

This discussion paper is/has been under review for the journal Geoscientific Model Development (GMD). Please refer to the corresponding final paper in GMD if available.



A Global Carbon Assimilation System using a modified EnKF assimilation method

S. Zhang¹, X. Zheng¹, Z. Chen¹, B. Dan¹, J. M. Chen^{2,3}, X. Yi¹, L. Wang⁴, and G. Wu¹

¹College of Global Change and Earth System Science, Beijing Normal University, Beijing, China

²Department of Geography, University of Toronto, Toronto, Canada

³International Institute for Earth System Science, Nanjing University, Nanjing, China

⁴Department of Statistics, University of Manitoba, Winnipeg, Canada

Received: 18 August 2014 – Accepted: 18 September 2014 – Published: 1 October 2014

Correspondence to: X. Zheng (x.zheng@bnu.edu.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.

6519

Abstract

A Global Carbon Assimilation System based on Ensemble Kalman filter (GCAS-EK) is developed for assimilating atmospheric CO₂ abundance data into an ecosystem model to simultaneously estimate the surface carbon fluxes and atmospheric CO₂ distribution. This assimilation approach is based on the ensemble Kalman filter (EnKF), but with several new developments, including using analysis states to iteratively estimate ensemble forecast errors, and a maximum likelihood estimation of the inflation factors of the forecast and observation errors. The proposed assimilation approach is tested in observing system simulation experiments and then used to estimate the terrestrial ecosystem carbon fluxes and atmospheric CO₂ distributions from 2002 to 2008. The results showed that this assimilation approach can effectively reduce the biases and uncertainties of the carbon fluxes simulated by the ecosystem model.

1 Introduction

The carbon dioxide concentration in the atmosphere plays an essential role in the study of global change for its potential to warm up the atmosphere and the surface. A better estimation of carbon fluxes over global ecosystems would help better understand each nation's contribution to the global warming and improve the global warming science.

In the past decade, many efforts have been made to estimate the surface CO₂ fluxes using both atmosphere-based top-down and land-based bottom-up methods. CarbonTracker (Peters et al., 2005, 2007) may be one of the most advanced among these efforts. It uses an ensemble square root filter to assimilate atmospheric CO₂ mole fractions into an ecosystem model coupled with an atmospheric transport model. The model state vectors in CarbonTracker are carbon fluxes within 5 weeks. However, Miyazaki et al. (2011) and Kang et al. (2011) argued that the state vectors should also include atmospheric CO₂ concentration, because the observed CO₂ consists of both existing atmosphere CO₂ and recently released carbon fluxes, so including CO₂



6520

concentration in the state vectors should improve the estimation of carbon fluxes. However, their efforts mainly focus on studying the performance of the assimilation methodology and observation settings by using idealized models only, not on assimilating real observations.

5 It is well known that correct estimation of the forecast error statistics is crucial for the accuracy of any data assimilation algorithm. In all existing EnKF assimilations for estimating carbon fluxes, the ensemble forecast errors are estimated by perturbed forecasts minus their ensemble mean. However, the definition of the perturbed forecast errors is the perturbed forecast states minus the true state. Motivated by the fact
10 that the analysis state is a better estimate of the true state than the forecast state, Wu et al. (2013) proposed a new estimator for the perturbed forecast errors using the perturbed forecast states minus the analysis state. Moreover, they used a simulation study to demonstrate that the new estimator can lead to better assimilations for models with large errors. Since the model errors of ecosystem models are generally large, the
15 new estimation of the perturbed forecast errors is potentially useful to improve EnKF assimilation for estimating carbon fluxes.

Besides forecast errors, the observation errors need also be accurately estimated. In the majority schemes for estimating carbon fluxes, including CarbonTracker, the observation error variances are not estimated but empirically assigned. The quality of
20 the estimation of observation error variances critically depends on whether the forecast error covariance matrix is appropriately estimated or not (Desroziers et al., 2005). However, appropriate estimation of the forecast error covariance matrix is a challenge in real applications.

In this paper we propose several modifications to the conventional EnKF for as-
25 simulating atmospheric CO₂ observations into ecosystem models. Firstly, the model state is set as a combination of the surface carbon fluxes and atmospheric CO₂ concentration as suggested by Miyazaki et al. (2011) and Kang et al. (2011). Secondly, the analysis state is used to adaptively estimate forecast errors as suggested by Wu et al. (2013) and Zheng et al. (2013). Thirdly, both forecast and observation errors are



6521

inflated as suggested by Liang et al. (2012). This modified EnKF is used to assimilate real CO₂ concentration data into the Boreal Ecosystem Productivity Simulator (BEPS, Chen et al., 1999; Liu et al., 1999; Mo et al., 2008) for estimating real world terrestrial carbon fluxes with 3 hourly and 1° × 1° resolution from 2002 to 2008.

5 This paper consists of 7 sections. The models and data used in this study are introduced in Sect. 2, while the methodology is in Sect. 3. Section 4 presents the observing system simulation experiment results. A real data application of the proposed methodology is presented in Sect. 5. Conclusions and discussions are given in Sects. 6 and 7 respectively.

10 2 Models and data

2.1 Surface carbon flux models

The surface carbon fluxes mainly arise from fossil fuel combustion, vegetation fire, oceanic uptake and biosphere. In this study, only the surface carbon fluxes from biosphere are simulated using BEPS, the rest are taken from datasets of CarbonTracker.

