Overarching response:

We would like to thank the reviewer for the thoughtful review.

The focal point of this work is to bring to the notice of the scientific community the issue of covariance/correlation structure between parameters while performing parameter estimation. There is some work done by the scientific community in the area of ensemble based parameter estimation. We feel that the important issue of covariance has received scant attention. In this work we have demonstrated the importance of this covariance structure and introduced a technique to estimate the optimal covariance structure.

This work also tries to demonstrate the feasibility of estimation of spatially distributed parameters using real satellite data. As the reviewer has rightly pointed out we have made several simplifying assumptions. We are aware that these assumptions could lead to erroneous estimates. These assumptions include:

1. The aerosols over North Africa are purely composed of dust.
2. The aerosols lie in only one size bin.
3. There is no feedback from the aerosol to the meteorology.
4. The resolution of the model used is coarse (81kms).
5. The model is assumed perfect except for error in erodibility.
6. The dust boundary conditions hold only

approximately.

7. The observational errors used in MODIS data assimilation are slightly less than realistic errors.

These assumptions are clearly too simplistic. Therefore this work should be considered as a first step towards ensemble based estimation of dust source. Please note that we have avoided recommending the tuned map to be used in realistic/operational models. We have discussed these limitations of this work in the last two sections.

As pointed out in the last section there are a number of potential improvements that could be implemented in future work. This include relaxation of the assumptions made in this work. However it is not clear how the boundary layer meteorological state can be constrained given the paucity of meteorological observations in the north African domain.

Please find our responses to questions below.

Q. *How certain are the authors that the MODIS AOT (used in this paper) is determined by dust? This may be an incorrect assumption that can potentially seriously impact your results. Stating, as the authors do, that they are only considering the dust season is not sufficient.*

Response: This assumption holds true to a large extent in reality. Over Sahara we are certain that it is dust. Not only due to the landscape, but also meteorologically speaking the time period of study coincides with the Saharan dust maximum. If the study were conducted in, say, January, there is the possibility that biomass burning emission would enter the southern domain. Please see our "Overarching response" at the beginning of this document.

Q. *Why assume that erodibility is wrong but the threshold windspeed is not? In all likelihood, both are wrong and the newly estimated erodibility maps will be affected by this (correcting for incorrect threshholds where possibly erodibility is reasonably estimated)*

Response: The reviewer is correct in pointing out that the estimated erodibility could compensate for errors in threshold windspeed. For sure, the assumption that the threshold windspeed is correct is too simplistic. This work should be considered as a first step towards ensemble based estimation of spatially extended parameters in aerosol models, rather than an estimate that can be used in operational or realistic dust forecast. We have corrected the text to bring this issue to the notice of the reader. These corrections are mainly in paragraph 3 in section 6.1 and paragraphs 5 & 6 in section 7. Please see our "Overarching response" at the beginning of this document.

Q. *The major assumption here is that erodibility is constant in time (at least over several weeks). This raises at least three issues:*

Q. a. *Is the technique (a filter with an assimilation cycle of 24 hours) appropriate? Possibly longer cycles and time-averaged observations should be used.*

Response: It would be interesting to compare these results to those obtained from 4d-Var or a smoother approach. Given that dust production at a single location will ultimately impact AOD at numerous down-stream locations over time it is likely that such distributed-in-time techniques would be beneficial. Unfortunately, such experiments are beyond the scope of the current manuscript.

Q. b. *Do the erodibility maps derived from real observations truly converge with time for the filter used in this paper?*

Response: Please read our response (below) to question ***Q.*** *28861,12*.

Q. c. *If the erodibility is not constant (this possibility is hinted at in Section 7), what shall be the real application of the technique developed in this paper?*

Response: The technique presented in this work should work as long as the true correlation length scale does not change with time. This work shows that the filter is able to identify the optimal correlation length only if the imposed length scale in the prior is longer than the true length scale. The filter fails to identify the true length scale if the true length scale is longer than the imposed length scale. If the true length scale is changing with time a possible alternative approach could be explicitly estimating the correlation length by including it in the augmented state vector. Also one can imagine periodic retuning to account for non-constant erodibility.

