



# A joint data assimilation system (Tan-Tracker) to simultaneously estimate surface CO<sub>2</sub> fluxes and 3-D atmospheric CO<sub>2</sub> concentrations from observations

X. Tian<sup>1,2</sup>, Z. Xie<sup>3</sup>, Y. Liu<sup>4</sup>, Z. Cai<sup>4</sup>, Y. Fu<sup>5</sup>, H. Zhang<sup>6</sup>, and L. Feng<sup>7</sup>

<sup>1</sup>ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

<sup>2</sup>Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing, 210044, China

<sup>3</sup>LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

<sup>4</sup>LAGEO, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

<sup>5</sup>Climate Change Research Center (CCRC), Chinese Academy of Sciences, Beijing, China

<sup>6</sup>Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, China

<sup>7</sup>School of GeoSciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JN, UK

Correspondence to: X. Tian (tianxj@mail.iap.ac.cn)

Received: 4 September 2013 – Published in Atmos. Chem. Phys. Discuss.: 24 September 2013

Revised: 8 October 2014 – Accepted: 12 November 2014 – Published: 12 December 2014

**Abstract.** We have developed a novel framework (“Tan-Tracker”) for assimilating observations of atmospheric CO<sub>2</sub> concentrations, based on the POD-based (proper orthogonal decomposition) ensemble four-dimensional variational data assimilation method (PODEn4DVar). The high flexibility and the high computational efficiency of the PODEn4DVar approach allow us to include both the atmospheric CO<sub>2</sub> concentrations and the surface CO<sub>2</sub> fluxes as part of the large state vector to be simultaneously estimated from assimilation of atmospheric CO<sub>2</sub> observations. Compared to most modern top-down flux inversion approaches, where only surface fluxes are considered as control variables, one major advantage of our joint data assimilation system is that, in principle, no assumption on perfect transport models is needed. In addition, the possibility for Tan-Tracker to use a complete dynamic model to consistently describe the time evolution of CO<sub>2</sub> surface fluxes (CFs) and the atmospheric CO<sub>2</sub> concentrations represents a better use of observation information for recycling the analyses at each assimilation step in order to improve the forecasts for the following assimilations. An experimental Tan-Tracker system has been built based on a complete augmented dynamical model, where (1) the surface atmosphere CO<sub>2</sub> exchanges are prescribed by using a persistent forecasting model for the scaling factors of the first-guess net CO<sub>2</sub> surface fluxes and (2) the at-

mospheric CO<sub>2</sub> transport is simulated by using the GEOS-Chem three-dimensional global chemistry transport model. Observing system simulation experiments (OSSEs) for assimilating synthetic in situ observations of surface CO<sub>2</sub> concentrations are carefully designed to evaluate the effectiveness of the Tan-Tracker system. In particular, detailed comparisons are made with its simplified version (referred to as TT-S) with only CFs taken as the prognostic variables. It is found that our Tan-Tracker system is capable of outperforming TT-S with higher assimilation precision for both CO<sub>2</sub> concentrations and CO<sub>2</sub> fluxes, mainly due to the simultaneous estimation of CO<sub>2</sub> concentrations and CFs in our Tan-Tracker data assimilation system. A experiment for assimilating the real dry-air column CO<sub>2</sub> retrievals (XCO<sub>2</sub>) from the Japanese Greenhouse Gases Observation Satellite (GOSAT) further demonstrates its potential wide applications.

## 1 Introduction

Carbon cycle data assimilation systems offer a promising new tool for CO<sub>2</sub> surface flux (CF) inversion (e.g., Peters et al., 2005; Feng et al., 2009), which tends to yield CO<sub>2</sub> surface flux estimates by optimally combining information

from both chemistry transport model (CTM) simulations and atmospheric CO<sub>2</sub> observations. Previous studies have helped to improve our understanding of the contemporary carbon cycle (e.g., David et al., 2006; Peters et al., 2007; Feng et al., 2011; Kang et al., 2012). The ensemble Kalman filter (referred to as EnKF) has been widely adopted in carbon cycle data assimilation (e.g., Peters et al., 2007; Feng et al., 2009, 2011; Kang et al., 2012; Liu et al., 2012), largely due to its simple conceptual formulation and relative ease of implementation (Evensen, 2003). Peters et al. (2005) coupled the state-of-the-art atmospheric transport TM5 model (<http://www.projects.science.uu.nl/tm5/>) to the ensemble square root filter (EnSRF), which forms the “CarbonTracker” data assimilation system, and its CF inversion results are fairly consistent with the majority of carbon inventories reported by the first North American State of the Carbon Cycle Report (SOCCR) (Peters et al., 2007). In CarbonTracker, a simple persistence forecasting operator is taken as the forecast model to represent the surface CO<sub>2</sub> flux propagation. This implies that the CFs (actually the scaling factors) are essentially treated as the model (i.e., the simple persistence forecasting operator) prognostic variables. Inclusion of a CF dynamical model in CarbonTracker meant that any useful information for CFs’ improvement achieved by the current data assimilation procedure could be used in the next assimilation cycle, so that the observed information would not be wasted. However, the uncertainty of the initial CO<sub>2</sub> concentration fields has been ignored in CarbonTracker. In fact, this uncertainty has such a large effect on CF estimates that neglecting this effect might result in unpredictable consequences (Bousquet et al., 2000; McKinley et al., 2004; Peylin et al., 2005). Recently, Kang et al. (2011, 2012) also presented a simultaneous data assimilation system of surface CO<sub>2</sub> fluxes and atmospheric CO<sub>2</sub> concentrations by means of the local ensemble transform Kalman filter (LETKF-CDAS). Here “LETKF-CDAS” means the LETKF (i.e., the local ensemble transform Kalman filter)-based carbon cycle data assimilation system (referred to as CDAS). In LETKF-CDAS, the CFs were also treated as part of the model states (as in Peters et al., 2005) and essentially a simple persistence dynamical model is adopted to describe the CFs’ integration. Similarly, Feng et al. (2009) also developed an ensemble Kalman filter to estimate 8-day CO<sub>2</sub> surface fluxes over geographical regions globally from satellite measurements of CO<sub>2</sub>.

The four-dimensional variational data assimilation (4D-Var) method has also been introduced in this field (e.g., Baker et al., 2006a; Engelen et al., 2009). Compared with EnKF, 4D-Var has its own attractive features: for example, it has the ability to simultaneously assimilate the observations at multiple times to the analysis fields (Tian and Xie, 2012). Nevertheless, the needs of the adjoint model and the linearization of the forecast model limit the wider applications of 4D-Var. Tian et al. (2008b, 2011) proposed the POD-based (proper orthogonal decomposition) ensemble four-dimensional variational data assimilation method (PODEn4DVar) based on

the POD and ensemble forecasting techniques, which aims to exploit the strengths of the two forms (i.e., EnKF and 4D-Var) of data assimilation while simultaneously offsetting their respective weaknesses. In PODEn4DVar, the control (state) variables in the 4D-Var cost function appear explicitly so that the adjoint model is no longer needed and the data assimilation process is significantly simplified (Tian et al., 2008). Furthermore, PODEn4DVar largely retains the basic advantages of the traditional 4D-Var. Its feasibility and effectiveness are demonstrated in an idealized model with simulated observations (Tian et al., 2011; Tian and Xie, 2012). It is found that the PODEn4DVar performs better than both 4D-Var and EnKF, and with lower computational costs than the EnKF (Tian et al., 2011). This method has been successfully applied to land data assimilation (Tian et al., 2009, 2010). Furthermore, we have built a PODEn3DVar (the three-dimensional version of PODEn4DVar)-based radar assimilation system on the atmospheric transport WRF model platform (Pan et al., 2012). This WRF-based data assimilation system indicates its (PODEn4DVar) potential in the atmospheric transport data assimilation.

In this study, we report on a new development of a CF data assimilation system based on the PODEn4DVar approach, named Tan-Tracker (in Chinese, “Tan” means carbon). This system is developed by incorporating a joint PODEn4DVar assimilation framework into the GEOS-Chem model (V9-01-03, <http://acmg.seas.harvard.edu/geos/>). We choose an identity operator as the CF dynamical model to describe the CFs’ evolution and then utilize such a CF dynamical model to constitute an augmented dynamical model together with the GEOS-Chem atmospheric transport model. Therefore in this case, the large-scale state vector made up of both the CFs and CO<sub>2</sub> concentrations is assumed to be the prognostic variable, which will be simultaneously constrained by assimilation of atmospheric CO<sub>2</sub> concentration observations.