15 BEPS is a process-based ecosystem model mainly developed to simulate forest ecosystem carbon budget (Chen et al., 1999; Ju et al., 2006; Liu et al., 1999). For many reasons, such as the complexity of ecosystem processes, spatial–temporal variabilities, representative errors, parameters in process-based models often do not represent their true values when these models are used to calculate carbon budget over large areas and for long time periods (Mo et al., 2008). Errors in these parameters lead to biases of model results. In this study, we try to reduce biases of BEPS simulated carbon fluxes by incorporating atmospheric CO₂ concentration measurements with data assimilation methods. The prior carbon fluxes simulated by BEPS are at a spatial resolution of 1° × 1° and for every 3 h. On each model grid, BEPS calculates carbon fluxes
20 of 6 different ecosystem types and outputs the sum of them through weighting the areal
25



6522

fractions of the ecosystem types. Figure 1 shows the ecosystem types with the largest weight on each grid.

2.2 Atmospheric transport model

The global chemical transport Model for OZone And Related chemical Tracers (MOZART, Emmons et al., 2010) is used as the atmospheric transport model. In this study, MOZART is run at a horizontal resolution of approximately $2.8^\circ \times 2.8^\circ$ with 28 vertical levels. The forcing meteorology is from NCAR reanalysis of the National Centers for Environmental Prediction (NCEP) forecasts (Kalnay et al., 1996; Kistler et al., 2001). Since CO_2 is chemically inert in atmosphere, we turn off all the chemistry processes and leave only transport of CO_2 by atmospheric motions. Given the atmospheric CO_2 concentration in the previous week and the surface carbon fluxes in the current week, MOZART is used to forecast gridded atmospheric CO_2 concentration within the current week.

2.3 Observation

The atmospheric CO_2 concentration measurements collected and preprocessed by CarbonTracker are used in this study. It reflects the variability of the total surface carbon fluxes (i.e. fossil fuel combustion, vegetation fire, oceanic uptake and biosphere) as well as inter-exchange among the CO_2 existing in the atmosphere. The CO_2 dataset released on CarbonTracker's website (<http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/>) includes observations of two main types: the measurements of air samples at surface sites and in situ quasi-continuous CO_2 time series from towers. On average there are about 140 observations in every week during 2000 and 2010.

The observation error variance dataset is also released on CarbonTracker's website. They were subjectively chosen and manually tuned to fit into their models and observations (Peters et al., 2005, 2007). The observation sites were divided into six categories, each with their own assigned observation errors (see the document of CarbonTracker



6523

for details). These observation error variances are used as prior values in our system and are adaptively adjusted with the proposed assimilation scheme.

3 Methodology

Within t th week, let \mathbf{c}_t be a set of gridded atmospheric CO_2 concentrations every 3 h, \mathbf{f}_t be the set of prior carbon fluxes every 3 h, and $\boldsymbol{\lambda}_t$ be a set of factors defined as constants on areas and within a week for adjusting \mathbf{f}_t . Then, the model state is defined as $\mathbf{x}_t = (\mathbf{c}_t^T, \boldsymbol{\lambda}_t^T)^T$.

3.1 EnKF with error inflations

Using the notations of Ide et al. (1997), the first EnKF algorithm used in this study consists of the following three main steps:

1. Forecast step

The perturbed forecast states are estimated as

$$\boldsymbol{\lambda}_{t,i}^f = \frac{2}{3} + \frac{1}{3} \boldsymbol{\lambda}_{t-1,i}^a + \boldsymbol{\xi}_{t,i} \quad (1)$$

$$\mathbf{c}_{t,i}^f = G \left(\mathbf{c}_{t-1,i}^a, \boldsymbol{\lambda}_{t,i}^f \right) \quad (2)$$

where i represents an ensemble member, $\boldsymbol{\xi}_{t,i}$ are vectors sampled from a distribution with mean 0 and a given covariance matrix (taken from prior covariance structure in CarbonTracker, see the document of CarbonTracker and Peters et al., 2005, 2007), and G is the atmospheric transport operator which maps \mathbf{c}_{t-1} and the $\boldsymbol{\lambda}_t$ -adjusted \mathbf{f}_t onto gridded CO_2 concentration. Then the forecast state is estimated as

$$\mathbf{x}_t^f = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_{t,i}^f, \quad (3)$$



6524

where m is the ensemble size.

2. Error step

The ensemble forecast errors and the observation error covariance matrix are estimated as $\sqrt{\theta_t} \mathbf{X}_t^f$ and $\mu_t \mathbf{R}_t$ respectively, where

$$5 \quad \mathbf{X}_t^f = \begin{pmatrix} \mathbf{x}_{t,1}^f - \mathbf{x}_t^f & \mathbf{x}_{t,2}^f - \mathbf{x}_t^f & \cdots & \mathbf{x}_{t,m}^f - \mathbf{x}_t^f \end{pmatrix}, \quad (4)$$

and \mathbf{R}_t is the observation error variance matrix for CarbonTracker. θ_t and μ_t are the inflation factors (of the forecast error and the observation error respectively) which are estimated by minimizing the following objective function (Liang et al., 2012; Zheng, 2009):

$$10 \quad -2L_t(\theta, \mu) = \ln \left\{ \det \left(\frac{\theta}{m-1} \mathcal{H}_t(\mathbf{X}_t^f) \mathcal{H}_t(\mathbf{X}_t^f)^T + \mu \mathbf{R}_t \right) \right\} \\ + \left(\mathbf{y}_t^o - \mathcal{H}_t(\mathbf{x}_t^f) \right)^T \left(\frac{\theta}{m-1} \mathcal{H}_t(\mathbf{X}_t^f) \left(\mathcal{H}_t(\mathbf{X}_t^f) \right)^T + \mu \mathbf{R}_t \right)^{-1} \left(\mathbf{y}_t^o - \mathcal{H}_t(\mathbf{x}_t^f) \right) \quad (5)$$

where \mathbf{y}_t^o is the vector of atmospheric CO₂ concentration measurements, \mathcal{H}_t is a linear observation operator, which interpolates gridded CO₂ concentrations at observation times and locations.

