

1 Jan. 5, 2014.

2

3 *Atmos. Chem. Phys.*

4

5 RE: Manuscript Number: acp-2014-496

6

7

8 Dear Editors:

9

10 Thank you very much for your kind decision letter on our paper entitled “**A Regional**
11 **Carbon Data Assimilation System and its Preliminary Evaluation in East Asia**”
12 (**acp-2014-496**). We are grateful for the helpful comments from you and the reviewers.
13 We have changed the manuscript according to the reviewer’s suggestions. And our
14 responses to his comments are detailed in the following document.

15 As today is the deadline, we return the manuscript only revised by us today. All
16 the scientific questions have been resolved in the manuscript (Please see details in it).

17 We have asked a native speaker, Colin Smith, the editor of **Advances in**
18 **Atmospheric Sciences** (<http://159.226.119.58/aas/EN/column/column316.shtml>) to
19 edit this manuscript before I submitted it in June (see the following attachment). Since
20 there are many changes after reviewing, we asked Mr. Colin Smith to edit our
21 manuscript again. However, he was in a holiday last week. He said he can return my
22 manuscript before Jan. 11. So please do me a favor to give us another week to edit this
23 manuscript. We can upload this manuscript as soon as Colin finishes editing. Thank
24 you very much.

25 We are looking forward to hearing from you soon. Thanks again.

26

27 Sincerely Yours,

28

29 Peng Zhen

30

31

1
2

Certificate

LucidPapers

www.lucidpapers.com



Reference number: IAP2014078

Date: 12 June 2014

Corresponding author: Zhen Peng

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

This document certifies that the above-detailed manuscript was edited by a native English-speaking expert at LucidPapers on the date stated.

Signed:



Colin Smith
Chief Editor
LucidPapers

Email: colin.smith@lucidpapers.com
Website: <http://www.lucidpapers.com>

Address:

85 Cotmore Close
Thame
Oxfordshire OX9 3ND
United Kingdom

3
4
5
6
7
8
9
10
11
12
13

1 **Response to the Reviewer’s comments**

2 We are very grateful to the reviewer for his helpful and careful reviews. His
3 thoughtful comments (in bold style below) have helped us to further improve our
4 manuscript. Following his suggestion, we have revised some expressions what are
5 improper, plot some figures again to emphasize the newness in our work. The changes
6 listed below have been incorporated into a final version of the manuscript.

7

8 **1) Page 9 line 11: $j=t-M, \dots$, t is not needed as λ only at $j=t$ is being updated.**

9 I have changed the mistake in Line 11, Page 9.

10 **2) Page 11 line 8 and line 9: both P and S represent the background error**
11 **covariance matrix, and S has the index j from $t-M$ to t . However, when j is**
12 **at t , S is identical to P . You only need one symbol to denote the background**
13 **error covariance matrix.**

14 Yes. It is better to delete one symbol of P and S . In the revised manuscript, I
15 replaced $P_{t,t|t-1}^e$ with $S_{t,t|t-1}^e$.

16 **3) Page 16 line 14 and 15: are you trying to say F_{fire} has zero values except at**
17 **model level 2~5?**

18 NO. What I mean is that the values of the fire flux, $F_{\text{fire}}(x, y, z, t)$, are zero at
19 model 6~15, while they all are not zero at model 1~5. We have revised this expression
20 in the edited manuscript.

21 **4) Page 4 line 4, page 20 line 12 and page 25 line 21: “though” should be**
22 **“through”.**

23 I have revised the mistakes.

24 **5) Page 25 from line 2 to line 13: the explanation here is difficult to understand.**

25 In our investigation, CFI-CMAQ could in general reproduce true fluxes at grid
26 scales with acceptable bias in most model area. But in large cities with strong local
27 anthropogenic CO_2 emissions, it failed to show improvements. Recently, we
28 compared the ensemble spread of $\lambda_{i,t|t-1}^a$ in large cities with those in remote areas. It
29 showed that the values of the ensemble spread of $\lambda_{i,t|t-1}^a$ in large cities are much

1 larger than those in remote areas. We have addressed this in Line 8 to Line 21 in Page
2 24.

3 **6) Page 25 line 17: the descriptions here do not match what were plotted in**
4 **Fig.11.**

5 It is Fig. 12 and I have revised the mistake.

6 **7) Page 27: in your reference case with $\lambda^p=1$, if your window size is around 5**
7 **weeks as recommended by Peters et al(2007), can you get the better results?**

8 Though the 9 days lag-window in the reference experiment is too short compared
9 to the 5 weeks lag-window recommended by Peters et al(2007), this reference
10 experiment could illustrate the limitation by only using the smoothing operator as the
11 persistence dynamical model. If the lag-window was around 5 weeks, we could get
12 better results because there were more observations in every assimilation cycle.
13 However, the results could not be better than those obtained by CFI-CMAQ because
14 most grids have no observations (refer to Fig. 3a) and the signal-to-noise problem still
15 remained. We have addressed this statement in Line 6 to Line 13 in Page 27.

16 **8) Page 40: Flowchart did not reflect the fact that the flux forcing used for \hat{C}^f**
17 **with hat and C^f without hat is different.**

18 They are different in implementation. For $\hat{C}_{i,t}^f$, the flux forcing is F_t^* , which
19 is the prescribed net CO₂ surface flux. We use the red arrows to indicate it. For $C_{i,t}^f$,
20 the flux forcing is $F_{i,t|t-1}^a$, which is calculated as: $F_{i,t|t-1}^a = \lambda_{i,t|t-1}^a F_t^*$ and we use the
21 blue arrows to indicate it.

