

## Comment on **essd-2021-296**

Anonymous Referee #1

---

Referee comment on "Full-coverage 1 km daily ambient PM<sub>2.5</sub> and O<sub>3</sub> 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-RC1>, 2021

---

### General comments:

In the manuscript titled as "Full-coverage 1 km Daily Ambient PM<sub>2.5</sub> and O<sub>3</sub> 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 PM<sub>2.5</sub> and O<sub>3</sub> 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 PM<sub>2.5</sub> and O<sub>3</sub> 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 1<sup>st</sup> specific comment).

### 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 PM<sub>2.5</sub> and O<sub>3</sub> 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 utilize at least one out of the three approaches or other appropriate method to evaluate the hindcasting accuracy.

- One novelty of this study is the fine spatial resolution of 1 \* 1 km. I recommend the authors conduct some cross-validation 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 conduct cross-validations based on different spatial resolutions.

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.
- 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.