Q. *28840,15: A few more references to aerosol modelling and its uncertainties would be welcome, e.g. AEROCOM results. Also, both Huneeus et al 2012 (ACP) and Schutgens et al 2012 (Remote Sensing) developed Kalman filters/smoothers for estimation of aerosol emissions. Huneeus et al. contains a long list of papers discussing emission uncertainties. There are other papers that attempt to* estimate aerosol emissions, a topic very close to your own (e.g. Dubovik et al. 2008), using different techniques.

Response: We have added references to Huneeus, 2012; Schutgens, 2012; Dubovik, 2008 and papers discussing aerosol modeling and emission uncertainties etc including AEROCOM results (Cakmur et al, 2006; Cooke & Wilson, 1996; Lavoue, D., 2000; de Meij et al, 2006; Textor, 2007).

Q. *28841,5: The paragraph after this list seems to belong to other sections: methodology, results and summary.*

Response: This paragraph was put in to give a preview of methodology and main results to the reader before she/he reads the whole paper. We have removed this paragraph.

Q. *28842,13: What is the highest atmospheric level in this model?*

Response: The highest level is 31 km. This information is included towards the end of paragraph 1, section 2.

Q. *28842: It is not clear what happens to the meteorological variables. Obviously they are calculated by the model but are they also nudged to any meteorology? Section 3 seems to suggest so "The OSSE is run using the meteorology of June/July 2009" 28846,21.*

Response: As such the meteorological variables are not nudged to any meteorology. However the ensemble boundary conditions used every 6 hours are the analysis from NOGAPS and hence contain observational information. These analyses represent the best estimate of the environment given the FNMOC operational DA scheme.

Q. *28843,10: I find the explanation of erodibility a bit confusing. If erodibility is only a surface weighting function, why is it called erodibility? If it is related to a surface's ability to emit dust , why is it's value between 0 and 1? The definition of erodibility is not clear (but should be as it is central to the paper). As far as k (Eq 1) is concerned: if it is determined from a fit to observations, assumed values of erodibility must have been used. What values were used and how certain are they?*

Response: The erodibility gives the fraction of the grid box covered by dust. We have added this sentence in the second paragraph following equation 1. Erodibility is also called dust fraction. A value of 1 means that 100% of the grid box is covered by dust. We call it a "dust fraction" but really even individual dust emitters have differing susceptibilities or emission factors based on such things as the availability of saltators.

Researchers have arrived at the value of k used in equation 1 after undertaking extensive and carefully designed field campaigns. These campaigns were carried out in the United States. Please see the details in (Gillette and Passi, 1988). This formula (Equation 1) and the value of k used has been used by other scientists and it has been seen that it gives good simulations of dust events.

Q. *28844,2: AOD is not a measure of the amount of dust, it is merely an indication of the amount of dust over regions where dust dominates AOD.*

Response: Throughout this work we have made the simplifying assumption that over the North African domain dust overwhelms other type of aerosols. Please read "Overarching response" at the beginning of this document.

Q. *AOD is often used but an odd term nevertheless. i suggest AOT: Thickness refers to the complete aerosol layer. Depth actually refers to a level within that layer.*

Response: The reviewer is absolutely correct in this definition. However, AOD in colloquial diction is more often used in literature than AOT and hence we have continued with that convention. However, we will keep this in mind in future papers.

Q. *28844,5: Do you have a reference to the source of this value of* *extinction? If it is based on e.g. Mie calculations, what sort of particle (distribution) was assumed? I gather the mean size is 2mu?*

Response: The size of 2 micron is used. We have mentioned this (last line of the second paragraph following Equation 1). The references are:

Reid, J. S., E. Reid, A. Walker, S. J. Piketh, S. S. Cliff, A. Mandoos, S. Tsay, and T. F. Eck (2008), Dynamics of Southwest Asian dust particle size characteristics with implications for global dust research, *J. Geophys. Res.*, *113*, D14212, doi:10.1029/2007JD009752.

Reid, J. S., H.H. Jonsson, H. B. Maring, A.A. Smirnov, D.L. Savoie, S.S. Cliff, E.A. Reid, M.M. Meier, O. Dubovik, and S-C Tsay (2003), Comparison of size and morphological measurements of coarse mode dust particles from Africa, *J. Geophys. Res.,* 108(D19), 8593, doi:10.1029/2002JD002485.

