Response to Referee A. Benedetti's Comments

Thank you, Angela for taking the time to review the paper. We very much appreciate all of your comments. Please see our responses below.

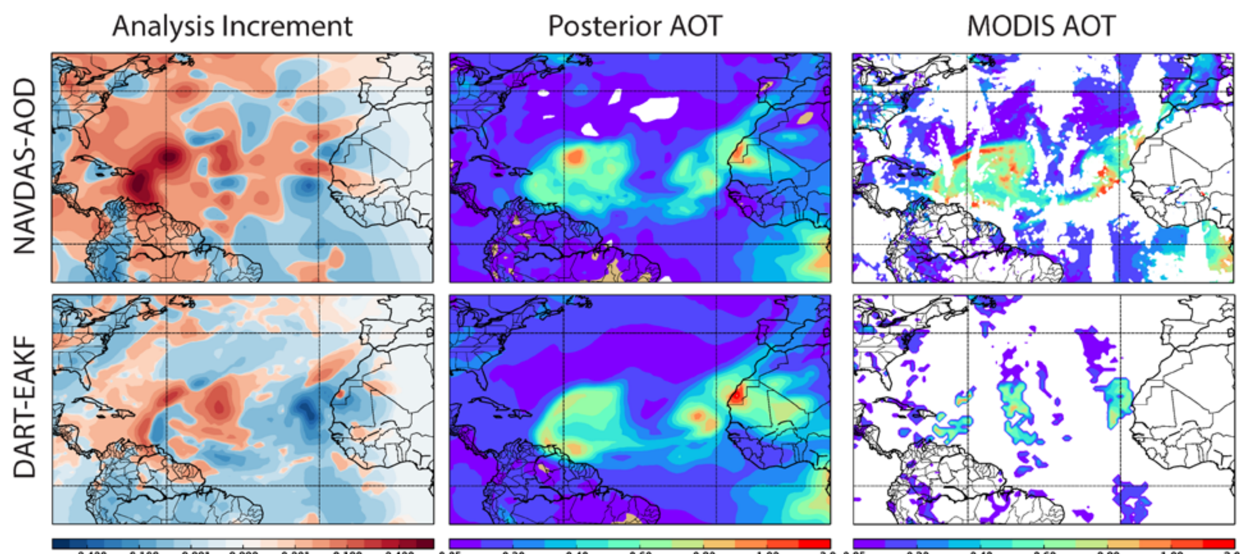
1. Page 4 Line 33.I would say "research" arena rather than "operational" arena as to my knowledge at the moment there are no operational ensemble systems for aerosols (although the situation may soon change).

   Response: We agree, we will update this statement to make it accurate. Thanks.

   Manuscript change: A core rationale for developing ENAAPS was to experiment with ensemble data assimilation techniques which have been successfully **implemented at operational centers on an experimental basis** (e.g., Sekiyama et al. 2010).

2. Page 5 Line 10. Here, like elsewhere where the comparison between the ensemble and variational systems was made, I thought it would be good to see the background error covariance matrices for the ensemble and the variational system side by side. Perhaps, if possible, for future work as well, it would be interesting showing the increments from a single observation experiment to show how the different background error statistics affect the distribution of the increments and spread to neighbouring points the information from a single observation

   Response: We agree that showing the analysis increments and error covariances would be helpful. We do show analysis increments for our Saharan dust case in Figure 15, but we can add a few other examples in the supplementary material. Yes, we agree that some single observation experiments would be nice to show as well. We might not show those types of experiments in this paper since it is already quite long, but we definitely will in subsequent papers.



Analysis increments (posterior-prior AOT) and posterior AOT fields from the 2DVar NAVDAS-AOD and the DART-EAKF for a dust event on August 2, 2013 (18Z).

3. Page 6 Line 22. How is the adaptive inflation estimated? Is it based on first guess departures? I know that the reader can look up the references, but just a sentence to explain briefly what the estimation is based on would be welcome.

   Response: We added some more information on the adaptive inflation in the methods section. Thank you for the suggestion.

