P. Rayner (Referee) prayner@unimelb.edu.au

Interactive comment on "A Global Carbon Assimilation System using a modified EnKF assimilation method" by S. Zhang et al.

Acknowledgements
The authors gratefully acknowledge Prof P. Rayner for his constructive and relevant comments, which helped greatly in improving the quality of this manuscript. We have checked our work carefully according to these comments and made the suggested changes.

General Comments

This paper presents a new inversion method for CO2 fluxes and applies it to the period 2002–2008. The novelty in the method is the inclusion of CO2 concentration in the state vector allowing the relaxation of the perfect model assumption for transport. The ensemble method used makes this large augmentation of the state vector possible. The explicit treatment of transport error as part of the forecast error also allows better treatment of the observational error since this is now much closer to the observations (previously it was dominated by errors in the transport model). The paper also introduces to Ensemble Kalman Filter inversions the techniques of objective estimation of covariance scaling parameters. These are called inflation parameters in this study but play the same role as the scaling parameters of Michalak et al. (2005). Incidentally I think this paper should be cited. No doubt the authors came to their objective function via the KF literature but a citation would point out the familiarity of the approach to the conventional atmospheric inverse community.

The paper makes an important methodological contribution. It is well written and, most pleasingly, the algorithm is clearly enough described that it could be copied by someone with reasonable knowledge of the field. Analysis of the results is less developed but this is GMD and hopefully this can be taken up at a later date. I have no overall suggestions for the paper but do suggest a couple of small extra pieces of analysis in the specific comments below.

Our reply:

Thank you for your valuable comments.

The inflation parameters in this study do play the same role as the scaling parameters in *Michalak et al. (2005).*

We have cited *Michalak's paper* in the revised version and added the following sentence in 2) Error Step of Section 3.1:

"Michalak et al. (2005) used a similar objective function for estimating the statistical parameters in the atmospheric inverse problems of surface fluxes."

Reference

Michalak, A. M., Hirsch, A., Bruhwiler, L., Gurney, K. R., Peters, W., and Tans, P. P.: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, J. Geophys. Res. [Atmos.], 110, D24107, 10.1029/2005JD005970, 2005.

Specific Comments

overall *It would be good to list the size of the state vector in various*

configurations (with and without concentration).

Our reply:

Thank you for your comment.

The size of the state vector without concentration is 145 (size of scaling factors $\boldsymbol{\lambda}_t$) and the size of the state vector with concentration is 145 (size of scaling factors $\boldsymbol{\lambda}_t$) + 128 × 64 × 28 × 8 × 7 (size of concentration: lon × lat × lev × times/day × days).

We have listed the size of the state vector at the beginning of Section 3 in the revised manuscript:

"*The size of the state vector in this study is 128×64×28×8×7 ($\mathbf{c}_t$: lon×lat×lev×times/day×days) plus 145 ($\boldsymbol{\lambda}_t$).*"

Eq. (1) *Can you justify the 2/3 1/3 split? See later comment for why this might be important.*

Our reply:

Thank you for your comment.

Actually we choose the (2/3,1/3) split by trial tests. We have tested 7 values of a in the following formula,

$$\lambda_{t,i}^{f} = a\lambda_{t-1,i}^{a} + (1-a) + \sqrt{1-a^2}\,\zeta_i$$

The forecast CO_2 concentrations in 2002 and 2003 are compared to the measurements in the following steps. First, the monthly means are calculated at each site (for example, Fig. R1 shows the monthly means of

forecast minus measurement at site TAP_01D0).

