Real-data assimilation experiment with a joint data assimilation system: assimilating carbon dioxide mole fraction measurements from the Greenhouse gases Observing Satellite

HAN Ruia,b, TIAN Xiang-Junb, FU Yuc and CAI Zhao-Nand

Abstract

The performance of a joint data assimilation system (Tan-Tracker), which is based on the PODEn4Dvar assimilation method, in assimilating Greenhouse gases Observing SATellite (GOSAT) carbon dioxide (CO2) data, was evaluated. Atmospheric 3D CO2 concentrations and CO2 surface fluxes (CFs) from 2010 were simulated using a global chemistry transport model (GEOS-Chem). The Tan-Tracker system used the simulated CO2 concentrations and fluxes as a background field and assimilated the GOSAT column average dry-air mole fraction of CO2 (XCO2) data to optimize CO2 concentrations and CFs in the same assimilation window. Monthly simulated XCO2 (XCO2,Sim) and assimilated XCO2 (XCO2,TT) data retrieved at different satellite scan positions were compared with GOSAT-observed XCO2 (XCO2,Obs) data. The average RMSE between the monthly XCO2,Sim and XCO2,Obs data was significantly (30%) lower than the average RMSE between XCO2,Sim and XCO2,Obs. Specifically, reductions in error were found for the positions of northern Africa (the Sahara), the Indian peninsula, southern Africa, southern North America, and western Australia. The difference between the correlation coefficients of the XCO2,Sim and XCO2,Obs and those of the XCO2,TT and XCO2,Obs was only small. In general, the Tan-Tracker system performed very well after assimilating the GOSAT data.

1. Introduction

Atmospheric carbon dioxide (CO2) is one of the most important greenhouse gases (GHGs) and its increase since pre-industrial times has directly led to global warming (IPCC 2013). The increase in atmospheric CO2 is mainly caused by anthropogenic emissions. Moreover, as a long-lived GHG, atmospheric CO2 can only be partially removed by natural processes, which in turn also contributes to atmospheric CO2 levels. Natural processes, such as the carbon cycle, are sophisticated interactions forming a complicated feedback loop. The key to understanding these interactions is to determine the spatiotemporal distribution of atmospheric CO2 concentrations and land CO2 fluxes (CFs) (Protocol 1997).

There are several methods used to study the spatial and temporal distribution of CO2 concentrations. Of these, surface observation is one of the most important. In the 1980s, the WMO’s Global Atmosphere Watch program was established to monitor the long-lived GHGs that are directly affected by anthropogenic activities. However, there are currently fewer than 200 monitoring stations...
around the world and the records on average are shorter than 30 years. Due to these limitations, it is difficult to analyze spatiotemporal distributions of CO$_2$ using only surface observations (Reuter et al. 2011). Recently, satellite data have become more widely used by researchers due to the increased spatial and temporal resolution of satellite observations from the Greenhouse gases Observing SATellite (GOSAT) and the Orbiting Carbon Observatory 2 (Yokota et al. 2009; Nassar et al. 2010; Cogan et al. 2012). Previous studies have used global or regional Chemistry Transport Models (CTMs) to obtain continuous simulated spatiotemporal features of global atmospheric CO$_2$ concentrations and CFs (Cogan et al. 2012; Chen, Zhù, and Zeng 2013). However, there is still an essential and urgent need to improve model simulation efficiency and accuracy.

Carbon cycle data assimilation systems are promising new tools for precisely simulating atmospheric CO$_2$ concentrations and CFs. These systems tend to yield estimates of CO$_2$ surface flux by combining information from both CTM simulations and atmospheric CO$_2$ observations (Peters et al. 2005; Tian et al. 2014). Peters et al. (2005) developed the Carbon-Tracker data assimilation system, which was coupled to the Tracer Model, version 5, using an ensemble square root filter to assimilate surface CO$_2$ concentration observations. The CF inversion results from the Carbon-Tracker system were consistent with the majority of carbon inventories reported by the first North American State of the Carbon Cycle Report (Peters et al. 2005). Zhang et al. (2014) assimilated CONTRAIL data with the Carbon-Tracker system and dramatically improved the accuracy of simulated concentrations while reducing the uncertainty of CFs in Asia by 20%. However, the accuracy of this simulation was largely dependent on the algorithm, the dynamical model, and surface observations. Due to the limitations of the algorithm, the deficiency of a suitable dynamical model and insufficient surface observations, attempts to construct an efficient and accurate assimilation system still represent a challenging task.

Tian et al. (2014) reported a new CF data assimilation system (Tan-Tracker) that was developed by incorporating a joint PODeN4DVar (Tian, Xie, and Sun 2011) assimilation framework into the GEOS-Chem model (V9-01-03; Suntharalingam et al. 2004; Nassar et al. 2010). In their study, an identity operator was chosen as the CF dynamical model to describe CF evolution, and then this CF dynamical model was utilized to create an augmented dynamical model using the GEOS-Chem atmospheric transport model. Therefore, the large-scale state vector of both CFs and CO$_2$ concentrations is set as the prognostic variable, which is simultaneously constrained by assimilation of atmospheric CO$_2$ concentration observations. Using this approach, the Tan-Tracker data assimilation system is expected to be powerful and efficient.

In this study, we assimilated column average dry-air mole fraction of CO$_2$ ($X_{CO_2}$) GOSAT observations from January 2010 to October 2010 into the GEOS-Chem model using the Tan-Tracker system, and compared the results with GOSAT observations. Details of the Tan-Tracker data assimilation system are presented next, in Section 2, followed by a description of the experiment and results in Section 3. Section 4 summarizes the conclusions that can be drawn from this study.

## 2. Tan-Tracker data assimilation system

The Tan-Tracker data assimilation system is a joint assimilation system capable of optimizing model states and parameters simultaneously from noisy measurements through the PODeN4DVar approach developed by Tian et al. (2014).

We initialized Tan-Tracker by running the GEOS-Chem twice to generate the assimilation inputs. The background run started from the previous best analysis $C_a$ forced by the prior CFs $F_a^b$ to determine the background CO$_2$ concentration fields $C^b_0$. This was used to prepare the rth background joint vector $(\lambda_b, C_{m, b})^r$, where $\lambda_b$ is the scaling factor between the prior CFs and the updated CFs. The sampling run followed the background run but with a different run length and adopted a 4D moving sampling strategy (Wang et al. 2010) to produce the joint vector ensemble $(\lambda_m^r, C_{m, b})^r$. This is explained in detail in Tian et al. (2014).

The CF dynamical sub-model, together with the CTM (GEOS-Chem), shapes the Tan-Tracker dynamical model:

$$\begin{align*}
M &= \begin{pmatrix}
M_{CF} \\
CTM
\end{pmatrix}
\end{align*}$$

Here, the flux persistence-forecasting model denotes $M_{CF} = I$ ($I$, identity matrix), which is chosen as the CF evolution dynamical sub-model (based on Peters et al. 2005).

The Tan-Tracker system assimilates $X_{CO_2}$ directly using the following observation operator (Equation (2)), which explains the link between the observational variable $X_{CO_2}$ and GEOS-Chem simulated 3D CO$_2$ concentrations (Feng et al. 2009; Tian et al. 2014):

$$X_{CO_2} = X_{CO_2, a} + h^T \mathbf{A} (C_m - C_a),$$

where $h$ is a pressure weighting function, $\mathbf{A}$ is the full averaging kernel matrix, $C_a$ is the prior CO$_2$ profile, $X_{CO_2, a}$ is the associated column mole fraction, and $C_m$ is the model-calculated CO$_2$ profile. We generated ensemble simulated observations, $C^p_m$ and background simulated observations, $C^b_m$, by applying the observation operator $H$. 


to the ensemble CO$_2$ concentration, $C_{en}$ and background CO$_2$ concentration, $C_{bg}$, respectively.

At this point, the background joint vector $(\lambda_{bg}, C_{bg})^T$, the joint vector ensemble $(\lambda_{en}, C_{en})^T$, the background simulated observations ($C_{bg}$), the ensemble simulated observations ($C_{en}$) and the real CO$_2$ concentrations ($C_{obs}$) have been determined and can be inputted into the PODEn4DVar assimilation processor to yield the best analysis joint vector $(\lambda_{a}, C_{a})^T$ and optimized CFs ($F_a = \lambda_a F^a$). The best analysis is the initial condition for the next assimilation cycle.

3. Real-data assimilation experiment with retrieved satellite data ($X_{CO_2}$)

The efficiency and accuracy of the Tan-Tracker data assimilation system was comprehensively evaluated using a well-designed real-data assimilation experiment applied to data from 2010. The setup of the experiment and results are described below.

