**Atmospheric
Chemistry
and Physics
Discussions**

Interactive comment on "Estimation of aerosol particle distributions with Kalman Filtering – Part 1: Theory, general aspects and statistical validity" by T. Viskari et al.

Anonymous Referee #1

Received and published: 5 September 2012

Estimation of aerosol particle distributions with Kalman filtering – Part 1: Theory, general aspects and statistical validity

T. Viskari, E.Asmi, P. Kolmonen, H. Vuollekoski, T. Petäjä, H, Järvinen

General comments

In this paper, the authors describe a new application for the Kalman filter: estimation of particle size distributions. Whereas typically in atmospheric science data-assimilation techniques are used in global or regional models to spread information from local observations throughout the grid, in this paper it is used to interpret the raw measurements by Differential Mobility Particle Sizers. These measurements are related to the

actual particle size distribution (PSD) through an instrument response function. The Kalman filter offers a new approach at obtaining the PSD over standard mathematical inversion. The authors describe the basic theory of the extended Kalman filter, the observations, the box model that will forecast particle size distributions and several (numerical) experiments that were conducted. On the whole, the paper is clearly written although several details need further clarification.

Validation of the methodology is done by comparison of the results with the Kalman filter to those of a straightforward inversion scheme. Although this comparison shows that remaining biases due to the Kalman filter are smaller than those due to the inversion, I am a bit dubious whether that warrants the conclusion that the Kalman filter is more accurate. I'd like the authors to address the following points:

1. The validation compares raw measurements by the DMPS to simulated raw measurements derived from the particle size distributions. However, those particle size distributions were retrieved (by either Kalman filtering or inversion) using those same raw measurements. No independent observations are used for validation. It is not obvious that the algorithms have found the appropriate solution (size distribution) and not one that is unphysical but nevertheless minimizes the differences between real and simulated raw measurements. The same issue must have confronted the developers of the inversion algorithm, so I suppose quite a bit of research has already been done. In any case, the authors should either supply a validation based on independent observations or present strong arguments why their current validation is sufficient.

2. The tests for the two algorithms are not identical as observations are handled differently (in particular in the overlap region of the two DMPS). This begs the question whether any differences in results are due to the algorithms or the observations (I realize that only EKF can handle observations from both DMPS in the overlap). Have the authors conducted experiments where the EKF uses the same observations as the inversion?

3. Related to both previous points: why are the differences between raw measurement and simulated measurement not zero for the inversion? I guess this is because the raw measurements in the overlap are averaged before inversion, but the validation is done with the original raw measurements?

Finally, it remains rather vague why the authors choose to use Kalman filtering instead of e.g. regularization techniques which seem more suited to their problem. Especially as the authors themselves point out that the time-information from the forward model calculations is detrimental to the solution when sudden changes occur (Sect 5.2.1). The main advantage that the Kalman filter has over the inversion is that it can handle observations from both instruments in the overlap, while the inversion requires averaging of those observations. The authors may want to stress this point.

I think the Kalman filter (or other data assimilation tools) may become very useful in interpreting DMPS raw measurements in a configuration where model parameters are estimated. I'd like to hear the authors' opinion on the feasibility of such an approach.

Specific comments

p. 18857, time evolution updating and observation updating might be replaced by the more common terms forecast and analysis. Observation updating is a confusing term anyway as no observations are updated.

p. 18857, eq 1, I suggest you use M(x), instead of Mx to stress non-linearity. Also, why do you not include an error term here as you do in eq 2, that makes the text more logical.

p. 18857, line 15: Q represents more unaccounted physics and chemistry and transport etc. it really represents structural model errors. Calling it system noise is not recommended because 1) your model may be biased (hence necessity of error term in eq 1); it gives the suggestion of numerical noise or some other implementation related issue. Also mention that Q can not be specified. Finally, Bk is an approximation

C6624

because the linear-tangent model is used.

p. 18857 line 20: "Here H is a possibly non-linear observation operator, which produces the observation counterpart corresponding to the prior state". replace "observation counterpart" by "simulated observation"

p. 18857 eq 3: H(x) instead of Hx

section 2. The way I understand EKF, it is just KF but you use a non-linear model. Nothing really changes from KF, except the full model is used for forecast and the tangent linear for analysis. Please point this out explicitly. This also has consequences because the Kalman filter is only exact and optimal for a linear model. Please point this out as well.

section 3.1 it would be good if the authors detail measurement errors in yr (!, including temporal representation errors, see last paragraph of 3.1), dominant error sources in inversion, final error estimates for x and validation by independent observations. I'm guessing this has already been done by other groups so a few lines and some references will suffice.

p. 18859 line 25: is this transfer function the same as R? If not, what is their relation? "Transfer functions for both DMPSs are integrated separately for this diameter grid": why? Do you mean that within each size range, the transfer functions are integrated?

section 3.2 How is the model initialized? What additional information is needed to run the model? For which time and spatial scales is it considered valid? Has the model been evaluated against observations (> results?)?

