Comment on essd-2021-326
Anonymous Referee #1

Referee comment on "A global land aerosol fine-mode fraction dataset (2001–2020) retrieved from MODIS using hybrid physical and deep learning approaches" by Xing Yan et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-326-RC1, 2021

This paper constructed a global aerosol fine mode fraction (FMF) dataset using combined physical and statistical methods. The dataset showed overall superior performance over existing satellite products. This is a good paper, providing a useful set of information to study the distribution and potentially climate impact of anthropogenic aerosols. I recommend publication of the paper after addressing a few questions. Although my recommendation is a major revision, the comments should not be difficulty to fix, but I think will increase of the credibility of the method and data.

Specific comments:

- A 10 fold cross validation seems too simple. It would be interesting to see how the method perform on different sites or different years. Therefore, an out of site validation (e.g., predicting FMF at AERONET sites whose data are not used for training) or out of period validation (i.e., using part of the time series as training and the rest as validation) is highly recommended to increase the robustness of the results.
- How are the input variables selected? Are they all necessary or there might be some redundant inputs? Since MODIS AE is used as input, I don’t see how some meteorology variables, such as boundary layer height, winds, RH are physically related to FMF.
- Although the proposed model produces the best FMF among different satellite products, there is still an obvious bias compared with AERONET. Also at some sites the
agreement is low, such as Australia, North Africa, Mid East, etc. These all seem to be desert regions, could this be related to surface, or optical property assumption for non-spherical particles? Some more in-depth error analysis would be helpful to fully evaluate the performance of the technique.