Geoscientific
Model Development

A global carbon assimilation system using a modified ensemble Kalman filter

S. Zhang[1,2]**, X. Zheng**[1,2]**, J. M. Chen**[2,3,4]**, Z. Chen**[1,2]**, B. Dan**[1,2]**, X. Yi**[1,2]**, L. Wang**[5]**, and G. Wu**[1,2]

[1]College of Global Change and Earth System Science, Beijing Normal University, Beijing, China
[2]Joint Center for Global Change Studies, Beijing, China
[3]Department of Geography, University of Toronto, Toronto, Canada
[4]International Institute for Earth System Science, Nanjing University, Nanjing, China
[5]Department of Statistics, University of Manitoba, Winnipeg, Canada

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

Abstract. A Global Carbon Assimilation System based on the ensemble Kalman filter (GCAS-EK) is developed for assimilating atmospheric CO_2 data into an ecosystem model to simultaneously estimate the surface carbon fluxes and atmospheric CO_2 distribution. This assimilation approach is similar to CarbonTracker, but with several new developments, including inclusion of atmospheric CO_2 concentration in state vectors, using the ensemble Kalman filter (EnKF) with 1-week assimilation windows, 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 used to estimate the terrestrial ecosystem carbon fluxes and atmospheric CO_2 distributions from 2002 to 2008. The results show 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 in better understanding each nation's contribution to global warming and improve global warming science.

In the past decade, many efforts have been made to estimate the surface CO_2 fluxes using both atmosphere-based top-down and land-based bottom-up methods. Carbon-Tracker (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 CO_2 mole fractions into an ecosystem model coupled with an atmospheric transport model.

The model state vectors in CarbonTracker are carbon fluxes only. However, the observed CO_2 consists of both initial state of atmosphere CO_2 and recently released carbon fluxes, and therefore including CO_2 concentration in the state vectors should improve the estimation of initial atmosphere CO_2 (Miyazaki et al., 2011). This could lead to further improvement of carbon flux estimation. Kang et al. (2011) and Liu et al. (2012) also added CO_2 concentrations to the state vectors due to their strong correlations with weather variables that are simultaneously assimilated. 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.

The length of the assimilation window in CarbonTracker is 5 weeks. This would include CO_2 observations far from the analysis time. However, this may not necessarily improve the flux analysis compared to an instantaneous analysis due to the attenuation of the detailed information as discussed by Enting (2002). A shorter assimilation window reduces the attenuation of observed CO_2 information, because the analysis system can use near-surface CO_2 observations before the transport of CO_2 blurs out the essential information of near-surface CO_2 forcing (Kang et al., 2012).

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 ensemble Kalman filter (EnKF) assimilations for estimating carbon fluxes, the ensemble forecast errors are estimated by the difference of perturbed forecasts and their ensemble mean. The perturbed forecast errors are defined as the perturbed forecast states minus the true state. Motivated by the fact 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 by using the difference between the perturbed forecast states and the analysis state. Moreover, they demonstrated through a simulation study that the new estimator can lead to better assimilations for models with large errors. Since the errors of ecosystem models are generally large, the new estimation of the perturbed forecast errors is potentially useful to improve EnKF assimilation for estimating carbon fluxes.

Besides forecast errors, the observation errors also need to be accurately estimated. In the majority of schemes for estimating carbon fluxes, including CarbonTracker, the observation error variances are not estimated but empirically assigned. The quality of the estimation of observation error variances critically depends on whether the forecast error covariance matrix is appropriately estimated (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 assimilating atmospheric CO_2 observations into ecosystem models. First, the model state contains both the surface carbon fluxes and atmospheric CO_2 concentration as suggested by Miyazaki et al. (2011), Kang et al. (2011) and Liu et al. (2012). Second, the analysis state is used to adaptively estimate forecast errors as suggested by Wu et al. (2013) and Zheng et al. (2013), and both forecast and observation errors are inflated as suggested by Liang et al. (2012). Finally, the 1-week assimilation window is tested against longer windows. This modified EnKF is used to assimilate real CO_2 concentration data into the Boreal Ecosystem Productivity Simulator (BEPS; Chen et al., 1999; Liu et al., 1999; Mo et al., 2008) for estimating the real terrestrial carbon fluxes with 3-hourly and $1° \times 1°$ resolution from 2002 to 2008.

This paper consists of six sections. The models and data used in this study are introduced in Sect. 2, while the methodology is described in Sect. 3. Section 4 presents the validations of the new methodologies using the real observing system. A real data application of the proposed methodology is presented in Sect. 5. Conclusions and discussions are given in Sect. 6.

2 Models and data

2.1 Surface carbon flux models

The surface carbon fluxes mainly arise from fossil fuel combustion, vegetation fire, oceanic exchange, and biosphere. In this study, only the surface carbon fluxes from biosphere are simulated using BEPS, while the rests are taken from data sets of CarbonTracker 2011 (http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/).

BEPS is a process-based ecosystem model mainly developed to simulate forest ecosystem carbon budgets (Chen et al., 1999; Ju et al., 2006; Liu et al., 1999). For many reasons, including the complexity of ecosystem processes, spatial–temporal variabilities, and representative errors, parameters in process-based models often do not represent their true values when these models are used to calculate carbon budgets over large areas or for long time periods (Mo et al., 2008). Errors in these parameters lead to biases in model results (other uncertainties, such as lack of knowledge on historical land use change and land management, also have influence on model results). In this study, we try to reduce biases in the BEPS-simulated carbon fluxes by incorporating atmospheric CO_2 concentration measurements with data assimilation methods. The prior carbon fluxes simulated by BEPS are at a spatial resolution of $1° \times 1°$ and for every 1 h. On each model grid, BEPS calculates carbon fluxes of six different plant function types and outputs the sum of them through weighting the fluxes against areal fractions of the plant function types. Figure 1 shows the plant function types with the largest weight on each grid.

The vegetation fire flux is taken from CarbonTracker 2011 data set, which is modeled using the Carnegie–Ames–Stanford Approach (CASA) biosphere model (Potter et al., 1993) based on the Global Fire Emission Database (GFED) (van der Werf et al., 2006).

The oceanic CO_2 flux is taken from CarbonTracker 2011 optimized results, whose a priori estimates are based on two different data sets: namely, the ocean inversion flux result (Jacobson et al., 2007) and $p$$CO_2$-Clim prior estimate derived from the climatology of seawater $p$$CO_2$ (Takahashi et al., 2009).

The fossil fuel combustion estimate is the data set preprocessed by CarbonTracker 2011 from the global total fossil fuel emission of the Carbon Dioxide Information and Analysis Center (CDIAC) (Boden et al., 2011) and the Open-source Data Inventory of Anthropogenic CO_2 emission (ODIAC) data set (Oda and Maksyutov, 2011).