Q. *28844,6: Am I right in assuming that dust is represented by a single size bin (or mode with fixed size and width) in the model? Please discuss this. How realistic is this (various deposition processes vary in efficiency a lot from 0.5 to 10 micron). There is no feedback of dust on the meteorology? Finally, what is the impact of other aerosol? Your model does not include those? They nevertheless exist in real life.*

Response: Yes, we have assumed that the dust is represented by a single size bin of 2 micron. In reality the dust has different sizes and as pointed out by the reviewer has different depositions. There is no feedback of dust on the meteorology. We assume that dust is the only aerosol

over this domain. Please see our "Overarching response" at the beginning of this document.
We have inserted last three paragraphs in section 7 to clarify these issues.

Q. *28845,5: Eq (2) and before seem to ignore dust emitted in previous time steps but not advected. If this dust is included in the transport term, then please explicitly mention this.*

Response: The dust emitted locally but not advected is included in the transport term. A sentence is inserted in the paragraph following equation(2) to reflect this.

Q. *28845,10-20: I suggest to move this to where you initially define and discuss erodibility.*

Response: We moved line 10-14 to the paragraph where we initially defined erodibility (see second paragraph after equation 1). We did not move the remaining lines because they have continuity with the next paragraph "In the current work…". Also the reader does not know what is AOD at the place we have initially defined erodibility.

Q. *28846, section 3: the text is confusing because the authors hop from topic to topic within one paragraph. I would suggest separate paragraphs in the following order: 1) ensemble DA and estimation of erodibility. give Kalman filter eq, discuss generation of ensemble (erodibility), discuss spin-up and assimilation cycles 2) treatment of boundary conditions in ensemble DA framework,* *including boundary conditions for aerosol 3) OSSE, assumptions on erodibility and threshhold windspeed (values, perturbations), generation of synthetic observations*.

Response: We have inserted a separate paragraph (paragraph 2 in section 3) to discuss ensemble boundary and initial conditions. This paragraph reads, "The meteorological boundary and initial conditions are obtained from Navy Operational Global Atmospheric Prediction System (NOGAPS) global model (Hogan & Rosmond, 1991). Ensemble analysis boundary conditions are used every 6 hours. These ensemble analysis are obtained by the local Ensemble transform technique (McLay, et al., 2010). The ensemble analysis is used as initial conditions so that each ensemble member is a different realization of meteorology. Since each ensemble members corresponds to a different realization of initial and boundary conditions the advection (that is wind) is different for each ensemble member. The resulting spread in the boundary layer wind is of the order of 0.7 m/s. For the lateral dust boundary conditions we are assuming that dust does not enter the domain, which is quite large. For the period of our study there is no dust storm east of the Arabian peninsula. So these dust boundary conditions approximately hold. This approximation may impact the estimation results in the real data experiments but it does not impact the OSSE results in any way."

As for the other details, we think that they become apparent as the paper proceeds. For example, the generation of ensemble of erodibility is discussed in section 3, paragraph 5 and also in section 5, paragraph 1. Though both, sections 3 and 5 discuss OSSE, in section 3 the erodibility is uncorrelated and in section 5 the erodibility is correlated. Hence it would be confusing to describe the generation of erodibility ensemble only at the start of section 3. Similarly with observations, in section 3.2 the AOD is observed at all grid points while in sections 4 and 5 it is observed at 20% of the grid points.

Q. *28846,9: It is a bit confusing that the set-up of the OSSE is discussed at the same time as the treatment of the meteorological boundary conditions. They are really different topics, please deal with them in different paragraphs. By the way, where do the different realizations of boundary conditions come from? NOGAPS is not an ensemble DA as far as I know. I suggest a separate paragraph (subsection?) to explain the treatment of boundary conditions in an ensemble DA context, not just the meteorology but also the aerosol itself.*

Response: We have inserted a separate paragraph (paragraph 2 in section 3) to discuss ensemble boundary and initial conditions. (Please read response to the previous question).

Q. *28848,24: "These values are chosen after experimentation with different values." what does this mean? How were those values chosen? Did you thoroughly examine parameter space or merely consider a few (likely ?) values?*

Response: We examined a few likely values. It is very time consuming to thoroughly examine the parameter space.

