Reviewer 1 :

General comments:

In the manuscript titled as “Full-coverage 1 km Daily Ambient PM2.5 and O3 Concentrations of China in 2005-2017 Based on Multi-variable Random Forest Model Supplementary materials”, Ma, Ban, Wang et al. develop and present a database of fine-resolution of PM2.5 and O3 across China, during 2005-2017. Generally speaking, the method of the dataset is solid, and the dataset can be useful for studies on health effects of air pollution in China. The results are novel, since rare studies present the estimates of PM2.5 and O3 together. Comprehensively comparing the two key air pollutants in China is of interest. Before publication, only issue needs to be addressed first, is the lack of evaluation on historically hindcasting concentrations before 2013 (see the 1st specific comment).

Response: Thank you for your comments. We have added some analyze to evaluate the historical concentration before 2013. Details are listed as follows.

Specific comments:

As is known, there is no nationwide monitoring network of air pollution before 2013. The product predicted the concentrations in the period of 2005-2012, which is good and novel. However, the accuracy needs to be evaluated first. There are several possible solutions. First, the method of year-by-year cross-validation has been utilized. For instance, when leaving the data in 2013 out, the cross-validation evaluate the corresponding predictions using the measurements from 2014 to 2017. This can somehow evaluate hindcasting accuracy. Second, the authors can also collect the values of PM2.5 and O3 before 2013, from the published literature, and use those values as a referent to evaluate the model. Third, the authors can utilize the monitoring data in specific sites, including US embassy monitors, Hong Kong monitoring networks or Taiwan monitoring networks. Those datasets provided historical concentrations of PM2.5 and O3 for free. I recommend the authors to utilized as least as one out of the three approaches or other appropriate method to evaluate the hindcasting accuracy.

Response: Thank you for your comments. Firstly, we already used the temporal validation to evaluate the credibility in temporal scale. The corresponding results were showed in the Fig. 3 in manuscript. Through built model based on the data of 90% randomly selected date, we found that for $PM_{2.5}$, the daily temporal R^2 was 0.49 in test set, monthly and yearly R^2 were 0.65 and 0.76; for O_3, the temporal R^2 were 0.58, 0.63 and 0.56 in daily, monthly and yearly level, respectively.

Secondly, from 2005 to 2012, only the US embassies in Beijing, Shenyang, Chengdu, Guangzhou and Shanghai had air quality monitoring stations. Embassy set of air quality monitoring site provides the surface concentration of $PM_{2.5}$ monitoring value can be used to verify the accuracy of simulation results before 2013. Therefore, we estimated the $PM_{2.5}$ concentration of the five stations from 2005 to 2017, and conducted a fitting analysis with the measured values of the stations. A total of 10204 samples from 5 sites participated in the verification of historical simulation values, among which the sample size before 2013 was 2489, which was less than the sample size after 2013. The results of fitting analysis (Fig. 1) show that the test-R^2 of simulation results and measured values before 2013 is 0.45, and the slope of fitting line is 0.43, which is lower than the test-R^2 of fitting after 2013 (0.86). While for O_3-8hmax, we did not find the reliable data resource for historical validation. As more data becomes available and shared in the future, we may be able to further validate the historical O_3-8hmax data in this study.

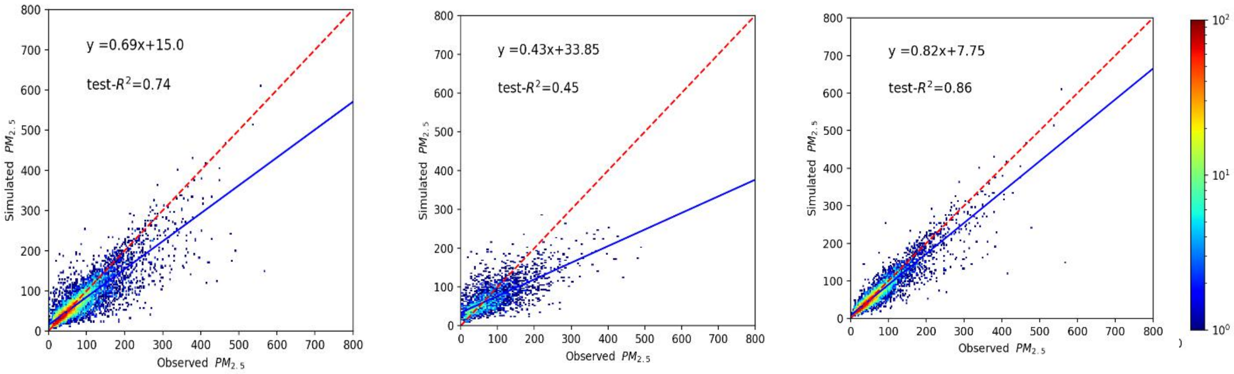


Fig. 1 The density plots of modeling and observed $PM_{2.5}$ concentration in US embassies in Beijing, Shenyang, Chengdu, Guangzhou and Shanghai during 2005-2012

From left to right respectively show the comparison between measured and simulated values of the site from 2005 to 2017 (N = 10204), measured and simulated values of the site from 2005 to 2012 (N = 2489), and measured and simulated values of the site from 2013 to 2017 (N = 7715).

Thirdly, we conducted the year-by-year cross validation for $PM_{2.5}$ and O_3. Because of the large training data, we only built models using training data during 2013 to 2016, and using training data in 2017 as validation (Fig. 2).

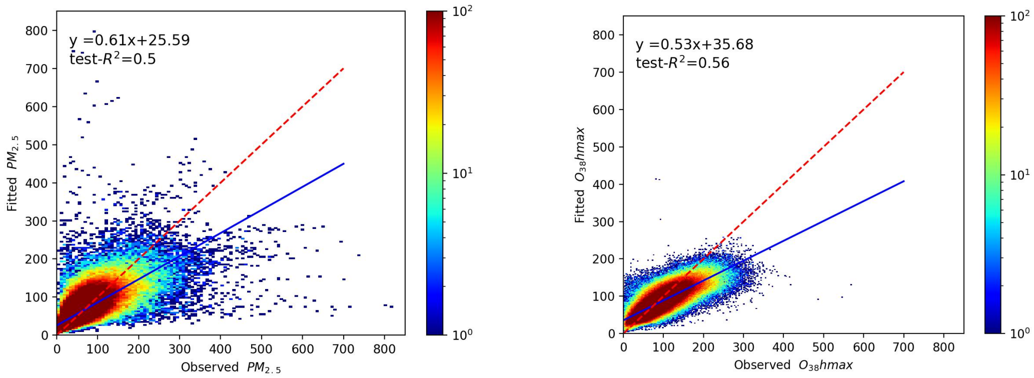


Fig. 2 The density plots of $PM_{2.5}$ (A) and O_3-8hmax in 2017(B)

Lastly, we have sort out the basic situation of relevant simulation research in Tables S6. But few studies provided specific concentrations in article or shared link of modeling data. Therefore, it is difficult for us to compare specific values. However, by comparing the research results, we believe that the results of this study are credible.