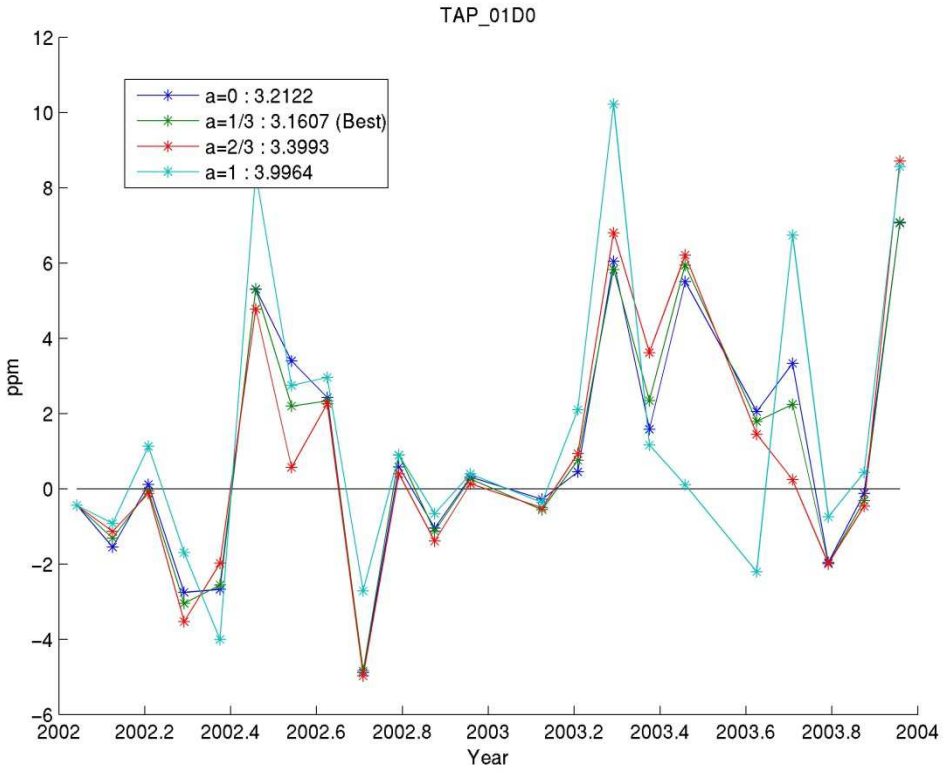

Figure R1. Residuals of monthly mean forecast minus measurement on site TAP_01D0 for four cases: a=0,1/3,2/3,1. The numbers in the legend are root mean square errors of monthly means.

Then we can define a root mean square error at individual sites as

$$r_{site} = \sqrt{\frac{1}{M}\sum_{month=1}^{M} r_{site,month}^{\ 2}}$$

where $r_{site,month}$ is the monthly mean of forecast minus measurement and M is the number of months when there are observations. Finally for all the sites in 2002 and 2003, we use the following relative root mean square error to test different choices of parameter a

$$r_a = \sqrt{\frac{1}{S}\sum_{site=1}^{S} \frac{r_{site}^{\ 2}}{v_{site}^{\ 2}}}$$

where $v_{site}^{\ 2}$ is the given error variance for each site and S is the

number of sites. The results are listed in Table R1. We can see that in 5 cases a=0,1/6,1/3,1/2,2/3 perform similarly while a=1 performs the worst among all cases. The performance of the case when a=5/6 is between the cases of a=2/3 and a=1. We then chose the median value a=1/3 between a=0 and a=2/3 in our formula. Furthermore, the inflation on forecast error covariance will decrease the impact of different choices of coefficient a .

Table R1. Overall relative root mean square error for 7 cases.

	a=0	a=1/6	a=1/3	a=1/2	a=2/3	a=5/6	a=1
$\sqrt{\frac{1}{S}\sum_{site=1}^{S}\frac{r_{site}^{\ 2}}{v_{site}^{\ 2}}}$ (dimensionless)	1.07	1.07	1.06	1.06	1.05	1.12	1.21

Sec 3.2 *We need a little more discussion on the relationship between the iteration of the forecast and analyzed state and the tuning of the inflation parameters. This tuning is set up to ensure that the assumed and actual statistics of departures and innovations are consistent with those assumed in the relevant covariances.*

Our reply:

Thank you for your comment.

