Author's response

Sergey Skachko

April 8, 2014

We appreciate very much the Anonymous Referee's comments. We have answered all questions. Each answer starts with "ANSWER:". We have kept the original Referee's comments in Bold.

0.1 Referee 1

1. In section 2.4, the paper describes the method to generate initial ensemble members and model errors in EnKF. It is not clear from the paper whether the initial ensemble members were created at the beginning of every 24-hour assimilation cycle or only at the beginning of the 6-month run. If it is at the beginning of every 24-hour assimilation cycle, the EnKF implemented in this paper did not take the advantage of the flow-dependent error covariance from one-assimilation cycle to the next. If it is at the beginning of the 6- month run, then it is not clear whether the constant magnitude model error is sufficient to overcome the filter divergence. I suggest the authors to clarify this point in this text.

ANSWER: We followed the suggestion of the referee and added the explanation to the second paragraphe of the Sect 2.4: 'The EnKF uses flow-dependent ensemble forecast error covariance (12) evolving in time with the ensemble. On the contrary, 4D-Var reinitializes the background error covariance every 24 h.'

2. EnKF can easily calculate the uncertainty of the analysis field along with the mean state, while 4D-Var would need approximation to obtain analysis uncertainty. I suggest authors adding some discussion about the uncertainty estimation from the EnKF, and adding the range of OMF statistics from EnKF when compared to 4D-Var estimates, such as in Figures 3, 4, 5, and 6.

ANSWER: The aim of the present paper is to compare properly both systems on their best possible level of performance in the form as they are usually used. As far as 4D-Var does not provide the analysis uncertainty, we decided not to include this discussion in the paper. However, we estimated the EnKF analysis error obtained in our experiments. This estimation is performed using the variance of the ensemble of analyses. It was found that the analysis error in the observation space is less than 0.5% from the background state, which makes its graphical representation difficult. The analysis error remains stable in the whole model domain and does not exceed 3.5% from the background state during the whole experiment.

3. The authors manually tuned the magnitude of model error term and the observation error in the paper. In the conclusion, the authors argued that EnKF requires more tuning than 4D-Var to get comparable performance even with relatively easier implementation. It is known that EnKF needs inflation in the background error covariance to avoid filter divergence, equivalent to the model error term in this paper, Anderson, J. L (2007b, 2009) and Miyoshi (2011) have discussed adaptive inflation. Li et al. (2009) discussed estimating the adaptive inflation factor and observation error covariance simultaneously in EnKF. I recommend the authors implementing the adaptive inflation method and observation error estimation strategy to their EnKF. The implementation of these methods would significantly re-

duce the tuning time for EnKF. If these methods could not apply in their EnKF, the authors should add some comments on why.

ANSWER: We agree with the referee that the automative calibration procedure would be usefull in the EnKF system, especially for the future work using full stratospheric chemistry set and instantenious assimilation of different chemical species. We added the following phrase to the last paragraphe of conclusions: 'The application of the EnKF method to the full-chemistry model may require a careful tuning procedure for each chemical species, a task that can be time consuming. Hence an adaptive calibration procedure of the error covariances (similar to Li et al. (2009)) should be implemented.'

Minor comments: There are several places with spelling and grammar mistakes. 1. Line 25 on Page 357: "....the differences (bias) between observations and forecasts, as well as the and their standard deviation", should be: "....the differences (bias) between observations and forecasts, as well as their standard deviation".

ANSWER: Done.

2. Line 13 on page 359: "the 4D-Var providing values slightly lower that those from the EnKF". It should be: "the 4D-Var providing values slightly lower than those from the EnKF".

ANSWER: Done.

3. Line 23 on page 360: "the standard deviation is smoother that displayed by the χ^2", should be: "the standard deviation is smoother than displayed by the χ^2",

ANSWER: Done

0.2 Referee 2

Major comments: 1. The following argument: "Since the biases are markedly similar, they have most probably the same causes: these can be deficiencies in the model and in the observation dataset, but not in the assimilation algorithm nor in the error calibration." can be specious because similar biases is not sufficient to prove that the assimilation and error calibration are correct. I suggest removing the last part after "but".

ANSWER: We are agree with the referee and have removed this part.

2. Please explain the experimental setup [preferably between 3 and 3.1] for the two strategies. Is EnKF restarted?

ANSWER: No, it uses flow-dependent background covariance from the beginning of the integration and is never restarted.

Is 4D-Var started from the previous forecast?

ANSWER: Yes, it is restarted using 24 h forecast of the previous 4D-Var analysis. But, the 4D-Var background error statistics is not cycled from one 24h assimilation window to the next.

Also please provide more details about the model and observations; how sparse are the observations? Some details are provided in Sec. 2.1

ANSWER: Sect.2.1 contains the information about the numerics, the temporal and spatial resolution, advection scheme used for the model used in the study. In this experiment, it is run as a pure tracer transport model, and no further details are needed. Sect.3.1 describes the spacing and the number of EOS Aura-MLS ozone profiles. It is a dense and complete global coverage in a 24 h period which is not shown here.

3. Consider analyzing the the 4D-Var and EnKF solutions against each other. OmF is an indicator of how far are they from the observations, but if the two are indeed close than that would be an interesting conclusion.

ANSWER: Fig. 7 of the article shows this comparison at a given time and for monthly means. We showed that individual analyses of EnKF and 4D-Var may have differences. However, the monthly mean show any important difference.