3. Analysis step

The perturbed analysis states are estimated as

$$20 \quad \mathbf{x}_{t,i}^a = \mathbf{x}_{t,i}^f + \sqrt{\theta_t} \mathbf{X}_t^f \left[(m-1) \mathbf{I} + \mathcal{H}_t \left(\sqrt{\theta_t} \mathbf{X}_t^f \right)^T \left(\mu_t \mathbf{R}_t \right)^{-1} \mathcal{H}_t \left(\sqrt{\theta_t} \mathbf{X}_t^f \right) \right]^{-1} \\ \cdot \left(\mathcal{H}_t \left(\sqrt{\theta_t} \mathbf{X}_t^f \right) \right)^T \left(\mu_t \mathbf{R}_t \right)^{-1} \left(\mathbf{y}_t - \mathcal{H}_t \left(\mathbf{x}_{t,i}^f \right) + \boldsymbol{\varepsilon}_{t,i} \right) \quad (6)$$

6525

where $\boldsymbol{\varepsilon}_{t,i}$ is a normal random variable with mean zero and covariance matrix $\mu_t \mathbf{R}_t$ (Burgers et al., 1998). The analysis state \mathbf{x}_t^a is estimated as

$$\mathbf{x}_t^a = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_{t,i}^a \quad (7)$$

5 Finally, set $t = t + 1$ and return to step (1) for the assimilation at next time step.

The assimilated surface carbon fluxes are from all sources because the observed CO₂ concentrations are arising from all sources. Then the surface carbon fluxes from biosphere are estimated by the assimilated total carbon fluxes minus carbon fluxes from other sources supplied by the forcing data.

10 3.2 Constructing error statistics using analysis

Let \mathbf{x}_t^\dagger be the true state. The ensemble forecast error should be defined as $\mathbf{x}_{t,i}^f - \mathbf{x}_t^\dagger$. However, \mathbf{x}_t^\dagger is estimated by \mathbf{x}_t^f in Eq. (4). Since \mathbf{x}_t^a is derived by assimilating observations into the model, it should be a better estimate of \mathbf{x}_t^\dagger than \mathbf{x}_t^f , especially when the model error is large (Wu et al., 2013). Therefore after the analysis step (3) in Sect. 3.1, it is suggested to return to the error step (2), and substitute \mathbf{x}_t^f in Eq. (4) by \mathbf{x}_t^a . This procedure is repeated until the corresponding objective function (Eq. 5) converges (Wu et al., 2013; Zheng et al., 2013). In this study, if the difference between the minima of $-2L_t(\theta, \mu)$ at n th and $n + 1$ th iterations is less one 1, then the iteration is stopped. A flowchart of the proposed assimilation scheme in this study is shown in Fig. 2.

Figure 3 shows that the RMSEs of the estimated carbon fluxes (Eq. 8) in all experiments are reduced on all 11 regions compared with the RMSEs of the prior carbon fluxes. Similarly, Fig. 4 shows that the RMSEs of the estimated in situ atmospheric CO₂ concentrations (Eq. 10) in all experiments are also reduced in all months compared with the RMSEs of the prior CO₂ concentrations. These facts indicate that the data assimilation schemes studied in this paper are useful. Moreover, the RMSEs from Experiment 3 (True R) is smaller than that from Experiment 1 (Wrong R). This suggests that correct specification of the observation error variances is important in improving the assimilation results.

The mean value of the estimated standard observation errors in Experiment 2 (Wrong R + Inf) is 0.262 ppmv. Although it is larger than the true value 0.2 ppmv, but is much smaller than the incorrectly specified value 0.5 ppmv. Nevertheless, the RMSEs from Experiment 2 (Wrong R + Inf) and Experiment 3 (True R) are comparable (see Figs. 3 and 4).

4.3 Constructing forecast error statistics using the analysis states

In this section, we test the methodology of using the analysis state to construct forecast error statistics described in Sect. 3.2. The corresponding Experiment 4 is referred to as “Wrong R + Inf + Anl”.

Figure 3 shows that the RMSEs of the carbon fluxes assimilated in Experiment 4 (Wrong R + Inf + Anl) are the smallest for all regions and in all experiments. Similarly, Fig. 4 shows that the RMSEs of the in situ atmospheric CO₂ concentrations assimilated in Experiment 4 are the smallest for all months except September. These results indicate that the methodology of using the analysis state to construct forecast error statistics described in Sect. 3.2 can improve the assimilation results. The mean value of the estimated standard observation errors in Experiment 4 (Wrong R + Inf + Anl) is 0.268 ppmv, which is virtually identical to that in Experiment 2 (Wrong R + Inf).

6529

4.4 Validation

For investigating the sensitivity of the inflation factors toward the choice of ensemble size, the time series of inflation factors are calculated with the ensemble size 200, and are plotted against the estimated inflation factors with the the ensemble size 150 (not shown here). In fact they are very close, suggesting that the ensemble size 150 used in both CarbonTracker and this study is reasonable.

The analysis spread and RMSE are also calculated for each region using Eq. (9) and they are very close (not shown here). This indicates that the **member** of analysis states may have the same distribution as the true states.

By including CO₂ concentration in the state vectors, the initial value of gridded CO₂ concentration at every week is the analysis states of CO₂ concentration at previous week. However, when excluding CO₂ concentration from the state vectors, the initial value of gridded CO₂ concentration can also be estimated by using atmospheric transport model forced by the assimilated surface carbon fluxes at previous week. Sensitivity experiments were carried out to compare these two approaches. The RMSEs of analysis in the experiment without including CO₂ concentration in state vectors are close to those in Wrong R + Inf + Anl, which suggests that including CO₂ concentration in state vectors may not significantly improve the assimilation results. However, in Wrong R + Inf + Anl the initial states of gridded atmospheric CO₂ concentration need not be re-estimated, thus the computational cost of atmospheric transport model is about half of that in experiment without including CO₂ concentration, which is one advantage of including CO₂ concentration.



