Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-116-RC4, 2020

Geoscientific Model Development Discussions
Open Access

Interactive comment on "The impacts of uncertainties in emissions on aerosol data assimilation and short-term $PM_{2.5}$ predictions in CMAQ v5.2.1 over East Asia" *by* Sojin Lee et al.

Anonymous Referee #4

Received and published: 25 June 2020

Review of

The impacts of uncertainties in emissions on aerosol data assimilation and short-term PM2.5 predictions in CMAQ v5.2.1 over East Asia

Overview:

The authors extent the NMC method to calculate back error statistics (i.e covariances, BEC) for a regional PM2.5 data assimilation system (3DVAR) to account also for the uncertainty of the emission. Using the new BEC in the DA system, they show a better agreement between the observations with the analysis and improved forecast skills (1day forecasts) for the period of the KORUS AQ measurement campaign (14.5-16.6.

2016)

General remarks

The NMC method, applied in the more common way, quantifies the uncertainty of the tracer transport from of the meteorological forecast error. But, the transport error is only one contribution to the model (background in the DA terminology) errors of air quality models. Including the uncertainty of the emissions in the BEC is therefore an interesting scientific objective. The represented method leads to increased background error standard deviations and increased vertical and horizontal length scales.

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.

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 represenativeness 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

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.

Specific comments:

Title:

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

Abstract:

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

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

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.

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

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

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 vari-

ances and covariances as expected from the definitions of these statistical parameters.

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)

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

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.

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

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

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

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

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-116, 2020.