

The Chinese carbon cycle data-assimilation system (Tan-Tracker)

Xiangjun Tian · Zhenghui Xie · Zhaonan Cai ·
Yi Liu · Yu Fu · Huifang Zhang

Received: 12 September 2013 / Accepted: 13 November 2013 / Published online: 15 March 2014
© Science China Press and Springer-Verlag Berlin Heidelberg 2014

Abstract In this study, the Chinese carbon cycle data-assimilation system Tan-Tracker is developed based on the atmospheric chemical transport model (GEOS-Chem) platform. Tan-Tracker is a dual-pass data-assimilation system in which both CO₂ concentrations and CO₂ fluxes are simultaneously assimilated from atmospheric observations. It has several advantages, including its advanced data-assimilation method, its highly efficient computing performance, and its simultaneous assimilation of CO₂ concentrations and CO₂ fluxes. Preliminary observing system simulation experiments demonstrate its robust performance with high assimilation precision, making full use of observations. The Tan-Tracker system can only assimilate in situ observations for the moment. In the future, we hope to extend Tan-Tracker with functions for

using satellite measurements, which will form the quasi-operational Chinese carbon cycle data-assimilation system.

Keywords Tan-Tracker · CO₂ surface flux · Dual-pass assimilation

1 Introduction

Rapid developments in the global ground- and aircraft-based observation network are leading to new advances in carbon cycle studies. These advances are being further accelerated by the addition of space-borne CO₂ observations (i.e., X_{CO₂}) from several carbon-observing satellites (e.g., SCIAMACHY [1], GOSAT [2], the NASA orbiting carbon observatory (OCO), and the Chinese Tan-SAT, to be launched in 2015). Carbon-observing satellites could provide near-continuous X_{CO₂} observations globally, significantly enhancing our ability to measure atmospheric CO₂. Carbon cycle data assimilation usually adopts the “top-down” method to determine CO₂ surface fluxes (CFs) [3], which makes measurements of atmospheric CO₂ concentrations the constraining factor.

Most previous atmospheric transport inversions of CO₂ aimed to solve a problem with several years (~10–20) of monthly fluxes, at a limited number (22–100) of large regions [4–9]. Those inversion results are so coarse that they can only describe the variations in CFs at the continental and seasonal scales, but not at finer scales.

More recently, substantial efforts (e.g., a geostatistical inversion presented by Michalak et al. [10] and the grid-scale inversions of Kaminski et al. [11], Houweling et al. [12], Rödenbeck et al. [13], and Peylin et al. [14]) have been devoted to improving the development of CF inversion or assimilation. However, high computing costs have

SPECIAL TOPIC: Greenhouse Gas Observation From Space: Theory and Application

X. Tian
ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

X. Tian · Z. Xie
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

Z. Cai (✉) · Y. Liu
LAGEO, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
e-mail: caizhaonan@mail.iap.ac.cn

Y. Fu
Climate Change Research Center (CCRC), Chinese Academy of Sciences, Beijing 100029, China

H. Zhang
Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

limited further developments, and wider applications of the aforementioned inversion methods. Particularly, the most computationally expensive step in these inversions is establishing the linear relationship (i.e., the so-called “observation operator”) between the unknown CFs and atmospheric CO₂ observations. Multiple simulations with expensive tracer transport models are usually required in this step, whose maximum iteration is equal to the number of unknown fluxes, or the number of observations, depending on which is smaller. Meanwhile, these inversions have shown heavy dependences on the adjoining tracer transport models, resulting in their rather poor transportability. Furthermore, the disadvantages of these inversion methods are becoming more apparent with the growing number and variety of observations and the need for greater accuracy in CF inversion results.

