Atmos. Chem. Phys. Discuss., 13, 24755–24784, 2013
www.atmos-chem-phys-discuss.net/13/24755/2013/
doi:10.5194/acpd-13-24755-2013

Atmospheric Chemistry and Physics Discussions
Open Access

A joint data assimilation system (Tan-Tracker) to simultaneously estimate surface CO_2 fluxes and 3-D atmospheric CO_2 concentrations from observations

X. Tian[1,2], Z. Xie[2], Y. Liu[3], Z. Cai[3], Y. Fu[4], H. Zhang[5], and L. Feng[6]

[1]ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
[2]LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[3]LAGEO, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[4]Climate Change Research Center(CCRC), Chinese Academy of Sciences, Beijing, China
[5]Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, China
[6]School of GeoSciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JN, UK

Received: 4 September 2013 – Accepted: 4 September 2013 – Published: 24 September 2013

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

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

Abstract

To quantitatively estimate CO_2 surface fluxes (CFs) from atmospheric observations, a joint data assimilation system ("Tan-Tracker") is developed by incorporating a joint data assimilation framework into the GEOS-Chem atmospheric transport model. In Tan-Tracker, we choose an identity operator as the CF dynamical model to describe the CFs' evolution, which constitutes an augmented dynamical model together with the GEOS-Chem atmospheric transport model. In this case, the large-scale vector made up of CFs and CO_2 concentrations is taken as the prognostic variable for the augmented dynamical model. And thus both CO_2 concentrations and CFs are jointly assimilated by using the atmospheric observations (e.g., the in-situ observations or satellite measurements). In contrast, in the traditional joint data assimilation frameworks, CFs are usually treated as the model parameters and form a state-parameter augmented vector jointly with CO_2 concentrations. The absence of a CF dynamical model will certainly result in a large waste of observed information since any useful information for CFs' improvement achieved by the current data assimilation procedure could not be used in the next assimilation cycle. Observing system simulation experiments (OSSEs) are carefully designed to evaluate the Tan-Tracker system in comparison to 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_2 concentrations and CO_2 fluxes, mainly due to the simultaneous assimilation of CO_2 concentrations and CFs in our Tan-Tracker data assimilation system.

1 Introduction

Carbon cycle data assimilation systems offer a promising new tool for CO_2 surface flux (CF) inversion (e.g., Peters et al., 2005; Feng et al., 2009), which tends to yield CO_2 surface flux estimates by optimally combining information from both chemistry transport model (CTM) simulations and atmospheric CO_2 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 (Evesen, 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_2 flux propagation since there is no other suitable dynamical model available. Unfortunately, the uncertainty of the initial CO_2 concentration fields is almost ignored in CarbonTracker. Actually, this uncertainty has such a large effect on CF estimates that neglecting its effect might result in unpredictable consequences (Bousquet et al., 2000; McKinley et al., 2004; Peylin et al., 2005). Recently, Kang et al. (2011, 2012) presented a simultaneous data assimilation of surface CO_2 fluxes and atmospheric CO_2 concentrations by means of the Local Ensemble Transform Kalman Filter (LETKF-CDAS). In LETKF-CDAS, the state vector is composed of meteorological variables, atmospheric CO_2 and CFs. In this application, the CFs should are treated as the model forcing (or boundary condition) rather than model states (as in Peters et al., 2005) and there is no other dynamical model adopted in LETKF-CDAS to describe CFs' integration. This implies that the CFs are essentially treated as the model (i.e., the CTM) parameters, which constitute a state-parameter augmented vector together with the model prognostic variables (Tian et al., 2008a; Tian and Xie, 2008). The exclusion of a CF dynamical model will mean that any useful information for CFs' improvement achieved by the current data assimilation procedure could not be used in the next assimilation cycle, meaning that the observed information

ACPD
13, 24755–24784, 2013

is underutilized. Similarly, Feng et al. (2009) also developed an ensemble Kalman filter to estimate 8 day CO_2 surface fluxes over geographical regions globally from satellite measurements of CO_2.

The four-dimensional variational data assimilation (4DVar) method has also been introduced in this field (e.g., Baker et al., 2006; Engelen et al., 2009). Compared with EnKF, 4DVar has its own attractive features: e.g., 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 4DVar. 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 4DVar) of data assimilation while simultaneously off-setting their respective weaknesses. In PODEn4DVar, the control (state) variables in the 4DVar cost function appear explicit 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 4DVar. 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 4DVar and EnKF 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 3-dimensional case of PODEn4DVar)-based radar assimilation system on the atmospheric transport WRF model platform (Pan et al., 2012), which demonstrates its robust performance in the atmospheric transport data assimilation.

In this study, we further 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. So in this case, the large-scale vector made up of CFs and CO_2 concentrations is supposed to be the prognostic variable, which is designed to be simultaneously constrained by assimilation of atmospheric CO_2 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. It is followed by 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 (Sect. 3). Finally, some summary and concluding remarks are provided in Sect. 4.

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 noisy measurements through classical filters (e.g., the dual UKF or EnKF) (Tian et al., 2008; Tian and Xie, 2008; Kang et al., 2011, 2012). Tian et al. (2009) expanded the dual-pass assimilation strategy to the PODEn4DVar approach and built a PODEn4DVar-based dual-pass microwave land data assimilation system (Tian et al., 2010). Similar to the usual joint assimilation schemes, the augmented vector used in LETKF-CDAS is also a state-parameter augmented one and the CFs are treated as the model parameters. But it should be noted that the prognostic variable used in Tan-Tracker is the large-scale vector made up of CFs and CO_2 concentrations, whose evolutions are according to the augmented dynamical model consisting of an identity operator and the CTM.

2.1 The Tan-Tracker joint assimilation framework

Considering the first guessed net CO_2 surface exchange $F^*(x,y,t)$, an usual ensemble-based assimilation system (e.g., CarbonTracker) should basically begin from a prepared ensemble of NCFs $F_{i,g}(i = 1,\cdots,N)$

$$F_{i,g}(t) = \lambda_{i,g} F_g^*(t), \tag{1}$$

where λ_g represents a set of linear scaling factors (Peters et al., 2007) for each day and each grid (g) to be assimilated. Usually, the CTM would integrate and produce the 3-D CO_2 concentration ensemble $U_{m,i}$ ($i = 1,\cdots,N$) N times derived by the ensemble of CFs $F_{i,g}(t)$ from the same initial background CO_2 concentration field. However, for Tan-Tracker, we seek a more innovative way to accomplish its implementation. Figure 1 shows the flowchart of the Tan-Tracker joint assimilation system: Tan-Tracker is initiated by a two CTM runs: one is the background run (the blue part in Fig. 1) and the other is the sampling run (the red part in Fig. 1).

