Response to reviewer 4

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 and suggestions.

General comment Given the structure of the 3DVAR cost function, any increase in BEC will lead to the analysis being closer to the observations, and further away from the model. Therefore, the reported better fit of the analysis with PM2.5 observations is consistent. However, there is the danger of statistical overfitting especially when the observation errors are potentially chosen to be too small. To convincingly demonstrate the improvement in the analysis requires a cross-validation approach: The randomly selected subset of the observations which is used for the evaluation should not be assimilated. I strongly recommend to carry out such a test to demonstrate the impact of the new BEC, especially the for the length scales, in a scientifically clean way.

Response To answer the review's comments, we carried out an additional analysis experiment with 20% of independent observations which were randomly taken out of the observations in the DA process and were used only for comparison purpose. Please, refer to pp. 9:262–pp. 9:268 and pp. 11:333–pp. 11:340 in the revised manuscript.

General Comment Besides the BEC, the choice of the observation error standard deviation influences the match between the assimilated observations and the analysis. The authors should therefore provide more details how the representativeness error (i.e. of the station observations for the corresponding model grid box) of the observation has been considered. It seems that the presented approach only accounts for the instrument error but not the representativeness error. It would be interesting to see if decreasing observation error SD has a similar influence on the analysis than increasing BEC. BEC and observations error need to be discussed in together as their relative differences determines the match of the observations with the analysis.

Response We totally agree with your comment that the balance between observation error (including the representativeness error) and background error determines the performances of data assimilation and short-term $PM_{2.5}$ predictions. However, we did not consider the representativeness error (also known as “sub-grid problem”) in current study. As far as we understand, several groups (including authors’ lab) have tried to evaluate these errors using an artificial intelligence technique. We have added a brief discussion of this point. Please, refer to pp. 13:400 – pp. 13:405.

General comment While it is acknowledged that the uncertainty of the emissions should be considered in the background error statistics, this quantification is not easy. Comparing two different emission data set, as done by the authors, will be dominated by the biases between the two data sets. The paper should provide more evidence that the documented in increase in BEC is a consistent estimate of the (unbiased) uncertainty of the emissions. The reader wants to know what the resulting uncertainty estimate of the emissions is.

Response Again, we agree with reviewer’s opinion. The biases between the two emission inventories dominate the comparisons of different emissions. Therefore, we removed the mean of perturbations for estimating BEC using GEN-BE v2.0. It is impossible to know the uncertainties of all chemical species in emission inventories. Please, note that the differences in the two emissions were not artificially created, but they were based on two fully independent emissions that had been established with two different techniques and statistical factors (such as emission factors and profiles, activity data, pollution control efficiency, regulation penetration, and economic growth rates) in East Asia. Given such difficulty in estimating the uncertainty in single emission inventory, we decided to use two independent emissions in our study to take into account the uncertainty in emissions in East Asia. Relevant discussions are introduced in our revised manuscript, too. Please, check out pp. 5:161 – pp. 5:167.

Comment The title suggests that the reader will get an information about the uncertainty of the emissions. This seems not the case.

Response Readers will get some information on “the impacts of uncertainty in emissions on data assimilation and short-term prediction in East Asia”. We have changed the title as: “An investigation into the impacts of uncertainties in emissions on aerosol data assimilation and short-term $PM_{2.5}$ predictions using CMAQ v5.2.1”.

Comment l 28: Please provide a quantification of the increase in the BE SD by taking the emission uncertainty into account.

Response We provided the quantification of the BEC SD. Please, refer to pp. 1:23–pp. 1:24.

Comment l 53 please discuss here also the representativeness error with respect to your model resolution

Response As mentioned above, we did not consider the representativeness error, but we have added a brief discussion on that errors. Please, see pp. 13:400 – pp. 13:405.

Comment l 130 Please comment if any additional temporal profiles (diurnal cycle, weekly cycle etc) were applied to the emissions during the simulation. The temporal variability might be a large source for the uncertainty of the emissions.