In Sect. 2, we describe our Tan-Tracker data assimilation system, including the Tan-Tracker joint assimilation framework, a simple review of the PODEn4DVar assimilation approach and its coupling with the joint assimilation framework, and its covariance localization scheme. The following section (Sect. 3) shows observing system simulation experiments (OSSEs) for the evaluations of the Tan-Tracker system in comparison to its simplified version only taking CFs as the prognostic variables. Furthermore, another assimilation experiment for assimilation of real spaceborne CO<sub>2</sub> dry-air mole fraction observations (*X*CO<sub>2</sub>) indicates potential wider applications of this new proposed Tan-Tracker system (Sect. 4). Finally, a summary and concluding remarks are provided in Sect. 5.

## 2 The Tan-Tracker joint data assimilation system

Joint or dual-pass assimilation schemes have been utilized to optimize model states and parameters simultaneously from



ilation cycle

$$F_b^s(t) = \lambda_{b,r} F_s^*(t), \quad (3)$$

where  $t = 1, \dots, L_s$ ;  $L_s (= L_{\text{Pre}} + L_a + L_{\text{Pos}})$  is the length of the sampling window; and  $L_{\text{Pre}}$  and  $L_{\text{Pos}}$  are the lengths of the Pre-Assim and Post-Assim windows, respectively (see Fig. 2), over the sampling window to yield the sampling CO<sub>2</sub> concentration series  $U_i^s$  ( $i = 1, \dots, L_s$ , and  $U_1^s = U_b^s$ ). Next, a 4-D moving sampling strategy (Fig. 2; Wang et al., 2010) is adopted to create the large-scale vector ensemble  $(\lambda_{m,i}, U_{m,i})^T$  ( $i = 1, \dots, N$ ,  $N = L_s - L_a + 1$ ) as follows:

$$(\lambda_{m,i}, U_{m,i})^T = \mathbf{X}_i^s = \begin{bmatrix} \frac{F_b^s(i)}{F_a^*(1)} \\ \vdots \\ \frac{F_b^s(i+L_a-1)}{F_a^*(L_a)} \\ U_i^s \\ \vdots \\ U_{i+L_a-1}^s \end{bmatrix}. \quad (4)$$

As a result the large-scale joint state vector  $(\lambda, U)^T$  is viewed as the prognostic variable in Tan-Tracker, with the identity operator (4) chosen to be the CF dynamical sub-model to describe the CFs' evolution:

$$M_{\text{CF}} = \mathbf{I}, \quad (5)$$

where  $\mathbf{I}$  is the identity matrix. This CF persistence forecasting model (4) follows Peters et al. (2005) and assumes that the prior (or background) scaling factors  $\lambda_{b,r+1}$  for the next assimilation cycle [( $r+1$ )th] are equal to the optimized scaling factors  $\lambda_{a,r}$  of the current ( $r$ )th assimilation cycle. In the actual implementations, the following dynamical model (5) is applied to the linear scaling factors,  $\lambda$

$$\lambda_{b,r+1} = \frac{1}{L_o} \sum_{j=1}^{L_o} \lambda_{a,r}^j, \quad (6)$$

where  $L_o$  is the length of the optimized window (Fig. 2) and  $\lambda_{a,r}^j$  are the daily optimized scaling factors  $\lambda_{a,j}$  ( $j = 1, \dots, L_o$ ). The CF dynamical sub-model  $M_{\text{CF}}$  is thus utilized to constitute the augmented dynamical model

$$M = \begin{pmatrix} \mathbf{I} \\ \text{CTM} \end{pmatrix} \quad (7)$$

for Tan-Tracker together with the CTM (GEOS-Chem) model. By applying the observation operator  $H$  to the modeled CO<sub>2</sub> concentrations  $U_{m,i}$  and the background CO<sub>2</sub> concentrations  $U_b$ , we can obtain the ensemble simulated observations  $U_m^o$  and the background simulated observations  $U_b^o$  as follows:

$$U_{m,i}^o = H(U_{m,i}) \quad (8)$$

and

$$U_b^o = H(U_b). \quad (9)$$

So far, the background joint vector  $(\lambda_b, U_b)^T$ , the joint vector ensemble  $(\lambda_{m,i}, U_{m,i})^T$ , Eqs. (8) and (9) and the real CO<sub>2</sub> measurements  $U_b^o$  would be input to the PODEn4DVar assimilation processor, which yields the assimilated  $(\lambda_a, U_a)^T$  and the optimized CFs  $F_a = \lambda_a F^*$  as a result.

In conclusion, Tan-Tracker works as follows: two CTM runs forced by the background CFs' series are firstly achieved over the assimilation window and the sampling window, respectively: the background run is used to prepare the background joint vector, and the sampling run is used to produce the joint vector ensemble by applying a 4-D moving strategy (Wang et al., 2010) to the sampling simulations throughout the sampling window. The background joint vector and the joint vector ensemble are then input into the PODEn4DVar processor, in which the usual observation operator (e.g., the interpolation function to interpolate the model gridded variables to the in situ observations) compares the simulated CO<sub>2</sub> concentrations with the observed according to the 4D-Var cost function: the CO<sub>2</sub> concentrations are assimilated to initialize the next assimilation cycle. Meanwhile, the scaling factors  $\lambda$  in the optimized window are also optimized and used for the next assimilation cycle through Eq. (5).

## 2.2 The PODEn4DVar and its coupling with the joint assimilation framework

The PODEn4DVar approach is born out of the incremental format of the 4D-Var cost function

$$J(\mathbf{x}') = \frac{1}{2} (\mathbf{x}')^T \mathbf{B}^{-1} (\mathbf{x}') + \frac{1}{2} [y'(\mathbf{x}') - y'_{\text{obs}}]^T \mathbf{R}^{-1} [y'(\mathbf{x}') - y'_{\text{obs}}], \quad (10)$$

where  $\mathbf{x}' = \mathbf{x} - \mathbf{x}_b$  is the perturbation of the background field  $\mathbf{x}_b$  at the initial time  $t_0$ ,

$$\mathbf{y}'_{\text{obs}} = \begin{bmatrix} y'_{\text{obs},1} \\ y'_{\text{obs},2} \\ \vdots \\ y'_{\text{obs},S} \end{bmatrix}, \quad (11)$$

$$\mathbf{y}' = y'(\mathbf{x}') = \begin{bmatrix} (\mathbf{y}_1)' \\ (\mathbf{y}_2)' \\ \vdots \\ (\mathbf{y}_S)' \end{bmatrix}, \quad (12)$$

$$(\mathbf{y}_k)' = y_k(\mathbf{x}_b + \mathbf{x}') - y_k(\mathbf{x}_b), \quad (13)$$

$$\mathbf{y}'_{\text{obs},k} = y_{\text{obs},k} - y_k(\mathbf{x}_b), \quad (14)$$

$$y_k = H_k(M_{t_0 \rightarrow t_k}(\mathbf{x})), \quad (15)$$

and

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}_S \end{bmatrix}. \quad (16)$$

Here index  $k$  denotes the observation time; the superscript T stands for a transpose;  $b$  represents background values;  $S$  is the total observational time steps in the observational window;  $H_k$  acts as the observation operator; and matrices  $\mathbf{R}_k$  and  $\mathbf{B}$  are the observational and background error covariances, respectively.

With the prepared background field  $\mathbf{x}_b$ , the initial model perturbations (MPs)  $\mathbf{x}'(\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_N)$ , the simulated observation perturbations  $\mathbf{y}'(\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_N)$ , the observational increments  $\mathbf{y}'_{\text{obs},k}$ , and the background and observational error covariances  $\mathbf{B}$  and  $\mathbf{R}_k$ , the final PODEn4DVar analysis solution  $\mathbf{x}_a$  without localization of its analysis error covariance  $P^a$  is formulated through some necessary calculations (see Tian et al., 2010, 2011, for more details) as

$$\mathbf{x}_a = \quad (17a)$$

$$\mathbf{x}_b + \mathbf{x}'\mathbf{V}[(N-1)\mathbf{I} + \mathbf{P}_y^T\mathbf{R}^{-1}\mathbf{P}_y]^{-1}\mathbf{P}_y^T\mathbf{R}^{-1}\mathbf{y}'_{\text{obs}}$$

and

$$P^a = P_x P_a^* P_x^T, \quad (17b)$$

where  $P_a^* = [(N-1)\mathbf{I} + \mathbf{P}_y^T\mathbf{R}^{-1}\mathbf{P}_y]^{-1}$  and  $\mathbf{V}$  is derivable from  $(\mathbf{y}')^T\mathbf{y}' = \mathbf{V}\Lambda^2\mathbf{V}^T$  and  $\mathbf{P}_y = \mathbf{y}'\mathbf{V}$ . To clarify, the background covariance  $\mathbf{B}$  is approximately estimated by  $\mathbf{B} = \frac{\mathbf{P}_x\mathbf{P}_x^T}{N-1}(\mathbf{P}_x = \mathbf{x}'\mathbf{V})$  in formulating PODEn4DVar.