22 **9) Page 45 Fig.7: recommend to remove (e) and plot $F^p - F^a$ with bar in (f) for**
23 **better comparison with (d).**

24 We deleted Fig. 7e and plot $\overline{F_t^p - F_t^a}$ in Fig. 7 f in the previous manuscript.
25 Please see Fig. 7 in the newly revised version.

1 **10) Page 47 Fig.9: recommend to remove (b) and plot F^p/F^a in (c) for better**
2 **comparison with (a).**

3 We deleted Fig. 9b and plot F_t^*/F_t^p in Fig. 9a and F_t^a/F_t^p in Fig. 9c in the
4 previous manuscript. The white part here indicates the ratios where the absolute
5 values of monthly mean F_t^p are larger than 0.1. Please see Fig. 9 in the newly
6 revised version. Beside, we changed the descriptions about F_t^p/F_t^* in Line 14, Page
7 23 to Line 7, Page 24 in the previous manuscript with the descriptions about F_t^*/F_t^p
8 in Line 13 to 16, Page 23 in the newly revised version.

9 **11) Page 50 Fig.12: recommend to remove (b).**

10 We deleted Fig. 12b.

11 **12) Page 51 Fig.13: recommend to remove (b).**

12 We deleted Fig. 13b.

13

14

15

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

3 Zhen Peng^{*1}, Meigen Zhang^{*2}, Xingxia Kou^{2,3}, Xiangjun Tian⁴, and Xiaoguang Ma⁴

4 ¹*School of Atmospheric Sciences, Nanjing University, Nanjing 210093, China*

5 ²*State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry,*
6 *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

7 ³*University of Chinese Academy of Sciences, Beijing 100049, China*

8 ⁴*Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*
9

*Corresponding author: pengzhen@nju.edu.cn;

*Corresponding author: mgzhang@mail.iap.ac.cn;

ABSTRACT

1
2 In order to optimize surface CO₂ fluxes at grid scales, a regional surface CO₂ flux
3 inversion system (Carbon Flux Inversion system and Community Multi-scale Air
4 Quality, CFI-CMAQ) has been developed by applying the ensemble Kalman filter
5 (EnKF) to constrain the CO₂ concentrations and applying the ensemble Kalman
6 smoother (EnKS) to optimize the surface CO₂ fluxes. The smoothing operator is
7 associated with the atmospheric transport model to constitute a persistence dynamical
8 model to forecast the surface CO₂ flux scaling factors. In this implementation, the
9 ‘signal-to-noise’ problem can be avoided; plus, any useful observed information
10 achieved by the current assimilation cycle can be transferred into the next assimilation
11 cycle. Thus, the surface CO₂ fluxes can be optimized as a whole at the grid scale in
12 CFI-CMAQ. The performance of CFI-CMAQ was quantitatively evaluated through a
13 set of Observing System Simulation Experiments (OSSEs) by assimilating CO₂
14 retrievals from GOSAT (Greenhouse Gases Observing Satellite). The results showed
15 that the CO₂ concentration assimilation using EnKF could constrain the CO₂
16 concentration effectively, illustrating that the simultaneous assimilation of CO₂
17 concentrations can provide convincing CO₂ initial analysis fields for CO₂ flux
18 inversion. In addition, the CO₂ flux optimization using EnKS demonstrated that
19 CFI-CMAQ could in general reproduce true fluxes at grid scales with acceptable bias.
20 Two further sets of numerical experiments were conducted to investigate the
21 sensitivities of the inflation factor of scaling factors and the smoother window. The
22 results showed that the ability of CFI-CMAQ to optimize CO₂ fluxes greatly relied on

1 the choice of the inflation factor. However, the smoother window had a slight
2 influence on the optimized results. CFI-CMAQ performed very well even with a short
3 lag-window (e.g. 3 days).

4

5 **1 Introduction**

6 Considerable progress has been made in recent years to reduce the uncertainties of
7 surface CO₂ flux estimates through the use of an advanced data assimilation technique
8 (e.g., Chevallier et al., 2005, 2007a and 2007b; Baker et al., 2006; Engelen et al., 2009;
9 Liu et al., 2012). Feng et al. (2009) showed that the uncertainties of surface CO₂ flux
10 estimates can be reduced significantly by assimilating OCO X_{CO₂} measurements.
11 Peters et al. (2005, 2007, 2009) developed a surface CO₂ flux inversion system,
12 CarbonTracker, by incorporating the ensemble square-root filter (EnSRF) into the
13 atmospheric transport TM5 model. And the inversion results obtained by assimilating
14 in situ surface CO₂ observations are in excellent agreement with a wide collection of
15 carbon inventories that form the basis of the first North American State of the Carbon
16 Cycle Report (SOCCR) (Peters et al., 2007). CarbonTracker is also well used to
17 constrain the surface CO₂ fluxes over Europe and Asia (eg., Zhang et al., 2014a,
18 2014b). Kang et al. (2012) presented a simultaneous data assimilation of surface CO₂
19 fluxes and atmospheric CO₂ concentrations along with meteorological variables using
20 the Local Ensemble Transform Kalman Filter (LETKF). They indicated that an
21 accurate estimation of the evolving surface fluxes can be gained even without any a
22 priori information. Recently, Tian et al. (2013) developed a new surface CO₂ flux data

1 assimilation system, Tan-Tracker, by incorporating a joint PODEn4DVar assimilation
2 framework into the GEOS-Chem model on the basis of Peters et al. (2005, 2007) and
3 Kang et al. (2011, 2012). They discussed in detail that the assimilation of CO₂ surface
4 fluxes could be improved **through** the use of simultaneous assimilation of CO₂
5 concentrations and CO₂ surface fluxes. Despite the rigor of data assimilation theory,
6 current CO₂ flux-inversion methods still face many challenging scientific problems,
7 such as: (1) the well-known ‘signal-to-noise’ problem (NRC, 2010); (2) large
8 inaccuracies in chemical transport models (e.g., Prather et al., 2008); (3) vast
9 computational expenses (e.g., Feng et al., 2009); and (4) the sparseness of observation
10 data (e.g., Gurney et al., 2002).