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 $2.8° \times 2.8°$
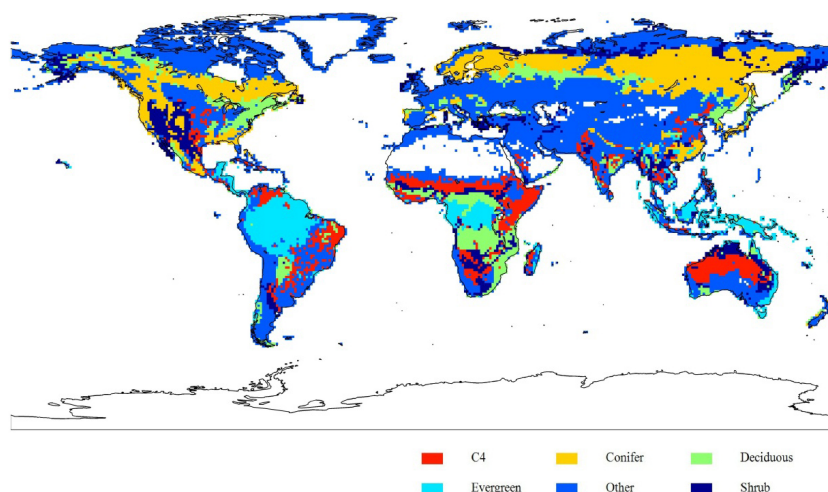
Figure 1. Land areas of six plant function types used in ecosystem model BEPS.

with 28 vertical levels. The forcing meteorology is from the National Center For Atmospheric Research (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 the atmosphere, we turn off all the chemical 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 Observation Package (ObsPack) data product (Masarie et al., 2014) are used in this study (Product Version: obspack_co2_1_CARBONTRACKER_CT2013_2014-05-08). The selected CO_2 measurements on 92 sites include observations of two main types: the measurements of air samples at surface sites and in situ quasi-continuous CO_2 time series from towers. Since some stations have multiple observations within a week, on average there are about 140 observations every week during 2002 and 2008. Six laboratories (NOAA Global Monitoring Division, Commonwealth Scientific and Industrial Research Organization, National Center For Atmospheric Research, Environment Canada, Lawrence Berkeley National Laboratory and Instituto de Pesquisas Energéticas e Nucleares) provided these measurements and information of observation sites used in this study is listed in Table 1. CO_2 concentration measurements reflect 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 CO_2 air mass in the initial atmosphere.

The observation error variances are also provided in obspack_co2_1_CARBONTRACKER_CT2013_2014-05-08. They were subjectively chosen and manually tuned to fit into specific atmospheric transport models and observations (Peters et al., 2005, 2007). Since these values depend on the atmospheric transport model used in a carbon data assimilation system, they are just used as prior values for this study and will be adaptively adjusted with the proposed assimilation scheme.

3 Methodology

Within tth week, let c_t be a set of gridded atmospheric CO_2 concentrations every 3 h, f_t be the set of prior carbon fluxes every 3 h, and λ_t be a set of factors defined as constants on areas and within a week for adjusting f_t. Then, the model state is defined as $x_t = \left(c_t^T, \lambda_t^T\right)^T$. In this study, only land surface carbon fluxes need to be adjusted. The partition of the adjustment factors (i.e. λ_t) is based on 11 Transcom regions (Gurney et al., 2004) and 19 Olson ecosystem types, as in CarbonTracker. Thus, the size of the state vector in this study is $128 \times 64 \times 28 \times 8 \times 7$ (c_t: lon $\times$ lat $\times$ lev $\times$ times/day $\times$ days) plus 145 (λ_t). We refer to this data assimilation scheme as the Global Carbon Assimilation System based on the ensemble Kalman filter (GCAS-EK).

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:

Table 1. Listed are 92 observation sites used in this study. r refers to prescribed observation error (μmol μmol^{-1}).

