A dual-pass global carbon cycle data assimilation system Tan-Tracker (v1) to estimate surface CO_2 fluxes and 3D atmospheric CO_2 concentrations from spaceborne measurements of atmospheric CO_2

Rui Han[1,2], Xiangjun Tian[1,2,3]

[1]International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
[2]University of Chinese Academy of Sciences, Beijing, 100049, China
[3]Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, 210044, China

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

Abstract. Here we introduce a new version of the global carbon cycle data assimilation system, Tan-Tracker (v1), which is based on the Nonlinear Least Squares Four-dimensional Variational Data Assimilation algorithm (NLS-4DVar) and the Goddard Earth Observing System atmospheric chemistry transport model (GEOS-Chem). Using a dual-pass assimilation framework that consists of a carbon dioxide (CO_2) assimilation pass and a flux assimilation pass, we assimilated the atmosphere column-averaged CO_2 dry air mole fraction (XCO_2), while sequentially optimizing the CO_2 concentration and surface carbon flux via different length windows with the same initial time. When the CO_2 assimilation pass is first performed, a shorter window of 3 days is applied to reduce the influence of the background flux on the initial CO_2 concentration. This allows us to obtain a better initial CO_2 concentration to drive subsequent flux assimilation passes. In the following flux assimilation pass, a properly elongated window of 2 weeks absorbs enough observations to reduce the influence of the initial CO_2 concentration deviation on the flux, resulting in better surface fluxes. In contrast, the joint assimilation system Tan-Tracker (v0) uses the same assimilation window for optimization of CO_2 concentration and flux, making the uncertainties in CO_2 concentration and flux indistinguishable. The proper orthogonal decomposition (POD)-4DVar algorithm applied with the older system is only a rough approximation of ~~the~~ one~~ step~~ iteration of the NLS-4DVar algorithm; thus, it can be difficult to fully resolve the nonlinear relationship between flux and CO_2 concentration. In this study, we designed a set of observation system simulation experiments to assimilate artificial XCO_2 observations, in an attempt to verify the performance of the newly developed dual-pass Tan-Tracker (v1). Compared with the prior and joint system, the dual-pass system provided a better representation of the spatiotemporal distribution of the true flux and true CO_2 concentration. We performed sensitivity tests of the flux assimilation window length and number of NLS-4DVar assimilation iterations. Our results indicated that the appropriate flux assimilation window length (14 days) and the appropriate number of NLS-4DVar maximum iterations (three) could be used to achieve optimal results. Thus, the Tan-Tracker (v1) system, based on a novel dual-pass assimilation framework, provides more accurate surface flux inversion estimates and is ultimately a better tool for carbon cycle research.

1. Introduction

Since the Industrial Revolution, humans have consumed fossil fuels and emitted large amounts of carbon dioxide (CO_2). About 50% of the CO_2 remains in the atmosphere. The continuous rise in global atmospheric CO_2 concentrations breaks the radiation balance of the Earth system, resulting in global climate change. The remaining CO_2 is absorbed by the terrestrial ecosystem and oceans; however, there are still many uncertainties associated with these absorption mechanisms (Ballantyne et al., 2012; Le Quéré et al., 2017). Determining the appropriate carbon budget for the Earth's ecosystem and oceans is important for the development of relevant climate policies and predictions of future scenarios, having been the focus of extensive carbon cycle research (Stocker et al., 2013). In recent years, there has been an increase in multi-source atmospheric CO_2 concentration measurements and model development. The surface carbon flux inversion method, especially carbon cycle data assimilation, obtained by combining model and atmospheric CO_2 information, has made great progress ~~in carbon cycle data assimilation~~ (Peters et al., 2005; Peters et al., 2007; Tian et al., 2014; Deng et al., 2016; Feng et al., 2016; Basu et al., 2013; Basu et al., 2018).

Many ~~have~~ attempts~~ed~~ have been made, assimilating atmospheric CO_2 measurements, to optimize surface carbon flux ~~measurements~~. For example, Carbon-Tracker (Peters et al., 2005; Peters et al., 2007) is a well-designed carbon assimilation system that uses Transport Model 5 (TM5) and the ensemble Kalman filter (EnKF) method (Evensen, 1994) to assimilate *in situ* CO_2 observations. The Carbon Cycle Data Assimilation System (CCDAS) (Rayner et al., 2005;~~)~~ Kaminski et al., 2013) couples the Biosphere Energy-Transfer HYdrosphere (BETHY) model (Kaminski and Heimann, 2001) with the atmospheric transport model TM2, to assimilate satellite observations of photosynthetically active radiation and atmospheric CO_2 concentration observations; the approach is a two-step process, in which the parameters of the carbon cycle model are first optimized to improve surface flux ~~measurement~~ accuracy. Tan-Tracker (v0) (Tian et al., 2014) uses the Goddard Earth Observing System atmospheric chemistry transport model (GEOS-Chem) and the identity matrix as a joint dynamical model; a proper orthogonal decomposition (POD)-based four-dimensional variational assimilation algorithm (POD-4DVar) (Tian et al., 2011) is combined with a joint assimilation framework to integrate *in situ* CO_2 concentration observations, with simultaneous optimization of the CO_2 concentration and flux. This method has obtained good results; however, there are still some problems associated with the joint assimilation framework. The same window lengths make uncertainties of CO_2 and surface flux indistinguishable ~~limit the ability to distinguish the CO_2 concentration from the flux~~. Additionally, the POD-4DVar algorithm is only a rough approximation of ~~a~~ one ~~step~~ iteration of the Nonlinear Least Squares (NLS)-4DVar algorithm (Tian and Feng, 2015; Tian et al., 2018). Although the above assimilation system has achieved reasonable results, the sparse and uneven spatial distributions of *in situ* stations greatly limit the flux optimization accuracy. Several unconventional data assimilation ~~techniques~~ attempts have been explored. For example, Zhang et al. (2014) conducted an assimilation of the aircraft observation Comprehensive Observation Network for Trace gases by Airline (CONTRAL) based on Carbon-Tracker. With the launch of the Greenhouse gases Observing SATellite (GOSAT) (Kuze et al., 2009) and the Orbiting Carbon Observatory-2 (OCO-2) satellite (Crisp et al., 2017), satellite data assimilation experiments have also been conducted based on the atmosphere column-averaged CO_2 dry air mole fraction (XCO_2) at higher temporal and spatial resolutions. Basu et al. (2013) used TM5 4DVar to assimilate GOSAT observations, and showed that satellite data provided an effective constraint for surface carbon source–sink inversion. Tian et al. (2014) used Tan-Tracker (v0) to conduct GOSAT observation assimilations using a set of observing system simulation experiments (OSSEs), and found that the optimized CO_2 concentration and flux showed expected results. Deng et al. (2016) used the GEOS-Chem and the 4DVar method to simultaneously assimilate GOSAT observations of the land and ocean. This method provided a better representation of the CO_2 surface flux than others that used only terrestrial observations; additionally, the results indicated that increasing the observation coverage further improved the sensitivity of surface flux inversion~~ measurements~~. Feng et al. (2016) used the EnKF to assimilate GOSAT observations in Europe; the flux inversion results obtained displayed a larger amplitude change than those using ~~an~~ *in situ* stations. Basu et al. (2018) applied 4DVar OSSEs to OCO-2 observations with multiple atmospheric transport models; they showed that the wider global coverage provided by OCO-2 observations enabled better surface flux representation than *in situ* observations;~~.~~ o~~O~~verall, flux results depend on the ~~atmospheric chemical transmission mode~~ atmospheric chemistry transport model used. The abovementioned assimilation attempts using satellite data have reduced the uncertainty associated with flux inversion ~~measurements~~ and provided some insight into surface carbon flux mechanisms. However, the assimilation of satellite column-average concentration observations of XCO_2 is still in the exploratory stage.

Based on GEOS-Chem and NLS-4DVar (Tian et al., 2018) assimilation of XCO_2 satellite observations, we introduce the Tan-Tracker (v1) carbon cycle data assimilation system. The novel dual-pass data assimilation framework consists of a CO_2 assimilation pass and a flux assimilation pass, which have the same initial time but different assimilation window lengths. Specifically, the first performed CO_2 assimilation pass uses a shorter window of 3 days to reduce the influence of background flux on the initial CO_2~~ readings~~. By minimizing the initial CO_2 deviation, better initial CO_2 concentrations are derived for the subsequent flux assimilation pass. In the following flux assimilation pass, a properly elongated window of 2 weeks absorbs enough observations to reduce the influence of the initial CO_2 concentration deviation on the flux, resulting in a better representation of the surface flux. Compared with the joint Tan-Tracker (v0) assimilation system, the Tan-Tracker (v1) system uses a dual-pass framework to mitigate the effects of the initial CO_2 concentration on surface flux, while using a more advanced assimilation algorithm, NLS-4DVar, to improve the accuracy of the optimized flux results.

This paper is divided into four sections. Section 2 introduces the method and the framework of the Tan-Tracker (v1) system and its coupling to the NLS-4DVar algorithm. In Section 3, we describe the OSSE design using OCO-2 observations, and compare Tan-Tracker (v1), Tan-Tracker (v0), and control experimental results to true results. The flux obtained using Tan-Tracker (v1) exhibited a total spatiotemporal flux distribution and optimized CO_2 concentration that were closer to those of the true flux. ~~A summary~~ Discussions and conclusions are presented in Section 4 and 5.

2. Methods and Systems

2.1 Dual-pass Tan-Tracker (v1) assimilation system framework

The dual-pass carbon cycle data assimilation system Tan-Tracker (v1) is divided into two assimilation passes: a CO_2 assimilation pass and a flux assimilation pass, in addition to an update section (Fig. 1). Based on the NLS-4DVar (Tian and Feng, 2015; Tian et al., 2018) assimilation method~~ for~~, assimilating satellite column-average CO_2 concentration measurements of XCO_2, we optimized the CO_2 concentration and surface CO_2 flux in different lengths of assimilation windows with the same initial time t_0 of CO_2 concentration. First, the CO_2 assimilation pass is implemented. The shorter 3-day window reduces the influence of background flux on the ~~initial~~ CO_2 ~~measurements~~ evolution, minimizing the initial CO_2 deviation to obtain a better initial CO_2 concentration to drive the flux assimilation pass. In the following flux assimilation pass, a properly elongated window of 2 weeks absorbs enough observations to reduce the influence of the initial CO_2 concentration deviation on the flux. As such, t~~T~~he evolution of the CO_2 concentration in the assimilation window is dominated by the background flux~~ for~~ to improve~~d~~ the accuracy of surface flux inversion~~ measurements~~. The update section guarantees a connection between the two adjacent assimilation windows, in which the initial CO_2 concentration and background flux of the CO_2 assimilation pass are provided for the next window, allowing the background flux and flux ensembles of the flux assimilation pass to be updated.

The CO_2 assimilation pass is shown in the blue portion of Figure 1. Given that NLS-4DVar is an ensemble-based ~~hybrid~~ assimilation algorithm, we first prepared a set of 3-day-length CO_2 concentration ensembles, $\mathbf{U}_{s,i}, (i = 1, \cdots, N)$ (see Section 2.3), where S denotes the ensembles and N is the ensemble number. In the CO_2 assimilation pass, we used N = 160. Starting from the background initial CO_2 $\mathbf{U}_{b,t_0}$ forcing by the background flux:

$$\mathbf{F}_b = \boldsymbol{\lambda}_b \times \mathbf{F}^*, \tag{1}$$

where $\mathbf{F}^*$ is the prior flux and $\boldsymbol{\lambda}_b$ is a linear scale factor (Peters et al., 2005; Tian et al., 2014) for the assimilation window, we simulated the 3-day CO_2 concentration $\mathbf{U}_b$ used as the background CO_2. Note that for one certain assimilation cycle, “background flux” is different to “prior flux” as shown in Eq. 1; “background flux” served as the assimilation background field where “prior flux” means prior flux data sets. H_k is a satellite XCO_2 observation operator, as given in Eq. (31). Putting $\mathbf{U}_s, \mathbf{U}_b, H_k$ together with observations $X_{CO_2,Obs}$ into the NLS-4DVar processor, we can obtain an optimized initial CO_2 $\mathbf{U}_{a,t_0}$, to be used as the initial CO_2 of the flux assimilation pass.

