Comment on egusphere-2022-578

Anonymous Referee #1

Referee comment on "A deep learning approach to increase the value of satellite data for PM2.5 monitoring in China" by Bo Li et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-578-RC1, 2022

Major Comments:

The introduction lacks background on why and how PM2.5 is created, removed from the atmosphere, or where it lingers in the atmosphere. This needs to be expanded upon along with its connection to AOD. Further, it needs to be discussed why AOD may not be an appropriate proxy for PM2.5 at times and how this has been a limiting factor in the past to using AOD to effectively monitor PM2.5 from space.

The setup of the model, data, and testing of the model is difficult to understand. The methodology portion needs major restructuring to understand what data was used, at what scale, what meteorological inputs were used, if they were all at the same grid sized or kept somehow at their native resolutions, what time steps were included of each, how MODIS AOD at a single time step is integrated into the series of AOD from Himiwari-8, etc.

Further, the section on the model configuration is extremely muddled. I do not undestand how k-means was used, what a contingency table is, how the sensitivity analysis fits into the data/model configuration, and why only 10% of data is used as a test case instead of the standard 20% test, 20% validate, and 60% train. It seems as if they only use 10% to test their model, and given their numbers it wouldn't be surprising if that meant the model was overfit and not enough variability in the test samples existed to find that. They state that they did cross validation but the accuracy is never shown. The number of samples is never stated nor is the resolution or the exact inputs of the model clearly stated. The authors claim that they are predicting PM2.5 on an hourly timescale, but it is never clearly stated if that is what they actually trained their model to do. They use sensitivity analysis to test what inputs to use in their model then somehow also use that analysis to verify their model.

Minor comments:

Line 55: A map would be useful to understand where the locations are since I assume these were used as the true labels in training/testing.

Line 73: What limites these past studies from being able to fill gaps in time?

Line 88: Why is AOD only available for 33% of China?

Line 91: What causes haze in rural areas?

Line 97: Spell out what ST-NN stands for.

Line 115: Define WRF as an acryonym as it is used later.

Line 118: A table would be useful of all the inputs since the meteorological inputs are never clearly stated.

Line 135: I do not understand why the Pearson correlation is used to find"containd imension of time"?

Line 136: What is CNEMC and why/how is a Chi-suqared test used?

Line 143: Why is a k-means used? What does the discreteness of variables mean in this context? What and why is a contingency table used?

Line 147: Where are these layers used? Why? How do they affect model performance?

Line 161: Hw many samples? Are they all put into a common grid? What is the time resolution?

Section 2.4 Sensitivity Analysis: What do the levels mean? What does this section mean?

Line 231: What is 26 micrograms m3 in terms of percentage? Or how does it compared to usual values? Is that significant?

Line 234: Why is the model good at nighttime prediction?

Line 240: What does it mean the data are influenced by meteorological and aerosol data at .05?

Line 264: What are the AOD conditions in cloudy scenes? How does the model predict without AOD if it is one of the main predictands?

Line 306: Have any other studies ever used a NN or RF to predict PM2.5?

Line 316: How do your inputs compare to past studies?

Line 323: How good of a proxy is AOD for PM2.5 or PM10?