4. Similar conclusions as expressed in argument 1. above may be obtained if the assimilation is constrained by data. Details in point 2. may clarify this item; however, leaving out observations may truly elucidate this point. Moreover, leaving out observations may also be used in a cross-validation experiment and will enhance any conclusions from 2.

ANSWER: The aim of the paper is to compare both assimilation systems on the same level of performance and assimilating the identical observations. So, we did not think remove observations from our experiments.

5. The calibration of the observation error covariance in EnKF and 4D-Var can have Bayesian interpretation. In this case, r is a stationary random variable with it's own uninformative prior or a hyperparameter. This variable is calibrated first, and its maximum likelihood value is used. This means that the error model for the observations is assumed to be additive and the actual errors to be unknown and part of the inverse problem and it should be stated as such.

ANSWER: By multiplying the observation error covariance by r, we continue to assume that the observation error is uncorrelated and additive to the state.

6. EnKF and 4D-Var use different statistical assumptions. The most problematic in this comparison is the model error, and therefore are directly comparable unless there is no model error. This does not invalidate the conlusions of this study; however, the conclusions can be overstated. In particular, the use of the same r in both strategies may be inadequate.

ANSWER: We modified the following part of the introduction: 'In the 4D-Var scheme, the evolution of forecast error within the assimilation window is computed by the model (whether it is accurate and appropriate or not) and is used generally as a strong constraint. By contrast the EnKF relaxes this assumption into a weak constraint by adding a model error covariance to the analysis error covariance which becomes dynamically-propagated. Hence, the model error covariance is of great importance for the filter performance.' Besides, we have the first paragraphe on the page 362 regarding the mentionned problem.

7. The posterior covariance of EnKF - readily available - is not used to assess its correct forecast. This is just a suggestion, but I realize the challenge in processing such datasets.

ANSWER: The OmF statistics is computed as a zonal and temporal mean, whereas the forecast error covariance matrix is instantenious. This is why, we decided to use the χ^2 diagnostic allowing to compare the error covariances at the observation points.

9. The covariance tuning for EnKF may become unstable based on the results presented herein. How would it be affected if the assimilation would have been started on September 1st where there is a distinctive growth in the chi factor?

ANSWER: The error covariance calibration was performed using the whole experiment period. And the obtained values of the adjustible parameters infer a stable χ^2 values.

Graphics comments: Suggest using Julian date (or day of year) in on the x axis in Fig. 2, 5, 6.

ANSWER: Our choice of using month rather than the number of days from the beginning of the year seems to be more useful to identify geophysical events like the beginning of the ozone hole, for example.

Typographical errors: P 352, L 23: missing "to be"

ANSWER: Done

0.3 Referee 3

Major comments: 1. The overall conclusion from this paper on the comparison between 4D-Var and EnKF suggests no significant difference between the two data assimilation methods. However, Fig. 5 shows 4D-Var outperforms EnKF only very slightly but quite consistently. If the statistical hypothesis testing (or statistical inference) considers the correlations in time series, statistical significance may be obtained to support 4D-Var's advantage. However, it appears that the paper does not consider correlations in statistical inference, this does not seem a wise choice to make the comparison of time series. I would suggest performing statistical inference with correlations considered, that may lead to different conclusion.

ANSWER: Fig.4 shows the OmF statistics and the statistical T- and F-tests for all pressure levels during the period of one month, where the difference in performance of the two systems is the most important. And the statistical tests show that there is no significant difference for confidence interval of 95% between the 4D-Var and EnKF OmF's at any pressure level. Fig.5 shows the same OmF statistics but for a particular pressure level during the whole experiment. The T- and F-tests may be performed for any given day of the temporal error evolution on Fig.5, but they provide similar results (not shown): the difference between the 4D-Var and EnKF OmF's is insignificant for the same confidence interval. Moreover, considering the correlations for the standard deviations and biases may be difficult to interpret.

Minor comments: 1. P.340, L.2, "The" -¿ "An"

ANSWER: Done

2. Eq. (18), the notations of rho and the Schur product do not look precise, need revisions. Eq. (17) assumes that rho has the same shape as the B matrix, and that the open circle operator indicates the element-wise product. Eq. (18) uses the same rho and open circle operator but applied to HBH, that has the matrix shape of the R matrix, not B. The same applies to Eq. (19). Also, Eq. (18) is an approximation, should not use the equal sign.

ANSWER: Done. We added indexes m and o to the matrix ρ and added the following sentence to the text: 'The indexes m and o are introduced to show that the dimension of the matrix ρ corresponds to the model and observation space dimensions, when the Schur product is applied to the matrix $\mathbf{B}_e\mathbf{H}^T$ and $\mathbf{HB}_e\mathbf{H}^T$, respectively.'

3. P.352, L.13, zeta in Eq. (4) is not an analysis increment, but control variables.

ANSWER: Yes. We corrected the sentence.

4. P.352, L.24, "mathematical" -¿ "statistical" ?

ANSWER: Yes.

5. P.372, Fig. 2, the values of alpha and r in figure caption are not consistent with legend and main text.

ANSWER: We checked the text.

6. P.357, L.11, I do not understand why 48 analyses. Does this mean EnKF analysis is computed 48 times during the 24-h period? It is necessary to clarify what "48" means.

ANSWER: The model time step is 0.5 h. So, during 24 h, model performs analyses 48 times. The text is checked.