11 The ‘signal-to-noise’ problem is one of the most challenging issue for an
12 ensemble-based CO₂ flux inversion system due to the fact that surface CO₂ fluxes are
13 the model forcing (or boundary condition), rather than model states (like CO₂
14 concentrations), of the chemistry transport model (CTM). In the absence of a suitable
15 dynamical model to describe the evolution of the surface CO₂ fluxes, most CO₂
16 flux-inversion studies have traditionally ignored the uncertainty of anthropogenic and
17 other CO₂ emissions and focused on the optimization of natural (i.e., biospheric and
18 oceanic) CO₂ emissions at the ecological scale (e.g., Deng et al., 2007; Feng et al.,
19 2009; Peters et al., 2005, 2007; Jiang et al., 2013; Peylin et al., 2013).

20 This compromise is acceptable to some extent. Indeed, the total amount of
21 anthropogenic CO₂ emissions can be estimated by relatively well-documented global
22 fuel-consumption data with a small degree of uncertainty (Boden et al., 2011). And

1 the uncertainties involved in the total amount of anthropogenic CO₂ emissions are
2 much smaller than those related to natural emissions. However, their spatial
3 distribution, strength and temporal development still remain elusive, because of their
4 inherent non-uniformities (Andres et al., 2012; Gurney et al., 2009). Marland (2008)
5 pointed out that even a tiny amount of uncertainty, i.e., 0.9%, in one of the leading
6 emitter countries like the U.S. is equivalent to the total emissions of the smaller
7 emitter countries in the world. Furthermore, the usual values of anthropogenic CO₂
8 emissions in chemical transport models have thus far been simply interpolated from
9 very coarse monthly-mean fuel consumption data. Therefore, great uncertainty in the
10 spatiotemporal distributions of anthropogenic emissions likely exists, which could
11 reduce the accuracy of CO₂ concentration simulations and subsequently increase the
12 inaccuracy of natural CO₂ flux inversion results. In addition, current research
13 approaches tend only to assimilate natural CO₂ emissions at the ecological scale,
14 which is far from sufficient. Therefore, surface CO₂ fluxes should be constrained as a
15 whole at finer scale.

16 In CarbonTracker (Peters et al., 2007), a smoothing operator is innovatively
17 applied as the persistence forecast model. In that application, the surface CO₂ fluxes
18 can be treated as the model states and the observed information ingested by the
19 current assimilation cycle can be used in the next assimilation cycle effectively.
20 However, the ‘signal-to-noise’ problem is not yet resolved, and thus CarbonTracker
21 also has to assimilate natural CO₂ emissions at the ecological scale only. In
22 Tan-Tracker (Tian et al., 2013), a 4-D moving sampling strategy (Wang et al., 2010)

1 is used to generate the flux ensemble members, and so the surface CO₂ fluxes can be
2 optimized as a whole at the grid scale. In the present reported work, the persistence
3 dynamical model taken by Peters et al. (2005) was further developed for the purpose
4 of resolving the ‘signal-to-noise’ problem to optimize the surface CO₂ fluxes as a
5 whole at the grid scale. This process is described in detail in section 2 of this paper.

6 The surface CO₂ flux inversion system presented in this paper was developed by
7 simultaneous optimizing the surface CO₂ fluxes and constraining the CO₂
8 concentrations. As we know, assimilating CO₂ observations from multiple sources can
9 improve the accuracy of simulation results (e.g., Miyazaki, 2009; Liu et al., 2009,
10 2011, 2012; Tangborn et al, 2013; Huang et al., 2014). In addition, previous studies
11 showed that the simultaneous assimilation of CO₂ concentrations and surface CO₂
12 fluxes can largely eliminate the uncertainty in initial CO₂ concentrations on the CO₂
13 evolution (Kang et al., 2012; Tian et al., 2013). Therefore, we also use the
14 simultaneous assimilation framework and the ensemble Kalman filter (EnKF) was
15 used to constrain CO₂ concentrations and the ensemble Kalman smoother (EnKS) was
16 used to optimize surface CO₂ fluxes. Since the regional chemical transport models,
17 compared to global models, have some advantages to reproduce the effects of
18 meso–micro–scale transport on atmospheric CO₂ distributions (Ahmadov et al., 2009,
19 Pillai et al., 2010; Kretschmer et al., 2011), we choose a regional model, Regional
20 Atmospheric Modeling System and Community Multi-scale Air Quality
21 (RAMS-CMAQ) (Zhang et al. 2002, 2003, 2007; Kou et al. 2013; Liu et al., 2013;
22 Huang et al. 2014), to develop this inversion system. For simplicity, this system is

1 referred to as CFI-CMAQ (Carbon Flux Inversion system and Community Multi-scale
2 Air Quality).