In the background run, the CTM (GEOS-Chem) integrates over the assimilation window (= the optimized window + the lag-window + the observational window, see Fig. 2) to produce the background CO_2 concentration fields U_b forced by the background CF series

$$F_b^a(t) = \lambda_b(t) F_a^*(t), \quad (t = 1,\cdots,L_a) \tag{2}$$

which is used to prepare the background joint vector $(\lambda_b, U_b)^T$. Here L_a is the length of the assimilation window and λ_b is the background scaling factor optimized through last assimilation cycle. The assimilation window consists of an optimized window (one week), a lag-window (five weeks) and an observational window (one week). In each assimilation cycle, the observations in the observational window will be used to assimilate the joint prognostic variables $(\lambda, U)^T$ in the optimized window. F_b^a (F_b^s) denotes the background CF series over the assimilation (sampling) window. And F_a^*(F_s^*) represents the first guessed CF series over the assimilation (sampling) window.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |
ACPD
13, 24755–24784, 2013
A joint carbon cycle data assimilation system (Tan-Tracker)
X. Tian et al.

Correspondingly, in the sampling run, we run the CTM from the background CO_2 concentration field U_b^s at the beginning of the sampling window (= the Pre-Assim window + the Assimilation window + the Post-Assim Window) (Fig. 2) driven by the background CF series $F_b^s(t) = \lambda_b(t)F_s^*(t)$ ($t = 1,\cdots,L_s$, $L_s = L_{Pre} + L_a + L_{Pos}$ is the length of the sampling window and L_{Pre}, L_{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_2 concentration series U_i^s ($i = 1,\cdots,L_s$, 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 $\left(\lambda_{m,i}, U_{m,i}\right)^T$ ($i = 1,\cdots,N$, $N = L_s - L_a + 1$) as follows

$$\left(\lambda_{m,i}, U_{m,i}\right)^T = \mathbf{X}_i^s = \begin{bmatrix} \frac{F_b^s(i)}{F_a^*(1)} \\ \vdots \\ \frac{F_b^s(i+L_a-1)}{F_a^*(L_a)} \\ U_i^s \\ \vdots \\ U_{i+L_a-1}^s \end{bmatrix}, \tag{3}$$

So in this case, the large-scale vector $(\lambda, U)^T$ is viewed as the prognostic variable for Tan-Tracker. And we choose the following identity operator (Eq. 4) as the CF dynamical sub-model to describe CFs' evolution

$$M_{CF} = \mathbf{I}, \tag{4}$$

where $\mathbf{I}$ is the identity matrix. This CF persistence forecasting model (Eq. 4) follows Peters et al. (2005) and assumes the background CFs for one time step are equal to the optimized CFs of the previous time step. In the actual implementations, the following dynamical model (Eq. 5) is applied to the linear scaling factors λ

$$\lambda_b(t+1) = \frac{1}{L_o}\sum_{i=1}^{L_o}\lambda_{a,i}, \tag{5}$$

where L_o is the length of the optimized window (Fig. 2). The CF dynamical sub-model M_{CF} is thus utilized to constitute the augmented dynamical model

$$M = \begin{pmatrix} \mathbf{I} \\ \mathrm{CTM} \end{pmatrix} \tag{6}$$

for Tan-Tracker together with the CTM (GEOS-Chem) model. By applying the observation operator H to the modeled CO_2 concentrations $U_{m,i}$ and the background CO_2 concentrations U_b, we can obtain the ensemble simulated observations U_m^o and the background simulated observations U_b^o as follows

$$U_{m,i}^o = H\left(U_{m,i}\right), \tag{7}$$

and

$$U_b^o = H(U_b), \tag{8}$$

So far, the background joint vector $(\lambda_b, U_b)^T$, the joint vector ensemble $\left(\lambda_{m,i}, U_{m,i}\right)^T$, the background simulated observations U_b^o, the ensemble simulated observations U_m^o, the ensemble CO_2 simulations U_m and the real CO_2 measurements U_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

ACPD
13, 24755–24784, 2013

A joint carbon cycle data assimilation system (Tan-Tracker)

X. Tian et al.

the sampling run is used to produce the joint vector ensemble by means of 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_2 concentrations with the observed according to the 4DVar cost function: the CO_2 concentrations are assimilated to initialize the next assimilation cycle. Meanwhile, the scaling factors λ 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 4DVar cost function

$$J(\boldsymbol{x}') = \frac{1}{2}(\boldsymbol{x}')\mathbf{B}^{-1}(\boldsymbol{x}') + \frac{1}{2}\left[y'(\boldsymbol{x}') - \boldsymbol{y}'_{\text{obs}}\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[y'(\boldsymbol{x}') - \boldsymbol{y}'_{\text{obs}}\right], \tag{9}$$

where $\boldsymbol{x}' = \boldsymbol{x} - \boldsymbol{x}_{\text{b}}$ is the perturbation of the background field $\boldsymbol{x}_{\text{b}}$ at the initial time t_0,

$$\boldsymbol{y}'_{\text{obs}} = \begin{bmatrix} \boldsymbol{y}'_{\text{obs},1} \\ \boldsymbol{y}'_{\text{obs},2} \\ \vdots \\ \boldsymbol{y}'_{\text{obs},S} \end{bmatrix}, \tag{10}$$

$$\boldsymbol{y}' = y'(\boldsymbol{x}') = \begin{bmatrix} (\boldsymbol{y}_1)' \\ (\boldsymbol{y}_2)' \\ \vdots \\ (\boldsymbol{y}_S)' \end{bmatrix}, \tag{11}$$

$$(\boldsymbol{y}_k)' = y_k(\boldsymbol{x}_{\text{b}} + \boldsymbol{x}') - y_k(\boldsymbol{x}_{\text{b}}), \tag{12}$$

$$\boldsymbol{y}'_{\text{obs},k} = \boldsymbol{y}_{\text{obs},k} - y_k(\boldsymbol{x}_{\text{b}}), \tag{13}$$