Q. *28849,3: Depending on the situation, you adjust either individual members or the mean. I guess the mean in an ensemble system can only be adjusted through its members. A bit more detail here will be appreciated.*

Response: In the EAKF the mean is updated and then the forecast ensemble is transformed into the analysis ensemble. We have not given the details to keep length of the manuscript manageable.

Q. *28849,23: Instead of calling this alfa-up, why no alfa-posterior? The paper already has an alfa-prior and alfa-up may be thought to be related to something upstream.*

Response: Though both *update* and *posterior* are widely used in literature we feel that update is more intuitive in that it suggests the estimate is being *corrected* by observations. As far as the confusion with upstream is concerned this symbol has been hardly used in the remainder of the paper. We have mentioned *update* explicitly wherever the need arises.

Q. *228849, Eq 3: this is the Kalman filter equation under this paper's simplifications. Please mention this. Better yet, refer to the full equation which you quote in an earlier section.*

Response: We have inserted a sentence after the equation 3 which makes this clear.

Q. *28850,1: "error variance is given by var (AODobs) which is set to 10 % of the mean observation" This information should really be in a separate paragraph, in an earlier section where the DA system is discussed. Also, even though you are working with synthetic obs and are free to choose your obs errors, I doubt 10% is a realistic estimate for real errors, certainly for obs over land.*

Response: We have moved this information to the introductory part (paragraph 5, section 3). In the OSSE the observational error is 10%. For the real data experiments the error is more than 10% and is given by (0.15 AOD units + 10%). Our indication is that these errors are largely spatially correlated, (e.g. Shi et al., 2012 in review). So while there is a large mean bias, we ignore it for the purposes of demonstrated the technique. Please see:

Shi, Y., Zhang, J., Reid, J. S., Hyer, E. J., and Hsu, N. C.: Critical evaluation of the MODIS Deep Blue aerosol optical depth product for data assimilation over North Africa, Atmos. Meas. Tech. Discuss., 5, 7815-7865, doi:10.5194/amtd-5-7815-2012, 2012.

Q. *28850,11: "the uncertainty in the AODprior ensemble is due to the uncertainties in local α, local uâ ́LU ̊, upstream α, upstream uâ ́LU ̊ and winds". Why do winds differ among ensemble members? Only because of boundary conditions or are there additional reasons as well? How large are the variations across the ensemble? It would be good to discuss this earlier, in section 2 or when discussing the ensemble DA system.*

Response: The winds differ among ensemble members because each ensemble member uses a different boundary condition. We have inserted this information in the section 3, paragraph 2.

Q. *28851,1-7: Again, most of this should have been introduced in a methodology section, where localization is also discussed. Later, one can then simply refer to that section while discussing figure 2.*

Response: We are assuming that some of the researchers reading this paper will be newcomers to the field of data assimilation/parameter estimation. Therefore we feel that the reader should have seen equations (3) and (4) to appreciate *spurious covariances* and their harmful effect on estimation. Of course for a reader who has worked in the field of data assimilation it is more convenient to have these concepts briefly explained in the introductory part and then referred to later in the text.

Q. *28851,10: "In this work the mean of the simulated observations is not perturbed. The difference between the observation mean and prior AOD mean is termed the "innovation" " Should have been defined & discussed before.*

Response: This paragraph as been deleted in response to the next question.

Q. *28851,10-25: In my opinion, this explanation is not needed. It follows from the equations. At the very least, I suggest shortening* *it*.

Response: We have deleted this explanation.

Q. *28852,3: "The covariance between α and AOD at the* observed *grid point informs what part of the innovation is used in the increment." There are no parts to the innovation. I think I understand what the authors intend, but this sentence needs rephrasing. Also, spurious covariances are discussed once more. It is better to introduce this terminology early on and later show examples. At present, the explanation is very haphazard.*

Response: We have replaced the sentence by:

"The covariance between α and AOD at the observed grid point determines the weight given to the innovation in the calculation of the increment" .

We hope that this is a better usage. We have defined *spurious covariances* earlier in the section. In these sentences we are only emphasizing that small ensemble size leads to more inaccurate estimates of covariances if the true covariance is small.