In the flux assimilation pass (the red portion shown in Fig. 1), we assume that there is no error in anthropogenic emissions, and only optimize the terrestrial ecosystems flux and oceans flux:

$$\mathbf{F}^* = \mathbf{F}^*_{bio} + \mathbf{F}^*_{oce}, \tag{2}$$

where $\mathbf{F}^*$ is the prior flux, with *bio* referring to the flux from the terrestrial biosphere, and *oce* representing the flux from the ocean. Starting from the optimized initial CO_2 $\mathbf{U}_{a,t_0}$, forcing by a set of prepared flux ensembles:

$$\mathbf{F}_{s,i} = \boldsymbol{\lambda}_{s,i} \times \mathbf{F}^{*}, \left(i = 1, \cdots, N\right), \tag{3}$$

we obtain a set of 2-week CO_2 ensembles $\mathbf{U}_{s,i}, \left(i = 1, \cdots, N\right)$, where $\boldsymbol{\lambda}_{s,i} \left(i = 1, \cdots, N\right)$ is a set of scale factors (see Section 2.3). ~~Using an optimization variable for the flux and c~~Considering computational cost, we chose N = 36. Simultaneously, starting from the background initial CO_2 $\mathbf{U}_{b,t_0}$ forcing by the background flux $\mathbf{F}_b = \boldsymbol{\lambda}_b \times \mathbf{F}^{*}$, we simulated the 2-week CO_2 concentration $\mathbf{U}_b$ as background CO_2. Putting $\boldsymbol{\lambda}_s$, $\mathbf{U}_s$, $\boldsymbol{\lambda}_b$, $\mathbf{U}_b$, H_k, and observations $X_{CO_2,Obs}$ into the NLS-4DVar processor, we can obtain the optimized scale factor $\boldsymbol{\lambda}_a$, with the optimized flux given by $\mathbf{F}_a = \boldsymbol{\lambda}_a \times \mathbf{F}^{*}$.

The update section is shown as the black portion of Figure 1. Starting from the ~~optimized~~ background initial CO_2 $\mathbf{U}_{b,t_0,r}$ (To guarantee the system's mass-balance) of the rth assimilation cycle forced~~ing~~ by optimized fluxes $\mathbf{F}_{a,r}$, and integrating through the window of the flux assimilation pass to the end, we obtain the background initial CO_2 concentration $\mathbf{U}_{b.t_0,r+1}$ of the (r+1)th assimilation cycle. Unlike the joint Tan-Tracker (v0) system, the background initial CO_2 concentration of Tan-Tracker (v1) is obtained by a running model, as opposed to a direct assimilation, thus eliminating the problem of CO_2 over-optimization. Similar to the approach of Peters (2007), the (r+1)th background flux, $\mathbf{F}_{b,r+1} = \boldsymbol{\lambda}_{b,r+1} \times \mathbf{F}^{*}_{r+1}$, is applied using the mean value of the two previous time steps' scale factors a:

$$\boldsymbol{\lambda}_{b,r+1} = (\boldsymbol{\lambda}_{a,r} + \boldsymbol{\lambda}_{a,r-1} + 1)/3. \tag{4}$$

2.2 Coupling of NLS-4DVar with Tan-Tracker (v1) assimilation framework

The NLS-4DVar algorithm is used to solve the optimal initial perturbation $\mathbf{x}'_a$ to satisfy the incremental form of the 4DVar cost function:

$$J\left(\mathbf{x}'\right) = \frac{1}{2}\left(\mathbf{x}'\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{x}'\right) + \frac{1}{2}\sum_{k=0}^{S}\left[L'_k\left(\mathbf{x}'\right) - \mathbf{y}'_{obs,k}\right]^{\mathrm{T}} \mathbf{R}_k^{-1}\left[L'_k\left(\mathbf{x}'\right) - \mathbf{y}'_{obs,k}\right], \tag{5}$$

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

$$L'_k\left(\mathbf{x}'\right) = L_k\left(\mathbf{x}_b + \mathbf{x}'\right) - L_k\left(\mathbf{x}_b\right), \tag{6}$$

$$\mathbf{y}'_{obs,k} = \mathbf{y}_{obs,k} - L_k\left(\mathbf{x}_b\right), \tag{7}$$

$$L_k = H_k M_{t_0 \to t_k}, \tag{8}$$

where the superscript $\mathbf{T}$ is the matrix transpose, the subscript b is the background value, $\mathbf{y}_{obs,k}$ is the observation at time $t_k, k = 0, 1, \cdots, S$, H_k is the observation operator, $M_{t_0 \to t_k}$ is the nonlinear forecast model integrating from t_0 to t_k, and $\mathbf{B}$ and $\mathbf{R}_k$ are the background error and observational error covariance matrices, respectively. For simplicity, $\mathbf{R} = diag(\mathbf{R}_0, \mathbf{R}_1, \cdots, \mathbf{R}_S)$.

As an ensemble-based assimilation approach, NLS-4DVar (Tian and Feng, 2015; Tian et al., 2018) assumes that the optimal analysis increment $\mathbf{x}_a^{'}$ can be expressed by a linear combination of the pre-prepared initial perturbations (IPs):

$$\mathbf{x}_a^{'} = \mathbf{P}_x \boldsymbol{\beta}, \tag{9}$$

where $\mathbf{P}_x = \left(\mathbf{x}_1^{'}, \mathbf{x}_2^{'}, \cdots, \mathbf{x}_N^{'}\right)$ are the initial perturbations, $\mathbf{x}_i^{'} = \mathbf{x}_i - \mathbf{x}_b, \left(i = 1, 2, \cdots, N\right)$, N is the ensemble number, and $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_N\right)$. We can replace the background error covariance matrix $\mathbf{B}$ with an ensemble perturbation estimate:

$$\mathbf{B}_e = \frac{\mathbf{P}_x \mathbf{P}_x^{\mathbf{T}}}{N-1}, \tag{10}$$

Furthermore, symmetric $\mathbf{R}$ has the Cholesky factorization,

$$\mathbf{R} = \mathbf{R}_+^{1/2} \left(\mathbf{R}_+^{1/2}\right)^{\mathbf{T}}, \tag{11}$$

Substituting Eqs. (9), (10) and (11) into Eq. (5), it can be rewritten as follows (Dennis and Schnabel, 1996),

$$J(\boldsymbol{\beta}) = \frac{1}{2} Q(\boldsymbol{\beta})^{\mathbf{T}} Q(\boldsymbol{\beta}), \tag{12}$$

$$Q(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{R}_+^{1/2} \left[L^{'}\left(\mathbf{P}_x \boldsymbol{\beta}\right) - \mathrm{y}_{obs}^{'} \right] \\ \sqrt{N-1}\boldsymbol{\beta} \end{pmatrix}. \tag{13}$$

Thinking approximations (Tian and Feng, 2015):

$$L^{'}\left(\mathbf{x}_j^{'}\right) = \mathbf{y}_j^{'} \approx \mathbf{H}^{'}\mathbf{M}^{'}\left(\mathbf{x}_j^{'}\right), \tag{14}$$

and

$$L^{'}\left(\mathbf{P}_x \boldsymbol{\beta}\right) \approx \mathbf{H}^{'}\mathbf{M}^{'}\left(\mathbf{P}_x \boldsymbol{\beta}\right) \approx \mathbf{P}_y \boldsymbol{\beta}, \tag{15}$$

~~,~~ where $\mathbf{P}_y = \left(\mathbf{y}_1', \mathbf{y}_2', \cdots, \mathbf{y}_N'\right)$, are the observation perturbations (OPs), $\mathbf{y}_j' = L'\left(\mathbf{x}_j'\right), \left(j = 1, 2, \cdots, N\right)$, and $L^{'} = \left(L_0^{'\mathbf{T}}, L_1^{'\mathbf{T}}, \cdots, L_S^{'\mathbf{T}}\right)$. ~~t~~The first-derivate matrix (or Jacobian matrix) $J_{ac}Q(\boldsymbol{\beta})$ of $Q(\boldsymbol{\beta})$ can be computed approximately as follows,

$$J_{ac}Q(\boldsymbol{\beta}) = \frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \approx \begin{pmatrix} \mathbf{R}_+^{1/2}\mathbf{P}_y \\ \sqrt{N-1}\mathbf{I} \end{pmatrix}, \tag{16}$$

where $\mathbf{I}$ denotes the $N \times N$ identity matrix. The Gauss-Newton iteration for the non-linear least squares problem (12) is defined by (Dennis and Schnabel, 1996):

$$\boldsymbol{\beta}^i = \boldsymbol{\beta}^{i-1} - \left[\left(J_{ac}Q(\boldsymbol{\beta}^{i-1})\right)^{\mathbf{T}} \left(J_{ac}Q(\boldsymbol{\beta}^{i-1})\right)\right]^{-1} \left(J_{ac}Q(\boldsymbol{\beta}^{i-1})\right)^{\mathbf{T}} Q(\boldsymbol{\beta}^{i-1}), \tag{17}$$

Substituting Eqs. (13) and (16) into Eq. (17), the cost function Eq. (5) can be rewritten as the least squares form of the control variable $\boldsymbol{\beta}$ (Tian and Feng, 2015) :

$$\boldsymbol{\beta}^{i}=\boldsymbol{\beta}^{i-1}+\left(\mathbf{P}_{y}^{*}\right)^{\mathbf{T}}L^{'}\left(\mathbf{x}_{a}^{'i-1}\right)+\left(\mathbf{P}_{y}^{\#}\right)^{\mathbf{T}}\mathbf{R}^{-1}\left[\mathrm{y}_{obs}^{'}-L^{'}\left(\mathbf{x}_{a}^{'i-1}\right)\right], \tag{18}$$

$$\left(\mathbf{P}_{y}^{*}\right)^{\mathbf{T}}=-\left(N-1\right)\left[\mathbf{P}_{y}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{P}_{y}+\left(N-1\right)\mathrm{I}\right]^{-1}\left[\mathbf{P}_{y}^{\mathbf{T}}\mathbf{P}_{y}\right]^{-1}\mathbf{P}_{y}^{\mathbf{T}}, \tag{19}$$

$$\left(\mathbf{P}_{y}^{\#}\right)^{\mathbf{T}}=\left[\mathbf{P}_{y}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{P}_{y}+\left(N-1\right)\mathrm{I}\right]^{-1}\mathbf{P}_{y}^{\mathbf{T}}. \tag{20}$$

Here, $i=1,2,\cdots,\mathrm{I}_{\max}$, where $\mathrm{I}_{\max}$ is the maximum NLS-4DVar iteration number~~=~~, $\mathbf{P}_{y}=\left(\mathbf{y}_{1}',\mathbf{y}_{2}',\cdots,\mathbf{y}_{N}'\right)$, ~~are the observation perturbations (OPs),~~ $\mathbf{y}_{j}'=L'\left(\mathbf{x}_{j}'\right),\left(j=1,2,\cdots,N\right)$, ~~and~~ $L^{'}=\left(L_{0}^{'\mathbf{T}},L_{1}^{'\mathbf{T}},\cdots,L_{S}^{'\mathbf{T}}\right)$.

Using an ensemble-estimated $\mathbf{B}_{e}$ to replace the background error covariance matrix $\mathbf{B}$ will bring a spurious correlation that can be eliminated by a localization scheme. An efficient local correlation matrix decomposition approach (Zhang and Tian, 2018) can be used to quickly assimilate a large number of observations while ensuring the assimilation results, especially for satellite data assimilation with high spatiotemporal resolution. Its implementation in NLS-4DVar is as follows:

$$\boldsymbol{\beta}^{i}=\boldsymbol{\beta}^{i-1}+\left(\mathbf{P}_{y,\rho}^{*}\right)^{\mathbf{T}}L^{'}\left(\mathbf{x}_{a}^{'i-1}\right)+\left(\mathbf{P}_{y,\rho}^{\#}\right)^{\mathbf{T}}\mathbf{R}^{-1}\left[\mathrm{y}_{obs}^{'}-L^{'}\left(\mathbf{x}_{a}^{'i-1}\right)\right], \tag{21}$$

$$\mathbf{x}_{a}^{'i}=\mathbf{x}_{a}^{'i-1}+\mathbf{P}_{x,\rho}\left(\mathbf{P}_{y,\rho}^{*}\right)^{\mathbf{T}}L^{'}\left(\mathbf{x}_{a}^{'i-1}\right)+\mathbf{P}_{x,\rho}\left(\mathbf{P}_{y,\rho}^{\#}\right)^{\mathbf{T}}\mathbf{R}^{-1}\left[\mathrm{y}_{obs}^{'}-L^{'}\left(\mathbf{x}_{a}^{'i-1}\right)\right], \tag{22}$$

and

$$\mathbf{P}_{x,\rho}=\left(\boldsymbol{\rho}_{m}<e>\mathbf{P}_{x}\right)=\left(\boldsymbol{\rho}_{m}\circ\mathbf{P}_{x,1}^{*},\boldsymbol{\rho}_{m}\circ\mathbf{P}_{x,2}^{*},\cdots,\boldsymbol{\rho}_{m}\circ\mathbf{P}_{x,N}^{*}\right), \tag{23}$$

$$\mathbf{P}_{y,\rho}^{*}=\left(\boldsymbol{\rho}_{o}<e>\mathbf{P}_{y}^{*}\right),\ \ \mathbf{P}_{y,\rho}^{\#}=\left(\boldsymbol{\rho}_{o}<e>\mathbf{P}_{y}^{\#}\right). \tag{24}$$

Here, symbol $<e>$ is given in Eq. (23), symbol "$\circ$" is the Schür product, each column of $\mathbf{P}_{x,j}^{*}\in\mathbb{R}^{n_m\times r}\,(j=1,2,\cdots,N)$ is the same as the jth column of $\mathbf{P}_{x}$, $\boldsymbol{\rho}_{m}\in\mathbb{R}^{n_m\times r}$ is the decomposition matrix of the model grids spatial correlation matrix, and $\boldsymbol{\rho}_{o}\in\mathbb{R}^{n_o\times r}$ is extracted from the decomposition matrix of the correlation matrix:

$$\mathbf{C}_{mo}\approx\boldsymbol{\rho}_{m}\boldsymbol{\rho}_{o}^{\mathbf{T}}. \tag{25}$$

$\mathbf{C}_{mo}\in\mathbb{R}^{n_m\times n_o}$ is the correlation matrix between the model grids and observation positions constructed by the following fifth-order piecewise rational function (Gaspari and Cohn, 1999):

$$\mathbf{C}_{mo}(i,j)=\mathbf{C}_{0}(d_{i,j}/d), \tag{26}$$

where $\mathbf{C}_{0}$ is defined as

$$\mathbf{C}_{0}(l)=\begin{cases}-\frac{1}{4}l^{5}+\frac{1}{2}l^{4}+\frac{5}{8}l^{3}-\frac{5}{3}l^{2}+1, & 0\le l\le 1\\ \frac{1}{12}l^{5}-\frac{1}{2}l^{4}+\frac{5}{8}l^{3}+\frac{5}{3}l^{2}-5l+4-\frac{2}{3}l^{-1}, & 1<l\le 2\\ 0, & 2<l\end{cases} \tag{27}$$

$l=\frac{d_{i,j}}{d}$, d is the localization radii, $d_{i,j}$ is the spatial spherical distance between ith model grid and jth observation, n_m is the model grid number, n_o is the observation number, and r is the number of selected truncation modes.