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

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

15

16 **2 Framework of the regional surface CO₂ flux inversion system**

17 Supposed we have the prescribed net CO₂ surface flux, $F^*(x, y, z, t)$, which can be
18 released from a climate model or be generated by others methods, our ultimate goal is
19 to optimize $F^*(x, y, z, t)$ by assimilating CO₂ observations from various platforms.
20 As an ensemble-based assimilation system, CFI-CMAQ was also developed by
21 applying a set of linear multiplication factors, similar to the approach by Peters et al.
22 (2007) and Tian et al. (2013). The i th ensemble member of the surface

1 fluxes, $F_i(x, y, z, t)$, from an N -member ensemble can be described by

$$2 \quad F_i(x, y, z, t) = \lambda_i(x, y, z, t)F^*(x, y, z, t), \quad (i = 1, \dots, N), \quad (1)$$

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

8 At each optimization cycle, CFI-CMAQ includes the following four parts in turn
 9 (see Fig. 1): (1) forecasting of the linear scaling factors at time t , $\lambda_{i,t|t-1}^a$; (2)
 10 optimization of the scaling factors in the smoother window ,
 11 $(\lambda_{i,t-M|t-1}^a, \lambda_{i,t-M+1|t-1}^a, \dots, \lambda_{i,j|t-1}^a, \dots, \lambda_{i,t-1|t-1}^a, \lambda_{i,t|t-1}^a)$, by EnKS, Where
 12 $\lambda_{i,j|t-1}^a$ ($j = t-1-M, \dots, t-1$) refer to analyzed quantities from the previous
 13 assimilation cycle at time j (see Fig. 1), $|t-1$ means that these factors have been
 14 updated by using observations before time $t-1$, and the superscript a refers to the
 15 analyzed; (3) updating of the fluxes in the smoother window ,
 16 $(F_{i,t-M|t-1}^a, F_{i,t-M+1|t-1}^a, \dots, F_{i,j|t-1}^a, \dots, F_{i,t-1|t-1}^a, F_{i,t|t-1}^a)$; and (4) assimilation of the forecast
 17 CO₂ concentration fields at time t , $C_i^f(x, y, z, t)$ (referred to as $C_{i,t}^f$, and the
 18 superscript f refers to the forecast or the background), by EnKF. A flowchart
 19 illustrating CFI-CMAQ is presented in Fig. 2. The assimilation procedure is addressed
 20 in detail below. In addition, the observation operator is introduced, particularly for use
 21 of the GOSAT X_{CO₂} data in Sect. 2.4. Furthermore, covariance inflation and
 22 localization techniques applied in CFI-CMAQ are introduced briefly in Sect. 2.5.

1 **2.1 Forecasting the linear scaling factors at time t , $\lambda_{i,t|t-1}^a$**

2 In the previous assimilation cycle $t-1-M \sim t-1$ (see Fig. 1), the optimized scaling factors
3 in the smoother window are $(\lambda_{i,t-1-M|t-1}^a, \lambda_{i,t-M|t-1}^a, \lambda_{i,t-M+1|t-1}^a, \dots, \lambda_{i,j|t-1}^a, \dots, \lambda_{i,t-1|t-1}^a)$ and
4 the assimilated CO₂ concentration fields at time $t-1$ are $C_i^a(x, y, z, t-1)$ (referred to as
5 $C_{i,t-1}^a$). In the current assimilation cycle $t-M \sim t$, the scaling factors in the current
6 smoother window are $(\lambda_{i,t-M|t-1}^a, \lambda_{i,t-M+1|t-1}^a, \dots, \lambda_{i,j|t-1}^a, \dots, \lambda_{i,t-1|t-1}^a, \lambda_{i,t|t-1}^a)$ and the forecast
7 CO₂ concentration fields at time t are $C_{i,t}^f$.

8 In order to pass the useful observed information onto the next assimilation cycle
9 effectively, following Peters et al. (2007) the smoothing operator is applied as part of
10 the persistence dynamical model to calculate the linear scaling factors $\lambda_{i,t|t-1}^a$,

$$11 \quad \lambda_{i,t|t-1}^a = \frac{\left(\sum_{j=t-M}^{t-1} \lambda_{i,j|t-1}^a + \lambda_{i,t|t-1}^p \right)}{M+1}, \quad (i=1, \dots, N), \quad (2)$$

12 where $\lambda_{i,t|t-1}^p$ refers to the prior values of the linear scaling factors at time t . The
13 superscript p refers to the prior. This operation represents a smoothing over all the
14 time steps in the smoother window (see Fig. 1), thus dampening variations in the
15 forecast of $\lambda_{i,t|t-1}^a$ in time.

16 In order to generate $\lambda_{i,t|t-1}^p$, the atmospheric transport model (CMAQ) is applied
17 and a set of ensemble forecast experiments are carried out. It integrates from time
18 $t-1$ to t to produce the CO₂ concentration fields $\hat{C}_i^f(x, y, z, t)$ (referred to as $\hat{C}_{i,t}^f$
19 hereafter to distinguish from $C_{i,t}^f$) forced by the prescribed net CO₂ surface flux

20 F_t^* with $C_{i,t-1}^a$ as initial conditions. Then, the ratio $\kappa_{i,t} = \hat{C}_{i,t}^f / \overline{\hat{C}_{i,t}^f}$ is calculated,

21 where $\overline{\hat{C}_{i,t}^f} = \frac{1}{N} \sum_{i=1}^N \hat{C}_{i,t}^f$. Supposed that $\lambda_{i,t|t-1}^p = \kappa_{i,t}$ due to the fact that the surface

1 CO₂ fluxes correlate with its concentrations, the values for $\lambda_{i,t|t-1}^p$ are obtained and
2 then $\lambda_{i,t|t-1}^a$ can finally be calculated (see the part with red arrows in the flowchart in
3 Fig. 2).