5 Application and results

In this section we use the data assimilation methods described in Sect. 3 to estimate the land surface carbon fluxes from 2002 to 2008. The fossil fuel, forest fire and ocean surface carbon flux forcing fields of the atmospheric transport model MOZART are all

6530



taken from CarbonTracker website. The initial atmospheric CO₂ concentration field is also from CarbonTracker products. The prior land surface carbon fluxes are simulated by the ecosystem model BEPS. In this study, only land surface carbon fluxes need to be adjusted. The partition of the adjustment factors (i.e. λ_t in Sect. 3) is based on 11 TransCom regions (Gurney et al., 2004) and 19 Olson ecosystem types, as in CarbonTracker. Prior observation error covariance matrix is adopted from CarbonTracker. We refer to this data assimilation scheme as Global Carbon Assimilation System using Ensemble Kalman filter (GCAS-EK).

5.1 Adjustment to total carbon budget of BEPS

We first carry out a control run starting from 1 January 2002 with no adjustment of prior fluxes. The simulated CO₂ concentrations are interpolated at measurement times and locations, and compared with real observations in the year 2005. The result is shown in Fig. 5a. It shows that the simulated concentrations have a bias of 2.945 ppmv and a RMSE of 4.525 ppmv, which implies an underestimation of carbon sinks by BEPS. With the ecosystem fluxes estimated by GCAS-EK, we carry out another control run and comparisons. The bias and RMSE are reduced to 0.967 and 3.675 ppmv, respectively (Fig. 5b).

We have to point out that the underestimation of carbon sinks by BEPS is conditioned on the estimated carbon fluxes released by fossil fuel and fire, even if the ocean fluxes used in our assimilation system are accurate. As described in Sect. 2, the observed variability of CO₂ concentration is due to the variability of carbon fluxes from all sources, including fossil fuel combustion, vegetation fire, oceanic uptake and biosphere. If carbon sources not from biosphere are underestimated, the carbon sinks from biosphere simulated by BEPS would also be underestimated. Nevertheless, our adjustment to carbon sinks simulated by BEPS appears reasonable.

6531

5.2 Multiyear average of the global carbon flux distribution

The distribution of the average global carbon budget from 2002 to 2008 is shown in Fig. 6. The two spatial patterns of carbon fluxes related to BEPS (Fig. 6a and b) are similar. However, they are quite different from that of CarbonTracker products (Fig. 6c).

In the North American region, CarbonTracker exhibits a nearly west-east strip of the carbon sink (Fig. 6c), while the carbon sinks assimilated or simulated by BEPS are mainly distributed in the east of 95° W (Fig. 6a and b). In the central Africa near the southern edge of Sahara desert, CarbonTracker simulates a strong carbon sink (Fig. 6c), but BEPS simulates a weak sink (Fig. 6a), while the assimilated result shows a weak source (Fig. 6b). In India, CarbonTracker simulates a moderate carbon source (Fig. 6c), while carbon sinks are simulated and assimilated by BEPS (Fig. 6a and b). In Australian Northern Territory, CarbonTracker simulates a carbon sink (Fig. 6c), while the other two produce carbon sources (Fig. 6a and b). In North American Temperate and Eurasia Boreal, the assimilated carbon sink is clearly larger than that simulated.

Carbon budgets are calculated based on the BEPS ecosystem types and the 11 TransCom regions (Fig. 7). Similar to the global distribution maps (Fig. 6), the assimilated BEPS carbon budgets (Fig. 7) have almost the same property in sources or sinks with that simulated by BEPS. However for the C4 and the shrub in Australia, BEPS simulates carbon sources while CarbonTracker shows carbon sinks. Moreover in North America, there is a large carbon sink increase of the assimilated over the BEPS simulated. Further diagnostic (not shown here) reveals that, between October and April, the carbon sinks estimated by CarbonTracker are much larger than that estimated by GCAS-EK. But between May and September, the carbon sinks estimated by CarbonTracker and GCAS-EK are very close.

6532

5.3 Interannual and seasonal variations

The interannual variations of the global total carbon budgets are shown in Fig. 8. It shows that CarbonTracker predicts the largest multiyear average carbon sink ($-3.89 \text{ Pg C year}^{-1}$), compared with the smallest one simulated by BEPS ($-2.23 \text{ Pg C year}^{-1}$). The assimilated mean carbon sink ($-3.87 \text{ Pg C year}^{-1}$) is virtually identical to that estimated by CarbonTracker. The carbon sinks simulated by BEPS and predicted by CarbonTracker obviously have more interannual oscillation than that assimilated by GCAS-EK.

The monthly changes of the multiyear-averaged carbon budgets before and after the assimilation of BEPS results are compared with that by CarbonTracker in Fig. 9. Clearly, the seasonal variability of the carbon budgets by CarbonTracker is the largest. The assimilated fluxes based on BEPS have larger sinks in the summer and smaller sources in the winter than those before the assimilation.

6 Discussion

6.1 Comparison with CarbonTracker

Including CO_2 concentration in the state vector implies that an atmospheric transport model is part of the forecast operator, not part of the observation operator (such as in CarbonTracker). In this framework, the forecast operator comprises an atmospheric transport model and forecast of adjusted factors (Eqs. 1 and 2). The observation operator is the linear operator which interpolates gridded atmospheric CO_2 concentration onto the observation points. Moreover for remotely sensed CO_2 cylinder concentration data, the observation operator can be chosen as a weighted average of gridded atmosphere CO_2 concentrations.

