The method recommended by the authors for the missing value filling to hourly $PM_{2.5}$ data is interesting. It could be useful for relevant study.

Reply: Thank you for your insightful comments and valuable suggestions which help a lot in improving the manuscript. All your raised concerns (in black) have been properly and adequately addressed in our revised manuscript and point-to-point responses (in blue) can be found below.

Some concerns remain as following, which might be considered to further improve the method.

(a) Because the $PM_{2.5}$ diurnal variation could vary largely from day to day, is it possible that some typical classification of $PM_{2.5}$ diurnal variation could be established and considered, which should be helpful if one can determine the general pattern of $PM_{2.5}$ diurnal variation for the interested day and make more adequate filling for the missing $PM_{2.5}$ data.

Reply: Thanks for your constructive suggestions. Actually, what you suggested is our ultimate goal that we intentionally focused on the analysis of the diurnal variation pattern of $PM_{2.5}$. However, the observed salient data gaps in using our retrieved $PM_{2.5}$ time series became a big obstacle and this is also the motivation of the development of this gap filling method. In the next step, we will attempt to extract the general pattern of $PM_{2.5}$ diurnal variation in space and time using the gap filled time series and then use such general patterns to better deal with data gaps present in future data records. In short, your insightful suggestion provides us new perspective to use $PM_{2.5}$ diurnal variation pattern to better deal with $PM_{2.5}$ data gaps in the future. Also, we have discussed this perspective in our revised manuscript.

(b) The $PM_{2.5}$ diurnal variation could be related to some specific meteorological factors as well as their diurnal evolution. Is it possible that the diurnal variation of specific meteorological factors be considered within the authors recommended missing value filling method?

Reply: Thanks for your insightful comments. Incorporating the diurnal cycle of some related factors such as the mixing layer depth would be a big plus in advancing the estimation of the $PM_{2.5}$ missing values. Nevertheless, such an endeavor is still subject to the following two constraints: (1) the lack of high temporal resolution data (e.g., hourly mixing layer depth) and (2) the relationship between $PM_{2.5}$ and the related factor (that is, to what extent the diurnal variation pattern of PM2.5 can be explained by the diurnal cycle of the related factor). To figure out the mechanism, more explicit model simulations are anticipated, which is out of the scope of our current study. However, it deserves more in-depth analysis in the future. Such an endeavor has been discussed in the revised manuscript to provide new envisions to the improvement of this gap filling method.

(c) What is the applicability of the method? Especially for the different spatial distribution of the air quality monitoring stations which are condense over eastern China but sparse over western part of the country.

Reply: Good point! The question you mentioned does matter the accuracy of the proposed method and we have discussed this issue in our revised manuscript in terms of the impact of number of neighboring stations on the prediction accuracy (Fig.9b). For stations with at least one neighboring monitor (like in the west of China), our method is still able to recover the missing value with good accuracy ($R>0.7$). This lies in the principle that both spatial and temporal neighborhood information are used to reconstruct the diurnal cycle of $PM_{2.5}$. Such effect is also corroborated by our most recent paper (Bai et al., *Environmental Pollution*, 2019, doi: 10.1016/j.envpol.2019.113047) that $PM_{2.5}$ has a good autocorrelation in adjacent space and time domain. The prediction accuracy could be relatively poor for those isolated stations (no neighboring station) given the lack of spatial neighborhood information, and such effect might be mitigated by incorporating other related factors as you suggested in the future. We have discussed this issue in this revision to bridge the readership gap.

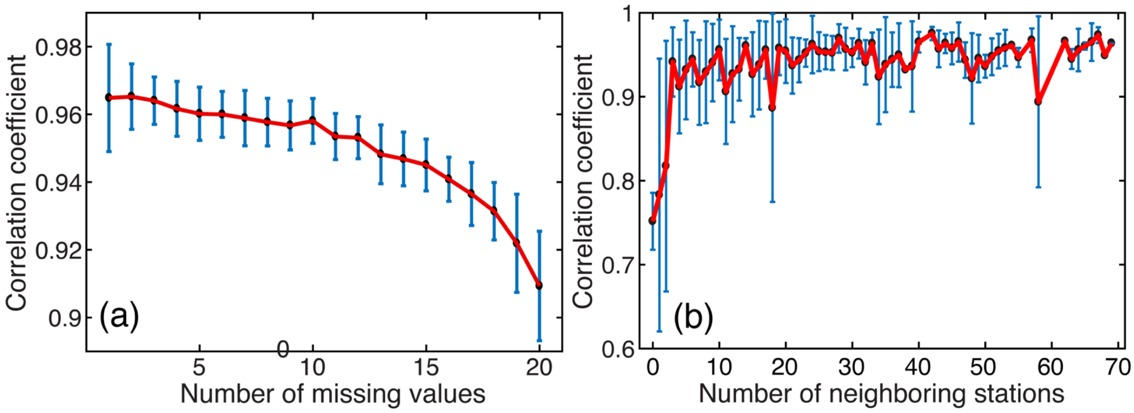

Figure 9. Impacts of the number of missing values present in hourly PM2.5 records for every 24-h (a) and the total number of neighboring stations within 100 km (b) on the performance of the proposed gap filling method. The error bars denote one standard deviation of each value from the mean on each side.