4 The way the prior scaling factor $\lambda_{i,t|t-1}^p$ is updated by associating with the
5 atmospheric transport model is the main improvement over the one used in
6 CarbonTracker (Peters et al, 2007). In CarbonTracker, all $\lambda_{i,t|t-1}^p$ are set to 1 (Peters et
7 al., 2007). The distribution of the ensemble members of the linear scaling factors at
8 time t , $\lambda_{i,t|t-1}^p$, are finally dependent on the distribution of the previous scaling factors
9 because Eq. (2) is a linear smoothing operator. In this study, the values of $\lambda_{i,t|t-1}^p$ are
10 updated by associating with the atmospheric transport model. It is important to note
11 that $\lambda_{i,t|t-1}^p$ in this study are rand fields with mean 1. However, the distribution of
12 $\lambda_{i,t|t-1}^a$ are dependent on the distribution of all the scaling factors in the smoother
13 window. An OSSE was designed to illustrate the difference between our method and
14 the one where $\lambda_{i,t|t-1}^p$ are set to 1 in Sect. 3

15 It is also important to note that, similar to Peters et al. (2007), this dynamical
16 model equation still does not include an error term in the dynamical model, and the
17 model error cannot yet be estimated. However, the covariance inflation is applied to
18 compensate for model errors before optimization, which is addressed in section 2.5.

19 **2.2 Optimizing the scaling factors in the smoother window by EnKS**

20 Substituting $\lambda_{i,t|t-1}^a$ into Eq. (1), the i th member of the surface fluxes at time t ,
21 $F_{i,t|t-1}^a$, can be generated. Then forced by $F_{i,t|t-1}^a$, CMAQ was run from time $t-1$ to
22 t to produce the background concentration field $C_{i,t}^f$ with $C_{i,t-1}^a$ as initial

1 conditions.

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

$$6 \quad \boldsymbol{\lambda}_{i,j|t}^a = \boldsymbol{\lambda}_{i,j|t-1}^a + \mathbf{K}_{j,t|t-1}^e (\mathbf{y}_t^{\text{obs}} - \mathbf{y}_{i,t}^f + \boldsymbol{v}_{i,t}), \quad (i=1, \dots, N, j=t-M, \dots, t), \quad (3)$$

$$7 \quad \mathbf{K}_{j,t|t-1}^e = \mathbf{S}_{j,t|t-1}^e H^T (H \mathbf{S}_{t,t|t-1}^e H^T + \mathbf{R})^{-1}, \quad (4)$$

$$8 \quad \mathbf{S}_{j,t|t-1}^e = \frac{1}{N-1} \sum_{i=1}^N [\boldsymbol{\lambda}_{i,j|t-1}^a - \overline{\boldsymbol{\lambda}_{i,j|t-1}^a}] [\boldsymbol{\lambda}_{i,t|t-1}^a - \overline{\boldsymbol{\lambda}_{i,t|t-1}^a}]^T, \quad (5)$$

$$9 \quad \mathbf{S}_{t,t|t-1}^e = \frac{1}{N-1} \sum_{i=1}^N [\boldsymbol{\lambda}_{i,t|t-1}^a - \overline{\boldsymbol{\lambda}_{i,t|t-1}^a}] [\boldsymbol{\lambda}_{i,t|t-1}^a - \overline{\boldsymbol{\lambda}_{i,t|t-1}^a}]^T, \quad (6)$$

$$10 \quad \mathbf{y}_{i,t}^f = H(\boldsymbol{\varphi}_{t-1 \rightarrow t}(\boldsymbol{\lambda}_{i,t|t-1}^a)) = H(\mathbf{C}_{i,t}^f), \quad (7)$$

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

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

$$21 \quad \mathbf{S}_{j,t|t-1}^e H^T = \frac{1}{N-1} \sum_{i=1}^N [\boldsymbol{\lambda}_{i,j|t-1}^a - \overline{\boldsymbol{\lambda}_{i,j|t-1}^a}] [\mathbf{y}_{i,t}^f - \overline{\mathbf{y}_t^f}]^T, \quad (8)$$

$$1 \quad HS_{t,t|t-1}^e H^T = \frac{1}{N-1} \sum_{i=1}^N [\mathbf{y}_{i,t}^f - \overline{\mathbf{y}}_t^f][\mathbf{y}_{i,t}^f - \overline{\mathbf{y}}_t^f]^T, \quad (9)$$

$$2 \quad \overline{\mathbf{y}}_t^f = H(\overline{\mathbf{C}}_t^f) = H\left(\frac{1}{N} \sum_{i=1}^N \mathbf{C}_{i,t}^f\right). \quad (10)$$

3 After EnKS, $(\lambda_{i,t-M|t}^a, \lambda_{i,t-M+1|t}^a, \dots, \lambda_{i,j|t}^a, \dots, \lambda_{i,t-1|t}^a, \lambda_{i,t}^a)$ are gained. Then the
 4 corresponding fluxes in the smoother window
 5 $(\mathbf{F}_{i,t-M|t}^a, \mathbf{F}_{i,t-M+1|t}^a, \dots, \mathbf{F}_{i,j|t}^a, \dots, \mathbf{F}_{i,t-1|t}^a, \mathbf{F}_{i,t}^a)$ can be gained (see the part with green arrows
 6 in the flowchart in Fig. 2) by substituting $(\lambda_{i,t-M|t}^a, \lambda_{i,t-M+1|t}^a, \dots, \lambda_{i,j|t}^a, \dots, \lambda_{i,t-1|t}^a, \lambda_{i,t}^a)$ into
 7 Eq. (1).

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

$$10 \quad \overline{\mathbf{F}}_{i,j|t}^a = \frac{1}{N} \sum_{i=1}^N \mathbf{F}_{i,j|t}^a, \quad (j = t-M, \dots, t), \quad (11)$$

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

14 **2.3 Assimilating the CO₂ concentration fields at time t by EnKF**

15 The analysis of CO₂ concentrations fields at time t in the EnKF scheme is updated via

$$16 \quad \mathbf{C}_{i,t}^a = \mathbf{C}_{i,t}^f + \mathbf{K}(\mathbf{y}_t^{\text{obs}} - \mathbf{y}_t^f + \mathbf{v}_{i,t}), \quad (12)$$

$$17 \quad \mathbf{K} = \mathbf{P}^f H^T (H \mathbf{P}^f H^T + \mathbf{R})^{-1}, \quad (13)$$

18 where \mathbf{K} is the Kalman gain matrix of EnKF, \mathbf{P}^f is the background error
 19 covariance among the background CO₂ concentration fields $\mathbf{C}_{i,t}^f$.