Since atmospheric CO_2 concentration is not model variable in CarbonTracker, the observation operator in this study is different from that in CarbonTracker. In

6533

CarbonTracker, the observation operator is atmospheric transport model coupled with a linear spatial interpolation operator which maps surface CO_2 fluxes to atmospheric CO_2 concentration observation network. Then its observation error comprises the following three components: (1) measurement error covariance, (2) representation error covariance; and (3) model transport error covariance. In this study, the observation operator is only the linear spatial interpolation operator. So the observation error in our experiment only comprises components (1) and (2).

The mean value of the estimated μ_t for inflating the prior observation error variances is 0.74. This indicates that the estimated observation error variances are smaller than that used in CarbonTracker. This may be due to that model transport errors are not included in our observation errors, but are included in the observation errors for CarbonTracker.

6.2 Forecast of adjusted factors

From the extensive experiments conducted in this study, we find that the spatial pattern of assimilated fluxes is highly correlated with the spatial pattern of prior fluxes. This is due to the fact that the unconditional expectation of the analysis $E[\lambda_{t,j}^a]$ is 1, which could be attributed to the setting of forecast of adjusted factors (Eq. 1). Then the probability of shifting between carbon sources and sinks is small. It means that GCAS-EK generally trusts the spatial pattern simulated by the ecosystem model.

To avoid this problem, more flexible adjustment of f_t with more adjusted factors may be considered. However, the increased number of the adjusted factors results in increased degrees of freedom of adjustment model. To fit such kind of model more abundance of observations may be required. For surface flask observations with a total number of about 100 every week over the entire globe, the number of adjusted factors has to be carefully controlled. Therefore, under the current density of ground-based observation network, improving the accuracy of the ecosystem model producing the prior fluxes may be more feasible strategy to improve the surface flux estimation.

6534

6.3 Length of the assimilation time window

Different lengths of the assimilation time window are used in various systems (5 weeks in CarbonTracker, 3 and 7 days in Miyazaki et al., 2011, and 6 h in Kang et al., 2012). We choose one week as the length in our methodology for the following two reasons.

5 Firstly, since most surface stations only have weekly observations, we need at least one week data to cover the globe. Secondly, beyond one week the model errors of MOZART and BEPS may be significant, but they are very difficult to quantify.

7 Conclusions

We propose a methodology to assimilate atmospheric CO₂ concentration into surface carbon fluxes simulated by an ecosystem model. In our framework CO₂ concentration is included in the state vector, both forecast and observation errors are inflated and forecast error statistics are estimated in an adaptive procedure using the analysis states. Generally speaking, these adaptive estimations improve the accuracy of assimilated error statistics in EnKF, which leads to further improvement in the accuracy of analysis states. Importantly, pre-assigned values of the observation error variance are improved if these adaptive procedures are applied.

Four simulation experiments were carried out to show the effectiveness of the proposed methodology. In the first two experiments, we assumed the observation error variances were incorrectly specified and compared assimilation results with and without using inflations on the observation error statistics. The third experiment, in which the observation error variances were supposed to be known, served as a benchmark of how the observation error variances were estimated using our methodology. The fourth experiment showed the effectiveness of using analysis states to further improve the estimation of forecast error. The results from all experiments met our expectation and increased our confidence in applying the improved EnKF to assimilate real observations. The application of the methodology to real data shows that the assimilated

6535

carbon fluxes by GCAS-EK are comparable to those reported by CarbonTracker. However, there are significant regional differences between carbon flux distributions assimilated by GCAS-EK and CarbonTracker, which may be attributed to the differences between the ecosystem models and the assimilation methodologies.

5 In our future study, we will investigate the sensitivity of assimilation results to the accuracy of ecological models and transport models. Also, more observation datasets, such as remote sensing data, will be introduced into the GCAS-EK.

Acknowledgements. This work was supported by National Program on Key Basic Research Project of China (Grant Nos. 2010CB950703 and 2010CB951604), and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

- Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, 126, 1719–1724, doi:10.1175/1520-0493(1998)126<1719:asitek>2.0.co;2, 1998.
- 15 Chen, J. M., Liu, J., Cihlar, J., and Goulden, M. L.: Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications, *Ecol. Model.*, 124, 99–119, doi:10.1016/S0304-3800(99)00156-8, 1999.
- Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, *Q. J. Roy. Meteor. Soc.*, 131, 3385–3396, 2005.
- 20 Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), *Geosci. Model Dev.*, 3, 43–67, doi:10.5194/gmd-3-43-2010, 2010.
- 25 Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Peylin, P., Prather, M., and Taguchi, S.: Transcom 3 inversion intercomparison: model mean