One novelty of this study is the fine spatial resolution of 1 * 1 km. I recommend the authors conduct some cross-valuation analyses to show advantages of this novelty. For instance, the authors can aggregate the fine-resolution data into different levels, e.g., 5 * 5 km, 10 * 10 km, or etc., and then conducted cross-validations based on different spatial resolutions.

Response: Thank you for your comments. We used the buffer analysis to selected the standard grid around the monitoring sites within 5 km radius, and average the modeling value in these selected grids, and using these values to compare with the true monitoring values. We used the data of 2017 as an example. It can be seen from the results of the case data that the 1km scale simulation data produced in this study still shows a good degree of fitting with the real point after being fused to the 5km scale, although it is slightly lower than the 1km scale simulation data (Fig. 3).

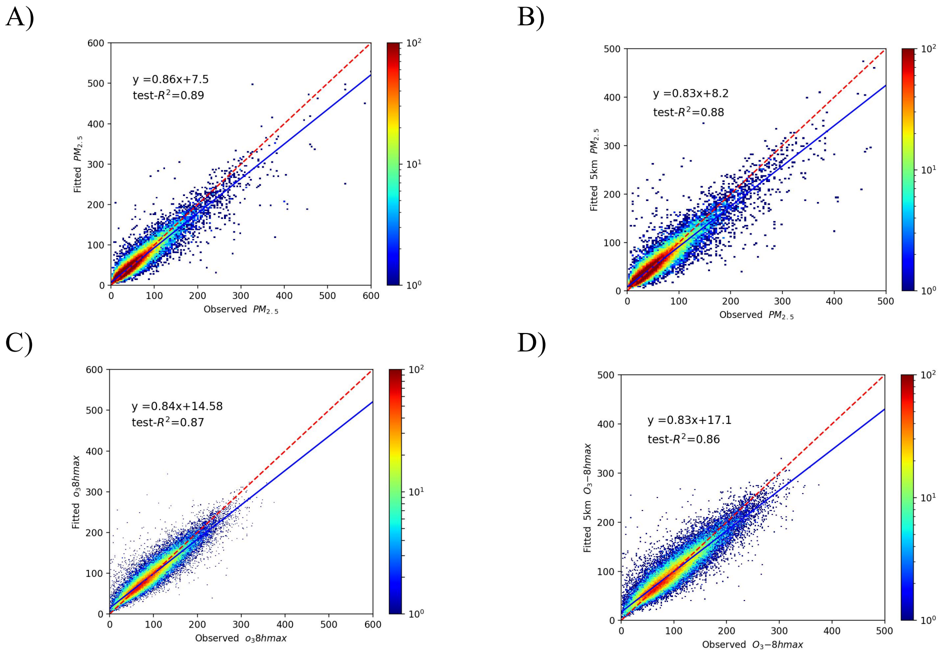


Fig. 3 The density plots of $PM_{2.5}$ and O_3-8hmax in 2017 with 1km and 5km resolution

A and B show the density plots of $PM_{2.5}$ in 2017 with 1km and 5km resolution; C and D show the density plots of O_3-8hmax in 2017 with 1km and 5km resolution.

Technical issues:

I recommend not to use the term, simulation to mention the outputs of the RF models. Maybe, prediction or estimation is appropriate. Simulation is often utilized to refer the direct outputs from the chemical transport models.

Response: Thank you for your comments. We have switched the “simulation” into “estimations”, as well as the “simulate” into “estimate” in the whole text.

The cross-validation results for O3 in daily, monthly, and yearly scale are reported as 0.58, 0.63 and 0.53, respectively. Is there any explain for why the accuracy in monthly scale is higher than 0.53, which is opposite to our expectation. Usually, if we aggregate more estimates, we expect to reduce more random errors and thus improve the accuracy.

Response: Thank you for your comments. We have noticed the lower performance of O_3-8hmax in yearly level at first, and we have carefully checked our simulation models and validation process to confirm this result. Possible reason for the lower yearly level but higher daily level may owing to the natural feature of ambient ozone, that O_3-8hmax, to some extent, represents the "extreme value" of a day; the annual mean value of O_3 concentration represents a more general level of concentration, therefore the relationship between the predictors and O_3 may be erased.

Reviewer 2：

General comments

This study aimed to estimate the 1-km resolution PM2.5 and O3 concentrations with a random forest model covering China. Spatiotemporal variations in PM2.5 and O3 distributions during 2005-2017 were further characterized. The modelling workflow is reasonable; however, the method section lacks some critical information and the source data can hardly support the spatiotemporal resolution of the predictions. The result presentation could be improved and some discussion on data uncertainties are needed.

Response: Thank you for your comments. We would revise the manuscript according to your advices, next are some responses of your comments.

Specific comments

In the abstract section, the author presented the model fitting R2 from "sample-based division method" in line 31-33 but it is not clear if the R2 is from test data or from cross validation. Additionally, the model fitting R2 normally means the R2 during model fitting stage with the model fitting dataset. Please clarify the R2 here and throughout the manuscript.

Response: Thank you for your comments. The "sample-based division method" R^2 means the R^2 from the test data. We would add more explanation about the "R^2" in the abstract. As for the manuscript, we explained R^2 when it first appeared so that readers could understand that R^2 would refer to R^2 of the test set; then we replaced R^2 with test-R^2.