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

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

$$2 \quad \mathbf{HP}^f \mathbf{H}^T = \frac{1}{N-1} \sum_{i=1}^N [\mathbf{y}_{i,t}^f - \overline{\mathbf{y}}_t^f]^T [\mathbf{y}_{i,t}^f - \overline{\mathbf{y}}_t^f]^T, \quad (15)$$

$$3 \quad \overline{\mathbf{C}}_t^f = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{i,t}^f \quad (16)$$

4 Finally, the ensemble mean values of the assimilated CO₂ concentrations fields can be
5 gained via,

$$6 \quad \overline{\mathbf{C}}_t^a = \frac{1}{N} \sum_{i=1}^N \mathbf{C}_{i,t}^a \quad (17)$$

7 where $\overline{\mathbf{C}}_t^a$ is regarded as the final analyzing concentration field.

8 **2.4 The observation operator**

9 As mentioned above, the observation operator $H(\cdot)$ transforms the state variable
10 from model space into observation space. Usually, it is the spatial bilinear interpolator
11 for traditional ground-based observations. Since the GOSAT X_{CO_2} retrieval is a
12 weighted CO₂ column average, the simulated X_{CO_2} should be calculated with the same
13 weighted column average method (Connor et al., 2008; Crisp et al., 2010, 2012;
14 O'Dell et al, 2012). So, the observation operator to assimilate the GOSAT X_{CO_2}
15 retrieval is

$$16 \quad \mathbf{y}_{i,t}^f = H(\boldsymbol{\varphi}_{t-1 \rightarrow t}(\boldsymbol{\lambda}_{i,t/t-1}^a)) = H(\mathbf{C}_{i,t}^f) = \mathbf{y}^{\text{priori}} + \mathbf{h}^T \mathbf{a}_{\text{CO}_2} (S(\mathbf{C}_{i,t}^f) - \mathbf{f}^{\text{priori}}), \quad (18)$$

17 where $\mathbf{y}_{i,t}^f$ is the simulated X_{CO_2} ; $\mathbf{y}^{\text{priori}}$ is the a priori CO₂ column average used in
18 the GOSAT X_{CO_2} retrieval process; $S(\cdot)$ is the spatial bilinear interpolation operator
19 that interpolates the simulated fields to the GOSAT X_{CO_2} locations to obtain the
20 simulated CO₂ vertical profiles there; $\mathbf{f}^{\text{priori}}$ is the a priori CO₂ vertical profile used

1 in the retrieval process; h is the pressure weighting function, which indicates the
 2 contribution of the retrieved value from each layer of the atmosphere; and \mathbf{a}_{CO_2} is
 3 the normalized averaging kernel.

4 **2.5 Covariance inflation and localization**

5 In order to keep the ensemble spread of the CO₂ concentrations at a certain level and
 6 compensate for transport model error to prevent filter divergence, covariance inflation
 7 is applied before updating the CO₂ concentrations. So,

$$8 \quad (\mathbf{C}_{i,t}^f)_{\text{new}} = \alpha(\mathbf{C}_{i,t}^f - \overline{\mathbf{C}_{i,t}^f}) + \overline{\mathbf{C}_{i,t}^f}, \quad (19)$$

9 where α is the inflation factor of CO₂ concentrations and $(\mathbf{C}_{i,t}^f)_{\text{new}}$ is the final field
 10 used for data assimilation.

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

$$13 \quad (\lambda_{i,t|t-1}^p)_{\text{new}} = \beta(\lambda_{i,t|t-1}^p - \overline{\lambda_{i,t|t-1}^p}) + \overline{\lambda_{i,t|t-1}^p}, \quad (20)$$

14 where β is the inflation factor of scaling factors and $(\lambda_{i,t|t-1}^p)_{\text{new}}$ is the final scaling
 15 factors used for data assimilation.

16 In addition, the Schur product is utilized to filter the remote correlation resulting
 17 from the spurious long-range correlations (Houtekamer and Mitchell 2001). So, the

18 Kalman gain matrix $\mathbf{K}_{j,t|t-1}^e$ and \mathbf{K} are updated via,

$$19 \quad \mathbf{K}_{j,t|t-1}^e = [(\boldsymbol{\rho} \circ \mathbf{S}_{j,t|t-1}^e)H^T (H(\boldsymbol{\rho} \circ \mathbf{P}_{t,t|t-1}^e)H^T + \mathbf{R})^{-1}], \quad (21)$$

$$20 \quad \mathbf{K} = [(\boldsymbol{\rho} \circ \mathbf{P}^f)H^T][(H(\boldsymbol{\rho} \circ \mathbf{P}^f)H^T + \mathbf{R})^{-1}], \quad (22)$$

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

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

where c is the element of the localization Schur radius. The matrix ρ 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 $6654 \times 5440 \text{ km}^2$ on a rotated polar stereographic map projection centered at $(35.0^\circ \text{N}, 116.0^\circ \text{E})$, with a horizontal grid resolution of $64 \times 64 \text{ km}^2$ 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

1 fluxes included biosphere–atmosphere CO₂ fluxes, ocean–atmosphere CO₂ fluxes,
 2 anthropogenic emissions, and biomass-burning emissions (Kou et al., 2013),

$$3 \quad F^P(x, y, z, t) = F_{\text{bio}}(x, y, z, t) + F_{\text{oce}}(x, y, z, t) + F_{\text{ff}}(x, y, z, t) + F_{\text{fire}}(x, y, z, t), \quad (24)$$

