Reply on RC2
Kaixu Bai et al.

Author comment on "LGHAP: a Long-term Gap-free High-resolution Air Pollutants concentration dataset derived via tensor flow based multimodal data fusion" by Kaixu Bai et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-404-AC2, 2022

By fusing multisource aerosol data with different resolutions via a tensor completion approach, the authors generated a spatially complete AOD dataset with daily 1-km resolution in China from 2000 to 2020, from which gap-free PM2.5 and PM10 concentration data were estimated using an ensemble learning method. Such gap free and high-resolution aerosol data would greatly benefit aerosol radiative effect diagnose and PM2.5 pollution exposure assessment as well as haze pollution management. Overall, the paper is well written and the quality of these datasets are properly validated. Some minor revisions are required before the acceptance of the current manuscript.

Reply: Thank you for your copious comments and suggestions, which greatly help improve the quality of this manuscript. We will the manuscript by following your comments, and our point-to-point reply is given right beneath each individual comment.

Specific comments:

1. Section 2.1: Given there exist many versatile AOD products, why did the authors choose to use these six AOD datasets, how about those from geostationary satellites, any specific reasons?

Reply: Thanks for your insightful comments. In the current release, we only used six gridded AOD datasets retrieved from polar orbiting satellites with a relatively long temporal coverage (>5 years) as observational data to reconstruct AOD fields, whereas AOD retrievals from geostationary satellites were not employed. The main reasons can be attributed to the follows. First, only geostationary satellites such as FY-4 and Hamawari-8 have a coverage of terrestrial over China. However, there is no operational AOD product publicly available from FY-4 till now. For Hamawari-8, the AOD product cannot provide observations over the northwest regions of China. Secondly, AODs retrieved from geostationary data are often at different local solar time, differing from AODs derived from polar satellites. Given these potential drawbacks, AODs from geostationary satellites were not applied when generating the current dataset. We appreciate your constructive suggestions and will attempt to include geostationary AODs in our future datasets.

2. Section 2.2: since atmospheric visibility data are continuously gauged, why did the authors only use data before 2014? Please clarify this in the manuscript.
Atmospheric visibility data from 2000 to 2013 were used as a critical indicator to infer pollution levels at a site level before 2014, at which the national ambient air quality monitoring network has not been established and thus site-based PM2.5 observations were lacking. After 2014, given the availability of substantial PM2.5 concentration observations, visibility data were not used anymore. This not only helps reduce the computational burden but avoid the propagation of uncertainty from visibility-inferred PM2.5. Another issue is tied to the massive instrument replacement and changes in the observing criterion of visibility after 2014 in China. Consequently, we only used visibility data before 2014. We will clarify this in our revised manuscript.

3. The full name of key parameters should be given in figure 1 captions

Reply: Thanks for your suggestion! We will provide the full names of these abbreviations when revising the manuscript.

4. Lines 205,314: PM should be PMx

Reply: Thanks!! We will change it to PMx.

5. Figure 2: the gap filled AOD in SC had a relatively low accuracy compared with other regions, what are the possible reasons?

Reply: Thanks!! This is mainly due to limited satellite-based AOD retrievals in Sichuan basin due to frequent and extensive clouds over there. In other words, few observational AODs are available for tensor completion and the reconstructed results are thus dominated by historical AOD observations and numerical AOD simulations. These collectively result in large uncertainty to the reconstructed AOD fields in Sichuan basin. We will discuss this issue in the results section when revising the manuscript.

6. What is the unit of shading value in figures 2 and 6? e.g., %?

Reply: As indicated in the colorbar, the shading values show the number of scatters (no unit), and thus the larger the value, the more the data points falling within the given location.

7. Figure 3: had these AERONET AOD observations been used as input when filling gaps in satellite AOD retrievals?

Reply: All AERONET AOD observations were not incorporated as inputs when reconstructing AOD fields. Rather, these observational AODs were simply used as the ground truth to validate our reconstructed AOD fields. We will emphasize this point when revising the manuscript.

8. Figure 5: how about the contribution of in-situ measurements?

Reply: Thanks!! In Figure 5, we only estimated the contribution of gridded AOD products, and the contribution of in situ measurements were not calculated because the used method needs to account for the number of valid observations. Compared with gridded AOD products, the volume of in situ measurements is a bit small, and the estimated contribution is thus incomparable to those derived from gridded products. Given this reason, we did not calculate the contribution of in situ measurements.

9. Wintertime (September to February) should be winter-half year. Please check it though manuscript

Reply: Thanks for pointing out, and we will make essential corrections when revising the
manuscript.