3.1. Experimental setup

The global 3D chemical transport model GEOS-Chem (version: 9-01-03) was used to simulate the atmospheric CO$_2$ concentration. This version of GEOS-Chem has a horizontal resolution of 2° × 2.5° (latitude × longitude) and 47 hybrid eta levels up to 0.01 hPa, which is driven by GEOS-5 meteorological fields. These are assimilated meteorological data from the GEOS of the NASA Global Modeling and Assimilation Office.

The original GEOS-Chem CO$_2$ simulation was designed by Suntharalingam et al. (2004) and updated by Nassar et al. (2010). The prescribed CO$_2$ fluxes used in the model include monthly fossil fuel burning and cement production CO$_2$ emissions, monthly biomass burning, climatological biofuel burning, monthly ocean exchange, three-hourly biospheric fluxes, annual climatological terrestrial biosphere exchange, chemical production of CO$_2$ from the atmospheric oxidation of other carbon species, and monthly emissions from shipping and aviation CO$_2$ emissions. A detailed description of the basic model input data can be found in Tian et al. (2014).

The experiment began with a two-year spin-up from 1 January 2008, with a globally uniform 3D CO$_2$ concentration field of 383.76 ppm. This was used since the annual mean CO$_2$ concentration in 2007 was 383.76 ppm at Mauna Loa, which is the marine surface site of the NOAA-ESRL (http://www.esrl.noaa.gov/gmd/obop/mlo/). This two-year spin-up is a key step to maintaining mass and energy balance, while also allowing model transport, sources, and sinks to develop global spatial patterns. After the spin-up, the acquired CO$_2$ concentration field is used to drive the real-data assimilation experiment.

The space-borne observations used in our experiment were retrieved from GOSAT, which was launched in 2009. We chose the GOSAT Level 2 data; that is, the $X_{CO_2}$. This can be found in the Atmospheric CO$_2$ Observations from Space (ACOS) data product, version 3.3. We applied the recommended data screening criteria and bias correction technique based on the ‘ACOS Level 2 Standard Product Data User’s Guide’ (http://disc.sci.gsfc.nasa.gov/acdisc/documentation/ACOS_v3.3_DataUsersGuide.pdf).

In order to guarantee high-quality assimilated data, we only retained $X_{CO_2}$ data with standard error deviations of less than 1.0 ppm. We initially performed CO$_2$ simulations from 1 January 2010 to 31 October 2010 without assimilation (referred to as ‘Sim’), and then we performed the CO$_2$ simulation with the Tan-Tracker assimilation system (referred to as ‘TT’) from 1 January 2010 to 31 October 2010. Due to the five-week lag window within the assimilation window, 13 months’ worth of data from November 2009 to November 2010 were used to ensure continuity.

3.2. Results and discussion

The global distribution of GOSAT-retrieved satellite data, $X_{CO_2}$, after quality control, demonstrated obvious spatiotemporal distribution patterns, such as lower concentrations of CO$_2$ at high latitudes in contrast with higher concentrations in equatorial regions. High concentrations of CO$_2$ (approximately 388–391 ppm) were found in the NH, while CO$_2$ concentrations in the SH were lower (approximately 385–388 ppm). The temporal distribution of CO$_2$ concentrations in the NH clearly followed a seasonal cycle, increasing with time and reaching a peak concentration of 388 ppm at the end of April or beginning of May prior to the maturation period that began in June, and then reducing with time to approximately 382 ppm in late autumn before rising again.

To examine the performance of Tan-Tracker, the $X_{CO_2}$ derived from Sim (referred to as $X_{CO_2,Sim}$) and TT (referred to as $X_{CO_2,TT}$) were compared with GOSAT observations. The CO$_2$ concentrations obtained from Sim and TT were converted into $X_{CO_2}$, i.e., $X_{CO_2,Sim}$ and $X_{CO_2,TT}$, by first calculating CO$_2$ profiles that were the same levels as GOSAT data profiles, and then calculating $X_{CO_2}$ using Equation (2). The distributions of $X_{CO_2,Sim}$ and $X_{CO_2,TT}$ at GOSAT satellite geographic coordinates are illustrated in Figures 1 and 2, respectively, which display data from January, April, July, and October 2010. Both of the distributions had spatiotemporal patterns that were comparable to the observed data. As shown in Figure 3, high deviations between the observations, $X_{CO_2,Obs}$ and the simulations, $X_{CO_2,Sim}$, can be seen in northern Africa (Sahara), the Indian peninsula, southern Africa, southern North America, and western Australia, within the range of 2–6 ppm (Figure...
Figure 1. Monthly distributions of GEOS-Chem simulated $X_{CO_2}$ data ($X_{CO_2,Sim}$) at GOSAT satellite geographic coordinates over January, April, July, and October 2010.

Figure 2. Monthly distributions of Tan-Tracker assimilated $X_{CO_2}$ data ($X_{CO_2,TT}$) at GOSAT satellite geographic coordinates over January, April, July, and October 2010.
Figure 3. Monthly distributions of the difference between retrieved GOSAT satellite $X_{\text{CO}_2}$ data (observation; $X_{\text{CO}_2,\text{obs}}$) and the GEOS-Chem simulated $X_{\text{CO}_2}$ data (Sim; $X_{\text{CO}_2,\text{sim}}$), calculated as $X_{\text{CO}_2,\text{obs}} - X_{\text{CO}_2,\text{sim}}$ at GOSAT satellite geographic coordinates over January, April, July, and October 2010.

Figure 4. Monthly distributions of the difference between retrieved GOSAT satellite $X_{\text{CO}_2}$ data (observation; $X_{\text{CO}_2,\text{obs}}$) and the Tan-Tracker assimilated $X_{\text{CO}_2}$ data (TT; $X_{\text{CO}_2,\text{tt}}$), calculated as $X_{\text{CO}_2,\text{obs}} - X_{\text{CO}_2,\text{tt}}$ at GOSAT satellite geographic coordinates over January, April, July, and October 2010.
The high levels of uncertainty at these places were predominantly caused by the deficiency of observations and insufficient knowledge of biophysical and physical processes. The biases between GEOS-Chem and GOSAT found in this study are similar to those reported by Cogan et al. (2012). Relative to the differences shown in Figure 3, the biases of XCO2,obs, and XCO2,TT decrease dramatically to about 1 to 4 ppm, especially in Africa (Sahara) and the Indian peninsula (Figure 4), indicating that assimilating GOSAT observations can effectively eliminate spatial and temporal variation between the model and observations.

The RMSE and correlation coefficients (CCs) between monthly time series of XCO2,model (XCO2,Sim and XCO2,TT) and XCO2,Obs, are shown in Figure 5. The RMSE between XCO2,Sim and XCO2,TT ranged from 1.75 to 3.0 ppm and the CC ranged from 0.25 (November) to 0.85 (April), both having an annual cycle. Relative to only a slight increase in the CC, the RMSE between XCO2,TT and XCO2,Obs decreased sharply by 20% to 40% to about 1.25 to 1.75 ppm. These results illustrate that assimilating GOSAT satellite observations through the Tan-Tracker system tangibly improves the performance of model simulations by reducing spatial error, RMSE and improving the CC to yield more accurate 3D CO2 concentrations and CFs.

4. Conclusions

This study assessed a Chinese carbon cycle data assimilation system (Tan-Tracker) by evaluating the performance during a real-data assimilation experiment. The Tan-Tracker system was initially developed based on an advanced hybrid assimilation approach (PODEn4DVar), as part of the preparation for the launch of the Chinese CO2 observation satellite, TanSat (Liu et al. 2012; Tian et al. 2014). The spatiotemporal distributions of the simulated data, XCO2,model with and without assimilation, XCO2,TT and XCO2,Sim were compared with GOSAT XCO2 data (XCO2,Obs) at each satellite scan position. The results demonstrate that assimilation markedly reduces the RMSE and slightly improves the CC. Overall, our real-data assimilation experiment demonstrated that the Tan-Tracker system performs well when implementing GOSAT data assimilation.

Acknowledgements

We would like to thank the two anonymous reviewers for their critical comments and suggestions, which helped to improve the manuscript. Also, we are very grateful to the ACOS and GOSAT teams for the availability of GOSAT observations.

Funding

This work was partially supported by the National High Technology Research and Development Program of China [grant number 2013AA122002], the National Natural Science Foundation of China [grant numbers 41575100 and 91437220], the Knowledge Innovation Program of the Chinese Academy of Sciences [grant number KZCX2-EW-QN207], and the Special Fund for Meteorological Scientific Research in Public Interest [grant number GYHY201506002].

References

Cogan, A. J., H. Boesch, R. J. Parker, L. Feng, P. I. Palmer, J.-F. L. Blavier, N. M. Deutscher et al. 2012. "Atmospheric Carbon Dioxide Retrieved from the Greenhouse Gases Observing Satellite (GOSAT): Comparison with Ground-Based TCCON Observations and GEOS-Chem Model Calculations." Journal of Geophysical Research 117 (D21). doi:10.1029/2012JD018087.

Chen, Z. H., J. Zhu, and N. Zeng. 2013. "Improved Simulation of Regional CO2 Surface Concentrations Using GEOS-Chem and Fluxes from VEGAS." Atmospheric Chemistry and Physics 13 (15): 7607–7618.

Feng, L., P. I. Palmer, H. Bösch and S. Dance. 2009. "Estimating Surface CO2 Fluxes from Space-borne CO2 Dry Air Mole Fraction Observations Using an Ensemble Kalman Filter." Atmospheric Chemistry and Physics 9 (8): 2619–2633.

IPCC. 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.

Liu, J., I. Fung, E. Kalnay, J. S. Kang, E. T. Olsen, and L. Chen. 2012. "Simultaneous Assimilation of AIRS XCO2 and Meteorological Observations in a Carbon Climate Model with an Ensemble Kalman Filter." Journal of Geophysical Research 117 (D5). doi:10.1029/2011JD016642.

Nassar, R., D. B. A. Jones, P. Suntharalingam, J. M. Chen, R. J. Andres, K. J. Wecht, R. M. Yantosca, et al. 2010. "Modeling Global Atmospheric CO2 with Improved Emission Inventories and CO2 Production from the Oxidation of Other Carbon
Species." Geoscientific Model Development Discussions 3 (3): 889–948.
Peters, W., J. B. Miller, J. Whitaker, A. S. Denning, A. Hirsch, M. C. Krol, D. Zupanski, L. Bruhwiler, and P. P. Tans. 2005. "An Ensemble Data Assimilation System to Estimate CO\textsubscript{2} Surface Fluxes from Atmospheric Trace Gas Observations." Journal of Geophysical Research 110 (D24). doi:10.1029/2005JD006157.
Protocol, K. 1997. United Nations Framework Convention on Climate Change. Kyoto: Kyoto Protocol.
Reuter, M., H. Bovensmann, M. Buchwitz, J. P. Burrows, B. J. Connor, N. M. Deutscher, D. W. T. Griffith, et al. 2011. "Retrieval of Atmospheric CO\textsubscript{2} with Enhanced Accuracy and Precision from SCIAMACHY: Validation with FTS Measurements and Comparison with Model Results." Journal of Geophysical Research: Atmospheres 1984–2012 116 (D4). doi:10.1029/2010JD015047.
Suntharalingam, P., D. J. Jacob, P. I. Palmer, J. A. Logan, R. M. Yantosca, Y. Xiao, M. J. Evans, et al. 2004. "Improved Quantification of Chinese Carbon Fluxes Using CO\textsubscript{2}/CO Correlations in Asian Outflow." Journal of Geophysical Research 109 (D18). doi:10.1029/2003JD004362.
Tian, X., Z. Xie, and Q. Sun. 2011. "A POD-based Ensemble Four-dimensional Variational Assimilation Method." Tellus A 63 (4): 805–816.
Tian, X., Z. Xie, Y. Liu, Z. Cai, Y. Fu, H. Zhang, and L. Feng. 2014. "A Joint Data Assimilation System (Tan-Tracker) to Simultaneously Estimate Surface CO\textsubscript{2} Fluxes and 3-D Atmospheric CO\textsubscript{2} Concentrations from Observations." Atmospheric Chemistry and Physics 14 (23): 13281–13293.
Wang, B., J. Liu, S. Wang, W. Cheng, J. Liu, C. Liu, Q. Xiao and Y. Kuo. 2010. "An Economical Approach to Four-dimensional Variational Data Assimilation." Advances in Atmospheric Sciences 27: 715–727.
Yokota, T., Y. Yoshida, N. Eguchi, Y. Ota, T. Tanaka, H. Watanabe, and S. Maksyutov. 2009. "Global Concentrations of CO\textsubscript{2} and CH\textsubscript{4} Retrieved from GOSAT: First Preliminary Results." SOLA 5: 160–163.
Zhang, H. F., B. Z. Chen, T. Machida, T. Machida, H. Matsueda, Y. Sawa, Y. Fukuyama, et al. 2014. "Estimating Asian Terrestrial Carbon Fluxes from CONTRAIL Aircraft and Surface CO\textsubscript{2} Observations for the Period 2006–2010." Atmospheric Chemistry and Physics 14 (11): 5807–5824.