As discussed in *Tarantola* (2005), we can calculate the χ^2 statistic of the analysis state for testing the error covariance constructed in this study,

$$\chi^2_{2,Iter} = \left(\mathbf{y}_t^{\mathrm{o}} - \mathcal{H}_t\left(\mathbf{x}_t^{\mathrm{f}}\right)\right)^{\mathrm{T}} \left(\frac{\theta}{m-1}\mathcal{H}_t\left(\widetilde{\mathbf{X}}_t^{\mathrm{f}}\right)\mathcal{H}_t\left(\widetilde{\mathbf{X}}_t^{\mathrm{f}}\right)^{\mathrm{T}} + \mu\mathbf{R}_t\right)^{-1} \left(\mathbf{y}_t^{\mathrm{o}} - \mathcal{H}_t\left(\mathbf{x}_t^{\mathrm{f}}\right)\right)$$

where

$$\widetilde{\mathbf{X}}_t^{\mathrm{f}} = \left(\mathbf{x}_{t,1}^{\mathrm{f}} - \mathbf{x}_t^{\mathrm{a}} \quad \mathbf{x}_{t,2}^{\mathrm{f}} - \mathbf{x}_t^{\mathrm{a}} \quad \cdots \quad \mathbf{x}_{t,m}^{\mathrm{f}} - \mathbf{x}_t^{\mathrm{a}}\right).$$

and θ, μ are the estimated inflation factors for the week associated with $\widetilde{\mathbf{X}}_t^{\mathrm{f}}$. $\chi^2_{2,Iter}$ should be distributed according to the Chi-square probability density with n_{obs} degree of freedom, where n_{obs} is the number of observations within *t*th week. Since the mean and the variance of $\chi^2_{2,Iter}/n_{obs}$ are 1 and $2/n_{obs}$, respectively, the value of $\chi^2_{2,Iter}/n_{obs}$ should be close to 1.

The Chi-square statistics for the error covariance matrices without using the analysis state can be defined similarly, but with $\widetilde{\mathbf{X}}_t^{\mathrm{f}}$ replaced by $\mathbf{X}_t^{\mathrm{f}}$. They are denoted as χ_0^2, χ_1^2 and χ_2^2 for the cases of no inflation, inflation on forecast error only and inflation on both forecast and observation errors, respectively. The closer χ_j^2/n_{obs}, $j = 0,1,2$ to 1 is, the better the corresponding error statistics.

