Review result of "Estimation of $PM_{2.5}$ Concentration in China Using Linear Hybrid Machine Learning Model." (AMT-2021-64) by Song et al.

Response to RC3:

referee's comments are given in blue,

our responses are given in red.

RC3: The authors presented a new perspective to derive hourly PM2.5 concentrations from Himawari-8 satellite in China by combing different AI methods. This study is overall good, and the results are generally well presented.

Response: We would like to thank the editor and referee for carefully reading the manuscript and providing detailed and constructive comments, which have helped a lot in improving the manuscript. We quote each comment below, followed by our response.

RC3: My first concern is that the authors used all the data samples collected at the same locations having ground-based measurements using the cross-validation method, but the $PM_{2.5}$ predictions are not evaluated at locations where ground-based measurements are unavailable. Thus, I suggest adding an additional validation to test the spatial prediction ability of your model based on the monitoring stations using the cross-validation method.

Response: We strongly agree with the comment. We have added the additional validation based on the monitoring stations. The results are shown in Fig. 3 (E), with a decrease in accuracy. In future studies, therefore, we should add better spatial predictor features.

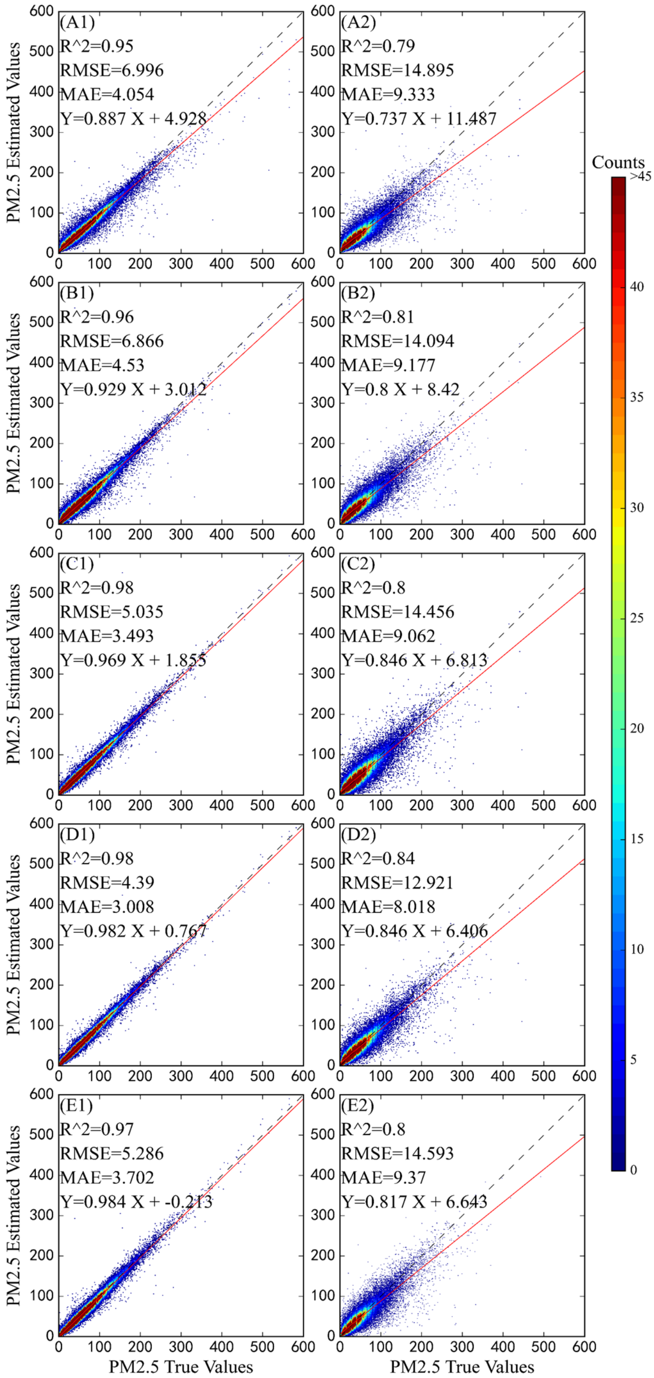

Figure 3 Accuracy of model Fitting and Validation (A: RF, B: GBRT, C: DNN, D: RGD-LHMLM (Based on sample), E: RGD-LHMLM (Based on site))

RC3: My other concern is that the purpose of this study is to derive hourly $PM_{2.5}$ concentrations from geostationary satellites. However, the spatial analysis is performed on a monthly scale (Section 4.3), which will largely reduce the sense of the current study. Thus, it is suggested to add more analysis on PM diurnal variations across China.

The advantage that AHI can provide high temporal resolution data is also discussed, but for some reasons it was not included in the previous version of the manuscript. In the revised manuscript we have added this content. The results are shown in the figure below.

