Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-116-RC3, 2020

Geoscientific Model Development Discussions

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 #3

Received and published: 21 June 2020

This paper aims at improving the assimilation of PM2.5 observations in chemical transport model by accounting for uncertainties in emission inventories which is a major source of uncertainties for atmospheric composition forecast. The approach consists in accounting for emission uncertainties in the modelling of the background error covariance (BEC) matrix. The BEC is estimated using a modified version of the NMC statistical method which is widely used in NWP. NMC consists in approximating the background error exploiting the differences between forecasts of different lengths that verify at the same time. While in the standard NMC, the forecasts differ only from their meteorological initialization, in the modified NMC which is used in this work, emission fluxes are spatially perturbated to account for their uncertainties. This method is

applied to the Community Multi-scale Air Quality (CMAQ) model using a 3D-VAR assimilation scheme applied to PM 2.5 observations. An experiment is done over East Asia using surface PM 2.5 observations in China and South Korea. The use of the new BEC leads to more accurate PM 2.5 estimates (increase in correlation and reduced biases) assessed against independent ground observations.

This paper addresses an important issue for PM2.5 forecast. However, it is not clear what is the actual contribution of this work since the method employed to adapt the BEC was already established in Kumar et al. (2019) for satellite AOD assimilation. Most results on the impacts of the incorporation of emission uncertainties on the background error variance and the horizontal/vertical length scales have been discussed in in Kumar et al. (2019). The differences include a distinct study area, distinct emission datasets and the assimilated variable which is PM2.5 instead of AOD. While the application of the method to a distinct experiment has a potential for publication, the present paper needs further developments including a better demonstration of its scientific contributions, more accurate explanations on the methodology and a deeper analysis and discussion of the results.

General remarks ——————

As it stands, the paper needs substantial improvements to fully understand the rationales and the key findings of this work. The scientific contribution of the paper with respect to previous studies should be better emphasized. While the overall goal is to test whether accounting for emission uncertainties in BEC modelling leads to more accurate predictions of PM 2.5, two or three discussions points should be identified in Introduction and clearly addressed in a separate section. The Introduction needs to provide more detailed and accurate background on BEC modelling and emission uncertainty quantification including a consolidated review of literature on these aspects. What are the other approaches to model BEC ? What are the limitations of the NMC method ? An alternative approach is to rely on ensemble of analyses using random perturbations of emission. Some authors have also investigated the possibility to incorporate the emission variables in the control vector and retrieve them using data assimilation. The use of references is not accurate enough. Some definitions (e.g. data assimilation, categories of uncertainties in CTM simulations) are not accurate enough or incorrect (e.g. background error covariance matrix). The structure of Section 2 (data and methods) needs to be improved. Some aspects of the methodology are not clear or incomplete to be fully understood and be reproductible. The paper does not provide a clear understanding of the conventional NMC versus the new NMC method. The paragraph on page 6 and Fig 4. do not provide enough details to understand how the NMC method was implemented. What are the differences between Met1 and Met2 ?. What are the actual differences between the three NMC implementations ? How the emission uncertainties have been incorporated in the NMC ? Part of the answers are in the text, but it requires too many efforts for the reader to find out.

I suggest having separate results and discussion section. I provide here some questions that could help to build the discussion:

- what are the limitations of the proposed NMC approach ?

- The variability in emission fluxes is accounted for using two inventories, is it enough to represent the spatiotemporal uncertainties in emission fluxes ? Would not be preferable to use spatial and temporal perturbations of emission fluxes based on a priori probability distribution function per type of emission.

- What are the benefits and limitations of assimilating PM2.5 compared to satellite AOD ? Is PM2.5 observation more directly related to the model variable ? What is the impact on the application of this method to over regions or historical periods which may not have PM2.5 observations.

Detailed remarks ————————-

Introduction:

- first paragraph: I suggest to give some key references for direct and indirect radiative

effects of aerosols.

- line 44: why limitation to GEO satellites ? there are a lot of atmospheric composition observations derived from LEO satellites (e.g: MODIS/TERRA, AQUA for AOD)

- “The inaccuracy of CTM simulations has been associated with uncertainties in emissions of primary air pollutants and meteorological fields as well as omissions of photochemical reactions occurring in chemical mechanisms (Han et al., 2013, 2015; Kim et al., 2017a; Song et al., 2012).”: The authors provide here specific examples of source of uncertainties impacting CTM simulations. I suggest to give the main categories of source of uncertainties: drivers and forcing variables (emission inventories, meteorological fields for offline simulation, land cover . . .), model structure (e.g. photochemical reactions, more or less realistic representation of atmospheric chemistry. . .), model parameters.

- line 48-57: The definition of data assimilation is well known and there is no need to repeat it here. This paragraph is somehow vague. Data assimilation in NWP context has mainly two goals:provide the best estimate of initial condition and provide an estimate of the uncertainties associated with the initial state, that could include emission uncertainties.

- The list of references is sometimes too long, Authors should select two or three references for a given statement and try to be more accurate.

- line 58: The background error covariance matrix is a key component of both sequential and variational methods.

- observation errors: Observation errors include gross error (e.g. cloud detection for aod satellite), measurement errors, representativity, observation operator uncertainties

- the background error is different than the model error.

*/The model error is the departure between the true atmospheric state at time k and the model prediction. The model error is represented by a dedicated model error co-

variance matrix. In strong constraint 4dvar the model is assumed to be perfect and the model error is neglected. In Kalman filter, the model error covariance matrix needs to be specified.