In the Tan-Tracker (v1) assimilation system, the optimization variables for different assimilation passes differ. In the CO_2 assimilation pass, the optimized state variable $\mathbf{x}$ is the CO_2 concentration $\mathbf{U}$, and $\mathbf{x}_a^{'}$ is the increase in the initial CO_2 concentration. IPs $\mathbf{P}_x$ are the initial perturbations of the CO_2 concentration, and OPs $\mathbf{P}_y$ are the perturbations of simulated XCO_2 within the 3-day window; H_k is the observation operator of XCO_2 given in Eq. (31). For the flux assimilation pass, state variable $\mathbf{x}$ is the scale factor $\boldsymbol{\lambda}$, and $\mathbf{x}_a^{'}$ is the increase in the scale factor within the window. IPs $\mathbf{P}_x$ and OPs $\mathbf{P}_y$ are the scale factor perturbations and simulated XCO_2 perturbations, respectively, within the 2-week window. At this point, the observation operator H_k can be considered as a two-part chemistry transport model and the observation operator of the column-average concentration XCO_2.

2.3 Ensemble generation and update of the Tan-Tracker (v1) assimilation system

The NLS-4DVar assimilation algorithm is an ensemble-based algorithm that is used to approximate the analysis incremental solution space with the ensemble perturbation sample space. As such, the generation and update of the ensemble samples are essential for assimilation accuracy. According to the characteristics of CO_2 and the flux assimilation pass, we designed different sampling and updating methods for the Tan-Tracker (v1) assimilation system.

A historical moving sampling scheme (Wang et al., 2010; Tian et al., 2014) was used in the CO_2 assimilation pass to select samples from a long-term historical CO_2 simulation, and a resampling scheme was used in the new assimilation window. The advantage of selecting samples from the historical simulation is that the appropriate sample size can be selected to ensure good results at a low computational cost. In this study, N = 160 was selected in the experiments to achieve a better assimilation effect.

To ensure better flux results and minimize computational cost, we chose an ensemble number of N = 36 in the flux assimilation pass, integrating from the same initial CO_2 concentration within each window; all ensembles ran throughout the entire assimilation process. The ensemble generation scheme of the flux assimilation pass combines the history sampling and ensemble update. The historical sampling was applied to the initial window, and the N = 36 initial ensemble members were selected by a moving strategy. Ensemble samples of subsequent windows were obtained using the ensemble update given by the Local Ensemble Transform Kalman Filter (Hunt et al., 2007; Tian and Xie, 2012):

$$\mathbf{P}_x^a=\mathbf{P}_x\mathbf{T}, \tag{28}$$

where $\mathbf{P}_x^a$ represents the updated ensemble perturbations, and the transformation matrix $\mathbf{T}$ is given by

$$\mathbf{T}=\left[\left(N-1\right)\mathbf{P}^*\right]^{(1/2)}, \tag{29}$$

with

$$\mathbf{P}^{*}=\left[\mathbf{P}_{y}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{P}_{y}+\left(N-1\right)\mathrm{I}\right]^{-1}. \tag{30}$$

Equation (28) indicates that the updated ensemble perturbation $\mathbf{P}_x^a$ can be obtained from the initial perturbation $\mathbf{P}_x$ and a transformation matrix $\mathbf{T}$.

As the assimilation cycle progresses, the above ensemble update method usually reduces the dispersion of ensemble samples (Wang and Bishop, 2003), leading to an approximate distortion of the ensemble space $\mathbf{P}_x^a$ with respect to the solution space $\mathbf{x}_a^{'}$; this ultimately causes the assimilation to fail. Therefore, we used an inflation factor $\sqrt{\eta}$ (see Zheng et al. (2013) for more details) with the ensemble perturbation $\mathbf{P}_x^a$, in which $\sqrt{\eta}\mathbf{P}_x^a$ maintained the dispersion of the ensemble samples; this is referred to as adaptive ensemble inflation.

3. Observing System Simulation Experiments

3.1 Model settings and observations

The Tan-Tracker (v1) carbon cycle data assimilation system is based on the global three-dimensional (3D) atmospheric chemistry model GEOS-Chem (version: v11-01, http://acmg.seas.harvard.edu/geos), driven by meteorological inputs of Modern-Era Retrospective analysis for Research and Applications (MERRA-2) from the GEOS of the National Aeronautics and Space Administration (NASA, United States) Global Modeling and Assimilation Office. The original GEOS-Chem CO_2 simulation was developed by Suntharalingam et al. (2004). A major update to the CO_2 simulation was completed by Nassar et al. (2010). The latest update to the CO_2 simulation was developed by Nassar et al. (2013) and appears in GEOS-Chem v10-01, which was released in 2015. In the following experiments, we used the same spatiotemporal resolution: a horizontal resolution of 2° × 2.5° (latitude × longitude), 47 vertical layers, a chemical time step of 20 min, a transmission time step of 10 min, and an output time of 3 h for the CO_2 concentration.

The fluxes used to drive GEOS-Chem for the CO_2 simulation were integrated and provided by the Harvard–NASA Emissions Component (HEMCO) model (Keller et al., 2014). There are seven emission inputs from the following sources: fossil fuel, ocean exchange, terrestrial ecosystem fluxes, biomass burning, ships, aviation, and chemical oxidation. Fossil fuel emissions were acquired from the Open-source Data Inventory of Anthropogenic CO_2 (ODIAC) (Oda and Maksyutov, 2011) daily emissions data. Ocean exchange emissions were obtained from daily scaling data by Takahashi et al. (2009). Terrestrial ecosystem fluxes, specifically balanced biosphere exchange with a seasonal cycle but zero net annual uptake, were taken from the hourly data provided by the Simple Biosphere (SBI3) model (Baker et al., 2006; Messerschmidt et al., 2013). Biomass burning emissions were obtained from the Global Fire Emissions Database v4 (GFED4) (Randerson et al., 2018) daily biomass burning data. Ship emissions were based on monthly scaling data from Endresen et al. (2007). Aviation emissions were derived from monthly scaling data (Olsen et al., 2013) from the Aviation Emissions Inventory Code (AEIC) (Simone et al., 2013). Sources of carbonaceous compound oxidation were taken from monthly data provided by Nassar et al. (2010).

The observations used in the OSSEs are based on real OCO-2 satellite column-average concentration XCO_2 data (Crisp et al., 2017), data version v8r (https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_V8r/summary; OCO-2 Science Team, 2017). From Connor's (2008) algorithm, we constructed an OCO-2 satellite XCO_2 observation operator, representing a projection from 3D atmospheric CO_2 concentrations to satellite column-average concentration:

$$X_{CO_2} = X_{CO_2,ap} + \mathbf{A} \cdot \left(U - U_{ap} \right), \tag{31}$$

where $X_{CO_2,ap}$ is the prior column-average concentration, $\mathbf{A}$ is the column-averaging kernel matrix, U_{ap} is the prior CO_2 profile, and U is the profile of the 3D atmospheric CO_2 concentration in each pressure layer of the prior CO_2 profile, used here as the interpolation result of the GEOS-Chem simulation profile.

When constructing the satellite observation $X_{CO_2,O}$, we retained the prior CO_2 profile, the prior column-average concentration, the column-averaging kernel matrix, the pressure profile, quality control parameters, and time and position information. Only the column concentration value $X_{CO_2,Obs}$ and the uncertainty $X_{CO_2,un}$ from real observations were updated. The simulated truth profile U_t was applied to Eq. (31) to obtain the simulated true column-average concentration:

$$X_{CO_2,t} = X_{CO_2,ap} + \mathbf{A} \cdot \left(U_t - U_{ap} \right). \tag{32}$$

By adding a normal distribution random error $X_{CO_2,err} \sim N\left(0, \mu\right)$ (Wang et al., 2010) instead of the observation uncertainty, we were able to determine the simulated column-average concentration:

$$X_{CO_2,O} = X_{CO_2,t} + X_{CO_2,err}, \tag{33}$$

instead of $X_{CO_2,Obs}$, which was used as artificial observations in the OSSEs.

The artificial observations were controlled for quality to ensure that the OSSEs were reasonable and close to the actual situation. We set the quality-control parameter Warn level = 0 (representing the 50% best data) and used an observation – background (O-B) 3σ quality control scheme for data culling. A comparison of artificial data before and after quality control in the first window (2 weeks) of the flux assimilation pass is shown in Figure 2. It is worth noting that Tan-Tracker (v1) does not use data thinning or regional average observations for assimilation, but instead applies the NLS-4DVar algorithm based on an efficient localization scheme (Zhang and Tian, 2018); this allows this tracking system to absorb large amounts of observations in a short period of time (about 10^5 per window in the CO_2 assimilation pass and about 10^6 per window in the flux assimilation pass).

3.2 Observing system simulation experiments

The initial atmospheric CO_2 concentration represents the state of the atmospheric carbon pool at the initial time, which is important for the simulation of CO_2 concentration and flux inversion. All of the initial atmospheric CO_2 concentrations in the following experiments were from the Carbon-Tracker 2017 global CO_2 concentration (Peters et al., 2007;

https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/), interpolated from a global resolution of 2° × 3° (latitude × longitude), with 25 vertical layers, to the GEOS-Chem model grid resolution.

We designed a set of OSSEs as shown in Table 1. Experimental *True* represents the true simulation, starting from the initial CO_2 of the Carbon-Tracker global CO_2 at time 20151101 (for short: CT20151101), running from 20151101 to 20161231. Forcing was driven by true fluxes: the terrestrial ecosystem flux of SIB3 in 2010 and the Takahashi ocean flux in 2010. Artificial observations $X_{CO_2,O}$ of *True* were constructed as discussed in Section 3.1. We also designed a background simulation control run (denoted as *Ctrl*), an assimilation experiment Tan-Tracker (v0) (denoted as *TT_v0*), and an assimilation experiment Tan-Tracker (v1) (denoted as *TT_v1*), with the same initial CO_2 CT20160101, the same running time from 20160101 to 20161231, and the same prior ~~(background)~~ fluxes: the terrestrial ecosystem flux of SIB3 in 2009 and the Takahashi ocean flux in 2009. The rest of the model settings remained the same as in *True*, with the difference being that *TT_v0* and *TT_v1* were assimilation experiments, assimilating artificial observations $X_{CO_2,O}$ of *True*. Note that the comparison between TT_v1 and TT_v0 is performed between two well-developed assimilation systems with their respective optimal parameters.

The settings and parameters for *TT_v0* can be found in Tan-Tracker (Tian et al., 2014), where only observation data, model versions, and prior flux replacements were performed. A comparison of the parameter settings of *TT_v0* and *TT_v1* is shown in Table 2. After the sensitivity test, the localization radii of the CO_2 assimilation pass and flux assimilation pass were both selected to be 2000 km, and the localization truncation modes numbers were $r_x = 50$ and $r_y = 30$ (see Zhang and Tian (2018) for details regarding the selection of localization-related parameters).

3.3 Analysis of results

3.3.1 CO_2 concentration

The CO_2 concentration reflects the state of the atmospheric carbon pool and can be used as a basic indicator for verification in flux inversion. Here, we analyzed the CO_2 concentration results in detail from the time series and spatial distributions. We used the time series of the daily root-mean-square error (RMSE) and the time series of the mean deviations to characterize the deviations of *Ctrl*, *TT_v0*, and *TT_v1* from *True* (Fig. 3). Overall, the indicators in Figure 3 showed that the results after the assimilation were better than the ~~background~~ prior results.

