Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-147-RC2, 2017


Geoscientific Model Development Discussions
Open Access
EGU



# *Interactive comment on* "3D-Var versus Optimal Interpolation for Aerosol Assimilation: a Case Study over the Contiguous United States" *by* Youhua Tang et al.

**Anonymous Referee #2**

Received and published: 16 August 2017

General recommendation

The general topic of the paper is a comparison of the performance of two different data assimilation methods when using the same air quality model and AOD/PM2.5 observations. Although both data assimilation methods have been used with air quality models before, the models themselves have been different (CMAQ and GOCART) and so a direct comparison on their relative performance has not been possible. This paper discusses the different set up of each of the data assimilation schemes and looks at the results of a single assimilation cycle and then the more general performance over the month of July 2011. The differences between the results of the two data assimilation



Printer-friendly version

Discussion paper

schemes are linked to the difference in the set-up of the methods, in particular for the single assimilation cycle.

Overall the paper contains interesting and relevant information and results. Although the data assimilation methods are not new concepts or ideas, to see them compared in a more consistent manner provides useful information for anyone looking to establish an air quality model that utilises observations. Particularly in regards to the different scenarios where the schemes show stronger or weaker performance. My main criticism of the paper is that the descriptions of how the actual experiments have been set-up are not sufficient for someone without an in-depth knowledge of the models and/or the data assimilation code to understand exactly what has been done. I detail my specific queries in the major comments below.

Major comments

1. Section 2.1 – page 3, lines 13 to 16 and 23 to 24. There are two queries I have here about the OI set up that I believe are related. (i) The first is that I'm not clear how the background uncertainties have been formulated. I've also looked at the referenced paper but still have questions and so I think that a brief but slightly more comprehensive description would be useful here. As far as I can tell from the referenced paper, the background uncertainties have been made using a free running model and comparing to the observations. If I look at Figure 2 from that paper then this only gives values at observation locations. Is this correct? Generally B is formulated to give uncertainty values across the whole of the model domain not just at point locations, so why is this not done in this case? You also mention the diurnal variations – does this mean you use a different B matrix for the different assimilation cycles? This also only addresses the diagonal variances or uncertainties of the B matrix and not the cross correlations and this is my second question. (ii) You mention that the OI adjustment is made in each 11x11 grid horizontally and it's effect expands up to the PBL. This would normally be done through the B matrix, is this the case here? If so is it a cut-off Gaussian that has been used for the horizontal correlations, or a uniform distribution? Similarly for the

vertical correlations – you state that the surface PM2.5 OI increment is applied from the surface to the height of the PBL. Is this done after an increment is calculated by OI or through the formulation of the B matrix? Is it a constant adjustment at each model level or does it taper as you reach the PBL? If so how does it taper? All this is relevant because you attribute the differences in the schemes largely to this formulation of the B matrix.

2. Section 2.1 – page 4, lines 1 to 4. Again this question addresses the B matrix but now for the GSI scheme. The referenced paper was informative and from this I understand that the horizontal and vertical correlations that make up the B matrix are Gaussian (and again it might be worth mentioning that here) and I assume the GSI package has been used to create these? Although it talks about optimisation in the referenced paper, I couldn't find anything to suggest these length-scales should be the same across the domain. Are the background errors and length scales the same across the domain in this case? It is unclear from the description given whether these plots are a single point representation or not. I also wasn't totally clear what is being used as the control variable for the AOD. Is it just ASO4J? You also mention that other aerosol species have proportional model or background errors. Were these just computed for interest or do they also feed in to the modelled AOD?

3. Section 2.2 – page 5, lines 17-40. The final issue that I really struggled to understand was how the CMAQ forecast model and the GSI data assimilation method actually interact. Perhaps a schematic attached to this section would be more informative, as it's clearly quite a complex process. Are the species given in the forecast by CMAQ, translated to the variables required by the CRTM's GOCART aerosol as stated in Table 1, then run through the forward model calculation to give AOD, assimilated by GSI and an increment returned for each of the GOCART species? How is this increment divided back in to the CMAQ variables? Once this is done I assume that CMAQ is used to forecast the species on to the next assimilation time? Is the B matrix you describe also a plug in to the GSI? Is the resolution the same for both models? If a





Printer-friendly version

Discussion paper

figure could be created that shows this process it would add a lot to the understanding of this section

4. Section 3.1 – page 7. How do you distinguish the effect of assimilating AOD from the effect of the PM2.5 measurements? Are three different experiments run (one with just PM2.5, one with AOD and one with both)? Given that the GSI method solves a cost function, I'm not clear how you distinguish the effects from two different observation types in one experiment. For the OI, as far as I understand an increment is given by the AOD and PM2.5 observations separately, although this needs to be described/clarified in further detail (see my question 1 above), and so I can potentially envisage how this could be done with one experiment. However, this needs to be made clear.

Minor comments

1. Abstract – page 1, line 26. I think perhaps 'background error uncertainties and the horizontal/vertical length-scales of the covariances' would be a more accurate description.

2. Figure 1. Are these errors and length-scales static quantities used across the domain or for one specific point.

3. Figure 4 is in the wrong order.

4. Page 9, line 14 - 'assimilation for the elevated layers blow PBL.'. Should be below.

5. Page 9, line 10 - 'steeper' not 'steepen'

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-147, 2017.



Printer-friendly version

Discussion paper