Site code	Lat (°)	Long (°)	r	Lab	Site code	Lat (°)	Long (°)	r	Lab
ABP_01D0	−12.27	−38.17	2.50	NOAA*	MID_01D0	28.21	−177.38	1.50	NOAA
ABP_26D0	−12.27	−38.17	2.50	IPEN*	MKN_01D0	−0.05	37.30	2.50	NOAA
ALT_01D0	82.45	−62.51	1.50	NOAA	MLO_01C0_02LST	19.54	−155.58	0.75	NOAA
ALT_06C0_14LST	82.45	−62.51	2.50	EC*	MLO_01D0	19.54	−155.58	1.50	NOAA
AMT_01C3_14LST	45.03	−68.68	3.00	NOAA	MQA_02D0	−54.48	158.97	0.75	CSIRO
AMT_01P0	45.03	−68.68	3.00	NOAA	NMB_01D0	−23.58	15.03	2.50	NOAA
ASC_01D0	−7.97	−14.40	0.75	NOAA	NWR_01D0	40.05	−105.58	1.50	NOAA
ASK_01D0	23.18	5.42	1.50	NOAA	NWR_03C0_02LST	40.05	−105.58	3.00	NCAR*
AZR_01D0	38.77	−27.38	1.50	NOAA	OBN_01D0	55.11	36.60	7.50	NOAA
BAL_01D0	55.35	17.22	7.50	NOAA	OXK_01D0	50.03	11.80	2.50	NOAA
BAO_01C3_14LST	40.05	−105.00	3.00	NOAA	PAL_01D0	67.97	24.12	2.50	NOAA
BAO_01P0	40.05	−105.00	3.00	NOAA	POC_01D1	−0.39	−132.32	0.75	NOAA
BHD_01D0	−41.41	174.87	1.50	NOAA	PSA_01D0	−64.92	−64.00	0.75	NOAA
BKT_01D0	−0.20	100.32	7.50	NOAA	PTA_01D0	38.95	−123.74	7.50	NOAA
BME_01D0	32.37	−64.65	1.50	NOAA	RPB_01D0	13.17	−59.43	1.50	NOAA
BMW_01D0	32.27	−64.88	1.50	NOAA	SCT_01C3_14LST	33.41	−81.83	3.00	NOAA
BRW_01C0_14LST	71.32	−156.61	2.50	NOAA	SEY_01D0	−4.67	55.17	0.75	NOAA
BRW_01D0	71.32	−156.61	1.50	NOAA	SGP_01D0	36.80	−97.50	2.50	NOAA
BSC_01D0	44.17	28.68	7.50	NOAA	SGP_64C3_16LST	36.80	−97.50	3.00	LBNL*
CBA_01D0	55.21	−162.72	1.50	NOAA	SHM_01D0	52.72	174.10	2.50	NOAA
CDL_06C0_14LST	53.99	−105.12	3.00	EC	SIS_02D0	60.17	−1.17	2.50	CSIRO
CFA_02D0	−19.28	147.06	2.50	CSIRO*	SMO_01C0_14LST	−14.25	−170.56	0.75	NOAA
CGO_01D0	−40.68	144.69	0.75	NOAA	SMO_01D0	−14.25	−170.56	1.50	NOAA
CGO_02D0	−40.68	144.69	0.75	CSIRO	SNP_01C3_02LST	38.62	−78.35	3.00	NOAA
CHR_01D0	1.70	−157.17	0.75	NOAA	SPL_03C0_02LST	40.45	−106.73	3.00	NCAR
CRZ_01D0	−46.45	51.85	0.75	NOAA	SPO_01C0_14LST	−89.98	−24.80	0.75	NOAA
CYA_02D0	−66.28	110.52	0.75	CSIRO	SPO_01D0	−89.98	−24.80	1.50	NOAA
EGB_06C0_14LST	44.23	−79.78	3.00	EC	STM_01D0	66.00	2.00	1.50	NOAA
EIC_01D0	−27.15	−109.45	7.50	NOAA	STR_01P0	37.76	−122.45	3.00	NOAA
ETL_06C0_14LST	54.35	−104.98	3.00	EC	SUM_01D0	72.58	−38.48	1.50	NOAA
FSD_06C0_14LST	49.88	−81.57	3.00	EC	SYO_01D0	−69.00	39.58	0.75	NOAA
GMI_01D0	13.43	144.78	1.50	NOAA	TAP_01D0	36.73	126.13	7.50	NOAA
HBA_01D0	−75.58	−26.50	0.75	NOAA	TDF_01D0	−54.87	−68.48	0.75	NOAA
HPB_01D0	47.80	11.01	7.50	NOAA	THD_01D0	41.05	−124.15	2.50	NOAA
HUN_01D0	46.95	16.65	7.50	NOAA	UTA_01D0	39.90	−113.72	2.50	NOAA
ICE_01D0	63.40	−20.29	1.50	NOAA	UUM_01D0	44.45	111.10	2.50	NOAA
KEY_01D0	25.67	−80.16	2.50	NOAA	WBI_01C3_14LST	41.72	−91.35	3.00	NOAA
KUM_01D0	19.52	−154.82	1.50	NOAA	WBI_01P0	41.72	−91.35	3.00	NOAA
KZD_01D0	44.06	76.82	2.50	NOAA	WGC_01C3_14LST	38.27	−121.49	3.00	NOAA
KZM_01D0	43.25	77.88	2.50	NOAA	WGC_01P0	38.27	−121.49	3.00	NOAA
LEF_01C3_14LST	45.95	−90.27	3.00	NOAA	WIS_01D0	31.13	34.88	2.50	NOAA
LEF_01P0	45.95	−90.27	3.00	NOAA	WKT_01C3_14LST	31.31	−97.33	3.00	NOAA
LLB_06C0_14LST	54.95	−112.45	3.00	EC	WKT_01P0	31.31	−97.33	3.00	NOAA
LMP_01D0	35.52	12.62	1.50	NOAA	WLG_01D0	36.29	100.90	1.50	NOAA
MAA_02D0	−67.62	62.87	0.75	CSIRO	WSA_06C0_14LST	49.93	−60.02	3.00	EC
MHD_01D0	53.33	−9.90	2.50	NOAA	ZEP_01D0	78.90	11.88	1.50	NOAA

* NOAA: NOAA Global Monitoring Division; CSIRO: Commonwealth Scientific and Industrial Research Organization; NCAR: National Center For Atmospheric Research; EC: Environment Canada; IPEN: Instituto de Pesquisas Energéticas e Nucleares; LBNL: Lawrence Berkeley National Laboratory.

3.1.1 Forecast step

The perturbed forecast states are estimated as

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

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

where i represents an ensemble member, $\boldsymbol{\xi}_{t,i}$ are vectors sampled from a distribution with mean zero 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 $\boldsymbol{c}_{t-1}$ and the $\boldsymbol{\lambda}_t$ adjusted $\boldsymbol{f}_t$ onto gridded CO_2 concentration. Then the forecast state is estimated as

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

where m is the ensemble size.

3.1.2 Error step

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

$$\boldsymbol{X}_t^{\mathrm{f}} = \left(\boldsymbol{x}_{t,1}^{\mathrm{f}} - \boldsymbol{x}_t^{\mathrm{f}} \quad \boldsymbol{x}_{t,2}^{\mathrm{f}} - \boldsymbol{x}_t^{\mathrm{f}} \quad \ldots \quad \boldsymbol{x}_{t,m}^{\mathrm{f}} - \boldsymbol{x}_t^{\mathrm{f}}\right), \qquad (4)$$

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

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

where $\boldsymbol{y}_t^{\mathrm{o}}$ is the vector of atmospheric CO_2 concentration measurements, $\mathcal{H}_t$ is a linear observation operator, which interpolates gridded CO_2 concentrations at observation times and locations. Michalak et al. (2005) used a similar objective function for estimating the statistical parameters in the atmospheric inverse problems of surface fluxes.
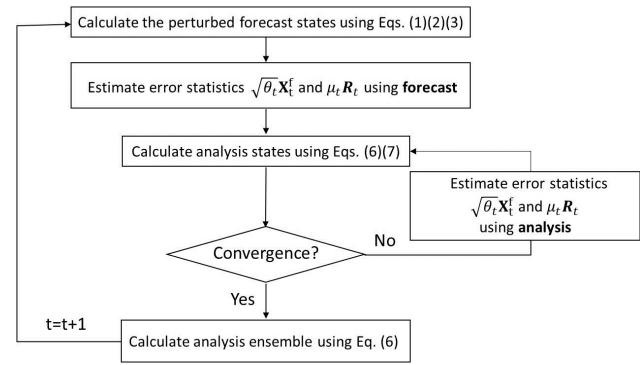

Figure 2. Flowchart of modified ensemble Kalman filter.

3.1.3 Analysis step

The perturbed analysis states are estimated as

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

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

$$\boldsymbol{x}_t^{\mathrm{a}} = \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{x}_{t,i}^{\mathrm{a}}. \qquad (7)$$

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

The assimilated surface carbon fluxes are from all sources because the observed CO_2 concentrations arise from all sources. Then, the surface carbon fluxes from the biosphere are estimated by the assimilated total carbon fluxes minus carbon fluxes from other sources supplied by the forcing data.