In particular, in Tan-Tracker,

$$\mathbf{y}'_{\text{obs},k} = U_{o,k} - U_b^o \quad (18)$$

and

$$\mathbf{y}' = U_m^o - U_b^o, \quad (19)$$

where  $U_b^o = H(U_b)$ . Here we mark

$$H = \begin{bmatrix} H_1 & 0 & \cdots & 0 \\ 0 & H_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_S \end{bmatrix}. \quad (20)$$

As mentioned, the model state to be optimized is the joint vector  $(\lambda, U)^T$ , which indicates

$$\mathbf{x}_b = (\lambda_b, U_b)^T \quad (21)$$

and

$$\mathbf{x}' = (\lambda_m, U_m)^T - (\lambda_b, U_b)^T \quad (22)$$

in Tan-Tracker.

We have realized the coupling between the joint assimilation framework with the PODEn4DVar assimilation processor through Eqs. (18–22) (see the green part of Fig. 1).

### 2.3 Covariance localization

As an ensemble-based assimilation system, Tan-Tracker also utilizes the covariance localization techniques to ameliorate the contaminations resulting from the spurious long-range correlations (Houtekamer and Mitchell, 2001). It uses the following exponential decay of the covariance structure with distance between state and observational variables (Gaspari and Cohn, 1999),

$$\rho_h[i, j] = e^{-d_{i,j}/d_0}, \quad (23)$$

to calculate the elements  $\rho_h[i, j]$  of the matrix  $\rho_h[L_x \times L_y]$ , where  $L_x$  and  $L_y$  are the lengths of the state vector  $\mathbf{x}$  and the observational vector  $\mathbf{y}$ , respectively;  $d_{i,j}$  is the distance between the  $i$ th state and the  $j$ th observation locations and  $d_0$  is the horizontal covariance localization Schur radius.

Consequently, the covariance localization in Tan-Tracker can be implemented by calculating the Schur product  $\circ$  (i.e., piecewise multiplication) as follows (Greybush et al., 2011):

$$\mathbf{x}_a = \quad (24)$$

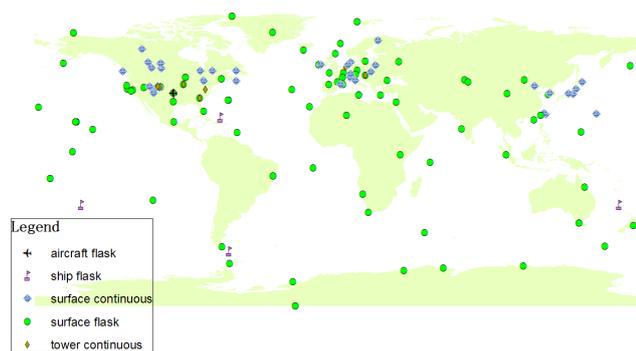
$$\mathbf{x}_b + \rho_h \circ \left\{ \mathbf{x}'\mathbf{V}[(N-1)\mathbf{I} + \mathbf{P}_y^T\mathbf{R}^{-1}\mathbf{P}_y]^{-1}\mathbf{P}_y^T\mathbf{R}^{-1} \right\} \mathbf{y}'_{\text{obs}}.$$

## 3 OSSEs for the evaluations of Tan-Tracker

In this section, Tan-Tracker will be comprehensively evaluated through a group of well-designed global observing system simulation experiments (OSSEs) over a given assimilation period.

### 3.1 Experimental setup

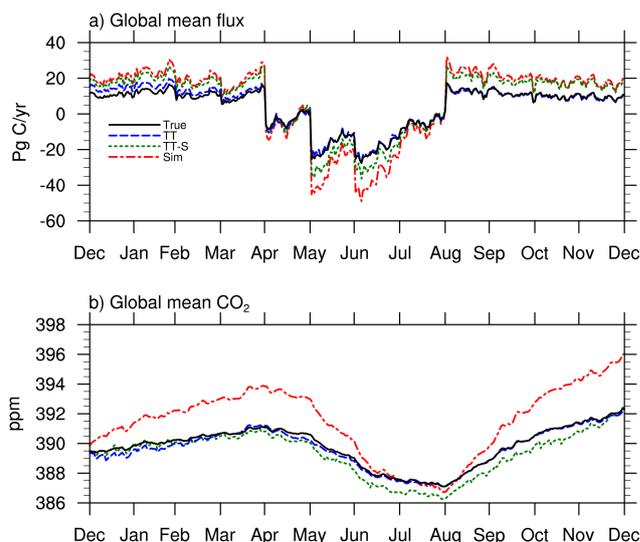
We simulate atmospheric CO<sub>2</sub> concentrations using the global three-dimensional chemical transport model GEOS-Chem (version 9-01-03, <http://acmg.seas.harvard.edu/geos/>) driven by the assimilated meteorological data from the Goddard Earth Observing System (GEOS) of the NASA Global Modeling and Assimilation Office. The version of the model we use is driven by the GEOS-5 meteorological fields with a horizontal resolution of 2° latitude by 2.5° longitude and 47 vertical layers up to 0.01hPa. The original GEOS-Chem CO<sub>2</sub> simulation was described in Suntharalingam et al. (2004) and updated by Nassar et al. (2010). Our simulations include CO<sub>2</sub> fluxes from monthly fossil fuel burning and cement production CO<sub>2</sub> emissions from the Carbon Dioxide Information Analysis Center (CDIAC) inventory for year 2009 (Andres et al., 2010), monthly biomass burning from the third version of the Global Fire Emission Database (GFEDv3) for 2010 (van der Werf et al., 2010; Mu et al., 2011), climatological biofuel burning (Yevich and Logan, 2003), monthly ocean exchange (Takahashi et al., 2009), 3-hourly biospheric fluxes from the Carnegie–Ames–Stanford Approach (CASA) model for 2000 (Olsen and Randerson, 2004), annual climatology terrestrial biosphere exchange based on TransCom



**Figure 3.** The observational sites used in this study.

CO<sub>2</sub> inversion results adjusted with GFEDv2 fire emissions (Baker et al., 2006b; van der Werf et al., 2006), the chemical production of CO<sub>2</sub> from the atmospheric oxidation of other carbon species (Nassar et al., 2010), the monthly emissions from shipping (Olivier and Berdowski, 2001), and aviation CO<sub>2</sub> emissions (Friedl, 1997; Sausen and Schuman, 2000; Kim et al., 2005, 2007; Wilkerson et al., 2010). For this work, our model simulation was initialized on 01 January 2008 with a globally uniform 3-D CO<sub>2</sub> field of 383.76 ppm. According to the record of NOAA-ESRL Mauna Loa Observatory in Hawaii (<http://www.esrl.noaa.gov/gmd/ccgg/>), which is a marine surface site, the annual mean CO<sub>2</sub> at Mauna Loa in 2007 was 383.76 ppm, with monthly means of 383.89 ppm in December 2007 and 385.44 ppm in January 2008. A 2-year spin-up simulation from this initialized state allows for model transport, sources and sinks to develop the global spatial patterns of CO<sub>2</sub>; this approach was evaluated in Nassar et al. (2010). After the spin-up run, the obtained CO<sub>2</sub> fields were used to drive the observing system simulation experiments. In all the following OSSEs, we firstly assume the default surface CO<sub>2</sub> fluxes released with the GEOS-Chem model as the true CF series  $F_{\text{True}}$ . Then we run the GEOS-Chem model, driven by the true CF series  $F_{\text{True}}$ , to obtain the true CO<sub>2</sub> concentration results from 1 January 2010 to 31 December 2010 (i.e., the assimilation period). The artificial CO<sub>2</sub> observations are thus generated every day by sampling the daily true CO<sub>2</sub> concentrations through adding small random noise (whose error variance is 0.01 ppm<sup>2</sup>) through the 136 observational sites used in this study (Fig. 3). The first-guess CF series  $F^*$  are set to  $1.8F_{\text{True}}$ , which drives the GEOS-Chem model at the same resolution (2° latitude × 2.5° longitude) to produce the background CO<sub>2</sub> simulations from the spun-up equilibrium state.