6536

- results for the estimation of seasonal carbon sources and sinks, *Global Biogeochem. Cy.*, 18, GB1010, doi:10.1029/2003GB002111, 2004.
- Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified notation for data assimilation: operational, sequential and variational, *J. Meteorol. Soc. Jpn.*, 75, 181–189, 1997.
- 5 Ju, W. M., Chen, J. M., Black, T. A., Barr, A. G., Liu, J., and Chen, B. Z.: Modelling multi-year coupled carbon and water fluxes in a boreal aspen forest, *Agr. Forest Meteorol.*, 140, 136–151, doi:10.1016/j.agrformet.2006.08.008, 2006.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M.,
10 Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *B. Am. Meteorol. Soc.*, 77, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.
- Kang, J. S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T., and Ide, K.: “Variable localization” in an ensemble Kalman filter: application to the carbon cycle data assimilation, *J. Geophys. Res.*, 116, D09110, doi:10.1029/2010JD014673, 2011.
- 15 Kang, J. S., Kalnay, E., Miyoshi, T., Liu, J., and Fung, I.: Estimation of surface carbon fluxes with an advanced data assimilation methodology, *J. Geophys. Res.*, 117, D24101, doi:10.1029/2012JD018259, 2012.
- Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W.,
20 Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation, *B. Am. Meteorol. Soc.*, 82, 247–267, doi:10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2, 2001.
- Liang, X., Zheng, X., Zhang, S., Wu, G., Dai, Y., and Li, Y.: Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation, *Q. J. Roy. Meteor. Soc.*, 138, 263–273, doi:10.1002/qj.912, 2012.
- 25 Liu, J., Chen, J. M., Cihlar, J., and Chen, W.: Net primary productivity distribution in the BOREAS region from a process model using satellite and surface data, *J. Geophys. Res.*, 104, 27735–27754, doi:10.1029/1999JD900768, 1999.
- Miyazaki, K., Maki, T., Patra, P., and Nakazawa, T.: Assessing the impact of satellite, aircraft,
30 and surface observations on CO₂ flux estimation using an ensemble-based 4-D data assimilation system, *J. Geophys. Res.*, 116, D16306, doi:10.1029/2010JD015366, 2011.

6537

- Mo, X. G., Chen, J. M., Ju, W. M., and Black, T. A.: Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter, *Ecol. Model.*, 217, 157–173, doi:10.1016/j.ecolmodel.2008.06.021, 2008.
- 5 Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO₂ surface fluxes from atmospheric trace gas observations, *J. Geophys. Res.*, 110, D24304, doi:10.1029/2005JD006157, 2005.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K.,
10 Miller, J. B., Bruhwiler, L. M., Petron, G., Hirsch, A. I., Worthy, D. E., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: carbonTracker, *P. Natl. Acad. Sci. USA*, 104, 18925–18930, doi:10.1073/pnas.0708986104, 2007.
- Wu, G., Zheng, X., Wang, L., Zhang, S., Liang, X., and Li, Y.: A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation, *Q. J. Roy. Meteor. Soc.*,
15 139, 795–804, doi:10.1002/qj.2000, 2013.
- Zheng, X.: An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation, *Adv. Atmos. Sci.*, 26, 154–160, doi:10.1007/s00376-009-0154-5, 2009.
- Zheng, X., Wu, G., Zhang, S., Liang, X., Dai, Y., and Li, Y.: Using analysis state to construct a forecast error covariance matrix in ensemble Kalman filter assimilation, *Adv. Atmos. Sci.*,
20 30, 1303–1312, doi:10.1007/s00376-012-2133-5, 2013.

6538

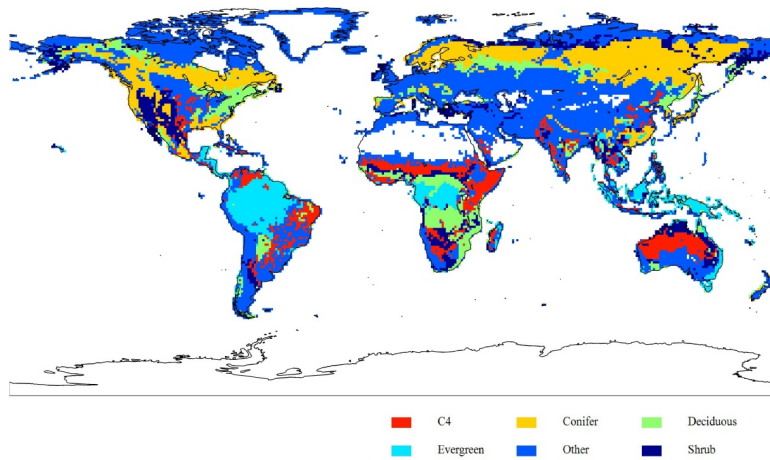


Figure 1. Global distribution of ecosystem types in BEPS.

6539

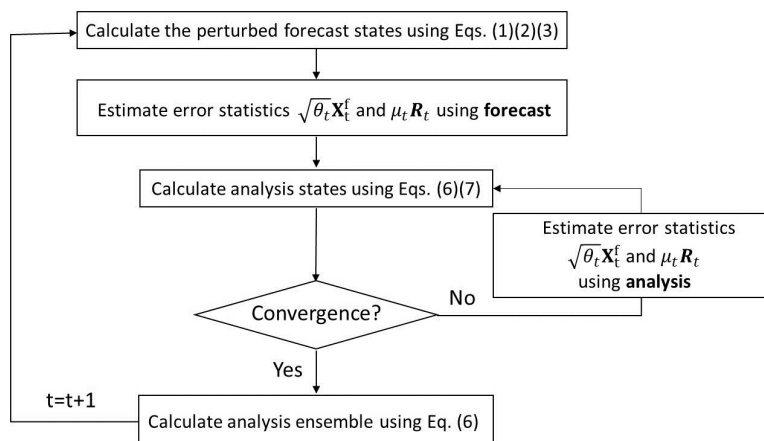


Figure 2. Flowchart of the proposed adaptive procedure.

