Comment on essd-2021-326
Anonymous Referee #3

In this manuscript, the authors developed and presented the new dataset of fine-mode fraction over global land during 2001-2020. The method they proposed is a hybrid physical and deep learning method, which is the physical model output calibration with DL. Generally, this FMF dataset can be useful for studies of anthropogenic aerosol and also the PM2.5 estimation. The paper is written in a consistent workflow and the inter-comparisons in terms of different methods and official products are very comprehensive. However, I think some concerns and issues need to be addressed.

Major comments:

- We know that in deep learning modeling, the test dataset should be independent of the training dataset to avoid the data leakage, therefore my first concern is that the validation for FMF is independent. In Figure 3a, the authors use the AERONET data for training and testing. A more rigorous validation should be added. My suggestion is to conduct independent validation by FMF from other sources of FMF observations, or the Phy-DL FMF is only reliable over the pixels with AERONET sites.
- My second concern is the use of meteorological data, which are very different inputs compared to previous studies. I can see the meteorological data are widely used in fine particulate (PM2.5) retrievals because meteorology has statistical correlation and physical interaction with the PM2.5. While in this study, the author simply explained their reason as “the impact of meteorological factors”, and what is this “impact” to make them use the meteorological data is not mentioned. For example, the temperature is introduced as input, I don’t see the influence of temperature on FMF or how it can improve the retrieval accuracy.
- For the physical model, although LUT-SDA has been used in other studies before, it is better for authors to emphasize its disadvantage with more details, rather than just listing its applications.
- Last I think the uncertainty of this Phy-DL FMF should have a more in-depth discussion. The major content in Results is the inter-comparisons of different results in terms of different methods and FMF products, which are good, but there should be more discussion on the sources of uncertainty of this Phy-DL FMF. For example, the Figure 6 compared performance over different land types and barren land has the worst
performance for all three FMFs, so the authors could discuss how the physical characteristics of barren land affected the retrieval performance.

Minor comments:

- in Line 44. The “performed previously,; currently” should be “performed previously; currently”
- Figure 1. AATSR and VIIRS also provide FMF products but they were not discussed in this paper. The reason for using MODIS rather than AATSR and VIIRS should also be mentioned.
- in Line 56. The Yonsei Aerosol Retrieval algorithm (Choi et al., 2016) is not for MODIS land-based FMF retrievals, it is for GOCI.
- in Line 103. There is no direct relative humidity data ERA5. Usually it is calculated from dew point temperature.
- in Line 104. The “ERA5” mentioned for the first time without given the full name.
- in Line 210. The “yr” mentioned for the first time without given the full name.
- in 3.2. Both past and present tense showed in this part when describing the same thing. For example, “in India, FMFs are noticeably higher in autumn and winter, especially in Northern India (i.e., the Indo-Gangetic Plain), where the FMF was greater than 0.87”, and they should be either past or present tense.
- in Figure 6a. Why do you choose to compare the result in bins of FMF?
- in Line 318 and 320. “eastern China” or “Eastern China”?
- I noticed the plurals appeared randomly, for example, it is “AERONET FMF” or “AERONET FMFs”. Make sure the plurals are consistent in the paper.