The status quo has gradually shifted since the advancement of data assimilation from numerical weather prediction (NWP) was introduced into the field of carbon cycle studies. The field has moved from assimilation of trace gas concentrations [15–20] to trace gas flux inversion [21–23]. Peters et al. [24, 25] introduced a data-assimilation system (hereafter referred to as CarbonTracker) based on the ensemble square root filter (EnSRF) method. In CarbonTracker, the atmospheric transport model is used to link the unknown CFs with the atmospheric CO₂ observations. Furthermore, the CFs integration is supposed to follow a simple, persistence model (i.e., $M = \mathbf{I}$, where \mathbf{I} is the identity matrix). Unfortunately, the uncertainty of the initial CO₂ concentration fields is completely ignored in CarbonTracker. Recently, Kang et al. [26, 27] began to simultaneously assimilate CFs and atmospheric CO₂ concentrations by means of the Local Ensemble Transform Kalman Filter. In their studies, the CFs are actually treated as the model (i.e., the CTM) parameters to constitute a state-parameter augmented vector together with the CO₂ concentrations. Feng et al. [28] used the revised, unbiased Ensemble Transform Kalman Filter (ETKF) algorithm (referred to as ETKF-CDAS) to estimate CFs from spaceborne CO₂ dry-air mole fraction observations. Similarly, the impacts of the CO₂ concentration uncertainty on CF inversion have not yet been appropriately taken into account in ETKF-CDAS.

It is thus necessary, although difficult, to eliminate the impacts of CO₂ concentration uncertainty on CF assimilation. However, this issue could be addressed using the strategy of dual-pass assimilation. To this end, the Chinese carbon data assimilation data-assimilation system, named Tan-Tracker was constructed by incorporating a dual-pass assimilation framework into the GEOS-Chem atmospheric chemical transport model (<http://acmg.seas.harvard.edu/geos/>). An advanced hybrid assimilation method (i.e., PODen4DVar [29–32]) is adopted as the core assimilation

algorithm for Tan-Tracker, and the whole assimilation is divided into two assimilation passes: the CO₂ concentration assimilation pass and the CF assimilation pass. In the CO₂ concentration assimilation pass, the GEOS-Chem model plays the role of the forecast model, while the CO₂ concentration is the prognostic variable, designed to be assimilated through the measurements of CO₂. Simultaneously (at the same time step), the CFs are also optimized with atmospheric CO₂ observations in the CF assimilation pass, where an identity operator acts as the CF dynamical model, and the GEOS-Chem model acts as the major contribution to the observation operator for linking the CFs with CO₂ observations. Some of the most recent developments of Tan-Tracker, and its future development planning are presented in this paper.

2 The Chinese carbon data-assimilation system (Tan-Tracker)

A flowchart of the Tan-Tracker dual-pass assimilation system is presented in Fig. 1. Tan-Tracker begins from an ensemble of N CFs (i.e., the net CO₂ surface exchanges) $F_{i,g}$ ($i = 1, \dots, N$)

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

where $\lambda_{i,g}$ represents a set of linear scaling factors [24, 25] for each day and each grid (g) to be assimilated in the CF assimilation pass, and F_g^* represents the initial estimated values. Subsequently, the CTM (GEOS-Chem) integrates and produces 3D CO₂ concentration profiles ($\mathbf{U}_{m,i}$) over the assimilation window (=the optimized window + the lag-window + the observational window; see the lower panel in of Fig. 1) N times, derived by the ensemble of CFs ($F_{i,g}(x, y, t)$) from the same initial background CO₂ concentration field. By applying the observation operator (H_k) to the modeled CO₂ concentrations ($\mathbf{U}_{m,i}$), we obtain the simulated observations (\mathbf{U}_m^o) as

$$\mathbf{U}_{m,i}^o = H_k(\mathbf{U}_{m,i}), \quad (2)$$

where k ($k = 1, \dots, S$) denotes the observational time (daily), and S is the total observational time in the observational window. The whole assimilation flow is divided into two passes: the CO₂ concentration assimilation pass and the CF assimilation pass. The PODen4DVar approach is the core algorithm for both assimilation passes. In the CO₂ concentration assimilation pass, the background CO₂ concentration field (\mathbf{U}_b), the simulated observations (\mathbf{U}_m^o), the ensemble CO₂ simulations (\mathbf{U}_m), and the real CO₂ measurements (\mathbf{U}_o) are the basic input variables. In this pass, the CTM (GEOS-Chem) is the forecast model, and the final optimized variable is the 3D CO₂

concentration. In the CF assimilation pass, the background CF (F_b), the ensemble CFs (F_i), the background CO₂ concentration fields (U_b), the simulated observations (U_m^o), and the real CO₂ measurements (U_o) are input to the PODEn4DVar assimilation processor, which yields the assimilated linear scaling factor (λ_a), and thus the assimilated CFs (F_a). In the CF assimilation pass, the linear scaling factor (λ) is viewed as the prognostic variable derived by a simple form of persistence forecasting,