Relevant text: According to our sample-based division method, the daily, monthly and yearly estimations of $PM_{2.5}$ from test datasets gave average model fitting R^2 values of 0.85, 0.88 and 0.90, respectively; these R^2 values were 0.77, 0.77, and 0.69 for O_3-8hmax, respectively. (Line 33-36)

We construct the main model using the training set with a 10-fold cross-validation. Since the data in the test set is not used in the main model, "true model performance" can be verified. The coefficient of determination (R^2) of main model on test set (test-R^2), and the verification indicators of model uncertainty, the root mean square error (RMSE) and mean absolute error (MAE) are calculated for the $PM_{2.5}$ and O_3-8hmax model, respectively. The monthly and yearly test-R^2 are also calculated. (Line 171-177)

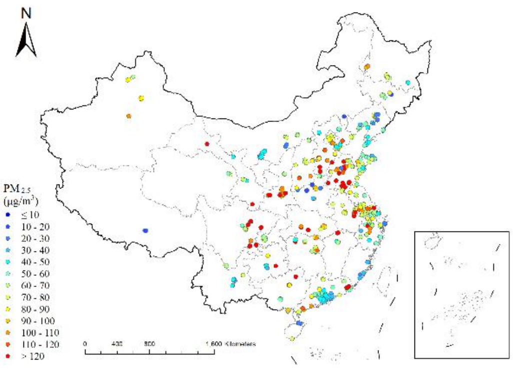
Line 55-56: Why did the high pollution events and unsatisfactory pollution control bring difficulties to capture pollution distribution? Are the pollution level higher than the monitor measurement range?

Response: Thank you for your comments. What we try to explain is that “despite the implementation of control policies, there are still $PM_{2.5}$ pollution events and ozone pollution exist. It is necessary to have simulation data to understand the overall pollution situation”. There are some ambiguities in these sentences, which would be revised.

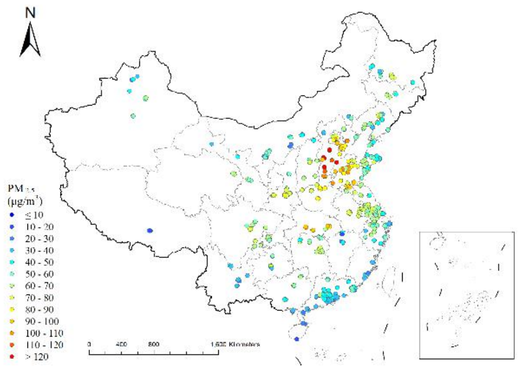
Relevant text: However, the occasional pollution events, as well as the short development history of air quality monitoring network, have brought many difficulties to accurately capture the temporal and spatial patterns of $PM_{2.5}$ and O_3 concentrations. (Line 57-59)

Fig. 1: The number of air quality monitors in China kept increasing during 2013-2015 and there are much more monitors in 2017 compared to 2013. This figure shows the 2017 monitors but describes it as average measurement concentrations during 2013-2017. How to deal with the monitors that are not available in 2013-2014?

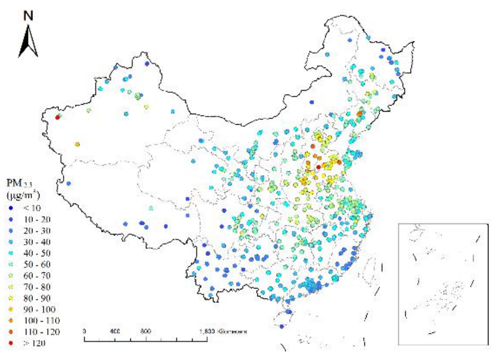
Response: Thank you for your comments. We did not consider your suggestion before, so we made the following modifications: First, we changed the title of the picture in the body to “Station distribution in China and average ground monitoring concentration based on the available data of $PM_{2.5}$ (A) and O_3-8hmax (B) from 2013 to 2017” to avoid possible ambiguity. Secondly, we visualized the yearly distribution of $PM_{2.5}$ and O_3 monitoring values, and placed the figures in supplementary materials (Fig. S1 and Fig. S2) to enable readers to have a deeper understanding of the basic situation of monitoring values. We also put these figures as follows (Fig. 1 and Fig. 2).



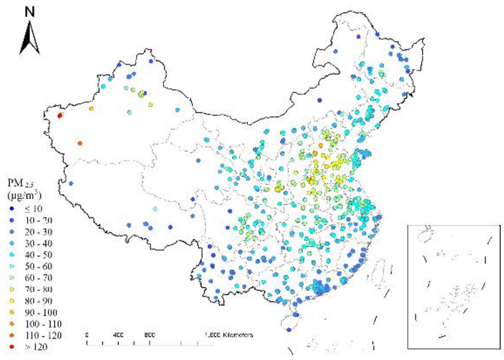
2013 annual



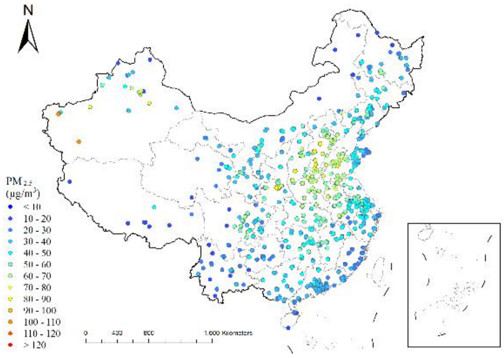
2014 annual



2015 annual

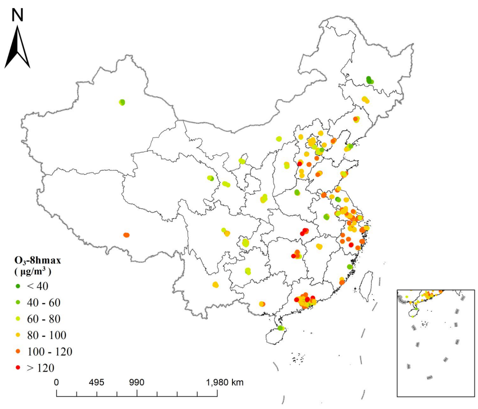


2016 annual

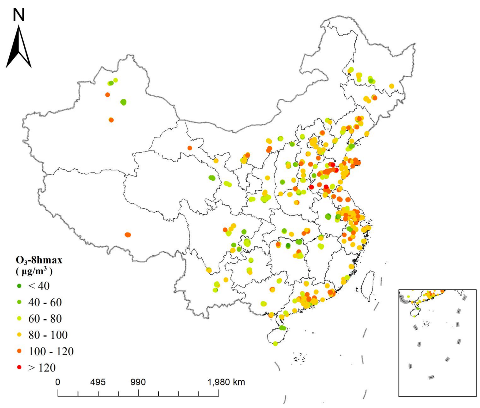


2017 annual

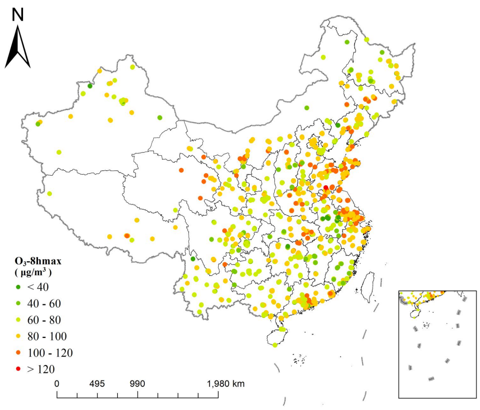
Fig. 1 Station distribution in China and average ground monitoring concentration of $PM_{2.5}$ during 2013-2017



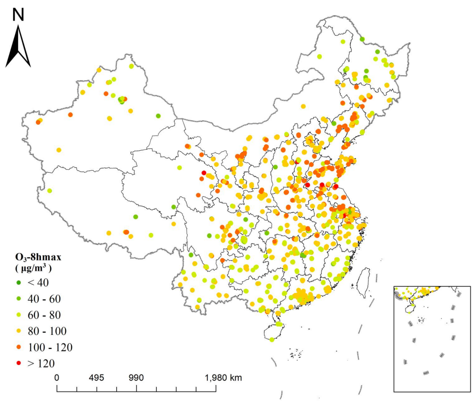
2013 annual



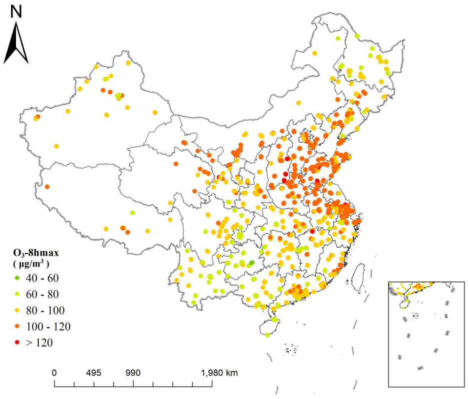
2014 annual



2015 annual



2016 annual



2017 annual

Fig. 2 Station distribution in China and average ground monitoring concentration of O_3-8hmax during 2013-2017

Relevant text: Daily average $PM_{2.5}$ and O_3 daily maximum of 8h-average concentration (O_3-8hmax) monitoring data of 1479 sites in 2013-2017 was obtained (Fig. 1; Fig. S1 and Fig. S2). (Line 91-93)

Line 81-82: the reference Wei et al. 2021, which was cited in the result section, constructed the datasets with longer time series and 1-km resolution. What is the advantage of this work compared to previous works?

Response: Thank you for your comments. First, we produce the modeling data of $PM_{2.5}$ and O_3-8hmax in the same time with high model performance. This allows us to simultaneously understand the temporal trends and spatial characteristic of two major pollutants harmful to health; The homologous simulation data will also avoid possible bias when applied to subsequent epidemiological studies. Second, this study is inconsistent with Wei et al 's research method (Random Forest Model and Space-Time Extra-Trees model), but similar research results have been obtained, which can be mutually verified to a certain extent.

Line 95-96: The resolution is 1-km but the author did not provide the projection information.

Response: Thank you for your comments. We have implemented related information in the manuscript. The projection system is Albers Conical Equal Area Projection. Details about the projection are as follows:

Projected Coordinate System: 1

Projection: Albers

False_Easting: 0.00000000

False_Northing: 0.00000000

Central_Meridian: 105.00000000

Standard_Parallel_1: 25.00000000

Standard_Parallel_2: 47.00000000

Latitude_Of_Origin: 0.00000000

Linear Unit: Meter

Geographic Coordinate System: GCS_WGS_1984

Datum: D_WGS_1984

Prime Meridian: Greenwich

Angular Unit: Degree

Relevant text: The coordinate system of the grid is WGS-84; and the projection of the grid is the Albers Conical Equal Area Projection. (Line 94-96)

Figure 3 and Figure 4: The PM2.5 ranges of the yearly plots are much smaller than those of the daily plots. Please shrink the x- and y-axis.

Response: Thank you for your comments. We have revised the Fig. 3 and Fig. 4 in the manuscript. The figures are also showed as follows (Fig. 3 and Fig. 4).