ACPD
13, 24755–24784, 2013

$$y_k = H_k\left(M_{t_0 \to t_k}(\boldsymbol{x})\right), \tag{14}$$

and

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}_S \end{bmatrix}. \tag{15}$$

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 $\boldsymbol{x}_\mathrm{b}$, the initial model perturbations (MPs) $\boldsymbol{x}'(\boldsymbol{x}'_1, \boldsymbol{x}'_2, \cdots, \boldsymbol{x}'_N)$, the simulated observation perturbations $\boldsymbol{y}'(\boldsymbol{y}'_1, \boldsymbol{y}'_2, \cdots, \boldsymbol{y}'_N)$, the observational increments $\boldsymbol{y}'_{\mathrm{obs},k}$, and the background and observational error covariances $\mathbf{B}$ and $\mathbf{R}_k$, the final PODEn4DVar analysis solution $\boldsymbol{x}_\mathrm{a}$ without localization is formulated through some necessary calculations (see Tian et al., 2010, 2011 for more details) as

$$\boldsymbol{x}_\mathrm{a} = \boldsymbol{x}_\mathrm{b} + \boldsymbol{x}'\mathbf{V}\left[(N-1)\mathbf{I} + \mathbf{P}_y^\mathrm{T}\mathbf{R}^{-1}\mathbf{P}_y\right]^{-1}\mathbf{P}_y^\mathrm{T}\mathbf{R}^{-1}\boldsymbol{y}'_\mathrm{obs}, \tag{16}$$

where $\mathbf{V}$ is derivable from $\left(\boldsymbol{y}'\right)^\mathrm{T}\boldsymbol{y}' = \mathbf{V}\Lambda^2\mathbf{V}^\mathrm{T}$ and $\mathbf{P}_y = \boldsymbol{y}'\mathbf{V}$. To clarify, the background covariance $\mathbf{B}$ is approximately estimated by $\mathbf{B} = \frac{\mathbf{P}_x\mathbf{P}_x^\mathrm{T}}{N-1}(\mathbf{P}_x = \boldsymbol{x}'\mathbf{V})$ in formulating PODEn4DVar.

Especially, in Tan-Tracker,

$$\boldsymbol{y}'_{\mathrm{obs},k} = U_{\mathrm{o},k} - U_\mathrm{b}^\mathrm{o}, \tag{17}$$

and

$$\boldsymbol{y}' = U_\mathrm{m}^\mathrm{o} - U_\mathrm{b}^\mathrm{o}, \tag{18}$$

ACPD
13, 24755–24784, 2013

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

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

$$H = \begin{bmatrix} H_1 & 0 & \cdots & 0 \\ 0 & H_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_S \end{bmatrix}. \quad (19)$$

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

$$\boldsymbol{x}_b = (\lambda_b, U_b)^T, \quad (20)$$

and

$$\boldsymbol{x}' = (\lambda_m, U_m)^T - (\lambda_b, U_b)^T, \quad (21)$$

in Tan-Tracker.

We have realized the coupling between the joint assimilation framework with the PO-DEn4DVar assimilation processor through Eqs. (17–21) (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)

$$\rho_h[i,j] = e^{-d_{i,j}/d_0} \quad (22)$$

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 $\boldsymbol{x}$ and the observational vector $\boldsymbol{y}$, 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.

ACPD
13, 24755–24784, 2013

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

$$\boldsymbol{x}_\mathrm{a} = \boldsymbol{x}_\mathrm{b} + \rho_h \circ \left\{ \boldsymbol{x}'\mathbf{V}\left[(N-1)\mathbf{I} + \mathbf{P}_y^\mathrm{T}\mathbf{R}^{-1}\mathbf{P}_y\right]^{-1}\mathbf{P}_y^\mathrm{T}\mathbf{R}^{-1}\right\} \boldsymbol{y}'_\mathrm{obs}. \tag{23}$$

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_2 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 0.01 hPa. The original GEOS-Chem CO_2 simulation was described in Suntharalingam et al. (2004) and updated by Nassar et al. (2010). Our simulations include CO_2 fluxes from fossil fuel burning and cement production, biomass burning, biofuel burning, ocean exchange, the terrestrial biosphere exchange, the chemical production of CO_2 from the atmospheric oxidation of other carbon species, as well as the emissions from shipping and aviation (Nassar et al., 2010). For this work, our model simulation was initialized on 1 January 2008 with a globally-uniform 3-D CO_2 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_2 at Mauna Loa in 2007 was 383.76 ppm, with monthly means of 383.89 ppm in

ACPD
13, 24755–24784, 2013

A joint carbon cycle data assimilation system (Tan-Tracker)

X. Tian et al.

December 2007 and 385.44 ppm in January 2008. A two year spin-up simulation from this initialized state allows model transport, sources and sinks to develop the global spatial patterns of CO_2, and this approach was evaluated in Nassar et al. (2010). After the spin-up run, the obtained CO_2 fields were used to drive the observing system simulation experiments. In all the following OSSEs, we firstly assume the default surface CO_2 fluxes released with the GEOS-Chem model as the true CF series F_{True}. Then we ran the GEOS-Chem model, driven by the true CF series F_{True} to obtain the true CO_2 concentration results from 1 January 2010 to 31 December 2010 (i.e., the assimilation period). The artificial CO_2 observations are thus generated by sampling the daily true CO_2 concentrations with adding small random noise through the 136 observational sites used in this study. The background (first guessed) CF series F_{b} are set to $1.8F_{\mathrm{True}}$, which drive the GEOS-Chem model at the same resolution (2° latitude × 2.5° longitude) to produce the background CO_2 simulations from the spun-up equilibrium state.