For the EKF, did the authors develop a tangent linear version of the model, or are model results linearized 'on the fly'?

section 3.3

p. 18861 line 20: "The observation operator was tested and validated by comparing

C6625

raw observations to the values computed by H from a size distribution obtained with a mathematical inversion." Is this not a circular argument? The raw observations have to be inverted (R^{-1}) and interpolated to the model grid (P^{-1}) before you can use eq 7.

sect 4.1

p. 18862 line 5: The authors state that the linear model is valid on time-scales of 30 min only, so carrying B forward repeatedly with the linear model should cause large errors and will be detrimental to the assimilation. Do they somehow condition the posterior B on the posterior x at each timestep?

p. 18862 line 20: "As a consequence, the state error covariance may become gradually smaller and smaller, as was the case here". I know this effect exist in purely linear KF as well. In ensemble KF, inflation techniques are used to combat it, much like this paper does. Please mention this. To what extent is this problem amplified by the linear model for B?

p. 18863 line 20: "but representativeness and observation operator errors need statistical material, i.e. innovation sequences produced with EKF". I do not see how one can reliably determine those errors from data assimilation experiments. After all, various error sources all contribute to the statistics of the innovation. Either argue your case or remove this line.

p. 18863 line 25: The authors also assume that observation errors are uncorrelated across sizes. This seems unrealistic due to the size categorization use in DMPS. Is this assumption purely practical or is it also approximately valid (maybe other error sources are more important)? Please discuss.

p. 18864 line 5: Do you actually mean that the errors in the smallest and largest size bins are correlated? Or do you mean that the errors in the smallest size bins are correlated and that the errors in the largest size bins are correlated? I guess you mean the latter. See also my comment before on observational error correlation.

C6626

p. 18865 line 10: Again, how do you initialize the model? Why are model errors chosen to be 30%?

p. 18865 line 15: EKF results are smoother because solutions are constrained by the time evolution of the model. Had the inversion used any kind of regularization, I guess it would also look more smooth?

p. 18865 line 20: "The differences in the total number concentrations are partially due to the diameters for xEKF and xINV not being the same, which makes it difficult to limit xEKF to the same diameter range than xINV" Sorry, but I don't understand what is meant here.

p. 18866 line 5: I think this even suggests that there are (unsurprisingly) biases in the observations. If the observations for DMSP I and II were unbiased with properly assigned errors, results in the overlap should be better than outside (because one uses more observations), shouldn't it (unless error magnitudes are much larger in the overlap)?

p. 18866 line 10: I am surprised that the inversion allows larger biases. You "validate" your results against the observations that you originally inverted. It seems the inversion allows less freedom in this (hence has larger errors) than EKF. But it is actually EKF that adds a lot of model information to the estimate of x. Please discuss. One thing that seems different is how you use the data in the overlap. For EKF you use all measurements by both instruments, but for the inversion, you use averaged data in the overlap, I believe.

Results in section 5 are not a proper validation as you do not use independent observations. Since the EKF introduces model information, it may actually yield worse results than the inversion when tested against independent observations.

p. 18867 line 10: so to what extent are the observations that you use for EKF and the inversion different? This should be specified in detail because it seems to have a

C6627

significant impact. Why did the authors choose to use different observations? As a result it becomes impossible to separate data sampling effects from algorithm effects.

For a proper assimilation system, remaining biases should be smaller than the remaining standard deviations. It would appear that on this score, EKF fares worse than the inversion as it has smaller biases and smaller stddev. What is the authors' take on this? Also, are stddev from raw measurements vs EKF similar to posterior stddevs? This is another important test for EKF. See also your point vi in Section 6.

Section 5.2.1: The (incorrect) temporal information from the model when air masses or so change make me question the appropriateness of EKF for this problem. Wouldn't an inversion with regularization be a better choice of algorithm? E.g. Philips-Thikonov regularization using assumed smoothness of the size distribution.

Section 6: I do not disagree with the points the authors make. They are, in a sense, general comments valid for a properly functioning KF. I feel the authors have not proven this is a properly working EKF as there are no independent observations to verify results. This impacts their statements i) and iii). In this particular paper, it seems the main advantage EKF holds over inversion is ii) better handling of multi-instrument retrievals. On the other hand v) and vi) seem rather unconvincing: due to the nature of a box-model it will be impossible to properly account for changing air masses etc.

Technical corrections

p 18861 line 15: "The interpolation matrix P is resolved" The authors mean: "The actual interpolation was performed". I have no idea what resolving a matrix implies, especially as there is no explicit equation.

p 18862 line 15: "source term" Please use "error term" as source term often refers to emissions.

C6628