Interactive comment on "A homogenized daily *in situ* PM$_{2.5}$ concentration dataset from national air quality monitoring network in China" *by* Kaixu Bai et al.

Kaixu Bai et al.

jpguocams@gmail.com

Received and published: 12 October 2020

Responses to Reviewer #1 comments

The PM2.5 data has been widely used for human exposure risk assessment and air quality management. However, as the author said, given the absence of an open access and quality assured in situ PM2.5 concentration dataset in China, it is urgent need to open a stable and reliable PM2.5 data access method. This paper attempted to generate a long-term coherent in situ PM2.5 concentration dataset for scientific community to use in future applications. Methods involving missing value reconstruction, change point detection, and bias adjustment were applied sequentially to deal with data gaps

and inhomogeneities in raw PM2.5 observations. It is a nice and well-organized paper with a clear focus. In my opinion, there are some minor problems need to be solved before publishing. My biggest concern is whether the data set will continue to be updated. I suggest that the author add a statement in the conclusion, stating the update frequency and download link of the homogenized PM2.5 datasets. In the change points detection, how long is the breakpoint interval?

Reply: Thank you for your valuable comments and suggestions in helping improve the quality of this manuscript. Our point-by-point replies to your concerns and comments were summarized as follows:

1)"My biggest concern is whether the data set will continue to be updated"

Reply: The homogenized in situ PM2.5 concentration dataset will be regularly updated for every six-month based on our newly retrieved data records, and the extended dataset is also freely accessible per the user's request. A full dataset will be then published online on PANGAEA once we have one-year's new measurements.

2) "I suggest that the author add a statement in the conclusion, stating the update frequency and download link of the homogenized PM2.5 datasets"

Reply: Per your suggestion, we will clearly state the updating frequency of the dataset in our revised manuscript.

3) "In the change points detection, how long is the breakpoint interval?"

Reply: The PMT method was hereby applied to detect possible break points in each PM2.5 concentration time series in reference to the generated reference series. As the default configuration in the RHtests v4 software package, a length scale of 5 was defined as the minimum interval between two possible change points, which means that no change point would be detected from the 5 adjacent observations. More technique details of PMT method can be found in the following reference, which has been also cited in section 3.

References: Wang, X.L. Accounting for Autocorrelation in Detecting Mean Shifts in Climate Data Series Using the Penalized Maximal t or F Test. J. Appl. Meteorol. Climatol. 2008, 47, 2423–2444, doi:10.1175/2008JAMC1741.1.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-121, 2020.