The performance of our Tan-Tracker system is examined by comparing with the simplified version (referred to as TT-S) only taking CFs as the prognostic variables. TT-S is somewhat similar to CarbonTracker except for replacing the ensemble square root filter (EnSRF) by the PODEn4DVar approach and using the GEOS-Chem model instead of the TM5 model. Similarly to CarbonTracker, the GEOS-Chem model in TT-S is actually the observation operator linking the CFs with CO_2 observations. Since the CO_2 concentrations are not assimilated together with the CFs in TT-S, we use the optimized CO_2 concentrations obtained by the GEOS-Chem model simulations with last assimilated CFs as the initial field at the beginning of the assimilation window for each assimilation cycle. All the assimilation processes are initialed by the GEOS-Chem model with the background CF series F_{b} (= $1.8F_{\mathrm{True}}$) and conducted continuously by assimilating the daily pseudo-observations throughout the assimilation period.

In all the OSSEs, the default lag-window is five weeks, the observational-window and optimized-window are both one week. The referenced 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 several sensitivity numerical experiments; the results of which are presented in Sect. 3.2.

3.2 Experimental results

To evaluate the Tan-Tracker's performance in a general view, time series of the daily global mean fluxes and CO_2 concentrations from the background simulations, the TT-S and the TT (Tan-Tracker) assimilations are compared with the true simulations in Fig. 3. Not surprisingly, the background simulations (referred to as Sim) are doomed to deviate seriously from the "true" simulations due to the predetermined background CF series F_b $(= 1.8F_{True})$. Remarkably, since both the CO_2 concentrations and CFs are simultaneously assimilated under the joint assimilation framework, it could largely eliminate the uncertainty of the initial CO_2 concentrations on the CO_2 evolution during the assimilation window and maximize the observations' potential. Probably for this reason, it shows that Tan-Tracker works beautifully throughout the whole assimilation period, especially after the first few months, which can be considered a spin-up period. On the contrary, the performance of TT-S is not very robust and its assimilated errors don't a trend of becoming less even though its performance behaves substantially better than the background simulation case: obviously, the impacts of the CO_2 concentration have not been taken into full consideration in the TT-S system and there must be some non-negligeable errors remaining in the TT-S-optimized CO_2 concentrations (Fig. 3b). The resulting errors in the initial CO_2 concentrations will in turn contaminate the TT-S assimilation of 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.

Similar to Peters et al. (2005), we also aggregated the daily, gridded (2° 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 4 shows the 24

ACPD
13, 24755–24784, 2013

"super-regions" aggregated mean 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 three: CT-09: Tropical Asia, CT-12: North Pacific Temperate, and CT-20: Southern Ocean, whose absolute values are very small (Fig. 4a). And furthermore, as far as the CO_2 concentration is concerned, the superior performance of Tan-Tracker beyond TT-S is increasingly obvious (Fig. 4b): the differences between the "truth" and the TT-assimilated 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 CO_2 concentrations and CFs is indispensible. The time series of daily mean fluxes and CO_2 concentrations from the four selected super-regions (Temperate North America, Europe, Boreal Eurasia, and Southern Ocean) are shown in Figs. 5 and 6. 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_2 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 superregion during October–December, 2010: here the TT-S-optimized CO_2 concentrations are even worse than the background simulations (Fig. 6d).

To evaluate the performance of our Tan-Tracker data assimilations system comprehensively, we show the root mean-square (RMS) errors for the daily, gridded (2° latitude × 2.5° longitude) TT- and TT-S-assimilated fluxes from 1 July to 31 December 2010 in Fig. 7. Complementarily, their corresponding RMS errors for the assimilated (optimized) CO_2 concentrations are also shown in Fig. 8. Compared with the Tan-Tracker case, larger RMS errors ($> 300 \times 10^{-11}\,\mathrm{kg\,C\,m^{-2}\,s^{-1}}$) for the TT-S-assimilated fluxes can be found in the central parts of South America, most of East Asia, and South of Africa (Fig. 7b). Encouragingly, the TT-assimilated flux RMS errors are largely kept at a very low level ($\leq 80 \times 10^{-11}\,\mathrm{kg\,C\,m^{-2}\,s^{-1}}$), in which relatively larger RMS errors (but still much less than the TT-S-assimilated) appear only in a very small area in the central parts of

ACPD
13, 24755–24784, 2013

South America (Fig. 7a). Naturally, a parallel circumstance is also replayed in the CO_2 concentration case (Fig. 8). Evidently, a relatively definite conclusion can be drawn that the uncertainty of the initial CO_2 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 contribution to its excellent performance (Tian et al., 2011). Of course, the imbalance of CFs and 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 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 km as the default or referenced radius. Figure 9 shows time series of the daily global CO_2 concentrations and fluxes from the "truth", and the TT assimilations using the three different horizontal localization radii (d_0 = 900, 1450, and 2000 km). Therefore, we can roughly judge that the Tan-Tracker system could do a good job with its horizontal localization radius around 900 km. Nevertheless, two extremely inappropriate localization radii (d_0 = 100, and 5000 km) are also tested in our experiments (but not shown here), whose poor performance declares the choice of an appropriate covariance localization radius is essential to its (Tan-Tracker) successful implementation.

Finally, to investigate the impacts of sample sizes on the Tan-Tracker assimilation results, we also conduct another group of Tan-Tracker assimilation experiments with the ensemble number $N = 60$, 106, and 150, respectively. Figure 10 shows that the 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 (≥ 100) could generally guarantee the robust performance of our system.

ACPD
13, 24755–24784, 2013

4 Summary and concluding remarks

In this study, a Chinese carbon cycle data assimilation system (i.e., Tan-Tracker) is preliminarily developed based on an advanced hybrid assimilation approach (PODEn4DVar), which is part of the preparation for the launch of the Chinese carbon dioxide observation satellite (TanSat) (Liu et al., 2012; Cai et al., 2013). 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 vector made up of CFs and CO_2 concentrations is actually the prognostic variable, which is designed to be simultaneously assimilated through the measurements of CO_2. As a step towards the application of Tan-Tracker, we carefully designed several groups of observing system simulation experiments (OSSEs) to evaluate the Tan-Tracker's performance comprehensively in comparison to its simplified version (TT-S) only taking CFs as the prognostic variables. It is found that the simultaneous assimilation of CO_2 concentrations and CFs plays a vital role to enhance the Tan-Tracker system performance. In our Tan-Tracker system, CO_2 concentration is assimilated continually accompanying with CF assimilation. The contamination on the Tan-Tracker performance resulted from the uncertainty of the CO_2 concentration evolution would be thus gradually eliminated, which consequently improves CF assimilation.