The performance of our Tan-Tracker system is examined by comparison with the simplified version (referred to as TT-S), taking only CFs as the prognostic variables. TT-S is somewhat similar to CarbonTracker except that the ensemble square root filter (EnSRF) has been replaced by the PODen4DVar approach and the GEOS-Chem model is used instead of the TM5 model. Similar to CarbonTracker, the

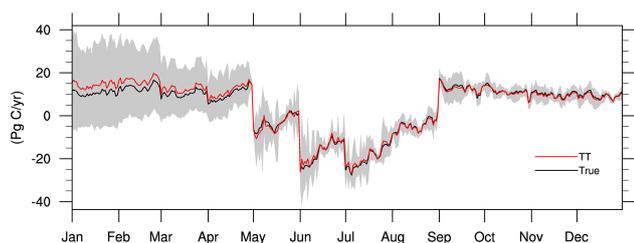


**Figure 4.** Time series of the global mean (a) CO<sub>2</sub> surface fluxes and (b) CO<sub>2</sub> concentrations from the “truth”, simulations, TT-S (the simplified version of Tan-Tracker) and TT (Tan-Tracker) assimilations from 1 January to 31 December 2010.

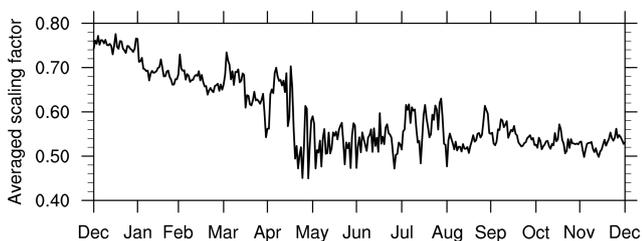
GEOS-Chem model in TT-S is actually the observation operator linking the CFs with CO<sub>2</sub> observations. In TT-S, since the CO<sub>2</sub> concentrations are not assimilated together with the CFs, we first obtain the optimized scaling factors through assimilating CO<sub>2</sub> observations, and thus the CO<sub>2</sub> concentrations are also updated by the GEOS-Chem modeling forced by the optimized CFs. All the assimilation processes are initiated by the GEOS-Chem model with first-guess CF series  $F^*$  ( $= 1.8F_{\text{True}}$ ) and conducted continuously by assimilating the daily pseudo-observations throughout the assimilation period. In all the assimilation experiments, the scaling factors are initiated from  $\lambda_{b,0}(i, j) = 1.0$  (where  $i$  and  $j$  are the longitude and latitude indexes, respectively, and 0 denotes the  $r$ th ( $= 0$ ) assimilation cycle). In all the OSSEs, the default lag window is 5 weeks, and the observational window and optimized window are both 1 week. The reference ensemble size  $N$  is 106 and the standard localization radius  $d_0$  is 900 km. Changes in the assimilation parameters might influence the assimilation performance. We further investigate the effects of the length of the horizontal localization Schur radius and the ensemble size in Tan-Tracker by means of several sensitivity numerical experiments, the results of which are presented in Sect. 3.2. In all assimilation experiments, we use the adaptive inflation technique proposed by Li et al. (2009).

### 3.2 Experimental results

To evaluate Tan-Tracker’s performance in a general view, time series of the daily global mean fluxes and CO<sub>2</sub> concentrations from the background simulations, the TT-S and the TT (Tan-Tracker) assimilations are compared with the true

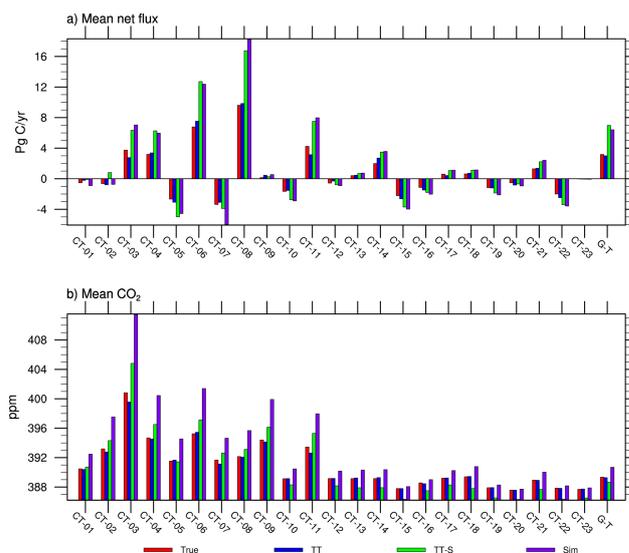


**Figure 5.** Time series of the posterior uncertainties (shaded areas) of the analyzed surface fluxes (TT) from 1 January to 31 December 2010.



**Figure 6.** Time series of the averaged scaling factors from 1 January to 31 December 2010.

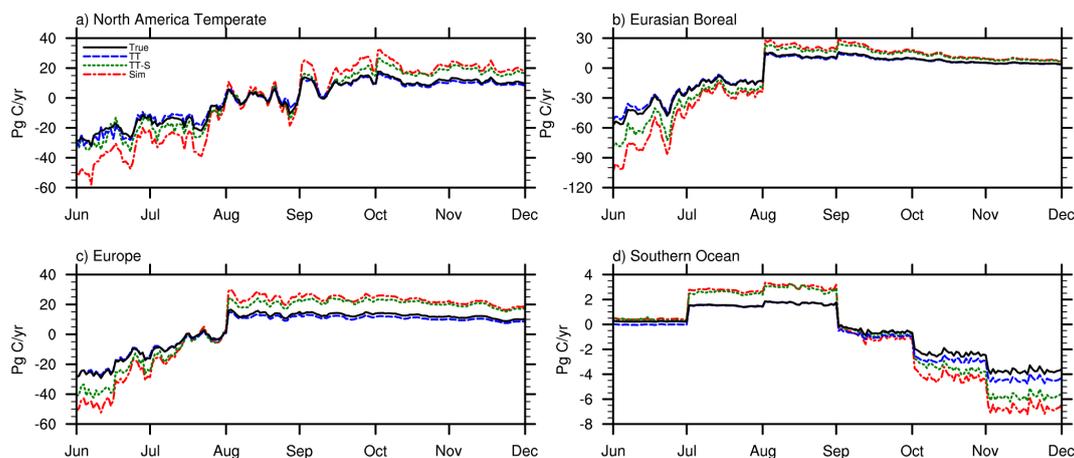
simulations in Fig. 4. Not surprisingly, the background simulations (referred to as Sim) will inevitably deviate seriously from the “true” simulations due to the predetermined background CF series  $F_b (= 1.8F_{\text{True}})$ . Remarkably, since both the  $\text{CO}_2$  concentrations and CFs are simultaneously assimilated under the joint assimilation framework, it could largely eliminate the uncertainty of the initial  $\text{CO}_2$  concentrations on the  $\text{CO}_2$  evolution during the assimilation window and maximize the observations’ potential. Probably for this reason, Fig. 4 shows that Tan-Tracker works very well throughout the whole assimilation period, especially after the first few months, which can be considered a spin-up period. However, the performance of TT-S is not very robust and its assimilated errors do not show a trend of becoming less even though its performance seems to be substantially better than the background simulation case: obviously, the impacts of the  $\text{CO}_2$  concentration have not been taken into full consideration in the TT-S system and there must be some non-negligible errors remaining in the TT-S-optimized  $\text{CO}_2$  concentrations (Fig. 4b). The resulting errors in the initial  $\text{CO}_2$  concentrations will in turn contaminate the TT-S assimilation of  $\text{CO}_2$  fluxes for the next assimilation cycle. In the following discussions, we focus on the results only during the latter half of the year 2010 and thus remove the spin-up period occurring in the first half of the year. Figure 5 also shows that the posterior uncertainties of the analyzed CFs are gradually decreased with assimilation of  $\text{CO}_2$  observations. Furthermore, Fig. 6 shows time series of the daily globally averaged scaling factor. The daily averaged scaling factor is also decreased