The daily RMSE of XCO_2 between the simulation/assimilation and artificial observations (Fig. 3a), representing the change in column-average concentration at the observed position, provides a comparison between O-B and the difference between observations and assimilation (O-A) to explain the effectiveness of the assimilation. The results in Figure 3a showed that the two versions of the Tan-Tracker carbon cycle data assimilation system effectively absorbed observations for flux optimization, with the *TT_v1* showing slightly better performance than *TT_v0* and superior performance with respect to that of *Ctrl*.

Daily RMSE (Fig. 3b) and the daily mean bias (Fig. 3c) of the atmospheric 3D CO_2 concentration between the simulation/assimilation results and *True* reflect the changes in the atmospheric carbon pool. Figure 3b shows the deviation of the simulation/assimilation results from *True*. The deviation between *Ctrl* and *True* decreased from 1.4 to 0.4 ppm at the initial time from January to February, and remained low (0.4 ppm) from March to June; this showed that the initial concentration deviation was reduced gradually, which could be considered as a model spin-up process. The deviation increased from July to September from 0.6 to 1.0 ppm, which indicated that there was a large deviation between the prior flux and the true flux in the Northern Hemisphere growth season. Finally, the deviation from October to December fell back to 0.3–0.4 ppm, indicating a decrease in the deviation between the prior flux and the true flux in the non-growth season of the Northern Hemisphere. The deviation between *TT_v1* and *True* decreased from 1.4 to 0.2 ppm at the initial time from January to February, maintaining a lower value of 0.2 ppm from March to June. After a slight increase to 0.2–0.6 ppm from July to September, the deviation between *TT_v1* and *True* finally fell back to 0.2 ppm from October to December.

Figure 3c shows the daily mean bias between the simulation/assimilation results and *True*. The daily mean bias of *TT_v1* dropped rapidly from −0.4 to 0 ppm and then remained low (-0.05 to 0.05 ppm); this performance was superior to that of *Ctrl*, which showed a larger bias amplitude. Thus, *TT_v1* exhibited a faster spin-up convergence speed and a smaller deviation over the entire simulation time than *Ctrl*; these improvements were attributed to an adjustment in the optimized flux. The effect of the initial CO_2 optimized by the CO_2 assimilation pass occurred only at the initial time of each window, thus only a small adjustment to the state of the atmospheric carbon pool, and mainly served to improve the accuracy of the optimized flux. This was achieved given the good continuity of the CO_2 results (Figs. 3b and 3c). The results of *TT_v0* were better than those of *Ctrl* but slightly inferior to those of *TT_v1*.

Figure 4 shows the spatial deviation between the simulation/assimilation results and *True* based on the RMSE spatial distribution of the vertical-averaged CO_2 concentration grid time series. Figure 4a displays the RMSE spatial distribution between *Ctrl* and *True*. Large values over land appeared in Western Siberia (1.0–1.2 ppm) and Eastern Siberia, Eastern Central Asia, Eastern North America, and Central South America (0.8–1.0 ppm). Large values over the ocean appeared in the Northern Hemisphere, with an increasing bias trend from the Southern Hemisphere to the Northern Hemisphere (0.2–0.5 ppm). The results of *TT_v1* were better than those of *Ctrl*, with a large bias over land of 0.3–0.5 ppm; the increasing bias trend over the ocean was lower at 0.1–0.2 ppm. The results of *TT_v0* were better than those of *Ctrl* and slightly inferior to those of *TT_v1*.

3.3.2 Flux

In real assimilation experiments, CO_2 concentration results can be used as the main objective indicator of flux evaluation due to the lack of a real flux. However, in OSSEs, we can quantitatively analyze the prior flux ~~quantitatively and~~, optimized ~~the~~ flux and real flux to give the most direct judgment. Below we present a detailed analysis of flux using time series, annual total amounts, and regional distributions.

Figure 5 shows the time series of the simulation/assimilation results of the monthly global total ecosystem, the ocean flux, and their deviations from *True*. Notably, similar to the spin-up process of the numerical model simulation, the first 4 months corresponded to the spin-up process of the flux assimilation pass. During the early stages of the spin-up phase (from January to February), a larger portion of the optimized flux increment was used to adjust the initial CO_2 concentration deviation from the true simulation. As a result, the deviation between the optimized flux and the true flux was larger than the prior value (Fig. 5b); however, the CO_2 concentration deviation continued to decrease (Fig. 3b). As the assimilation progressed, the concentration deviation became more stable. At this time point, the uncertainties in CO_2 concentration and flux could not be distinguished; as such, the assimilation continued to run, allowing for adjustments to the flux and concentration. Finally, the deviations caused during the corresponding flux and concentration optimization processes were minimized. Here, we mainly discuss the flux results from May to December after reaching equilibrium.

The prior flux (*Ctrl*) was in good agreement with the true flux (*True*) (Fig. 5a). Additionally, a significant seasonal cycle was evident (Fig. 5a). April to September is the growing season of the Northern Hemisphere, when the total flux of the global terrestrial ecosystem and oceans is negative, reaching its lowest value in July and August. From October to March, corresponding to the non-growth season in the Northern Hemisphere, the global flux was positive, and there was no obvious monthly change. The main deviation of the prior flux (*Ctrl* in Fig. 5b) appeared in the Northern Hemisphere growing season from June to August, reaching −4.0 PgC yr^{-1}. In addition, there was a significant deviation of about 0.2 PgC yr^{-1} during the non-growth season of the Northern Hemisphere from October to December. The *TT_v1* optimized flux of the dual-pass system showed significant improvement over *Ctrl*. The deviation was reduced to 0.0 PgC yr^{-1} from June to August, and the deviation decreased to 0.1 PgC yr^{-1} from October to December (Fig. 5b). The results from *TT_v0* were better than those of *Ctrl*, but slightly inferior to those of *TT_v1*. Table 3 and Figure 6 show the assimilation/simulation deviations of the terrestrial ecosystem flux, ocean flux, and global total flux from *True* from May to December. Compared with *Ctrl*, the results of *TT_v1* were better optimized for the terrestrial ecosystem flux and slightly improved for the ocean flux. In addition, the results of *TT_v0* were better than those of *Ctrl*, but slightly inferior to those of *TT_v1* for the terrestrial ecosystem flux and slightly superior to those of *TT_v1* for ocean flux.

We used the TransCom "super-regions" (Gurney et al., 2002) to calculate the regional total flux. Figure 7 shows the flux results of 11 land regions and the deviation from *True*. The results of *TT_v1* had a positive effect on each region relative to the prior flux of *Ctrl*, with significant improvements in the mid-to-high latitudes of North America, Europe, and Eurasia, and the mid-latitudes of South America and Australia. The results in the equatorial region of South America and Asia did not show significant improvements. The prior flux in Africa was close to the true value; an increase was not obvious in the data. *TT_v0* showed slightly improved results compared with *Ctrl*, but both were inferior to the performance of *TT_v1*.

3.3.3 Sensitivity experiments

The parameters of the carbon cycle data assimilation system Tan-Tracker (v1) are listed in Table 2. The main parameters are the assimilation window length~~, and~~ the maximum NLS-4DVar assimilation iteration number and the localization radii of the flux assimilation pass, as described below.

The flux assimilation pass window length determines the influence of the initial CO_2 concentration and the time of transmission, thus affecting the flux inversion. The sensitivity experiments of the assimilation window were used to select a window length of 7 days (denoted as *v1_07*), 14 days (denoted as *v1_14*), or 30 days (denoted as *v1_30*); the other *TT_v1* parameters remained unchanged. The flux and concentration results are shown in Figure 8. From the time series of the total flux (Fig. 8a,) it could be concluded that the assimilation experiments of all three windows had positive effects; however, the assimilation results of *v1_14* were better than those of *v1_07*, which was better than those of *v1_30*. The CO_2 concentration results (Fig. 8c) showed that the assimilation experiments of all three windows had positive effects. The assimilation results of *v1_07* were roughly equivalent to those of *v1_14*, both of which were better than those of *v1_30*. Thus, flux assimilation pass is sensitive to the length of the assimilation window. Note that, 14-days flux assimilation pass window length is close to those adopted by some other published inversion systems, such as the one week length of Carbon-Tracker (Peters et al., 2007), the one-month length of Basu et al. (2013) also the 7-days length of Tan-Tracker (v0) (Tian et al., 2014). The window of the appropriate length (14 days) had a small initial CO_2 concentration deviation, the appropriate integration time, and was closest to the OCO-2 satellite 16-day regression period, i.e., it was possible to absorb more observations to obtain good flux inversion results.

As the maximum NLS-4DVar iteration number increases, the assimilation results tend to converge, especially for solving the problem of high nonlinear systems. However, the computational cost increases with the number of iterations. The sensitivity experiments of the maximum NLS-4DVar iteration number selected one (*Imax = 1*), two (*Imax = 2*), and three (*Imax = 3*) iterations, with the remaining parameters retaining the values of *TT_v1*. The resulting flux and concentration results are shown in Figure 9. The time series of the monthly total flux (Fig. 9a) and the CO_2 concentration (Fig. 9c) results showed that the assimilation results improved and tended to converge quickly as the number of maximum NLS-4DVar iterations increased. Considering the computational cost, we chose three maximum NLS-4DVar iterations as the final solution.

The sensitivity experiments of the flux assimilation pass localization radii were used to select a localization radius of 1000 km (denoted as Loc-1k), 2000 km (denoted as Loc-2k), or 4000 km (denoted as Loc-4k); the other TT_v1 parameters remained unchanged. The flux and concentration results are shown in Fig. 10. The time series of the monthly total flux (Fig. 10a) and the CO_2 concentration (Fig. 10c) results showed that the assimilation results is better with 2000 km localization radii. It is reasonable to be longer than 900km in Carbon-Tracker (Peters et al., 2005) and Tan-Tracker (v0) because of a shorter model integration time meaning lower error from remote location.

4. Discussion

For each assimilation cycle, the simulated CO_2 concentration errors originated from both the initial CO_2 and the background flux errors. These errors entangled with the model evolution, which is indeed difficult to optimize the CO_2 concentrations and fluxes altogether. Dual-pass assimilation system Tan-Tracker (v1) was proposed to proper distinguish the errors and reduce their influences. The CO_2 assimilation pass with a shorter length (3-days) window is firstly utilized to assimilate the initial CO_2 concentrations with little influence of the background fluxes. This allows us to initiate the subsequent flux assimilation pass from the optimal initial CO_2 concentration. A properly elongated 2-weeks length is specially designed to incorporate enough observations for surface fluxes. Through the above optimization steps, initial CO_2 errors and background flux errors are properly distinguished; CO_2 concentration and surface CO_2 flux are mutually adjusted and optimized. In the ensemble-based Tan-Tracker system, the uncertainties are described by the ensembles in the NLS-4DVar, which are further optimized with the ensembles update. In a summary, dual-pass presents a proper way controlling both CO_2 initial condition and flux successively.

Mass balance is important for a carbon cycle assimilation system. In Tan-Tracker (v1), CO_2 assimilation pass is a directly change to the atmospheric carbon pool and will result in flux bias if accumulated through the whole assimilation process. To avoid this, the update section starts from the background initial CO_2; As a result, the analysis CO_2 concentrations are forced by the model and optimized fluxes only, starting from background initial CO_2 of first window. This also means chemical transport model can impose continuous constraint on flux and CO_2 without truncation error. In other words, optimized flux is not only the best-fitting of current window constrained by observations under low initial CO_2 error, but the best-fitting of the whole assimilation progress constrained by model and mass balance.

5. Conclusion

We designed a new version of a carbon cycle data assimilation system, Tan-Tracker (v1), based on the atmospheric chemical transport model GEOS-Chem and an advanced NLS-4DVar data assimilation algorithm. Using a dual-pass assimilation framework consisting of a CO_2 assimilation pass and a flux assimilation pass, we assimilated atmospheric CO_2 observations to obtain an optimized representation of the surface carbon flux. Compared with the joint assimilation system Tan-Tracker (v0), the dual-pass assimilation system Tan-Tracker (v1) innovatively uses a dual-pass assimilation framework to successively optimize CO_2 concentration and surface carbon flux in different assimilation passes. Optimization of the CO_2 concentration uses a shorter assimilation window to reduce the effects of background flux for a more accurate initial CO_2 concentration ~~measurement~~. Flux optimization uses a longer assimilation window, allowing the system to absorb enough observations to optimize the flux while reducing the effects of the initial CO_2 concentration deviation, resulting in more accurate surface flux estimates.

We designed a set of OSSEs based on OCO-2 satellite data, which we compared with the Tan-Tracker (v0) joint assimilation system. The Tan-Tracker (v1) performance was superior to that of Tan-Tracker (v0) in resolving the CO_2 concentration and surface flux estimates, and was far better than prior ~~direct background~~ simulations. Thus, the dual-pass assimilation strategy offers an advantage in satellite carbon cycle data assimilation. The results of the sensitivity experiment of window length and maximum NLS-4DVar assimilation iterations showed that the appropriate window length (14 days) and a greater number of iterations (three), as permitted by the computational cost, provides better assimilation results.

