Geosci. Model Dev. Discuss., 6, C1191–C1194, 2013
www.geosci-model-dev-discuss.net/6/C1191/2013/

Geoscientific Model Development Discussions

Interactive comment on "Regional scale ozone data assimilation using an Ensemble Kalman Filter and the CHIMERE Chemical-Transport Model" *by* B. Gaubert et al.

Anonymous Referee #3

Received and published: 29 July 2013

1 General comments

The manuscript describes results form experiments with an Ensemble Kalmanf Filter to assimilate surface ozone concentrations in the regional CTM Chimere. Although these type of experiments are hard to reporoduce without having the complete modelling, assimilation system, and data sets available, the authors provide enough information on the methode, the choosen settings, and the evaluation diagnostics to have a good idea on what has been done and how somebody with similar tools available coold obtain similar results. Whenever possible, references are made to other papers to show what kind of settings are used in other studies for to some extend arbitrary numbers of

length scales and error characteristics, which gives trust that the assimilation system is well-defined.

A particular strong point is the emphasis on the diurnal cycle of concentrations and diagnostics. The majority of the figures show diurnal cycles, and within the system, hourly profiles for assimilation parameters are set based on diagnostic profiles obtained for the day before. This focus might get some more attention in the abstract and conclusions, since to my opinion this is one of the novelties of this application.

The experiments described have been performed for a time period of 10 days only. Although this does not hamper the evaluation and the conclusions, some outlook on the performance over a longer time period would be useful. Is the system in this case in particular tuned for a high-ozone episode for example ? Is the performance in area with low ozone (west of the domain) expected to be typical for 'normal' conditions ?

2 Specific comments

- Eq. 5 and 6. It is not always clear over which 'p' observations the statistics are computed. From the text it seems that within the asismilatin these numbers are computed over all assimilated observations available at a single hour, but for some evaluations of the overall performance it seems also to include the 10 days of the experiment. Could this
- p 3047, line 14. Only ozone fields are included in the ensemble. With 24 ensemble members, the cost of propagation of the ensemble is probably less than the cost of propagation of the full model, is this correct? If this is correct, it makes the system a very cheap assimilation tool. Please add some lines on the computational costs of the system.
- p. 3048, line 16: *"., we first use for all types of stations an observation error*

standard devaiton of 5 ppb, ...". I guess the "second" choice is the "R_EXPLICIT" experiment in section 5.4, but this is not imediately clear. A table showing the essential differences between the various experiments would be useful.

- p 3049, lines 16-18. I guess that what is refered to as the "correction" means the innovation from forecast to analysis field. But the spatial extend of the innovation is limitted by the local analysis range of 250 km. Thus, the observed change over the North Sea is the result of transport (this is also what is mentioned in the conclusion at p. 3057 lines 25-27). The same result could be obtained with an optimal interpolation method. I would say that extension of innovation with the flow is only possible if the ensemble size is large enough to avoid spurious correlations and no form of localization is used. Please clarify this statement.
- p 3051, line 5. The temporal profile of the noise in the reference run just a switch between 10% and 20% for day and night. How would the results for the "MOD_DESR" look like ?
- p 3052, line 1. It seems obvious that in a region with only a single observation station the spatial impact of this site is large: there are no other observations available that could counter act. I think the main issue here is the presence of a model bias that is persistent over a large area: if this is present, then it is indeed sufficient to have a single station, but only if the spatial scales in the BECM are large too.
- p 3056, eq 8 . The formulation of a spatiall dependend observation error in Eq. 8 feels a bit as violating the idea of a Kalman filter. The spatial relations between grid cells are supposed to be described in the P matrix in eq. 3, while the relation between observations is in the R matrix (which are usually set to zero). The impact of an observation site on a grid cell further away is small if the gain matrix is small, and this is most controlled by the P matrix; eventually localization is applied to explicitly limit this spatial correlation at large distance. The observation

error matrix R is not changed by a localization. But in the choosen formulation, R has become a mixture of observation properties and spatial (physical) properties. It almost feels as if the authors have constructed a fixed gain, that is very similar to what you get from an optimal interpolation. What benefit from the ensemble is left here ? Please clarify.

3 Technical corrections

- p. 3035, line 24: are the '`\dots`' intended ?

Interactive comment on Geosci. Model Dev. Discuss., 6, 3033, 2013.