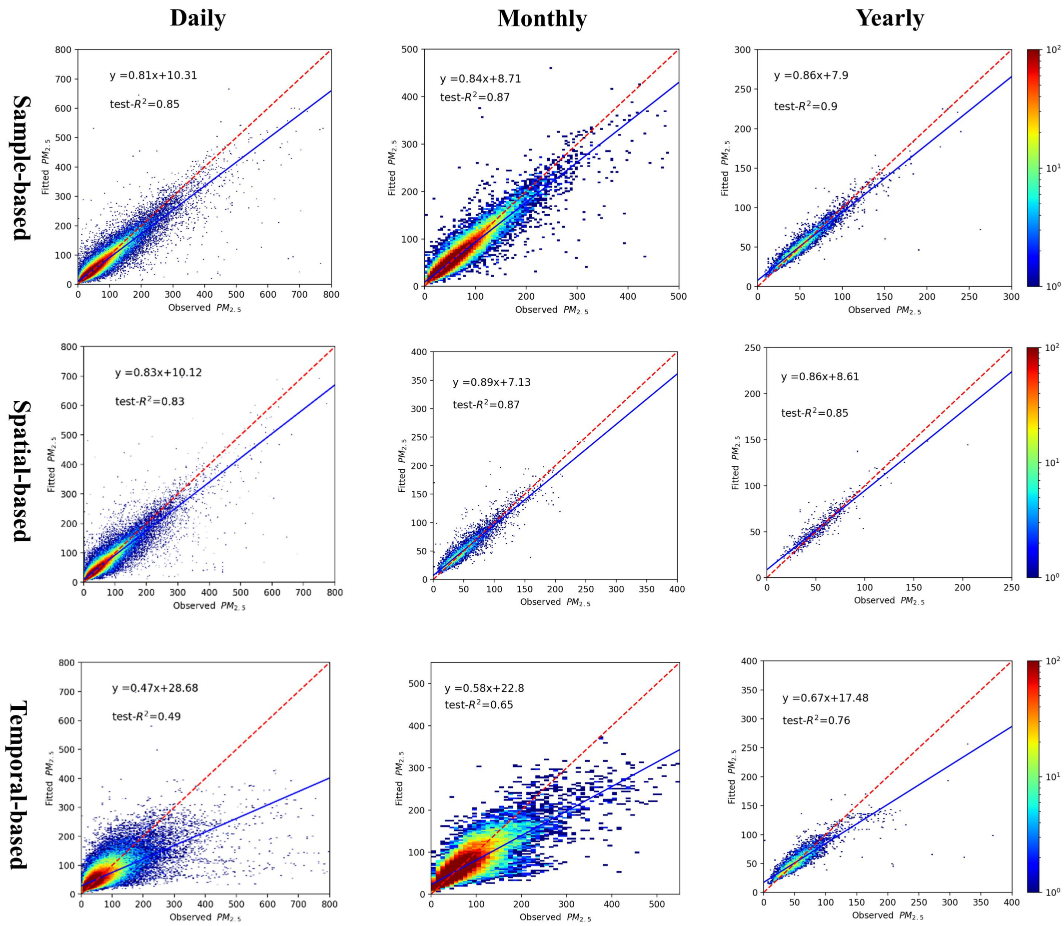


Fig. 3 The density plot of $PM_{2.5}$ model

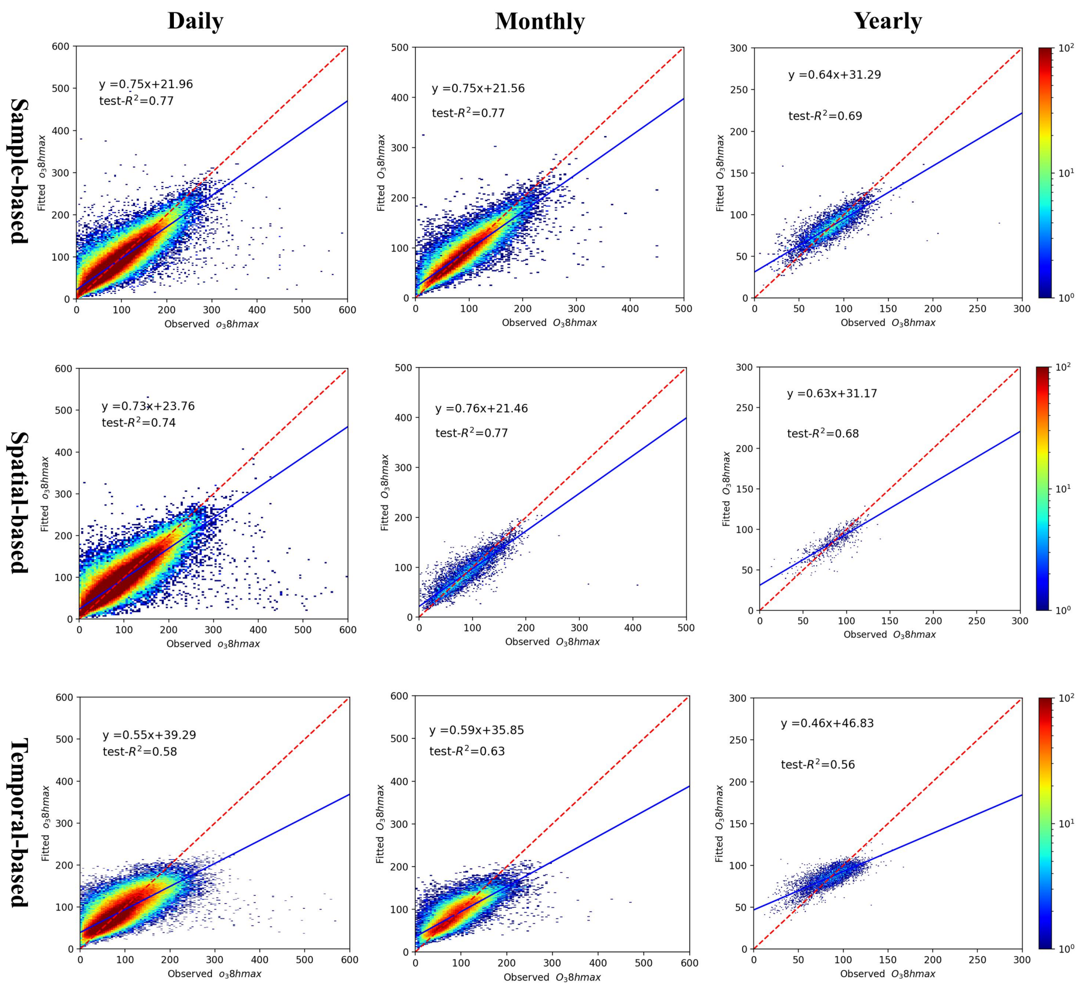


Fig. 4 The density plot of O_3-8hmax model

Line 105-107: Why did the author use Aqua AOD but not Terra AOD for PM2.5 modeling? What percent of satellite AOD are missing? How did the author deal with the missing satellite retrievals to get full-coverage daily dataset? And what is the performance of the gap-filling method?

Response: Thank you for your comments. Both Aqua AOD and Terra AQD are the best aerosol optical depth products for near-real-time aerosol data assimilation. A study in the U.S. showed that Aqua AOD and Terra AOD showed similar coverage rates, and the combination of Aqua AOD and Terra AOD significantly improved data coverage, thereby improving the accuracy of $PM_{2.5}$ simulations[1]; this maybe the focus of further researches. In addition, a study based on Beijing showed that the correlation between Aqua AOD and $PM_{2.5}$ was higher than Terra AOD: The R^2 of

Terra AOD and $PM_{2.5}$ was 0.53 at nine urban sites and 0.34 at three suburban and sub-suburbs sites. The average R^2 of Aqua AOD and $PM_{2.5}$ was 0.62 at 9 stations in the urban area, and 0.53 at 3 stations in the suburbs and sub-suburbs[2]. In addition, the research team has previously carried out relevant gap studies based on Aqua AOD data[3], and has certain experience on data basis and technology basis. Therefore, Aqua AOD data was finally selected.

Due to MODIS satellite orbit interval, cloud cover, high reflectivity (such as snow and ice cover) and limitations of different inversion algorithms, AOD data have a high missing rate. Especially in the west of China or in winter, the data coverage rate is even less than 10%[4], which is difficult to be directly used for model simulation. Therefore, it is necessary to carry out AOD data supplement. In this study, interpolation was considered to complement the missing AOD. The inverse distance weight interpolation method refers to that the similarity of two objects decreases with the increase of the distance between them. The distance between the interpolation points and the sample point is taken as the weight to carry out the weighted average. The closer the sample is to the interpolation point; the more weight is given to it. Due to the large area of study and the large amount of data, the inverse distance weight interpolation method is less difficult to implement than the existing interpolation methods (such as Kriging interpolation), with intuitive effect and high efficiency, and can quickly and comprehensively complete the missing AOD. AOD is filled by the IDW third-party library in Python. In this study, the original data was processed in batch by ENVI5.3+IDL remote sensing professional processing software. After geometric correction, splicing and cutting steps, it was processed into WGS84 coordinate system and TIF data format. ArcPy was used to extract the values to the standard grid, and then interpolation was carried out to obtain the national standard grid data of aerosol optical thickness. The simulation effect is good (Fig. 5). The brief introduction of process of AOD have been added into the Methods.