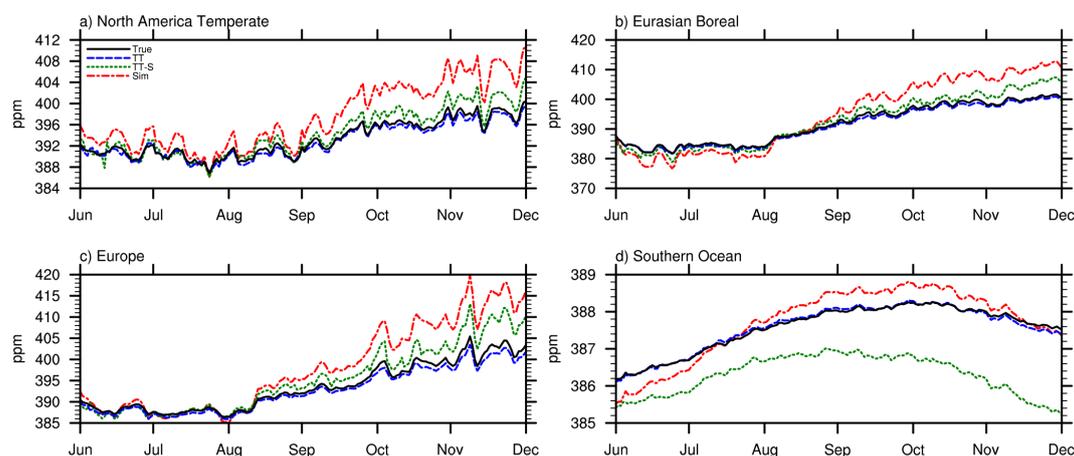
**Figure 7.** (a) Mean  $\text{CO}_2$  surface fluxes and (b)  $\text{CO}_2$  concentration from the “truth”, simulations, TT-S (the simplified version of Tan-Tracker) and TT (Tan-Tracker) assimilations aggregated to TransCom regions (i.e., CT-01: North America Boreal; CT-02: North America Temperate; CT-03: South America Tropical; CT-04: South America Temperate; CT-05: Northern Africa; CT-06: Southern Africa; CT-07: Eurasia Boreal; CT-08: Eurasia Temperate; CT-09: Tropical Asia; CT-10: Australia; CT-11: Europe; CT-12: North Pacific Temperate; CT-13: West Pacific Tropical; CT-14: East Pacific Tropical; CT-15: South Pacific Temperate; CT-16: Northern Ocean; CT-17: North Atlantic Temperate; CT-18: Atlantic Tropics; CT-19: South Atlantic Temperate; CT-20: Southern Ocean; CT-21: India Tropical; CT-22: South India Tropical; CT-23: Zero Flux Regions; G-T: Global Total) during the period from 1 June to 31 December 2010.

and becomes close to  $\sim 0.56$  (i.e.,  $1/1.8$ ) with small fluctuations during the latter half of the year 2010.

Similar to Peters et al. (2005), we also aggregated the daily, gridded ( $2^\circ$  latitude  $\times$   $2.5^\circ$  longitude) simulation and assimilation results to 24 “super-regions” corresponding to the TransCom 3 regions given by Gurney et al. (2002). Figure 7 shows the 24 super-regions’ aggregated mean  $\text{CO}_2$  concentrations and fluxes during the latter half of the year 2010. Generally, Tan-Tracker is able to reproduce the true fluxes well and its superiority dominates most of the 24 super-regions except for 3 – CT-09 (Tropical Asia), CT-12 (North Pacific Temperate) and CT-20 (Southern Ocean) – whose absolute values are very small (Fig. 7a). Furthermore, as far as the  $\text{CO}_2$  concentration is concerned, the superior performance of Tan-Tracker beyond TT-S is increasingly obvious (Fig. 7b): the differences between the “truth” and the TT-assimilated  $\text{CO}_2$  concentrations are much less than those between the TT-S-assimilated and the “truth” in the overwhelming majority of cases, which illustrates once more that the simultaneous assimilation of  $\text{CO}_2$  concentrations and CFs is indispensable. The time series of daily mean



**Figure 8.** Time series of the daily mean CO<sub>2</sub> surface fluxes from the “truth”, simulations, TT-S (the simplified version of Tan-Tracker) and TT (Tan-Tracker) assimilations aggregated to the selected four TransCom regions (i.e., CT-02, CT-07, CT-11 and CT-20) during the period from 1 July to 31 December 2010.

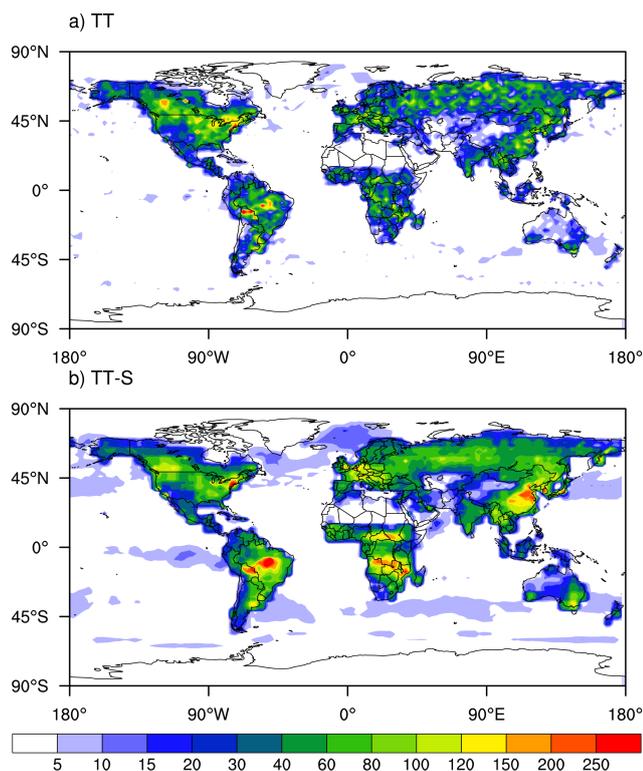


**Figure 9.** Same as Fig. 8 but for CO<sub>2</sub> concentrations.

fluxes and CO<sub>2</sub> concentrations from the four selected super-regions (Temperate North America, Europe, Boreal Eurasia, and Southern Ocean) are shown in Figs. 8 and 9. Similar to the global mean case shown in Fig. 3, the ability of our assimilation system to represent the variations of seasonal peak-to-trough amplitudes of CO<sub>2</sub> concentrations and fluxes is expressed thoroughly and demonstrates its power to make full use of the observations. Comparatively speaking, the ability of the TT-S system is considerably inferior to Tan-Tracker, especially in the Southern Ocean super-region during October–December, 2010: here the TT-S-optimized CO<sub>2</sub> concentrations are even worse than the background simulations (Fig. 9d).

To evaluate the performance of our Tan-Tracker data assimilations system comprehensively, we show the root-mean-square errors (RMSEs) for the daily, gridded (2° latitude × 2.5° longitude) TT- and TT-S-assimilated fluxes from 1 July to 31 December 2010 in Fig. 10. In

addition, their corresponding RMSEs for the assimilated (optimized) CO<sub>2</sub> concentrations are also shown in Fig. 11. Compared with the Tan-Tracker case, larger RMSEs ( $> 300 \times 10^{-11} \text{ kg C m}^{-2} \text{ s}^{-1}$ ) for the TT-S-assimilated fluxes can be found in the central parts of South America, most of East Asia, and southern Africa (Fig. 10b). Encouragingly, the TT-assimilated flux RMSEs are largely kept at a very low level ( $\leq 80 \times 10^{-11} \text{ kg C m}^{-2} \text{ s}^{-1}$ ), in which relatively larger RMSEs (but still much less than the TT-S-assimilated) appear only in a very small area in the central parts of South America (Fig. 10a). Naturally, a parallel circumstance is also replayed in the CO<sub>2</sub> concentration case (Fig. 11). Evidently, a relatively definite conclusion can be drawn that the uncertainty of the initial CO<sub>2</sub> concentrations cannot be ignored and the joint assimilation framework contributes a lot to the final Tan-Tracker performance. Moreover, the application of the advanced hybrid assimilation approach (i.e., PODen4DVar) would definitely make a positive con-

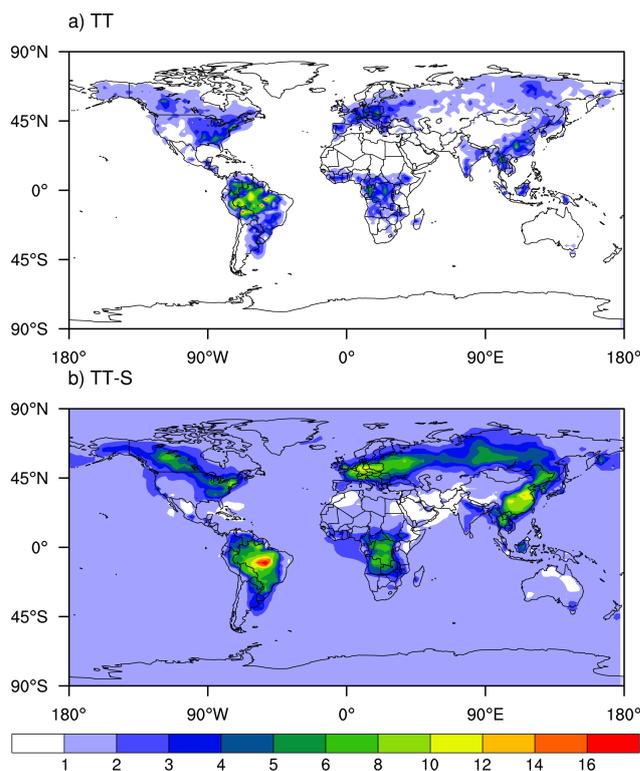


**Figure 10.** Root-mean-square errors (RMSEs) (units are  $10^{-11} \text{ kg C m}^{-2} \text{ s}^{-1}$ ) for the daily (a) TT- and (b) TT-S-assimilated  $\text{CO}_2$  surface fluxes during the period from 1 July to 31 December 2010.

tribution to its excellent performance (Tian et al., 2011). Of course, the imbalance of CFs and  $\text{CO}_2$  concentrations in TT-S partly explains its inferior performance.