Response Good points! But, there is no detailed information available in our domain, as far as we know (although we used a temporal profile of the CREATE in our study!). This is obviously a topic we should work on more in the future. We have made some comments on the temporal profiles of emission inventories. Please, see pp. 5:144–pp. 5:145.

Comment l 149 Theoretically speaking, the error in H is the representativeness error

Response We have added a sentence into pp. 6:187.

Comment l 154 Please explain if the PM components are also modified by the DA or only the diagnostic PM2.5 field.

Response Diagnostic $PM_{2.5}$ fields estimated by Eq. 1 was modified by the DA. Regarding this point, please see pp. 7:198–pp. 7:206.

Comment l 185 Please provide more details here. PM2.5 simulations using the same emissions but different meteorological fields (i.e from different forecast lead times) can expected to be unbiased (i.e. inly a random error). However, this will not be the case if different emission data are used. How does the biases in the emissions turn into increased SD of the background errors. Did you remove these biases in the calculation of the variances and covariances as expected from the definitions of these statistical parameters.

Response: Once GEN_BE v2.0 used for estimating BEC starts to calculate the perturbations, it removes the mean of perturbations (i.e., "Stage 1" in the program). After that, the calculations of the variances and covariances start. Not direct bias correction was made on emissions because we selected $PM_{2.5}$ as the control variable. The perturbations are from simulated $PM_{2.5}$ variations caused by the changes in meteorology and/or emission inventories.

Comment l 214 I am not sure I understand this formulae. The term seems completely dominated by the 50 microgram/m3 constant value. Are you saying the SD of the observation errors is more or less 50 micorgram/m3 all the time? That would be quite a lot. Please compare the observation error SD against the SD of the background error (2, 4 or 8 microgram, see Fig 5)

Response We have corrected the value which is originally 1.5 μg m^{-3} from Schwartz et al (2012). The value given in the original manuscript was just a typo in the previous manuscript, and we used the value of 1.5 in our real work. Thank you for your comment and correction!

Comment l 224 Please express these numbers also in percent w.r.t to typical PM2.5 values.

Response We could not find the line you made a comment on. Could you provide the line? We will then correct the numbers.

Comment l 278 Please see my general comment. I think it is necessary to use independent (i.e. not used in the assimilation) observations to estimate the quality of the analysis.

Response As we answer in **General comment**, we carried out additional comparison of DA results with 20% of independent observations, which were randomly taken out of the observations and were then used for comparison purpose. Please, refer to pp. 9:262–pp. 9:268 and pp. 11:333–pp. 11:340 in the revised manuscript.

Comment L 298 Please provide more detail, how the analysis for diagnostic PM25 (formulae 2) is converted back into the prognostic aerosol variables.

Response We have added a sentence into pp. 7:198–pp. 7:206.

Comment l 325 Please provide a quantitative information about the assumed or inferred estimate of emission uncertainty.

Response It is difficult to find a way to quantify the emission uncertainty. As far as we are aware, it is almost impossible job to quantify the uncertainty of a bottom-up emission inventory. Again, given the difficulty in estimating the uncertainty in single emission inventory, we decided to use two independent emissions in our study to take into account the uncertainty in emissions in East Asia. Relevant discussions are introduced in our revised manuscript. Please, refer to pp. 5:161–pp. 6:167.

Comment l 328 A proper cross validation is needed to avoid overfitting (see general comment).

Response Again, please refer to added/modified paragraphs (pp. 9:262–pp. 9:268 and pp. 11:333–pp. 11:340).

Comment l 345 Please compare the SD of your BEC with the one derived with ensemble methods quoted in the literature.

Response The references you mentioned are NWP studies. The SDs in those literatures are all related to the meteorological variables. We think that the ensemble method used in recent NWP studies may be applicable to chemical DA. That is going to be our future topic to investigate.