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 extend Tan-Tracker with functions for utilizing satellite measurements. This goal could be achieved by expanding the observation operator to link the 1-D CO_2 concentration profiles with XCO_2. In light of the Chinese Tan-Sat having been not yet launched, we will focus our proposed Tan-Tracker on GOSAT (the Greenhouse Gases Observing SATellite) measurements of CO_2 as substitute data.

Acknowledgements. We would like to acknowledge Annemarie Fraser's, Paul Palmer's, Ross Bannister's, Wouter Peters's and Ross Bannister's constructive comments on the manuscript.

This work was supported by the National High Technology Research and Development Program of China (Grant No. 2013AA122002), the National Natural Science Foundation of China (Grant No. 41075076), and the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No. KZCX2-EW-QN207).

References

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.

Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO_2, Tellus B, 58, 359–365, 2006.

Cai, Z., Liu, Y., and Yang, D.: Sensitivity studies for the retrieval of XCO_2 from simulated Chinese Carbon Satellite (TanSat) measurements: a linear error analysis, Sci. China Ser. D, doi:10.1007/s11430-013-4707-1, in press, 2013.

Engelen, R. J., Serrar, S., and Chevallier, F.: Four-dimensional data assimilation of atmospheric CO_2 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_2 fluxes from space-borne CO_2 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_2 using ground-based, aircraft, and space-borne data, Atmos. Chem. Phys., 11, 2789–2803, doi:10.5194/acp-11-2789-2011, 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.

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., 117, D05309, doi:10.1029/2011JD016642, 2012.

Liu, Y., Yang, D., and Cai, Z.: A retrieval algorithm for TanSat XCO_2 observation: retrieval experiments using GOSAT data, Chinese Sci. Bull., 58, 1520–1523, 2013.

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_2 with improved emission inventories and CO_2 production from the oxidation of other carbon species, Geosci. Model Dev., 3, 689–716, doi:10.5194/gmd-3-689-2010, 2010.

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., 110, D24304, doi:10.1029/2005JD006157, 2005.

McKinley, G. A., Follows, M. J., and Marshall, J.: Mechanisms of air–sea CO_2 flux variability in the equatorial Pacific and the North Atlantic, Global Biogeochem. Cy., 18, GB2011, doi:10.1029/2003GB002179, 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., 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 inversions of atmospheric CO_2 data, J. Geophys. Res., 107, 4385, doi:10.1029/2001JD000857, 2002.

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 Sachese, G. W.: Improved quantification of Chinese carbon fluxes using CO_2/CO correlations in Asian outflow, J. Geophys. Res., 109, D18S18, doi:10.1029/2003JD004362, 2004.

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., 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.

Tian, X., Xie, Z., and Sun, Q.: A POD-based ensemble four dimensional variational assimilation method, Tellus A, 63, 805–816, 2011.

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.

Zhang, H., Xue, J., Zhuang, S., Zhu, G., and Zhu, Z.: GRAPeS 3D-Var data assimilation system ideal experiments, Acta Meteorol. Sin., 62, 31–41, 2004.

ACPD
13, 24755–24784, 2013

A joint carbon cycle data assimilation system (Tan-Tracker)

X. Tian et al.

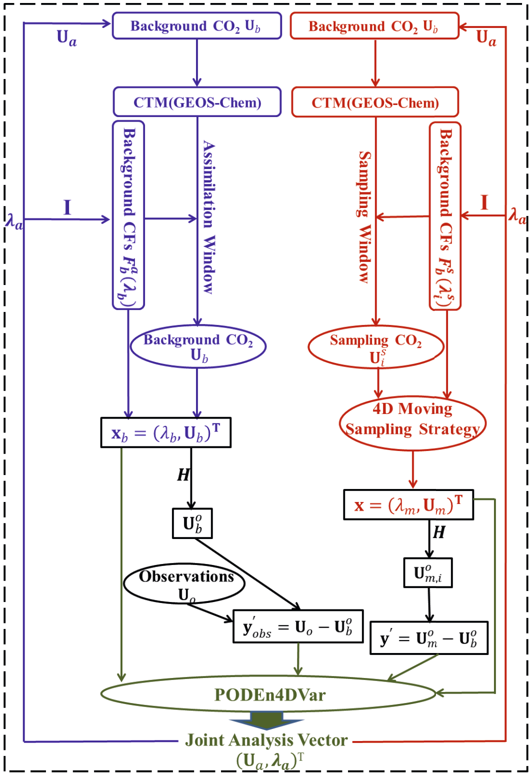

Fig. 1. Flowchart of the Tan-Tracker joint data assimilation system.

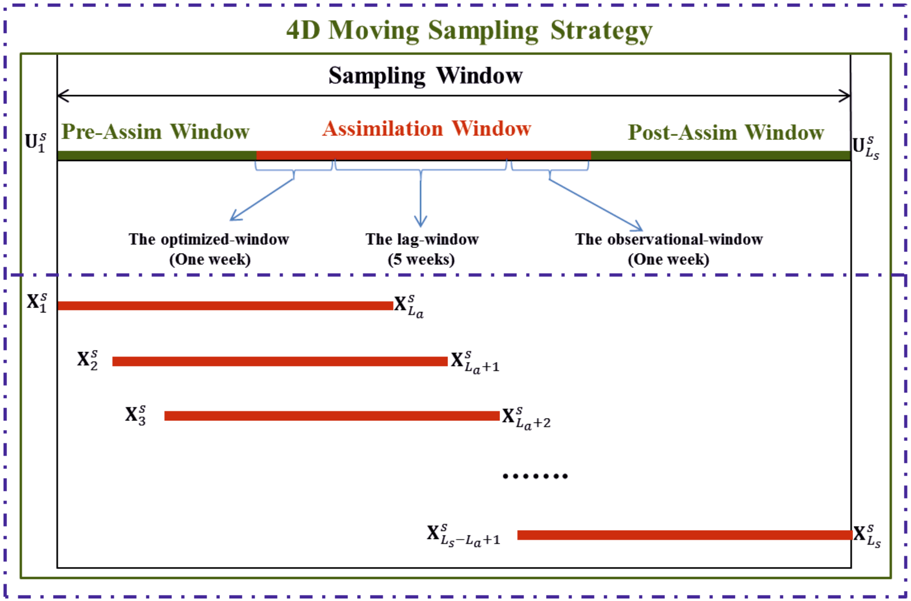

Fig. 2. The 4-D moving sampling strategy.