6540

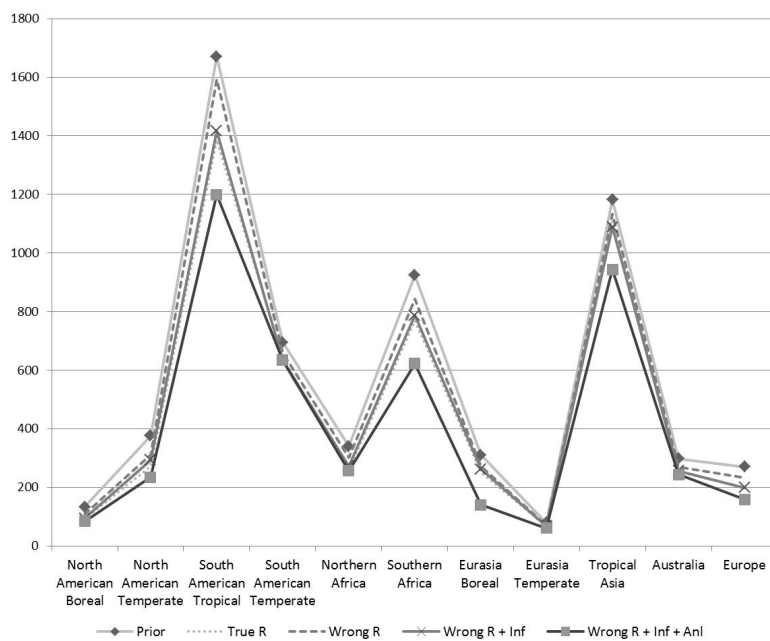


Figure 3. RMSEs ($\text{g C m}^{-2} \text{ year}^{-1}$) of carbon fluxes in four experiments: (1) the “standard” EnKF with incorrectly specified observation errors (Wrong R); (2) the “standard” EnKF with correctly specified observation errors (True R); (3) EnKF with inflation on both forecast and observation errors (Wrong $R + \text{Inf}$); (4) using analysis to further improve the estimation of forecast error statistics (Wrong $R + \text{Inf} + \text{Anl}$).

6541

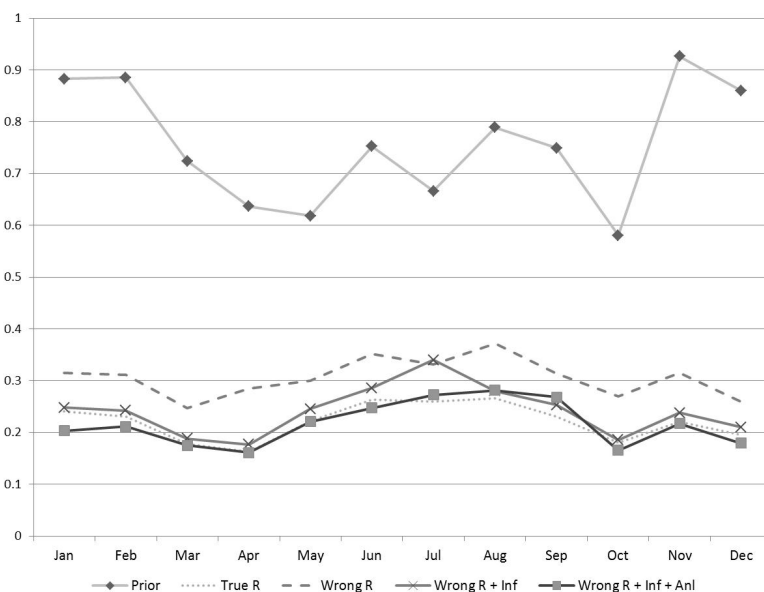


Figure 4. Posterior RMSEs (ppmv) of concentrations on observation sites in four experiments.

6542

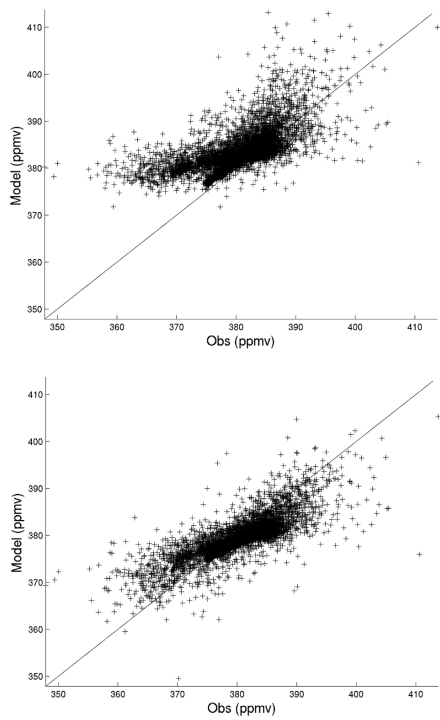


Figure 5. Comparisons between real observations and simulated concentrations by control runs: **(a)** control run forcing by prior carbon fluxes; **(b)** control run forcing by assimilated carbon fluxes by GCAS-EK. Both simulations start from 1 January 2002 and all simulated concentrations at observation locations and times in 2005 are compared here.

6543

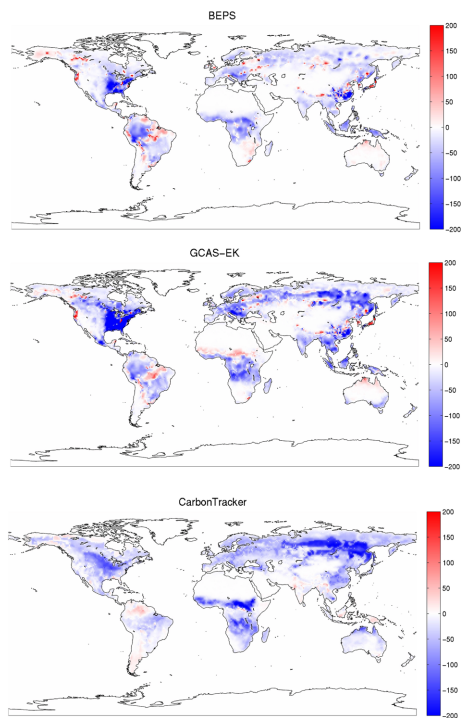


Figure 6. Global carbon budget (g C m^{-2}) distributions on multiyear average from 2002 to 2008: **(a)** prior carbon fluxes simulated by BEPS; **(b)** assimilated carbon fluxes by GCAS-EK; **(c)** CarbonTracker estimated carbon fluxes.

