the referee for his/her comment. This statement is intended to express that the observations from existing atmospheric satellites (e.g., Terra, Himawari-8, and Suomi-NPP) can be adopted. We have deleted it in the manuscript.

Q1.4: L45 This statement needs appropriate citation.

Response: Thank the referee for his/her comment. This statement is inaccurate and has been reworded in the manuscript.

The main revision is as follows:

With regard to CTMs, the uncertainties of the emission inventories could be large in some areas (Li et al., 2017b) and it will consume time and energy to collect the necessary information for simulation (Chu et al., 2016). The approaches based on remote sensing satellites have been greatly developed in recent years (Sorek-Hamer et al., 2020).

References:

Sorek-Hamer, M., Chatfield, R., & Liu, Y. (2020). Strategies for using satellite-based products in modeling PM2.5 and short-term pollution episodes. Environment International, 144, 106057.

Chu, Y., Liu, Y., Li, X., Liu, Z., Lu, H., Lu, Y., ... & Liu, F. (2016). A review on predicting ground PM2.5 concentration using satellite aerosol optical depth. Atmosphere, 7(10), 129.

Q1.5: Section 2.2 This section is far too lacking in details to interpret or reproduce the work presented in this paper. The authors should explicitly state the variables used in the main body of the manuscript.

Response: Thank the referee for his/her comment. Since numerous variates (30) are introduced in the proposed approach, the specific description of them was considered tedious. To be more clear, the explicit statement about the variates used in our study has been appended in Section 3.2, following the estimation function.

The main revision is as follows:

The general scheme of the model for estimating the ambient concentrations of $PM_{2.5}$ and PM_{10} can be expressed as Eq. (1).

$$C_{PM} = f(VM_P, VM_{CC}, VA_{O3}, VA_{MF}, VA_{GF}, DOY) \quad (1)$$

where C_{PM} signifies the estimated ambient concentrations of $PM_{2.5}$ and PM_{10}. f denotes the estimation function for the ambient concentrations of $PM_{2.5}$ and PM_{10} based on LGBM. VM_P and VM_{CC} indicate the main variates of the precursors and chemical compositions, respectively, for $PM_{2.5}$ and PM_{10}. VA_{O3}, VA_{MF}, and VA_{GF} represent the auxiliary variates of the O_3 from TROPOMI, meteorological factors, and geographical factors, respectively; DOY is the day of year. To be specific, VM_P consists of nitrogendioxide_tropospheric_column and sulfurdioxide_total_vertical_column_1km; VM_{CC} includes Black Carbon Column Mass Density, Organic Carbon Column Mass Density, Nitrate Column Mass Density, SO4 Column Mass Density, Dust Column Mass Density, Ammonium Column Mass Density, and Sea Salt Column Mass Density; VM_{MF} covers 10-meter Specific Humidity, 10-meter Air Temperature, 10-meter Eastward Wind, 10-meter Northward Wind, Total Precipitable Water Vapor, Pbltop Pressure, Surface Pressure, Planetary Boundary Layer Height, Air Density at Surface, Surface Velocity Scale, and Evaporation from Turbulence; and VA_{GF} contains 1_km_16_days_NDVI, the fractions of forest, savanna, grassland, cropland, urban, and aridland, road density, and population density.

Q1.6: L120 This link does not work for me.

Response: Thank the referee for his/her comment. This website requires the Microsoft Silverlight (> 4.0) (https://www.microsoft.com/getsilverlight/get-started/install/default) and the screenshot of it has been presented in Figure r4. In addition, the air quality data is also available at http://www.cnemc.cn/.

Figure r4: Screenshot of http://106.37.208.233:20035.

Q1.7: Section 3.2 How were the hyperparameters selected? Was there any optimization or search algorithm applied here?

Response: Thank the referee for his/her comment. Since the matched samples (e.g., 31*742932 for $PM_{2.5}$) are large and the training procedure requires plenty of time, the random search based on cross validation is adopted in our study for some key hyperparameters (e.g., num_leaves and learning_rate).