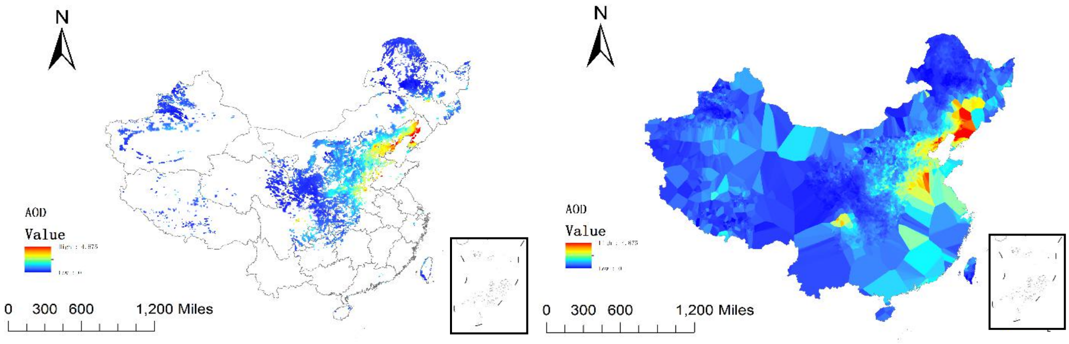


Fig. 5 MYD04_3K coverage in China and results after interpolation on January 1, 2017

Relevant text: Briefly, most of the model variables are processed into 1km×1km resolution based on the standard grid using interpolation methods (such as inverse distance weighted and bilinear algorithm) in ArcGIS 10.2 and Python 2.7. For example, AOD is processed by ENVI 5.3+IDL and extracted into standard grid using ArcPy, then the inverse distance weighted interpolation is carried out to obtain the 1km×1km resolution data. (Line 115-120)

The 2*2.5 degree GEOS-Chem simulations were used for 1-km resolution O3 modeling. Since the spatial resolution of GEOS-Chem simulation was too coarse compared to the O3 prediction and it ranked the second most important predictor, I doubt if the prediction could truly reflect O3 variations at local scale. Actually, none of the predictors for O3 modeling provides sufficient spatiotemporal information on variations at the 1-km daily resolution. The design of the O3 model is not solid.

Response: Thank you for your comments. GEOS-Chem model output is an important feature to show the process of ozone formation and dissipation, however, the limitation caused by spatial resolution of GEOS-Chem is inevitable. In the future, more refined and accurate data of predictors can effectively improve the accuracy of $PM_{2.5}$ simulation. At present, random forest model is one of the statistical methods that can precisely capture the nonlinear relationship between predictors and ambient ozone. Furthermore, the model features in this study are consist with the formation mechanism of ambient ozone, and have been used in previous modeling studies. The results of varied validation method also proved the credibility of modeling data.

Line 113-114: The gridded GDP data needs a citation. The GDP data only cover year 2005 and 2010, how did the author assign GDP of other years? Similarly, the road data is of year 2016 but the road map of year 2005 could be considerably different from the road map of year 2016. How did the author consider this issue?

Response: Thank you for your comments. The specific information of data resource is showed in Table S1. The road map and GDP data are collected from Resource and Environment Science and Data Center (http://www.resdc.cn), a reliable platform for obtaining geographic data resources at the national level; and the data we have is the best we can get. According to the feature importance in the study, we found the GDP and road map did not show great influence in both $PM_{2.5}$ model (0.007 for GDP and 0.01 for road map) and O_3-8hmax model (1.18% for GDP and 1.8% for road map). The impact of mismatched data years may be small.

Figure 4: the slopes of the daily and monthly plots are lower than 0.8, and the slopes of the yearly plots are lower than 0.7, indicating system bias.

Response: Thank you for your advice. We thought the possible reason is that due to the indicator we chose, O_3-8hmax, to some extent, represents the "extreme value" of a day; so, when it is calculated to the "annual mean" scale, the relationship between the predictors and O_3 may be erased.

Section 3.2: How to calculate the feature importance and what does the “Value” in Table S4 mean? Why did the Value is in digital number in Table S4-1 but in percentage in Table S4-2. The values of some predictors, e.g. High speed road and Railway, are very low. Why did the author keep them in the model? The author used a whole section to discuss the importance of predictors, thus the Table S4-1 and Table S4-2 could be move to the main text.

Response: Thank you for your advice. The “Value” means the feature importance, produced by random forest model, showed the importance of model features for the modeling of $PM_{2.5}$ and O_3. We modified the header and unified the expression of the results of the two tables. We hope that variables in the model can represent the formation mechanism of $PM_{2.5}$ and ozone in a relatively complete way. Furthermore, considering that low-importance variables still contribute to the model and the complexity of the random forest model is not high, we chose to retain all variables will not affect the model training difficulty and model running speed. Our previous study used the same strategy and achieved high model performance[5]. In the future, if near-real-time simulation is required, we will consider setting up conditions to screen model variables. As for the position of the importance ranking table, since we have made a detailed summary in the main body, and considering the length of the article, we still choose to place the importance ranking table in the supplementary materials.

Figure 6: Figure S1 and Figure S2: This study produce 1-km PM2.5 and O3 data products, but only showed the national map and the quality of these figures could not reflect any local scale characteristics. Please zoom in at key regions to give the readers more details.

Response: Thank you for your advice. We have implemented the local scale map of Beijing-Tianjin-Hebei, Yangtze River Delta and Pearl River Delta (Fig. 6 and Fig. 7) into the Supplementary, and the temporal and spatial distribution trend is also explained.

