**Geoscientific
Model Development
Discussions**

Interactive comment on "A simulation study of the ensemble-based data assimilation of satellite-borne lidar aerosol observations" *by* T. T. Sekiyama et al.

Anonymous Referee #1

Received and published: 20 August 2012

General comments

In this interesting paper, the authors present a framework for testing an ensemble-based data assimilation (DA) system (first presented in other papers) for the global aerosol model MASINGAR by use of simulated CALIOP LIDAR observations. DA is becoming more and more important for aerosol modelling, but the sparseness of observations and the complicated relation between observations and typical prognostic variables in an aerosol model, make validation very difficult. OSSEs present a possible way forward and this paper may be the first paper to try to do so. The authors also introduce a new method, Object-Based Diagnostic Evaluation (MODE) of comparing

aerosol model results, that compares patterns rather than model values. Although first results with MODE seem a bit ambiguous (MODE is sometimes surprisingly insensitive), the authors are commended for taking this approach.The authors have limited their study to South-East Asia, although global analysis definitely seems possible.

The authors describe in detail and with clarity their data assimilation system, the simulated CALIOP observations, and the various tests they do. In comparison, MODE gets a bit less attention even though it is probably the most novel aspect of their analysis. References are provided though.

My main concerns are three-fold. Firstly, the OSSEs presented in this paper use a nearly identical twin setup and as such are only an (important) first step. The authors should be clearer about this. Secondly, the authors provide no evidence (neither mathematical proof nor proper experiments) that their DA system is suitable for emission estimation (note that this is not the only purpose of their system). I recommend they frame their statements regarding emission estimation more cautiously. Thirdly, although MODE shows clear improvement of the DA analysis over a free run, it seems rather insensitive when applied to several other experiments. This may limit the usefulness of MODE.

Specific comments

p. 1878, line 15: how can the authors be sure their emission estimate is better? Fig. 16 seems to be inconclusive. In the paper they do not give mathematical evidence of the appropriateness of their 4D-LETKF for emission estimation, nor do they conduct experiments particularly aimed at validating emissions (see also later comments). I suggest removing this statement.

p.1878, line 20: the authors suggest here and in several places of the paper that their experiments suggest that nudged meteorology is a limiting factor in their results. As I see it, no evidence is presented for this. Since many reasons can be imagined for what they call 'limited degree of freedom', more cautious phrasing is recommended.

p. 1880, line 10: Depending on the assimilation scheme, column-averaged observations can yield constraints on aerosol profiles. This is true at least for DA that takes flow information into account (4DVAR,EnKF). By the way, Schutgens et al. (2010, ACP) also assimilated AE. Also, LIDAR observations will likely miss a substantical part of the boundary layer. Please mention this.

p. 1881, line 5: I wonder whether CALIOP allows sufficient resolution of boundary layer (BL)? Since MODIS AOT will be determined mostly by BL aerosol, there is a chance the authors are comparing apples and pears. This also has consequences for their analysis later on. Please comment on this.

p. 1882, line 1: Ideally a different model should be used for the nature run than for DA. Please mention this here.

p. 1885, line 20: What is meant here by 'losing covariance among members'? Adding perturbations will itself cause the covariances to be without (physical) meaning.

p. 1886, line 1: I realise that 4D-LETKF is often called a smoother by its developers and users. But it seems to me that this is not a real smoother. It is just a DA scheme that takes observations within a time-window into consideration, as opposed to a single time. Granted, this leads to smoothing properties. But 4D-LETKF has a single analysis within that time-window, while a smoother would have several (the latest and several timesteps before that).

line 20: The observations are no longer independent, which is a major requirement of DA (or requires adjusting R, which is impossible in this case). How do you justify using observations twice?

p. 1888, line 20: I still do not understand these perturbations. After each analysis the authors perturb both α and the mixing ratio? Are there spatial or temporal correlations among these perturbations? Authors describe this in some more detail on p. 1893 but still relative size of perturbations or inflation eludes me. Why were the values

mentioned on p. 1893 chosen?

Fig 2: why are there negative correlations between emission and concentration? In a linear model this is not to be expected. The non-linearity of MASINGAR is probably due to dynamics which are constrained anyway by the nudging. Is there any non-linear component to the aerosol simulation?

p. 1891, line 5: 'require erroneous assumptions' Actually, the retrieval requires correct assumptions, not erroneous ones! Sadly, this is not always achieved. Still, the authors should rephrase this. The authors should also mention that assumptions are still required to construct the observation operator.

line 20: Another approach would be to use spheroids for dust, using e.g. scattering codes by Dubovik or Mishchenko. note also you assume that Mie calculations still give a good approximation to extinction, mention this explicitly. Do you have references for the 50sr value for the LIDAR ratio?

p. 1892, line 5: Why are oc, bc and seasalt not considered in the assimilation. Especially oc is likely to be important over land in Asia. Seasalt is mainly confined to BL so may not figure much in LIDAR observations. Still, it may have its impact.

line 10: Actually 3dVar and 4dvar do not require a linear obs operator. It depends on the solution method employed.

p. 1893, line 5: so after each analysis step, you perturb the fields? Is the perturbation different each time? I wonder if this does not destroy any covariant information the authors may have in their flow? Relative speaking, how large are these perturbations?

p. 1893, line 10 I doubt that your model is very non-linear or chaotic. After all you nudge the meteorology. Point in case is that the authors perturb the velocity field without negative repercussions.

