Response to reviewer 2

We appreciate reviewer's thoughtful comments and suggestions, which are greatly helpful for us to improve our manuscript. The manuscript has been revised to accommodate the reviewer's comments.

General summary This paper assesses the impacts of anthropogenic emission uncertainties on aerosol data assimilation and short-term $PM_{2.5}$ predictions in East Asia. Two different anthropogenic emission inventories are used to calculate background error covariance statistical parameters. This is an interesting topic and such studies are welcome in light of the emerging needs of air quality forecasting systems especially in the developing countries. However, I have three major concerns listed below.

General comment First, it is not clear how the authors created the new NMC background error statistical parameters. The NMC method can use only two forecasts but at line 189, the authors said they used four different forecasts to calculate the background error statistical parameters. Please explain in detail how you did that?

Response For the calculations of the statistical parameters of the NMC method and our method in this study, the options of "NMC" and "ENS" were used, respectively, in GEN_BE v2.0. Regarding this point, please, refer to pp. 7:221–pp. 7:223 in the revised manuscript. Also, in order to further clarify this point, we have changed the term of "NEW NMC" to "NMC+EMIS" throughout the manuscript. We also modified Fig. 4. and added Appendix. A to remove such confusion.

General comment Second, it is not clear how the 24 h forecasts have been performed. Are they launched everyday after assimilation at 00, 06, 12 and 18 UTC? It is difficult to follow the discussion without answering this question.

Response The 24-hour forecasts were launched every 6-hour after assimilation at 00, 06, 12 and 18 UTC (four times per day). As far as we understand, this is a typical way that air quality prediction is conducted. By doing this, the benefits of assimilations can be accumulated, because each 24-h prediction uses the assimilated initial conditions using observations and

background file assimilated before 6-hour. Please, refer to modified Fig. 3 in the revised manuscript.

General comment I was also surprised to see that the benefits of assimilation are diminishing so rapidly. IOA, R, and RMSE go back to the CNTL experiment level in less than 12 hours. Another surprising feature is that MB in the new NMC method is larger up to 6 hours of forecast and decreases with lead time. This is unexpected because one would expect the assimilation to bring the model close to the observations at the initial time and errors would start to grow with lead time as model deficiencies start to take over. Could you please explain this?

Response One of the reasons for the rapid diminishing of the benefit may be that the performances of our base model (CTL) were not too good to keep the benefits of data assimilation, particularly in East China. Please refer to the RMSEs of CTL over East China are ~80% higher than these over South Korea (see Table 2 and Table S3). We think that if the CTL simulation could capture the spatial and temporal distributions of $PM_{2.5}$ with more accurate model inputs, e.g., better emissions and finer model spatial resolution over China, our method would show better performances**.**

General comment Third, I did not find details of the independent observations used for evaluation. From the discussion, it appears that all sites were used for assimilation as well as evaluation in Figures 6 and 7. If that is the case, the results will be artificially good for the assimilation experiments.

Response To answer the review's comment, we carried out an additional comparison of DA results with 20% of independent observations which were taken out and were then used only for comparison purpose. Regarding this point, please, refer to pp. 9:262–pp. 9:268 and pp. 11:333–pp. 11:340 in the revised manuscript.

Other Specific and Minor Comments

Comment Line 19: spell out PM. Since you are just focusing on PM2.5, you should use PM2.5 here rather than PM.

Response We have spelled out PM in that line. Please, check out pp. 1:19.

Comment Line 28: Do you mean 44% stations showed smaller negative bias.

Response Here, what we tried to mean was that the negative biases were reduced with the new method (NMC+EMS), compared with those from the NMC method. We have modified the sentence as followed: "Our method also showed closer agreement with the observations in 24-hour $PM_{2.5}$ predictions than the conventional method (in particular, with a ~44 % reduction of negative biases)".

Line 37: Remove "a" before significant.

Response We have corrected the sentence. Please, check out pp. 2:41.

Line 38: I guess you meant to say "high PM levels affect the radiation budget in Asia."

Response We have modified the sentence. Please refer to pp. 2:41–pp. 2:43.

Comment Line 45: You should also highlight that inaccuracy also results from errors in initial conditions because that's what you improve with chemical data assimilation.

Response We have modified the sentence. Please, see pp. 2:48–pp. 2:50.

Comment Line 51: Remove "the results of"

Response We removed it. Please, see pp. 2:55–pp. 2:57.

Comment Line 126: Change "fewer" to "smaller"

Response We have changed it. Please, check out pp. 5:154–pp. 5:156.