Figure 6 shows the scatterplot fitted with the inversion results of the mixed model from 9:00-17:00 Local Time. The model R^2 ranged from 0.556 to 0.88 at different times. Except for 17:00 when the model had the worst performance, the model R^2 exceeded 0.7 at other times, indicating that the model had a good performance. The optimal performance time is 13:00, R^2 is 0.88. According to the results, the hourly differences in model performance were significant.

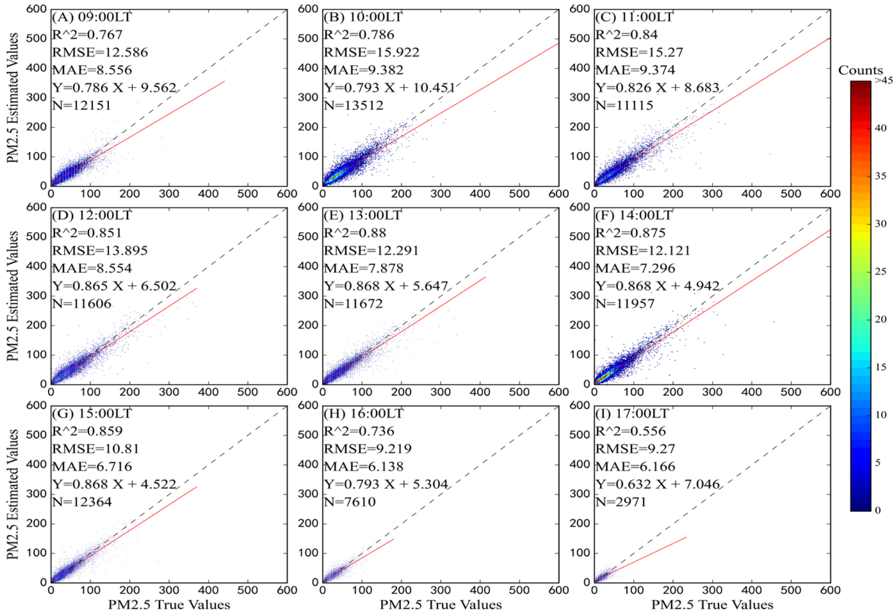

Figure 6 Hourly validation of model performance

The temporal distribution of $PM_{2.5}$ is shown in Figure 10, The $PM_{2.5}$ concentration began to rise from 9:00, and peaked at 55.65μg/m3 between 10:00 and 11:00 every day. After that, it maintained a high concentration until 15:00, and began to decrease. In the most polluted areas of China, the peak concentration of $PM_{2.5}$ can reach 85.05μg/m^3, while the peak in the less polluted areas is only about 40μg/m^3. On a national scale, daily $PM_{2.5}$ concentrations fluctuate little.

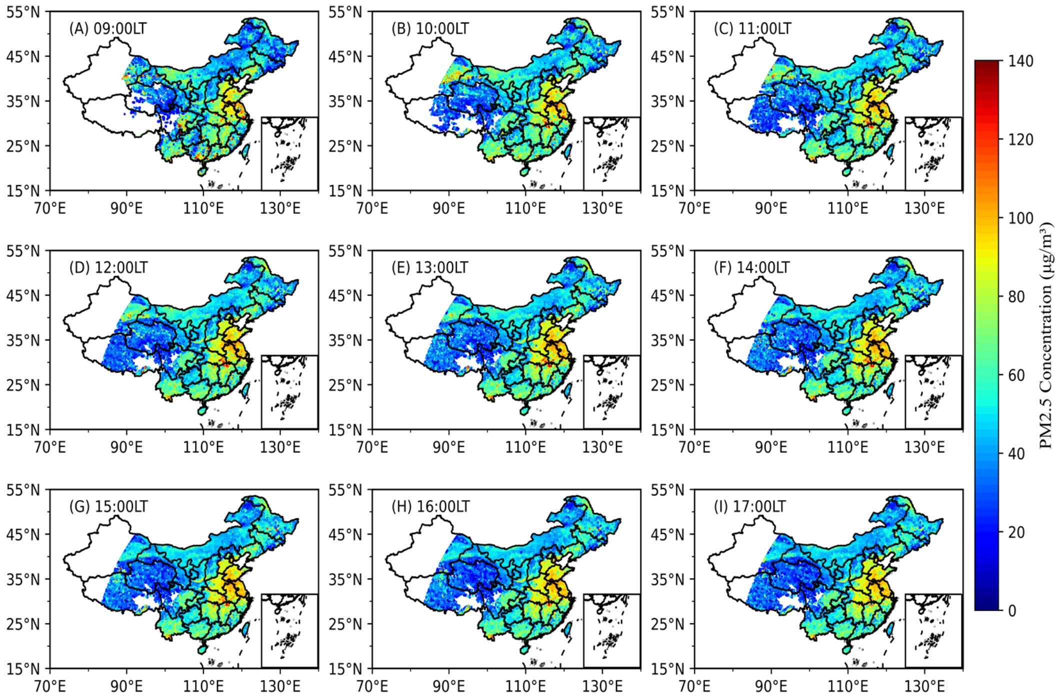

Figure 10 Hourly distribution of $PM_{2.5}$ in China in 2019

RC3: The authors are suggested to update the literature by summarizing more recent studies on $PM_{2.5}$ estimations using sun-synchronous and geostationary orbit satellites, especially those focusing on the whole of China. Below references may help you found more information on various recent studies to help enrich your study.