ACPD
13, 24755–24784, 2013

A joint carbon cycle data assimilation system (Tan-Tracker)

X. Tian et al.

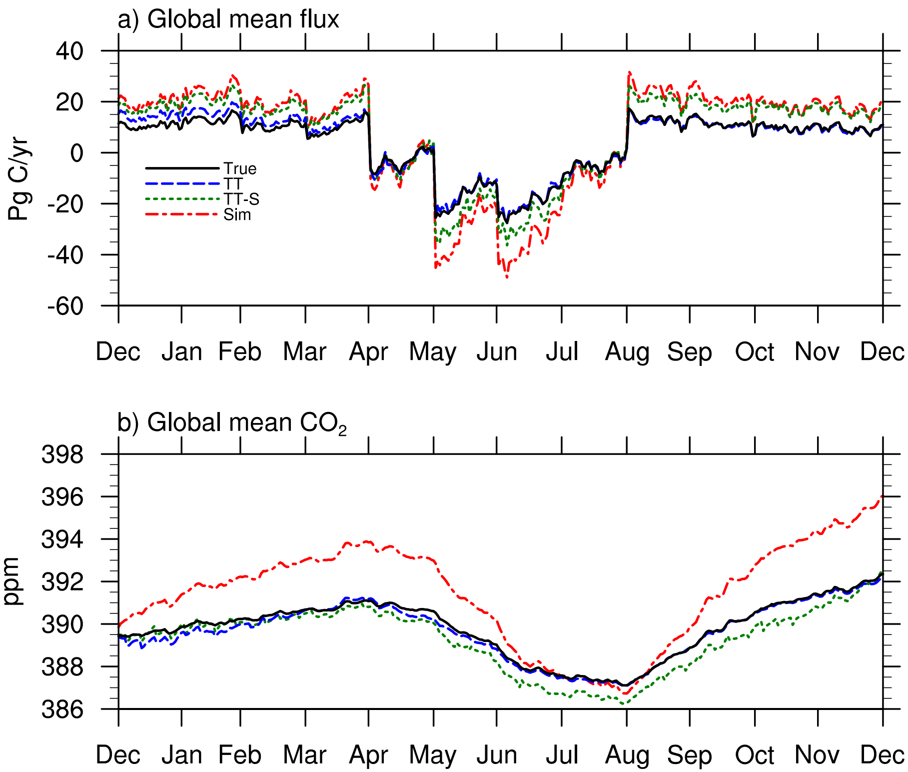

Fig. 3. Time series of the global mean **(a)** CO_2 surface fluxes and **(b)** CO_2 concentrations from the "truth", simulations, TT-S (the simplified version of Tan-Tracker) and TT (Tan-Tracker) assimilations from 1 January to 31 December 2010.

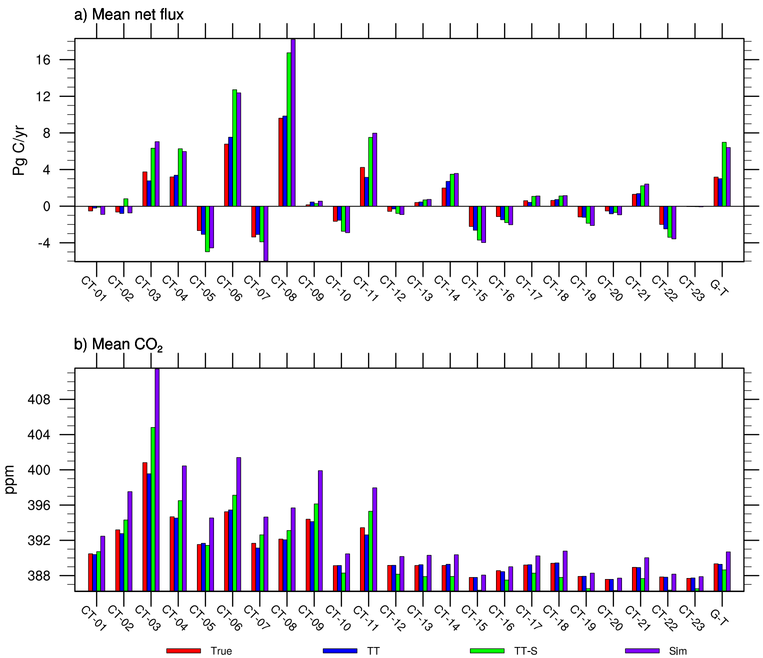

Fig. 4. (a) Mean CO_2 surface fluxes and **(b)** CO_2 concentration from the "truth", simulations, TT-S (the simplified version of Tan-Tracker) and TT (Tan-Tracker) assimilations aggravated to TransCom regions (i.e., CT-01: North America Boreal, CT-02: North America Temperate, CT-03: South America Tropical, CT-04: South America Temperate, CT-05: Northern Africa, CT-06: Southern Africa, CT-07: Eurasian Borea, CT-08: Eurasian 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: Indian Tropical, CT-22: South Indian Tropical, CT-23: Zero Flux Regions, G-T: Global Total) during the period from 1 June–31 December 2010.