3.2 Constructing error statistics using analysis

Let $\boldsymbol{x}_t^{\mathrm{t}}$ be the true state. Then the ensemble forecast error should be defined as $\boldsymbol{x}_{t,i}^{\mathrm{f}} - \boldsymbol{x}_t^{\mathrm{t}}$. However, $\boldsymbol{x}_t^{\mathrm{t}}$ is estimated by $\boldsymbol{x}_t^{\mathrm{f}}$ in Eq. (4). Since $\boldsymbol{x}_t^{\mathrm{a}}$ is derived by assimilating observations into the model, it is a better estimate of $\boldsymbol{x}_t^{\mathrm{t}}$ than $\boldsymbol{x}_t^{\mathrm{f}}$, 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 $\boldsymbol{x}_t^{\mathrm{f}}$ in Eq. (4) by $\boldsymbol{x}_t^{\mathrm{a}}$. This procedure is repeated until the corresponding objective function (Eq. 5) converges (Wu et al., 2013; Zheng et al., 2013). In this study, the iteration is stopped when the difference between the minima of $-2L_t(\theta,\mu)$ at nth and $n+1$th iterations is less than 1. A flowchart of the proposed assimilation scheme is shown in Fig. 2.
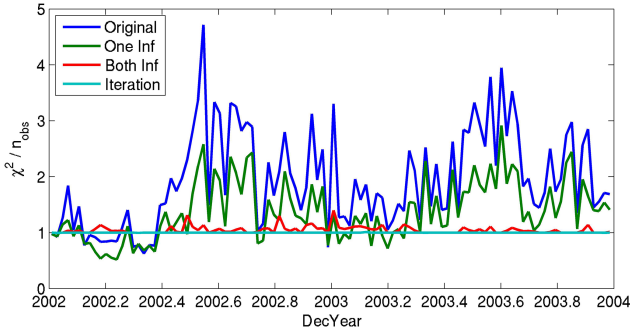
Figure 3. χ^2 statistics of the analysis state for four estimates of error covariance. Original refers to the case without inflations, one Inf refers to the case with inflation on forecast error covariance only, both Inf refers to the case with inflations on both forecast and observation error covariances and iteration refers to the case with both inflations and further using analysis to improve forecast error statistics. The closer $\chi^2/n_{\rm obs}$ is to 1, the better the corresponding error estimates.

3.3 Removing carbon mass imbalance

In this study, the background CO_2 concentration field at the beginning of a week is the analysis state at the end of the previous week. It is then updated using the observations within the week; therefore, the estimated CO_2 concentration at the beginning of the week is different from that at the end of the previous week. This results in inexact carbon mass balance. To remove this imbalance, a corrected atmospheric CO_2 concentration is generated using the sequential forecast of CO_2 concentration with the optimized carbon fluxes from the very beginning of the entire assimilation period. The corrected CO_2 concentration is denoted by $c_t^{\rm ca}$.

3.4 Validation statistics

Chi-square statistics (Tarantola, 2005) are used to test the error covariances constructed in this study. For the tth week, it is defined as

$$
\chi^2_{2,\rm Iter} = \left(\mathbf{y}_t^{\rm o} - \mathcal{H}_t \left(\mathbf{x}_t^{\rm f} \right) \right)^T
$$
$$
\left(\frac{\theta}{m-1} \mathcal{H}_t \left(\widetilde{\mathbf{X}_t^{\rm f}} \right) \mathcal{H}_t \left(\widetilde{\mathbf{X}_t^{\rm f}} \right)^T + \mu \mathbf{R}_t \right)^{-1}
$$
$$
\left(\mathbf{y}_t^{\rm o} - \mathcal{H}_t \left(\mathbf{x}_t^{\rm f} \right) \right),
\tag{8}
$$

where

$$
\widetilde{\mathbf{X}_t^{\rm f}} = \left(\mathbf{x}_{t,1}^{\rm f} - \mathbf{x}_t^{\rm a} \quad \mathbf{x}_{t,2}^{\rm f} - \mathbf{x}_t^{\rm a} \quad \ldots \quad \mathbf{x}_{t,m}^{\rm f} - \mathbf{x}_t^{\rm a} \right),
\tag{9}
$$

and θ, μ are the estimated inflation factors for the week. If the forecast and observation error covariance matrix are correctly estimated, $\chi^2_{2,\rm Iter}$ follows a Chi-square distribution with $n_{\rm obs}$ degrees of freedom, where $n_{\rm obs}$ is the number of observations within tth week. Since the mean and the vari-
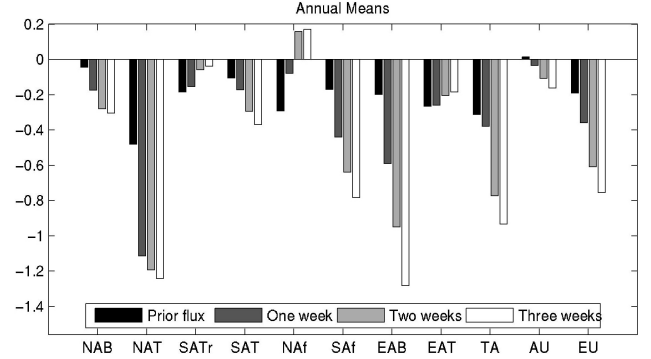

Figure 4. Annual means of carbon budgets (PgC yr^{-1}) on 11 Transcom regions in four different cases. Four cases are associated with prior values modeled with ecosystem model BEPS, assimilated results using GCAS-EK with 1-week assimilation windows, 2-week windows and 3-week windows. In all, 11 regions in x axis refer to North American boreal (NAB), North American temperate (NAT), South American tropical (SATr), South American temperate (SAT), northern Africa (NAf), southern Africa (SAf), Eurasia boreal (EAB), Eurasia temperate (EAT), tropical Asia (TA), Australia (AU) and Europe (EU), respectively.

ance of $\chi^2_{2,\rm Iter}/n_{\rm obs}$ are 1 and $2/n_{\rm obs}$, respectively, the value of $\chi^2_{2,\rm Iter}/n_{\rm obs}$ should be close to 1.

The Chi-square statistics for the error covariance matrices without using the analysis state can be defined similarly to Eq. (8), but with $\widetilde{\mathbf{X}_t^{\rm f}}$ replaced by $\mathbf{X}_t^{\rm f}$. They are denoted as χ^2_0, χ^2_1 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 $\chi^2_j/n_{\rm obs}$, $j = 0, 1, 2$ to 1 is, the better the corresponding error statistics.

The root mean square error (RMSE) of estimated CO_2 observations is defined as

$$
\sqrt{\frac{1}{L} \sum_{i,l} \left(y_i^{\rm ca}(l) - y_i^{\rm o}(l) \right)^2},
\tag{10}
$$

where $y_i^{\rm ca}(l)$ is generated by interpolating $c_t^{\rm ca}$ to the observation site l and time i, and L is the total number of the CO_2 concentration observations during the entire assimilation period. The smaller RMSE means better assimilation scheme.

4 Discussions on methodology

4.1 Error covariance statistics

To validate the construction of error statistics used in this study, we plot the weekly time series of $\chi^2_{2,\rm Iter}/n_{\rm obs}$ (Eq. 8) from 2002 to 2003 in Fig. 3, which shows that the values are remarkably close to 1. In contrast, the weekly time series of $\chi^2_0/n_{\rm obs}$, $\chi^2_1/n_{\rm obs}$, and $\chi^2_2/n_{\rm obs}$ (for the cases of no inflation, inflation on forecast error only, and inflation on both forecast and observation errors) are not as close to 1 as $\chi^2_{2,\rm Iter}/n_{\rm obs}$.
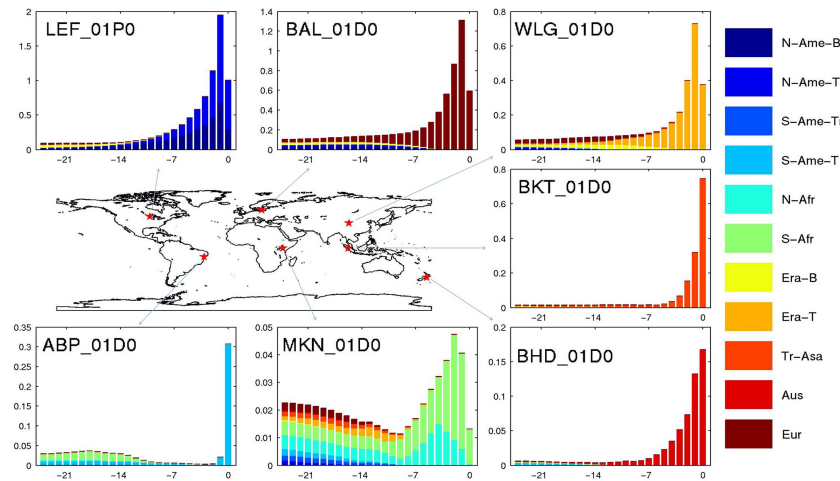
Figure 5. Mean components of CO_2 concentration at observation sites (Site IDs: LEF_01P0, BAL_01D0, WLG_01D0, BKT_01D0, BHD_01D0, MKN_01D0 and ABP_01D0) from 11 Transcom regions in each of the 25 days before the observation time; x axis refers to days before the observation time; y axis refers to the amount of CO_2 concentration in ppm. Different colors within a bar refer to CO_2 concentration from 11 different Transcom regions; 11 regions refer to North American boreal (N-Ame-B), North American temperate (N-Ame-T), South American tropical (S-Ame-Tr), South American temperate (S-Ame-T), northern Africa (N-Afr), southern Africa (S-Afr), Eurasia boreal (Era-B), Eurasia temperate (Era-T), tropical Asia (Tr-Asa), Australia (Aus) and Europe (Eur), respectively.

This indicates that the construction of error statistics using the analysis state iteratively (Sect. 3.2) is effective for correctly estimating the error statistics.

Figure 3 also shows that χ_2^2/n_{obs} is closer to 1 than χ_1^2/n_{obs}, and both are closer to 1 than χ_0^2/n_{obs}. This suggests that the inflations on forecast error and observation error are also both effective in improving the estimation of error statistics.

4.2 Inclusion of CO_2 concentration in state vectors

In this study, the CO_2 concentration is included in state vectors. The benefit of this inclusion needs to be tested against the traditional approach without this inclusion. This issue is studied with the 1-week assimilation window.

For this purpose we design a comparative experiment as follows. In every week, the CO_2 concentration (i.e. c) is not updated (Eq. 6). Instead the analysis CO_2 concentration is derived by sequentially predicting atmospheric CO_2 concentration forced by the updated flux within the week. The carbon mass is automatically balanced in this experiment. The results show that RMSE of the analysis CO_2 concentration observations (Eq. 10) is 8.5 % larger than that of the corrected analysis CO_2 concentration described in Sect. 3.3. This suggests that inclusion of CO_2 concentration in state vectors can significantly alter the CO_2 mass balance and may have an advantage in optimizing the surface CO_2 flux.

If the CO_2 concentration is not included in state vectors, the analysis CO_2 concentration at the beginning of each week is just the analysis CO_2 concentration at the end of the previous week, so the CO_2 concentration observations within the current week are not used to optimize the CO_2 concen-

tration at the beginning of each week. However, when the CO_2 concentration is included in state vectors, all the observations within the current week and the previous weeks are used to estimate the CO_2 concentration at the beginning of the current week. So the CO_2 concentration at the beginning of each week estimated by inclusion of CO_2 concentration in state vectors could be more accurate than estimated without inclusion. Therefore, the estimated flux associated with the updated CO_2 concentration at the beginning of the current week should have better quality. This is more clearly demonstrated by smaller RMSE in Eq. (10) with the inclusion than that without the inclusion.

4.3 Length of assimilation window

Different lengths of the assimilation 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 the 1-week assimilation window in our methodology for the following reasons. First, since most surface stations only have weekly observations, we need at least 1 week of data to cover the globe. Second, beyond 1 week the errors of the atmospheric transport model may be significant, and they are very difficult to quantify. Third, the detailed information of observations may be attenuated with time by atmospheric diffusion and advection (Enting, 2002).

For comparison to longer assimilation windows, the following alternative experiments with moving assimilation windows were carried out. In the first alternative experiment, the length of the moving window is set to be 2 weeks while the forecast time step is still 1 week. The CO_2 concentration observation system is still the same as that described in

Sect. 3, but is used to update the global carbon flux and the atmospheric CO_2 concentration within the current week and the previous week. This procedure is similar to Eq. (6), while the ensemble forecast state of the first week in the assimilation window is set as its ensemble analysis state at previous assimilation time step. Therefore, carbon fluxes and CO_2 concentration every week are optimized twice with the observations in the current week and the next week. The corrected analysis of CO_2 concentration is also retrieved from rerunning the atmospheric transport model as described in Sect. 3.3. The second alternative experiment is similar to the first one, but with the 3-week moving window.

