

Earth Syst. Sci. Data Discuss., author comment AC2
<https://doi.org/10.5194/essd-2021-296-AC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Runmei Ma et al.

Author comment on "Full-coverage 1□km daily ambient PM_{2.5} and O₃ concentrations of China in 2005–2017 based on a multi-variable random forest model" by Runmei Ma et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-296-AC2>, 2021

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

Response to Specific Comments:

1. 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 PM_{2.5} 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 PM_{2.5} 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 R² was 0.49 in test set, monthly and yearly R² were 0.65 and 0.76; for O₃, the temporal R² 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 in Supplement of Response) show that the test-R² 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-R² of fitting after 2013 (0.86). While for O₃-8hmax, 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 O₃-8hmax data in this study.

Thirdly, we conducted the year-by-year cross validation for PM_{2.5} and O₃-8hmax. 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. The test-R² for PM_{2.5} and O₃-8hmax are 0.50 and 0.56 (Fig. 2 in Supplement of Response).

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.

2. 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 in Supplement of Response).

Response to Technical Issues:

1. 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.

2. 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 O₃-8hmax 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 O₃-8hmax, to some extent, represents the "extreme value" of a day; the annual mean value of O₃ concentration represents a more general level of concentration, therefore the relationship between the predictors and O₃ may be erased.

Please also note the supplement to this comment:

<https://essd.copernicus.org/preprints/essd-2021-296/essd-2021-296-AC2-supplement.pdf>