Future work will focus on multi-satellite (e.g., OCO-2, GOSAT, and Tan-Sat) observations for long-term sequence assimilation, regional high-resolution nested assimilation, and analyses used to distinguish between anthropogenic and natural sources.

Author contribution

X. Tian: conception and design, data analysis and interpretation, manuscript writing, final approval of manuscript; R. Han: model simulation, data analysis and interpretation, manuscript writing.

Code and data availability

Initial global CO_2 concentrations are from Carbon-Tracker CT2017 results provided by NOAA ESRL, Boulder, Colorado, USA from the website at http://carbontracker.noaa.gov. Column-average concentration XCO2 data were produced by the OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center: data version v8r, https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_V8r/summary. Global three-dimensional (3D) atmospheric chemistry model GEOS-Chem version: v11-01 is available at: http://acmg.seas.harvard.edu/geos. The code of the assimilation algorithm NLS-4DVar and POD-4DVar used in the system was originally from Xiangjun Tian and can be free and anonymous accessed at: https://doi.org/10.5281/zenodo.2677887.

Acknowledgement

The work was partially supported by the National Key Research and Development Program of China(2016YFA0600203), the Natural Science Foundation of China (41575100) and the Key Research Program of Frontier Sciences, Chinese Academy of Sciences (CAS)(QYZDY-SSW-DQC012).

References

Baker, D. F., Law, R. M., Gurney, K. R., 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. Cycles, 20(1), 1988–2003, doi:10.1029/2004GB002439, 2006.

Ballantyne, A. P., Alden, C. B., Miller, J. B., Trans, P. P. and White, J. W. C.: Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years, Nature, 488(7409), 70–73, doi:10.1038/nature11299, 2012.

Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A. and Worthy, D.: Global CO_2 fluxes estimated from GOSAT retrievals of total column CO_2, Atmos. Chem. Phys., 13(17), 8695–8717, doi:10.5194/acp-13-8695-2013, 2013.

Basu, S., Baker, D. F., Chevallier, F., Patra, P. K., Liu, J. and Miller, J. B.: The impact of transport model differences on CO_2 surface flux estimates from OCO-2 retrievals of column average CO_2, Atmos. Chem. Phys., 18(10), 7189–7215, doi:10.5194/acp-18-7189-2018, 2018.

Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C. and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error analysis, J. Geophys. Res. Atmos., 113(5), 1–14, doi:10.1029/2006JD008336, 2008.

Crisp, D., Pollock, H., Rosenberg, R., Chapsky, L., Lee, R., Oyafuso, F., Frankenberg, C., Dell, C., Bruegge, C., Doran, G., Eldering, A., Fisher, B., Fu, D., Gunson, M., Mandrake, L., Osterman, G., Schwandner, F., Sun, K., Taylor, T., Wennberg, P. and Wunch, D.: The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products, Atmos. Meas. Tech., 10(1), 59–81, doi:10.5194/amt-10-59-2017, 2017.

Deng, F., Jones, D. B. A., O'Dell, C. W., Nassar, R. and Parazoo, N. C.: Combining GOSAT XCO_2 observations over land and ocean to improve regional CO_2 flux estimates, J. Geophys. Res. Atmos., 121(4), 1896–1913, doi:10.1002/2015JD024157, 2016.

Dennis, J. E. Jr. and Schnabel, R. B. 1996. Numerical Methods for Unconstrained Optimization and Nonlinear Equations (Classics in Applied Mathematics). SIAM, Philadelphia, p. 378. ISBN:0-89871-364-1.

Endresen, Ø., Sørgård, E., Behrens, H. L., Brett, P. O. and Isaksen, I. S. A.: A historical reconstruction of ships' fuel consumption and emissions, J. Geophys. Res. Atmos., 112(12), 1–17, doi:10.1029/2006JD007630, 2007.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10143, doi:10.1029/94JC00572, 1994.

Feng, L., Palmer, P. I., Parker, R. J., Deutscher, N. M., Feist, D. G., Kivi, R., Morino, I. and Sussmann, R.: Estimates of European uptake of CO_2 inferred from GOSAT XCO_2 retrievals: Sensitivity to measurement bias inside and outside Europe, Atmos. Chem. Phys., 16(3), 1289–1302, doi:10.5194/acp-16-1289-2016, 2016.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Q. J. R. Meteorol. Soc., 125(554), 723–757, https://doi.org/10.1002/qj.49712555417, 1999.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Yu-Han Chen, Philippe Ciais, Fan, S., Fungs, I. Y., Gloor, M., Heimann, M., Higuchi, K., Johns, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S. and Yuen, T. T. & C.-W.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415(Journal Article), 5, 2002.

Hunt, B. R., Kostelich, E. J. and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Phys. D Nonlinear Phenom., 230(1–2), 112–126, doi:10.1016/j.physd.2006.11.008, 2007.

Kaminski, T. and Heimann, M.: Inverse Modeling of Atmospheric Carbon Dioxide Fluxes, Science (80-.)., 294(5541), 259a–259, doi:10.1126/science.294.5541.259a, 2001.

Kaminski, T., Knorr, W., Schürmann, G., Scholze, M., Rayner, P. J., Zaehle, S., Blessing, S., Dorigo, W., Gayler, V., Giering, R., Gobron, N., Grant, J. P., Heimann, M., Hooker-Stroud, A., Houweling, S., Kato, T., Kattge, J., Kelley, D., Kemp, S., Koffi, E. N., Köstler, C., Mathieu, P. P., Pinty, B., Reick, C. H., Rödenbeck, C., Schnur, R., Scipal, K., Sebald, C., Stacke, T., Van Scheltinga, A. T., Vossbeck, M., Widmann, H. and Ziehn, T.: The BETHY/JSBACH Carbon Cycle Data Assimilation System: Experiences and challenges, J. Geophys. Res. Biogeosciences, 118(4), 1414–1426, doi:10.1002/jgrg.20118, 2013.

Keller, C. A., Long, M. S., Yantosca, R. M., Da Silva, A. M., Pawson, S. and Jacob, D. J.: HEMCO v1.0: A versatile, ESMF-compliant component for calculating emissions in atmospheric models, Geosci. Model Dev., 7(4), 1409–1417, doi:10.5194/gmd-7-1409-2014, 2014.

Kuze, A., Suto, H., Nakajima, M. and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Appl. Opt., 48(35), 6716, doi:10.1364/AO.48.006716, 2009.

Messerschmidt, J., Parazoo, N., Wunch, D., Deutscher, N. M., Roehl, C., Warneke, T. and Wennberg, P. O.: Evaluation of seasonal atmosphere-biosphere exchange estimations with TCCON measurements, Atmos. Chem. Phys., 13(10), 5103–5115, doi:10.5194/acp-13-5103-2013, 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(2), 689–716, doi:10.5194/gmd-3-689-2010, 2010.

Nassar, R., Napier-Linton, L., Gurney, K. R., Andres, R. J., Oda, T., Vogel, F. R. and Deng, F.: Improving the temporal and spatial distribution of CO_2 emissions from global fossil fuel emission data sets, J. Geophys. Res. Atmos., 118(2), 917–933, doi:10.1029/2012JD018196, 2013.

OCO-2 Science Team: OCO-2 Level 2 bias-corrected XCO2 and other select fields from the full-physics retrieval aggregated as daily files, Retrospective processing V8r, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), , doi:Accessed: Jan 2018, 10.5067/XR7ZWYSKP7D0, 2017.

Oda, T. and Maksyutov, S.: A very high-resolution (1km×1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys., 11(2), 543–556, doi:10.5194/acp-11-543-2011, 2011.

Olsen, S. C., Wuebbles, D. J. and Owen, B.: Comparison of global 3-D aviation emissions datasets, Atmos. Chem. Phys., 13(1), 429–441, doi:10.5194/acp-13-429-2013, 2013.

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(D24), 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., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C. and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker., Proc. Natl. Acad. Sci. U. S. A., 104(48), 18925–18930, doi:10.1073/pnas.0708986104, 2007.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Jackson, R. B. and others: Global carbon budget 2017, Earth Syst. Sci. Data Discuss., 1–79, doi:https://doi.org/10.5194/essd-10-405-2018, 2017.

Randerson, J. T., Werf, G. R. van der, Giglio, L., Collatz, G. J. and Kasibhatla, P. S.: Global Fire Emissions Database, Version 4, (GFEDv4).ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1293., 2018.

Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R. and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), Global Biogeochem. Cycles, 19(2), 1–20, doi:10.1029/2004GB002254, 2005.

Simone, N., Stettler, M., Eastham, S. and Barrett, S.: Aviation Emissions Inventory Code (AEIC) User Manual (R1), Laboratory for Aviation and the Environment, Massachusetts Institute of Technology, Report No: LAE-2013-001-N, www.LAE.MIT.edu., 2013.

Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P.: Climate change 2013 the physical science basis: Working Group I contribution to the fifth assessment report of the intergovernmental panel on climate change, Cambridge University Press., 2013.

Suntharalingam, P., Jacob, D. D., 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 quantificaion of Chinese carbon fluxes using CO_2/CO correlations in Asian outflow, J. Geophys. Res. Atmos., 109(18), 1–13, doi:10.1029/2003JD004362, 2004.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R. and de Baar, H. J. W.: Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO_2 flux over the global oceans, Deep. Res. Part II Top. Stud. Oceanogr., 56(8–10), 554–577, doi:10.1016/j.dsr2.2008.12.009, 2009.

Tian, X. and Feng, X.: A non-linear least squares enhanced POD-4DVar algorithm for data assimilation, Tellus, Ser. A Dyn. Meteorol. Oceanogr., 67(1), doi:10.3402/tellusa.v67.25340, 2015.

Tian, X. and Xie, Z.: Implementations of a square-root ensemble analysis and a hybrid localisation into the POD-based ensemble 4dvar, Tellus, Ser. A Dyn. Meteorol. Oceanogr., 64(1), doi:10.3402/tellusa.v64i0.18375, 2012.

Tian, X., Xie, Z. and Sun, Q.: A POD-based ensemble four-dimensional variational assimilation method, Tellus, Ser. A Dyn. Meteorol. Oceanogr., 63(4), 805–816, doi:10.1111/j.1600-0870.2011.00529.x, 2011.

Tian, X., Xie, Z., Liu, Y., Cai, Z., Fu, Y., Zhang, H. and Feng, L.: A joint data assimilation system (Tan-Tracker) to simultaneously estimate surface CO_2 fluxes and 3-D atmospheric CO_2 concentrations from observations, Atmos. Chem. Phys., 14(23), 13281–13293, doi:10.5194/acp-14-13281-2014, 2014.

Tian, X., Zhang, H., Feng, X. and Xie, Y.: Nonlinear Least Squares En4DVar to 4DEnVar Methods for Data Assimilation: Formulation, Analysis, and Preliminary Evaluation, Mon. Weather Rev., 146(1), 77–93, doi:10.1175/MWR-D-17-0050.1, 2018.

Wang, B., Liu, J., Wang, S., Cheng, W., Liu, J., Liu, C., Qingnong, X. and Ying-Hwa, K.: An Economical Approach to Four-dimensional Variational Data Assimilation, Adv. Atmos. Sci., 27(4), 715–727, doi:10.1007/s00376-009-9122-3.1.Introduction, 2010.

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

Zhang, H. and Tian, X.: An Efficient Local Correlation Matrix Decomposition Approach for the Localization Implementation of Ensemble-Based Assimilation Methods, J. Geophys. Res. Atmos., 123(7), 3556–3573, doi:10.1002/2017JD027999, 2018.

Zhang, H. F., Chen, B. Z., van der Laan-Luijkx, I. T., Machida, T., Matsueda, H., Sawa, Y., Fukuyama, Y., Langenfelds, R., Van Der Schoot, M., Xu, G., Yan, J. W., Cheng, M. L., Zhou, L. X., Tans, P. P. and Peters, W.: Estimating Asian terrestrial carbon fluxes from CONTRAIL aircraft and surface CO_2 observations for the period 2006–2010, Atmos. Chem. Phys., 14(11), 5807–5824, doi:10.5194/acp-14-5807-2014, 2014.

Zheng, X., Wu, G., Zhang, S., Liang, X., Dai, Y. and Li, Y.: Using analysis state to construct a forecast error covariance matrix in ensemble Kalman filter assimilation, Adv. Atmos. Sci., 30(5), 1303–1312, doi:10.1007/s00376-012-2133-5, 2013.

Table 1. Experimental setup of the observing system simulation experiments (OSSEs). SIB3 and Takahashi flux are as described in Section 3.1; the remaining flux in each experiment is the same.

Name	Running time	Initial CO_2	Flux *bio*	Flux *oce*
True	20151101~20161231	CT 20151101	SIB3 2010	Takahashi 2010
Ctrl	20160101~20161231	CT 20160101	SIB3 2009	Takahashi 2009
TT_v0				
TT_v1				

Table 2. Selection of assimilation parameters (parameters for *TT_v1* divided into the CO_2 assimilation pass and the flux assimilation pass).

Name	Window length(days)	Lag window (days)	Localization radius(km)	Localization parameters(r_x,r_y)	Iteration times
TT_v0	7	35(5 weeks)	900		1
CO_2 pass	3		2000	50，30	3
Flux pass	14(2 weeks)		2000	50，30	3

Table 3. Total flux from May to December and its deviation from *True* (unit: PgC yr^{-1}).

	True	*Ctrl*	*TT_v0*	*TT_v1*
Ocean	-2.03069	-1.87451	-1.88218	-1.87805
land ecosystem	-4.09202	-3.96772	-4.04544	-4.10416
total	-6.12273	-5.84224	-5.92764	-5.9822
		Ctrl	*TT_v0*	*TT_v1*
Ocean		0.15618	0.148511	0.152643
land ecosystem		0.124296	0.046578	-0.01214
total		0.280489	0.195089	0.140531

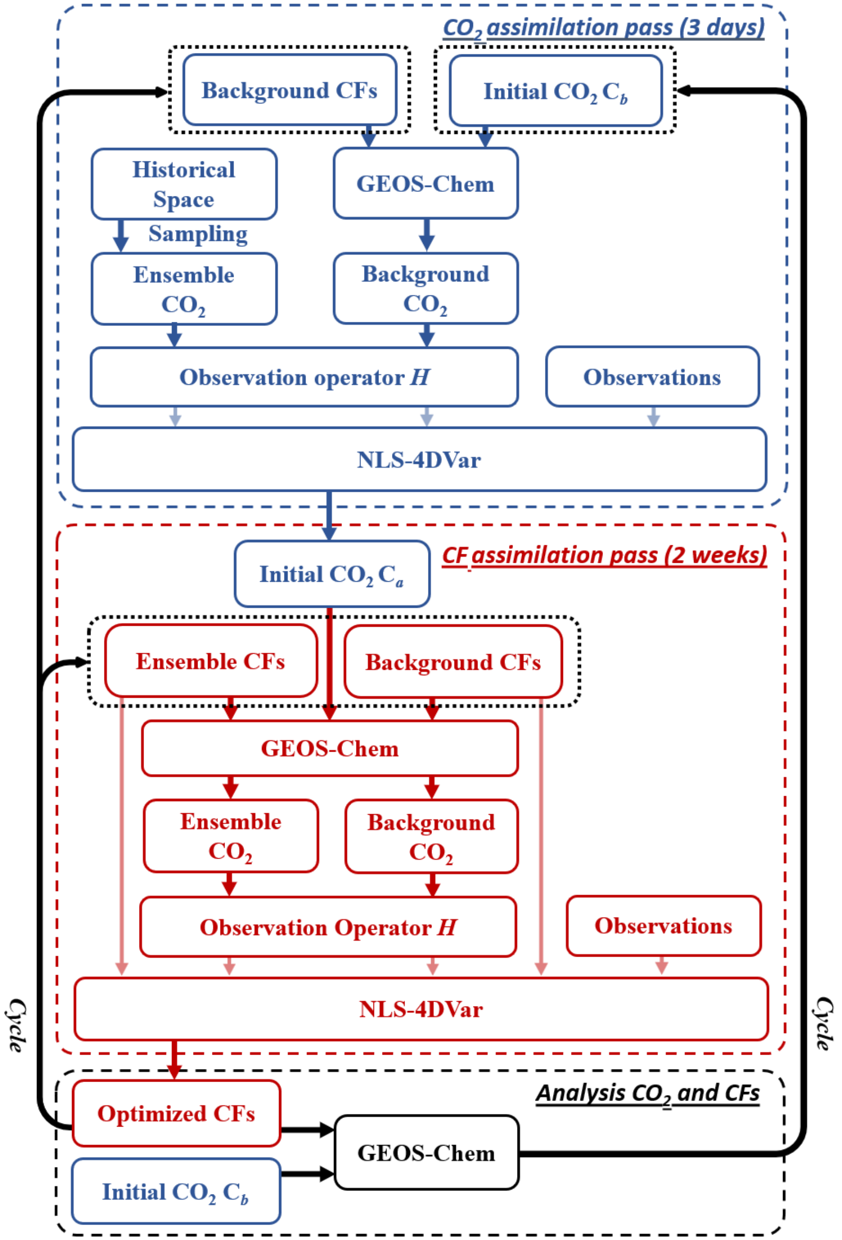

Figure 1. Dual-pass Tan-Tracker (v1) assimilation system framework.

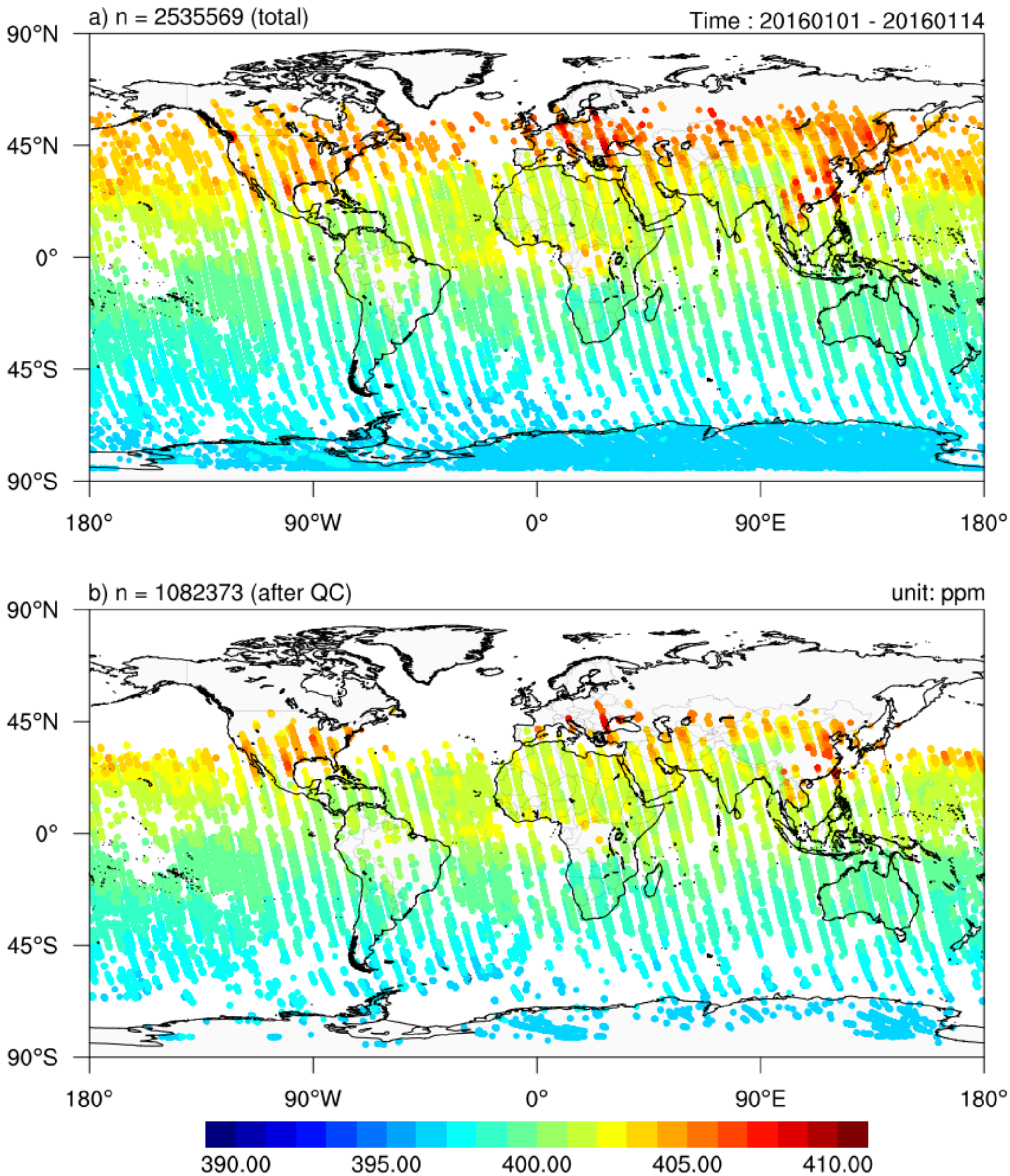

Figure 2. Spatial distribution of artificial observations $X_{CO_2,O}$ before and after quality control in the first window of the flux assimilation pass.

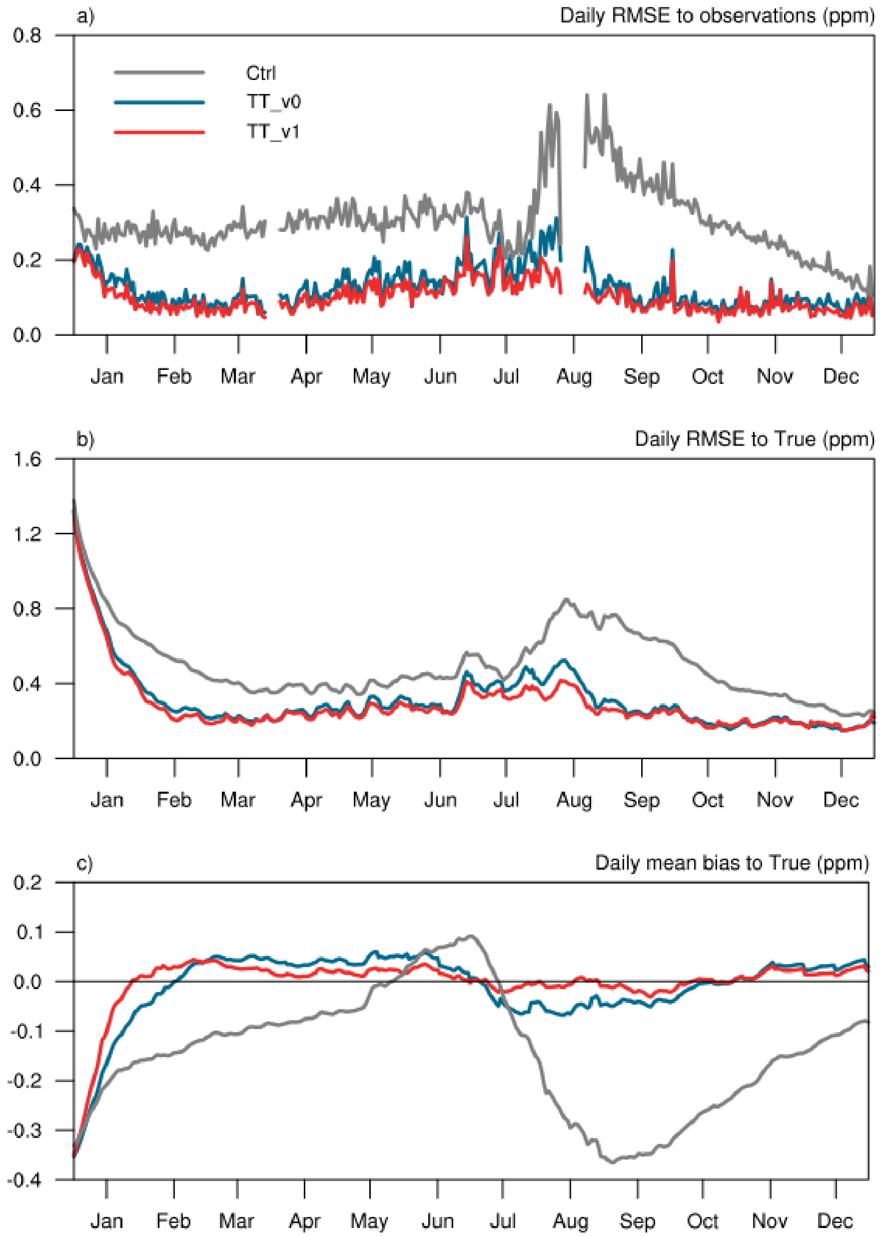

Figure 3. CO_2 assimilation results: a. daily root-mean-square error (RMSE) between assimilation/simulation and artificial observations $X_{CO_2,O}$; b. daily RMSE between assimilation/simulation and *True*; c. daily mean bias between assimilation/simulation and *True*.