Another group of experiments using the Tan-Tracker system with different horizontal localization radii ( $d_0 = 100, 900, 1450, 2000$  and  $5000 \text{ km}$ ) are also conducted to explore the sensitivity of our Tan-Tracker assimilation system to the variations of the horizontal radius. As suggested by Peters et al. (2005), we take  $900 \text{ km}$  as the default or reference radius. Figure 12 shows time series of the daily global  $\text{CO}_2$  concentrations and fluxes from the “truth” as well as the TT assimilations using the three different horizontal localization radii ( $d_0 = 900, 1450$  and  $2000 \text{ km}$ ). Therefore, we can roughly judge that the Tan-Tracker system could perform well with its horizontal localization radius around  $900 \text{ km}$ . Nevertheless, two extremely inappropriate localization radii ( $d_0 = 100$  and  $5000 \text{ km}$ ) are also tested in our experiments (but not shown here), whose poor performance demonstrates that the choice of an appropriate covariance localization radius is essential to Tan-Tracker’s successful implementation.

Finally, to investigate the impacts of sample sizes on Tan-Tracker’s assimilation results, we also conduct another group of Tan-Tracker assimilation experiments with the ensemble numbers  $N = 60, 106$  and  $150$ . Figure 13 shows that the



**Figure 11.** Same as Fig. 10 but for  $\text{CO}_2$  concentrations (units are ppm).

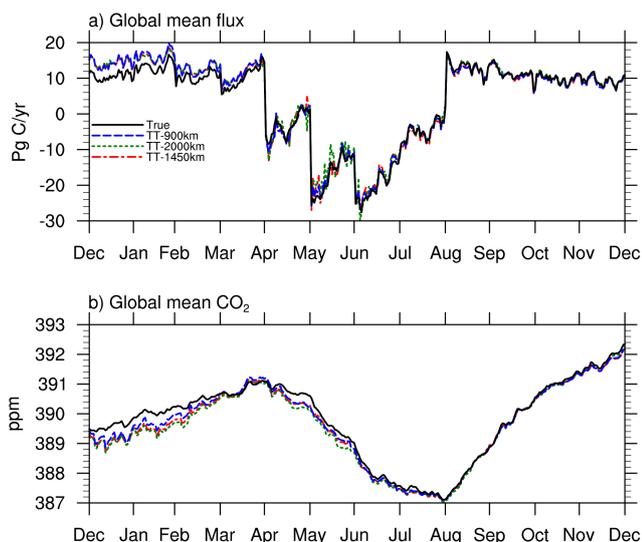
differences between the two assimilation experiments with  $N = 106$  and  $150$  are very small. However, if we decrease the ensemble number to  $60$  (not shown), the assimilation results become divergent. Synthesizing the above results, we can conclude that giving a certain number of sample sizes ( $\geq 100$ ) could generally guarantee the robust performance of our system.

#### 4 Real-data assimilation experiment with spaceborne observations

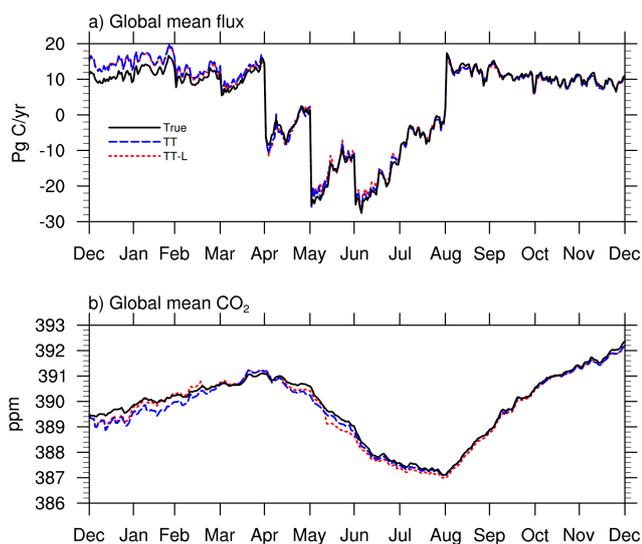
In this section, a preliminary real assimilation experiment is conducted by using spaceborne  $\text{CO}_2$  dry-air mole fraction observations to illustrate the potential applications of Tan-Tracker in real-data assimilation.

##### 4.1 Experimental setup

The basic experimental designs (such as the GEOS-Chem model, ensemble size, assimilation window, localization radius, etc.) are exactly the same as those adopted in Sect. 3. Nevertheless, in this real-data experiment, we took the default surface  $\text{CO}_2$  fluxes released with the GEOS-Chem model as the first-guess CF series  $F^*$  and used spaceborne  $\text{CO}_2$  dry-air mole fraction observations ( $X_{\text{CO}_2}$ ) instead of artificial  $\text{CO}_2$  observations. The spaceborne ob-

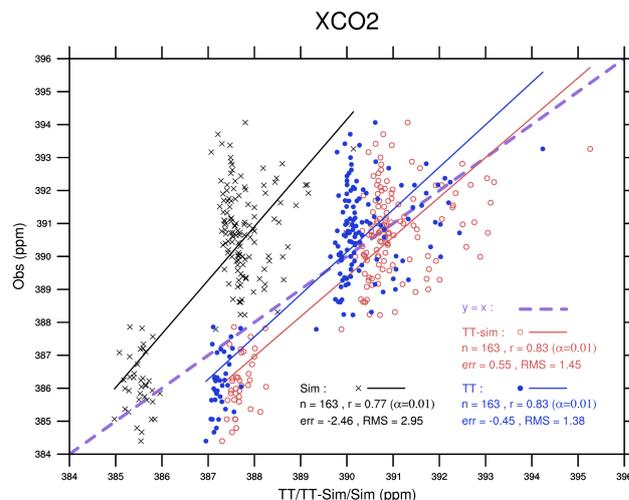


**Figure 12.** Time series of the daily global mean (a) CO<sub>2</sub> surface fluxes and (b) CO<sub>2</sub> concentrations from the “truth” and the TT (Tan-Tracker) assimilations using different covariance localization radii (900, 1450 and 2000 km), respectively, from 1 January to 31 December 2010.



**Figure 13.** Time series of the daily global mean (a) CO<sub>2</sub> surface fluxes and (b) CO<sub>2</sub> concentrations from the “true” and the TT (Tan-Tracker) assimilations with ensemble numbers  $N = 106$  and  $150$ , respectively, from 1 January to 31 December 2010.

observations used here are originated from the Japanese Greenhouse Gases Observing Satellite (GOSAT), which was launched into orbit in 2009. TANSO-FTS, onboard GOSAT, operates in the shortwave infrared band (SWIR) between 758 and 2080 nm and thermal infrared band (TIR) from 5.56 to 14.3  $\mu\text{m}$ , providing information on CO<sub>2</sub> and CH<sub>4</sub> in the atmosphere. Level 2 data or the so-called the column-averaged CO<sub>2</sub> dry-air mole fraction  $X_{\text{CO}_2}$  is



**Figure 14.** Comparisons between the observed  $X_{\text{CO}_2}$  and the open-loop GEOS-Chem-simulated (Sim), Tan-Tracker-assimilated (TT) and the TT-Sim (i.e., the GEOS-Chem model run without data assimilation forced by the TT-optimized CF series derived from the Tan-Tracker assimilation run with the TT-assimilated initial CO<sub>2</sub> fields at 1 January 2010) simulated  $X_{\text{CO}_2}$  on 15 March 2010.

taken from version 3.3 atmospheric CO<sub>2</sub> observations from space (ACOS) data product (O’Dell et al., 2012). Validation against ground-based TCCON data shows a mean bias less than 1.4 ppm; these biases can be further reduced by applying the recommended data screening criteria and bias correction technique (for more details please refer to the document “ACOS Level 2 Standard Product Data User’s Guide”, [http://disc.sci.gsfc.nasa.gov/acdisc/documentation/ACOS\\_v3.3\\_DataUsersGuide.pdf](http://disc.sci.gsfc.nasa.gov/acdisc/documentation/ACOS_v3.3_DataUsersGuide.pdf)). Furthermore, to guarantee the high quality of the assimilated data as much as possible, we discarded the  $X_{\text{CO}_2}$  data with observation errors  $\geq 0.75$  ppm.

