Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-95-AC3, 2016

Geoscientific Model Development Discussions

Interactive comment on "EnKF and 4D-Var Data Assimilation with a Chemistry Transport Model" *by* S. Skachko et al.

S. Skachko et al.

sergey.skachko@aeronomie.be

Received and published: 3 June 2016

Response to the Follow-up review of the Anonymous Referee 2

June 3, 2016

We thank the referee for the detailed revision of the paper. The author's responses are marked in blue.

General comments:

I am confused by the authors' response to my main concern, related to the difference in the window length used for the 4D-Var and EnKF experiments. In response to my first general comment, the authors' response is: "So we disagree that the difference in window length has such an impact in the context of chemical transport." Then, when I later bring up the same point again in relation to the discussion of the results, the authors' response is: "In the context of chemistry, the difference in data assimilation window lengths really has implications, as pointed out by the referee."

We should be more clear on this. The first mentioned sentence means that in the context of chemical tracer transport only (without chemistry system), there is no difference in using an EnKF with 30 min ensemble model forecasts and a model error term or a 4D-Var with 24 h assimilation window without model error term. This was shown in our previous article (Skachko et al 2014). The purpose of the present work is to reveal

the role of the chemistry system (including interactions between chemical species) in the context of our two data assimilations that are configured as they are normally used in chemical data assimilation applications: one model time step ensemble model forecasts within EnKF, and 12 - 24 h of 4D-Var assimilation window.

Also, I believe the authors' misinterpreted part of my first general comment. I made no suggestion that a hybrid 4D-EnVar experiment be performed, or even mentioned. What I did suggest was that a 4D-EnKF approach (with model error perturbations only applied at the beginning of each window to be equivalent with strong-constraint 4D-Var) should be considered and mentioned, since this would allow a longer window to be used for the EnKF. In this case, the analysis would be forced to simultaneously fit all of the observations distributed over a longer window, while still satisfying the model equations, as in 4D-Var.

I appreciate that the authors have tested two data assimilation methods in configurations as they are usually used for chemical applications. This point should be emphasized in the paper to justify the choice. However, it would be helpful to inform the reader that other configurations are possible that would reduce the differences between the two approaches (i.e. including model error in weakconstraint 4D-Var and using 4D covariances with a longer window in the EnKF). Otherwise, readers will conclude that one approach (i.e. EnKF or 4D-Var) is fundamentally better or worse than the other in some respects, whereas it is more likely the choice of how each approach was implemented that is more important.

The fourth paragraph of the introduction is modified as follows: "But how do the EnKF and the 4DVar methods compare when photochemical reactions are taken into account? Do the results depend on the assimilated chemical species? Using actual satellite datasets and operational configurations, what are their respective performances in terms of precision, accuracy and computational efficiency? What is the role of the practical implementation of each method, when the full description of the stratospheric chemistry is taken into account in the CTM. These are the main questions addressed

in this paper."

The conclusions start with: "We have conducted a comparison of an EnKF and 4DVar data assimilation system using a comprehensive stratospheric chemical transport model. We considered 4D-Var and EnKF configurations that are normally used for chemical data assimilation applications. Both data assimilation systems have online estimation of error variances based on the Desroziers' method and share the same correlation model for all prescribed error correlations (i.e. the background error covariance for 4D-Var, initial error and model error for EnKF) so that each data assimilation system is nearly optimal and can also be compared to each other. A previous comparison study by (Skachko et al. 2014) showed that for chemical tracer transport only both assimilation system provide results of essentially similar quality despite the difference in practical implementation of each method: the 4D-Var was applied in its strong constraint formulation with a 24 h assimilation window with the assumption of no model error over this period, whereas the EnKF was used to sequentially assimilate observations every 30 minutes with model error perturbations added every 30 minutes."

Then the following text is added at the end of our conclusions: "Another possibilities may be considered to properly compare two essentially different data assimilation systems. First, a 4D-EnKF approach, where model error perturbations only applied at the beginning of each 4D-Var assimilation window to be equivalent with a strong-constraint 4D-Var, may be considered. This would allow a longer assimilation window to be used for the EnKF. In this case, the analysis would be forced to simultaneously fit all of the observations distributed over a longer window, while still satisfying the model equations, as in 4D-Var. Second, the use of a weak-constraint 4D-Var including model error would also reduce the differences between two considered approaches. "

Specific comments:

In response to the third specific comment, your revised sentence seems imprecise: "For comparison purposes, we apply the same estimate procedure in the 4D-Var data

assimilation, where both, the background and observation error covariance matrices are estimated using the Desroziers' method." I presume it is only the scale factors for both covariance matrices that are estimated and not the full matrices? Please improve the wording.

The sentence is now written as: "For comparison purposes, we apply the same estimate procedure in the 4D-Var data assimilation, where both scale factors of the background and observation error covariance matrices are estimated using the Desroziers' method."

In response to the fifth specific comment, your revised sentence does not clear up my concern: "The second issue in EnKF with comprehensive atmospheric chemistry models is the spurious error, that occurs when species are weakly chemically related at the same location." The term "spurious error" is very ambiguous... how can error be spurious? I believe this is again where "error" is used in place of "error covariance". Only the "estimated error covariance" is spurious. [The word "error" on its own really should be reserved for the difference between an estimate and the truth and I don't think this is what is meant in this case. I realize that some published papers have used "error" to mean "error standard deviation" or "error covariance", but I believe this has needlessly caused confusion for some people in the DA community.]

The sentence is rewritten as follows: "The second issue in EnKF with comprehensive atmospheric chemistry models is the noise in the cross-covariance between species, that occurs when species are weakly chemically related at the same location."

References

Skachko, S., Errera, Q., Ménard, R., Christophe, Y., and Chabrillat, S.: Comparison of the ensemble Kalman filter and 4D-Var assimilation methods using a stratospheric tracer transport model, Geosci. Model Dev., 7, 1451-1465, doi:10.5194/gmd-7-1451-2014, 2014.