Comment Line 145: Does not GSI optimizes the background fields and creates analysis fields rather than optimizing analysis fields.

Response We have corrected the sentence. Please, see pp. 6:181–pp. 6:182.

Comment Line 148: R is

Response We have changed it. Please, check out pp. 6:185.

Comment Line 150: Change "status" to "space".

Response We have changed it. Please, check out pp. 6:186.

Comment Line 158: I guess you meant to say that a small tail of coarse mode distribution represents aerosols with diameter smaller than PM2.5.

Response Yes, we had tried to mean it. In order to further clarify the point, we corrected the sentences. Please, see pp. 7:195–pp. 7:197.

Comment Line 189: NMC method can use only two forecasts. Can you explain how you run NMC with 4 forecasts? Did you use the ensemble method of GEN_BE?

Response Yes, we used ensemble method in GEN_BE v2.0, as reviewer pointed out. Regarding this, please, refer to 7:221–pp. 8:223. In order to further clarify this, we have changed the term of "NEW NMC" to "NMC+EMIS" throughout the entire manuscript.

Comment Line 194: Add "generate" before the.

Response We have modified the sentence. Please, check out pp. 8:244.

Comment Figure 4: Why are there multiple arrows pointing from CMAQ2GENBE to green, red, and blue outlined boxes?

Response CMAQ2GENBE converts CMAQ outputs (NetCDF-format) into inputs for GEN-BE readable files (binary format). Two inputs from the CMAQ2GENBE were used for NMC BECs, and four inputs from the CMAQ2GENBE were used for NMC+EMIS. We have modified Fig. 4. and have added Appendix A to remove such confusion.

Comment Line 195: The authors state that they extracted background fields from the first simulation (Met. 1 + CREATE emissions) at assimilation times 00, 06, 12, and 18 UTC daily and used the "BKOBS2GSI" module to convert the background file. This is in contrast to the information shown in Figure 3 which shows that 06 hours prediction after assimilation at 00 UTC serves as initial condition for the assimilation and forecast starting at 06 UTC. This statement gives an impression that benefits of assimilation are not accumulated over time as suggested in Figure 3.

Response As we already discussed in a previous comment, the benefits of assimilations are accumulated, because each 24-h prediction uses the assimilated initial conditions using observations and background file that was assimilated before 6-hour. Please, see modified Fig. 3.

Comment Figure 5 and Line 227: Do you use the same background error statistics at 00, 06, 12, and 18 UTC? Are the profiles shown in Figure 5 averaged over all latitude bins? What was the latitude bin in your set-up?

Response We used the same background error statistics at 00, 06, 12, and 18 UTC. We selected a method of binning as sampling all the horizontal grids per each vertical level (i.e., an option of bin_type = 5 in GEN-BE v2.0). Please, refer to pp. 7:219–pp. 7:221.

Comment Line 243-244: It is also reflecting that perturbations used for NMC calculations do not account for uncertainties in vertical transport.

Response We have modified the sentence, based on your comment. Please, check out pp. 10:293–pp. 10:295.

Comment Line 246-247: I guess the unit of horizontal length scale is km here.

Response We have added the unit of horizontal length scale (km) into the sentence. Please, check out pp. 10:298.

Comment Figures 6 and 7: Looks like the color bar for RMSE is not correct. RMSE values can't be too low for the MB values shown in the middle panel. Maybe, you just copied the correlation coefficient color bar to RMSE.

Response Indeed, we made the mistakes. We corrected the color bar for RMSE in Fig. 6 and 7. Thank you for your corrections!

Comment Table 2: Statistical metrics in New NMC are only slightly better compared to conventional NMC. Mean Bias in New NMC is even slightly worse than conventional NMC. This indicates that emission perturbations did not have a very large impact on the analysis fields. This is somewhat unexpected considering large differences between the standard deviation between NMC and New NMC. Could you explain the reason behind this?

Response One of the reasons may be that our base model (CTL) does not capture the high variability of $PM_{2.5}$ over East China. We note that the RMSEs of CTL over East China are ~80% higher than those over South Korea (see Table 2 and Table S3). Despite the larger differences between the standard deviations between NMC and NMC+EMIS, it seems that our new method is limited to overcome the uncertainty in CTL simulations. We expect that if the CTL simulation could capture the spatial and temporal distributions of $PM_{2.5}$ with more accurate model inputs, e.g., better emissions and finer model spatial resolution over China, the NMC+EMIS simulation could show a better performance, as in the case of South Korea.