The linear trends for the observations, the estimates with 1-week, 2-week and 3-week moving windows are 2.14, 2.17, 1.59, and 1.13 ppm yr^{-1}, respectively. It seems that the longer the moving window is, the larger difference is the long-term growth rate to the measurements. For further investigating the reason, the annual mean carbon budgets on 11 Transcom regions are shown in Fig. 4. It can be found that the longer the moving window is, the larger are the carbon budget adjustments. Long windows result in underestimation of the corresponding long-term growth rate.

To further investigate the long time and long distance impact of atmospheric transport on CO_2 observations, components of CO_2 concentration at observation sites associated with different Transcom regions in each day before their observation times are calculated in the following way. For a given region and some day before the observation time, prior fluxes on other regions and in other days are all masked. Then the atmospheric transport model can be run with a homogeneous initial atmospheric CO_2 concentration and forced by the masked fluxes to obtain the corresponding CO_2 concentration components.

These components at individual sites are then averaged in time to investigate general impacts of carbon fluxes from different sources. The results at 7 selected sites are shown in Fig. 5. For these sites, CO_2 concentrations resulting from carbon fluxes within 25 days are mainly from local carbon fluxes within 7 days (although mostly within 3 days). Carbon fluxes beyond seven days or regions far from the observation locations have very small impacts, indicating that they have little information in observations (i.e. the contribution is less than observation error), even if the atmospheric transport model is accurate. Actually majority of observations (approximately 49) over continental sites used in this study have similar properties to these seven sites. If the errors of the transport and ecosystem models are considered, the information of fluxes 1 week before may be even more difficult to estimate.

The setting of length of the assimilation window is closely related to spatial and temporal localizations of forecast errors. For the observation network and the atmospheric transport model used in this study, the 1-week assimilation window seems most suitable.
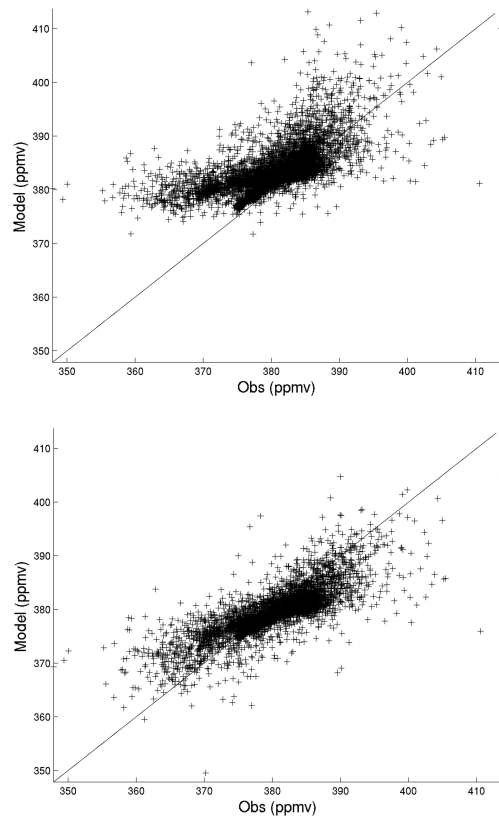

Figure 6. Comparisons between real observations and simulated concentrations by control runs: (top) control run forcing by prior carbon fluxes; (bottom) control run forcing by assimilated carbon fluxes by GCAS-EK. Both simulations start from 1 January 2002 and all simulated concentrations at observation locations and times in 2005 are compared here.

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.

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 CO_2 concentrations are interpolated at observation times and locations, and compared with real observations in the year 2005. The result in Fig. 6 (top) shows that the simulated concentrations have a bias of 2.945 ppm and an RMSE of 4.525 ppm, implying an underestimation of carbon sinks by BEPS. Using GCAS-EK to estimate the ecosystem fluxes, we carry out another control run and comparisons. The bias and RMSE are reduced to 0.967 and 3.675 ppm, respectively (Fig. 6, bottom).

It is worthwhile to point out that the underestimation of carbon sinks by BEPS is conditioned on the estimated carbon
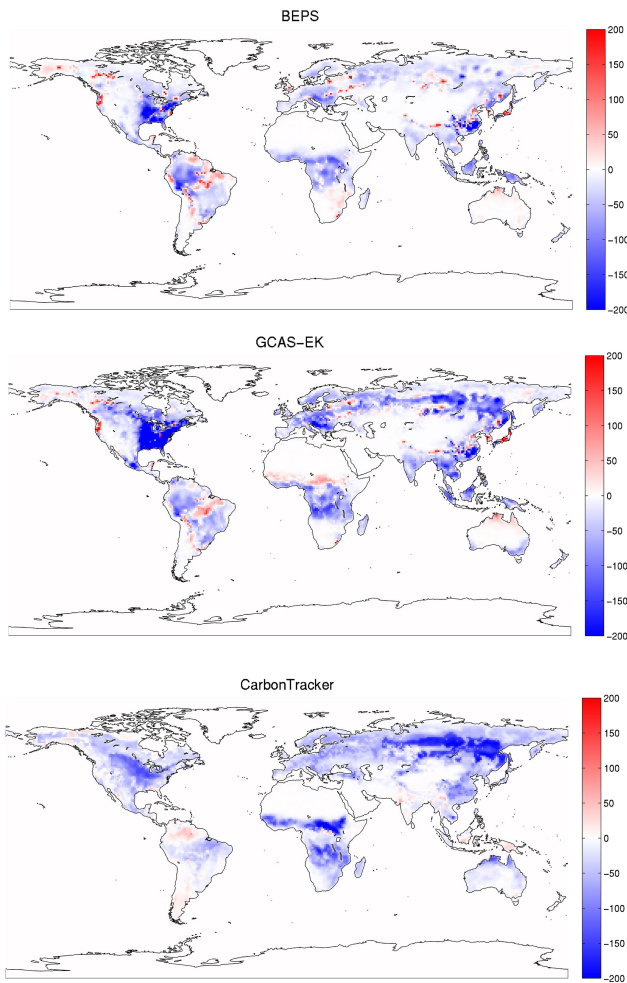
Figure 7. Global carbon budget (gC m^{-2}) distributions on multiyear average from 2002 to 2008: prior carbon fluxes simulated by BEPS; assimilated carbon fluxes by GCAS-EK; CarbonTracker 2011 estimated carbon fluxes.

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 CO_2 concentration is due to the variability of carbon fluxes from all sources, including fossil fuel combustion, vegetation fire, oceanic uptake, and biosphere exchange. If non-biospheric carbon sources are underestimated, the carbon sinks from the biosphere simulated by BEPS would also be underestimated. Nevertheless, our adjustment to carbon sinks simulated by BEPS appears reasonable.