*/The background error is the error associated to the short-term forecast. In some assimilation system the background can be a climatology and not an output of the model. Part of the error in the background is due to the model but it can also be generated by other sources of uncertainties such as emission inventories. When the BEC is flow-dependent or in sequential assimilation scheme, the BEC is updated at each cycle and thus it is also influenced by the observation error used in the previous analysis.

*/I suggest to give here the main role of the BEC in terms of information spreading, information smoothing and balance properties.

- BEC modelling: Most methods derive the statistics of the background error from the departure between the observation and the background (expel: Hollingsworth, A., and P. Lonnberg, 1986), or using a surrogate quantity whose error statistics can be a good approximation of the unknown background errors (such as NMC). More recent approaches rely on ensembles of analyses. I suggest to provide here more background information on existing approaches to model the BEC including their advantages and limitations.

- line 69 : “Among the greatest sources of errors in CTM simulations (e.g., Elbern et al., 2000; Wu et al., 2008) are the uncertainties of emission inventories”. This sentence should be moved to the paragraph listing the sources of uncertainties affecting CTM simulations.

- Paragraph on methods to account for uncertainties in emission inventories: Not enough background is given on this central aspect of this paper. There are studies that have attempted to include the emission fluxes in the control vector and estimate them using ensemble data assimilation approach.

- need to clarify that PM2.5 is the output variable targeted in this work. What is the rationale for choosing PM2.5 instead of AOD ?

Section 2:

The structure of Section 2 needs substantial revision.

Section 2.1 includes several aspects that should be included in separate subsections, I suggest the following structure:

a/Study site and observations: âĂć the second paragraph of Section 2.1 concerns the description of the study âĂć it is not clear how the observations used for validation and data assimilation were selected âĂć What is the vertical footprint of the measured PM2.5 ? How does it compare with the modelled value ?

b/model:

b1/model description: a short description of the CMAQ CTM is missing. Providing the version of the aerosol and chemistry module is not enough, key references are missing. I suggest to give the main characteristics of aerosol and chemistry schemes (e.g. number of species and reactions for the chemistry, list of aerosol species for the aerosol scheme) along the main characteristics of the atmospheric transport model: (which type of advection scheme is used)

b2/model configuration: it should address time and spatial resolution, coupling between WRT and CMAQ, temporal period, location, output variables.

c/Emission datasets: Can you justify why these two data sets have been selected for this work .

d/Data assimilation and BEC modelling: Since BEC modelling is a central aspect of the methodology, a dedicated section should explain how it is parametrized and how the NMC method is used to estimate the BEC parameters in this work.

e/Experiment design: This section should include the statements given from line 178 to

199. A table summarizing all the experiments/simulation could be helpful. The various cases of implementation of the NMC method need some clarifications.

f/Validation methodology

- a clear definition of PM2.5 is missing: what is the vertical footprint of PM2.5 ? What are the differences between the modelled and the observed PM2.5 ?

- Some aspects of the methodology are not clear or not accurate enough

- Section 2.1

line 123: which conserving method ? please give a reference

line 125-130: this belongs to Results and not to Methodology section. “The differences in South Korea are relatively small, except for CO in the MIX emission inventory.”: Are you talking about the differences between the two databases ? I do not understand “except for CO in the MIX inventory”

line 130-136 on the use of MEGAN. Why are you using LAI from MODIS and GVF from VIIRS ? Are these variables required to drive MEGAN ? There is a possible inconsistency between LAI from MODIS and GVF from VIIRS ? Can you comment on it ?

- Section 2.2

the description of the cost function (l14-151) is a bit confusing. x is the control vector. x and xb contain the same variables (both are of the same size). x is the analysis and xb is the background. Are you also assimilating other variables which drive the chemistry or the transport model ? line 166 redundancy with Introduction line 172, not accurate definition of S: S represents the background error and its diagonal components are the standard deviation of the error of the background. What are the differences between the measured and the simulated PM2.5 ?

- Section 2.3:

line 211, eq 4: How a and b values have been chosen? line 213: replace ‘second criterion’ by ‘Eq 4 criterion’ line 213-215: this belong to the data assimilation section/BEC description. âĂć Some parts of the methodology are lacking such as the selection of observations for data assimilation versus validation.

Section 3

Section 3 should be dedicated to the presentation of the results. A separate section should address the discussions points. I shortly review the results but further review of them should be done if the manuscript is considered for publication.

- line 221-222: “To estimate the influence of the two . . .” : this belongs to the previous section

- why incorporating emission uncertainties in BEC should influence the vertical distribution for PM 2.5 ?

- l248: “In the DA process, the horizontal length scale determines PM2.5 increases in the horizontal spread of analysis” I do not understand this statement. The horizontal length scale refers to the horizontal correlation of PM2.5

- line 258: “The characteristics of the vertical and horizontal length scales, however, have not been fully explained in this study, thus requiring future”: The authors should further discuss this aspect and provide possible explanations.

- Section 3.2 first Paragraph: This belongs to methodology and should be described in the experiment design section.

Conclusion

The last two paragraph should be developed in a separate discussion Section. Part of it should also be used as background information in Introduction. I can see also some redundant ideas from the Kumar et al, 2019 paper.

Technical, writing ——————————-

- Lack of references in several part of the papers

- The use of a large number of acronyms makes the reading somehow very difficult.

- Result description needs to be improved, some sentences are confusing.

- the style is frequently not appropriate with a lot of uncertain and long sentences: for example “We found that the new approach exhibited a tendency to generate substantially increased standard deviations” , a tendency to generate . . . , I suggest using more direct sentences.

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