

Atmos. Chem. Phys. Discuss., referee comment RC2
<https://doi.org/10.5194/acp-2020-1277-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on acp-2020-1277

Anonymous Referee #2

Referee comment on "Himawari-8-derived diurnal variations in ground-level PM_{2.5} pollution across China using the fast space-time Light Gradient Boosting Machine (LightGBM)" by Jing Wei et al., Atmos. Chem. Phys. Discuss.,
<https://doi.org/10.5194/acp-2020-1277-RC2>, 2021

General Comments

The authors developed a new space-time Light Gradient Boosting Machine (STLG) model for estimating ground-level PM_{2.5} concentration across China from Himawari-8/AHI AOD product and compared the hourly PM_{2.5} estimates with the ground measurements. The results demonstrated that the STLG model is more accurate than other tree-based machine learning models and previous studies as well as superior with faster learning speed and reduced memory consumption. It also suggests the importance of introducing spatio-temporal information into the development of the PM_{2.5}-AOD relationship. However, some doubts remain about the reliability of the hourly PM_{2.5} estimates and their diurnal variation shown in this manuscript. I hope the authors will address the following major criticisms.

Major Criticisms

- In addition to the Himawari-8 AOD, many other variables shown in Figure S1 were used in the STLG model as features to estimate ground-level PM_{2.5} pollution. However, the manuscript does not discuss the contribution of each variable to the estimation. There should be a quantitative discussion of which variables contribute to PM_{2.5} estimation and to what extent. For example, it may be useful to calculate the importance of each variable in the model, or to compare the RMSEs for different combination of feature variables.
- Previous studies have showed the time-dependent bias in the Himawari-8 AOD product used in this study (e.g., Wei et al., 2019b). It is therefore preferable to quantify the temporal dependence of bias in the hourly PM_{2.5} estimation before discussing the

diurnal variations of PM2.5. The manuscript seems to show the temporal dependence of the RMSE (e.g., Figure 2 and Table 1), but not the temporal dependence of the bias.

- In Table 1, the hourly mean concentrations of PM2.5 reach their maximum at 10 a.m., whereas they are lowered at sunrise or sunset. However, the RMSE of PM2.5 estimates is as large as the magnitude of these diurnal variations, and the RMSE appears to be proportional to the value of PM2.5. I believe the authors should ensure that the diurnal variations of PM2.5 derived from Himawari-8 are reasonable compared to ground-based measurements.

Minor Criticisms

- Please spell out the abbreviations: MISR, MODIS and VIIRS in Lines 49-50; RMSE and MAE in Line 92; NDVI, STRM and DEM in Lines 114; MEIC in Line 107;
- Line 37: 'Zhang et al., 2017' is missing from References.
- Line 39: 'Sun et al., 2014' is missing from References.
- Line 72: Which dose 'Zhang et al., 2019' refer to 'Zhang, Q. et al., 2019' or 'Zhang, T. et al., 2019' in Reference?
- Line 159: 'Rodriguez et al., 2010' is missing from References.
- Line 209: '0' should be removed.
- Lines 226: Please explain 'a harsher environment and more intense human activities' in more detail.

(Tables)

- Table 1: Is the uncertainty ' ± 49.31 ' at BTH/9:00 correct value? It seems to be abnormally large compared to the others.

(Figures)

- Labels for y-axis in Figures 2, 3, 6, S2, S3 and S4: The authors labeled the y-axis of Figures 2, 3, 6, S2, S3 and S4 with different labels such as "Model-fitted PM2.5", "Station-based-CV PM2.5", "Model-CV PM2.5" and "Estimated PM2.5". However, some difference may exist, but they should all represent the PM2.5 estimated by the STL model. If so, I believe it would be clearer to use the same label, such as "Estimated PM2.5".
- Captions of Figures 2, 3, 4, 5, 8, S2, S3 and S4: Although Figures 2, 3, 4, 5, 8, S2, S3 and S4 have several panels, explanations are missing from the captions. Please add

explanations of each panel to the captions, as shown in Figures 6 and 7.

(Supporting Information)

- Tables and Figures in the "Supporting Information" section: I wonder why the authors placed so many tables and figures in the " Supporting Information" section. They seem to be part of the main results of this study and used in the discussion demonstrating the advantages of the proposed STLB model. Unless the authors have a clear reason, it would be preferable to place these tables and figures in the full text of the manuscript.