Q. *28853,9: "The observational error is set equal to 10% of the mean observation, consistent with instrument accuracy". This is NOT consistent with instrument accuracy because AOD error from satellite, especially over land, does not depend on instrument calibration. Rather, retrieval assumptions (cloud free, surface albedo, particle species) determine these errors. Again, I think 10% is too small.*

Response: We have removed "consistent with instrument accuracy". We have changed the sentence to
"The observational error is set to 10% of the mean AOD observation. This observational error is motivated by AOD satellite data whose error is at least 10% of mean observation."
The error used for real data experiments is higher than 10%.

Q. *28854: The association of high erodability and high advection noise seems a bit tenuous. Advection noise depends more strongly on friction speed in neighboring grid points. Would it be more instructive to show the correlation between AOD and erodability in contours in Fig 4?*

Response: Advection noise depends both on friction velocity and erodibility. It does depend more strongly on friction velocity than on erodibility. (The dependence of AOD on erodibility is linear while on friction velocity is to the power 4) In the extreme case, say a particular region has very high friction velocity but zero erodibility then it will not be a source of advection noise. In W1 and W3 there is considerable overlap between high erodibility and high friction velocity contours. It is really the combination of high erodibility and high friction velocity that gives rise to the advective noise. We feel that showing the contours of correlation in Fig. 4 can potentially confuse the reader.

Q. *28855,16:"The reason for the bad covariances is a* *combination of the effect of advective noise and the small size of the ensemble". Ensemble size: yes. Advective noise: no. The advective "noise" is an integral part of the covariance and allows you to use down-stream observations to estimate erodibility. Model errors will cause bad covariances because e.g. transport is poorly represented. That is one likely reason that you need to use a cut-off radius.*

Response: If the ensemble size were very large than the filter would be far better at filtering out the advective noise. So for a large ensemble size the advective noise does not matter. However given a small ensemble size the advective noise could adversely affect covariances (apart from the model error affecting covariances). So we are saying that the advective noise in combination with small ensemble size is problematic. It is true that model errors (in transport) also give rise to bad covariances. The main comparison in this work is between correlated erodibility experiment and the uncorrelated erodibility experiment. The model error due to transport is present in both the experiments. Also just providing a cutoff radius did not improve the results; we had to additionally impose a correlation.

Q. *28856,18:"then constructing an ensemble-member by ensemble-member weighted average". I would suggest "a spatially-smoothed perturbation for each ensemble member separately".*

Response: We have corrected this.

Q. *28857: I find this an interesting result, that an optimal correlation length for erodability perturbations can be found through use of the filter itself. But this begs the question what exactly defines this correlation length. I believe the authors leave this question open. One issue are the synthetic observations. They are sparse but the method for sampling them is not explained clearly. However, I suppose it is really the spatial correlation in the true erodability map that determine this correlation length. Did the authors investigate this?*

Response: We have explained the method of generating synthetic observations in the first paragraph of section 4. We did not investigate the issue of the true correlation length scale in detail. We propose to do so in the future using a simplified tracer model. Based on the work we have done we think that this optimal length scale is principally dictated by the true length scale and advective length scale.

Q. *28858: As a follow-up: how can you be certain that results from tuning experiments for an OSSE can be applied to real data?*

Response: It cannot be guaranteed that the results from OSSE are directly applicable to experiments with real data. We have modified the last paragraph of section 5 to include this caveat.

"Though results from the OSSE experiments are not guaranteed to hold for experiments with real data, they do provide valuable insights into the tuning of erodibility …"

Q. *28858,15:"This confirms our hypothesis that correlating* *perturbations leads to decrease in advection noise thereby improving the covariance estimates". Can figures like 2c verify this? I would suggest to verify this hypothesis: is it possible that the advection noise does not decrease but that the reduction of degrees of freedom in your system increases the impact of observations?*

Response: It is hard to verify this hypothesis directly because apart from advective noise many other factors determine the AOD spread. The spread in the AOD is indeed lower in the correlated experiment compared to the uncorrelated experiment. We have modified the text to include the effect of reduction of degrees of freedom. Please see the paragraph 5, section 5.

Q. *28861,12:"After 28 update cycles the mean of tuned α converges to values shown in Fig. 8c.". it would be good to show that α truly converges. Given the number of simplifying assumptions you have made, it is possible that α does not converge but continuously adjusts as dictated by the observations (especially since you keep windspeed threshold constant). Proving that α converges in a large part of the domain, would add considerable strength to your argument. As the estimated α are much smaller than the operational α, care should be taken in defining a convergence criterion.*

Response: We have added figure 9. The following discussion about this figure is added in paragraph 3, section 6.1.

