The submitted manuscript well fits within the journal scope as it is describing a method to fill missing values in hourly PM2.5 concentrations for more than one thousand observational sites across China. Overall, the work is consistent and the method is well explained. Nevertheless, in my opinion, before publication, two points should be considered before publication

Reply: Thank you for your valuable comments and suggestions in helping improve the quality of this manuscript. The paper has been thoroughly revised according to your comments (in black), and please find the point-to-point responses (in blue) to your concerns below and refer to the revised paper for more detail.

1) The authors made a sensitivity study to assess how the number of neighbour stations impact the reconstruction of PM2.5 concentration. However, it might happen that the spatial distribution of the neighbour station might influence the final result, i.e. in case of equispatially distributed or spreade. I suggest to perform a sensivity test for a couple of cases taking as metric the sum of euclidean distances using the same number of stations for the same aerosol loading.

Reply: Good point. Per your kind suggestion, we checked the potential impacts of the number of neighboring stations and their spatial structure on the prediction accuracy of missing values, which is shown in Figure R1. It can be seen that the correlation coefficient does not changes dramatically with the increase of number of neighboring stations as well as the distance between the target station and the closest station. This means that the spatial pattern of neighboring station does not influence the performance of the proposed gap filling method. This is mainly due to the implementation of an optimization process (step 2 in our method) to identify similar observations rather than using all available observations for the reconstruction of PM2.5 diurnal cycle. In other words, the final input observations only contain those with similar diurnal variation pattern to the target observation, and the distance is thus not a critical influential factor when there exist abundant samples.

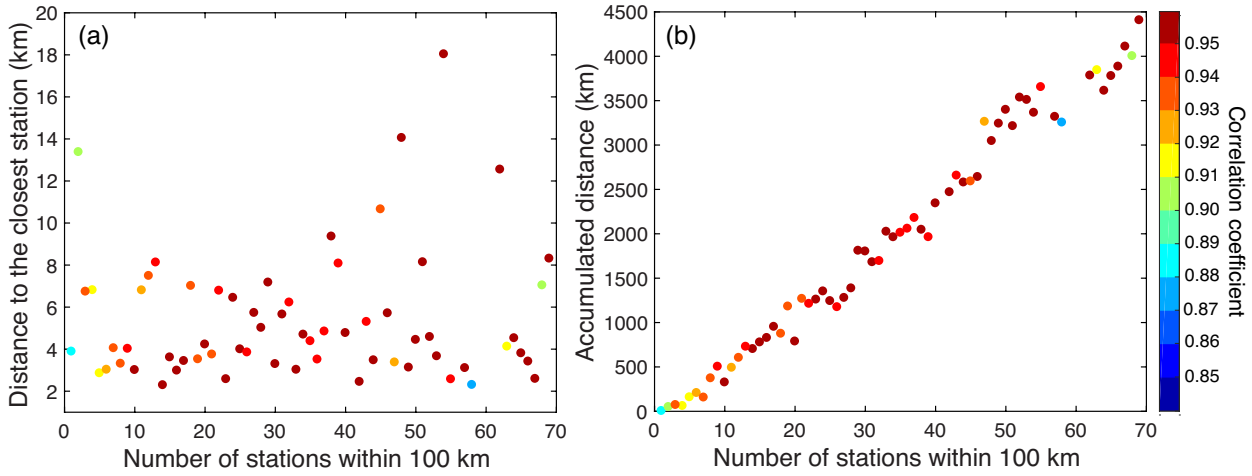

Figure R1. Impacts of number of neighboring stations and their spatial structure on the prediction accuracy of missing values.

2) it is missing how the measurement error is impacting the reconstruction as all the measurements are presented without error bars.

Reply: Thanks for your valuable comments. The impact of measurement error on the final accuracy of gap filling is not assessed in the current manuscript. The reasons are twofold: (1) The $PM_{2.5}$ data used in this study are gauged by the state-level monitors, so the quality of the data record is assured. (2) Our gap filling method mainly get involved in the usage of empirical orthogonal function (EOF) in order to reconstruct the diurnal variation pattern of $PM_{2.5}$, which would in turn cancel out the measurement errors (if any). Therefore, the measurement error would have little effect on the final results.

English should be revised as some sections are not very clear.

Reply: We have made essential corrections in this revised manuscript per your valuable suggestion.

Specific comments are available in the attached file.

Reply: Thanks for your valuable comments and suggestions. Except for the glitches and typos that have been corrected directly in our revision, the responses to several specific concerns are listed as follows.

Line 152: how those numbers (m and n) are determined? How the method accuracy changes changing those numbers?

Reply: *m* and *n* are determined by the given spatial (100 km) and temporal (7-day before and after t) window size, respectively. A cutoff value of 100 km and 7-day are used based on our recent results in which an optimal window size of 50 km and 3-day was found to attain a good autocorrelation of PM2.5 concentration in space and time, respectively (Bai et al., 2019). Here we enlarge (double) the both window sizes so as to have adequate samples for the construction of $\boldsymbol{X}_{p,t}^{m,n}$ while avoiding including all available samples, especially for those distant away. In general, these two window sizes would have little effect on the performance of the subsequent gap filling once they are large enough (at least greater than the identified optimal window sizes) to cover most similar observations nearby since a sorting scheme (step 2) will be further applied to identify observations with similar diurnal variation patterns to that of the target station. Such effect is also evidenced in Figure 9b that the prediction accuracy would not increase with the number of neighboring stations once there are more than 3 neighboring stations nearby. These more detailed discussions have been added in the revised manuscript to ease the readership.

Line 171-175: this part should be better explained.

Reply: This part regarding the EOF process for data gap filling has been explained by clarifying it in the context in this revision, which shows as follows:

"Reconstruct the local diurnal cycle of PM2.5: The diurnal cycle of PM2.5 at site $\boldsymbol{p}$ on date $\boldsymbol{t}$ (denoted as $\boldsymbol{\beta}_{\boldsymbol{p}}^{\boldsymbol{t}}$) was then reconstructed from the optimized PM2.5 neighborhood field $\widehat{\boldsymbol{X}^{\boldsymbol{k}}}$ using EOF in an iterative process similar to the DINEOF method (Beckers and Rixen, 2003). In our DCCEOF method, the target PM2.5 time series at site $\boldsymbol{p}$ on date $\boldsymbol{t}$ (denoted as $\boldsymbol{x}_{\boldsymbol{p}}^{\boldsymbol{t}}$) were also included to constrain the reconstruction of $\boldsymbol{\beta}_{\boldsymbol{p}}^{\boldsymbol{t}}$, and the whole field was then denoted as $\widetilde{\boldsymbol{X}}$.

$$\tilde{X} = \left\{x_p^t, \widehat{X^k}\right\} \tag{4}$$

In general, the EOF-based gap filling process can be outlined as follows: a) 20% of valid PM2.5 observations in $\widetilde{\boldsymbol{X}}$ were first held out for cross validation and then these data values were treated as gaps by replacing with nulls (i.e., missing value); b) given that a small amount of missing values would not significantly influence the leading EOF mode for the original data set, we may assign a first guess (here we used the mean value of valid data on each specific date) to the data points

where missing values are identified to initialize the EOF analysis; c) EOF analysis was performed on the previously generated background field (that is, $\tilde{\boldsymbol{X}}$ with gaps are filled with daily mean and denoted as $< \tilde{X} >$) in a form of singular value decomposition (SVD) and then data values at value-missing points were replaced by the reconstructed values using the first EOF mode. These processes can be expressed as:

$$[U, S, V] = svd(< \tilde{X} >) \tag{5}$$

$$X' = u_1 * s_1 * v_1 \tag{6}$$

where $< \tilde{X} >$ denotes the initial matrix in which the missing values were filled with daily means. U, S, and V are three matrices derived from SVD while u_1, s_1, and v_1 denote the SVD components in the first EOF mode."

Line 355: "largely from that of neighboring stations at the same time", how do you deal with this problem?

Reply: The proposed DCCEOF method is unable to deal with such issue once the diurnal variation pattern of neighbors differs largely from that of the target station. We have clearly stated this defect in our revised manuscript.

Line 360: how about the instrument precision?

Reply: The precision of $PM_{2.5}$ records have been introduced in section 4.1.

Line 409: how about the spatial distribution of the stations? How this impacts on final result? "The experimental results suggest that three neighboring stations within 100 km", does matter the their mutual location?

Reply: The topotaxy effect between these neighboring stations is not considered in our current method since we only accounted for the relative similarity between their diurnal variation patterns rather than their locations. In other words, whether the $PM_{2.5}$ observation measured at one station will be applied for gap filling does not depend on its location (see Figure S1); Rather, we only took the similarity of $PM_{2.5}$ observations between the target station and neighboring stations as a

measure to select similar observations for the subsequent diurnal cycle reconstruction. Discussions related to this issue has been added in the revised manuscript to bridge the readership gap.