4 where $F^P(x, y, z, t)$ (referred to as F_t^P) was the prior surface CO₂ flux;
 5 $F_{\text{bio}}(x, y, z, t)$ and $F_{\text{oce}}(x, y, z, t)$ were the biosphere–atmosphere and
 6 ocean–atmosphere CO₂ fluxes, respectively, which were obtained from the optimized
 7 results of CarbonTracker 2011 (Peters, 2007); $F_{\text{ff}}(x, y, z, t)$ was fossil fuel emissions,
 8 adopted from the Regional Emission inventory in ASia (REAS, 2005 Asia monthly
 9 mean emission inventory) with a spatial resolution of $0.5^\circ \times 0.5^\circ$ (Ohara et al., 2007);
 10 $F_{\text{fire}}(x, y, z, t)$ was biomass–burning emissions, provided by the monthly mean
 11 inventory at a spatial resolution of $0.5^\circ \times 0.5^\circ$ from the Global Fire Emissions
 12 Database, Version 3 (GFED v3) (Van der Werf et al., 2010). Among all these fluxes,
 13 $F_{\text{bio}}(x, y, z, t)$, $F_{\text{oce}}(x, y, z, t)$ and $F_{\text{ff}}(x, y, z, t)$ had nonzero values at model level 1,
 14 while they all were zeros at other 14 levels. However, $F_{\text{fire}}(x, y, z, t)$ had nonzero
 15 values at model level 1~5 and they were all zeros at other 10 levels. So, all fluxes in
 16 this paper were the function of (x, y, z, t) for convenience.

17 Firstly, the prior flux F_t^P was assumed as the true surface CO₂ flux in all of the
 18 following OSSEs. Forced by F_t^P , the RAMS-CMAQ model was run to produce the
 19 artificial true CO₂ concentration results $C^P(x, y, z, t)$ (refer to as C_t^P in the
 20 following). Then, the artificial GOSAT observations y_t^{obs} (or $X_{\text{CO}_2}^P$) were generated
 21 by substituting C_t^P into the observation operator in Eq. (18). The retrieval
 22 information of GOSAT $X_{\text{CO}_2}(y^{\text{priori}}, f^{\text{priori}}, h$ and $a_{\text{CO}_2})$ needed in Eq. (18) were

1 gained from the v2.9 Atmospheric CO₂ Observations from Space (ACOS) Level 2
 2 standard data products, which only utilized the SWIR observations. Only data
 3 classified into the “Good” category were utilized in this study. During the retrieval
 4 process, most of the soundings (such as data with a solar zenith angle greater than 85 °,
 5 or data not in clear sky conditions, or data collected over ocean but not in glint, etc.)
 6 were not processed, so typically data products for the “Good” category contained only
 7 10-100 soundings per satellite orbit (Osterman et al., 2011), and there were only 0~60
 8 samples per orbit in the study model domain generally. Fig. 3 (a) also showed the total
 9 number of “good” GOSAT X_{CO2} observations for each model grid in February in 2010.
 10 There were relatively more observations over most continental regions of the study
 11 domain except some regions in North-East and South China. The total numbers
 12 ranged from 1 to 8. However, there were almost no data over oceans of the study
 13 domain.

14 Secondly, the prescribed surface CO₂ fluxes series F_t^* were created by

$$15 \quad F_t^* = (1.8 + \delta(x, y, z, t))F_t^p, \quad (25)$$

16 where δ was a random number. They were standard normal distribution time series
 17 at each grid in the integration period of our numerical experiment. Driven by F_t^* , the
 18 RAMS-CMAQ model was integrated to obtain the CO₂ simulations
 19 $C^f(x, y, z, t)$ (referred to as C_t^f hereafter). Then, the column-averaged
 20 concentrations $X_{CO_2}^f$ were obtained using Eq. (18).

21 The performance of CFI-CMAQ was evaluated through a group of well-designed
 22 OSSEs. And the goal of each OSSE was to retrieve the true fluxes F_t^p from given

1 true observations $X_{CO_2}^p$ and “wrong” fluxes F_t^* . In all the OSSEs, we assimilated
2 artificial observations $X_{CO_2}^p$ about three times a day since GOSAT has about three
3 orbits in the study model domain. If there were some observations, CFI-CMAQ
4 paused to assimilate. Otherwise, it continued simulating. The default ensemble size N
5 was 48, the measurement errors were 1.5 ppmv, the standard localization Schur radius
6 c was 1280 km (20 grid spacing), and the covariance inflation factor of
7 concentrations α was 1.1. The referenced lag-window was 9 days and the
8 covariance inflation factor of the prior scaling factors β was 70. Since the smoother
9 window was very important for CO₂ transportation and β was a newly introduced
10 parameter, both these parameters were further investigated by several numerical
11 sensitivity experiments. The primary focus of this paper was to describe the
12 assimilation methodology, so all the numerical experiments started on 1 January 2010
13 and ended on 30 March 2010.