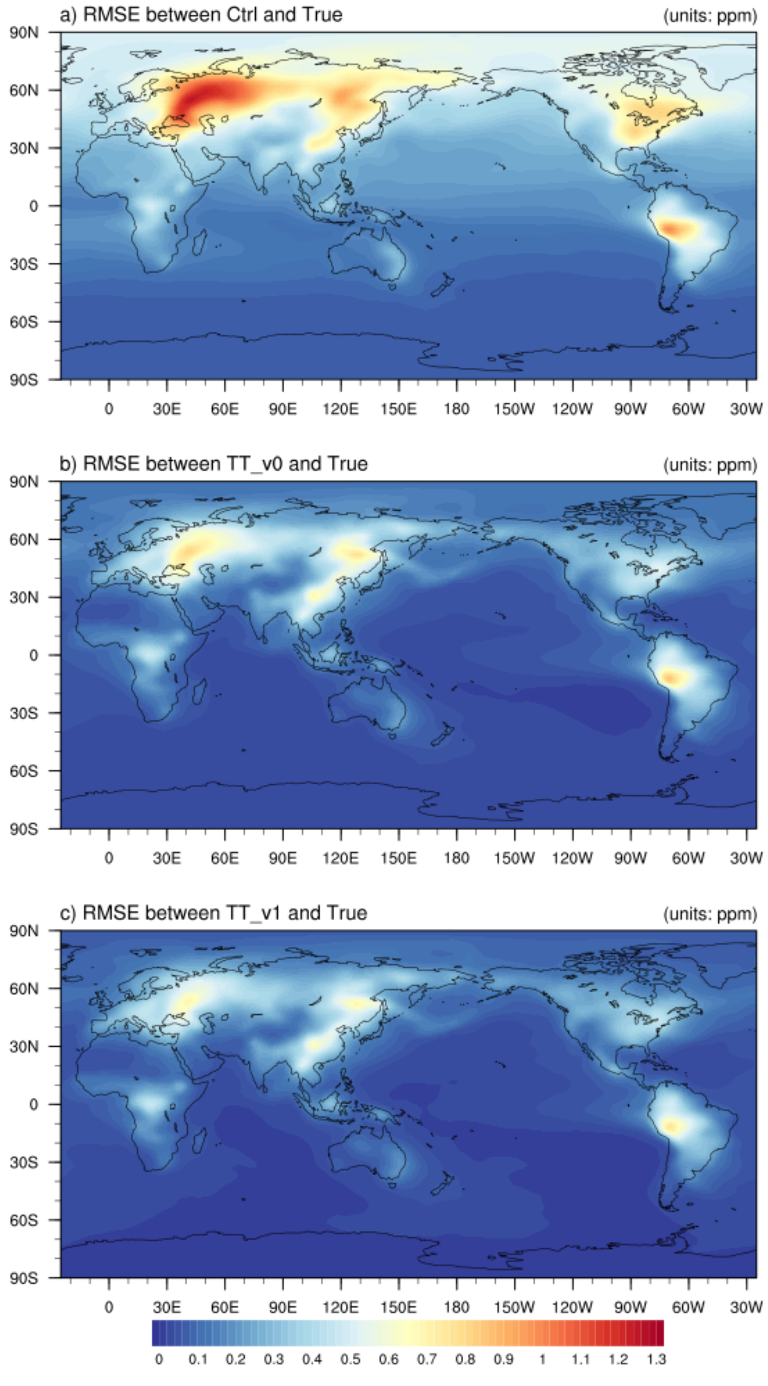

Figure 4. Root-mean-square error (RMSE) spatial distribution of vertical-averaged CO_2 concentration grid time series. RMSE between a. *Ctrl* and *True*; b. *TT_v0* and *True*; c. *TT_v1* and *True*.

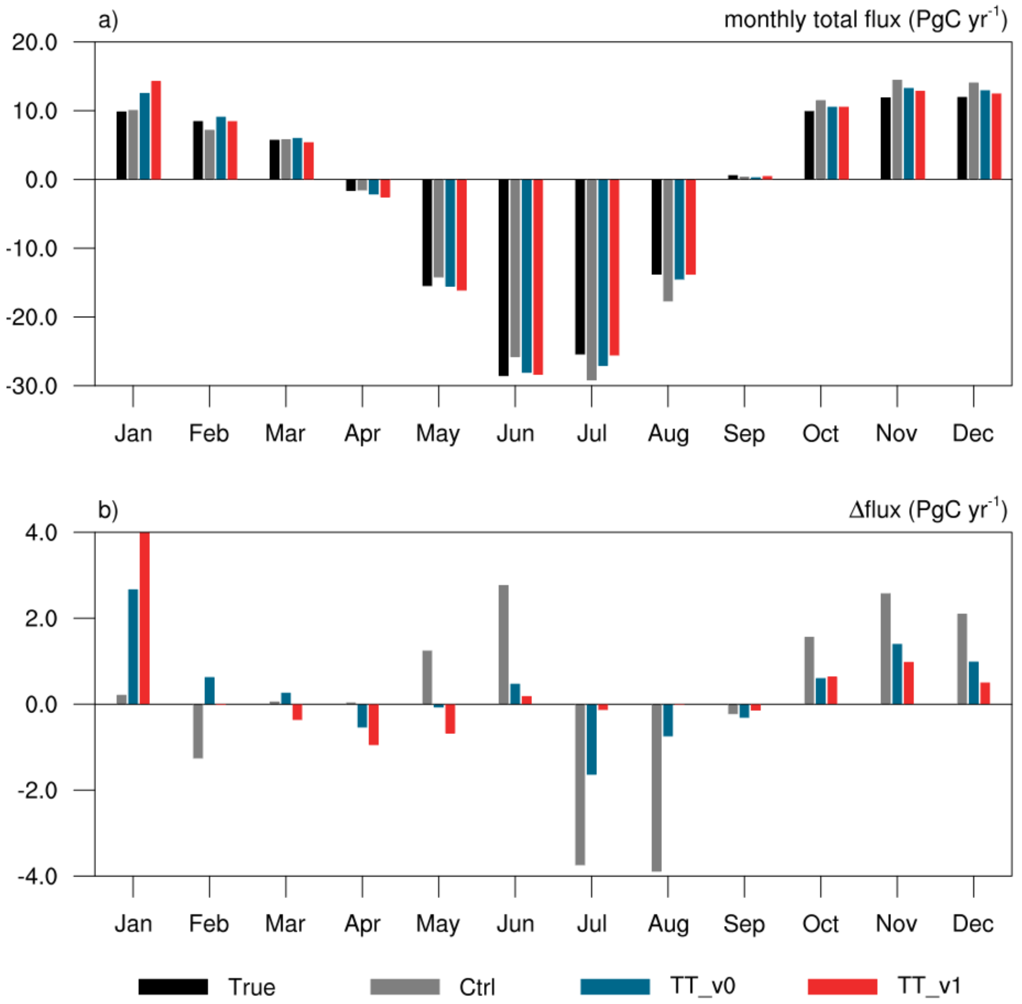

Figure 5. Time series simulation/assimilation results of the monthly global total ecosystem and ocean flux and their deviation from the truth *True*: a. monthly total flux; b. monthly delta flux.

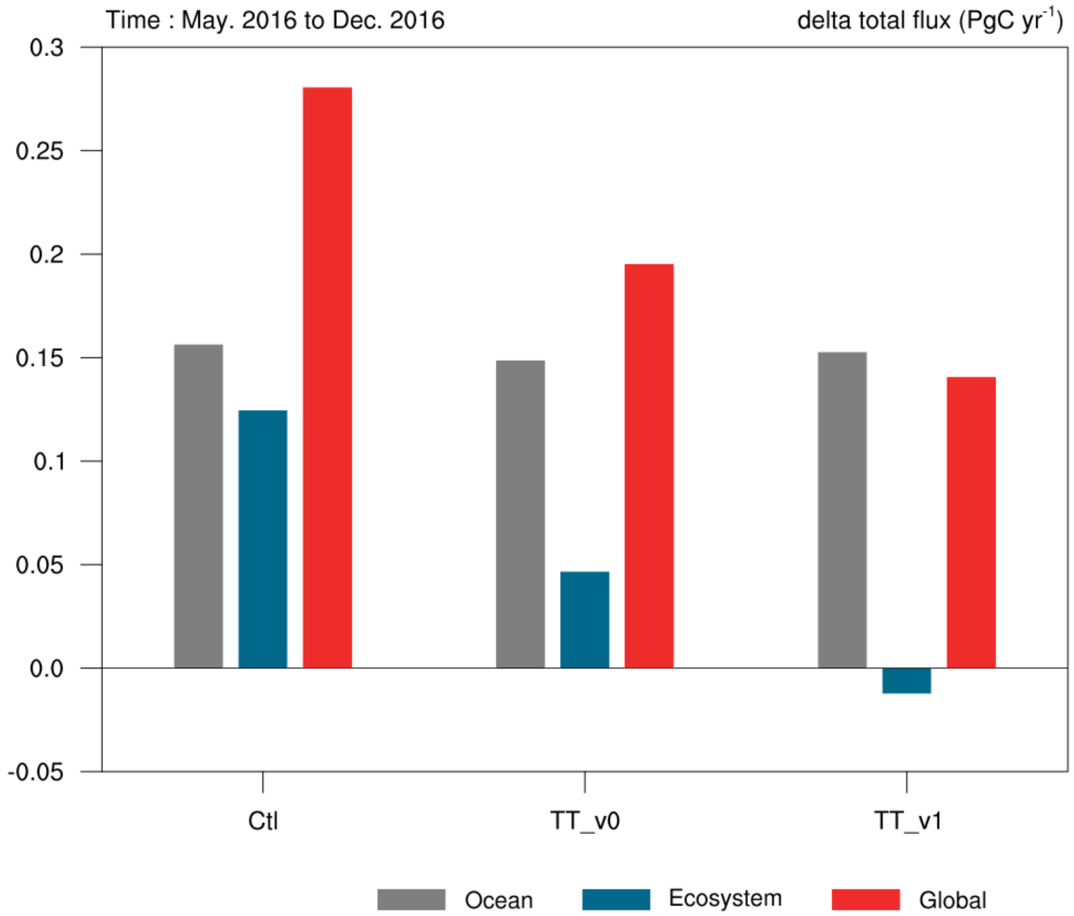

Figure 6. Total flux from May to December and its deviation from *True*.

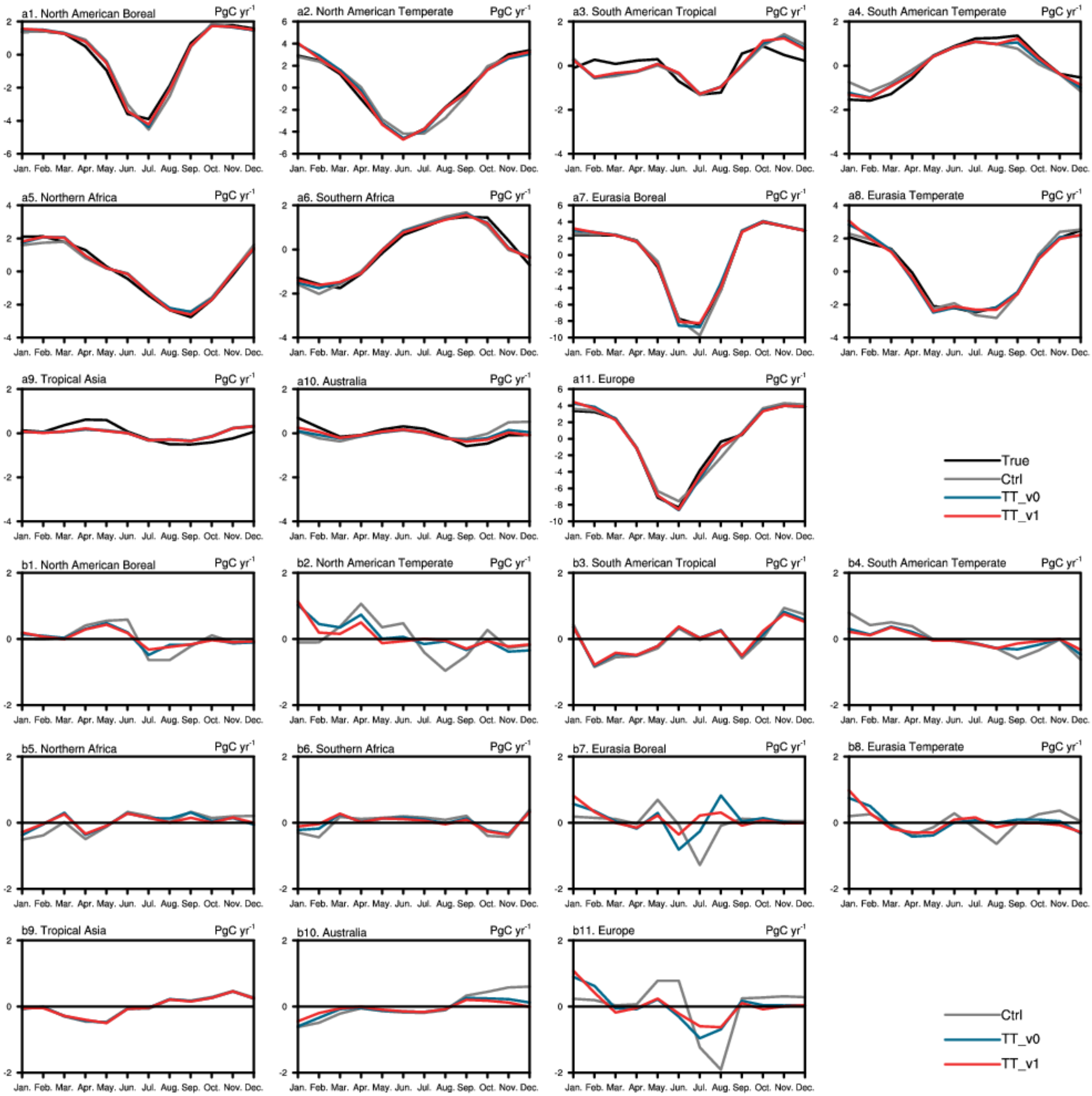

Figure 7. Monthly total flux of 11 land regions of TransCom "super-regions" and its deviation from *True*: a. flux of each region; b. deviation from *True*.