line 25: The observation error should also be discussed in the CALIOP section. Moreover, it needs to be explained. Measurement errors of attenuated backscatter are a few

%, so why an observation error of 40 and even 100%? What is the contribution of the (spatial) representation error? Are the other error sources? How is this representation error treated in your OSSE. What are the errors in depolarization?

p. 1896, line 5: The dynamcial core, the assumptions on deposition, sulfate chemistry, oc (and bc) emissions as well as the seasalt emission parametrisation are still the same. Even though you do perturb some aspects of your model for the nature run, these are only parametric perturbations, exactly the sort that your DA system was developed to deal with. Had the authors perturbed the OC simulation in the nature run (OC is not analysed) is would be a more realistic experiment. I feel this paragraph should stress more the similarities between the models. Merely calling this a fraternal twin experiment is not enough.

p. 1900, line 25: why do the authors not use extinction or attenuated backscatter? After all you assimilate CALIOP profiles. Again, I come back to the observations missing in the boundary layer. Is AOT really a good metric to validate assimilation of LIDAR data?

p. 1903, line 25: this is a new technique and potentially very interesting to aerosol modellers. Please be more specific in the strengths and weaknesses of this approach. In particular there seem to be parameters that one has to choose a-priori. How did you choose these? Did you test the sensitivity of your results vs these parameter values?

p. 1904, line 15:I don't understand the definition of ratio 75% of intensity, especially since the authors talk of the 'lesser of the analysis intensity'? Surely thy do not use the minimum analysis intensity, so what do they use? Also, if the ratio is either defined as a/b or b/a (reciprocal), does this not cause ambiguities when comparing two different experiments versus the nature run?

p. 1905, line 5: The free model run is a single run without any perturbations? As the authors are comparing against an ensemble DA, a free run might also imply the mean of the perturbed ensemble without assimilating observations.

p. 1909, line 5: the authors analyze atmospheric mixing ratios and emissions at the same time. Can this not also be a source of error in the emission estimation? There seems to a chance of balancing errors here. This may explain why the emission maps show little improvement but the area-averaged values do show improvement. Did the authors run the model with the newly estimated emissions to see if it improves AOT or backscatter?

p. 1910, line 1: I am surprised by the small impact most of your sensitivity studies have on the MODE analysis (table 7 and 8). This seems to suggest MODE may not be an appropriate tool to compare against the nature run. Note that both Schutgens et al. (2010b, ACP) and Yumimoto et al. (2011, GRL) find clear differences using a standard analysis and real (!) observations. In their case, the discrepancy between real and simulated flow field is probably larger than in this rather identical twin OSSE.

p. 1910, line 15: Authors claim that the nudged meteorology is responsible for large ensemble sizes having worse results than n=32, but this can not be concluded without identical twin experiments to show that results do improve for increasing ensemble size (as is expected). Did you check how different your meteorology really is? After all, the same dynamical core is used. It would appear that the noisiness of the n=32 ensemble somehow balances other errors. It is important the authors point this out.

p. 1911, line 10: both Schutgens et al. (2010b, ACP) and Yumimoto et al. (2011, GRL) find optimal horizontal localization lengths of about 2 gridboxes (at T42). Granted they used a different model (MIROC-SPRINTARS) and a somewhat different DA. Still the difference in results is striking and comments from the authors are invited.

line 25: Again, the lack of any difference (for sulfate) due to changes in observational error makes one wonder if MODE is an appropriate methodology. It would be helpful if the authors indicated how the ensemble spread behaves (ie without and with assimilation). For any good DA, one expects that analysis become more like the nature run for increasing obs error. Is that indeed the case?

p. 1912, line 1: 'the error is not statistically significant because the observations of the aerosol remote sensing instruments exhibit a large discrepancy with one another'. I do not understand this sentence. The authors are now claiming that 20% is not a good estimate for the error? Than why use it in the first place? 40% seems already quite a large error for attenuated backscatter observations, where the representation error probably dominates the 'retrieval' error.

p. 1912, line 20: it is surprising that an experiment with only 1/6 the amount of data yields similar results to the stdexp. Especially as Fig. 18 seems to argue the opposite. Are clouds and aerosol plumes in different locations? What is happening?

p. 1913, line 5: Why do the authors employ a threshhold in the first place? Since the obs operator includes molecular scattering, this seems not necessary?

I hope that in future work, the authors will consider tests to assess the impact of: - different OC emissions -sensor loss (what if no depolarization or 1064nm intensity is available).

p. 1914, line 5: 'Although some technical difficulties and limitations of OSSEs were revealed'. OSSEs seem very important tools to assess DA and the authors are commended for using them. It seems the summary would be a good place to summarize strengths and weaknesses of OSSEs.

p. 1915, line 10: 'In spite of the many controversies regarding OSSEs'. Please address these controversies.

Technical corrections p. 1878, line 1: 'caliop satellite was emulated'. Please rephrase using 'simulation'. Emulation is used in various contexts but seldom in that of sensor simulation.

p. 1889, line 20: ' in which dynamic tendencies are added with an 18-hour relaxation time constant' should be replaced by 'in which an extra tendency is added that relaxes the calculated flowfield to a reanalysis with an relaxation time constant of 18 hrs' (if my

interpretation is correct, otherwise, rephrase more clearly).

p. 1899, line 25: I take it that 'these observations' refers to real CALIOP observations and not to the simulated observations. Please prevent this ambiguity.

Interactive comment on Geosci. Model Dev. Discuss., 5, 1877, 2012.