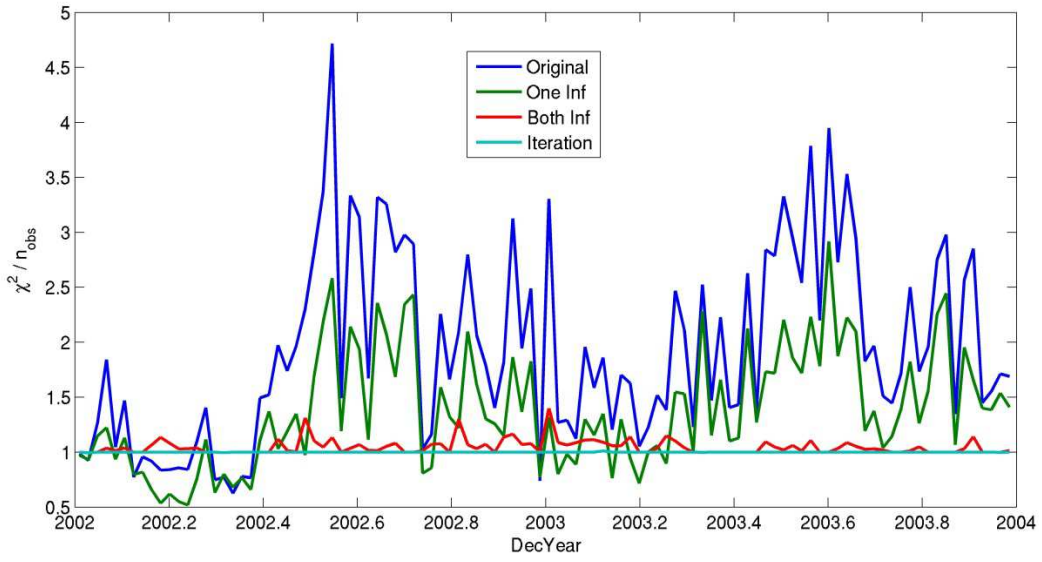

Figure R2. χ^2 statistic of the analysis state for four estimates of error covariance. "Original" refers to the case without inflations; "One Inf" refers to the case with inflation only on forecast error covariance; "Both Inf" refers to the case with inflations on both forecast and observation error covariance and "Iteration" refers to the case with both inflations and further using analysis to improve forecast error statistics. The closer χ^2/n_{obs} is to 1, the better the corresponding error estimates.

For validating the construction of error statistics used in this study, the weekly time series of $\chi^2_{2,Iter}/n_{obs}$ from 2002 to 2003 is shown in Fig. R2. It is remarkably close to 1. The weekly time series of χ^2_0/n_{obs}, χ^2_1/n_{obs} and χ^2_2/n_{obs} for the cases of no inflation, inflation on forecast error only and inflation on both forecast and observation errors are also shown in Fig. R2. All of them are not as close to 1 as that of $\chi^2_{2,Iter}/n_{obs}$. This indicates that the construction of error statistics using the analysis state iteratively is effective for correctly estimating the error statistics. Figure R2 also shows that χ^2_2/n_{obs} is closer to 1 than χ^2_1/n_{obs} is, and both are closer to 1 than χ^2_0/n_{obs} is. This suggests that the inflation on forecast

error and observation error are also both effective in improving the estimation of error statistics.

The above discussions have been added to the revised manuscript.

Reference.

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, Other Titles in Applied Mathematics, Society for Industrial and Applied Mathematics, 348 pp., 2005.

I'm not quite sure what consistency is enforced by the iteration in Sec 2.2 and am a little concerned that the observations might be implicitly used twice, once via the analyzed state now used to describe the forecast uncertainty then again in the update step. This probably reflects limited understanding on my part but I doubt I am alone.

Our reply:

Now we have given some proof of the effectiveness of using analysis to improve the estimation of forecast error covariance by Chi-square test. On the other hand, theoretically the basic assumption of EnKF assimilation is that the forecast and observation are statistically independent. In our iterative scheme, the ensemble forecast is always the same, that is, using observations to estimate the forecast uncertainty do not change the ensemble forecast, so this basic assumption is not violated. Furthermore, in all existing schemes for adaptive estimation of the inflation factor, observations are also used to estimate the forecast

uncertainty since it is the forecast uncertainty being inflated (e.g. *Anderson* (2007), *Li et al.* (2009), *Michalak et al.* (2005), *Miyoshi* (2011), *Wang and Bishop* (2003)). Therefore, we feel that using observations to estimate the forecast uncertainty is justified.

References.

Anderson, J. L.: An adaptive covariance inflation error correction algorithm for ensemble filters, Tellus Ser. A-dynamic Meteorology and Oceanography, 59(2), 210--224, doi:10.1111/j.1600-0870.2006.00216.x, 2007.

Li, H., E. Kalnay, and T. Miyoshi: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter, Q. J. R. Meteorolog. Soc., 135(639), 523--533, doi:10.1002/qj.371, 2009.

Michalak, A. M., Hirsch, A., Bruhwiler, L., Gurney, K. R., Peters, W., and Tans, P. P.: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, J. Geophys. Res. [Atmos.], 110, D24107, 10.1029/2005JD005970, 2005.

Miyoshi, T.: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter, Mon. Weather Rev., 139(5), 1519—1535, 2011.

Wang, X., and C. H. Bishop: A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes, J. Atmos. Sci., 60(9), 1140--1158, doi:10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2, 2003.

P6530 it's a fascinating idea that by hugely increasing the size of the state vector (including concentration) you can actually reduce the computational cost. Shouldn't this be compensated by requiring different ensemble sizes to span the much larger space?

Our reply:

Thank you for your comments.

