**Atmospheric
Chemistry
and Physics
Discussions**

Interactive comment on "Assimilation of IASI partial tropospheric columns with an Ensemble Kalman Filter over Europe" *by* A. Coman et al.

A. Coman et al.

adriana.coman@lisa.u-pec.fr

Received and published: 2 February 2012

Authors response to Remus Hanea

The authors thank the referee for all his remarks which allowed increasing the paper's clarity and interest.

Comments:

1 Introduction

1. Page 3, paragraph 35. The authors are describing the principal objective of the data assimilation in case of weather prediction as the improvement of the initial condition estimation. I think it is good to specify in the same place the different target for data

assimilation in atmospheric chemistry: state and parameter estimation. There is a big difference in the performance of certain methods/algortihms for data assimilation when applied to state estimation or to parameter estimation (especially for the EnKF). There is a general believe in the data assimilation community that for the state estimation the sequential methods (ensemble methods) are giving great results and for the parameter estimation the variational methods are more suitable. BUT, that does not mean that the sequential methods can not be applied successfully to parameter estimation problems. I think it is worthy to be mentioned and noted that the state and parameter estimation are two different problems with different issues to be taken care.

A small paragraph was included in order to mention the other objective of the data assimilation: parameter estimation. Even if this problem was not addressed in our study, we agree that it is worth mentioning its existence (see page 3).

"There are two important objectives for when we apply data assimilation. First, described above, is the state estimation problem which consists in finding the best estimate of the model state which best fits the model equations and the observed data. The second one is the parameter estimation. In this case, we want to improve estimates of a set of poorly known model parameters, the errors in the model being usually associated with uncertainties in the selected parameters. It is worth mentioning that it exists also a combined state and parameter estimation, where the two problems are addressed simultaneously. This issue can be solved very efficiently using ensemble or variational methods. In this study only the state estimation problem using ensemble methods is discussed."

2. Page 3, paragraph 55. ". . . together with a bias/rmse reduction . . . ". The abbreviation for rmse was not defined prior in the text.

Done

3. Page 4, paragraph 95. At the end of the Introduction the authors are presenting the whole summary of their papers and the structure of the paper. I do believe that

the paper shows clearly the added value of the IASI data in the assimilation process and the impact on the quality of the estimation. Therefore, the authors should be more bold in presenting their apport to the research in hand and replace verbs like ". . . the aim of the present study was to examine . . . " with ". . .was to show . . . " and drop the "eventual gain" word out the text, because the paper shows clearly the gain of assimilating the IASI data. It might send the wrong signal from the beginning of the paper, meaning that the results of the paper can be considered as a trial and not as a certainty.

The text was modified as suggested by the referee.

2 The Ensemble Kalman filter method

The EnKF is explained in Chapter 4, section4.1 presenting the main equations and introducing the issues that usually show when we work with a complex model and use real data in our assimilation processes. The use of an ensemble with small number of members leads to spurious correlations(nonphysical). Also , the use of a small ensemble size makes the assimilation of large amount of independent data very difficult. This is the case when using satellite data. One well known solution for these kind of problems is to use localization. But, even then there are issues in choosing the right type of localization (distance based, adaptive localization, etc.).

1. Page 11, paragraph 280. " Moreover, an ensemble with a limited number of members becomes unable to estimate the background error . . . . Thus "covariance localization" has become a very widely used technique . . . of the background covariance matrix." The terms background errors and background covariance matrix were not defined prior. The authors discussed about forecast and analyzed covariance matrix when explaining the EnKF. It is a bit confusing taking because the use of the term "background" for errors and covariances comes from the variational approaches and the terms forecast and analyzed are used usually in case of sequential methods.

We agree that the term "background" accompanying the error or the covariance ma-

trix is a little bit confusing, being especially used in the variational context. Thus, we changed it in "forecast" everywhere in the manuscript.

2. Page 11, paragraph 280 "Thus "covariance localization" has become . . . " and Page 12, paragraph 320 " Another option made here was to apply a local analysis (as mentioned above) in order to avoid spurious correlations ". The terms covariance localization and local analysis are two different ways to tackle localization (see Sakov, P, Bertino, L . 2011 "Relation between two common localization methods for the EnKF"). It is not obvious from the text which of the two methods were applied. Taking into account that the local patch is an important parameter for the assimilation in the presented paper, I think a more detailed presentation of the localization is needed to cover the basic issues, to present the two possible ways of handling localization and to explain the choices made in this paper.

Changes were made in the text following suggestions from the two referees: a short discussion about the two possible options for localization was presented and some details about the procedure selected in the present study, the local analysis. All this was placed at the end of the set-up assimilation section (pages 13-14, see below the cited text).