Relevant text: In key pollution areas, with the implementation of various air pollution prevention and control policies, $PM_{2.5}$ levels in the Beijing-Tianjin-Hebei region have dropped the most, but the overall concentration levels are still higher than those in the Yangtze River Delta and Pearl River Delta (Fig. S4). (Line 286-289)

The Beijing-Tianjin-Hebei region has shown an obvious upward trend since 2013; while the Pearl River Delta region change trend is not obvious (Fig. S6). (Line 294-295)

This spatial pattern barely changed during 2005-2017 (Fig. S3 and Fig. S5), but the temporal trend showed spatial characteristic (Fig. 6; Fig. S4 and S6). For $PM_{2.5}$ concentration, the key pollution areas were severely polluted during 2005-2013. The air pollution control measures of these regions were strict during 2013-2017, thus the decline was obvious, especially for the Beijing-Tianjin-Hebei region. For O_3-8hmax concentration, the growth rate was not obvious (except for the eastern part of Hubei Province) during 2005-2013. However, after 2013, there was a clear upward trend across the country, especially in the northern China.(Line 328-336)

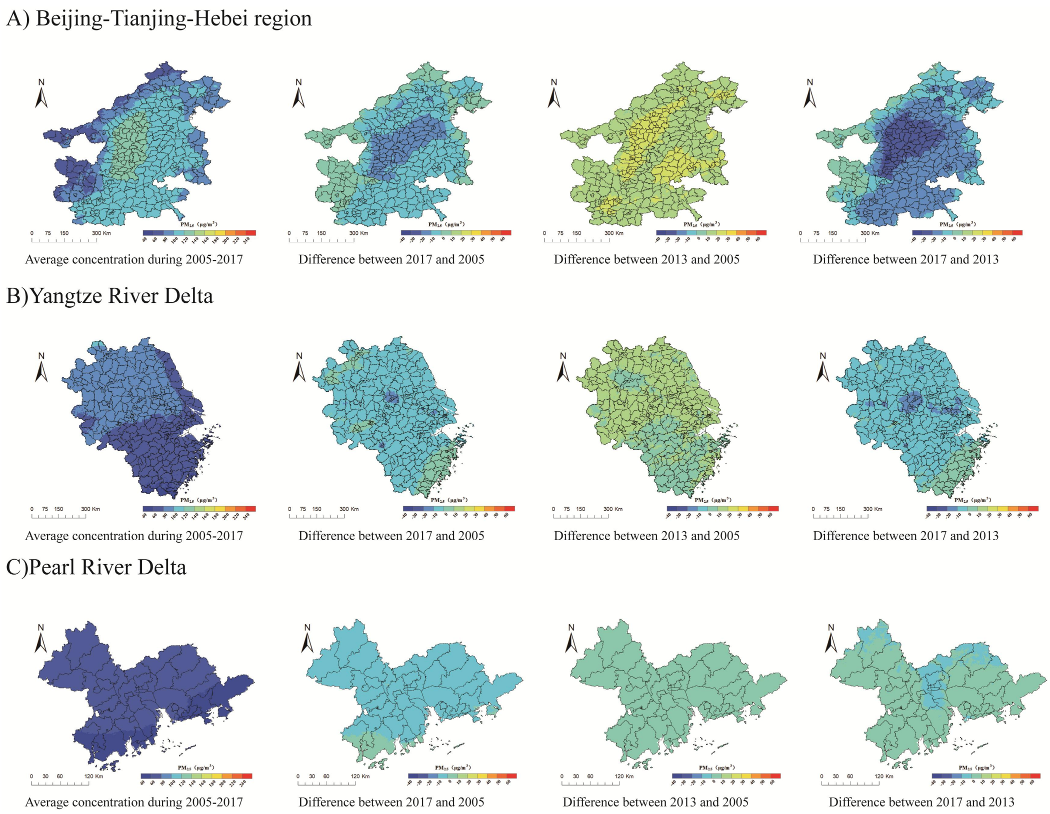


Fig. 6 The Simulated annual mean and difference of $PM_{2.5}$ concentration in A) Beijing-Tianjin-Hebei region, B) Yangtze River Delta, and C) Pearl River Delta during 2005 to 2017

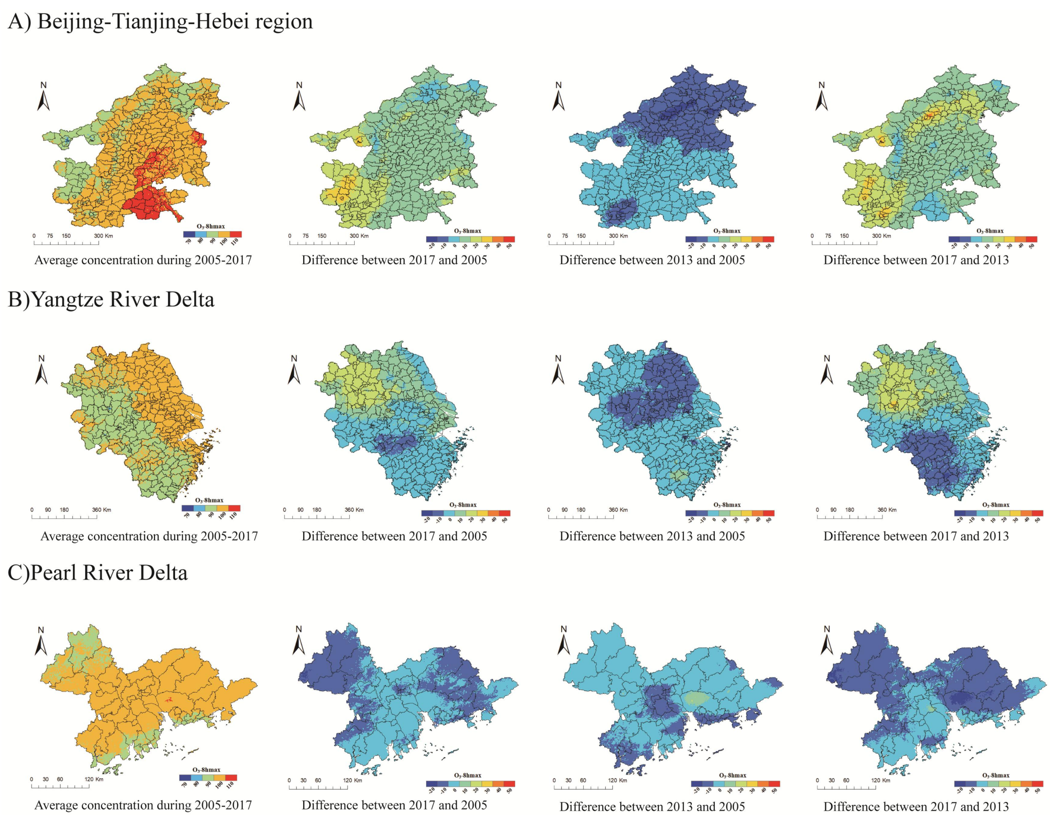


Fig. 7 The Simulated annual mean and difference of O_3-8hmax concentration in A) Beijing-Tianjin-Hebei region, B) Yangtze River Delta, and C) Pearl River Delta during 2005 to 2017

Figure 6: Figure S1, and Figure S3: The spatial patterns over the west China are weird. Figure S2: Please explain the extremely low O3 concentrations over Tibet on the 2016 map and the weird spatial pattern in West China on the 2017 map.

