Review of EnKF and 4D-Var Data Assimilation with a Chemistry Transport Model

by S. Skachko et al.

This paper presents a comparison of the EnKF and 4D-Var data assimilation methods applied to a chemistry transport model. These two methods have been used extensively in applications to numerical weather prediction (NWP), a field with which I am more familiar. It is interesting to see such an intercomparison for an atmospheric chemistry model for simultaneously assimilating observations of the concentrations of several chemical species.

General Comments:

#1 Set up of assimilation windows

One major concern I have is related to the basic setup of the experiments. Whereas intercomparisons in the context of NWP have generally used the same data assimilation window length for the two approaches (typically 6h), this study uses a different window length: 4D-Var is applied in its strong constraint formulation with a 24h window, whereas the EnKF is used to sequentially assimilate observations every 30 minutes (the time step length of the forecast model). In addition, model error perturbations are added every 30 minutes in the EnKF experiment, in contrast with the assumption of no model error over the 24h assimilation window in the 4D-Var setup. This difference in the configurations of the two methods seems to affect the conclusions of the study, as acknowledged by the authors near the end of the paper.

Alternative configurations could have possibly been chosen that would have reduced these differences. For example, the 4D-Var could have been implemented in its weak-constraint formulation, employing a model error term every 30 minutes based on the same covariance matrix as used for model error in the EnKF experiment. Alternatively, a 4D-EnKF approach could have been used with the EnKF in which ensembles of model solutions over a longer assimilation window are used to construct a 4D background error covariance matrix (as is done in most current implementations of both EnVar and the EnKF for NWP). In this case a window shorter than 24h (say 6h) would likely be preferable (and used for both EnKF and 4D-Var) to avoid problems with horizontal localization when a chemical species is advected within the time window by a distance comparable to the localization length scale. The authors should mention these two alternative approaches and discuss how they could have reduced the differences seen in some of their results.

As a consequence, the difference in window length also likely has a direct impact on the quality of the forecasts, since the 4D-Var forecasts are started from a state that may not be in as good agreement with the most recent observations as in the case of the EnKF. It would be useful to show the fit of the state used to initialize the 24h forecasts with respect to the observations

valid at the same time for the EnKF and 4D-Var experiments, or at least mention in the text how this fit differs between the two. I wonder if this explains some of the differences in results for HCl.

#2 Inter-species background error covariances

The discussion regarding the estimation and treatment of inter-species background error covariances is somewhat unclear. The authors state that because it seems suitable to use the same model error covariances for each species it follows that the main source of error is from the transport. However, then the authors claim that the cross-species background error covariances are weak and not sufficiently well estimated with the EnKF ensemble and therefore should be set to zero when assimilating the observations. It seems to me that if the errors of the different species are dominantly affected by the same common source (i.e. the transport), then their errors should be correlated. Or maybe this is not the case due to the dependence of the concentration errors from transport on the concentration spatial gradients, which may be very different for each chemical species. Please provide some comment on this.

Also, I think the authors need to stress how the relatively small size of the ensemble (20 members versus the usual size for NWP of O(100)) could affect their conclusions regarding the utility of the cross-species covariances. Presumably with a much larger ensemble they would be better estimated and therefore could be more useful.

<u>Specific comments:</u>

In several places the term “observation errors” is used where I believe “observation error variances” or covariance or statistics is really what is meant. Similarly, on page 3, line 2 I am guessing that “the covariance inflation” refers to the inflation of the “background error covariance”, please be more precise.

page 1, line 10 “…where we keep both estimates:” This is not clear. Is it meant that you keep both estimates fixed in time? Please clarify.

page 1, line 12 “…a single model error based…” It is difficult to decipher what is meant. I suppose this refers to using the same specified model error covariance matrix for each chemical species. Please make this more clear.

page 1, line 15 “…sampling noise errors…” and “These errors need to be filtered out…”. This is awkward wording. Better to talk about “sampling error due to the use of a small ensemble leading to spurious covariances” and “setting these spurious covariance to zero”.