   Manuscript change: An alternative method to a uniform multiplicative inflation is adaptive covariance inflation (Anderson 2009) which produces temporally and spatially varying inflation factors. **This approach is based on a Bayesian algorithm that estimates the inflation with time as part of the state update, using a normally distributed inflation factor associated with each element of the model state vector. An initial inflation factor of 1 (ie. no inflation) was set for all locations and a fixed standard deviation of 0.4 was used.**

4. Page 7 Line 12. 25% seems like a large perturbation, although later you say that it might be small for certain emissions (for example fires). How is this value assigned? I am surprised that location-dependent perturbations did not help with the ensemble performance, as you later mention that for localized sources the ensemble had the problem of over-correlating them. Perhaps the perturbations should be a function of the source spatial extension and intensity. I really do not know, just wondering.

   Response: In general, we have seemed to get the opposite reaction, that the source perturbation is small. We selected 25% as a pretty conservative estimate of the source perturbation with the expectation that we would evaluate the system performance and see what adjustments needed to be made. In general, we found that the system did pretty well in representing uncertainty (spread ~ RMSE) with this perturbations with the exception of fire-impacted regions. For these regions, we weren't getting enough spread, especially for high AOT events. This tells us that we probably need to increase the source perturbation for fires. Fire emissions are also highly uncertain, so needing perturbations larger than 25% for these emissions is not unexpected.

   The perturbations to the aerosol sources aren't location-dependent (we will work to make this clearer in the methods). We initially tried random perturbations that were drawn for each grid, however, we ruled this method out (page 28079, lines 20-21). The method that we did use in this work was to apply a randomly drawn perturbation for each aerosol source and for each ensemble member. This essentially creates large correlations between all emissions of aerosol of a given source-type (dust as an example), only limited by the localization. So for emissions in which the correlation lengthscale is smaller than the localization lengthscale (such as pollution sources), we identified issues. For events in which the correlation lengthscale is greater than the localization lengthscale, this method worked well (ie. large smoke and dust plumes). We have plans to reassess the source perturbations in future work to better deal with emissions for pollution events and small fires.

5. Page 11 Line 9. Please do explain briefly the methodology behind AI

   Response: We agree. We added some additional information in the methods section that we think will clarify the AI discussion in the rest of the paper.

6. Line 19. That points to model shortcomings which are not likely to be corrected with DA

   Response: Yes, we agree. We think that this is related to issues in the smoke emissions. The adaptive inflation is trying to correct for inconsistencies in the model prior and the observations by inflating. The smoke emissions are persistent during this time period and are likely contributing to the discrepancy between the model distribution and observations which leads to over-inflation by the AI and eventually, a crashing of the model. This is why we think we need to do some tuning to the smoke emissions and hope that this will alleviate the problem. We hope that once we tune the smoke emissions, the AI can be run without any measures for preventing inflation blow up.

   Manuscript change: The inflation factor blow up indicates that the discrepancy between the prior and observational distributions increased over time, producing unrealistic AOT values and aerosol mass concentrations, eventually leading the model to crash. **This type of behavior is indicative of model shortcomings related to smoke aerosol.**

7. Page 12 Line 11. Well phrased. This is another one of the issues related to the fact that the aerosol problem is under-constrained.

   Response: Thanks!

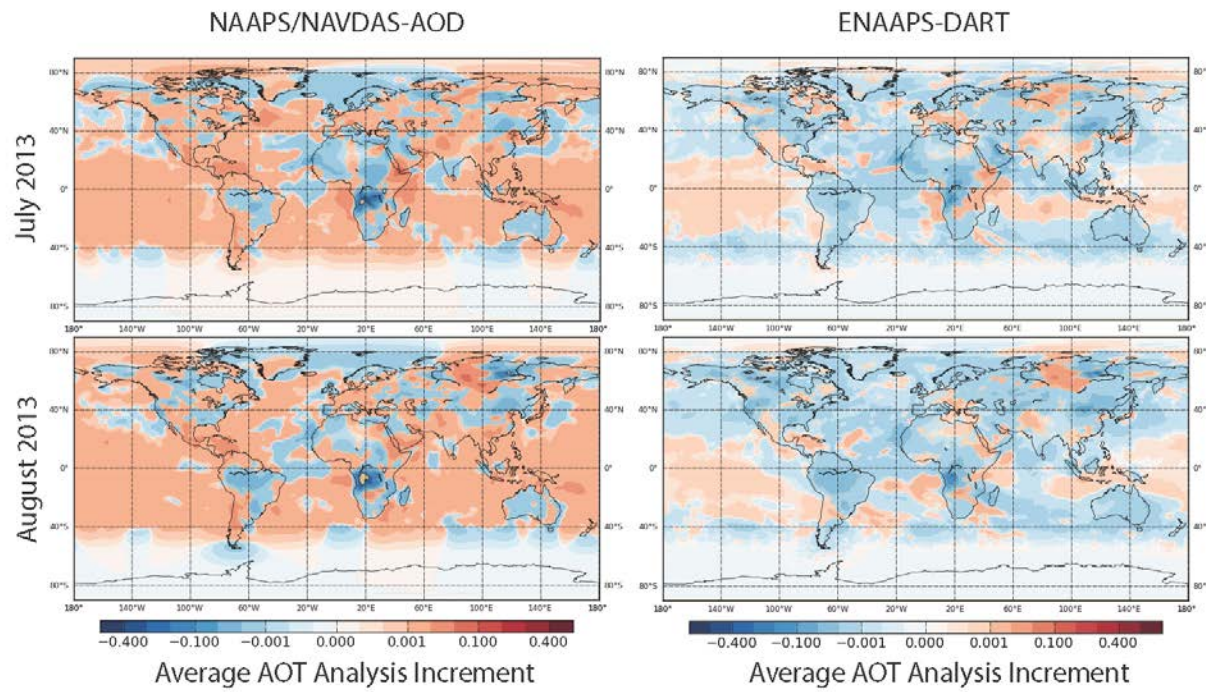
8. Page 15 Line 35. An interesting conclusion about the observation errors being too large for small AOTs. Perhaps the methodology of Desroziers et al (2005) could be applied to ascertain so in a more mathematical way. [Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005), Diagnosis of observation, background and analysis-error statistics in observation space. Q.J.R. Meteorol. Soc., 131: 3385–3396. doi:10.1256/qj.05.108]

   Response: We should clarify that we mean that the observational error is too large relative to the ensemble spread. This is probably more likely due to aerosol being a positive-definite, therefore, it is hard to get enough spread near-zero. We added some discussion on this point. However, this still has an important implication for data assimilation in that the observations won't have much impact. We have been discussing different ways to deal with this issue, including doing a data transform on the observations before assimilation.

9. Page 19 Line 12. The fact that the RMSE values of the two analysis are not statistically different might also mean that the system is driven more by the observations than the background, and perhaps the observations errors are too small. This may seem to contradict what said on page 15 line 35, but the two things may co-exist as the balance is to be obtained between the background errors and the observation errors and it is possible that the analysis draws too much to the observations (i.e. the background errors are large with respect to the observation errors). Again, perhaps an analysis of the departures of both the variational and ensemble analyses could offer some insight on this particular aspect.

   Response: Yes, we agree, the observations are pulling the priors in the two systems to similar values (although there are differences such as over ocean where we aren't verifying with AERONET). The ensemble system tends to produce larger AOT values (positive bias) and the observations in general pull the AOT lower. The deterministic system tends to produce smaller AOT values (negative bias) and the observations tend to pull the AOT higher (see figure below).

However, the analysis increments tend to be smaller for the ensemble system than the deterministic system. This means that the forecast error as specified for the 2DVar system is quite large and puts a lot of weight toward the obs. The forecast error determined by the ensemble is smaller and as a result has a smaller analysis increment (ie. less obs impact), this would indicate that the ensemble system is doing a better job in the short-term forecast (prior) at least at AERONET sites



10. Page 19 Line 15. Please use another verb other than "produced", like "displayed".

    Response: Thank you, we updated this and similar wording throughout the text.

11. Page 19, Line 39 As already mentioned, it would be good to see a plot of the background error covariance matrices for the variational and the ensemble system (single observation experiment increments would also do the job). Figure 15 shows some of this, but it would be good to have a dedicated single observation experiment.

    Response: We may reserve single observations experiments for subsequent papers since this paper is already quite lengthy.

12. Page 20 Line 16. To be fair to the variational system, it is definitely not tuned at all to capture sharp gradients. I presume the 2D-Var background error covariance matrix is spatially homogeneous, constant and with fixed correlation length. It seems to be asking too much of the system.

    Response: Yes, we agree that the 2DVar system won't be able to capture gradients based on how the error covariance matrix is defined. This is meant to demonstrate why an ensemble approach might be the chosen approach moving forward or at least should be part of the operational runs, especially as we begin to incorporate spatially-limited observations.

13. Page 24 Line 32. Have you looked what happens at longer forecast ranges than 24h?

    Response: No, not at this point in time. However, we have plans to implement this system semi-operationally and will begin to evaluate forecasts out to a few days.