

Interactive comment on “Improved 1-km-resolution PM_{2.5} estimates across China using the space-time extremely randomized trees” by Jing Wei et al.

Anonymous Referee #2

Received and published: 2 November 2019

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 PM_{2.5} observations, and other auxiliary data (e.g., meteorological data), to retrieve daily surface PM_{2.5} concentrations over China. The newly-developed model outperforms most of the previously reported models in capturing the spatiotemporal variations in surface PM_{2.5} concentrations and in finer spatial resolution. Overall, this manuscript is well organized with extensive evaluations on the model performance. 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

C1

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.

2. L201-202. It is possible that the limited impact of precipitation on PM_{2.5} estimates can be attributed to the fact that there's a high probability of missing AOD measurements on rainy days?

3. It is unclear to me how the authors compare monthly, seasonal, and annual mean PM_{2.5} retrievals with observed PM_{2.5} data. For example, for one grid with 100 days of valid daily PM_{2.5} retrieval, to compare annual mean PM_{2.5} retrieval with observation, did the authors calculate the corresponding 100-day mean PM_{2.5} observation or the 365-day mean PM_{2.5} observation for comparison?

4. L247-248. What's the reason for the overall underestimation of PM_{2.5} concentration in high polluted days by the STET 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?

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 PM_{2.5} concentration data was developed. Please rephrase this sentence.

7. The caption for Fig.9 is incorrect.

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

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

9. Would suggest thoroughly checking the manuscript to avoid grammar errors and

C2

make the manuscript more readable.

Interactive comment on Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2019-815>,
2019.