A regional carbon flux data assimilation system and its preliminary evaluation in East Asia

Z. Peng\textsuperscript{1}, M. Zhang\textsuperscript{2}, X. Kou\textsuperscript{2,3}, X. Tian\textsuperscript{4}, and X. Ma\textsuperscript{4}

\textsuperscript{1}School of Atmospheric Sciences, Nanjing University, Nanjing 210093, China
\textsuperscript{2}State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
\textsuperscript{3}University of Chinese Academy of Sciences, Beijing 100049, China
\textsuperscript{4}Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

Received: 21 June 2014 – Accepted: 25 July 2014 – Published: 8 August 2014

Correspondence to: Z. Peng (pengzhen@nju.edu.cn) and M. Zhang (mgzhang@mail.iap.ac.cn)

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

In order to optimize surface CO$_2$ fluxes at finer scales, a regional surface CO$_2$ flux inversion system (Carbon Flux Inversion system and Community Multi-scale Air Quality, CFI-CMAQ) has been developed by simultaneously assimilating CO$_2$ concentrations and surface CO$_2$ fluxes into the regional modeling system, CMAQ. The smoothing operator is associated with the atmospheric transport model to constitute a persistence dynamical model to forecast the surface CO$_2$ flux scaling factors. In this implementation, the “signal-to-noise” problem can be avoided; plus, any useful observed information achieved by the current assimilation cycle can be transferred into the next assimilation cycle. Thus, the surface CO$_2$ fluxes can be optimized as a whole at the grid scale in CFI-CMAQ. The performance of CFI-CMAQ was quantitatively evaluated through a set of Observing System Simulation Experiments (OSSEs) by assimilating CO$_2$ retrievals from GOSAT (Greenhouse Gases Observing Satellite). The results showed that the CO$_2$ concentration assimilation using the ensemble Kalman filter (EnKF) could constrain the CO$_2$ concentrations effectively, illustrating that the simultaneous assimilation of CO$_2$ concentrations can provide convincing CO$_2$ initial analysis fields for CO$_2$ flux inversion. In addition, the CO$_2$ flux optimization using the ensemble Kalman smoother (EnKS) demonstrated that CFI-CMAQ could in general reproduce true fluxes at finer scales with acceptable bias. Two further sets of numerical experiments were conducted to investigate the sensitivities of the inflation factor of scaling factors and the smoother window. The results showed that the ability of CFI-CMAQ to optimize CO$_2$ fluxes greatly relied on the choice of the inflation factor. However, the smoother window had a slight influence on the optimized results. CFI-CMAQ performed very well even with a short lag-window (e.g. 3 days).
1 Introduction

Atmospheric CO\textsubscript{2} concentrations are increasing due to fossil fuel combustion, cement production, and land use changes such as deforestation. As the most important anthropogenic greenhouse gas, the accumulation of CO\textsubscript{2} in the atmosphere is reported to be the largest human-induced driver of global warming (IPCC AR5 WG1: 2013), providing a substantial threat to the sustainable development of both the environment and mankind. Quantifying CO\textsubscript{2} sources and sinks is central to the successful implementation of carbon emissions reduction strategies aimed at mitigating the adverse effects of global warming. It is possible to improve the accuracy of surface CO\textsubscript{2} flux estimates through the use of an advanced data assimilation technique (e.g., Chevallier et al., 2005, 2007; Chevallier, 2007; Baker et al., 2006; Peters et al., 2007; Engelen et al., 2009; Feng et al., 2009; Kang et al., 2012; Liu et al., 2012). Feng et al. (2009) showed that the uncertainties of surface CO\textsubscript{2} flux estimates can be reduced significantly by assimilating OCO \(X_{\text{CO}_2}\) measurements. Peters et al. (2005, 2007, 2009) developed a surface CO\textsubscript{2} flux inversion system, CarbonTracker, by incorporating the ensemble square-root filter (EnSRF) into the atmospheric transport TM5 model. And the inversion results obtained by assimilating in situ surface CO\textsubscript{2} observations are in excellent agreement with a wide collection of carbon inventories that form the basis of the first North American State of the Carbon Cycle Report (SOCCR) (Peters et al., 2007). CarbonTracker is also well used to constrain the surface CO\textsubscript{2} flux over Europe and Asia (e.g., Zhang et al., 2014a, b). Kang et al. (2011, 2012) presented a simultaneous data assimilation of surface CO\textsubscript{2} fluxes and atmospheric CO\textsubscript{2} concentrations along with meteorological variables using the Local Ensemble Transform Kalman Filter (LETKF). They indicated that an accurate estimation of the evolving surface fluxes can be gained even without any a priori information. Recently, Tian et al. (2013) developed a new surface CO\textsubscript{2} flux data assimilation system, Tan-Tracker (in Chinese, “Tan” means carbon), by incorporating a joint PODEn4DVar assimilation framework into the GEOS-Chem model on the basis of Peters et al. (2005, 2007) and Kang et al. (2011, 2012).
They discussed in detail that the assimilation of CO$_2$ surface fluxes could be improved though the use of simultaneous assimilation of CO$_2$ concentrations and CO$_2$ surface fluxes due to the fact that the uncertainty of the evolution of CO$_2$ concentrations could be gradually eliminated. Despite the rigor of data assimilation theory, current CO$_2$ flux-inversion methods (e.g., CarbonTracker) still face many challenging scientific problems, such as: (1) the well-known “signal-to-noise” problem (NRC, 2010); (2) large inaccuracies in chemical transport models (e.g., Prather et al., 2008); (3) vast computational expenses (e.g., Feng et al., 2009); and (4) the sparseness of observation data (e.g., Gurney et al., 2002).

The “signal-to-noise” problem is one of the most challenging issue for an ensemble-based CO$_2$ flux inversion system due to the fact that surface CO$_2$ fluxes are the model forcing (or boundary condition), rather than model states (like CO$_2$ concentrations), of the chemistry transport model (CTM). In the absence of a suitable dynamical model to describe the evolution of the surface CO$_2$ fluxes, most CO$_2$ flux-inversion studies have traditionally ignored the uncertainty of anthropogenic and other CO$_2$ emissions and focused on the optimization of natural (i.e., biospheric and oceanic) CO$_2$ emissions at the ecological scale (e.g., Deng et al., 2007; Feng et al., 2009; Peters et al., 2005, 2007; Jiang et al., 2013; Peylin et al., 2013).

This compromise is acceptable to some extent. Indeed, the total amount of anthropogenic CO$_2$ emissions can be estimated by relatively well-documented global fuel-consumption data with a small degree of uncertainty (Boden et al., 2011). It was reported in IPCC (2007) that anthropogenic CO$_2$ emissions were the primary source contributing to the $7.2 \pm 0.3 \text{Pg C yr}^{-1}$ from 2000 to 2005; ocean–atmosphere exchanges were the largest sinks, absorbing $-2.2 \pm 0.5 \text{Pg C yr}^{-1}$, followed by biosphere–atmosphere exchanges, absorbing $-0.9 \pm 0.6 \text{Pg C yr}^{-1}$ during the same period. There is no doubt that the uncertainties involved in the total amount of anthropogenic CO$_2$ emissions are much smaller than those related to natural emissions.

However, their spatial distribution, strength and temporal development still remain elusive, because of their inherent non-uniformities (Andres et al., 2012; Gurney et al.,
With increasing scientific and political interest in recent years, information on anthropogenic CO$_2$ emissions is required at increasingly finer scales, in addition to quantifying their total amount. Marland (2008) pointed out that even a tiny amount of uncertainty, i.e., 0.9%, in one of the leading emitter countries like the US is equivalent to the total emissions of the smaller emitter countries in the world. Furthermore, the usual values of anthropogenic CO$_2$ emissions in chemical transport models have thus far been simply interpolated from very coarse monthly-mean fuel consumption data. Therefore, great uncertainty in the spatiotemporal distributions of anthropogenic emissions likely exists, which could reduce the accuracy of CO$_2$ concentration simulations and subsequently increase the inaccuracy of natural CO$_2$ flux inversion results. In addition, current research approaches tend only to assimilate natural CO$_2$ emissions at the ecological scale, which is far from sufficient. Therefore, surface CO$_2$ fluxes should be constrained as a whole at a finer scale.

In CarbonTracker (Peters et al., 2007), a smoothing operator is innovatively applied as the persistence forecast model. In that application, the surface CO$_2$ fluxes can be treated as the model states and the observed information ingested by the current assimilation cycle can be used in the next assimilation cycle effectively. However, the “signal-to-noise” problem is not yet resolved, and thus CarbonTracker also has to assimilate natural CO$_2$ emissions at the ecological scale only. In Tan-Tracker (Tian et al., 2013), a 4-D moving sampling strategy (Wang et al., 2010) is used to generate the flux ensemble members, and so the surface CO$_2$ fluxes can be optimized as a whole at the grid scale. In the present reported work, the persistence dynamical model taken by Peters et al. (2005) was further developed for the purpose of resolving the “signal-to-noise” problem, as well as transporting the useful observed information in the next assimilation cycle, in order to optimize the surface CO$_2$ fluxes as a whole at the grid scale. This process is described in detail in Sect. 2 of this paper.

The surface CO$_2$ flux inversion system presented in this paper was developed by incorporating a joint data assimilation framework into a regional modeling system to optimize the CO$_2$ fluxes at a finer scale. A joint assimilation framework – the simultane-
ous assimilation of CO₂ concentrations and surface CO₂ fluxes – was applied because assimilating CO₂ observations from multiple sources can improve the accuracy of simulation results (e.g., Miyazaki, 2009; Liu et al., 2009, 2011, 2012; Tangborn et al., 2013; Huang et al., 2014), and the simultaneous assimilation of CO₂ concentrations and surface CO₂ fluxes can largely eliminate the uncertainty in initial CO₂ concentrations on the CO₂ evolution (Kang et al., 2011, 2012; Tian et al., 2013). The data assimilation techniques applied were the ensemble Kalman filter (EnKF) and the ensemble Kalman smoother (EnKS). EnKF was used to constrain CO₂ concentrations and EnKS was used to optimize surface CO₂ fluxes. The regional transport modeling system applied was Regional Atmospheric Modeling System and Community Multi-scale Air Quality (RAMS-CMAQ). Previous studies (Ahmadov et al., 2009; Pillai et al., 2010; Kretschmer et al., 2011) have shown that regional chemical transport models with high resolution, compared to global models, can reduce model errors and may have the ability to reproduce the effects of meso–micro–scale transport on atmospheric CO₂ distributions. Besides, RAMS-CMAQ has been proven to be reliable (Zhang et al., 2002, 2003, 2007; Kou et al., 2013; Liu et al., 2013; Huang et al., 2014). Thus, RAMS-CMAQ was chosen to develop this regional surface CO₂ flux inversion system. For simplicity, this system is referred to as CFI-CMAQ (Carbon Flux Inversion system and Community Multi-scale Air Quality).

Since this is the first time of introducing CFI-CMAQ, we focus mainly on introducing the methodology in this paper. Nevertheless, in addition, Observing System Simulation Experiments (OSSEs) were designed to assess the system’s ability to optimize surface CO₂ fluxes. The information of GOSAT X_CO₂ are used to generate artificial observations because of the sparseness and heterogeneity of ground-based measurements.

The remainder of the paper is organized as follows. Section 2 describes the details of the regional surface CO₂ flux inversion system, CFI-CMAQ, including the developed persistence dynamical model, a simple review of the EnKS and EnKF assimilation approaches, and the process involved. The experimental designs are then introduced.
and the assimilation results shown in Sect. 3. Finally, a summary and conclusions are provided in Sect. 4.

2 Framework of the regional surface CO\(_2\) flux inversion system

As an ensemble-based assimilation system, CFI-CMAQ was also developed by applying a set of linear multiplication factors, similar to the approach by Peters et al. (2007) and Tian et al. (2013). Therefore, the background surface fluxes from an \(N\)-member ensemble can be described by

\[
F^b_i(x, y, z, t) = \lambda^b_i(x, y, z, t)F^0(x, y, z, t), \quad (i = 1, \ldots, N),
\]

where \(F^b_i(x, y, z, t)\) are the background fluxes that force the CTM to integrate to produce the background CO\(_2\) concentration fields, \(F^0(x, y, z, t)\) are the prior net CO\(_2\) surface exchanges, and \(\lambda^b_i(x, y, z, t)\) represent the background linear scaling factors (Peters et al., 2007; Tian et al., 2013) for each time and each grid to be optimized in the assimilation. The notation is standard: the superscript \(b\) refers to the background, and the subscript \(i\) refers to the \(i\)th ensemble member. In the following, \(\lambda^b_i(x, y, z, t)\) are referred to as \(\lambda^b_{i,t}\) for simplicity.

At each optimization cycle, CFI-CMAQ includes the following four parts: (1) forecasting of the background linear scaling factors \(\lambda^b_{i,t}\); (2) optimization of the scaling factors in the smoother window by EnKS (see Fig. 1); (3) updating of the flux in the smoother window; and (4) assimilation of the CO\(_2\) concentration fields at time \(t\) by EnKF. A flowchart illustrating CFI-CMAQ is presented in Fig. 2. The assimilation procedure is addressed in detail below. In addition, the observation operator is introduced, particularly for use of the GOSAT \(X_{\text{CO}_2}\) data in Sect. 2.5. Furthermore, covariance inflation and localization techniques applied in CFI-CMAQ are introduced briefly in Sect. 2.6.
2.1 Forecasting the background linear scaling factors

In order to pass the useful observed information onto the next assimilation cycle effectively, following Peters et al. (2007) the smoothing operator is applied as the persistence dynamical model to calculate the linear scaling factors $\lambda^b_{i,t}$,

$$\lambda^b_{i,t} = \left( \frac{\sum_{j=t-M+1}^{t-1} \lambda^a_{i,j|t-1} + \lambda^p_{i,t}}{M} \right), \quad (i = 1, \ldots, N, j = t - M + 1, \ldots, t - 1),$$

where $\lambda^a_{i,j|t-1}$ refers to analyzed quantities from the previous assimilation cycle at time $j$ (see Fig. 1) and $|t - 1$ means that these factors have been optimized by using observations at time $t - 1$; $\lambda^p_{i,t}$ refers to the prior values of the linear scaling factors at time $t$. The superscript $a$ refers to the analyzed, and $p$ refers to the prior. This operation represents a smoothing over all the time steps in the smoother window (see Fig. 1), thus dampening variations in the forecast of $\lambda^b_{i,t}$ in time.

In order to generate $\lambda^b_{i,t}$, the atmospheric transport model (CMAQ) is applied and a set of ensemble forecast experiments are carried out to obtain the prior values for $\lambda^p_{i,t}$. It integrates from time $t - 1$ to $t$ to produce the CO$_2$ concentration fields $\mathbf{C}^f_i(x,y,z,t)$ forced by the prior flux $F_0(x,y,z,t)$ with $\mathbf{C}^a_i(x,y,z,t - 1)$ as initial conditions. Then, the ratio $\kappa_{i,t} = \frac{C_i^f(x,y,z,t)}{C_i^a(x,y,z,t)}$ is calculated, where $C_i^f(x,y,z,t) = \frac{1}{N} \sum_{i=1}^{N} C_i^f(x,y,z,t)$. Suppose that $\lambda^p_{i,t} = \kappa_{i,t}$ due to the fact that the surface CO$_2$ fluxes correlate with its concentrations, the values for $\lambda^p_{i,t}$ are obtained and then $\lambda^b_{i,t}$ can finally be calculated (see the part with red arrows in the flowchart in Fig. 2). Hereafter, $C_i^f(x,y,z,t)$ are referred to as $C_i^f_{i,t}$ for simplicity.

The assumption $\lambda^p_{i,t} = \kappa_{i,t}$ is just a compromise from the complex conservation equation. In CarbonTracker, $\lambda^p_{i,t}$ are all set to 1 (Peters et al., 2007). In this study, the smooth-
ing operator is associated with the atmospheric transport model to constitute the persistence dynamical model to describe the evolution of the scaling factors. Through this process, the prior values of $\lambda_{i,t}^p$ can be forecast at the grid scale without any random noise. So do $\lambda_{i,t}^b$. Ultimately, the “signal-to-noise” problem can be resolved, and the surface CO$_2$ fluxes can then be optimized as a whole at the grid scale.

It is important to note that, similar to Peters et al. (2007), this dynamical model equation still does not include an error term in the dynamical model, and the model error cannot yet be estimated. However, the covariance inflation is applied to compensate for model errors before optimization, which is addressed in Sect. 2.6.

2.2 Optimizing the scaling factors in the smoother window by EnKS

In the current assimilation cycle $t - M \sim t$ (see Fig. 1), the scaling factors to be optimized in the smoother window are $(\lambda_{i,t-M|t-1}^a, \lambda_{i,t-M+1|t-1}^a, \ldots, \lambda_{i,j|t-1}^a, \ldots, \lambda_{i,j-1|t-1}^a, \lambda_{i,t|t-1}^a)$. Referring to $\lambda_{i,t}^b$ as $\lambda_{i,t|t-1}^a$ for convenience, the scaling factors to be optimized are $(\lambda_{i,t-M-1}^a, \lambda_{i,t-M+1|t-1}^a, \ldots, \lambda_{i,j-1|t-1}^a, \lambda_{i,j|t-1}^a, \lambda_{i,t|t-1}^a)$. Using the EnKS analysis technique, these scaling factors are updated in turn via

$$
\lambda_{i,t|t}^a = \lambda_{i,t|t-1}^a + K^e \left( y_{i,t}^{\text{obs}} - y_{i,t}^{f} + v_{i,t} \right), \quad (i = 1, \ldots, N, j = t - M + 1, \ldots, t - 1),$$

$$K^e = S^e_{j,t|t-1} H^T (H P^e_{t,t|t-1} H^T + R)^{-1},$$

$$S^e_{j,t|t-1} = \frac{1}{N - 1} \sum_{i=1}^{N} \left[ \lambda_{i,j|t-1}^a - \lambda_{i,j|t-1}^a \right] \left[ \lambda_{i,j|t-1}^a - \lambda_{i,j|t-1}^a \right]^T,$$

$$P^e_{t,t|t-1} = \frac{1}{N - 1} \sum_{i=1}^{N} \left[ \lambda_{i,t|t-1}^a - \lambda_{i,t|t-1}^a \right] \left[ \lambda_{i,t|t-1}^a - \lambda_{i,t|t-1}^a \right]^T,$$

$$y_{i,t}^{f} = H \left( M_{t-1\rightarrow t} \left( \lambda_{i,t|t-1}^a \right) \right),$$

20353
where $K^e$ is the Kalman gain matrix of EnKS, $S^e_{j,t|t-1}$ is the background error cross-covariance between the state vector $\lambda^a_{i,j|t-1}$ and $\lambda^a_{i,j|t-1}$, $P^e_{t,t|t-1}$ is the background error covariance of the state vector $\lambda^a_{i,t|t-1}$, $H(\cdot)$ is the observation operator that maps the state variable from model space into observation space, $\nu_{i,t}$ is a random normal distribution perturbation field with zero mean and $R$ standard deviation representing the measurement errors, $y^{\text{obs}}_t$ is the observation vector measured at time $t$ and $y^f_{i,t}$ is the simulated observations.

In actual implementations, it is unnecessary to calculate $S^e_{j,t|t-1}$ and $P^e_{t,t|t-1}$ separately. $S^e_{j,t|t-1}H^T$ and $H P^e_{t|t-1}H^T$ can be calculated as a whole by

$$S^e_{j,t|t-1}H^T = \frac{1}{N-1} \sum_{i=1}^{N} \left[ \lambda^a_{i,j|t-1} - \overline{\lambda^a_{i,j|t-1}} \right] \left[ y^f_{i,t} - \overline{y^f_t} \right]^T,$$

$$H P^e_{t|t-1}H^T = \frac{1}{N-1} \sum_{i=1}^{N} \left[ y^f_{i,t} - \overline{y^f_t} \right] \left[ y^f_{i,t} - \overline{y^f_t} \right]^T,$$

$$\overline{y^f_t} = \frac{1}{N-1} \sum_{i=1}^{N} y^f_{i,t}.$$  

2.3 Updating the flux in the smoother window

Substituting the optimized scaling factors $(\lambda^a_{i,t-M|t}, \lambda^a_{i,t-M+1|t}, \ldots, \lambda^a_{i,|t}, \ldots, \lambda^a_{i,t-1|t}, \lambda^a_{i,t|t})$ into Eq. (1), the flux in the smoother window can be updated.
2.4 Assimilating the CO$_2$ concentration fields at time $t$ by EnKF

The analysis of CO$_2$ concentrations fields at time $t$ in the EnKF scheme is updated via

$$C_{a,i,t} = C_{f,i,t} + K \left( y_{t}^{\text{obs}} - y_{t}^{f} + u_{i,t} \right), \quad (11)$$

$$K = P_{f}^{T}H^{T} \left( H P_{f}^{T}H^{T} + R \right)^{-1}, \quad (12)$$

where $K$ is the Kalman gain matrix of EnKF and $P_{f}^{T}H^{T}$ and $H P_{f}^{T}H^{T}$ can be calculated as a whole by

$$P_{f}^{T}H^{T} = \frac{1}{N-1} \sum_{i=1}^{N} \left[ C_{i,t}^{f} - \overline{C}_{i,t}^{f} \right] \left[ y_{i,t}^{f} - \overline{y}_{t}^{f} \right]^{T}, \quad (13)$$

$$H P_{f}^{T}H^{T} = \frac{1}{N-1} \sum_{i=1}^{N} \left[ y_{i,t}^{f} - \overline{y}_{t}^{f} \right]^{T} \left[ y_{i,t}^{f} - \overline{y}_{t}^{f} \right]. \quad (14)$$

2.5 The observation operator

As mentioned above, the observation operator $H(\cdot)$ transforms the state variable from model space into observation space. Usually, it is the spatial bilinear interpolator for traditional ground-based observations. Since the GOSAT $X_{\text{CO}_2}$ retrieval is a weighted CO$_2$ column average, the simulated $X_{\text{CO}_2}$ should be calculated with the same weighted column average method (Connor et al., 2008; Crisp et al., 2010, 2012; O’Dell et al., 2012). So, the observation operator to assimilate the GOSAT $X_{\text{CO}_2}$ retrieval is

$$y^{f} = H \left( M_{t-1 \rightarrow t} \left( \lambda^{a}_{i,t|t-1} \right) \right) = H \left( C_{i,t}^{f} \right) = y^{\text{priori}} + h^{T} a_{\text{CO}_2} \left( S \left( C_{i,t}^{f} \right) - f^{\text{priori}} \right), \quad (15)$$

where $y^{f}$ is the simulated $X_{\text{CO}_2}$; $y^{\text{priori}}$ is the a priori CO$_2$ column average used in the GOSAT $X_{\text{CO}_2}$ retrieval process; $S(\cdot)$ is the spatial bilinear interpolation operator that...
interpolates the simulated fields to the GOSAT $X_{\text{CO}_2}$ locations to obtain the simulated CO$_2$ vertical profiles there; $f^\text{priori}$ is the a priori CO$_2$ vertical profile used in the retrieval process; $h$ is the pressure weighting function, which indicates the contribution of the retrieved value from each layer of the atmosphere; and $a_{\text{CO}_2}$ is the normalized averaging kernel.

### 2.6 Covariance inflation and localization

In order to keep the ensemble spread of the CO$_2$ concentrations at a certain level and compensate for transport model error to prevent filter divergence, covariance inflation is applied before updating the CO$_2$ concentrations. So,

$$
(C_{i,t}^f)_{\text{new}} = \alpha \left( C_{i,t}^f - \overline{C_{i,t}^f} \right) + \overline{C_{i,t}^f},
$$

(16)

where $\alpha$ is the inflation factor of CO$_2$ concentrations and $(C_{i,t}^f)_{\text{new}}$ is the final field used for data assimilation.

Similarly, covariance inflation is also used to keep the ensemble spread of the prior scaling factors at a certain level and compensate for dynamical model error. So,

$$
(\lambda_{i,t}^p)_{\text{new}} = \beta \left( \lambda_{i,t}^p - \overline{\lambda_{i,t}^p} \right) + \overline{\lambda_{i,t}^p},
$$

(17)

where $\beta$ is the inflation factor of scaling factors and $(\lambda_{i,t}^p)_{\text{new}}$ is the final scaling factors used for data assimilation.

In addition, the Schur product is utilized to filter the remote correlation resulting from the spurious long-range correlations (Houtekamer and Mitchell, 2001). So, the Kalman
gain matrix $K^e$ and $K$ are updated via

$$K^e = \left( (\rho \circ S^e_{j,t,t-1}) H^T (H (\rho \circ P^e_{t,t,t-1}) H^T + R)^{-1} \right),$$

(18)

$$K = \left( (\rho \circ P^f) H^T \right) \left( (H (\rho \circ P^f) H^T + R)^{-1} \right),$$

(19)

where the filtering matrix $\rho$ is calculated using the formula

$$C_0(r,c) = \begin{cases} 
-\frac{1}{4} \left( \frac{|r|}{c} \right)^5 + \frac{1}{2} \left( \frac{|r|}{c} \right)^4 + \frac{5}{8} \left( \frac{|r|}{c} \right)^3 - \frac{5}{3} \left( \frac{|r|}{c} \right)^2 + 1, & 0 \leq |r| \leq c \\
\frac{1}{12} \left( \frac{|r|}{c} \right)^5 - \frac{1}{2} \left( \frac{|r|}{c} \right)^4 + \frac{5}{8} \left( \frac{|r|}{c} \right)^3 + \frac{5}{3} \left( \frac{|r|}{c} \right)^2 - 5 \left( \frac{|r|}{c} \right) + 4 - \frac{2}{3} \left( \frac{c}{|r|} \right), & c \leq |r| \leq 2c \\
0, & c \leq |r|
\end{cases},$$

(20)

where $c$ is the element of the localization Schur radius. The matrix $\rho$ can filter the small background error correlations associated with remote observations through the Schur product (Tian et al., 2011). And the Schur product tends to reduce the effect of those observations smoothly at intermediate distances due to the smooth and monotonically decreasing of the filtering matrix.

3 OSSEs for evaluation of CFI-CMAQ

A set of OSSEs were designed to quantitatively assess the performance of CFI-CMAQ. The setup of the experiments and the results are described in this section.

3.1 Experimental setup

The chemical transport model utilized was RAMS-CMAQ (Zhang et al., 2002), in which CO$_2$ was treated as an inert tracer. The model domain was 6654 km $\times$ 5440 km on a ro-
tated polar stereographic map projection centered at (35.0° N, 116.0° E), with a horizontal grid resolution of 64 km × 64 km and 15 vertical layers in the $\sigma_z$-coordinate system, unequally spaced from the surface to approximately 23 km. The initial fields and boundary conditions of the CO$_2$ concentrations were interpolated from the simulated CO$_2$ fields of CarbonTracker 2011 (Peters, 2007). The prior surface CO$_2$ fluxes included biosphere–atmosphere CO$_2$ fluxes, ocean–atmosphere CO$_2$ fluxes, anthropogenic emissions, and biomass-burning emissions (Kou et al., 2013),

$$F^0(x, y, z, t) = F_{\text{bio}}(x, y, z, t) + F_{\text{oce}}(x, y, z, t) + F_{\text{ff}}(x, y, z, t) + F_{\text{fire}}(x, y, z, t),$$  \hspace{1cm} (21)

where $F^0(x, y, z, t)$ was the prior surface CO$_2$ flux; $F_{\text{bio}}$ and $F_{\text{oce}}$ were the biosphere–atmosphere and ocean–atmosphere CO$_2$ fluxes, respectively, which were obtained from the optimized results of CarbonTracker 2011 (Peters, 2007); $F_{\text{ff}}$ was fossil fuel emissions, adopted from the Regional Emission inventory in ASia (REAS, 2005 Asia monthly mean emission inventory) with a spatial resolution of 0.5° × 0.5° (Ohara et al., 2007); $F_{\text{fire}}$ was biomass–burning emissions, provided by the monthly mean inventory at a spatial resolution of 0.5° × 0.5° from the Global Fire Emissions Database, Version 3 (GFED v3) (Van der Werf et al., 2010).

Firstly, the prior flux $F^0(x, y, z, t)$ was assumed as the true surface CO$_2$ flux. Forced by $F^0(x, y, z, t)$, the RAMS-CMAQ model was run to produce the artificial true CO$_2$ concentration results $C^0(x, y, z, t)$. Then, the artificial GOSAT observations $y_t^{\text{obs}}$ (or $X_t^{\text{CO}_2}$) were generated using the observation operator in Eq. (7).

Secondly, the background surface CO$_2$ fluxes series $F^b(x, y, z, t)$ was created by

$$F^b(x, y, z, t) = (1.8 + \delta(x, y, z, t))F^0(x, y, z, t),$$  \hspace{1cm} (22)

where $\delta$ was a random number. Driven by $F^b(x, y, z, t)$, the RAMS-CMAQ model was integrated to obtain the background CO$_2$ simulations $C^b(x, y, z, t)$. Then, the column-averaged concentrations $X^{\text{CO}_2}$ were obtained using Eq. (7).
The performance of CFI-CMAQ was evaluated through a group of well-designed OSSEs. In all the OSSEs, the default ensemble size $N$ was 48, the standard localization Schur radius $c$ was 1280 km (20 grid spacing), and the covariance inflation factor of concentrations $\alpha$ was 1.1. The referenced lag-window was 9 days and the covariance inflation factor of scaling factors $\beta$ was 70. Since the lag-window is very important for CO$_2$ transportation and $\beta$ is a newly introduced parameter, both these parameters were further investigated by several numerical sensitivity experiments. The primary focus of this paper is to describe the assimilation methodology, so all the numerical experiments started on 1 January and ended on 30 March.

3.2 Experimental results

Essentially, CFI-CMAQ includes two parts: one for the CO$_2$ concentration assimilation with EnKF, which can provide a convincing CO$_2$ initial analysis field for the next assimilation cycle; and the other for the CO$_2$ flux optimization with EnKS. The EnKF part for the CO$_2$ concentration assimilation is independent of the EnKS part for the CO$_2$ flux inversion, but the performance of the EnKS part will be greatly influenced by the validation of the EnKF part.

We begin by describing the impacts of assimilating artificial observations $X^0_{\text{CO}_2}$ on CO$_2$ simulations. As shown in Fig. 4a, b and d, the monthly mean values of the background CO$_2$ concentrations $C^b$ produced by the magnified surface CO$_2$ fluxes $F^b$ were much larger than those of the artificial true CO$_2$ concentrations $C^0$ produced by the prior surface CO$_2$ fluxes $F^0$ near the surface in February 2010. In the east and south of China especially, the magnitude of the difference between $C^0$ and $C^b$ was at least 6 ppmv. Also, as expected, the monthly mean $X^b_{\text{CO}_2}$ was much larger than the monthly mean artificial observations $X^0_{\text{CO}_2}$, and the magnitude of the difference between $X^0_{\text{CO}_2}$ and $X^b_{\text{CO}_2}$ reached 2 ppmv in the east and south of China (see Fig. 3b, c and e). However, the impact of magnifying surface CO$_2$ fluxes on the CO$_2$ concentrations was
primarily below the model-level 10 (approximately 6 km), and especially below level 7 (approximately 1.6 km). And above model level 10, the differences between $C^0$ and $C^b$ fell to zero. After assimilating $X^0_{CO_2}$, the analysis $CO_2$ field $C^a$ was much closer to $C^0$ (see Fig. 4c, e and f). The monthly mean difference between $C^0$ and $C^a$ ranged from −2 to 2 ppmv and the relative error $(C^0 - C^a)/C^0$ ranged from −1 to 1% in almost the entire model domain at model-level 1. The monthly mean differences between $C^0$ and $C^a$ were negligible above model-level 1. The monthly mean $X^a_{CO_2}$ was also closer to $X^0_{CO_2}$ and the difference between $X^0_{CO_2}$ and $X^a_{CO_2}$ ranged from −0.5 to 0.5 ppmv. In order to evaluate the general impact of assimilating $X^0_{CO_2}$ in the surface layer, time series of the daily mean $CO_2$ concentration extracted from the background simulations and the assimilations were compared with the artificial true simulations at four national background stations in China and their nearest large cities. As shown in Fig. 3a, Waliguan is 150 km away from Xining, Longfengshan is 180 km away from Haerbin, Shangdianzi is 150 km away from Beijing, and Linan is 50 km away from Hangzhou. The assimilated results are shown in Fig. 7. The background time series were much larger than the artificial true time series, especially at Shangdianzi, Beijing, Linan and Hangzhou, which are strongly influenced by local anthropogenic $CO_2$ emissions. After assimilating $X^0_{CO_2}$, the assimilated time series were very near to the true time series with negligible bias, as expected, at Waliguan, Xining, Shangdianzi, Beijing, Linan and Hangzhou, especially after the first 10 days, which can be considered the spin-up period. Meanwhile, the improvements at Longfengshan and Haerbin were limited due to the absence of observation data at those locations (see Fig. 3a). Nevertheless, in general, the substantial benefits to the $CO_2$ concentrations in the surface layer of assimilating GOSAT $X_{CO_2}$ with EnKF are clear. All the results illustrated that EnKF can provide a convincing $CO_2$ initial analysis field for $CO_2$ flux inversion.

The impacts of assimilating $X^0_{CO_2}$ on surface $CO_2$ fluxes were also highly impressive. On the whole, the background surface $CO_2$ fluxes $F^b$ were much larger than the prior true surface $CO_2$ fluxes $F^0$ in February 2010, especially in the east and south of
China. The monthly mean difference between $F^b$ and $F^0$ reached 0.5 µmole m$^{-2}$ s$^{-1}$ in Jing-Jin-Ji, the Yangtze River Delta, and Pearl River Delta Urban Circle because of the strong local anthropogenic CO$_2$ emissions (see Fig. 7a, b and d). After assimilating $X^0_{CO_2}$, the assimilated surface CO$_2$ fluxes $F^a$ decreased sharply. Thus, the monthly mean values of $F^a$ were much smaller than $F^b$ in most of the model domain in February 2010. The pattern of the difference between $F^a$ and $F^b$ was similar to that of the difference between $F^0$ and $F^b$ (see Fig. 7b–e). The assimilated surface CO$_2$ fluxes were also compared to the artificial true fluxes $F^0$, revealing that $F^a$ was equivalent to $F^0$ in most of the model domain. The monthly mean difference between $F^a$ and $F^0$ ranged from −0.01 to 0.01 µmole m$^{-2}$ s$^{-1}$ only (see Fig. 7f). In addition, the root-mean-square errors (RMSEs) of the assimilated flux members were analyzed. As shown in Fig. 8, the monthly mean RMSE was less than 0.05 µmole m$^{-2}$ s$^{-1}$ in most of the model domain, except in areas near to large cities such as Beijing, Shanghai and Guangzhou, indicating that the assimilated CO$_2$ fluxes were reliable.

In order to evaluate the ability of CFI-CMAQ to optimize the surface CO$_2$ fluxes comprehensively, the ratios of the monthly mean $F^0$ to the monthly mean $F^b$, which was similar to the monthly mean $\lambda^b_t$ (where $\lambda^b_t = \sum_{i=1}^{N} \lambda^b_{i,t}/N$), but not the same, were analyzed. In actual implementation, we only analyzed the ratios where the absolute values of the monthly mean $F^b$ were larger than 0.01, to avoid random noise. As shown in Fig. 9a, the ratios of the monthly mean $F^0$ to the monthly mean $F^b$ ranged from 0.5 to 0.65 in most of China, except in the Qinghai–Tibet Plateau, where the absolute values of the monthly mean $F^b$ were very small in February. The ratios varied greatly in the Indo-China Peninsula because of strong diurnal variation of CO$_2$ fluxes. In addition, the ratios of the monthly mean $F^a$ to the monthly mean $F^b$ and the ratios of the monthly mean $F^a$ to the monthly mean $F^0$ are shown in Fig. 9b and c, respectively. These two figures demonstrate that the impact of the assimilation of $X^0_{CO_2}$ by CFI-CMAQ on
CO_2 fluxes was great in the east and south of China in general, but the influence was negligible in Northeast China due to the lack of observation data.

In addition, time series of daily mean surface CO_2 fluxes extracted from F^b and F^a were compared with that from F^0 at four national background stations in China and their nearest large cities, similar to the CO_2 concentration assimilation. The results are shown in Fig. 9. The background time series were much larger than the artificial true time series, especially at Haerbin, Shangdianzi, Beijing, Linan and Hangzhou, which are strongly influenced by local anthropogenic CO_2 emissions. After assimilating X^0_{CO_2}, the assimilated time series were near to the true time series with acceptable bias, as expected, at Waliguan, Xining, Shangdianzi, Linan and Hangzhou after the 10 day spin-up period. However, the improvements at Longfengshan and Haerbin were negligible because of a lack of observations at these locations. Also, this inversion system failed to show improvements at Beijing because of an excessive impact of assimilation. Since the impact of assimilation X^0_{CO_2} by CFI-CMAQ on CO_2 fluxes was in general greater in the east and south of China than other model areas (see Figs. 7e and 9b), the time series of the daily mean CO_2 fluxes in that area averaged from F^a was compared with those from F^b and F^0, as well as their ratios (see Fig. 11). The two figures indicate that CFI-CMAQ could in general reproduce the true fluxes with acceptable bias.

A series of numerical experiments were conducted to investigate the sensitivity of the inflation factor of the scaling factors \( \beta \) and the lag-window. As shown in Fig. 11, CFI-CMAQ worked rather well for \( \beta = 60, 70, 75, 80 \). However, if \( \beta \) was much smaller than 50 (e.g. \( \beta = 10 \)), the impact of assimilation was small due to the small ensemble spread; or if \( \beta \) was much larger than 80 (e.g. \( \beta = 100 \)), the assimilated CO_2 fluxes deviated markedly from the “true” CO_2 fluxes. In other words, the performance of CFI-CMAQ greatly relies on the choice of \( \beta \). From the perspective of the lag-window, the differences among the four assimilation sensitivity experiments with lag-windows of 3, 6, 9 and 12 days were very small (see Fig. 12). Although Peters et al. (2007) indicated that the lag-window should be more than five weeks, it seems that the smoother window had a slight influence on the assimilated results for CFI-CMAQ. It is clear that the
assimilated results with a larger lag-window were better than those with a smaller lag-window; however, CFI-CMAQ performed very well even with a short lag-window (e.g. 3 days).

4 Summary and conclusions

A regional surface CO₂ flux inversion system, CFI-CMAQ, has been developed to optimize CO₂ fluxes at finer scales. It operates under a joint data assimilation framework, simultaneously assimilating CO₂ concentrations and surface CO₂ fluxes, similar to Kang et al. (2011, 2012) and Tian et al. (2013). The smoothing operator, which was first adopted by Peters et al. (2005) as the persistence dynamical model, is associated with the atmospheric transport model for the purpose of resolving the “signal-to-noise” problem, as well as transporting the useful observed information onto the next assimilation cycle, in order to optimize the surface CO₂ fluxes as a whole at the grid scale. In this application, the scaling factors to be optimized in the flux inversion system can be forecast at the grid scale without random noise. The OSSEs showed that the performance of CFI-CMAQ is effective and promising. In general, it could reproduce the true fluxes at finer scales with acceptable bias.

This study represents the first step in developing a regional surface CO₂ flux inversion system to optimize CO₂ fluxes over East Asia, particularly over China. In future, we intend to further develop the covariance localization techniques and inflation techniques to improve the performance of CFI-CMAQ. Furthermore, the uncertainty of the boundary conditions should be considered to improve the effectiveness of regional CO₂ flux optimization.

Acknowledgements. This work was supported by the Strategic Priority Research Program–Climate Change: Carbon Budget and Relevant Issues (XDA05040404), the National High Technology Research and Development Program of China (2013AA122002). CarbonTracker results used to generate the initial condition are provided by NOAA ESRL, Boulder, Colorado, USA from the website at http://carbontracker.noaa.gov. The numerical calculations in this paper
have been done on the IBM Blade cluster system in the High Performance Computing Center (HPCC) of Nanjing University.

References


A regional carbon flux data assimilation system

Z. Peng et al.


Pillai, D., Gerbig, C., Ahmadov, R., Rödenbeck, C., Kretschmer, R., Koch, T., Thompson, R., Neininger, B., and Lavrié, J. V.: High-resolution simulations of atmospheric CO₂ over complex


Figure 1. Schematic diagram of the smoother window.
Figure 2. Flowchart of the CFI-CMAQ system used to optimize surface CO₂ fluxes. The system includes the following four parts at each optimization cycle: (1) forecasting of the background linear scaling factors \( \lambda_{b,i,t} \) (red arrows); (2) optimization of the scaling factors in the smoother window \( M \) by EnKS (see Fig. 1) (blue arrows); (3) updating of the flux in the smoother window (green arrows); and (4) assimilation of the CO₂ concentration fields at time \( t \) by EnKF (black arrows).
Figure 3. (a) Total number of observations in February 2010 in the model grid. Each symbol indicates the total number of all GOSAT $X_{CO_2}$ measurements in the corresponding model grid. Monthly mean values of (b) $X_{CO_2}^0$, column mixing ratio of $C^0$; (c) $X_{CO_2}^b$, column mixing ratio of $C^b$; (d) $X_{CO_2}^a$, column mixing ratio of $C^a$; (e) $X_{CO_2}^0 - X_{CO_2}^b$; and (f) $X_{CO_2}^a - X_{CO_2}^b$. All column mixing ratios are column-averaged with real GOSAT $X_{CO_2}$ averaging kernels at GOSAT $X_{CO_2}$ locations. Each symbol indicates the monthly average value of all $X_{CO_2}$ estimates in the model grid.
Figure 4. Monthly mean values of (a) $C^0$, the artificial true simulations driven by the prior surface CO$_2$ fluxes $F^0$; (b) $C^b$, the background simulations driven by magnified surface CO$_2$ fluxes $F^b = 1.8 \cdot F^0$; (c) $C^a$, assimilations; (d) $C^0 - C^b$; (e) $C^0 - C^a$; and (f) $100 \cdot (C^0 - C^a)/C^0$ at model-level 1 in February 2010. Black lines EF and GH indicate the positions of the cross sections shown in Fig. 5.
Figure 5. Monthly mean cross sections of $C^0 - C^b$ along line (a) EF and (b) GH, and monthly mean cross sections of $C^0 - C^a$ along line (c) EF and (d) GH (cross section lines shown in Fig. 4d).
Figure 6. Daily mean time series of CO$_2$ concentrations at national background stations in China and their nearest large cities from 1 January to 20 March 2010 extracted from the artificial true simulations $C^0$ (black), background simulations $C^b$ (red), and assimilations (blue). All time series were interpolated to the observation locations by the spatial bilinear interpolator method. The sites used are (a) Waliguan (36.28° N, 100.91° E), (b) Xining (36.56° N, 101.74° E), (c) Longfengshan (44.73° N, 127.6° E), (d) Haerbin (45.75° N, 126.63° E), (e) Shangdianzi (40.65° N, 117.12° E), (f) Beijing (39.92° N, 116.46° E), (g) Linan (30.3° N, 119.73° E), and (h) Hangzhou (30.3° N, 120.2° E).
Figure 7. Monthly mean values in February 2010 of (a) $F^0$, the prior true surface CO$_2$ fluxes; (b) $F^b$, the background CO$_2$ fluxes, $F^b(x, y, z, t) = (1.8 + \delta)F^0(x, y, z, t)$; (c) $F^a$, the assimilated surface CO$_2$ fluxes; (d) $F^0 - F^b$; (e) $F^a - F^b$; and (f) $F^a - F^0$ (units: μmole m$^{-2}$ s$^{-1}$).
Figure 8. Monthly mean RMSEs of $F_a$ in February 2010 (units: μmole m$^{-2}$ s$^{-1}$).
Figure 9. (a) Ratios of monthly mean $F^0$ to monthly mean $F^b$; (b) ratios of monthly mean $F^a$ to monthly mean $F^b$; and (c) ratios of monthly mean $F^a$ to monthly mean $F^0$ in February 2010. The white part indicates the ratios where the absolute values of monthly mean $F^b$ are larger than 0.01, not analyzed in this study. The black square labeled I indicates the domain where surface CO$_2$ fluxes were used for the results presented in Fig. 12.
**Figure 10.** Daily mean time series of CO$_2$ fluxes at national background stations in China and their nearest large cities from 1 January to 20 March 2010 extracted from the prior true surface CO$_2$ fluxes $F^0$ (black), background CO$_2$ fluxes $F^b$ (red), and assimilated CO$_2$ fluxes $F^a$ (blue). All time series were interpolated to the observation locations by the spatial bilinear interpolator method. The sites used are (a) Waliguan, (b) Xining, (c) Longfengshan, (d) Haerbin, (e) Shangdianzi, (f) Beijing, (g) Linan, and (h) Hangzhou.
Figure 11. Time series of daily mean CO$_2$ fluxes averaged in domain I (shown in Fig. 9b) from 1 January to 20 March 2010 with the inflation factor of scaling factors $\beta = 70$, 75 and 80. The black dashed line is the time series averaged from $F^b$ and the black solid line is the time series averaged from $F^0$. 
Figure 12. Time series of daily mean CO$_2$ fluxes averaged in domain I (shown in Fig. 9b) from 1 January to 20 March 2010 with different smoother windows (3, 6, 9 and 12 days). The black dashed line is the time series averaged from $F^b$ and the black solid line is the time series averaged from $F^0$. 