Section 2.2: Line 15, Reference for Himawari-8 aerosol algorithm is needed.

Line 17: Below references provide a more comprehensive evaluation of Himawari-8 aerosol products in China.

Section 2.3: Reference for ERA5 reanalysis is needed.

References for these traditional ML or DL methods are needed.

Response: Many thanks for the references that were provided to our paper.

We have included it in the revised manuscript.

RC3: Lines 5-9: It is not clear to me how to determine the weight coefficients, and could you add more descriptions?

Response: The coefficient is determined by multiple linear regression model. Firstly, we use three sub-models to calculate the predicted value under the corresponding model. Then, multiple linear regressions are performed between the calculated predicted values and the label values in the original data. Finally, the output coefficients and intercepts of the multiple linear regression model are taken as the parameters of the weight coefficients.

RC3: Section 4.2.2: How about the accuracy of PM2.5 estimations for different hours?

Response: Figure 6 shows the scatterplot fitted with the inversion results of the mixed model from 9:00-17:00 Local Time. The model R^2 ranged from 0.556 to 0.88 at different times. Except for 17:00 when the model had the worst performance, the model R^2 exceeded 0.7 at other times, indicating that the model had a good performance. The optimal performance time is 13:00, R^2 is 0.88. According to the results, the hourly differences in model performance were significant.

RC3: Page 11, Lines 12-15, Page 12, and Page 13, Lines 1-4: May move to a new separate Discussion section.

Response: This is a very good comment, and we have adjusted it in the revised manuscript. The part pointed out by the Referee #3 has been taken as a separate subsection.

RC3: How about your model compared with those developed in previous studies using the Himawari-8 AOD products in China?

Response: We have compared other studies with our own and listed the results in Table 1:

Table 1

Model	R^2	RMSE	MAE	Reference
Stacking model	0.85	17.3	10.5	(Chen et al., 2019)
Two-stage random forests (YRD)	0.86	12.4	/	(Tang et al., 2019)
LME (BTH)	0.86	24.5	14.2	(Wang et al., 2017)
GTWR	0.78	20.10	/	(Xue et al., 2020)
STLG	0.85	13.62	8.49	(Wei et al., 2021)
RGD-LHMLM	0.84	12.92	8.01	This paper

According to the result of the table 1, the accuracy of our model is similar to other models, both of which can better complete the estimation of $PM_{2.5}$.

References

Chen, J. P., Yin, J. H., Zang, L., Zhang, T. X., and Zhao, M. D.: Stacking machine learning model for estimating hourly PM2.5 in China based on Himawari 8 aerosol optical depth data,Sci Total Environ, 697,https://doi.org/10.1016/j.scitotenv.2019.134021, 2019.

Tang, D., Liu, D. R., Tang, Y. L., Seyler, B. C., Deng, X. F., and Zhan, Y.: Comparison of GOCI and Himawari-8 aerosol optical depth for deriving full-coverage hourly PM2.5 across the Yangtze River Delta,Atmos Environ, 217,https://doi.org/10.1016/j.atmosenv.2019.116973, 2019.

Wang, W., Mao, F. Y., Du, L., Pan, Z. X., Gong, W., and Fang, S. H.: Deriving Hourly PM2.5 Concentrations from Himawari-8 AODs over Beijing-Tianjin-Hebei in China,Remote Sens-Basel,

9,https://doi.org/10.3390/rs9080858, 2017.
Wei, J., Li, Z., Pinker, R. T., Wang, J., Sun, L., Xue, W., Li, R., and Cribb, M.: Himawari-8-derived diurnal variations in ground-level PM2.5 pollution across China using the fast space-time Light Gradient Boosting Machine (LightGBM),Atmos. Chem. Phys., 21, 7863-7880,https://doi.org/10.5194/acp-21-7863-2021, 2021.
Xue, Y., Li, Y., Guang, J., Tugui, A., She, L., Qin, K., Fan, C., Che, Y. H., Xie, Y. Q., Wen, Y. N., and Wang, Z. X.: Hourly PM2.5 Estimation over Central and Eastern China Based on Himawari-8 Data,Remote Sens-Basel, 12,https://doi.org/10.3390/rs12050855, 2020.