ACPD
13, 24755–24784, 2013

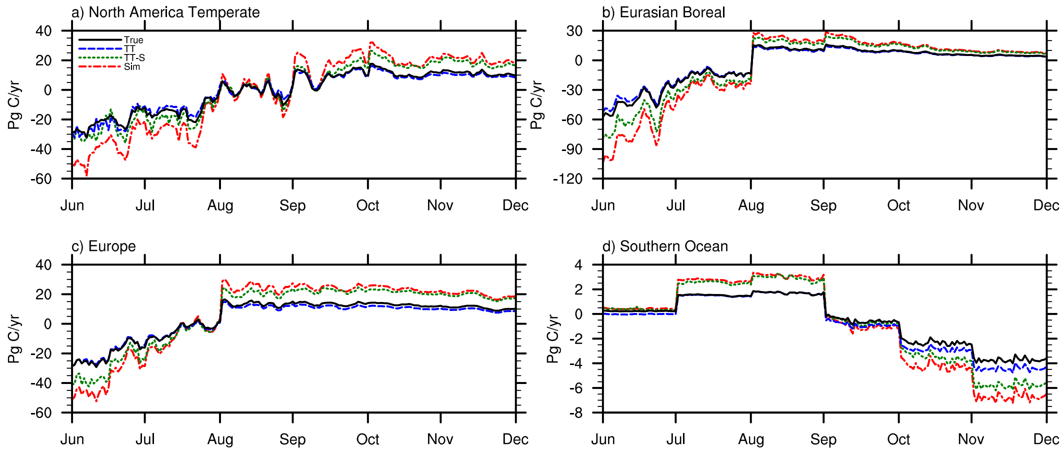

Fig. 5. Time series of the daily mean CO_2 surface fluxes from the "truth", simulations, TT-S (the simplified version of Tan-Tracker) and TT (Tan-Tracker) assimilations aggravated to the selected four TransCom regions (i.e., CT-02, CT-07, CT-11 and CT-20) during the period from 1 July–31 December 2010.

ACPD
13, 24755–24784, 2013

A joint carbon cycle data assimilation system (Tan-Tracker)

X. Tian et al.

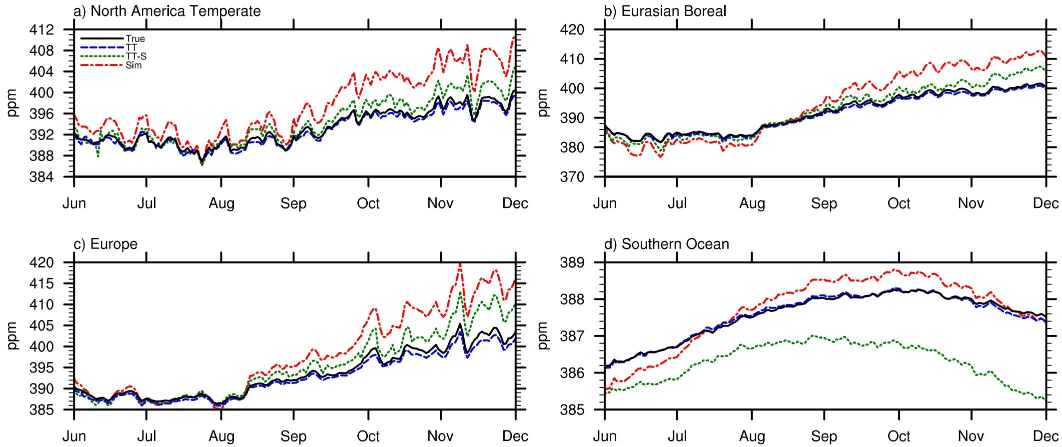

Fig. 6. Same as Fig. 5 but for CO_2 concentrations.

ACPD
13, 24755–24784, 2013

A joint carbon cycle data assimilation system (Tan-Tracker)

X. Tian et al.

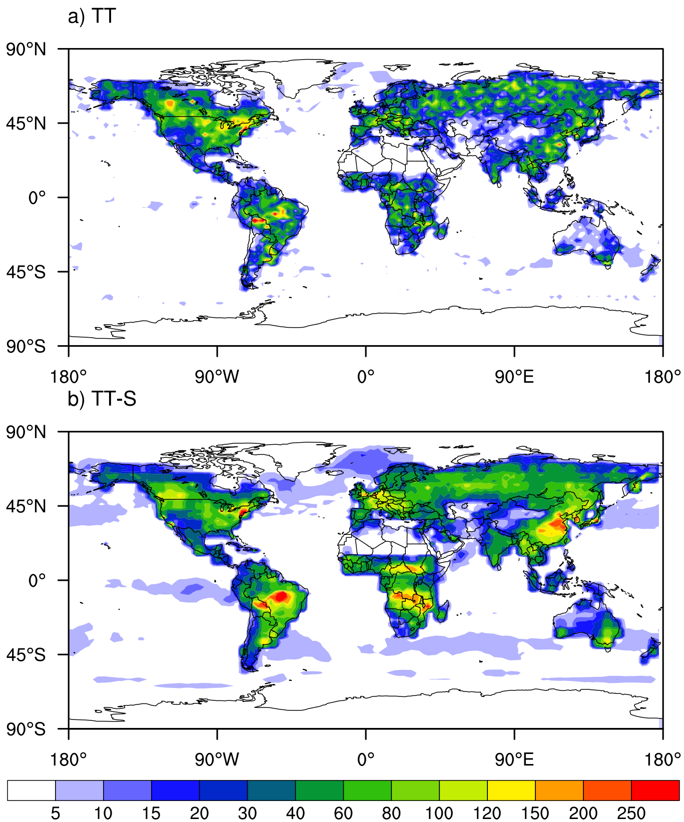

Fig. 7. Root mean-square (RMS) errors (units are 10^{-11} kg C m^{-2} s^{-1}) for the daily **(a)** TT- and **(b)** TT-S-assimilated CO_2 surface fluxes during the period from 1 July–31 December 2010.

ACPD
13, 24755–24784, 2013

A joint carbon cycle data assimilation system (Tan-Tracker)

X. Tian et al.

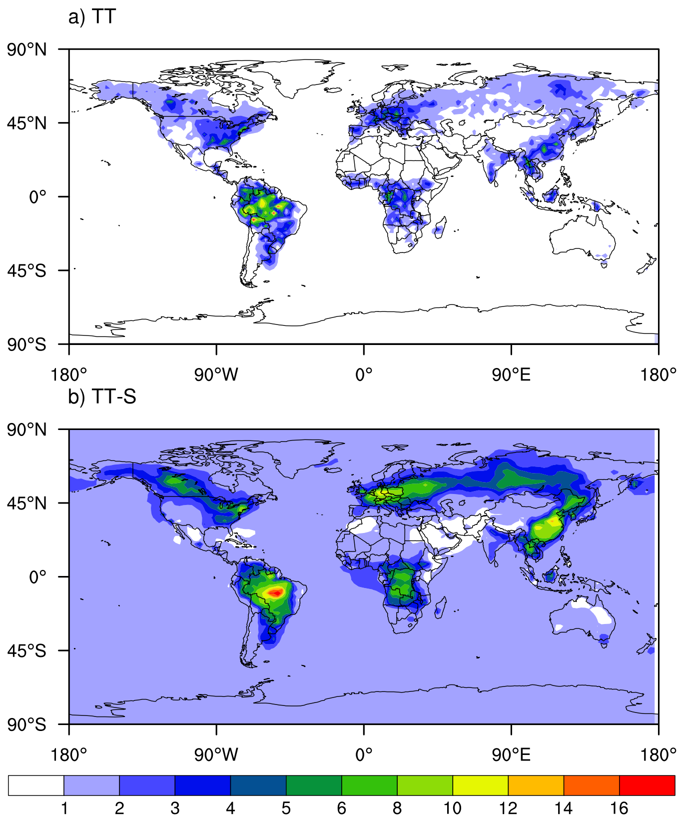

Fig. 8. Same as Fig. 7 but for CO_2 concentrations (units are ppm).