In order to assimilate the spaceborne  $X_{\text{CO}_2}$  directly, the following observation operator (Eq. 25) needs to be incorporated into Tan-Tracker to provide a link between the observational variable  $X_{\text{CO}_2}$  and the GEOS-Chem-simulated CO<sub>2</sub> concentrations (Feng et al., 2009):

$$X_{\text{CO}_2} = X_{\text{CO}_2,a} + h^T A (u_m - u_a), \quad (25)$$

where  $h$  is the pressure weighting function;  $A$  is the full averaging kernel matrix;  $U_a$  and  $X_{\text{CO}_2,a}$  are the prior CO<sub>2</sub> profile and the associated column amount, respectively; and  $u_m$  is the GEOS-Chem-produced CO<sub>2</sub> profile. The experiment period is from 1 January 2010 to 31 March 2010. In particular, we chose one arbitrary day’s (15 March 2010 in this experiment)  $X_{\text{CO}_2}$  data as the evaluation data set, which are designedly not assimilated in the experiments to provide an independent evaluation for the Tan-Tracker system.

## 4.2 Experimental results

The lack of reliable independent CF estimates derived from GOSAT  $XCO_2$  retrievals (Chevallier et al. 2014) forces us to seek an indirect way to evaluate the Tan-Tracker assimilations. Here, we performed a parallel free run of GEOS-Chem forward simulation without any data assimilation. Then, to examine Tan-Tracker's performance quantitatively, the simulated and assimilated  $CO_2$  dry-air mole fraction observations of  $XCO_2$  on 15 March 2010 were compared with the corresponding (independent) GOSAT observations. After the data quality control (observation error  $< 0.75$  ppm) implemented in this experiment, there are still 163 valid footprints left for system evaluation. Compared with the Sim case, the TT-assimilated  $XCO_2$  is improved considerably with higher correlation (0.83 vs. 0.77) and a smaller RMSE (1.38 ppm vs. 2.95 ppm). The GEOS-Chem model generally underestimates the  $XCO_2$  values by a substantial negative bias of  $-2.46$  ppm, where the mean bias is given by 
$$\text{err} = \frac{1}{163} \sum_{i=1}^{163} (XCO_2, i^{s(a)} - XCO_2, i^o),$$
 with  $XCO_2, i^{s(a)}$  and  $XCO_2, i^o$  being the simulated (assimilated) and observed  $XCO_2$  values for each valid footprint, respectively. However, the TT-assimilated case only has a very small bias ( $\text{err} = -0.45$  ppm). Obviously, the above discussions could only demonstrate that our Tan-Tracker system is capable of yielding fairly good  $CO_2$  concentration results. It is encouraging to find that the performance of the TT-Sim case is slightly inferior to the TT case (RMSE = 1.45 ppm and  $r = 0.83$ ), suggesting that Tan-Tracker does enhance the  $CO_2$  concentration and flux estimations. It provides a promising new tool for  $CO_2$  surface flux (CF) inversion. In addition, in Fig. 14,  $\alpha$  (0.01) is the confidence coefficient. Certainly, extra efforts should be made to give a more detailed assessment for Tan-Tracker satellite data assimilation, which will be provided in another study.

## 5 Summary and concluding remarks

In this study, a new carbon cycle data assimilation system (i.e., Tan-Tracker) is developed based on an advanced hybrid assimilation approach (PODEn4DVar), as a part of the preparation for the launch of the Chinese carbon dioxide observation satellite (TanSat) (Liu et al., 2012; Cai et al., 2014). Tan-Tracker adopts a joint data assimilation framework: a simple persistence model is chosen to describe the CFs' evolution, which acts as the CF dynamical sub-model and constitutes an augmented dynamical model together with the GEOS-Chem atmospheric transport model. In such an augmented dynamical model, the large-scale state vector made up of CFs and  $CO_2$  concentrations is actually the prognostic variable, which is designed to be simultaneously constrained by the observations of atmospheric  $CO_2$  concentrations. As a step towards the application of Tan-Tracker, we carefully designed several

groups of observing system simulation experiments (OSSEs) to comprehensively evaluate Tan-Tracker's performance in comparison to its simplified version (TT-S), taking only CFs as the prognostic variables. It is found that the simultaneous estimation of  $CO_2$  concentrations and CFs plays a vital role in enhancing the Tan-Tracker system's performance: contamination in Tan-Tracker's performance in CF estimation from the uncertainty in the  $CO_2$  concentration evolution has been gradually reduced through continuously fitting model  $CO_2$  concentration simulations to the observations.

Our future work will focus on the realization of  $XCO_2$  assimilation in the first version of Tan-Tracker, which is a key step to extending Tan-Tracker with functions for assimilating satellite measurements. This goal could be achieved by introducing the observation operator to link the  $CO_2$  concentration profiles with  $XCO_2$ . As the Chinese TanSat has not yet been launched, we will focus our proposed Tan-Tracker on GOSAT and OCO-2 (O'Dell et al., 2012) measurements of  $CO_2$ . Encouragingly, a preliminary real-data assimilation experiment conducted by using spaceborne (GOSAT) observations demonstrates its potential wider applications.

*Acknowledgements.* We would like to acknowledge Anemarie Fraser, Wouter Peters and Ross Bannister for constructive comments on the manuscript. The two anonymous reviewers are thanked for their critical comments and suggestions, which helped to improve the manuscript. This work was supported by the National High Technology Research and Development Program of China (grant no. 2013AA122002), the Knowledge Innovation Program of the Chinese Academy of Sciences (grant no. KZCX2-EW-QN207), the Special Fund for Meteorological Scientific Research in Public Interest (GYHY201306045) and the National Natural Science Foundation of China (grant no. 91437220).

Edited by: M. Heimann

## References

- Andres, R. J., Gregg, J. S., Losey, L., Marland, G., and Boden, T. A.: Monthly, global emissions of carbon dioxide from fossil fuel consumption, *Tellus* 63B, 309–327, 2011.
- Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric  $CO_2$ , *Tellus* B, 58, 359–365, 2006a.
- Baker, D. F., Law, R. M., Gurney, K. M., Rayner, P. Peylin, P. Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., and Zhu, Z.: TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional  $CO_2$  fluxes, 1988–2003, *Global Biogeochem. Cy.*, 20, GB1002, doi:10.1029/2004GB002439, 2006b.
- Bousquet, P., Peylin, P., Ciais, P., Le Quere, C., Friedlingstein, P., and Tans, P. P.: Regional changes in carbon dioxide fluxes of land and oceans since 1980, *Science*, 290, 1342–1346, 2000.

