Interactive comment on "Estimation of $PM_{2.5}$ Concentration in China Using Linear Hybrid Machine Learning Model" by Song et al.

Anonymous Reviewer

May 21, 2021

The study by Song et al. presents a linear hybrid machine learning model to estimate regional $PM_{2.5}$ distributions from Himawari-8 AOD observations. In the manuscript, the authors stated that the proposed RGD-LHMLM method outperforms than three conventional machine learning methods and can perform accurate estimations.

The topic fits well to the aims and scopes of AMT. Machine learning based methods have been widely used to estimate near-surface $PM_{2.5}$ using satellite AOD observations. As the authors stated in the manuscript, there have been a lot of studies about regional estimation of $PM_{2.5}$ over China. However, in my opinion, the presented material in this study do not sufficiently prove that the proposed method is superior to the other three conventional methods and that it can be used to "perform the seasonal evolution of pollutants ", "help control the local pollution", and "fit the $PM_{2.5}$ in the future". I would expect that the functionality of this hybrid method should be logic and provable, which means that the method has indeed learned and generalized mostly from the satellite AOD observations so that it can be used to estimate/predict unexpected $PM_{2.5}$ features in the future. Unfortunately, in the current manuscript, the authors only use the satellite and ground measurements in 2019 and do not specify any reason why the 2019 data can be considered to be representative. Or, is the focus of this study only on investigating spatiotemporal distributions of $PM_{2.5}$ concentration during 2019? If so, the scientific meaning should be addressed.

In Sections 2 and 3, information about training data quality and its error propagation, and details of the mixed model are missing. Without these details, I cannot justify the model performance. The reasons are given in Section 1.

Therefore, I would not recommend a publication based on the current manuscript. Besides, I do have a number of concerns that require feedback (see Section 1 below).

1 Specific comments

- The paper does not provide enough evidence to support the major conclusions. The proposed method does not have generality in terms of target period as the training relies fully on the Himawari-8 AOD data over 2019. What about for the $PM_{2.5}$ estimation in some other years? To have a completely new training? Since the authors did not perform any $PM_{2.5}$ estimation for other years, I'd like to ask whether the training data already includes all possible cases between satellite AOD and ground $PM_{2.5}$. Even if by including more satellite AOD datasets over a longer period, it can still be questionable whether the selected training data are considered to be representative.

- Section 2: Please include information about data quality of all datasets used for training (e.g., satellite AOD, ground-based data, meteorological data). The current training assumes that Himawari-8 AOD and ground $PM_{2.5}$ data are true values, which in reality, is not true. Thus, please discuss how much impact of their data quality on the model performance in a quantitative way, i.e., what is the error propagation of these training data?

- These machine learning based models are sort of "black boxes", which means that it would seem unclear what a physical relationship between input and output are learned, particularly to readers who are not familiar with $PM_{2.5}$ estimation. I would suggest to reformulate the beginning of Section 3 by adding mathematical explanation for such context.

- Section 3: Please specify explicitly the input/output of the training(s).

- Section 3: Please describe in detail the linear combination of the three optimal sub-models.

- Page 8, Line 13: According to Table 1, I do not notice any "significant" improvement from an individual sub-model to a linear-mixed model. I would prefer to say slightly improved, as can be seen also from Figure 3.

- Section 4: The current manuscript only discusses the monthly performance of the linear-mixed model. But as far as I know, the usage of geostationary data such as Himawari-8, is especially beneficial to improving the understanding of daily variation of $PM_{2.5}$. If this study focuses solely on the monthly/seasonal variation, why not use MODIS AOD data over a longer period?

- Figure 5: It seems that the estimated $PM_{2.5}$ are in general lower than the "true" values. Is this underestimation pattern related to Himawari-8 data? Please expand the relevant discussion.

- Figure 6: Please include importance of input parameters to DNN as well.
- Section 4: An error characterization of model estimation is missing. Please discuss (quantitatively if possible) error contributions of the input parameters (at least including dominant error sources) to the final output.
- Page 15, Line 19: Any examples of "other satellite data"? If other satellite observations are considered, how do you optimize the model training, as the current training is only based on Himawari data.