$$M = I, \tag{3}$$

where I is the identity matrix. This CF persistence forecasting model (3) follows Peters et al. [24, 25] and assumes that the background CFs for one time step are equal, optimized CFs of the previous time step. In actual implementations, the following dynamical model (4) is applied to the linear scaling factors λ ,

$$\lambda_b(t + 1) = \frac{1}{L_o} \sum_{i=1}^{L_o} \lambda_{a,i}, \tag{4}$$

which is similar to the simple persistence forecasting model (3), but represents smoothing of the optimized window [24, 25]. Here, subscript a refers to analyzed quantities in the optimized window from the last assimilation cycle, and L_o is the length of the optimized window.

In summary, Tan-Tracker works as follows: The CFs (and their perturbations) are first integrated through $F(x, y, t + 1) = \lambda(t + 1)F_b(x, y, t + 1)$ based on the updated λ (and its perturbations) from the previous assimilation cycle, which forces the CTM to produce the ensemble CO₂ concentration forecasts over the assimilation window. These forecasts are then input into the CO₂ concentration assimilation pass to conduct the CO₂ concentration assimilation process, where the usual observation operator (e.g., the interpolation function to interpolate the model gridded variables to the in situ observations) compares the simulated CO₂ concentrations with the observed according to the 4DVar cost function. In this pass, the CO₂ concentrations are assimilated for initializing the next assimilation cycle. Meanwhile, the scaling factors λ in the optimized window are optimized through the CF assimilation pass, which will be used for the next assimilation cycle through (4). In the CF assimilation pass, the simple persistence forecasting (3) for the linear scaling factor λ is treated as the forecast model, and the CTM acts as the major contribution to the observation operator. It should be noted that CarbonTracker essentially includes only the function of the CF assimilation pass in Tan-Tracker, and the uncertainty of the initial CO₂ concentration fields is not fully taken into account [24]. Particularly, the local implementation of the PODEn4DVar assimilation

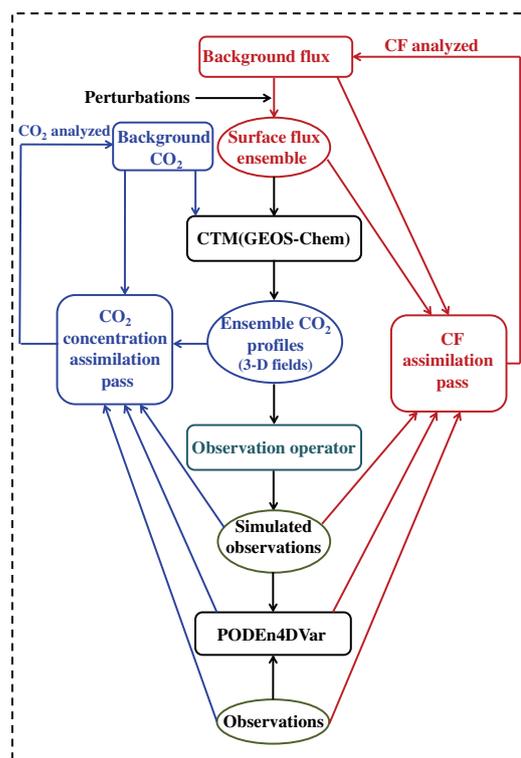


Fig. 1 Flowchart of the Chinese carbon cycle data-assimilation system

processor [31, 32] (<http://dx.doi.org/10.3402/tellusa.v64i0.18375>) and the elaborate sampling strategy proposed by Wang et al. [33] and adopted in Tan-Tracker lead to its highly efficient computing performance.

3 Preliminary evaluations of Tan-Tracker

Observing system simulation experiments (OSSEs) are a useful tool for the evaluation of a proposed data-assimilation system (or approach). Compared with the usual real-data-assimilation experiments, the OSSEs can be completely designed and controlled by the user. In addition to the “observational” data for assimilation, the “real” or “true” results can act as a standard to quantitatively analyze and evaluate the assimilation results. Thus, for a new or proposed data-assimilation system (or method), OSSEs are an ideal method to conduct evaluations.

We employ the GEOS-Chem atmospheric transport model as the forecast model for Tan-Tracker and assume the default surface CO₂ fluxes, output by GEOS-Chem, as the true CF series (F_{True}). The background CF series F_b are set to $1.8F_{True}$. Using F_{True} as a forcing, a 2-year (from Jan 1, 2008 to Jan 1, 2010) GEOS-Chem model spin-up at a resolution of 2° latitude × 2.5° longitude is implemented to reach an equilibrium state, which is then used to initiate

all the other model runs involved in this study. Following this, we integrate the GEOS-Chem model using F_{True} from January 1, 2010 to December 31, 2010 (i.e., the assimilation period) to obtain the “true” CO_2 concentration simulations. The atmospheric CO_2 “observations” are then generated by sampling the daily “true” CO_2 concentrations, adding small random noise through the 136 observational sites used in this study. The assimilation process is initialized by the GEOS-Chem model with the background CF series $F_b (= 1.8F_{\text{True}})$ and conducted continuously by assimilating the daily pseudo-observations throughout the assimilation period. The background simulation is also conducted by the GEOS-Chem model with the background CF series over the same period.

The time series of daily global mean fluxes and CO_2 concentrations from the background simulations (referred to as Sim), the Tan-Tracker (referred to as TT) assimilations, and the “true” simulations are shown in Fig. 2. Naturally, since the background CF series F_b are set to $1.8F_{\text{True}}$, the background simulations deviate from the “true” CF simulations (Fig. 2a). Correspondingly large gaps between the background and “true” simulations are observed in the CO_2 simulation series, with a maximum amplitude of ~ 2 ppm. In contrast, with observational atmospheric CO_2 assimilated continually into the evolutions of both the CO_2 concentrations and CFs, the TT-assimilated errors are quickly reduced during the first few months (about 3 months).

Because both the CO_2 concentrations and fluxes are simultaneously assimilated through their own state assimilation passes, under the dual-pass assimilation framework, the uncertainty of the initial CO_2 concentrations with respect to the CO_2 evolution, during the assimilation window, could be largely eliminated. It is mostly likely for this reason that Tan-Tracker works well throughout the whole assimilation period, especially after the first few months (about 3 months) of its spin-up. The TT assimilation curve agreed well with the “true” one for both the CO_2 concentrations, and the CFs. The transfer process of CO_2 concentration is quite slow, making it difficult to eliminate the uncertainty of the initial CO_2 fields by only adjusting the CFs through CF assimilation (in the usual carbon cycle data-assimilation systems such as CarbonTracker). This uncertainty would in turn evolve and accumulate increasingly with the assimilation, ultimately affecting the CF assimilation.

The robustness of our Tan-Tracker data assimilation is further verified by the root mean-square (RMS) errors for the daily gridded (2° latitude \times 2.5° longitude) TT-assimilated CFs from Jul 1 to Dec 31, 2010, as shown in Fig. 3a. The corresponding RMS errors for the TT-assimilated CO_2 concentrations are also shown in Fig. 3b. Encouragingly, the RMS errors can be controlled in a limited range for both the

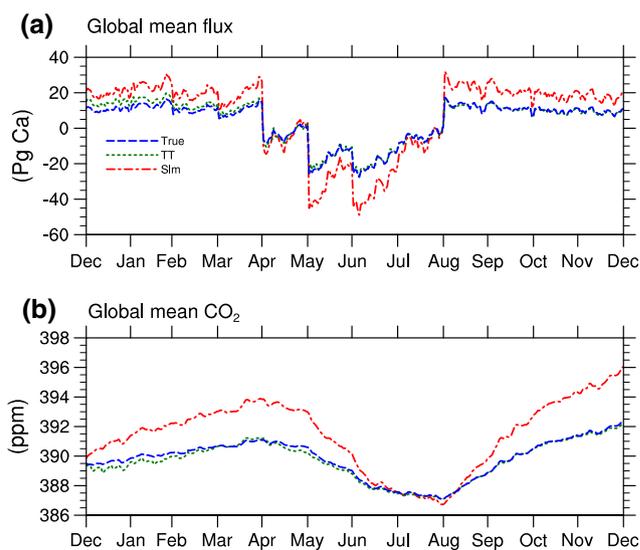


Fig. 2 Time series of the global mean. **a** Fluxes and **b** CO_2 concentrations from “true” data, simulations, and TT (Tan-Tracker) assimilations from Jan 1 to Dec 31, 2010

CFs and the CO_2 concentrations. Only relatively larger RMS errors appeared in a very small area in the central parts of South America (Fig. 3). Such good results over the globe demonstrate the strength of the Tan-Tracker system for estimating CO_2 concentrations and CFs simultaneously from atmospheric observations. Moreover, the application of the advanced hybrid assimilation approach (i.e., PODEn4DVar) would make a positive contribution to its excellent performance [29–32, 34, 35].

4 Conclusions and the development planning for Tan-Tracker

In this paper, the Chinese carbon cycle data-assimilation system Tan-Tracker was introduced. Tan-Tracker is a dual-pass assimilation system based on an advanced hybrid data-assimilation system (i.e., PODEn4DVar). Unlike the usual dual-pass frameworks consisting of a state assimilation pass and a parameter calibration pass, both optimization passes in Tan-Tracker are state assimilation passes, one for CO_2 concentration assimilation; and the other for CF assimilation. In the CO_2 concentration assimilation pass, the GEOS-Chem model plays the role of the forecast model, and CO_2 concentration is assimilated through the measurements of atmospheric CO_2 . Simultaneously, the CFs are also assimilated in the CF assimilation pass, in which an identity operator is chosen as the CF dynamical model, while the GEOS-Chem model provides a major contribution to the observation operator, linking the CFs with CO_2 observations. The robustness of our Tan-Tracker has been comprehensively

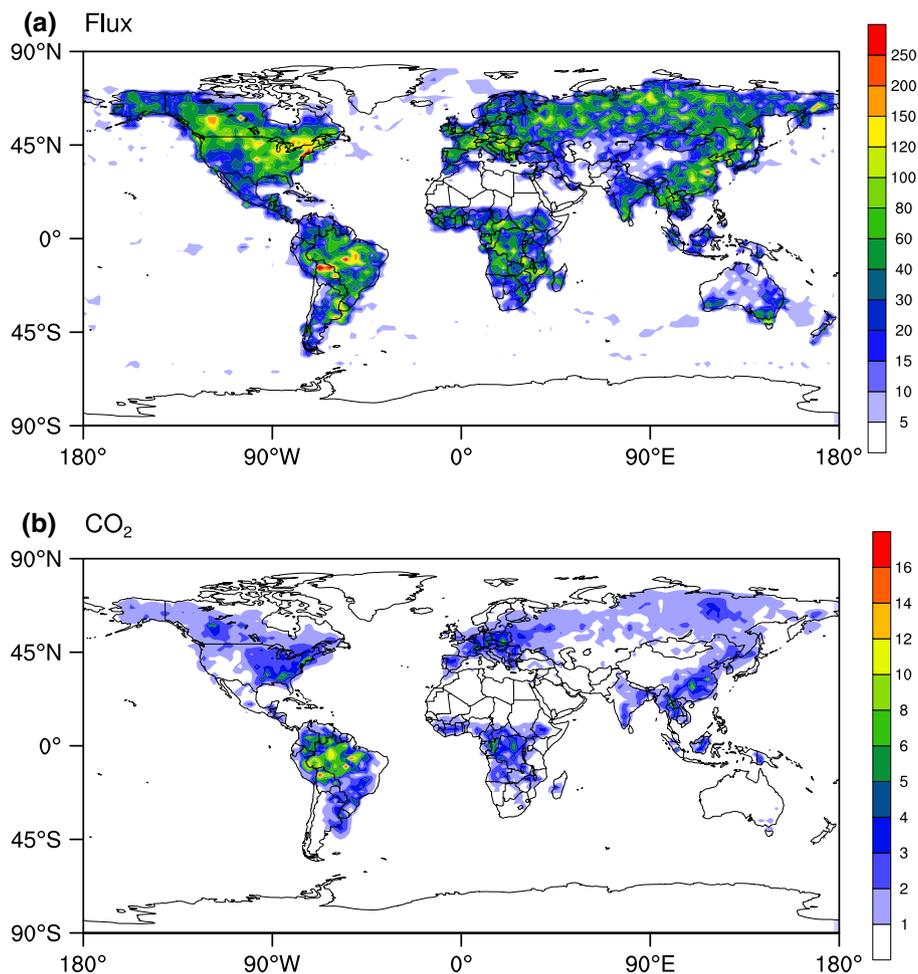


Fig. 3 RMS errors for the daily, gridded (2° latitude \times 2.5° longitude) TT-assimilated. **a** CFs (10^{-11} kg cm^{-2} s^{-1}) and **b** CO_2 concentrations (ppm) during the period from Jul 1 to Dec 31, 2010

verified in this study through a series of carefully designed OSSEs. It should be noted that our Tan-Tracker system is one of a series in preparation for the launch of the Chinese CO_2 observation satellite TanSat. According to the plan, Tan-Tracker will be moving ahead as follows:

Firstly, we need to include the assimilation of XCO_2 in Tan-Tracker, and thus equip it with a function to utilize satellite measurements.

Secondly, integrated assimilation for both the in situ and satellite observations needs to be achieved in Tan-Tracker.

Thirdly, a series of real-data experiments for Tan-Tracker should be carried out to critically assess its real application performance using the in situ observations and Greenhouse Gases Observing SATellite (GOSAT) measurements of XCO_2 .

Fourthly, the Earth simulator, self-developed by the Chinese Academy of Sciences, is expected to be incorporated into Tan-Tracker instead of the GEOS-Chem atmospheric transport model, resulting in a fully home-grown Chinese carbon cycle data-assimilation system.

Finally, the final version of Tan-Tracker should jointly assimilate meteorological and carbon observations.

Acknowledgments We thank Feng Chen for constructive discussions on the programming of Tan-Tracker. This work was supported by the Strategic Priority Research Program-Climatic Change: Carbon Budget and Relevant Issues (XDA05040200), the National High Technology Research and Development Program of China (Grant No. 2013AA122002), the National Natural Science Foundation of China (41075076), and the Knowledge Innovation Program of the Chinese Academy of Sciences (KZCX2-EW-QN207).

References

1. Bovensmann H, Burrows J, Buchwitz M et al (1999) SCIAM-ACHY: mission objectives and measurement modes. *J Atmos Sci* 56:127–150
2. Maksyutov S, Kadyrov N, Nakatsuka Y et al (2008) Projected impact of the GOSAT observations on regional CO_2 flux estimations as a function of total retrieval error. *J Remote Sensing Soc Jpn* 28:190–197
3. Enting I (2002) Inverse problems in atmospheric constituent transport. Cambridge University Press, Cambridge