References:

Bai, K., Li, K., Chang, N.-B., Gao, W., 2019. Advancing the prediction accuracy of satellite-based PM2.5 concentration mapping: A perspective of data mining through in situ PM2.5 measurements. ***Environ. Pollut.*** 254, 113047. https://doi.org/10.1016/j.envpol.2019.113047

(d) In the manuscript, the authors made cross validation for missing value filling for several hours, is it possible that there are missing value for a specific station for one day or several days? If this situation happens, how about the performance of the authors recommended method to make missing value filling?

Reply: In the current manuscript, we only deal with the days with missing values no more than 20 within 24 hours since the missing values are recovered by projecting the reconstructed diurnal cycle of $PM_{2.5}$ to the level of valid measurements on a specific day. For the situation with data missing for a whole day or several consecutive days, we did not recover the data given the lack of essential reference data values. Although there exists a possible way that the diurnal cycle of other related factors could be used, data amplitudes on different days may still differ from each other even in the presence of similar diurnal cycle pattern, and this is also the reason why we need to have several valid measurements for that specific day. This issue has been discussed in the revised manuscript to bridge the readership gap. Again, we highly appreciate your insightful comments in helping improve the quality of this manuscript.

Some specific comments are also listed below for the authors.

1. Line 60, "data cleaning processes", consider using more accurate wording to describe what the authors want to mention.

Reply: Per your kind suggestion, it has been reworded to "how data gaps were treated in their data exploration processes (e.g., integration and transformation)" to ease the readership.

2. Lines 70-71, it is better to directly give the disadvantages of "approaches of ignoring missing values or excluding records on days with missing values", rather than arbitrarily comment these approaches as "unreasonable".

Reply: Per your kind suggestion, the disadvantages have been clearly stated in the revised manuscript as: "Nevertheless, such a treatment on missingness (i.e., ignoring missing values or excluding records on days with missingness) would either introduce new bias to the aggregated $PM_{2.5}$ record or make the original $PM_{2.5}$ time series temporally discontinuous, especially when missingness occurs at some specific times (e.g., during severe pollution episodes)."

3. Table 1, the lines for the references are not quite clear, it is difficult to find which reference is corresponding to which method.

Reply: Thanks for pointing it out. We have enlarged the height of each row to make them more distinguishable.

4. Line 152, "$\boldsymbol{m}$ was defined as the number of stations within 100 km of the target station", as the authors mentioned about the "significant heterogeneity" of the $PM_{2.5}$

data, is the setting of "100 km" improperly greater in this context? $PM_{2.5}$ concentration can vary largely even within a small area. Moreover, the air quality monitoring stations are densely distributed over eastern China but sparsely over western part of China. Is there any special consideration should be taken on this issue?

Reply: Thanks for your insightful comments. We are aware of the fact that *m* and *n* are two critical factors associated with the performance of the proposed gap filling method since it determines the size of spatial and temporal neighborhood used for the reconstruction of the diurnal cycle of $PM_{2.5}$. In the current method, two empirically-determined invariant spatial and temporal window sizes of 100 km and 14 near-term days were used, but these two numbers have little effect on the final prediction accuracy of missing values. This is because these two numbers are simply used as a threshold to limit the number of samples to avoid the usage of all available data. Our recent study published in Environmental Pollution has revealed a proper spatial and temporal window size of autocorrelation of 50 km and 3-day in eastern China for $PM_{2.5}$. Therefore, a window size of 100 km and 14-day suffices to include adequate number of candidate samples in space and time for the reconstruction of $PM_{2.5}$ diurnal cycle. Most critically, the neighboring data confined to these two numbers are not directly used to reconstruct the diurnal cycle of $PM_{2.5}$; rather, we have proposed a set of constraints to pinpoint those similar samples from the whole dataset determined by 100 km and 14-day for the subsequent diurnal cycle reconstruction. Finally, only those samples have similar diurnal variation pattern will be used for the diurnal cycle reconstruction. We have clearly stated this in section 3, please refer to the second procedure (construct a compact $PM_{2.5}$ neighborhood field) on page 8 (Line 156-160) for more detail.

5. The day-to-day $PM_{2.5}$ diurnal variation could vary largely, which depends on whether it is a clean day or a severe polluted day, as well as the various weather conditions. The authors also mentioned this in Lines 302-304. While the method

the authors suggested only considers the diurnal variation of one week before and one week after the data missing day to be filled. Is it possible any variety in the diurnal variation of $PM_{2.5}$ can be considered in the recommended method? Also, more detailed classification and establishment of the typical patterns of $PM_{2.5}$ diurnal variation and adequate consideration of this issue could be very helpful to improve the data filling method suggested.

Reply: We appreciate your constructive suggestion. Same as the above question, the temporal window used here would not significantly affect our results since it is simply a cutoff value (large enough to include adequate samples) to limit the number of samples for the subsequent analysis. A compact neighborhood is further generated for the reconstruction of $PM_{2.5}$ diurnal cycle by only including similar samples rather than all the data samples. The classification of the typical patterns of $PM_{2.5}$ diurnal variation is quite a smart suggestion and we will try to account for this effect in the further to improve the current method. We have envisioned this perspective in the discussion section to broaden the possible improvement of the current method. Again, thanks for your insightful suggestion.

6. Figure 3, it is a little difficult to understand the variables illustrated. The result presented in each panel of the figure seems not match with the caption. The name of the x axis in Figure 3f could be better as “hour”.

Reply: Thanks for pointing these typos. Following your suggestions, we have revised the caption to match the figure. The name of the x axis in Figure 3f has been revised to “hour” per your suggestion.

7. Figure 4a, the 50th percentile of the mean relative differences generally remains constant around zero, does this mean that the 50th percentile is subjective of less influence from missing values?

Reply: The 50th percentile of the mean relative differences around zero just reveals the fact that missing values would result in random bias (half below zero and half above zero) to $PM_{2.5}$ daily averages. We have explained this effect in the manuscript in section 4.2.2 (Line 282-286).

8. Figure 6, the reconstructed diurnal $PM_{2.5}$ variation seems to be a smoothed average of the observations near the interested station within a week before and after the interested day, it cannot reconstruct any particular variation of $PM_{2.5}$ such as those at 19:00 local time in Figure 6e and at 08:00-09:00 local time in Figure 6f.

Reply: Yes, the reconstructed diurnal cycle of $PM_{2.5}$ is relatively smooth compared with the actual observations and thus some small variations cannot be fully recovered. This lies in the fact that the diurnal cycle of $PM_{2.5}$ is reconstructed from the spatial and temporal neighborhood using the EOF method and hence it mainly captures the dominant variation mode. We have clearly explained this defect in our manuscript in Lines 336-340.

9. Lines 409-411, because of the "significant heterogeneity" of the $PM_{2.5}$ spatial distribution, how about the spatial distribution of the diurnal pattern of $PM_{2.5}$ variation? Is it practical to consider the variability of $PM_{2.5}$ at the stations 100 km away to fill missing value of $PM_{2.5}$?

Reply: Thanks for your constructive comments. Yet the spatial distribution of the diurnal pattern of $PM_{2.5}$ variation in China has not been examined due to the discontinuous hourly $PM_{2.5}$ observations, we will investigate the diurnal pattern of $PM_{2.5}$ variation in China soon per your suggestion and try to identify the typical diurnal variation pattern to improve the current method. In our current method, we did not consider to use $PM_{2.5}$ data measured 100 km away for gap filling though there might exist similar variation patterns. This lies in the first principle of geography that data from closer stations are more similar than those distant away. On the other hand, the final prediction accuracy is not sensitive to the spatial window size if it is large enough to include three neighboring stations (Figure 9b).

10. Do Figure 10a and 10b reflect the same information from different perspectives? Is it possible just keep one figure to discuss the issue?

Reply: Not exactly. Actually, Figure 10a indicates the total number of missingness (percentage with respect to the total number of record) have been filled at each station whereas Figure 10b shows the number of days with missingness removed. As shown in Figure 10a, the removed total number of missingness seems to be low compared with the total number of samples. Nevertheless, Figure 10b indicates the percentage of how many days are without missingness after gap filling.

11. Lines 414-422 and Figure 10, have the authors done data filling for all the available $PM_{2.5}$ data over China with the recommended method? Is the evaluation presented here are based on data filling for the whole dataset of $PM_{2.5}$ available?

Reply: Yes, we have performed gap filling for each site-specific $PM_{2.5}$ record in China and the results shown in Figure 10 are based on the gap-filled data set. As indicated, data gaps still persist even after gap filling and this is mainly because we did not fill the gaps for the episodes with missingness continuing for a whole day or several consecutive days. Discussions with respect to this issue has been added to fill the readership gap.