"The estimates of α as a function of the update cycles at four different grid points is shown in Figure 9. For the grid point in panel(a) the estimate decreases from 0.3 to about 0.05. The convergence is not smooth but clearly the estimation corrects a bias in the first guess in the downward direction. Between cycles 10 and 28 the mean wiggles between 0.05 and 0.1 rather than staying at a constant value. This is because the estimated erodibility can compensate for other errors in the model like those in threshold velocity and advection. Similarly in panel (b) the estimation corrects the erodibility in an upward direction but does not remain constant. Panel (c) shows a case where the erodibility has clearly not converged. In panel (d) the estimate appears to converge between updates 10 and 15 but undergoes large variation after update 20. The estimation curves shown in these panels are representative of many locations in the domain. The assumption that the model is imperfect only in the erodibility is too simplistic. There are many other imperfections in the model. The estimate of erodibility inadvertently corrects for these imperfections. The imperfections in threshold velocity and near surface wind would have the highest impact on the estimate of erodibility because these control the dust flux. The friction velocity depends on the 10m wind. It is possible that the estimation corrects α to account for imperfection in the 10m wind. Therefore one has to exercise caution while interpreting the tuned map of erodibility."

Q. *28861,15: Can your system deal with significant underestimations of the emission flux? As α <=1 by construction, it would appear fortunate that the operational model overestimates dust emission.*

Response: The OSSE results shown in the paper correspond to (mean, spread)=(0.25,0.25). We have performed OSSE experiments with different values for the initial mean and spread. The tuned estimates are similar for mean as low as 0.1 and as high as 0.75. We see that the tuned estimates are similar irrespective of the mean and spread of the initial guess. However we have not experimented with different initial guesses for experiments with real data.

Q. *28864-28867: The point of this expose eludes me. In any case, a shorter text may be more convincing. I would suggest to show scatter plots of forecast AOT vs observation, for both estimated and operational erodability maps, possibly for individual regions. Combined with Figs 9 and 10, that would give a much better insight in the overall improvement than Fig 12.*

Response: Using these graphs we were trying to show that the tuned model results in a bias correction in the AOD. We have deleted several panels from this figure and also shortened the text. We have added the scatter plots for various sub regions as suggested. Please see figure 14 and the corresponding discussion in section 6.2.

"The regression coefficients a and b (which is the bias) are shown in the upper right corner of the panel. The blue line is the linear fit to the cyan dots which shows the tuned AOD. The tuned model reduces the bias from 0.56 to 0.31. Though the tuned model on an average overestimates the AOD by 0.31 it decreases the bias by 0.25 (compared to the operational model) which is a substantial improvement.

The panel (b) shows the scatter plot for the east Sahara region. In this region the tuned model increases the bias (0.12). The operational model has a small negative bias of -0.04. Panel (c) shows that the tuned model decreases the bias by 0.46 from 0.72 to 0.26 in the Arabian peninsula."

Q. *Section 7: I would suggest a fairly substantial rewrite of this section. As it is, it mainly discusses technical issues and ignores some of the major points: - what did the authors set out to do? - why is that new and interesting? - what methodology did they use? after this, a brief summary of results as well as possible improvements can be discussed.*

Response: We have rewritten this section. We have added the last four paragraphs. Paragraph 5 and 6 highlights the limitations of the results with real data and point to improvements that can be implemented in the future. The limitations emanate from the number of simplifying assumptions we have made (which have been pointed out by the reviewer). Paragraph 7 and 8 lays out possibilities for future improvements.

Q. *Textual comments 28846,6: "An OSSE is cast in as a perfect model experiment". Please rephrase.*

Response: We have rephrased this sentence. "The OSSE uses simulated observations drawn from the *perfect model experiment.*"

Q. *28850,17: "Out of the total spread of αprior" I guess this should be AODprior? 28864,6: MAE should be dMAE?*

Response: We have corrected these.

Q. *Figures: several figures could do with shorter captions which repeat part of the paper's text anyway.*

Response: We have shortened captions of several figures.