6544

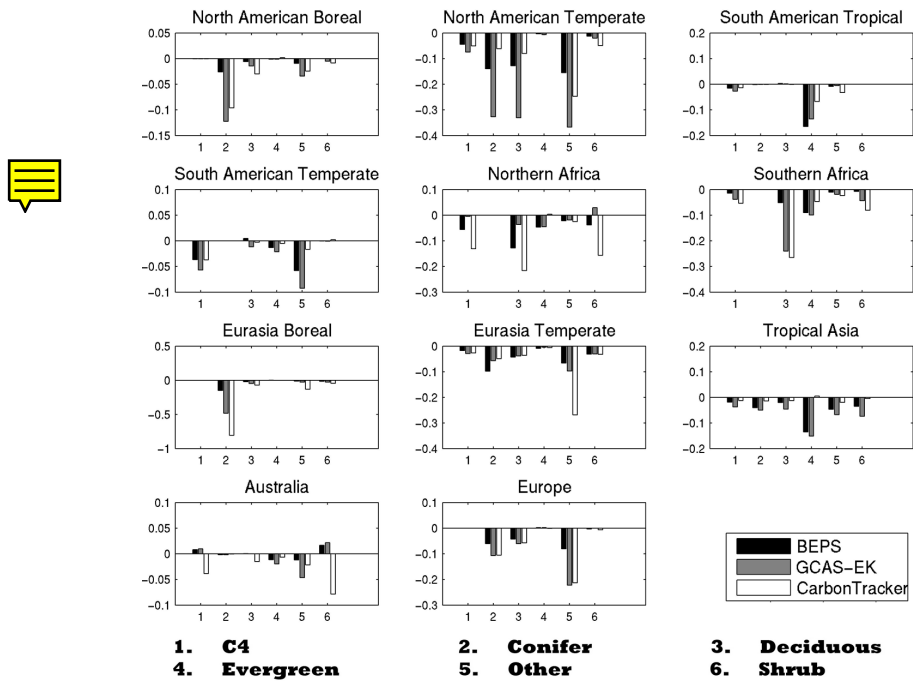


Figure 7. Average carbon budgets (Pg C year^{-1}) on areas of BEPS ecosystem types and TransCom regions from 2002 to 2008.

6545

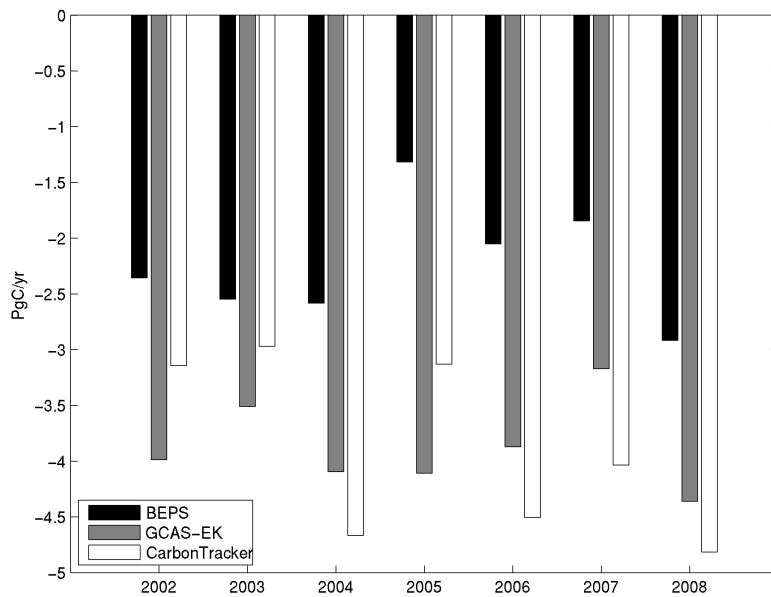


Figure 8. Comparison of the interannual change of global total carbon budget from 2002 to 2008.

6546

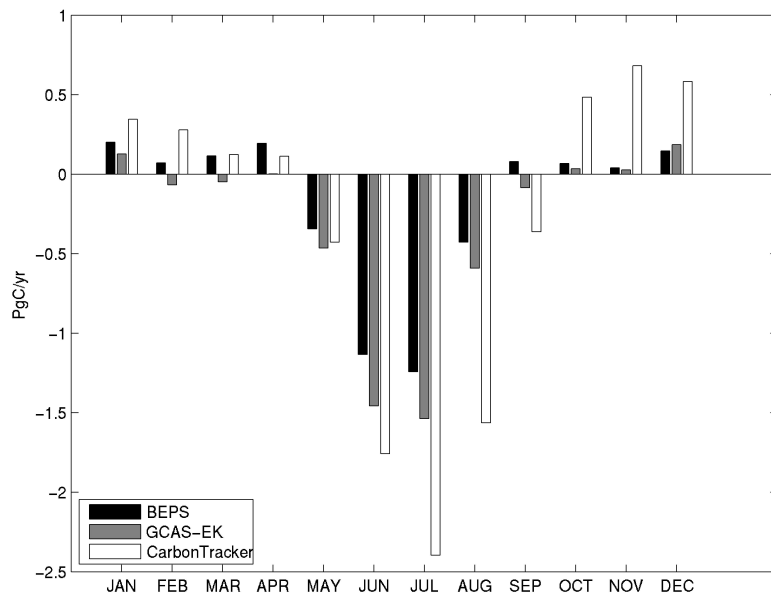


Figure 9. Comparison of multiyear average monthly change from 2002 to 2008.