We fully agree with you that if we increase hugely the size of the state vector, we have to increase the ensemble sizes.

However, since the size of scaling factor vector λ_t is 145 in this study, the degrees of freedom of surface flux sets are less than 145.

On the other hand, the concentrations mix rapidly by diffusion in one week. An intuitional example is given in Fig. R3. We started from one modeled concentration field in July 1st, 2003, and forecasted the concentration field in the following week without any carbon fluxes at land surface (i.e. zero boundary conditions). In this way the diffusion and advection of CO_2 existing in atmosphere at July 1st, 2003, can be investigated. We have plotted the lowest vertical model level since it is most strongly influenced by previous carbon fluxes and thus has largest varibilities at the starting time. It can be seen that after one week the concentration field becomes very smooth. Therefore, the atmospheric CO_2 concentration is mixed rapidly with time and it does not have as large degree of freedom as the size itself.

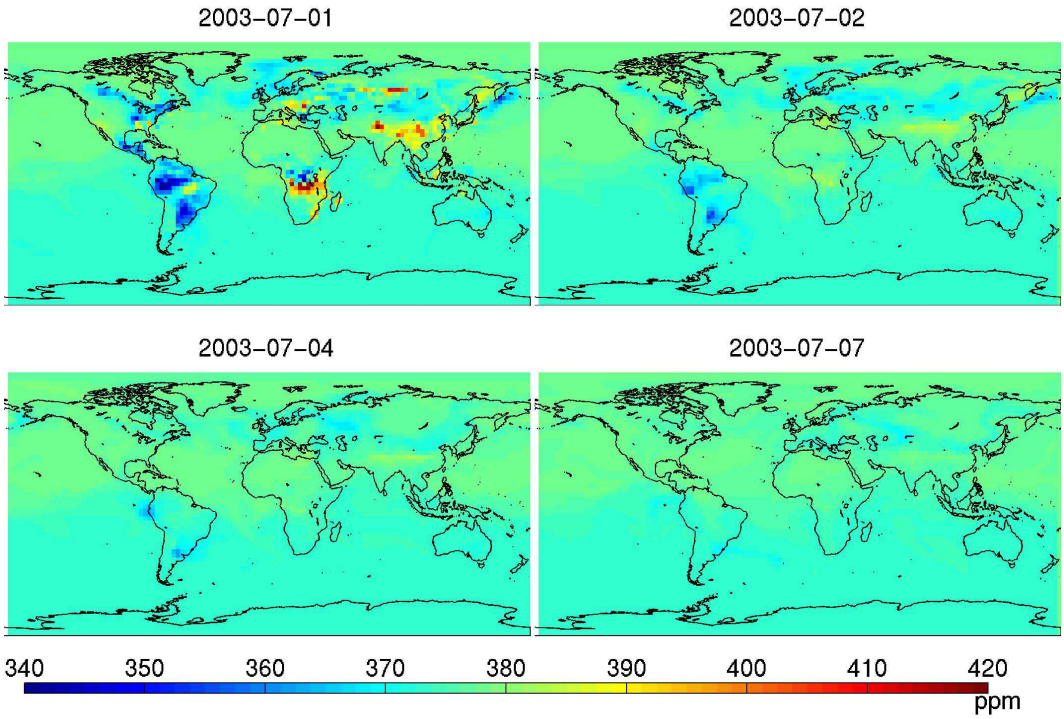

Figure R3. Forecast of concentration field in the lowest vertical model level in one week without carbon fluxes as boundary conditions.

Actually we determined the size of ensemble (150) by experiments. The difference of the assimilated carbon budgets in 2002 is within 10% and the patterns are very similar when we use different ensemble sizes of 150 and 200.

For all above reasons we chose 150 as the default ensemble size in GCAS-EK.

P6531 The bias in the simulation after analysis could be disturbing if it represents a miscalculation of the trend in concentration. Could you plot this bias as a function of time? If there is an error in the concentration trend this would suggest an error in the long-term fluxes.