Q1.8: Figure 5 The colors being split into only 5 bins makes assessing performance difficult. Consider using a continuous colorbar.

Response: Thank the referee for his/her comment. These figures have been redrawn in the manuscript.

An example of revision is as follows:

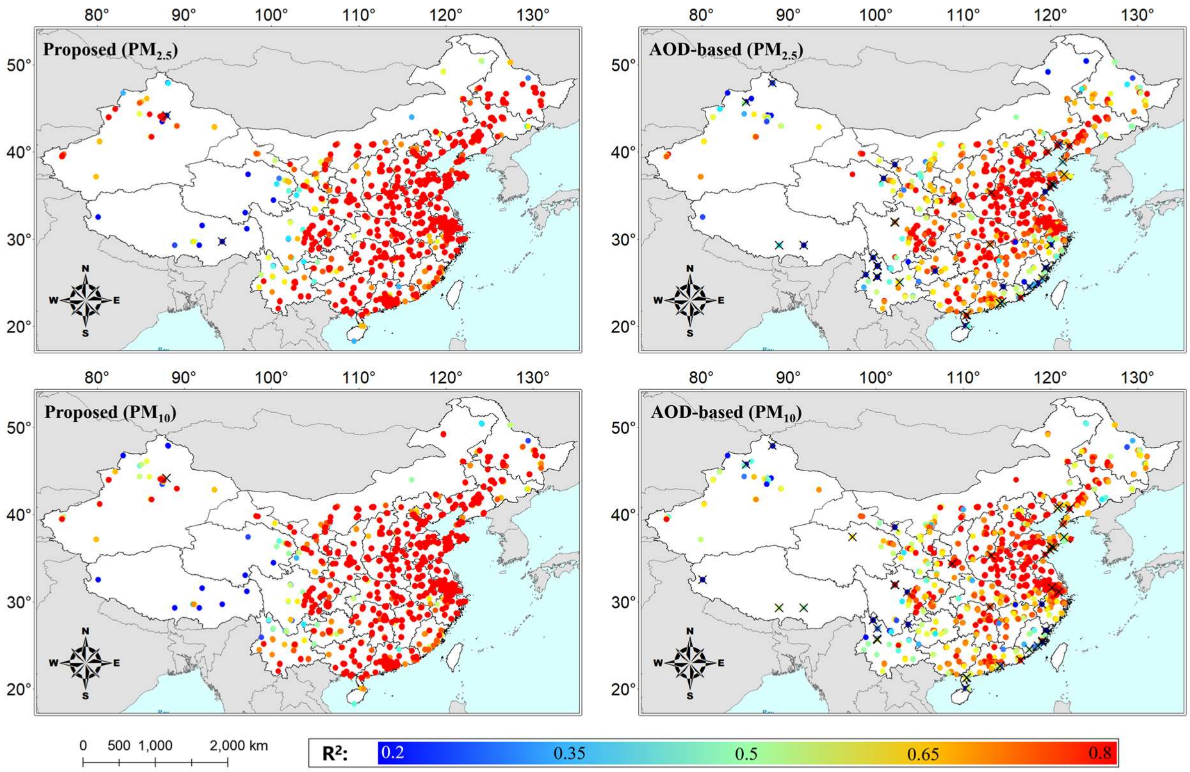

Figure 5. The spatial distribution of R^2s for the space-based CV at each matched grid over China. The black crosses denote that the significance levels (p) of the metrics are not less than 0.01 at these matched grids.

Q1.9: L329 The authors should explicitly show evidence for this incorrect estimation through sampling.

Response: Thank the referee for his/her comment. The annual validation results (space-based cross validation) of the estimated results (proposed and AOD-based) in 2019 over China are depicted in Figure r5. As can be seen, the annual performance of AOD-based is poor (large bias) due to the sampling that discards the missing values in the AOD product, with RPEs of 28.07% and 33.62% for $PM_{2.5}$ and PM_{10}, respectively. By contrast, the proposed approach performs well for the annual estimation.