page 1, line 20 “…not too small chemical life times…” This does not sound sufficiently quantitative. Could you instead say something like “chemical lifetimes longer than…” where you compare the lifetimes to some relevant time scale, e.g. the model time-step or assimilation window length.

page 2, line 6-7: It would make sense to also refer here to one of the first EnKF/4D-Var comparisons performed with real NWP systems done at the same center as one of the co-authors: Buehner et al. 2010 (2-part paper in MWR).

page 2, line 29-34: It is not clear why the issue of estimating error statistics for many chemical species is specific to the EnKF, as the text now seems to imply. Clearly this is equally important and challenging for the application of 4D-Var? Please clarify.

page 3, line 4 “background error” should be “background error covariance”.

page 4, line 13, 15: Define the acronyms PSC and MACC.

page 5, line 15: It would help to state here that the same background error correlations are used for each chemical species and that the between-species covariances are assumed to be zero.

page 6, line 8: gamma is really set to 5? According to equation 5 this means an observation is rejected if its innovation is larger than 2.2 stddev (i.e. sqrt(5) = 2.2). Maybe equation 5 should have gamma^2 instead of just gamma?

page 6, line 20: the parameter N (ensemble size) needs to be defined here, since this is where it first appears.

page 6, line 23: “…are normally distributed random numbers…” should probably be “…are vectors of independent normally distributed random numbers…” otherwise equations 6 and 7 do not make sense.

page 6, lines 26-29: The two sentences about the application of the Desroziers method seems out of place here, since the method has not yet been introduced. Considering moving this to section 2.5.

page 7, lines 8-9: “To this end, the EnKF algorithm accounts now for a new effective procedure to find a current local sub domain in the model space.” It is not at all clear what this sentence means. It must be better explained. Is the algorithm similar to that described by Houtekamer et al. (2014, MWR - “Parallel implementation of an EnKF”)?

page 8, line 3: I believe “variance” should be “standard deviation”. Please also check the entire paper to ensure standard deviation and variance are used correctly. Also, as already

mentioned, ensure the word “error” is not used in places where “error standard deviation” or “error statistics” or “error covariance” is actually what is meant. This is a common mistake that can be very confusing for some readers.

page 8, lines 32: “…and thus goes along in arguing that they represent some true error statistics.” This does not sound very solid as a logical argument. Consider improving. Similarly for the following 2 sentences. It is probably not necessary to make this assertion at this point in the text.

page 9, line 4: “…10 iterations…” Is this enough to obtain a substantial reduction in the amplitude of the cost function gradient (i.e. at least a factor of 10)?

page 9, section 3: Please provide some additional general information about the observations assimilated: how much of the globe is observed during 24h? what is the horizontal and vertical spatial resolution?

page 10, section 4.1 and verification in general: What observations are used for verification? For example, are all observations with a valid time within 1h or 3h or ?h of the valid time of the forecast used? Since you are always verifying 24h forecasts valid at 0000UTC each day (if I understood correctly), does this tend to focus the verification in only certain geographical regions due to the orbit of the satellite?

page 11, line 16-17: related to the first general comment above, I think the difference seen here between the 4D-Var and EnKF results may be due to the difference in assimilation window lengths, please add a comment here.

page 12, line 32: “This is due to the automatic rejection by the 4D-Var of most observations…” What does this mean? Is it because the 4D-Var cannot make the forecast model solution fit the observations over the 24h assimilation window (due to model error) or is it referring to some quality control procedure (which I though was deactivated for this chemical species at the pressures considered here)?

page 13, line 23: “…model error covariance…” should be “…background error covariance in observation space…”

page 13, line 26: “…localization of the error variance…” should be “…localization of the error covariance…”

page 15, last paragraph: Another limitation is that much fewer ensemble members were used as compared with typical NWP applications. This should be mentioned.