This is worth discussing since it's always seemed possible in these weak-constraint formulations that we might not match the long-term growth rate.

Our reply:

Thank you for your comments.

Following your advice, we calculated the long-term growth rate in different cases. Atmospheric CO_2 concentration is generated using the sequential forecast of CO_2 concentration with the prior and optimized carbon fluxes, respectively, from 2002 to 2008. The annual mean growth rate with optimized flux (2.17 ppm yr^{-1}) is much closer to observations (2.14 ppm yr^{-1}) than that with prior flux (3.13 ppm yr^{-1}), indicating that we have a good match with the long-term growth rate after optimization. The time series of the bias look similar to the scatter plot in Fig. 6 in the revised manuscript.

We have added the analysis of the long-term growth rates in the revised manuscript.

Sec 6.2] Some of the concern over low variability in may be explained by Eq. 1. The division by 3 should have the effect of strongly smoothing. What would happen if you replaced Eq. 1 with a pure random walk model?

Our reply:

Thank you for your comments.

Some comparisons between Eq. 1 (a=1/3) and a pure random walk model (a=0) can be found in our reply to Specific Comment 2. The overall relative root mean square errors of forecasted CO_2 observations are very close. The estimated annual carbon budget in 2002 and 2003 with the model with a=1/3 is 15% more on average than that with the model with a=0.

However these two models perform differently at individual sites. It can be found in Fig. R4 that which model performs better at each observation site. The sites at which the a=1/3 model performs better are mostly located in or closely to land areas. Since we focus on the optimization of ecosystem carbon fluxes, we prefer to use the strong smoothing model with a=1/3.

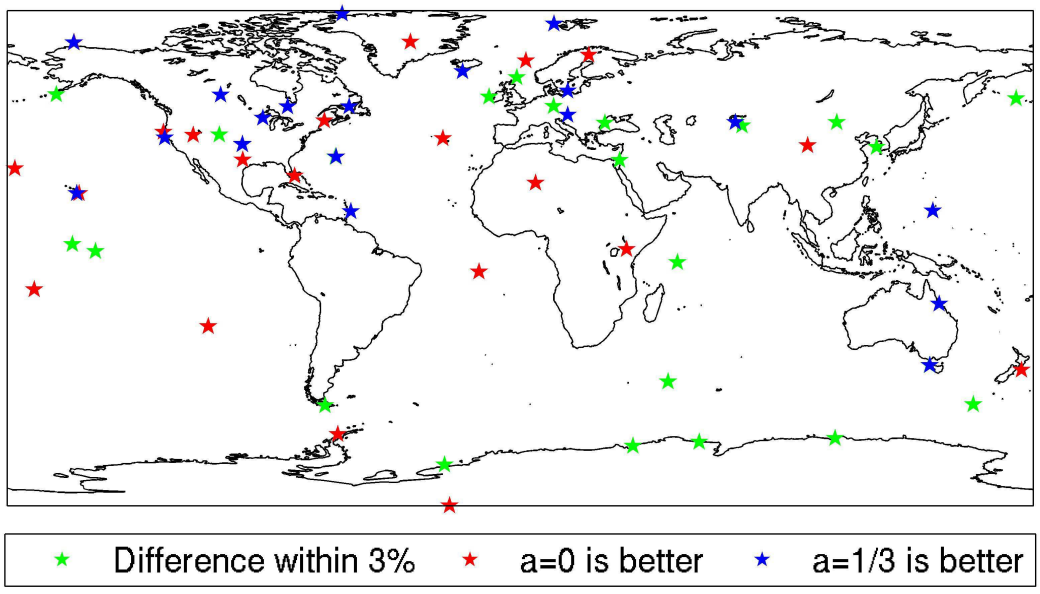

Figure R4. Performance of model a=0 and a=1/3 at different sites. The green stars indicate that these two models have almost equal performance. The red stars indicates that the model with a=0 performs much better at these sites than the model with a=1/3. The blue stars indicates that the model with a=1/3 performs much

better at these sites than the model with a=0.