3 Set-up of the assimilation experiments

1. Page 11, paragraph 290. "We allow the model to contain unknown errors and use the information both from the data and the model to improve the actual model state, which contains all the concentrations for all the species in each cell grid" So there is model error present in the assimilation process. It is clear that is not coming from the unknown parameters. But is not clear what really is the cause for the model error and how that is taking into account in the set-up of the assimilation. It is due to un-modeled physical phenomenons? or due to errors in the model when describing a certain chemical reaction? How the model error is taking into account? It is an additive white/colored noise applied to the forecast equation? I think that a more detailed

explanation for the set-up of the assimilation process is needed. EnKF is a extension of the classical Kalman filter approach for linear models. Therefore, the basic state space representation of the system and the observations plays an important role and can give a better understanding of all the elements involved. $x_{k+1} = F(x_k) + w_k$ (1) $y_k = H(x_k) + v_k$ The definitions for x as the the model state, F as the CHIMERE atmospheric chemistry model, w_k as the model error, H as the observation operator and v_k as the measurement error should be given and explained in more details. Is the observation operator from eq. 5 page 12 a non liner function? If yes, how do we deal with that?

The text cited in the above comment was already modified for the published version of the manuscript in ACPD (the referee's comments were made on the first version of the manuscript, before publication). A detailed paragraph concerning the state vector and the model error was added in the published version. More precisely, the state vector is composed of the concentrations of all the species present in the chemical model in all the grid cells and the model error is an additive white noise ("pseudo-random perturbation"), added to the ozone field, and taking into account the sum of the errors committed in the parameterization process and chemical reactions. Finally, the projection of the state vector by the observation operator described in the setup assimilation section was obtained only by linear transformations of the state vector.

2. As we have already discussed the reason for spurious correlations is a small number of members and that can be solved with localization. What if we increase the dimension of the ensemble to 200? The improvements in the RMSE for the estimated concentrations will improve enough and the spurious correlation will fade away? Is this practical? Even then, the huge amount of data to be assimilated will lead to the use of localization due to the difficulties of the EnKF to handle large sets of data. I think this discussion should be made more clear in the text.

As the referee already highlighted in his comment the principal problem is the presence of spurious correlations. This is the principal reason for selecting a localisation method. In our opinion, increasing the ensemble size will not produce more improvements in the

RMSE due to the saturation of the errors already mentioned in the new version of the manuscript (the sensitivity tests performed show already that the system, in actual configuration, reached a certain RMSE threshold). So, in our opinion, augmenting the size to 200 will not improve the system performance. As for the localisation the entire paragraph was modified (pages 13-14 cited below). We hope that the discussion is clearer now.

"An ensemble with a limited number of members cannot estimate accurately the forecast error across the entire state space due to spurious error correlations; therefore it is better to restrict the new information provided by the measurement to a local neighbourhood. In a local region, the ensemble size may be sufficient to represent a large portion of the state uncertainty (Szunyogh et al., 2005). "Localisation" has become a very widely used technique to filter out the spurious long-range correlations, and increase the rank of the forecast covariance matrix. This method requires existence of some "physically sensible norm" (Sakov and Bertino, 2010) to characterise the distance between model elements. A first method of localisation was described in Houtekamer and Mitchell (1998). They limit the influence of an observation on the analysis to a surrounding local region using a cutoff radius beyond which covariances between variables are assumed to be zero, also called "covariance localisation". Another localisation method, scheme-independent, is "local analysis" or localisation in model grid space which uses "local approximation of the state error covariance for each updated state vector element by building a virtual local spatial window around this element" (Sakov and Bertino, 2010). We apply here a "local analysis" in order to avoid spurious correlations in the forecast ensemble, which are introduced by the perturbation method for finite ensemble sizes, and which do not have any geophysical reality. The basic idea of this method is to perform the analysis in a given grid point using the observations within a local region centred at that point and this analysis is performed grid point by grid point. The radius of this region was fixed at 200 km, corresponding to the decorrelation length in the horizontal perturbations applied (following Boynard et al., 2010). The maximum number of observations to be assimilated was limited to 30 pixels. This

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

parameter was subject to sensitivity tests (see later). No vertical localisation was applied. In certain cases, this method can lead to discontinuities in the analysis when an observation is taken into account into a local window and not in the next one, when we move from updating one state vector element to another, but this problem is beyond the scope of this study. Note however that this unsuitable occurrence has been addressed in Hunt et al. (2007)." (also for the anonymous referee #2)

4 Results The evaluation of the whole data assimilation setup and the improvements of different species estimation is well written and explained and the added value of this exercise is clear. I have only one question regarding the possibility of forecasting with the model after the model state was optimally updated by the EnKF. There were any tests or experiments made where we move towards an ensemble prediction system and hope that the results obtained with the improved/optimal model (after data assimilation) can and will predict certain dangerous peaks of ozone concentration or strong smog episodes? Consequently, what will be the quality of these predictions related with the quality of our model?

The aim of this study was to show the improvements obtained in estimation of the ozone field when using IASI data. No experiment was performed in the prediction of ozone peaks using the model state updated by the EnSRF. This will constitute certainly the subject of a future study. The only test made was a prediction after 24 hours from the previous assimilation because the satellite data were available/assimilated one time per day at 9 a.m. The self consistency test: OmF versus OmA show a slight improvement (for the Eastern part of the domain) from one day to another. In the published version (ACPD) in conclusion section P 71 L 8-10 we stated: "Improvement of forecast has still to be demonstrated. The use of these data for operational and assimilation/forecast context shall be investigated in the framework of the GMES/MACC project."

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 26943, 2011.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper