Reviewer: 1

This study built a new space-time extremely randomized trees model (STET), which integrates information from satellite-based aerosol optical depth (AOD) measurements, ground-based PM2.5 observations, and other auxiliary data (e.g., meteorological data), to retrieve daily surface PM2.5 concentrations over China. The newly-developed model outperforms most of the previously reported models in capturing the spatiotemporal variations in surface PM2.5 concentrations and in finer spatial resolution. Overall, this manuscript is well organized with extensive evaluations on the model performance.
Response: We appreciate the time and effort you spent on this manuscript, and we have carefully revised our manuscript. The responses to the questions raised in your report are as follows.

There are some minor concerns that should be addressed before publication.
1. Eq. 1. It is not clear to me how the authors apply these equations. Did the authors apply the relationships between Terra- and Aqua-based AOD measurements to fill the missing AOD value for one sensor while another sensor has a valid measurement on the same day? Please clarify the usage of Eq. 1.
Response: We have replaced the regression method with the average approach according to Reviewer#2's suggestion, and we have clarified this in Section 3.1 of the revised manuscript as follows:
"Terra and Aqua MAIAC AOD retrievals are thus averaged for each pixel on each day to form a new dataset and enlarge the spatial coverage."

2. L201-202. It is possible that the limited impact of precipitation on PM2.5 estimates can be attributed to the fact that there's a high probability of missing AOD measurements on rainy days?
Response: Yes, that's the reason for the limited impact of precipitation on $PM_{2.5}$ estimates. We have added this as "This can be attributed to the high probability of missing AOD retrievals on rainy days." in Section 3.3 of the revised manuscript.

3. It is unclear to me how the authors compare monthly, seasonal, and annual mean PM2.5 retrievals with observed PM2.5 data. For example, for one grid with 100 days of valid daily PM2.5 retrieval, to compare annual mean PM2.5 retrieval with observation, did the authors calculate the corresponding 100-day mean PM2.5 observation or the 365-day mean PM2.5 observation for comparison?
Response: We compared the monthly, seasonal, and annual mean $PM_{2.5}$ retrievals with $PM_{2.5}$ observations using the same number of valid days. We have clarified this in the revised manuscript as follows:
"Synthetized $PM_{2.5}$ retrievals are validated against $PM_{2.5}$ surface observations by calculating the effective values from the same number of valid days at monthly, seasonal, and annual time scales (Figure 10)."

4. L247-248. What's the reason for the overall underestimation of PM2.5 concentration in high polluted days by the STET model?

Response: We have discussed potential reasons in Section 5.1 in the revised manuscript as follows:

"Potential causes are: 1) There are large estimation errors in AOD retrievals under severe pollution conditions in China (Wei et al., 2019c). This is further rooted to the fundamental limitations of satellite-based AOD retrievals, i.e., the non-linear to reflectance and the high sensitivity of the single-scattering albedo (Z. Li et al., 2009); 2) High AOD does not correspond to high $PM_{2.5}$ concentrations because their ratio is highly variable over space and time, affected by both natural and human factors; 3) The number of samples for high-pollution cases is small, hindering the ability to train the model."

5. L310-316. What's the possible impact of variations in the valid sample number of AOD measurement across seasons on the differences in model performance at the seasonal level?

Response: We have discussed the potential causes for the differences in the number of data samples and model performance at the seasonal level in Section 4.2.2 of the revised manuscript as follows:

"Results suggest that there are clear differences in the number of valid data samples because of the long-term snow/ice cover in winter and more frequent clouds in summer, resulting in an overall smaller number of samples than in the other two seasons. … The differences in model performance among the seasons are mainly attributed to seasonal variations in natural conditions and human activities. Meteorological conditions in summer favor the diffusion of pollutants but complicate the $PM_{2.5}$-AOD relationship (Su et al., 2018, 2020), whereas direct emissions of pollutants are greater in winter, resulting in severe air pollution."

6. L361-363. Results in this study cannot support the conclusion here (i.e., air quality improvement from clean air policies) as only one-year PM2.5 concentration data was developed. Please rephrase this sentence.

Response: We have removed this sentence from the manuscript.

7. The caption for Fig.9 is incorrect.

Response: We have corrected the caption in the revised manuscript.

8. L36. "cross-validation coefficient" is unclear here, please clarify whether it means correlation coefficient (R) or coefficient of determination (R2).

Response: We have clarified this in the revised manuscript.

9. Would suggest spelling out all statistical metrics (e.g., R2, RMSE, MAE, MRE) when you first mention them.

Response: Done.

10. Would suggest thoroughly checking the manuscript to avoid grammar errors and make the manuscript more readable.

Response: The manuscript has been more carefully edited by a native speaker.