- Cai, Z., Liu, Y., and Yang, D.: Sensitivity studies for the retrieval of XCO<sub>2</sub> from simulated Chinese Carbon Satellite (TanSat) measurements: a linear error analysis, *Sci. China Ser. D*, 57, 1919–1928, doi:10.1007/s11430-013-4707-1, 2014.
- Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Toward robust and consistent regional CO<sub>2</sub> flux estimates from in situ and spaceborne measurements of atmospheric CO<sub>2</sub>, *Geophys. Res. Lett.*, 41, 1065–1070, doi:10.1002/2013GL058772, 2014.
- Engelen, R. J., Serrar, S., and Chevallier, F.: Four-dimensional data assimilation of atmospheric CO<sub>2</sub> using AIRS observations, *J. Geophys. Res.*, 114, D03303, doi:10.1029/2008JD010739, 2009.
- Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003.
- Feng, L., Palmer, P. I., Bösch, H., and Dance, S.: Estimating surface CO<sub>2</sub> fluxes from space-borne CO<sub>2</sub> dry air mole fraction observations using an ensemble Kalman Filter, *Atmos. Chem. Phys.*, 9, 2619–2633, doi:10.5194/acp-9-2619-2009, 2009.
- Feng, L., Palmer, P. I., Yang, Y., Yantosca, R. M., Kawa, S. R., Paris, J.-D., Matsueda, H., and Machida, T.: Evaluating a 3-D transport model of atmospheric CO<sub>2</sub> using ground-based, aircraft, and space-borne data, *Atmos. Chem. Phys.*, 11, 2789–2803, doi:10.5194/acp-11-2789-2011, 2011.
- Friedl, R. R. (Ed.): Atmospheric Effects of Subsonic Aircraft: Interim Assessment Report of the Advanced Subsonic Technology Program, Ref. Publ. 1400, NASA, Greenbelt, Md., 168 pp., 1997.
- Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R.: Balance and Ensemble Kalman Filter Localization Techniques, *Mon Weather Rev.*, 139, 511–522, 2011.
- Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, 126, 796–811, 1998.
- 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 assimilation methodology, *J. Geophys. Res.*, 117, D24101, doi:10.1029/2012JD018259, 2012.
- Kim, B. Y., Fleming, G. G., Lee, J. J., Waitz, I. A., Clarke, J.-P., Balasubramanian, S., Malwitz, A., Klima, K., Locke, M., Holsclaw, C. A., Maurice, L. Q., and Gupta, M. L.: System for assessing Aviation's Global Emissions (SAGE), Part 1: Model description and inventory results, *Transport. Res. D-TRE*, 12, 325–346, 2007.
- Kim, B., Fleming, G., Balasubramanian, S., Malwitz, A., Lee, J., Waitz, I., Klima, K., Locke, M., Holsclaw, C., Morales, A., McQueen, E., and Gillette, W.: System for assessing Aviation's Global Emissions (SAGE) Federal Aviation Administration Office of Environment and Energy, Version 1.5, Global Aviation Emissions Inventories for 2000 through 2004 (FAA-EE-2005-02), September, 2005.
- Li, H., Kalnay, E., and Miyoshi, T.: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter, *Quart. J. Roy. Meteor. Soc.*, 135, 523–533, 2009.
- Liu, J., Fung, I., Kalnay, E., Kang, J.-S., Olsen, E. T., and Chen, L.: Simultaneous assimilation of AIRS Xco<sub>2</sub> and meteorological observations in a carbon climate model with an ensemble Kalman filter, *J. Geophys. Res.*, 117, D05309, doi:10.1029/2011JD016642, 2012.
- Liu, Y., Yang, D., and Cai, Z.: A retrieval algorithm for TanSat XCO<sub>2</sub> observation: retrieval experiments using GOSAT data, *Chinese Sci. Bull.*, 58, 1520–1523, 2013.
- McKinley, G. A., Follows, M. J., and Marshall, J.: Mechanisms of air–sea CO<sub>2</sub> flux variability in the equatorial Pacific and the North Atlantic, *Global Biogeochem. Cy.*, 18, GB2011, doi:10.1029/2003GB002179, 2004.
- Mu, M., Randerson, J. T., van der Werf, G. R., Giglio, L., Kasibhatla, P., Morton, D., Collatz, G. J., DeFries, R. S., Hyer, E. J., Prins, E. M., Griffith, D. W. T., Wunch, D., Toon, G. C., Sherlock, V., and Wennberg, P. O.: Daily and 3-hourly variability in global fire emissions and consequences for atmospheric model predictions of carbon monoxide, *J. Geophys. Res.-Atmos.*, 116, D24303, doi:10.1029/2011JD016245, 2011.
- Nassar, R., Jones, D. B. A., Suntharalingam, P., Chen, J. M., Andres, R. J., Wecht, K. J., Yantosca, R. M., Kulawik, S. S., Bowman, K. W., Worden, J. R., Machida, T., and Matsueda, H.: Modeling global atmospheric CO<sub>2</sub> with improved emission inventories and CO<sub>2</sub> production from the oxidation of other carbon species, *Geosci. Model Dev.*, 3, 689–716, doi:10.5194/gmd-3-689-2010, 2010.
- O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The ACOS CO<sub>2</sub> retrieval algorithm – Part 1: Description and validation against synthetic observations, *Atmos. Meas. Tech.*, 5, 99–121, doi:10.5194/amt-5-99-2012, 2012.
- Olivier, J. G. J. and Berdowski, J. J. M.: Global emissions sources and sinks, in: *The Climate System*, edited by: Berdowski, J., Guicherit, R., and Heij, B. J., 33–78, A. A. Balkema Publishers/Swets & Zeitlinger Publishers, Lisse, the Netherlands, 2001.
- Olsen, S. C. and Randerson, J. T.: Differences between surface and column atmospheric CO<sub>2</sub> and implications for carbon cycle research, *J. Geophys. Res.*, 109, D02301, doi:10.1029/2003JD003968, 2004.
- Pan, X., Tian, X., Li, X., Xie, Z., Shao, A., and Lu, C.: Assimilating Doppler radar radial velocity and reflectivity observations in the weather research and forecasting model by a proper orthogonal-decomposition-based ensemble, three-dimensional variational assimilation method, *J. Geophys. Res.*, 117, D17113, doi:10.1029/2012JD017684, 2012.
- 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<sub>2</sub> 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., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *P. Natl. Acad. Sci. USA*, 104, 18925–18930, 2007.
- Peylin, P., Baker, D., Sarmiento, J., Ciais, P., and Bousquet, P.: Influence of transport uncertainty on annual mean and seasonal in-

- versions of atmospheric CO<sub>2</sub> data, *J. Geophys. Res.*, 107, 4385, doi:10.1029/2001JD000857, 2002.
- Sausen, R. and Schumann, U.: Estimates of the Climate Response to Aircraft CO<sub>2</sub> and NO<sub>x</sub> Emissions Scenarios, *Climate Change*, 44, 27–58, 2000.
- Suntharalingam, P., Jacob, D. J., Palmer, P. I., Logan, J. A., Yantosca, R. M., Xiao, Y., Evans, M. J., Streets, D. G., Vay, S. L., and Sachse, G. W.: Improved quantification of Chinese carbon fluxes using CO<sub>2</sub>/CO correlations in Asian outflow, *J. Geophys. Res.*, 109, D18S18, doi:10.1029/2003JD004362, 2004.
- Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, A., Chipman, D. W., Hales, B. E., Friederich, G. E., Chavez, F., Sabine, C. L., Watson, A. J., Bakker, D. C. E., Schuster, E., 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<sub>2</sub>, and net sea-air CO<sub>2</sub> flux over the global oceans, *Deep-Sea Res. Pt. II*, doi:10.1016/j.dsr2.2008.12.009, 2009.
- Tian, X. and Xie, Z.: A land surface soil moisture data assimilation framework in consideration of the model subgrid-scale heterogeneity and soil water thawing and freezing, *Sci. China Ser. D*, 51, 992–1000, 2008.
- Tian, X., Xie, Z., and Dai, A.: A land surface soil moisture data assimilation system based on the dual-UKF method and the Community Land Model, *J. Geophys. Res.*, 113, D14127, doi:10.1029/2007JD009650, 2008a.
- Tian, X., Xie, Z., and Dai, A.: An ensemble-based explicit four-dimensional variational assimilation method, *J. Geophys. Res.*, 113, D21124, doi:10.1029/2008JD010358, 2008b.
- Tian, X., Xie, Z., and Sun, Q.: A POD-based ensemble four dimensional variational assimilation method, *Tellus A*, 63, 805–816, 2011.
- Tian, X., Xie, Z., Dai, A., Shi, C., Jia, B., Chen, F., and Yang, K.: A dual-pass variational data assimilation framework for estimating soil moisture profiles from AMSR-E microwave brightness temperature, *J. Geophys. Res.*, 114, D16102, doi:10.1029/2008JD011600, 2009.
- Tian, X., Xie, Z., Dai, A., Jia, B., and Shi, C.: A microwave land data assimilation system: Scheme and preliminary evaluation over China, *J. Geophys. Res.*, 115, D21113, doi:10.1029/2010JD014370, 2010.
- 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.
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys.*, 10, 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.
- Wang, B., Liu, J., Wang, S., Cheng, W., Liu, J., Liu, C., Xiao Q., and Kuo, Y.: An economical approach to four-dimensional variational data assimilation, *Adv. Atmos. Sci.*, 27, 715–727, doi:10.1007/s00376-009-9122-3, 2010.
- Wilkerson, J. T., Jacobson, M. Z., Malwitz, A., Balasubramanian, S., Wayson, R., Fleming, G., Naiman, A. D., and Lele, S. K.: Analysis of emission data from global commercial aviation: 2004 and 2006, *Atmos. Chem. Phys.*, 10, 6391–6408, doi:10.5194/acp-10-6391-2010, 2010.
- Yevich, R. and Logan, J. A.: An assessment of biofuel use and burning of agricultural waste in the developing world, *Global Biogeochem. Cy.*, 17, 1095, doi:10.1029/2002GB001952, 2003.