5.2 Multiyear average of the global carbon flux distribution

Figure 7 shows the distribution of the average global carbon budget from 2002 to 2008 where the two spatial patterns of carbon fluxes related to BEPS (BEPS and GCAS-EK) are

Figure 8. Annual mean carbon budgets (PgC yr^{-1}) on areas with six BEPS plant function types in Transcom regions from 2002 to 2008. The errors of GCAS-EK fluxes are the root mean square errors of the ensemble.

similar, although they are quite different from that of CarbonTracker 2011.

Carbon budgets are calculated based on the BEPS ecosystem types and the 11 Transcom regions (Fig. 8). Similar to the global distribution maps (Fig. 7), GCAS-EK carbon budgets (Fig. 8) have almost the same property in sources or sinks with that simulated by BEPS. However, they are quite different from that of CarbonTracker 2011 in many aspects. For example, for the C4 and the shrub in Australia, BEPS simulates carbon sources while CarbonTracker 2011 shows carbon sinks. Moreover, in North America, there is a large carbon sink increase of the GCAS-EK over the BEPS simulated. A further diagnostic (not shown here) reveals that, between October and April, the carbon sinks estimated by CarbonTracker 2011 are much larger than that estimated by GCAS-EK. But between May and September, the carbon sinks estimated by CarbonTracker 2011 and GCAS-EK are very close.

5.3 Interannual and seasonal variations

The interannual variations of the global total carbon budgets are shown in Fig. 9. It shows that CarbonTracker 2011 predicts the largest multiyear average carbon sink (-3.89 PgC yr^{-1}), compared with the smallest one simulated by BEPS (-2.23 PgC yr^{-1}). The assimilated mean carbon sink (-3.87 PgC yr^{-1}) is virtually identical to that estimated by CarbonTracker 2011. The carbon sinks simulated by BEPS and predicted by CarbonTracker 2011 obviously have more interannual oscillation than that assimilated by GCAS-EK.
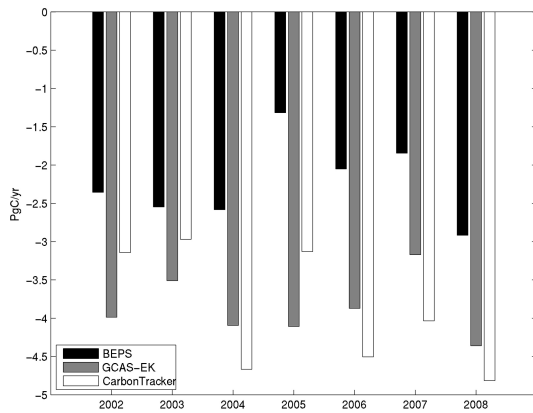
Figure 9. Comparison of interannual variations of global carbon budgets from 2002 to 2008 by three products: BEPS, GCAS-EK and CarbonTracker 2011.
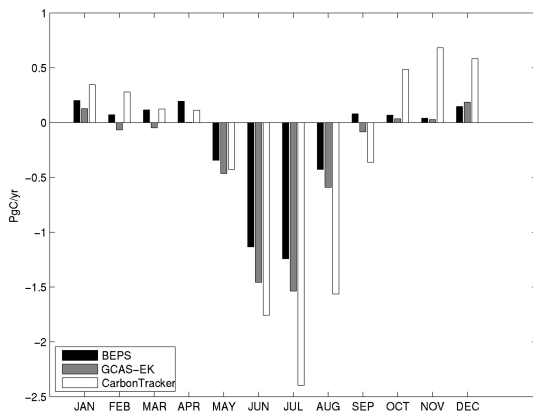

Figure 10. Comparison of multiyear average monthly variations from 2002 to 2008 by three products: BEPS, GCAS-EK and CarbonTracker 2011.

The monthly variations of the multiyear-averaged carbon budgets before and after the assimilation of BEPS results are compared with that by CarbonTracker 2011 in Fig. 10. Clearly, the seasonal variability of the carbon budgets by CarbonTracker 2011 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.

5.4 Comparison to other flux estimations

Two independent gridded carbon flux estimates are compared with GCAS-EK estimates.

The first independent data set is net carbon exchange of U.S. terrestrial ecosystems by Xiao et al. (2011) which is generated by integrating eddy covariance flux measurements and satellite observations from Moderate Resolution Imaging Spectroradiometer (MODIS), and is referred to as EC-MOD. The original data set is during 2002 to 2006 with spatial resolution of 1 km and temporal resolution of 8 day. For
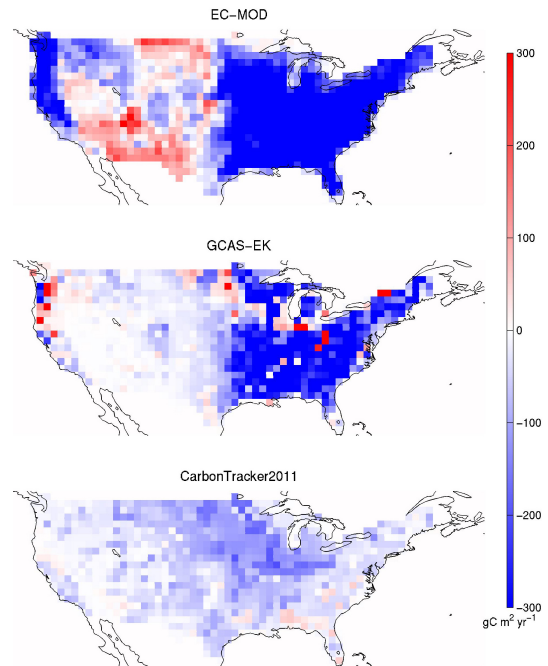

Figure 11. The distribution of averaged net ecosystem exchange ($gC\,m^{-2}\,yr^{-1}$) from 2002 to 2006 for conterminous U.S. by EC-MOD, GCAS-EK and CarbonTracker 2011, respectively. The pattern correlation coefficient is 0.47 between EC-MOD and GCAS-EK, and 0.22 between CarbonTracker 2011 and EC-MOD.

comparison, Xiao's data were grouped from 1 km to 1° spatial resolution. The carbon flux distributions of the multiyear average from 2002 to 2006 over United States are shown in Fig. 11 for Xiao's data, GCAS-EK and CarbonTracker 2011. It shows that spatial pattern of the flux assimilated by GCAS-EK is closer to Xiao's data (with spatial standard deviation $153\,gC\,m^2\,yr^{-1}$ and spatial correlation 0.47) than that by CarbonTracker 2011 (with spatial standard deviation $197\,gC\,m^2\,yr^{-1}$ and spatial correlation 0.22).