ACPD
13, 24755–24784, 2013

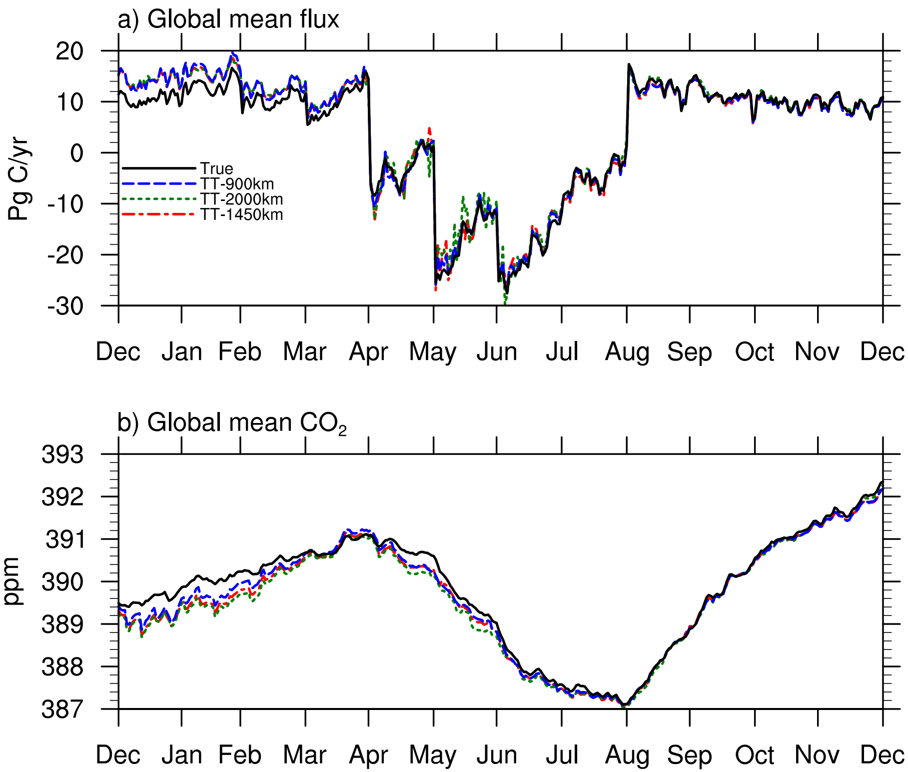

Fig. 9. Time series of the daily global mean **(a)** CO_2 surface fluxes and **(b)** CO_2 concentrations from the "truth", and the TT (Tan-Tracker) assimilations using different covariance localization radii (900 km,1450 km and 2000 km), respectively, from 1 Janyary to 31 December 2010.

ACPD
13, 24755–24784, 2013

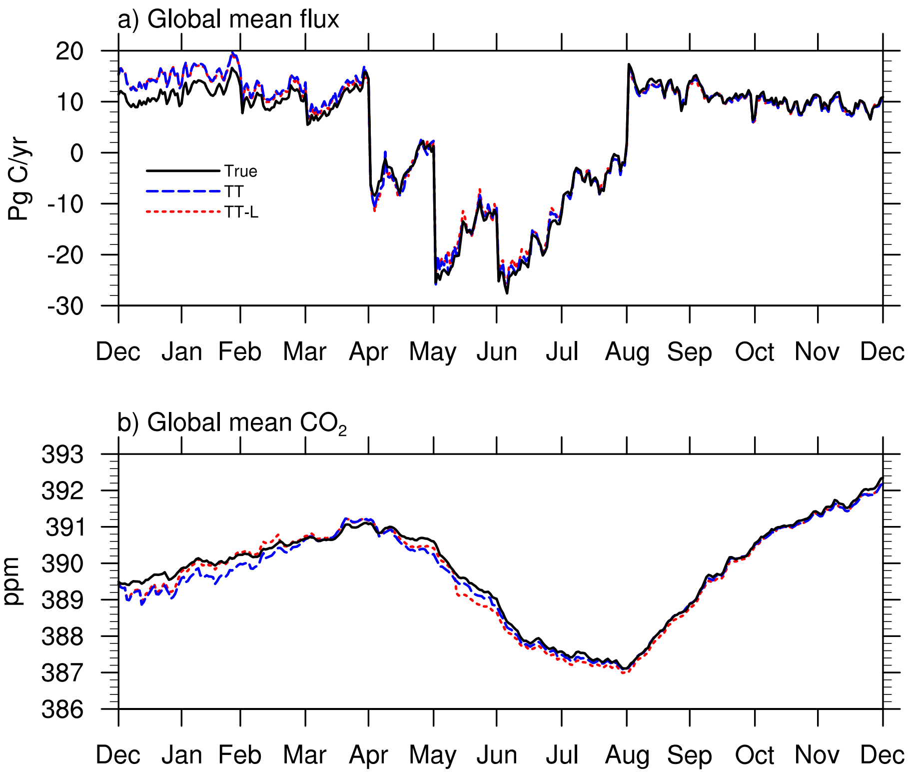

Fig. 10. Time series of the daily global mean **(a)** CO_2 surface fluxes and **(b)** CO_2 concentrations from the "true", and the TT (Tan-Tracker) assimilations with the ensemble number N = 106, and 150, respectively, from 1 January to 31 December 2010.

ACPD
13, 24755–24784, 2013

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀| |▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion