A Regional Carbon Data Assimilation System and its Preliminary Evaluation in East Asia

Zhen Peng[* 1], Meigen Zhang[* 2], Xingxia Kou[2, 3], Xiangjun Tian[4], and Xiaoguang Ma[4]

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

[*]Corresponding author: pengzhen@nju.edu.cn;
[*]Corresponding author: mgzhang@mail.iap.ac.cn;

ABSTRACT

In order to optimize surface CO_2 fluxes at gird 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 applying the ensemble Kalman filter (EnKF) to constrain the CO_2 concentrations and applying the ensemble Kalman smoother (EnKS) to optimize the surface CO_2 fluxes. 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 EnKF could constrain the CO_2 concentration 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 EnKS demonstrated that CFI-CMAQ could in general reproduce true fluxes at grid 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

Considerable progress has been made in recent years to reduce the uncertainties of surface CO_2 flux estimates through the use of an advanced data assimilation technique (e.g., Chevallier et al., 2005, 2007a and 2007b; Baker et al., 2006; Engelen et al., 2009; Liu et al., 2012). Feng et al. (2009) showed that the uncertainties of surface CO_2 flux estimates can be reduced significantly by assimilating OCO X_{CO2} measurements. Peters et al. (2005, 2007, 2009) developed a surface CO_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_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_2 fluxes over Europe and Asia (eg., Zhang et al., 2014a, 2014b). Kang et al. (2012) presented a simultaneous data assimilation of surface CO_2 fluxes and atmospheric CO_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_2 flux data assimilation system, Tan-Tracker, 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 through the use of simultaneous assimilation of CO_2 concentrations and CO_2 surface fluxes. Despite the rigor of data assimilation theory, current CO_2 flux-inversion methods 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). And 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., 2009). Marland (2008) pointed out that even a tiny amount of uncertainty, i.e., 0.9%, in one of the leading emitter countries like the U.S. 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 finer scale.

In CarbonTracker (Peters at 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 to optimize the surface CO_2 fluxes as a whole at the grid scale. This process is described in detail in section 2 of this paper.

The surface CO_2 flux inversion system presented in this paper was developed by simultaneous optimizing the surface CO_2 fluxes and constraining the CO_2 concentrations. As we know, assimilating CO_2 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). In addition, previous studies showed that the simultaneous assimilation of CO_2 concentrations and surface CO_2 fluxes can largely eliminate the uncertainty in initial CO_2 concentrations on the CO_2 evolution (Kang et al., 2012; Tian et al., 2013). Therefore, we also use the simultaneous assimilation framework and the ensemble Kalman filter (EnKF) was used to constrain CO_2 concentrations and the ensemble Kalman smoother (EnKS) was used to optimize surface CO_2 fluxes. Since the regional chemical transport models, compared to global models, have some advantages to reproduce the effects of meso–micro–scale transport on atmospheric CO_2 distributions (Ahmadov et al., 2009, Pillai et al., 2010; Kretschmer et al., 2011), we choose a regional model, Regional Atmospheric Modeling System and Community Multi-scale Air Quality (RAMS-CMAQ) (Zhang et al. 2002, 2003, 2007; Kou et al. 2013; Liu et al., 2013; Huang et al. 2014), to develop this 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 CO_2 fluxes. The retrieval 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 CO_2 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

Supposed we have the prescribed net CO_2 surface flux, $F^*(x, y, z, t)$, which can be released from a climate model or be generated by others methods, our ultimate goal is to optimize $F^*(x, y, z, t)$ by assimilating CO_2 observations from various platforms. 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). The *i*th ensemble member of the surface fluxes, $F_i(x,y,z,t)$, from an N-member ensemble can be described by

$$F_i(x,y,z,t) = \lambda_i(x,y,z,t)F^*(x,y,z,t), \quad (i=1,\cdots,N), \tag{1}$$

where $\lambda_i(x,y,z,t)$ represents the ith ensemble member of the linear scaling factors (Peters et al., 2007; Tian et al., 2013) for each time and each grid to be optimized in the assimilation. The notations are standard: the subscript i refers to the ith ensemble member. In the following, $\lambda_i(x,y,z,t)$ is referred to as $\boldsymbol{\lambda}_{i,t}$, $F^*(x,y,z,t)$ is referred to as $\boldsymbol{F}_t^*$, and $F_i(x,y,z,t)$ is referred to as $\boldsymbol{F}_{i,t}$ for simplicity.

At each optimization cycle, CFI-CMAQ includes the following four parts in turn (see Fig. 1): (1) forecasting of the linear scaling factors at time t, $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$; (2) optimization of the scaling factors in the smoother window, $(\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M+1|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t1|t-1})$, by EnKS, Where $\boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}$ $(j=t-1-M,\cdots,t-1)$ refer to analyzed quantities from the previous assimilation cycle at time j (see Fig. 1), $|t-1$ means that these factors have been updated by using observations before time $t-1$, and the superscript a refers to the analyzed; (3) updating of the fluxes in the smoother window, $(\boldsymbol{F}^{\mathrm{a}}_{i,t-M|t-1}, \boldsymbol{F}^{\mathrm{a}}_{i,t-M+1|t-1}, \cdots, \boldsymbol{F}^{\mathrm{a}}_{i,j|t-1}, \cdots, \boldsymbol{F}^{\mathrm{a}}_{i,t-1|t-1}, \boldsymbol{F}^{\mathrm{a}}_{i,t|t-1})$; and (4) assimilation of the forecast CO_2 concentration fields at time t, $C_i^{\mathrm{f}}(x,y,z,t)$ (referred to as $\boldsymbol{C}^{\mathrm{f}}_{i,t}$, and the superscript f refers to the forecast or the background), 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_{CO2} data in Sect. 2.4. Furthermore, covariance inflation and localization techniques applied in CFI-CMAQ are introduced briefly in Sect. 2.5.

2.1 Forecasting the linear scaling factors at time t, $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$

In the previous assimilation cycle t-1-M~t-1 (see Fig. 1), the optimized scaling factors in the smoother window are $(\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1-M|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M+1|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1|t-1})$ and the assimilated CO_2 concentration fields at time t-1 are $C^{\mathrm{a}}_i(x, y, z, t-1)$ (referred to as $\boldsymbol{C}^{\mathrm{a}}_{i,t-1}$). In the current assimilation cycle t-M~t, the scaling factors in the current smoother window are $(\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M+1|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1})$ and the forecast CO_2 concentration fields at time t are $\boldsymbol{C}^{\mathrm{f}}_{i,t}$.

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 part of the persistence dynamical model to calculate the linear scaling factors $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$,

$$\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1} = \frac{\left(\sum_{j=t-M}^{t-1} \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1} + \boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}\right)}{M+1}, \qquad (i = 1, \cdots, N), \tag{2}$$

where $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ refers to the prior values of the linear scaling factors at time t. The superscript 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 $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$ in time.

In order to generate $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$, the atmospheric transport model (CMAQ) is applied and a set of ensemble forecast experiments are carried out. It integrates from time $t-1$ to t to produce the CO_2 concentration fields $\hat{C}^{\mathrm{f}}_i(x, y, z, t)$ (referred to as $\hat{C}^{\mathrm{f}}_{i,t}$ hereafter to distinguish from $\boldsymbol{C}^{\mathrm{f}}_{i,t}$) forced by the prescribed net CO_2 surface flux $\boldsymbol{F}^*_t$ with $\boldsymbol{C}^{\mathrm{a}}_{i,t-1}$ as initial conditions. Then, the ratio $\boldsymbol{\kappa}_{i,t} = \hat{C}^{\mathrm{f}}_{i,t} \big/ \overline{\hat{C}^{\mathrm{f}}_{i,t}}$ is calculated, where $\overline{\hat{C}^{\mathrm{f}}_{i,t}} = \frac{1}{N}\sum_{i=1}^{N} \hat{C}^{\mathrm{f}}_{i,t}$. Supposed that $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1} = \boldsymbol{\kappa}_{i,t}$ due to the fact that the surface

CO_2 fluxes correlate with its concentrations, the values for $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ are obtained and then $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$ can finally be calculated (see the part with red arrows in the flowchart in Fig. 2).

The way the prior scaling factor $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ is updated by associating with the atmospheric transport model is the main improvement over the one used in CarbonTracker (Peters et al, 2007). In CarbonTracker, all $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ are set to 1 (Peters et al., 2007). The distribution of the ensemble members of the linear scaling factors at time t, $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$, are finally dependent on the distribution of the previous scaling factors because Eq. (2) is a linear smoothing operator. In this study, the values of $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ are updated by associating with the atmospheric transport model. It is important to note that $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ in this study are rand fields with mean 1. However, the distribution of $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$ are dependent on the distribution of all the scaling factors in the smoother window. An OSSE was designed to illustrate the difference between our method and the one where $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ are set to 1 in Sect. 3

It is also 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 section 2.5.

2.2 Optimizing the scaling factors in the smoother window by EnKS

Substituting $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$ into Eq. (1), the ith member of the surface fluxes at time t, $\boldsymbol{F}^{\mathrm{a}}_{i,t|t-1}$, can be generated. Then forced by $\boldsymbol{F}^{\mathrm{a}}_{i,t|t-1}$, CMAQ was run from time $t-1$ to t to produce the background concentration field $\boldsymbol{C}^{\mathrm{f}}_{i,t}$ with $\boldsymbol{C}^{\mathrm{a}}_{i,t-1}$ as initial conditions.

In the current assimilation cycle t-M~t (see Fig. 1), the scaling factors to be optimized in the smoother window are $(\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M+1|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1})$, as stated in the first paragraph of Sect. 2.1. Using the EnKS analysis technique, these scaling factors are updated in turn via

$$\boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t} = \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1} + \mathbf{K}^{\mathrm{e}}_{j,t|t-1}(\boldsymbol{y}^{\mathrm{obs}}_{t} - \boldsymbol{y}^{\mathrm{f}}_{i,t} + \boldsymbol{\upsilon}_{i,t}),\ (i = 1, \cdots, N,\ j = t-M, \cdots, t), \tag{3}$$

$$\mathbf{K}^{\mathrm{e}}_{j,t|t-1} = \mathbf{S}^{\mathrm{e}}_{j,t|t-1} H^{T} (H\mathbf{S}^{\mathrm{e}}_{t,t|t-1} H^{T} + \mathbf{R})^{-1}, \tag{4}$$

$$\mathbf{S}^{e}_{j,t|t-1} = \frac{1}{N-1}\sum_{i=1}^{N}[\boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1} - \overline{\boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}}][\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1} - \overline{\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}}]^{T}, \tag{5}$$

$$\mathbf{S}^{\mathrm{e}}_{t,t|t-1} = \frac{1}{N-1}\sum_{i=1}^{N}[\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1} - \overline{\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}}][\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1} - \overline{\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}}]^{T}, \tag{6}$$

$$\boldsymbol{y}^{\mathrm{f}}_{i,t} = H(\boldsymbol{\varphi}_{t-1\to t}(\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1})) = H(\boldsymbol{C}^{\mathrm{f}}_{i,t}), \tag{7}$$

where $\mathbf{K}^{\mathrm{e}}_{j,t|t-1}$ is the Kalman gain matrix of EnKS, $\boldsymbol{y}^{\mathrm{obs}}_{t}$ is the observation vector measured at time t and $\boldsymbol{y}^{\mathrm{f}}_{i,t}$ is the simulated values, $\boldsymbol{\upsilon}_{i,t}$ is a random normal distribution perturbation field with zero mean, $\mathbf{S}^{\mathrm{e}}_{j,t|t-1}$ is the background error cross-covariance between the state vector $\boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}$ and $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$, $\mathbf{S}^{\mathrm{e}}_{t,t|t-1}$ is the background error covariance of the state vector $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$, $H(\cdot)$ is the observation operator that maps the state variable from model space into observation space, $\mathbf{R}$ standard deviation representing the measurement errors, and $\boldsymbol{\varphi}(\cdot)$ is the atmospheric transport model.

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

$$\mathbf{S}^{\mathrm{e}}_{j,t|t-1}H^{T} = \frac{1}{N-1}\sum_{i=1}^{N}[\boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1} - \overline{\boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}}][\boldsymbol{y}^{\mathrm{f}}_{i,t} - \overline{\boldsymbol{y}^{\mathrm{f}}_{t}}]^{T}, \tag{8}$$

$$H\mathbf{S}^{\mathrm{e}}_{t,t|t-1}H^T = \frac{1}{N-1}\sum_{i=1}^{N}[\boldsymbol{y}^{\mathrm{f}}_{i,t} - \overline{\boldsymbol{y}^{\mathrm{f}}_t}][\boldsymbol{y}^{\mathrm{f}}_{i,t} - \overline{\boldsymbol{y}^{\mathrm{f}}_t}]^T \ , \tag{9}$$

$$\overline{\boldsymbol{y}^{\mathrm{f}}_t} = H(\overline{\boldsymbol{C}^{\mathrm{f}}_t}) = H(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{C}^{\mathrm{f}}_{i,t}) \ . \tag{10}$$

After EnKS, $(\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M|t}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M+1|t}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1|t}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t})$ are gained. Then the corresponding fluxes in the smoother window $(\boldsymbol{F}^{\mathrm{a}}_{i,t-M|t}, \boldsymbol{F}^{\mathrm{a}}_{i,t-M+1|t}, \cdots, \boldsymbol{F}^{\mathrm{a}}_{i,j|t}, \cdots, \boldsymbol{F}^{\mathrm{a}}_{i,t-1|t}, \boldsymbol{F}^{\mathrm{a}}_{i,t|t})$ can be gained (see the part with green arrows in the flowchart in Fig. 2) by substituting $(\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M|t}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M+1|t}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1|t}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t})$ into Eq. (1).

Then the ensemble mean values of the assimilated fluxes in the smoother window can be calculated via,

$$\overline{\boldsymbol{F}^{\mathrm{a}}_{i,j|t}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{F}^{\mathrm{a}}_{i,j|t}, \ (j = t-M, \cdots, t) \ , \tag{11}$$

Finally, those ensemble mean assimilated fluxes which are before the next smoother window and will not be updated by the succeeding observations are regarded as the final optimized fluxes. We referred them as $\overline{\boldsymbol{F}^{\mathrm{a}}_t}$ for simplicity.

2.3 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

$$\boldsymbol{C}^{\mathrm{a}}_{i,t} = \boldsymbol{C}^{\mathrm{f}}_{i,t} + \mathbf{K}(\boldsymbol{y}^{\mathrm{obs}}_t - \boldsymbol{y}^{\mathrm{f}}_t + \boldsymbol{\upsilon}_{i,t}) \ , \tag{12}$$

$$\mathbf{K} = \mathbf{P}^{\mathrm{f}} H^T (H\mathbf{P}^{\mathrm{f}} H^T + \mathbf{R})^{-1}, \tag{13}$$

where $\mathbf{K}$ is the Kalman gain matrix of EnKF, $\mathbf{P}^{\mathrm{f}}$ is the background error covariance among the background CO_2 concentration fields $\boldsymbol{C}^{\mathrm{f}}_{i,t}$.

In actually application, $\mathbf{P}^{\mathrm{f}} H^T$ and $H\mathbf{P}^{\mathrm{f}} H^T$ can be calculated as a whole by

$$\mathbf{P}^{\mathrm{f}} H^{T} = \frac{1}{N-1}\sum_{i=1}^{N}[\boldsymbol{C}_{i,t}^{\mathrm{f}} - \overline{\boldsymbol{C}_{t}^{\mathrm{f}}}][\boldsymbol{y}_{i,t}^{\mathrm{f}} - \overline{\boldsymbol{y}_{t}^{\mathrm{f}}}]^{T} \quad , \tag{14}$$

$$H\mathbf{P}^{\mathrm{f}} H^{T} = \frac{1}{N-1}\sum_{i=1}^{N}[\boldsymbol{y}_{i,t}^{\mathrm{f}} - \overline{\boldsymbol{y}_{t}^{\mathrm{f}}}]^{T}[\boldsymbol{y}_{i,t}^{\mathrm{f}} - \overline{\boldsymbol{y}_{t}^{\mathrm{f}}}]^{T} , \tag{15}$$

$$\overline{\boldsymbol{C}_{t}^{\mathrm{f}}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{C}_{i,t}^{\mathrm{f}} \tag{16}$$

Finally, the ensemble mean values of the assimilated CO_2 concentrations fields can be gained via,

$$\overline{\boldsymbol{C}_{t}^{\mathrm{a}}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{C}_{i,t}^{\mathrm{a}} \tag{17}$$

where $\overline{\boldsymbol{C}_{t}^{\mathrm{a}}}$ is regarded as the final analyzing concentration field.

2.4 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_{CO2} retrieval is a weighted CO_2 column average, the simulated X_{CO2} 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_{CO2} retrieval is

$$\boldsymbol{y}_{i,t}^{\mathrm{f}} = H(\boldsymbol{\varphi}_{t-1\to t}(\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{a}})) = H(\boldsymbol{C}_{i,t}^{\mathrm{f}}) = \boldsymbol{y}^{\mathrm{priori}} + \boldsymbol{h}^{T}\boldsymbol{a}_{\mathrm{CO2}}(S(\boldsymbol{C}_{i,t}^{\mathrm{f}}) - \boldsymbol{f}^{\mathrm{priori}}), \tag{18}$$

where $\boldsymbol{y}_{i,t}^{\mathrm{f}}$ is the simulated X_{CO2}; $\boldsymbol{y}^{\mathrm{priori}}$ is the a priori CO_2 column average used in the GOSAT X_{CO2} retrieval process; $S(\cdot)$ is the spatial bilinear interpolation operator that interpolates the simulated fields to the GOSAT $\mathrm{X}_{\mathrm{CO2}}$ locations to obtain the simulated CO_2 vertical profiles there; f^{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 $\boldsymbol{a}_{CO2}$ is the normalized averaging kernel.

2.5 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,

$$(\boldsymbol{C}_{i,t}^{\mathrm{f}})_{\mathrm{new}} = \alpha(\boldsymbol{C}_{i,t}^{\mathrm{f}} - \overline{\boldsymbol{C}_{i,t}^{\mathrm{f}}}) + \overline{\boldsymbol{C}_{i,t}^{\mathrm{f}}}, \tag{19}$$

where α is the inflation factor of CO_2 concentrations and $(\boldsymbol{C}_{i,t}^{\mathrm{f}})_{\mathrm{new}}$ 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,

$$(\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}})_{\mathrm{new}} = \beta(\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}} - \overline{\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}}}) + \overline{\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}}}, \tag{20}$$

where β is the inflation factor of scaling factors and $(\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}})_{\mathrm{new}}$ 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 $\mathbf{K}_{j,t|t-1}^{\mathrm{e}}$ and $\mathbf{K}$ are updated via,

$$\mathbf{K}_{j,t|t-1}^{\mathrm{e}} = [(\boldsymbol{\rho} \circ \mathbf{S}_{j,t|t-1}^{\mathrm{e}})H^T (H(\boldsymbol{\rho} \circ \mathbf{P}_{t,t|t-1}^{\mathrm{e}})H^T + \mathbf{R})^{-1}, \tag{21}$$

$$\mathbf{K} = [(\boldsymbol{\rho} \circ \mathbf{P}^{\mathrm{f}})H^T][(H(\boldsymbol{\rho} \circ \mathbf{P}^{\mathrm{f}})H^T + \mathbf{R}]^{-1}, \tag{22}$$

where the filtering matrix $\boldsymbol{\rho}$ is calculated using the formula

$$C_0(r,c)=\begin{cases}-\frac{1}{4}\left(\frac{|r|}{c}\right)^5+\frac{1}{2}\left(\frac{|r|}{c}\right)^4+\frac{5}{8}\left(\frac{|r|}{c}\right)^3-\frac{5}{3}\left(\frac{|r|}{c}\right)^2+1, & 0\le |r|\le c\\ \frac{1}{12}\left(\frac{|r|}{c}\right)^5-\frac{1}{2}\left(\frac{|r|}{c}\right)^4+\frac{5}{8}\left(\frac{|r|}{c}\right)^3+ & \\ \quad \frac{5}{3}\left(\frac{|r|}{c}\right)^2-5\left(\frac{|r|}{c}\right)+4-\frac{2}{3}\left(\frac{c}{|r|}\right), & \mathrm{c}\le |r|\le 2c\\ 0, & \mathrm{c}\le |r| \end{cases} \tag{23}$$

,

where c is the element of the localization Schur radius. The matrix $\boldsymbol{\rho}$ 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 × 5440 km^2 on a rotated polar stereographic map projection centered at (35.0°N, 116.0°E), with a horizontal grid resolution of 64 × 64 km^2 and 15 vertical layers in the σ_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),

$$F^{\mathrm{p}}(x,y,z,t)=F_{\mathrm{bio}}(x,y,z,t)+F_{\mathrm{oce}}(x,y,z,t)+F_{\mathrm{ff}}(x,y,z,t)+F_{\mathrm{fire}}(x,y,z,t), \quad (24)$$

where $F^{\mathrm{p}}(x,y,z,t)$ (referred to as $\boldsymbol{F}_t^{\mathrm{p}}$) was the prior surface CO_2 flux; $F_{\mathrm{bio}}(x,y,z,t)$ and $F_{\mathrm{oce}}(x,y,z,t)$ were the biosphere–atmosphere and ocean–atmosphere CO_2 fluxes, respectively, which were obtained from the optimized results of CarbonTracker 2011 (Peters, 2007); $F_{\mathrm{ff}}(x,y,z,t)$ was fossil fuel emissions, adopted from the Regional Emission inventory in ASia (REAS, 2005 Asia monthly mean emission inventory) with a spatial resolution of $0.5°\times0.5°$(Ohara et al., 2007); $F_{\mathrm{fire}}(x,y,z,t)$ was biomass–burning emissions, provided by the monthly mean inventory at a spatial resolution of $0.5° \times 0.5°$ from the Global Fire Emissions Database, Version 3 (GFED v3) (Van der Werf et al., 2010). Among all these fluxes, $F_{\mathrm{bio}}(x,y,z,t)$, $F_{\mathrm{oce}}(x,y,z,t)$ and $F_{\mathrm{ff}}(x,y,z,t)$ had nonzero values at model level 1, while they all were zeros at other 14 levels. However, $F_{\mathrm{fire}}(x,y,z,t)$ had nonzero values at model level 1~5 and they were all zeros at other 10 levels. So, all fluxes in this paper were the function of (x,y,z,t) for convenience.

Firstly, the prior flux $\boldsymbol{F}_t^{\mathrm{p}}$ was assumed as the true surface CO_2 flux in all of the following OSSEs. Forced by $\boldsymbol{F}_t^{\mathrm{p}}$, the RAMS-CMAQ model was run to produce the artificial true CO_2 concentration results $C^{p}(x,y,z,t)$ (refer to as $\boldsymbol{C}_t^{\mathrm{p}}$ in the following). Then, the artificial GOSAT observations $\boldsymbol{y}_{\mathrm{t}}^{\mathrm{obs}}$ (or X_{CO2}^{p}) were generated by substituting $\boldsymbol{C}_t^{\mathrm{p}}$ into the observation operator in Eq. (18). The retrieval information of GOSAT X_{CO2}($\boldsymbol{y}^{\mathrm{priori}}$, $f^{\,priori}$, h and $\boldsymbol{a}_{CO2}$) needed in Eq. (18) were gained from the v2.9 Atmospheric CO_2 Observations from Space (ACOS) Level 2 standard data products, which only utilized the SWIR observations. Only data classified into the "Good" category were utilized in this study. During the retrieval process, most of the soundings (such as data with a solar zenith angle greater than 85°, or data not in clear sky conditions, or data collected over ocean but not in glint, etc.) were not processed, so typically data products for the "Good" category contained only 10-100 soundings per satellite orbit (Osterman et al., 2011), and there were only 0~60 samples per orbit in the study model domain generally. Fig. 3 (a) also showed the total number of "good" GOSAT X_{CO2} observations for each model grid in February in 2010. There were relatively more observations over most continental regions of the study domain except some regions in North-East and South China. The total numbers ranged from 1 to 8. However, there were almost no data over oceans of the study domain.

Secondly, the prescribed surface CO_2 fluxes series $\boldsymbol{F}_t^*$ were created by

$$\boldsymbol{F}_t^* = (1.8 + \delta(x, y, z, t))\boldsymbol{F}_t^{\mathrm{p}}, \tag{25}$$

where δ was a random number. They were standard normal distribution time series at each grid in the integration period of our numerical experiment. Driven by $\boldsymbol{F}_t^*$, the RAMS-CMAQ model was integrated to obtain the CO_2 simulations $C^f(x, y,, z, t)$ (referred to as $\boldsymbol{C}_t^{\mathrm{f}}$ hereafter). Then, the column-averaged concentrations X_{CO2}^{f} were obtained using Eq. (18).

The performance of CFI-CMAQ was evaluated through a group of well-designed OSSEs. And the goal of each OSSE was to retrieve the true fluxes $\boldsymbol{F}_t^{\mathrm{p}}$ from given true observations X_{CO2}^{p} and "wrong" fluxes $\boldsymbol{F}_t^*$. In all the OSSEs, we assimilated artificial observations X_{CO2}^{p} about three times a day since GOSAT has about three orbits in the study model domain. If there were some observations, CFI-CMAQ paused to assimilate. Otherwise, it continued simulating. The default ensemble size N was 48, the measurement errors were 1.5 ppmv, the standard localization Schur radius c was 1280 km (20 grid spacing), and the covariance inflation factor of concentrations α was 1.1. The referenced lag-window was 9 days and the covariance inflation factor of the prior scaling factors β was 70. Since the smoother window was very important for CO_2 transportation and β was a newly introduced parameter, both these parameters were further investigated by several numerical sensitivity experiments. The primary focus of this paper was to describe the assimilation methodology, so all the numerical experiments started on 1 January 2010 and ended on 30 March 2010.

As for the initialization of CFI-CMAQ, only the ensemble of background concentration fields $C_{i,0}^{\mathrm{f}}$ needed to be initialized at the time $t=0$ because the values of $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{a}}$ were updated by using the persistence dynamical model. In practice, the mean concentration fields at $t=0$ are interpolated from the simulated CO2 fields of CarbonTracker 2011 (Peters, 2007). The ensemble members of the background concentration fields were created by adding random vectors. The mean values of the random vectors were zero and the variances were 2.5 percent of the mean concentration fields. Then the atmospheric transport model integrated from time $t=0$ to $t=1$ driven by $\boldsymbol{F}_t^*$ with $C_{i,0}^{\mathrm{f}}$ as initial conditions to produce the CO_2

concentration fields $\hat{C}^{\mathrm{f}}_{i,1}$. And then the first prior linear scaling factors, $\boldsymbol{\lambda}^{\mathrm{p}}_{i,1|0}$, could be calculated by applying $\hat{C}^{\mathrm{f}}_{i,1}$. Assumed $\boldsymbol{\lambda}^{\mathrm{a}}_{i,1|0} = \boldsymbol{\lambda}^{\mathrm{p}}_{i,1|0}$, $\boldsymbol{\lambda}^{\mathrm{a}}_{i,1|0}$ are gained finally. For the first assimilation cycle, the lag-window was only one (that is, only $\boldsymbol{\lambda}^{\mathrm{a}}_{i,1|0}$ needed to be optimized in the first assimilation cycle). And it increased for the first dozens of assimilation cycles until it reached M+1 as CFI-CMAQ continued to assimilate observations. Once the system was initialized, all future scaling factors could be created using the persistence dynamical model, which was associated the smoothing operator with the atmospheric transport model.

In order to illustrate the limitation by only using the smoothing operator as the persistence dynamical model to generate all future scaling factors, another OSSE (referred to as the reference experiment to distinguish it from the above-mentioned CFI-CMAQ OSSEs) was designed to optimize the surface CO_2 fluxes at grid scale. The reference experiment was under the same assimilation framework as CFI-CMAQ except that all $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ were set to 1 (Peters et al., 2007). Besides, the initialization procedure of the reference experiment was different from that of the CFI-CMAQ. In practice, both the ensemble of background concentration fields at $t=0$, $C^{\mathrm{f}}_{i,0}$, and the ensemble members of the scaling factors at $t=1, \boldsymbol{\lambda}^{\mathrm{a}}_{i,1|0}$, needed to be initialized because they could not generated by other ways (Peters et al., 2005). The initial concentration fields $C^{\mathrm{f}}_{i,0}$ were created using the same method as that was used to generate $C^{\mathrm{f}}_{i,0}$ for the CFI-CMAQ OSSEs. The ensemble members of the scaling factors $\boldsymbol{\lambda}^{\mathrm{a}}_{i,1|0}$ were rand fields. Their mean values were 1 and their variances were 0.1. In addition, in order to keep the ensemble spread of the scaling factors $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$ at a certain level and compensate for dynamical model error, covariance inflation was also used and the covariance inflation factor of the scaling factors $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$ was 1.6. All other parameters are the same as used in the CFI-CMAQ OSSEs. The ensemble size N was 48, the measurement errors were 1.5 ppmv, the standard localization Schur radius c was 1280 km, the covariance inflation factor of concentrations α was 1.1, and the lag-window was 9 days.

3.2 Experimental results

Essentially, the assimilation part of CFI-CMAQ includes two subsections: one for the CO_2 concentration assimilation with EnKF, which can provide a convincing CO_2 initial analysis fields for the next assimilation cycle; and the other for the CO_2 flux optimization with EnKS, which can provide better estimation of the scaling factors for the next time through the persistence dynamical model except for optimized CO_2 fluxes. The performance of the EnKF subsection will be greatly influenced by the validation of the EnKS subsection, or vice versa. Firstly, the performance of CFI-CMAQ will be quantitatively assessed in detail by using the assimilated results of a CFI-CMAQ OSSE, in which the lag-window was 9 days and β was 70. Then the sensitivities of β and the lag-window will be discussed in the following two paragraphs. And finally, the assimilation results of the reference experiment in which $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ were set to 1 will be described in brief at the end of this subsection.

We begin by describing the impacts of assimilating artificial observations X^{p}_{CO2} on CO_2 simulations by CFI-CMAQ. As shown in Figs. 4a, 4b and 4d, the monthly mean values of the background CO_2 concentrations $\boldsymbol{C}^{\mathrm{f}}_{t}$ produced by the magnified surface CO_2 fluxes $\boldsymbol{F}_t^{*}$ were much larger than those of the artificial true CO_2 concentrations $\boldsymbol{C}_t^{\mathrm{p}}$ produced by the prior surface CO_2 fluxes $\boldsymbol{F}_t^{\mathrm{p}}$ near the surface in February 2010. In the east and south of China especially, the magnitude of the difference between $\boldsymbol{C}_t^{\mathrm{p}}$ and $\boldsymbol{C}_t^{\mathrm{f}}$ was at least 6 ppmv. Also, as expected, the monthly mean X_{CO2}^{f} was much larger than the monthly mean artificial observations X_{CO2}^{p}, and the magnitude of the difference between X_{CO2}^{p} and X_{CO2}^{f} reached 2 ppmv in the east and south of China (see Figs. 3b, 3c and 3e). 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 model-level 7 (approximately 1.6 km). And above model-level 10, the differences between $\boldsymbol{C}_t^{\mathrm{p}}$ and $\boldsymbol{C}_t^{\mathrm{f}}$ fell to zero (see Fig. 5a and 5b). After assimilating X_{CO2}^{p}, the analysis CO_2 concentrations $\overline{\boldsymbol{C}_t^{\mathrm{a}}}$ was much closer to $\boldsymbol{C}_t^{\mathrm{p}}$ (see Figs. 4c, 4e and 4f). The monthly mean difference between $\boldsymbol{C}_t^{\mathrm{p}}$ and $\overline{\boldsymbol{C}_t^{\mathrm{a}}}$ ranged from −2 to 2 ppmv and the relative error ($\boldsymbol{C}_t^{\mathrm{p}}$ —$\overline{\boldsymbol{C}_t^{\mathrm{a}}}$)/ $\boldsymbol{C}_t^{\mathrm{p}}$ ranged from −1 to 1% in almost the entire model domain at model-level 1. The monthly mean differences between $\boldsymbol{C}_t^{\mathrm{p}}$ and $\overline{\boldsymbol{C}_t^{\mathrm{a}}}$ were negligible above model-level 2 (see Fig. 5c and 5d). The monthly mean X_{CO2}^{a} was also closer to X_{CO2}^{p} and the difference between X_{CO2}^{p} and X_{CO2}^{a} ranged from −0.5 to 0.5 ppmv. In order to evaluate the general impact of assimilating X_{CO2}^{p} 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. 6. 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_{CO2}^{p}, the assimilated time series were very close 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_{CO2} with EnKF are clear. All the results illustrated that CFI-CMAQ can provide a convincing CO_2 initial analysis fields for CO_2 flux inversion.

The impacts of assimilating X_{CO2}^{p} on surface CO_2 fluxes were also highly impressive by CFI-CMAQ. On the whole, the prescribed CO_2 surface fluxes $\boldsymbol{F}_t^*$ were much larger than the true surface CO_2 fluxes $\boldsymbol{F}_t^{\mathrm{p}}$ in February 2010, especially in the east and south of China. The monthly mean difference between $\boldsymbol{F}_t^*$ and $\boldsymbol{F}_t^{\mathrm{p}}$ reached 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 Figs. 7a, 7b and 7d). After assimilating X_{CO2}^{p}, the ensemble mean of the assimilated surface CO_2 fluxes $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ decreased sharply. Thus, the monthly mean values of $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ were much smaller than $\boldsymbol{F}_t^*$ in most of the model domain in February 2010. The pattern of the difference between $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ and $\boldsymbol{F}_t^*$ was similar to that of the difference between $\boldsymbol{F}_t^{\mathrm{p}}$ and $\boldsymbol{F}_t^*$ (see Fig. 7d). The ensemble mean of the assimilated surface CO_2 fluxes $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ were also compared to the artificial true fluxes $\boldsymbol{F}_t^{\mathrm{p}}$, revealing that $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ was equivalent to $\boldsymbol{F}_t^{\mathrm{p}}$ in most of the model domain. The monthly mean difference between $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ and $\boldsymbol{F}_t^{\mathrm{p}}$ ranged from −0.1 to 0.1 μmole m^{-2} s^{-1} only (see Fig. 7e). 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 0.5 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 $\boldsymbol{F}_t^*$ to the monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$ were analyzed. In actual implementation, we only analyzed the ratios where the absolute values of the monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$ were larger than 0.1, to avoid random noise. As shown in Fig. 9a, the ratios of the monthly mean $\boldsymbol{F}_t^*$ to the monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$ are about 1.8 in most of China, except in the Qinghai–Tibet Plateau, where the absolute values of the monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$ in February were very small and we did not analyze. In addition, the ratios of the monthly mean $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ to the monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$ are shown in Fig. 9b. This figure demonstrates that the impact of the assimilation of X_{CO2}^{p} 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.

Time series of daily mean surface CO_2 fluxes extracted from $\boldsymbol{F}_t^*$ and $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ were also compared with that from $\boldsymbol{F}_t^{\mathrm{p}}$ 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.10. 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_{CO2}^{p}, 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. One of the possible reasons was that the values of the ensemble spread of $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{a}}$ in Beijing area are too large (see Fig. 11c). Beijing was located in Jing-Jin-Ji Urban Circle, which had strong local anthropogenic CO_2 emissions during January to March. So the values of the ensemble spread of $\boldsymbol{C}_{i,t}^{\mathrm{f}}$ in Beijing area at model-level 1 could be much larger than those in other areas, which had weak local anthropogenic CO_2 emissions (see Fig. 11a). As a result, the values of the ensemble spread of $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}}$ before inflating in Beijing area are much larger than those in other areas with small local anthropogenic CO_2 emissions (see Fig. 11b). After inflating, the ensemble spread of $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}}$ in Beijing area could be too large, compare to those in other areas with small local anthropogenic CO_2 emissions (see Fig. 11c), which lead to the failure to reproduce the true fluxes in Beijing area. Later, CFI-CMAQ will be improved by optimizing the covariance inflation method.

Since the impact of assimilation X_{CO2}^{p} by CFI-CMAQ on CO_2 fluxes was in general greater in the east and south of China than other model areas (see Figs.7 and 9), the time series of the daily mean CO_2 fluxes in that area averaged from $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ was compared with those from $\boldsymbol{F}_t^*$ and $\boldsymbol{F}_t^{\mathrm{p}}$ (see Fig. 12). This figure indicates that CFI-CMAQ could in general reproduce the true fluxes with acceptable bias.

As stated in the above section, β was a newly introduced parameter. The prior scaling factors should have been inflated indirectly through the inflated CO2 concentration forecast. However, the values of the ensemble spread of $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}}$ before inflating were very small (ranging from 0 to 0.08 in most area at model-level 1, see Fig. 11b), though the values of the ensemble spread of $\boldsymbol{C}_{i,t}^{\mathrm{f}}$ after inflating could reach 1 to 14 ppmv in most area at model-level 1 (see Fig. 11a). So we had to inflate them again before using them into Eq. (2) . Fig. 11c showed the distribution of the ensemble spread of $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{a}}$ at model-level 1 at 00 UT on 1 March 2010 when $\beta = 70$. It showed that the values of the ensemble spread of $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{a}}$ ranged from 0.1 to 0.8 in most area. In order to investigate the sensitivity of the inflation factor of the scaling factors β, a series of numerical experiments were conducted. As shown in Fig. 12, CFI-CMAQ worked rather well for $\beta = 60$, 70, 75, 80. However, if β was much smaller than 50 (e.g. $\beta = 10$), the impact of assimilation was small due to the small ensemble spread; or if β 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 β.

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. 13). Although Peters et al. (2007) indicated that the lag-window should be more than five weeks, it seemed that the smoother window had a slight influence on the assimilated results for CFI-CMAQ. It was 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).

At the end of this subsection, the assimilation results of the reference experiment in which $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}}$ were set to 1 will be addressed briefly. The impact of assimilation X_{CO2}^{p} on CO_2 fluxes was disordered. The monthly mean values of the difference between the prior true surface CO_2 fluxes and the ensemble mean values of the assimilated surface CO_2 fluxes were irregular noise (see Fig. 14). The main reason is that all the elements of the scaling factors to be optimized in the smoother window are only random numbers. As stated in the above section, only $\boldsymbol{\lambda}_{i,1|0}^{\mathrm{a}}$ needed to be optimized in the first assimilation cycle. However, $\boldsymbol{\lambda}_{i,1|0}^{\mathrm{a}}$ were rand fields (in other words, all the elements of $\boldsymbol{\lambda}_{i,1|0}^{\mathrm{a}}$ are only random numbers) because they could not generated by other ways at the first time. So their spatial correlations were too small. The correlations between the scaling factors and the observations were also too small. Therefore it was impossible to systematically change the values of $\boldsymbol{\lambda}_{i,1|0}^{\mathrm{a}}$ in large areas where the observations located after assimilating observations at $t=1$. Thus the signal-to-noise problem arose. So the elements of $\boldsymbol{\lambda}_{i,1|1}^{\mathrm{a}}$ are only random numbers too. Though $\boldsymbol{\lambda}_{i,2|1}^{\mathrm{a}}$ could be generated automatically by the smoothing operator when all $\boldsymbol{\lambda}_{i,2|1}^{\mathrm{p}}$ were set to 1, the elements of $\boldsymbol{\lambda}_{i,2|1}^{\mathrm{a}}$ are random numbers too since the smoothing operator is only a linear operator. Similarly, it was impossible to systematically change the values of $\lambda_{i,1|1}^{a}$ and $\lambda_{i,2|1}^{a}$ in large areas after assimilating observations at $t=2$. As this inversion system continued assimilating observations, all future scaling factors could be created by the smoothing operator and then updated. But this inversion system could not ingest the observations effectively because all the elements of the scaling factors were always random numbers. Though the 9 days lag-window in the reference experiment is too short compared to the 5 weeks lag-window recommended by Peters et al(2007), this reference experiment could illustrate the limitation by only using the smoothing operator as the persistence dynamical model. If the lag-window was around 5 weeks, we could get better results because there were more observations in every assimilation cycle. However, the results could not be better than those obtained by CFI-CMAQ because most grids have no observations (refer to Fig. 3a) and the signal-to-noise problem still remained.

4 Summary and conclusions

A regional surface CO_2 flux inversion system, CFI-CMAQ, has been developed to optimize CO_2 fluxes at grid scales. It operates under a joint data assimilation framework by applying EnKF to constrain the CO_2 concentrations and applying EnKS to optimize the surface CO_2 flux, which is similar to Kang et al. (2011, 2012) and Tian et al. (2013). The persistence dynamical model, which was first introduced by Peters et al. (2007) by applying the smoothing operator to transport the useful observed information onto the next assimilation cycle, is further developed. We associated the smoothing operator with the atmospheric transport model to constitute the persistence dynamical model to forecast the surface CO2 flux scaling factors for the purpose of resolving the ‘signal-to-noise’ problem, as well as transporting the useful observed information onto the next assimilation cycle. 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 the grid scale with acceptable bias.

This study represents the first step in developing a regional surface CO_2 flux inversion system to optimize CO_2 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 CO_2 flux optimization.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 41130528), 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

Andres, R. J., Boden, T. A., Bréon, F. M., Ciais, P., Davis, S., Erickson, D., Gregg, J. S., Jacobson, A., Marland, G., Miller, J., Oda, T., Olivier, J. G. J., Raupach, M. R., Rayner, P. and

Treanton, K.: A synthesis of carbon dioxide emissions from fossil-fuel combustion, Biogeosciences, 9, 1845-1871. doi:10.5194/bg-9-1845-2012, 2012.

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

Boden, T. A., Marland, G., and Andres, R. J.: Global, regional, and national fossil-fuel CO_2 emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A, doi:10.3334/CDIAC/00001_V2011, 2011.

Chevallier, F. M. F., Peylin, P., Bousquet, S. S. P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO_2 sources and sinks from satellite observations: Method and application to TOVS data, J. Geophys. Res., 110, D24309, doi:10.1029/2005JD006390, 2005.

Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, J. Geophys. Res., 112, D09307, doi:10.1029/2006JD007375, 2007a.

Chevallier, F.: Impact of correlated observation errors on inverted CO_2 surface fluxes from OCO measurements, Geophys. Res. Lett., 34, L24804, doi:10.1029/2007GL030463, 2007b.

Connor, B. J., Bösch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error analysis, J. Geophys. Res., 113, D05305, doi:10.1029/2006JD008336, 2008.

Crisp, D., Bösch, H., Brown, L., Castano, R., Christi, M., Connor, B., Frankenberg, C., McDuffie, J., Miller, C. E., Natraj, V., O'Dell, C., O'Brien, D., Polonsky, I., Oyafuso, F., Thompson, D., Toon, G., and Spurr, R.: OCO (Orbiting Carbon Observatory)-2 Level 2 Full Physics Retrieval Algorithm Theoretical Basis, Tech. Rep. OCO D-65488, NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, version 1.0 Rev 4, http://disc.sci.gsfc.nasa.gov/acdisc/documentation/OCO-2_L2_FP_ATBD_v1_rev4_Nov10.pdf, (last access: August 4, 2014), 2010.

Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO_2 retrieval algorithm – Part II: Global X_{CO2} data characterization, Atmos. Meas. Tech., 5, 687-707, doi:10.5194/amt-5-687-2012, 2012.

Deng, F., Chen, J. M., Ishizawa, M., YUEN, C. W. A. I., Mo, G., Higuchi, K., Chan, D., and Maksyutov, S.: Global monthly CO2 flux inversion with a focus over North America, Tellus B, 59, 179–190, 2007.

Engelen, R. J., Serrar, S., and Chevallier, F.: Four-dimensional data assimilation of atmospheric CO2 using AIRS observations, J. Geophys. Res., 114, D03303, doi:10.1029/2008JD010739, 2009.

Feng, L., Palmer, P. I., Bösch, H., and Dance, S.: Estimating surface CO_2 fluxes from space-borne CO_2 dry air mole fraction observations using an ensemble Kalman Filter, Atmos. Chem. Phys., 9, 2619–2633, 2009.

Feng, L., Palmer, P. I., Yang, Y., Yantosca, R. M., Kawa, S. R., Paris, J.-D., Matsueda, H., and Machida, T.: Evaluating a 3-D transport model of atmospheric CO_2 using ground-based, aircraft, and space-borne data, Atmos. Chem. Phys., 11, 2789–2803, doi: 10.5194/acp-11-2789-2011, 2011.

Gurney, K. R., Law, R. L., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H, Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J, Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., Yuen, C. W.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, 2002

Gurney, K. R., Mendoza, D. L., Zhou, Y. Y., Fischer, M. L., Miller, C. C., Geethakumar, S. and Du Can, S. D.: High resolution fossil fuel combustion CO_2 emission fluxes for the United States, Environ. Sci. & Technol., 43, 5535-5541. doi:10.1021/es900806c, 2009.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Wea. Rev. 129: 123-137, 2001.

Huang, Z. K., Peng, Z., Liu, H. N., Zhang, M. G.: Development of CMAQ for East Asia CO2 data assimilation under an EnKF framework: a first result, Chinese Science Bulletin, 59: 3200-3208, doi: 10.1007/s11434-014-0348-9, 2014.

Jiang, F., Wang, H. W., Chen, J. M., Zhou, L. X., Ju, W. M., Ding, A. J., Liu, L. X., and

Peters, W.: Nested atmospheric inversion for the terrestrial carbon sources and sinks in China, Biogeosciences, 10, 5311-5324, doi:10.5194/bg-10-5311-2013, 2013.

Kang, J.-S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T., and Ide, K.: "Variable localization" in an ensemble Kalman filter: application to the carbon cycle data assimilation, J. Geophys. Res., 116, D09110, doi:10.1029/2010JD014673, 2011.

Kang, J.-S., Kalnay, E., Miyoshi, T., Liu, J., and Fung, I.: Estimation of surface carbon fluxes with an advanced data assimilation methodology, J. Geophys. Res., 117, D24101, doi:10.1029/2012JD018259, 2012.

Kou, X., Zhang, M., Peng, Z.: Numerical Simulation of CO_2 Concentrations in East Asia with RAMS-CMAQ, Atmos, Oceanic Sci Lett, 6, 179-184, 2013.

Kretschmer, R., Gerbig, C., Karstens, U., Koch, F.-T.: Error characterization of CO_2 vertical mixing in the atmospheric transport model WRF-VPRM, Atmos. Chem. Phys., 12, 2441–2458, 2012.

Liu, J., Fung, I., Kalnay, E., Kang, J.: CO_2 transport uncertainties from the uncertainties in meteorological fields, Geophys. Res. Lett., 38, L12808, doi: 10.1029/2011GL047213, 2011.

Liu, J., Fung, I., Kalnay, E., Kang, J.-S., Olsen, E. T., and Chen, L.: Simultaneous assimilation of AIRS Xco_2 and meteorological observations in a carbon climate model with an ensemble Kalman filter, J. Geophys. Res., 117, D05309, doi: 10.1029/2011JD016642, 2012.

Liu Z., Bambha, R. P., Pinto, J. P.: Toward verifying fossil fuel CO_2 emissions with the Community Multi-scale Air Quality (CMAQ) model: motivation, model description and initial simulation. J. Air & Waste Management Assoc., 64, 419-435, doi: 10.1080/10962247.2013.816642, 2013.

Marland, G.: Uncertainties in accounting for CO_2 from fossil fuels, J. of Indust. Ecol. 12, 136-139, doi:10.1111/j.1530-9290.2008.00014.x, 2008.

Miyazaki K.: Performance of a local ensemble transform Kalman filter for the analysis of atmospheric circulation and distribution of long-lived tracers under idealized conditions, J. Geophys. Res., 114, D19304, doi: 10.1029/2009JD011892, 2009.

O'Dell, C. W., Connor, B., B¨osch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C.,Wennberg, P. O., and Wunch, D.: The ACOS CO2 retrieval algorithm – Part 1: Description and validation against synthetic observations, Atmos. Meas. Tech., 5, 99–121, doi:10.5194/amt-5-99-2012, 2012.

Osterman, G., Martinez, E., Eldering, A., Avis, C.: ACOS Level 2 Standard Product Data User's Guide, v2.9, 2011.

Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO_2 surface fluxes from atmospheric trace gas observations, J. Geophys. Res., 110, D24304, doi: 10.1029/2005JD006157, 2005.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R.,Randerson, J. T., Wennberg, P. O., Krol, M. C., Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, P. Natl. Acad. Sci. USA, 104, 18925–18930, 2007.

Peters, W., KROL, M. C., Van Der WERF, G. R., Houwling, S., Jones, C. D., Hughes, J., Schaefer, K., Masarie, K. A., Jacobson, A. R., Miller, J. B., Cho, C. H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., Di Sarra, A. G., Piachentina, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H. A. J., Van Der Laan, S., Neubert, R. E. M., Jordan, A., Rodo, X., Morgui, J. –A., Vermeulen, A. T., Popa, E., Rozanski, M., Manning, A. C., Leuenberger, M., Uglietti, C., Dolman, A. J., Ciais, P., Heimann, M., Tans, P. P.: Seven years of recent Europenan net terrestrial carbon dioxide exchange constrained by atmisppheric observations, Global Change Biology, 16(4), 1365-2486, 2009.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO2 inversions, Biogeosciences, 10, 6699–6720, doi:10.5194/bg-10-6699-2013, 2013.

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 CO_2 over complex terrain – representing the Ochsenkopf mountain tall tower, Atmos. Chem. Phys., 11, 7445-7464, doi:10.5194/acp-11-7445-2011, 2011.

Prather, M., Zhu, X., Strahan, S.E., Steenrod, S., D., and Rodriguez, J., M.: Quantifying errors in trace species transport modeling. Proc. Natl. Acad. Sci. U. S. A. 105:19617-19621. doi:10.1073/pnas.0806541106, 2008.

National Research Council: Verifying greenhouse gas emissions: Methods to support international climate agreements, The National Academies Press, Washington, D.C. 2010.

Tian, X., Xie, Z., Sun, Q.: A POD-based ensemble four-dimensional variational assimilation method, Tellus A, 63, 805-816, 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. Discuss., 13, 24755-24784, 2013.

van der Werf, G. R., Randerson, J. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). Atmos. Chem. Phys., **10**, 11707–11735, doi: 10.5194/acp-10-11707-2010, 2010.

Wang, B., Liu, J., Wang, S., Cheng, W., Liu, J., Liu, C., Xiao Q., and Kuo, Y.: An economical approach to four-dimensional variational data assimilation, Adv. Atmos. Sci., 27, 715–727, doi:10.1007/s00376-009-9122-3, 2010.

Zhang, H. F., B. Z. Chen, I. T. van der Laan-Luijkx, J. Chen, G. Xu, J. W. Yan, L. X. Zhou, Y. Fukuyama, P. P. Tans, and W. Peters, Net terrestrial CO_2 exchange over China during 2001–2010 estimated with an ensemble data assimilation system for atmospheric CO_2, J. Geophys. Res. Atmos., 119, 3500–3515, doi:10.1002/2013JD021297, 2014.

Zhang, H. F., Chen, B. Z., 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, 5807-5824, doi:10.5194/acp-14-5807-2014, 2014.

Zhang, M., Uno, I., Sugata, S., Wang, Z., Byun, D., Akimoto, H.: Numerical study of boundary layer ozone transport and photochemical production in East Asia in the wintertime, Geophys. Res. Lett., 29(11), 40-1-40-4, doi: 10.1029/2001GL014368, 2002.

Zhang, M., Uno, I., Carmichael, G. R., Akimoto, H., Wang, Z., Tang, Y., Woo, J., Streets, D. G., Sachse, G. W., Avery, M. A., Weber, R. J., Talbot, R. W.: Large-scale structure of trace gas and aerosol distributions over the western Pacific Ocean during the Transport and Chemical Evolution Over the Pacific (TRACE-P) experiment, J. Geophys. Res., 108(D21), 8820, doi: 10.1029/2002JD002946, 2003.

Zhang, M., Gao, L., Ge, C., Xu, Y.: Simulation of nitrate aerosol concentrations over East Asia with the model system RAMS-CMAQ, Tellus B, 59, 372-380, 2007.

List of Figures

Fig. 1. Schematic diagram of the smoother window. $(\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1-M|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M+1|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1|t-1})$ are the optimized scaling factors in the smoother window and $\boldsymbol{C}^{\mathrm{a}}_{i,t-1}$ are the assimilated CO_2 concentrations fields at time $t-1$ in the previous assimilation cycle *t*-1-*M*~*t*-1. $(\boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-M+1|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,j|t-1}, \cdots, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t-1|t-1}, \boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1})$ are the scaling factors in the smoother window and $\boldsymbol{C}^{\mathrm{f}}_{i,t}$ are the forecast CO_2 concentrations fields at time t which need to be optimized in the current assimilation cycle *t*-*M*~*t*.

Fig. 2. Flowchart of the CFI-CMAQ system used to optimize surface CO_2 fluxes at each assimilation cycle. The system includes the following four parts in turn: (1) forecasting of the linear scaling factors $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$ (red arrows); (2) optimization of the scaling factors in the smoother window by EnKS (see Fig. 1) (blue arrows); (3) updating of the flux in the smoother window (green arrows); and (4) assimilation of the CO_2 concentration fields at time *t* by EnKF (black arrows).

Fig. 3. (a) Total number of observations in February 2010 in the model grid. Each symbol indicates the total number of all GOSAT X_{CO2} measurements in the corresponding model grid. Monthly mean values in February 2010 of (b) X^{p}_{CO2}, column mixing ratio of $\boldsymbol{C}^{\mathrm{p}}_{t}$; (c) X^{f}_{CO2}, column mixing ratio of $\boldsymbol{C}^{\mathrm{f}}_{t}$; (d) $\overline{X^{\mathrm{a}}_{CO2}}$, column mixing ratio of $\overline{\boldsymbol{C}^{\mathrm{a}}_{t}}$; (e) $X^{\mathrm{p}}_{CO2} - X^{\mathrm{f}}_{CO2}$; and (f) $X^{\mathrm{p}}_{CO2} - \overline{X^{\mathrm{a}}_{CO2}}$. All column mixing ratios are column-averaged with real GOSAT X_{CO2} averaging kernels at GOSAT X_{CO2} locations. Each symbol indicates the monthly average value of all X_{CO2}

estimates in the model grid. $\overline{\boldsymbol{C}_t^{\mathrm{a}}}$ are the ensemble mean values of the assimilated CO_2 concentrations fields of a CFI-CMAQ OSSE, in which the lag-window was 9 days and β was 70. And they are the same OSSE in Fig. 3 to Fig. 6.

Fig. 4. Monthly mean values of (a) $\boldsymbol{C}_t^{\mathrm{p}}$, the artificial true simulations driven by the prior surface CO_2 fluxes $\boldsymbol{F}_t^{\mathrm{p}}$; (b) $\boldsymbol{C}_t^{\mathrm{f}}$, the background simulations driven by magnified surface CO_2 fluxes $\boldsymbol{F}_t^{*} = (1.8+\delta(x,y,z,t))\boldsymbol{F}_t^{\mathrm{p}}$; (c) $\overline{\boldsymbol{C}_t^{\mathrm{a}}}$, the ensemble mean values of the assimilated CO_2 concentrations fields; (d) $\boldsymbol{C}_t^{\mathrm{p}} - \boldsymbol{C}_t^{\mathrm{f}}$; (e) $\boldsymbol{C}_t^{\mathrm{p}} - \overline{\boldsymbol{C}_t^{\mathrm{a}}}$; and (f) $100*(\boldsymbol{C}_t^{\mathrm{p}} - \overline{\boldsymbol{C}_t^{\mathrm{a}}})/\boldsymbol{C}_t^{\mathrm{p}}$ at model-level 1 in February 2010. Black lines EF and GH indicate the positions of the cross sections shown in Fig. 5.

Fig. 5. Monthly mean cross sections of $\boldsymbol{C}_t^{\mathrm{p}} - \boldsymbol{C}_t^{\mathrm{f}}$ along line (a) EF and (b) GH, and monthly mean cross sections of $\boldsymbol{C}_t^{\mathrm{p}} - \overline{\boldsymbol{C}_t^{\mathrm{a}}}$ along line (c) EF and (d) GH (cross section lines shown in Fig. 4d) in February 2010.

Fig. 6. Daily mean time series of CO_2 concentrations at national background stations in China and their nearest large cities from 1 Jan. to 20 Mar. 2010 extracted from the artificial true simulations $\boldsymbol{C}_t^{\mathrm{p}}$ (black), background simulations $\boldsymbol{C}_t^{\mathrm{f}}$ (red), and the ensemble mean values of the assimilated CO_2 concentrations fields $\overline{\boldsymbol{C}_t^{\mathrm{a}}}$ (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).

Fig. 7. Monthly mean values in February 2010 of (a) $\boldsymbol{F}_t^{\mathrm{p}}$, the prior true surface CO_2 fluxes; (b) $\boldsymbol{F}_t^{*}$, the prescribed CO_2 surface fluxes, $\boldsymbol{F}_t^{*} = (1.8 + \delta(x, y, z, t))\boldsymbol{F}_t^{\mathrm{p}}$; (c) $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$, the ensemble mean values of the assimilated surface CO_2 fluxes; (d) $\boldsymbol{F}_t^{\mathrm{p}} - \boldsymbol{F}_t^{*}$; and (e) $\boldsymbol{F}_t^{\mathrm{p}} - \overline{\boldsymbol{F}_t^{\mathrm{a}}}$ (units: μmole m^{-2} s^{-1}). $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ are the assimilated results of an CFI-CMAQ OSSE, in which the lag-window was 9 days and β was 70. And they are the same in Fig. 7 to Fig. 10.

Fig. 8. Monthly mean RMSEs of $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ in February 2010 (units: μmole m^{-2} s^{-1}).

Fig. 9. (a) Ratios of monthly mean $\boldsymbol{F}_t^{*}$ to monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$; and (b) ratios of monthly mean $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ to monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$ in Feb. 2010. The white part indicates the ratios where the absolute values of monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$ are larger than 0.1, 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 and 13.

Fig. 10. Daily mean time series of CO_2 fluxes at national background stations in China and their nearest large cities from 1 Jan to 20 Mar. 2010 extracted from the prior true surface CO_2 fluxes $\boldsymbol{F}_t^{\mathrm{p}}$ (black), the prescribed CO_2 surface fluxes $\boldsymbol{F}_t^{*}$ (red), and the assimilated CO_2 fluxes $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ (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.

Fig. 11. (a) Ensemble spread of $\boldsymbol{C}_{i,t}^{\mathrm{f}}$ after inflating; (b) ensemble spread of $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}}$ before inflating; (c) ensemble spread of $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{a}}$ at model-level 1 at 00 UT on 1 March 2010 when $\beta = 70$.

Fig. 12. Time series of daily mean CO_2 fluxes averaged in domain I (shown in Fig. 9a) from 1 Jan. to 20 Mar. 2010 with the inflation factor of scaling factors $\beta = 50$, 60, 70, 75 and 80. The black dashed line is the time series averaged from $\boldsymbol{F}_t^*$ and the black solid line is the time series averaged from $\boldsymbol{F}_t^{\mathrm{p}}$.

Fig. 13. Time series of daily mean CO_2 fluxes averaged in domain I (shown in Fig. 9a) from 1 Jan. to 20 Mar 2010 with different smoother windows (3, 6, 9 and 12 days). The black dashed line is the time series averaged from $\boldsymbol{F}_t^*$ and the black solid line is the time series averaged from $\boldsymbol{F}_t^{\mathrm{p}}$.

Fig. 14. Monthly mean values of the difference between the prior true surface CO_2 fluxes and the ensemble mean values of the assimilated surface CO_2 fluxes (units: μmole m^{-2} s^{-1}) of the reference experiment in which $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}}$ were set to 1.

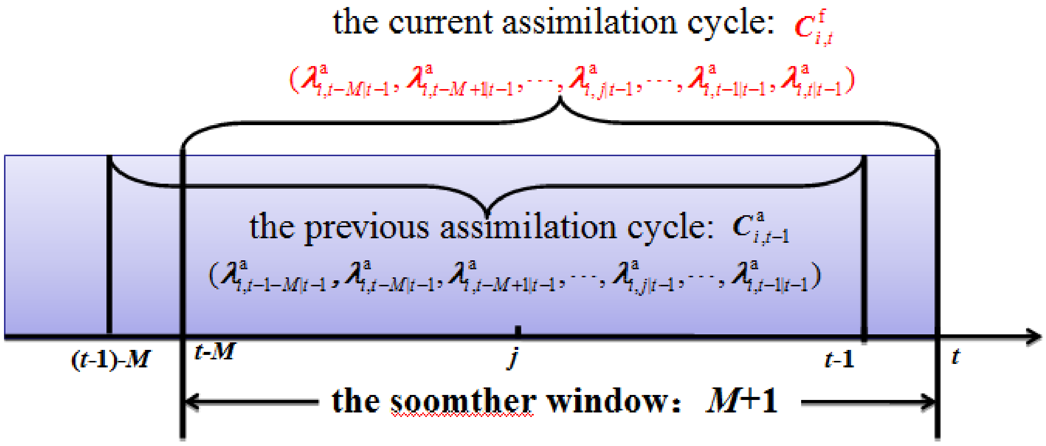

Fig. 1. Schematic diagram of the smoother window.

$(\lambda^{\mathrm{a}}_{i,t-1-M|t-1},\lambda^{\mathrm{a}}_{i,t-M|t-1},\lambda^{\mathrm{a}}_{i,t-M+1|t-1},\cdots,\lambda^{\mathrm{a}}_{i,j|t-1},\cdots,\lambda^{\mathrm{a}}_{i,t-1|t-1})$ are the optimized scaling factors in the smoother window and $\boldsymbol{C}^{\mathrm{a}}_{i,t-1}$ are the assimilated CO_2 concentrations fields at time $t-1$ in the previous assimilation cycle *t*-1-*M*~*t*-1. $(\lambda^{\mathrm{a}}_{i,t-M|t-1},\lambda^{\mathrm{a}}_{i,t-M+1|t-1},\cdots,\lambda^{\mathrm{a}}_{i,j|t-1},\cdots,\lambda^{\mathrm{a}}_{i,t-1|t-1},\lambda^{\mathrm{a}}_{i,t|t-1})$ are the scaling factors in the smoother window and $\boldsymbol{C}^{\mathrm{f}}_{i,t}$ are the forecast CO_2 concentrations fields at time t which need to be optimized in the current assimilation cycle *t*-*M*~*t*.

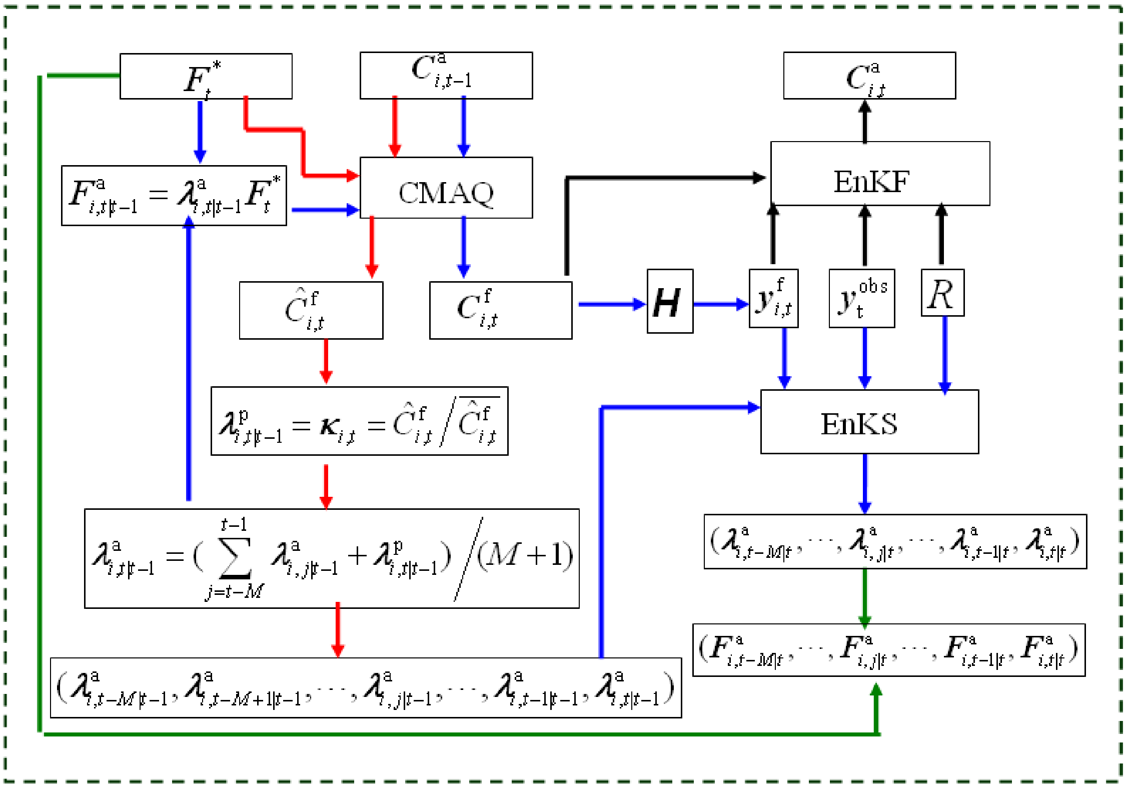

Fig. 2. Flowchart of the CFI-CMAQ system used to optimize surface CO_2 fluxes at each assimilation cycle. The system includes the following four parts in turn: (1) forecasting of the linear scaling factors $\lambda^{a}_{i,t|t-1}$ (red arrows); (2) optimization of the scaling factors in the smoother window by EnKS (see Fig. 1) (blue arrows); (3) updating of the flux in the smoother window (green arrows); and (4) assimilation of the CO_2 concentration fields at time t by EnKF (black arrows).

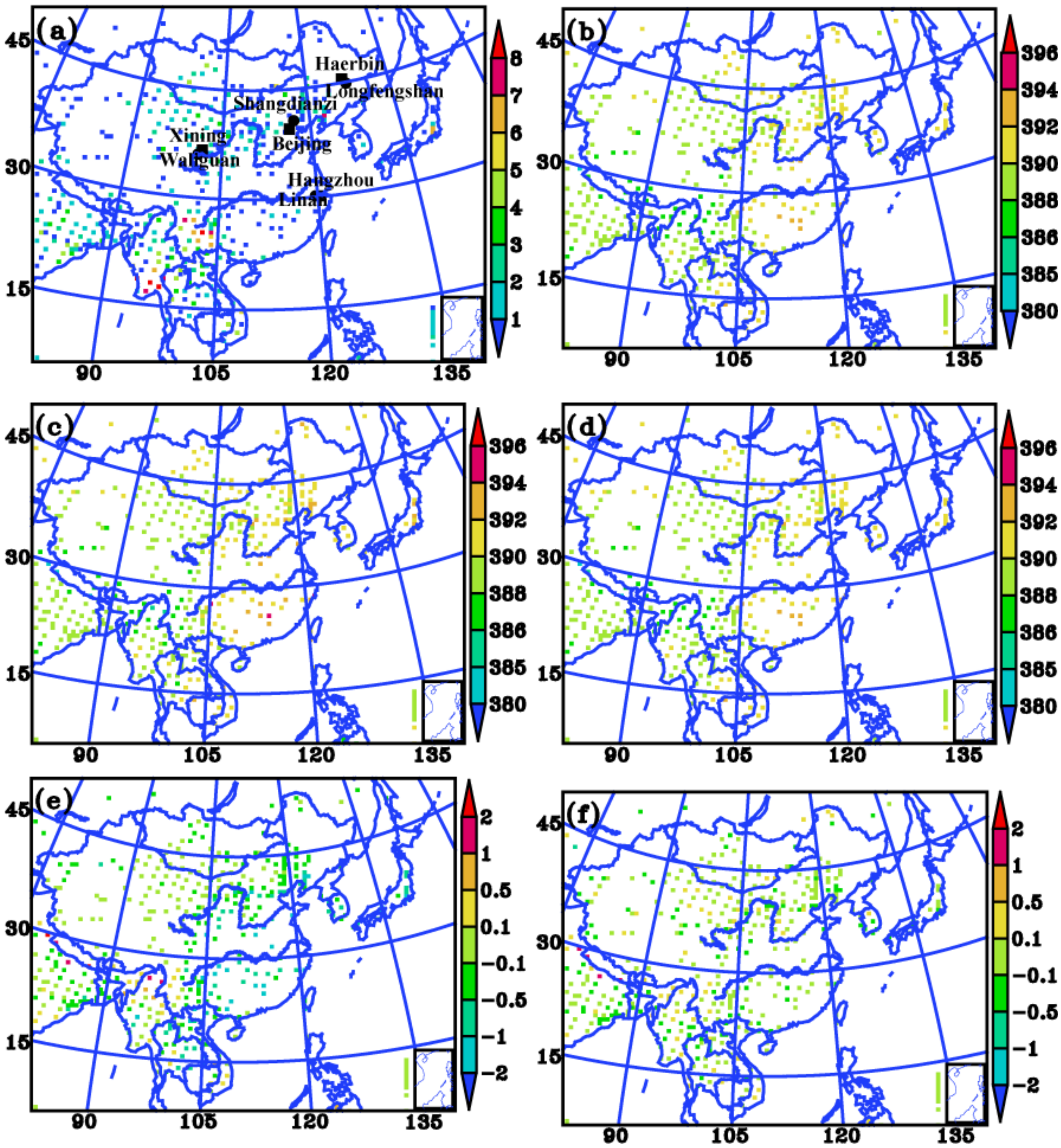

Fig. 3. (a) Total number of observations in February 2010 in the model grid. Each symbol indicates the total number of all GOSAT X_{CO2} measurements in the corresponding model grid. Monthly mean values in February 2010 of (b) X^{p}_{CO2}, column mixing ratio of $\boldsymbol{C}^{\mathrm{p}}_{t}$; (c) X^{f}_{CO2}, column mixing ratio of $\boldsymbol{C}^{\mathrm{f}}_{t}$; (d) $\overline{X^{\mathrm{a}}_{CO2}}$, column mixing ratio of $\overline{\boldsymbol{C}^{\mathrm{a}}_{t}}$; (e) $X^{\mathrm{p}}_{CO2} - X^{\mathrm{f}}_{CO2}$; and (f) $X^{\mathrm{p}}_{CO2} - \overline{X^{\mathrm{a}}_{CO2}}$. All column mixing ratios are column-averaged with real GOSAT X_{CO2} averaging kernels at GOSAT X_{CO2} locations. Each symbol indicates the monthly average value of all X_{CO2} estimates in the model grid. $\overline{\boldsymbol{C}^{\mathrm{a}}_{t}}$ are the ensemble mean values of the assimilated CO_2 concentrations fields of a CFI-CMAQ OSSE, in which the lag-window was 9 days and β was 70. And they are the same OSSE in Fig. 3 to Fig. 6.

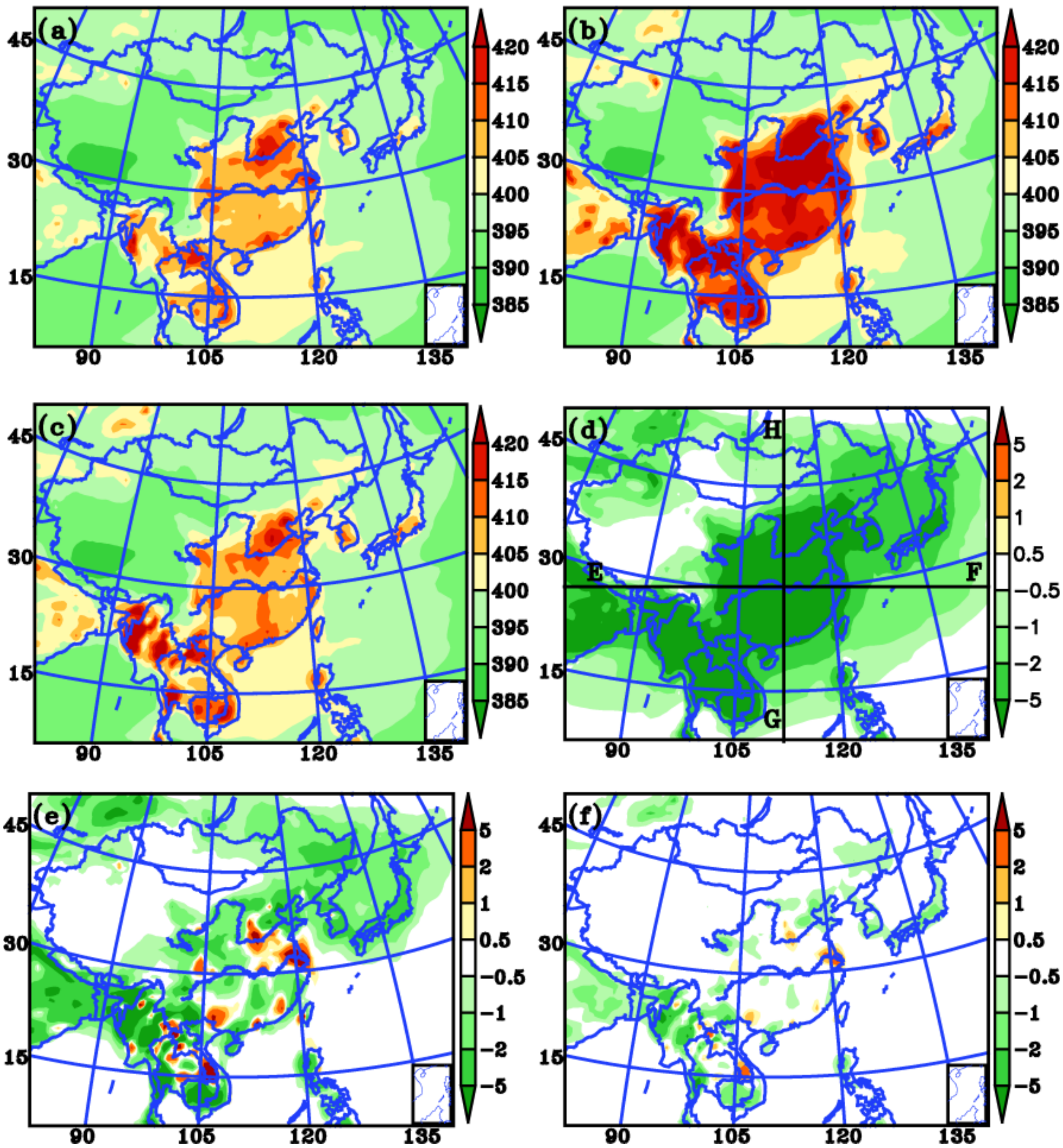

Fig. 4. Monthly mean values of (a) $\boldsymbol{C}_t^{\mathrm{p}}$, the artificial true simulations driven by the prior surface CO_2 fluxes $\boldsymbol{F}_t^{\mathrm{p}}$; (b) $\boldsymbol{C}_t^{\mathrm{f}}$, the background simulations driven by magnified surface CO_2 fluxes $\boldsymbol{F}_t^{*} = (1.8 + \delta(x, y, z, t))\boldsymbol{F}_t^{\mathrm{p}}$; (c) $\overline{\boldsymbol{C}_t^{\mathrm{a}}}$, the ensemble mean values of the assimilated CO_2 concentrations fields; (d) $\boldsymbol{C}_t^{\mathrm{p}} - \boldsymbol{C}_t^{\mathrm{f}}$; (e) $\boldsymbol{C}_t^{\mathrm{p}} - \overline{\boldsymbol{C}_t^{\mathrm{a}}}$; and (f) $100 * (\boldsymbol{C}_t^{\mathrm{p}} - \overline{\boldsymbol{C}_t^{\mathrm{a}}}) / \boldsymbol{C}_t^{\mathrm{p}}$ at model-level 1 in February 2010. Black lines EF and GH indicate the positions of the cross sections shown in Fig. 5.

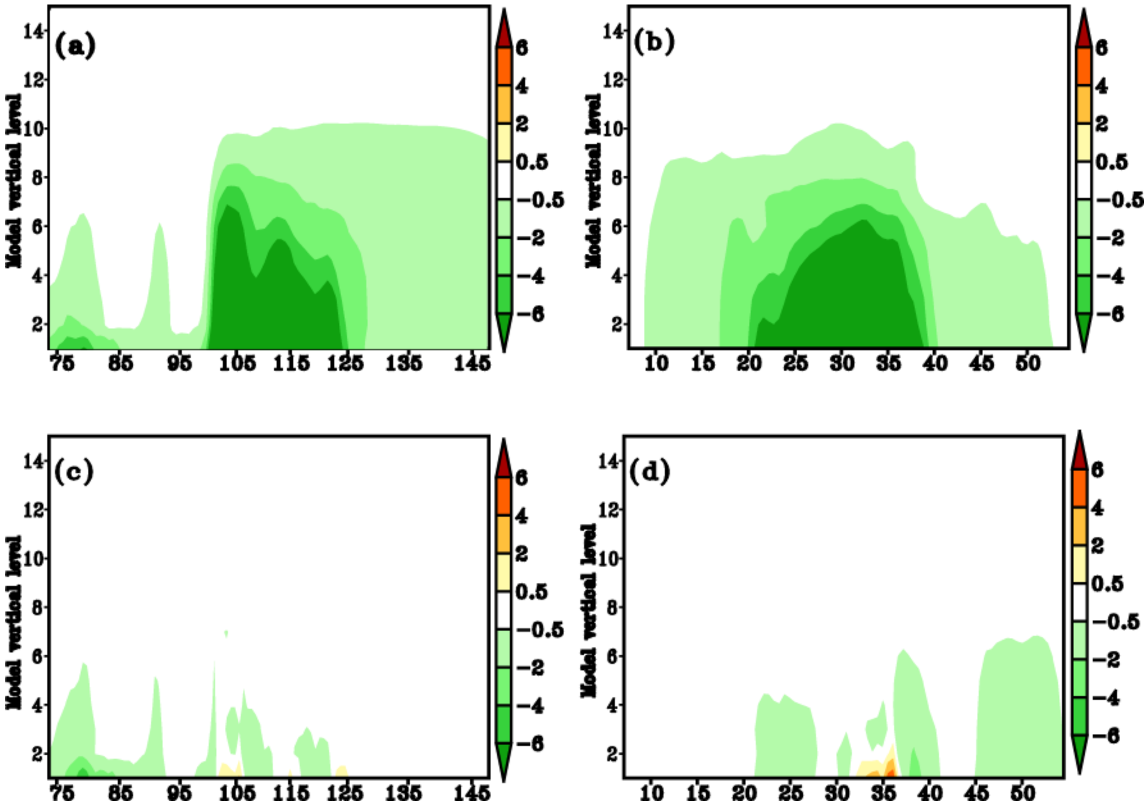

Fig. 5. Monthly mean cross sections of $\boldsymbol{C}_t^{\mathrm{p}} - \boldsymbol{C}_t^{\mathrm{f}}$ along line (a) EF and (b) GH, and monthly mean cross sections of $\boldsymbol{C}_t^{\mathrm{p}} - \overline{\boldsymbol{C}_t^{\mathrm{a}}}$ along line (c) EF and (d) GH (cross section lines shown in Fig. 4d) in February 2010.

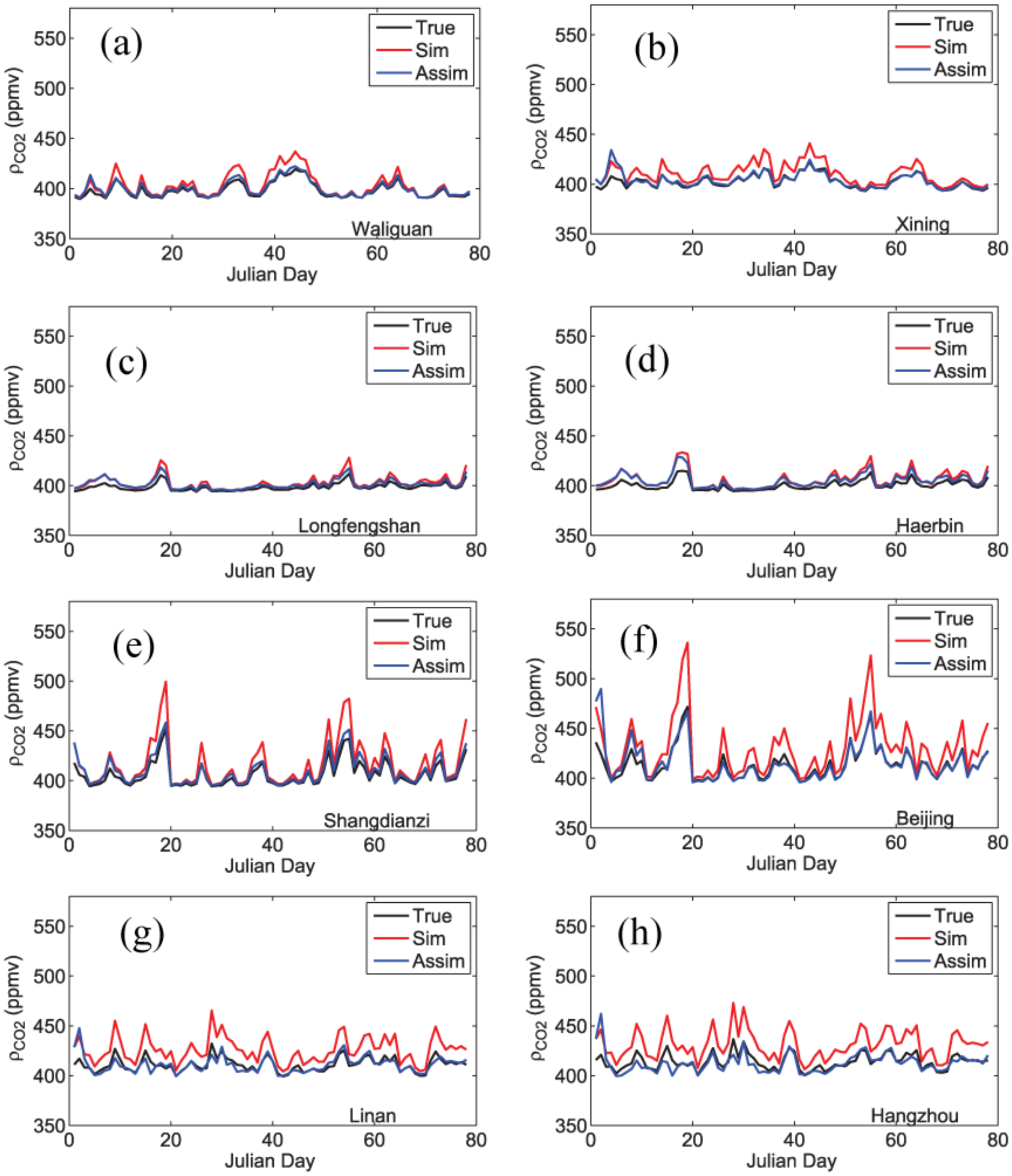

Fig. 6. Daily mean time series of CO_2 concentrations at national background stations in China and their nearest large cities from 1 Jan. to 20 Mar. 2010 extracted from the artificial true simulations $\boldsymbol{C}_t^{\mathrm{p}}$ (black), background simulations $\boldsymbol{C}_t^{\mathrm{f}}$ (red), and the ensemble mean values of the assimilated CO_2 concentrations fields $\overline{\boldsymbol{C}_t^{\mathrm{a}}}$ (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).

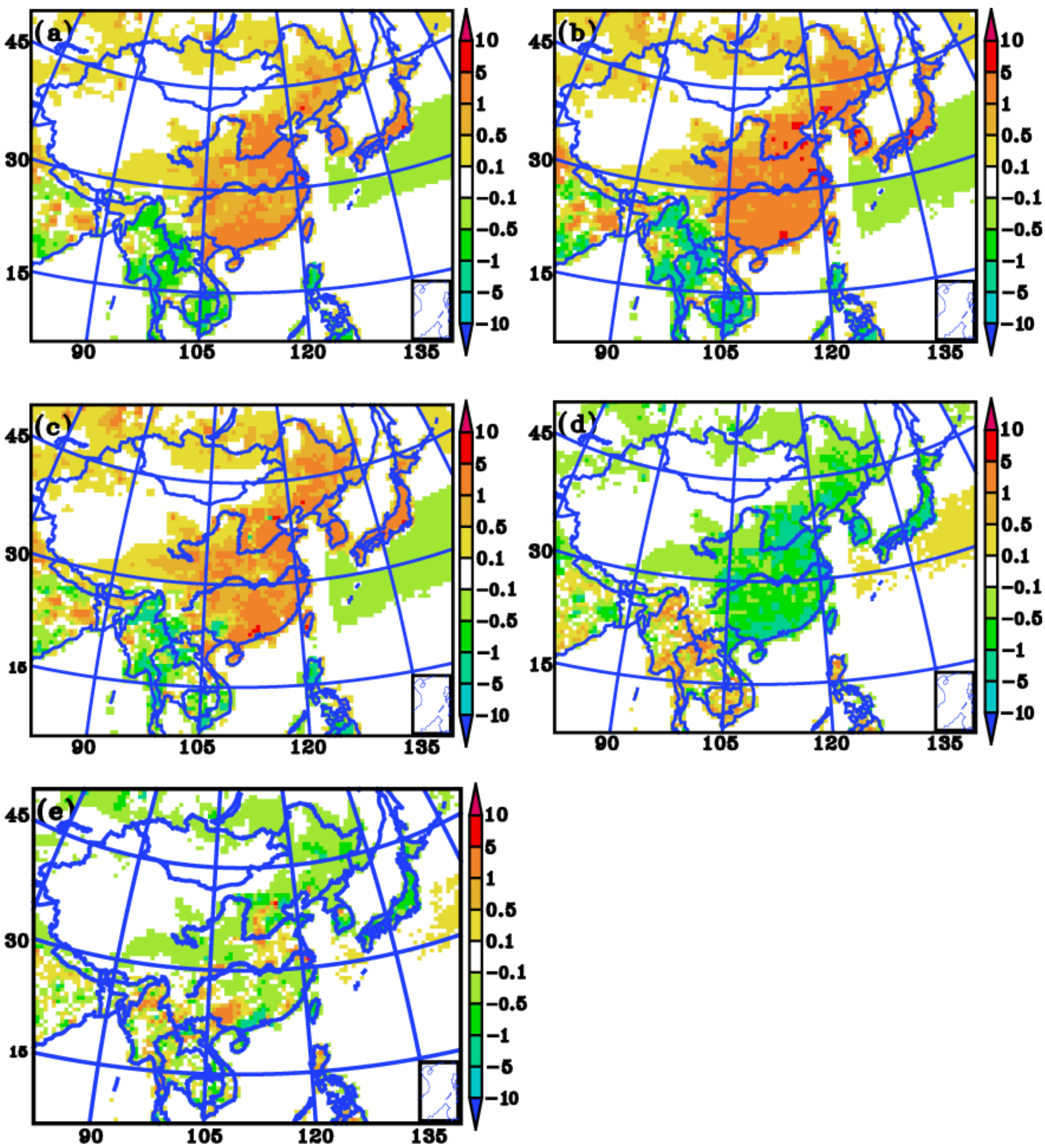

Fig. 7. Monthly mean values in February 2010 of (a) $\boldsymbol{F}_t^{\mathrm{p}}$, the prior true surface CO_2 fluxes; (b) $\boldsymbol{F}_t^{*}$, the prescribed CO_2 surface fluxes, $\boldsymbol{F}_t^{*} = (1.8+\delta(x,y,z,t))\boldsymbol{F}_t^{\mathrm{p}}$; (c) $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$, the ensemble mean values of the assimilated surface CO_2 fluxes; (d) $\boldsymbol{F}_t^{\mathrm{p}} - \boldsymbol{F}_t^{*}$; and (e) $\boldsymbol{F}_t^{\mathrm{p}} - \overline{\boldsymbol{F}_t^{\mathrm{a}}}$ (units: μmole m^{-2} s^{-1}). $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ are the assimilated results of an CFI-CMAQ OSSE, in which the lag-window was 9 days and β was 70. And they are the same in Fig. 7 to Fig. 10.

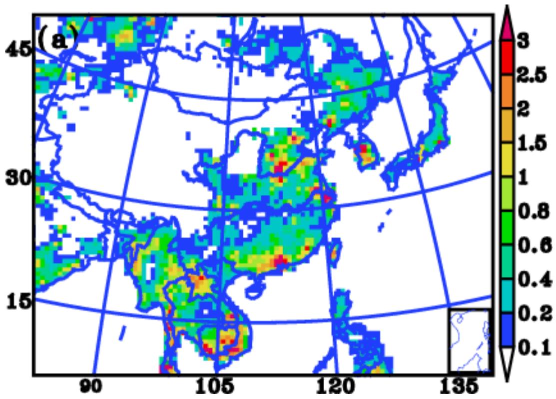

Fig. 8. Monthly mean RMSEs of $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ in February 2010 (units: μmole m^{-2} s^{-1}).

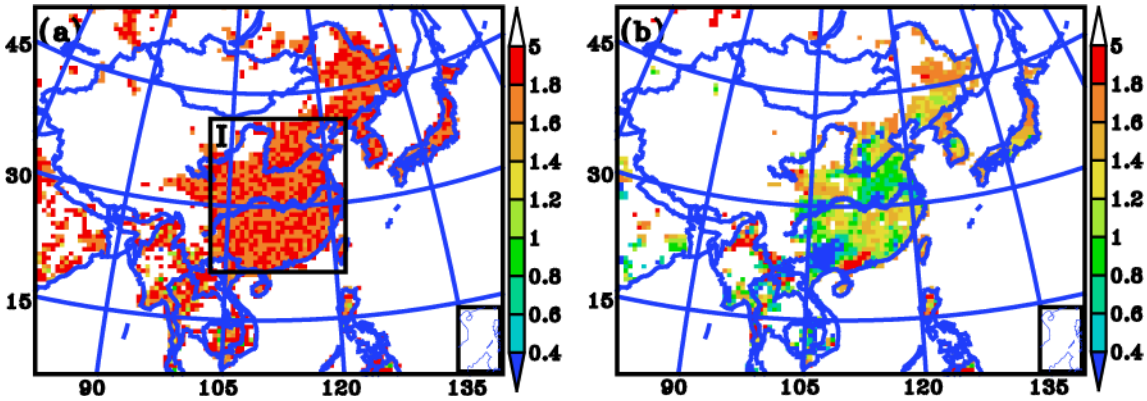

Fig. 9. (a) Ratios of monthly mean $\boldsymbol{F}_t^{*}$ to monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$; and (b) ratios of monthly mean $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ to monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$ in Feb. 2010. The white part indicates the ratios where the absolute values of monthly mean $\boldsymbol{F}_t^{\mathrm{p}}$ are larger than 0.1, 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 and 13.

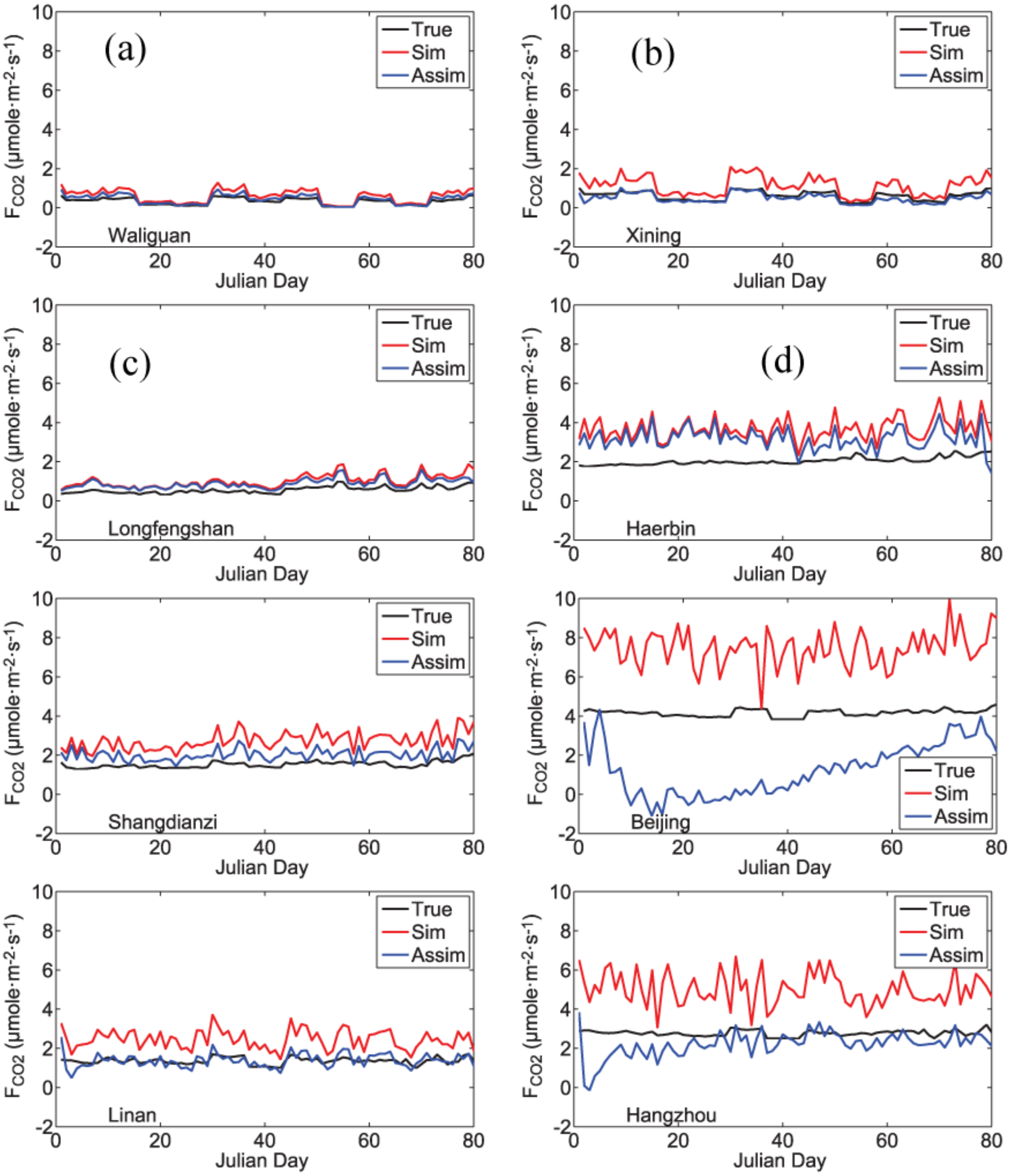

Fig. 10. Daily mean time series of CO_2 fluxes at national background stations in China and their nearest large cities from 1 Jan to 20 Mar. 2010 extracted from the prior true surface CO_2 fluxes $\boldsymbol{F}_t^{\mathrm{p}}$ (black), the prescribed CO_2 surface fluxes $\boldsymbol{F}_t^{*}$ (red), and the assimilated CO_2 fluxes $\overline{\boldsymbol{F}_t^{\mathrm{a}}}$ (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.

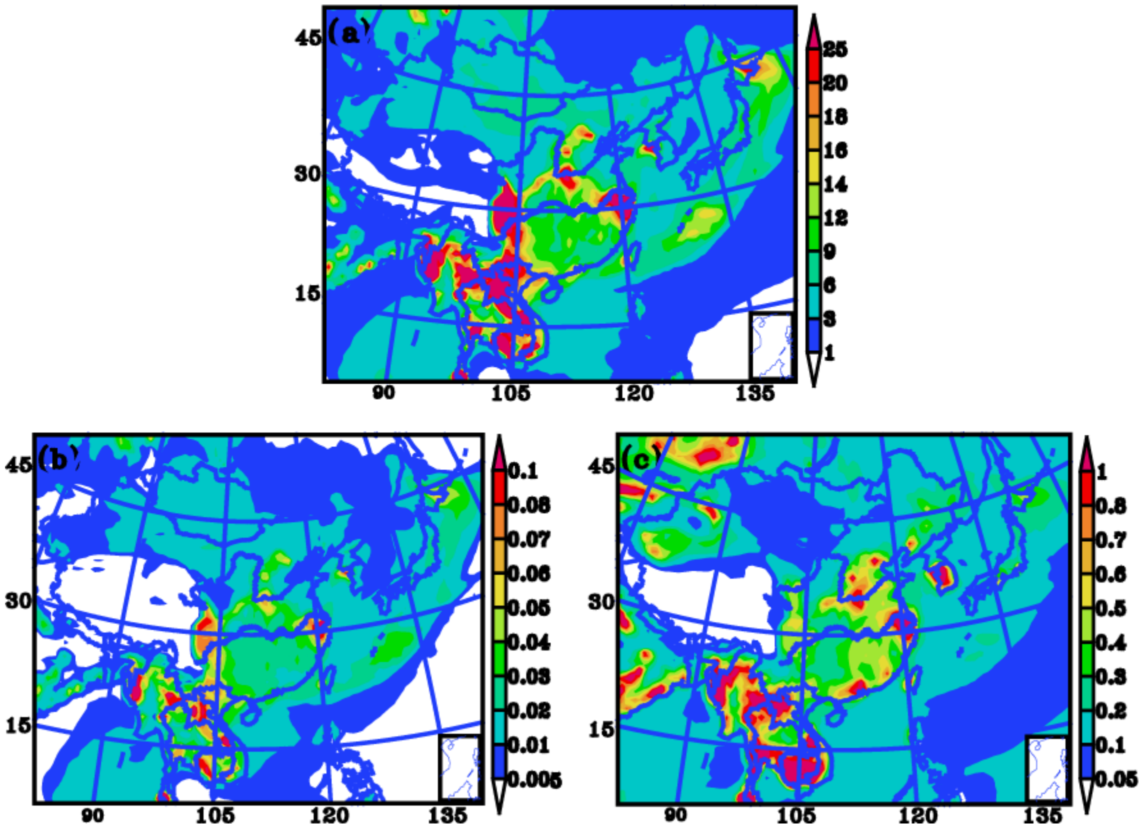

Fig. 11. (a) Ensemble spread of $\boldsymbol{C}^{\mathrm{f}}_{i,t}$ after inflating; (b) ensemble spread of $\boldsymbol{\lambda}^{\mathrm{p}}_{i,t|t-1}$ before inflating; (c) ensemble spread of $\boldsymbol{\lambda}^{\mathrm{a}}_{i,t|t-1}$ at model-level 1 at 00 UT on 1 March 2010 when $\beta = 70$.

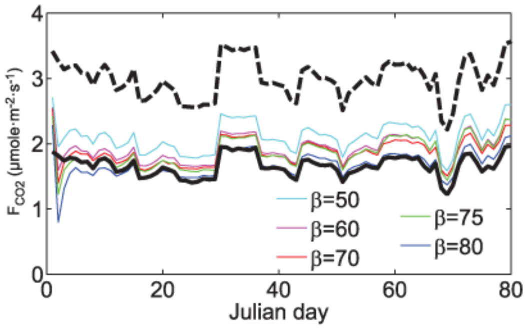

Fig. 12. Time series of daily mean CO_2 fluxes averaged in domain I (shown in Fig. 9a) from 1 Jan. to 20 Mar. 2010 with the inflation factor of scaling factors $\beta = 50$, 60, 70, 75 and 80. The black dashed line is the time series averaged from $\boldsymbol{F}_t^*$ and the black solid line is the time series averaged from $\boldsymbol{F}_t^p$.

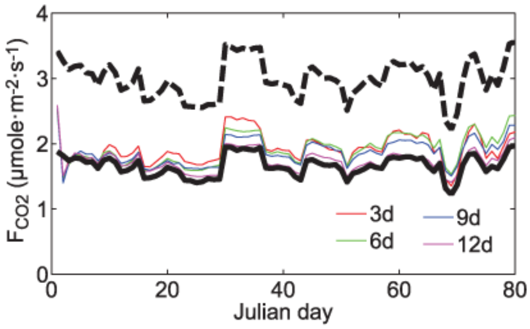

Fig. 13. Time series of daily mean CO_2 fluxes averaged in domain I (shown in Fig. 9a) from 1 Jan. to 20 Mar 2010 with different smoother windows (3, 6, 9 and 12 days). The black dashed line is the time series averaged from $\boldsymbol{F}_t^*$ and the black solid line is the time series averaged from $\boldsymbol{F}_t^{\mathrm{p}}$.

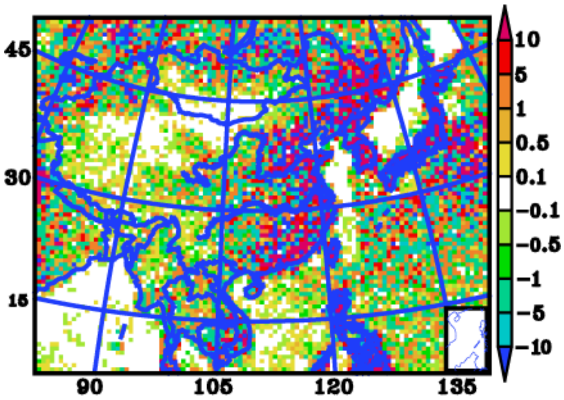

Fig. 14. Monthly mean values of the difference between the prior true surface CO_2 fluxes and the ensemble mean values of the assimilated surface CO_2 fluxes (units: μmole m^{-2} s^{-1}) of the reference experiment in which $\boldsymbol{\lambda}_{i,t|t-1}^{\mathrm{p}}$ were set to 1.