Response: Thank you for your advice. First, due to the Due to the sparseness of monitoring stations in northwest China, the variation trend of $PM_{2.5}$ and O_3 concentration is relatively poorly captured by the model. The sparseness of monitoring sites in Northwest region caused the uncertainty in modeling data. This trend is also reflected in other studies, some of which choose to cut out the sparse areas in the northwest site[6]. For the sake of data integrity, we still retain data from these areas. In the future, with the improvement of monitoring network, the simulation performance in northwest China will be improved effectively. We implemented a map to display the true concentration and modeling concentration in 2016 and 2017 for O_3-8hmax (Fig. 8). It can be found that the model performance is better in the area with monitoring stations, but the uncertainty is still large in the vast area without monitoring stations. Furthermore, we have adjusted Fig. S10 to make it better.

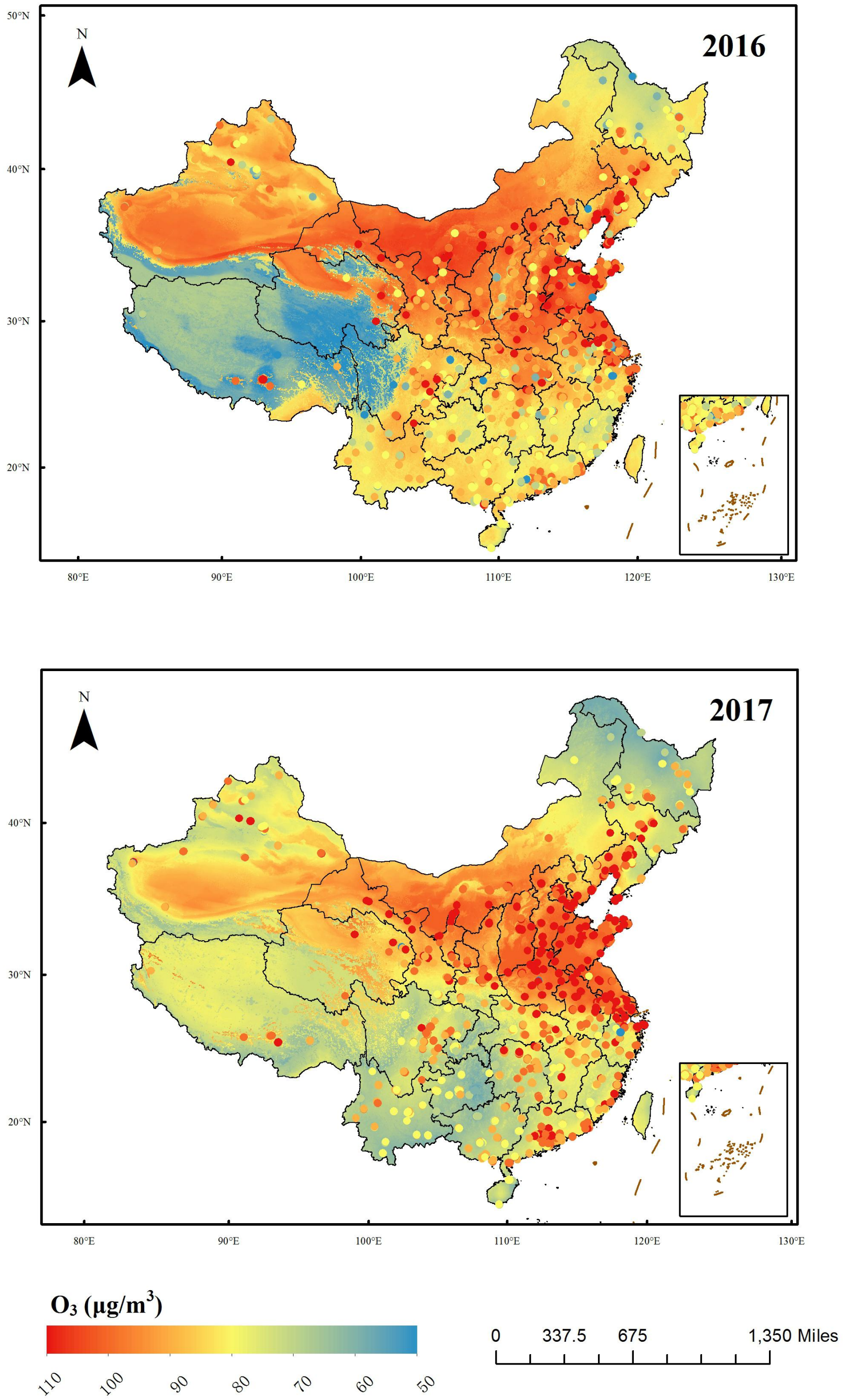


Fig. 8 The modeling and monitoring O_3-8hmax concentrations in China in 2016 and 2017

Reference

1. Kim, M., Zhang, X., Holt, J. B. & Liu, Y. Spatio-Temporal Variations in the Associations between Hourly PM2.5 and Aerosol Optical Depth (AOD) from MODIS Sensors on Terra and Aqua. *Health* **05**, 8–13 (2013).
2. Wang, W., Zhang, C., Zang, Z., Wang, T. & You, W. Comparative analysis between hourly PM2.5 concentration and MODIS 3 km aerosol optical depth derived from Terra and Aqua satellites in Beijing. *Journal of the Meteorological Sciences* **37**, 93–100 (2017).
3. Zhao, C. *et al.* High-resolution daily AOD estimated to full coverage using the random forest model approach in the Beijing-Tianjin-Hebei region. *Atmospheric Environment* **203**, 70–78 (2019).
4. Liu, Z., Xie, M., Tian, kun & Xie, xiaoxiao. Classification of PM2.5 for natural cities based on co-Kriging and head/tail break algorithms. *J Tsinghua Univ (Sci & Technol)* **57**, 555–560 (2017).
5. Ma, R., Ban, J., Wang, Q., Zhang, Y. & Li, T. Random Forest Model based Fine Scale Spatiotemporal O3 Trends in the Beijing-Tianjin-Hebei region in China, 2010 to 2017. *Environmental Pollution* 116635 (2021).
6. Zhang, X. Y., Zhao, L. M., Cheng, M. M. & Chen, D. M. Estimating Ground-Level Ozone Concentrations in Eastern China Using Satellite-Based Precursors. *IEEE Trans. Geosci. Remote Sensing* **58**, 4754–4763 (2020).