The carbon budgets estimated by GCAS-EK were also compared to those by Lauvaux et al. (2012), Penn State University (PSU) inversion and Colorado State University (CSU) inversion (Schuh et al., 2013) for the Mid-Continent Intensive (MCI) area from June to December 2007. The spatial patterns by GCAS-EK and CarbonTracker 2011 are similar to those estimated by PSU, CSU (Schuh et al., 2013) and Lauvaux et al. (2012) (not shown here). The regional-averaged carbon sinks estimated by GCAS-EK and by CarbonTracker 2011 are 0.19 and 0.26 PgC, respectively, while the averaged carbon sinks estimated by PSU and CSU (Schuh et al., 2013) and by Lauvaux et al. (2012) are between 0.14 and 0.18 PgC, which are closer to that estimated by GCAS-EK than that by CarbonTracker 2011.

Since the true values of carbon flux are unknown, the closeness to the independent gridded carbon flux estimates does not mean a better assimilation. However, these two ex-

amples indicate that the carbon fluxes estimated by GCAS-EK may provide some useful new information of global carbon flux estimation to the atmospheric inversion community. Therefore, the development of the new assimilation system is worthwhile.

6 Conclusion

We propose a methodology to assimilate atmospheric CO_2 concentration into surface carbon fluxes simulated by an ecosystem model. In our framework, CO_2 concentration is included in the state vector, and the assimilation window is restricted to 1 week. 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.

The application of the methodology to real data shows that the assimilated total carbon budgets by GCAS-EK are comparable to those reported by CarbonTracker 2011. However, there are significant regional differences between carbon flux distributions assimilated by GCAS-EK and CarbonTracker 2011, which may be attributed to the differences between the ecosystem models, atmospheric transport models, or the assimilation methodologies.

In our future study, we will investigate the sensitivity of assimilation results to the accuracy of ecosystem and transport models. Also, more observation data sets, such as remote-sensing CO_2 column data, will be introduced into the GCAS-EK.

References

Boden, T. A., Marland, G., and Andres, R. J.: Global, Regional, and National Fossil-Fuel CO_2 Emissions, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tenn., USA, doi:10.3334/CDIAC/00001_V2011, 2011.

Burgers, G., Jan van Leeuwen, P., 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.

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, doi:10.1256/qj.05.108, 2005.

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.

Enting, I. G.: Inverse Problems in Atmospheric Constituent Transport, Cambridge University Press, New York, 2002.

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 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.

Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L., and Gloor, M.: A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes, Global Biogeochem. Cy., 21, GB1019, doi:10.1029/2005GB002556, 2007.

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., 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.

Kang, J. S., Kalnay, E., Miyoshi, T., Liu, J., and Fung, I.: Estimation of surface carbon fluxes with an advanced data assim-

ilation methodology, J. Geophys. Res.-Atmos., 117, D24101, doi:10.1029/2012JD018259, 2012.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., 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.

Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., Sweeney, C., Diaz, L. I., Martins, D., Shepson, P. B., and Davis, K. J.: Constraining the CO_2 budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system, Atmos. Chem. Phys., 12, 337–354, doi:10.5194/acp-12-337-2012, 2012.

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.

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.-Atmos., 104, 27735–27754, doi:10.1029/1999JD900768, 1999.

Liu, J., Fung, I., Kalnay, E., Kang, J.-S., Olsen, E. T., and Chen, L.: Simultaneous assimilation of AIRS Xco2 and meteorological observations in a carbon climate model with an ensemble Kalman filter, J. Geophys. Res.-Atmos., 117, D05309, doi:10.1029/2011JD016642, 2012.

Masarie, K. A., Peters, W., Jacobson, A. R., and Tans, P. P.: ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements, Earth Syst. Sci. Data, 6, 375–384, doi:10.5194/essd-6-375-2014, 2014.

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, doi:10.1029/2005JD005970, 2005.

Miyazaki, K., Maki, T., Patra, P., and Nakazawa, T.: Assessing the impact of satellite, aircraft, and surface observations on CO_2 flux estimation using an ensemble-based 4-D data assimilation system, J. Geophys. Res.-Atmos., 116, D16306, doi:10.1029/2010JD015366, 2011.

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.

Oda, T. and Maksyutov, S.: A very high-resolution (1 km × 1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys., 11, 543–556, doi:10.5194/acp-11-543-2011, 2011.

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 CO_2 surface fluxes from atmospheric trace gas observations, J. Geophys. Res.-Atmos., 110, D24304, doi:10.1029/2005JD006157, 2005.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., 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.

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., and Klooster, S. A.: Terrestrial ecosystem production: A process model based on global satellite and surface data, Global Biogeochem. Cy., 7, 811–841, doi:10.1029/93GB02725, 1993.

Schuh, A. E., Lauvaux, T., West, T. O., Denning, A. S., Davis, K. J., Miles, N., Richardson, S., Uliasz, M., Lokupitiya, E., Cooley, D., Andrews, A., and Ogle, S.: Evaluating atmospheric CO_2 inversions at multiple scales over a highly inventoried agricultural landscape, Global Change Biol., 19, 1424–1439, doi:10.1111/gcb.12141, 2013.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R., and de Baar, H. J. W.: Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans, Deep-Sea Res. Pt. II, 56, 554–577, doi:10.1016/j.dsr2.2008.12.009, 2009.

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.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos. Chem. Phys., 6, 3423–3441, doi:10.5194/acp-6-3423-2006, 2006.

Wu, G. C., Zheng, X. G., Wang, L. Q., Zhang, S. P., Liang, X., and Li, Y.: A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation, Q. J. Roy. Meteor. Soc., 139, 795–804, doi:10.1002/Qj.2000, 2013.

Xiao, J., Zhuang, Q., Law, B. E., Baldocchi, D. D., Chen, J., Richardson, A. D., Melillo, J. M., Davis, K. J., Hollinger, D. Y., Wharton, S., Oren, R., Noormets, A., Fischer, M. L., Verma, S. B., Cook, D. R., Sun, G., McNulty, S., Wofsy, S. C., Bolstad, P. V., Burns, S. P., Curtis, P. S., Drake, B. G., Falk, M., Foster, D. R., Gu, L., Hadley, J. L., Katul, G. G., Litvak, M., Ma, S., Martin, T. A., Matamala, R., Meyers, T. P., Monson, R. K., Munger, J. W., Oechel, W. C., Paw, U. K. T., Schmid, H. P., Scott, R. L., Starr, G., Suyker, A. E., and Torn, M. S.: Assessing net ecosystem carbon exchange of U.S. terrestrial ecosystems by integrating eddy covariance flux measurements and satellite observations, Agr. Forest Meteorol., 151, 60–69, doi:10.1016/j.agrformet.2010.09.002, 2011.

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., 30, 1303–1312, doi:10.1007/s00376-012-2133-5, 2013.