4. Rayner PJ, Law RM (1999) The interannual variability of the global carbon cycle. *Tellus Ser B* 51:210–212
5. Bousquet P, Peylin P, Ciais P et al (2000) Regional changes in carbon dioxide fluxes of land and oceans since 1980. *Science* 290:1342–1346
6. Peylin P, Baker D, Sarmiento J et al (2000) Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO₂ data. *J Geophys Res* 107:4385
7. Gurney KR, Law RM, Denning AS et al (2002) Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models. *Nature* 415:626–630
8. Law RM, Chen YH, Gurney KR (2003) TransCom 3 CO₂ inversion intercomparison: 2. Sensitivity of annual mean results to data choices. *Tellus Ser B* 55:580–595
9. Maksyutov S, Machida T, Mukai H et al (2003) Effect of recent observations on Asian CO₂ flux estimates by transport model inversions. *Tellus Ser B* 55:522–529
10. Michalak AM, Bruhwiler L, Tans PP (2004) A geostatistical approach to surface flux estimation of atmospheric trace gases. *J Geophys Res* 109:D14109
11. Kaminski T, Heimann M, Giering R (1999) A coarse grid three dimensional global inverse model of the atmospheric transport: 2. Inversion of the transport of CO₂ in the 1980s. *J Geophys Res* 104:18555–18581
12. Houweling S, Kaminski T, Dentener F et al (1999) Inverse modeling of methane sources and sinks using the adjoint of a global transport model. *J Geophys Res* 104:26137–26160
13. Rodenbeck C, Houweling S, Gloor M et al (2003) CO₂ flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmos Chem Phys* 3:1919–1964
14. Peylin P, Rayner P, Bousquet P et al (2005) Daily CO₂ flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology. *Atmos Chem Phys Disc* 5:1647–1678
15. Lyster PM, Cohn SE, Menard R et al (1997) Parallel implementation of a Kalman filter for constituent data assimilation. *Mon Weather Rev* 125:1674–1686
16. Miller SM, Snell HE, Moncet JL (1999) Simultaneous retrieval of middle atmospheric temperature and trace gas species volume mixing ratios from Cryogenic Infrared Radiance Instrumentation for Shuttle (CIRRIS 1A). *J Geophys Res* 104:18697
17. Menard R, Cohn SE, Chang LP et al (2000) Assimilation of stratospheric chemical tracer observations using a Kalman filter. part I: formulation. *Mon Weather Rev* 128:2654–2671
18. Khattatov BV, Lamarque JF, Lyjak LV et al (2000) Assimilation of satellite observations of long-lived chemical species in global chemistry transport models. *J Geophys Res* 105:29135–29144
19. Eskes HJ, Van Velthoven PF, Valks PJ et al (2003) Assimilation of GOME total-ozone satellite observations in a three-dimensional tracer-transport model. *Q J R Meteorol Soc* 129:1663–1681
20. Stajner I, Wargan K (2004) Antarctic stratospheric ozone from the assimilation of occultation data. *Geophys Res Lett* 31:L18108
21. Kleiman G, Prinn RG (2000) Measurement and deduction of emissions of trichloroethene, tetrachloroethene, and trichloromethane (chloroform) in the northeastern United States and southeastern Canada. *J Geophys Res* 105:28875–28893
22. Petron G, Granier C, Khattatov B et al (2004) Monthly CO surface sources inventory based on the 2000–2001 MOPITT satellite data. *Geophys Res Lett* 31:L21107
23. Yudin VA, Petron G, Lamarque JF et al (2004) Assimilation of the 2000–2001 CO MOPITT retrievals with optimized surface emissions. *Geophys Res Lett* 31:L20105
24. Peters W, Miller JB, Whitaker J et al (2005) An ensemble data assimilation system to estimate CO₂ surface fluxes from atmospheric trace gas observations. *J Geophys Res* 110:D24304
25. Peters W, Jacobson AR, Sweeney C et al (2007) An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proc Natl Acad Sci USA* 104:18925–18930
26. Kang JS, Kalnay E, Miyoshi T et al (2012) Estimation of surface carbon fluxes with an advanced data assimilation methodology. *J Geophys Res* 117:D24101
27. Kang JS, Kalnay E, Liu J et al (2011) “Variable localization” in an ensemble Kalman filter: application to the carbon cycle data assimilation. *J Geophys Res* 116:D09110
28. Feng L, Palmer PI, Bosch H et al (2009) Estimating surface CO₂ fluxes from space-borne CO₂ dry air mole fraction observations using an ensemble Kalman filter. *Atmos Chem Phys* 9:2619–2633
29. Tian X, Xie Z, Dai A (2008) An ensemble-based explicit four-dimensional variational assimilation method. *J Geophys Res* 113:D21124
30. Tian X, Xie Z, Sun Q (2011) A POD-based ensemble four dimensional variational assimilation method. *Tellus A* 63A:805–816
31. Tian X, Xie Z (2012) Implementations of a square-root ensemble analysis and a hybrid localization into the POD-based ensemble 4DVar. *Tellus A* 64A:18375
32. Tian X (2013) A local implementation of the POD-based ensemble 4DVar with R-localization. *Atmos Oceanic Sci Lett* (in press)
33. Wang B, Liu J, Wang S et al (2010) An economical approach to four-dimensional variational data assimilation. *Adv Atmos Sci* 27:715–727
34. Tian X, Xie Z, Dai A et al (2009) A dual-pass variational data assimilation framework for estimating soil moisture profiles from AMSR-E microwave brightness temperature. *J Geophys Res* 114:D16102
35. Tian X, Xie Z, Dai A et al (2010) A microwave land data assimilation system: scheme and preliminary evaluation over China. *J Geophys Res* 115:D21113