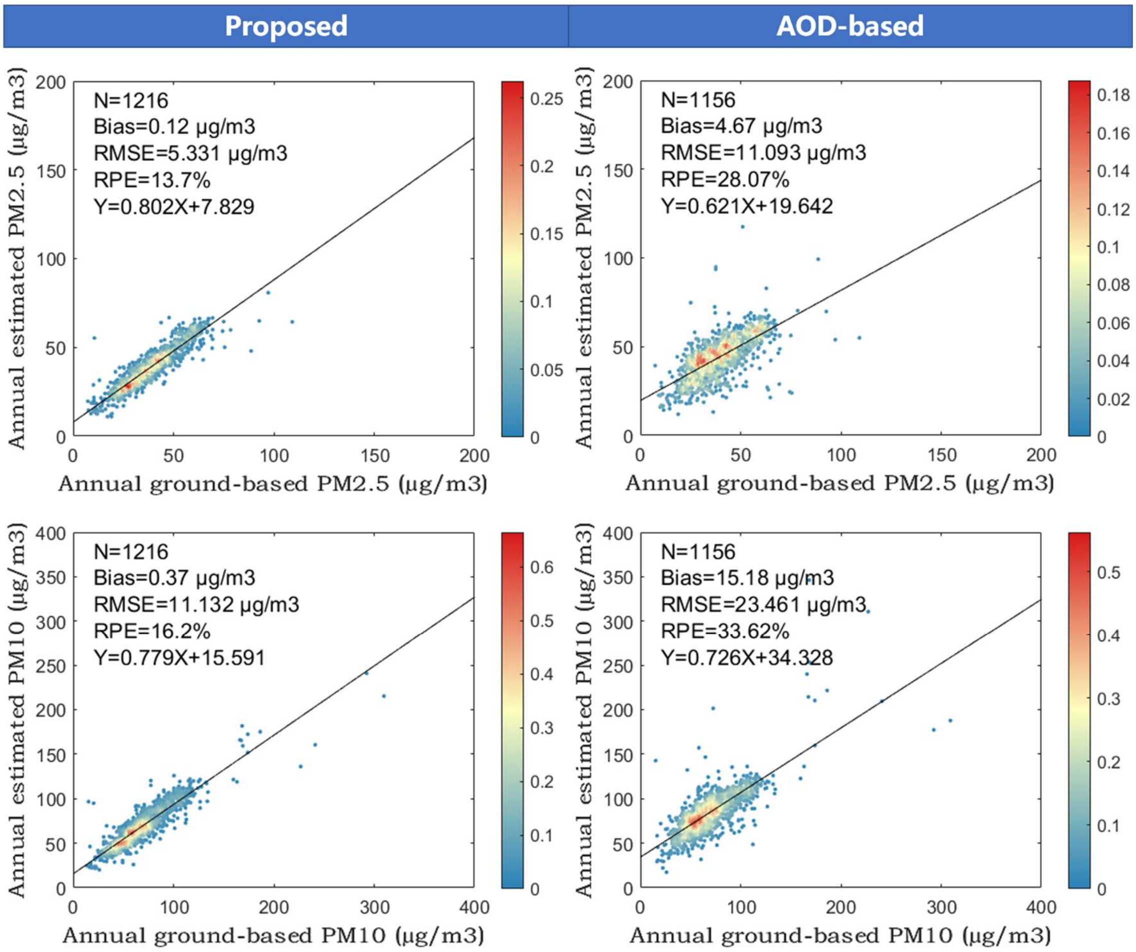

Figure r5: The density scatter plots of the annual validation results for $PM_{2.5}$ and PM_{10} in 2019 over China. The black solid line signifies the fitted line and the color bar denotes the density of samples. Y: annual estimated ambient concentrations of $PM_{2.5}$ and PM_{10}; X: annual ground-based ambient concentrations of $PM_{2.5}$ and PM_{10}.