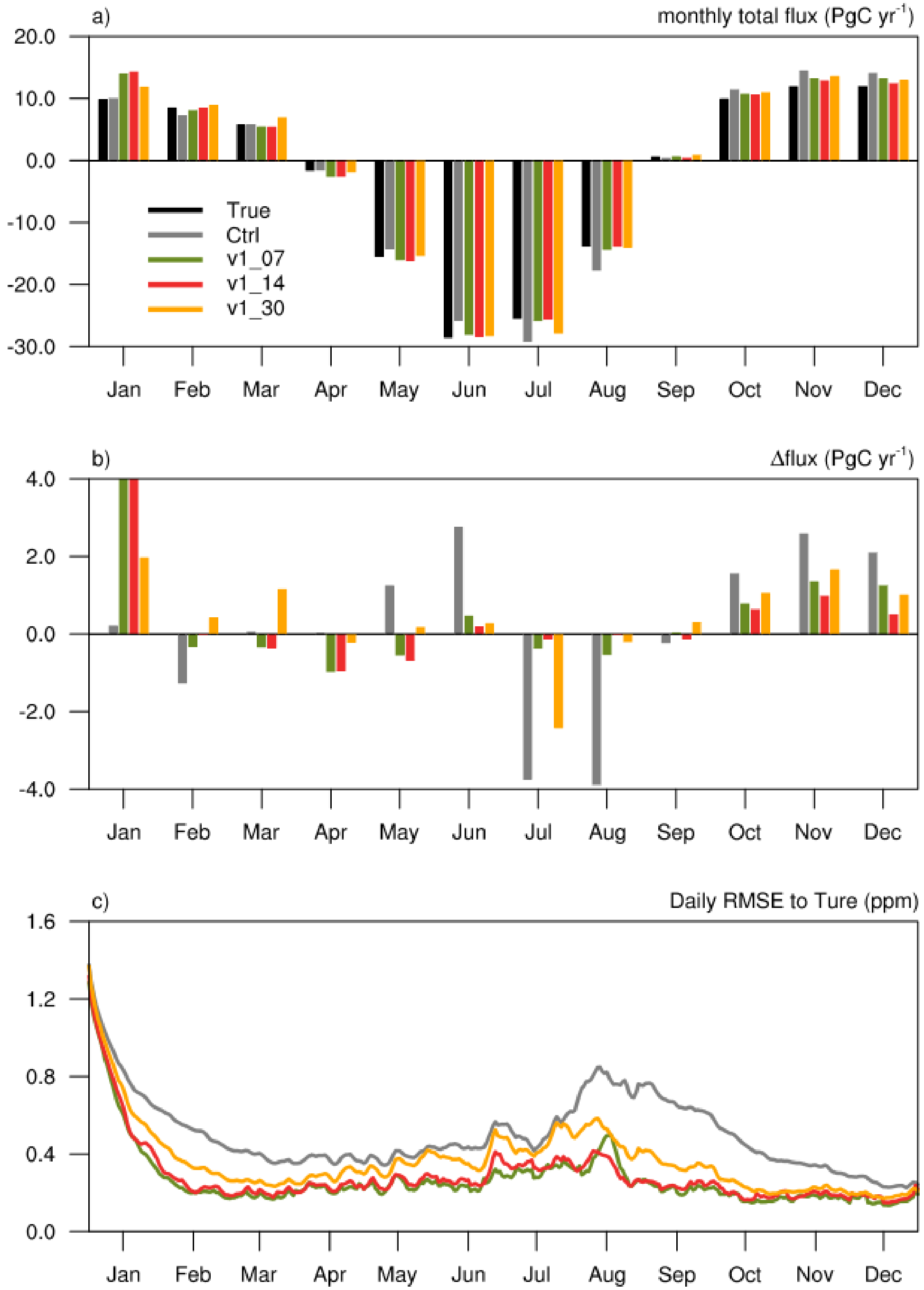

Figure 8. Window length sensitivity experiment results: a. monthly total flux; b. monthly total flux deviation; c. daily root-mean-square error (RMSE) of CO_2 concentration between the simulation/assimilation results and *True*.

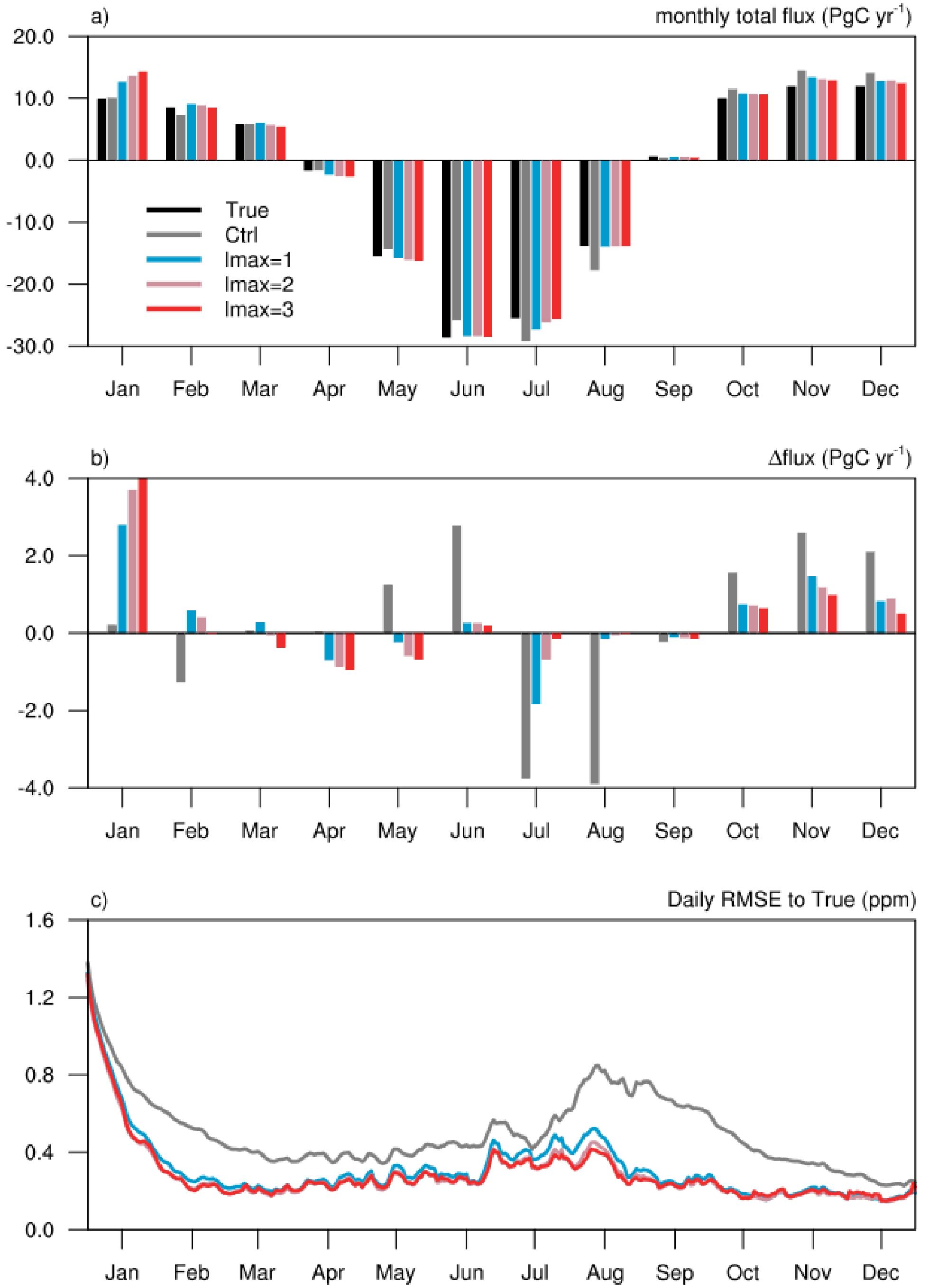

Figure 9. Maximum Nonlinear Least Squares Four-dimensional Variational Data Assimilation algorithm (NLS-4DVar) iteration sensitivity experimental results: a. monthly total flux; b. monthly total flux deviation; c. daily root-mean-square error (RMSE) of CO_2 concentration between the simulation/assimilation results and *True*.

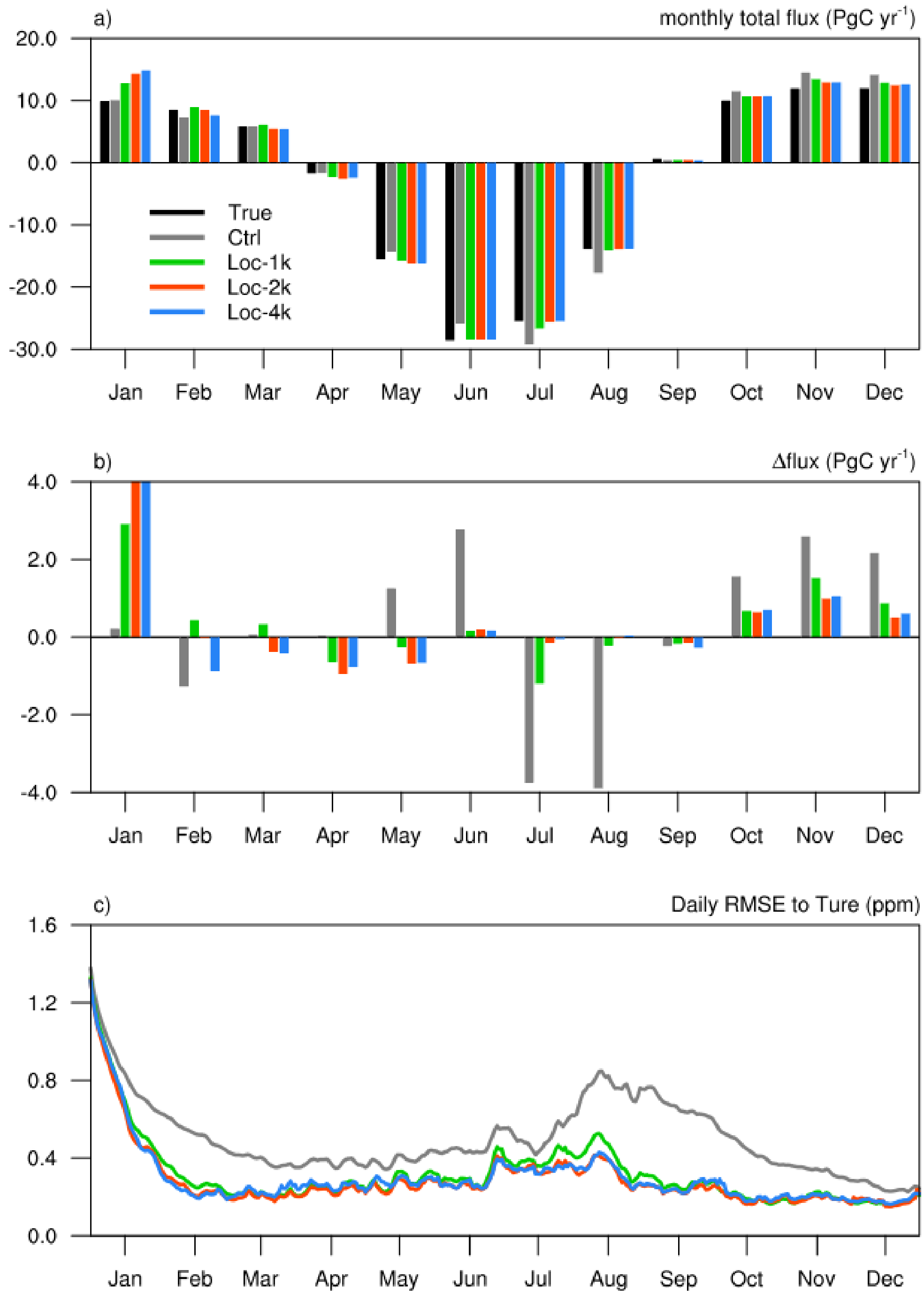

Figure 10. Localization radius sensitivity experiment results: a. monthly total flux; b. monthly total flux deviation; c. daily root-mean-square error (RMSE) of CO_2 concentration between the simulation/assimilation results and *True*.

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see:

http://www.textcheck.com/certificate/0ZjjbP