14 As for the initialization of CFI-CMAQ, only the ensemble of background
15 concentration fields $C_{i,0}^f$ needed to be initialized at the time $t=0$ because the
16 values of $\lambda_{i,t|t-1}^a$ were updated by using the persistence dynamical model. In practice,
17 the mean concentration fields at $t=0$ are interpolated from the simulated CO₂ fields
18 of CarbonTracker 2011 (Peters, 2007). The ensemble members of the background
19 concentration fields were created by adding random vectors. The mean values of the
20 random vectors were zero and the variances were 2.5 percent of the mean
21 concentration fields. Then the atmospheric transport model integrated from time
22 $t=0$ to $t=1$ driven by F_t^* with $C_{i,0}^f$ as initial conditions to produce the CO₂

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

9 In order to illustrate the limitation by only using the smoothing operator as the
10 persistence dynamical model to generate all future scaling factors, another OSSE
11 (referred to as the reference experiment to distinguish it from the above-mentioned
12 CFI-CMAQ OSSEs) was designed to optimize the surface CO₂ fluxes at grid scale.
13 The reference experiment was under the same assimilation framework as CFI-CMAQ
14 except that all $\lambda_{i,t|t-1}^p$ were set to 1 (Peters et al., 2007). Besides, the initialization
15 procedure of the reference experiment was different from that of the CFI-CMAQ. In
16 practice, both the ensemble of background concentration fields at $t=0$, $C_{i,0}^f$, and the
17 ensemble members of the scaling factors at $t=1$, $\lambda_{i,1|0}^a$, needed to be initialized
18 because they could not generated by other ways (Peters et al., 2005). The initial
19 concentration fields $C_{i,0}^f$ were created using the same method as that was used to
20 generate $C_{i,0}^f$ for the CFI-CMAQ OSSEs. The ensemble members of the scaling
21 factors $\lambda_{i,1|0}^a$ were rand fields. Their mean values were 1 and their variances were 0.1.
22 In addition, in order to keep the ensemble spread of the scaling factors $\lambda_{i,t|t-1}^a$ at a

1 certain level and compensate for dynamical model error, covariance inflation was also
2 used and the covariance inflation factor of the scaling factors $\lambda_{i,t|t-1}^a$ was 1.6. All
3 other parameters are the same as used in the CFI-CMAQ OSSEs. The ensemble size N
4 was 48, the measurement errors were 1.5 ppmv, the standard localization Schur radius
5 c was 1280 km, the covariance inflation factor of concentrations α was 1.1, and
6 the lag-window was 9 days.

7 **3.2 Experimental results**

8 Essentially, the assimilation part of CFI-CMAQ includes two subsections: one for the
9 CO₂ concentration assimilation with EnKF, which can provide a convincing CO₂
10 initial analysis fields for the next assimilation cycle; and the other for the CO₂ flux
11 optimization with EnKS, which can provide better estimation of the scaling factors for
12 the next time **through** the persistence dynamical model except for optimized CO₂
13 fluxes. The performance of the EnKF subsection will be greatly influenced by the
14 validation of the EnKS subsection, or vice versa. Firstly, the performance of
15 CFI-CMAQ will be quantitatively assessed in detail by using the assimilated results of
16 a CFI-CMAQ OSSE, in which the lag-window was 9 days and β was 70. Then the
17 sensitivities of β and the lag-window will be discussed in the following two
18 paragraphs. And finally, the assimilation results of the reference experiment in
19 which $\lambda_{i,t|t-1}^p$ were set to 1 will be described in brief at the end of this subsection.

20 We begin by describing the impacts of assimilating artificial observations $X_{CO_2}^p$
21 on CO₂ simulations by CFI-CMAQ. As shown in Figs. 4a, 4b and 4d, the monthly
22 mean values of the background CO₂ concentrations C_t^f produced by the magnified

1 surface CO₂ fluxes F_t^* were much larger than those of the artificial true CO₂
 2 concentrations C_t^p produced by the prior surface CO₂ fluxes F_t^p near the surface in
 3 February 2010. In the east and south of China especially, the magnitude of the
 4 difference between C_t^p and C_t^f was at least 6 ppmv. Also, as expected, the monthly
 5 mean $X_{CO_2}^f$ was much larger than the monthly mean artificial observations $X_{CO_2}^p$,
 6 and the magnitude of the difference between $X_{CO_2}^p$ and $X_{CO_2}^f$ reached 2 ppmv in
 7 the east and south of China (see Figs. 3b, 3c and 3e). However, the impact of
 8 magnifying surface CO₂ fluxes on the CO₂ concentrations was primarily below the
 9 model-level 10 (approximately 6 km), and especially below model-level 7
 10 (approximately 1.6 km). And above model-level 10, the differences between C_t^p and
 11 C_t^f fell to zero (see Fig. 5a and 5b). After assimilating $X_{CO_2}^p$, the analysis CO₂
 12 concentrations \overline{C}_t^a was much closer to C_t^p (see Figs. 4c, 4e and 4f). The monthly
 13 mean difference between C_t^p and \overline{C}_t^a ranged from -2 to 2 ppmv and the relative
 14 error $(C_t^p - \overline{C}_t^a) / C_t^p$ ranged from -1 to 1% in almost the entire model domain at
 15 model-level 1. The monthly mean differences between C_t^p and \overline{C}_t^a were negligible
 16 above model-level 2 (see Fig. 5c and 5d). The monthly mean $X_{CO_2}^a$ was also closer
 17 to $X_{CO_2}^p$ and the difference between $X_{CO_2}^p$ and $X_{CO_2}^a$ ranged from -0.5 to 0.5
 18 ppmv. In order to evaluate the general impact of assimilating $X_{CO_2}^p$ in the surface
 19 layer, time series of the daily mean CO₂ concentration extracted from the background
 20 simulations and the assimilations were compared with the artificial true simulations at
 21 four national background stations in China and their nearest large cities. As shown in
 22 Fig. 3a, Waliguan is 150 km away from Xining, Longfengshan is 180 km away from

1 Haerbin, Shangdianzi is 150 km away from Beijing, and Linan is 50 km away from
2 Hangzhou. The assimilated results are shown in Fig. 6. The background time series
3 were much larger than the artificial true time series, especially at Shangdianzi, Beijing,
4 Linan and Hangzhou, which are strongly influenced by local anthropogenic CO₂
5 emissions. After assimilating $X_{CO_2}^p$, the assimilated time series were very close to the
6 true time series with negligible bias, as expected, at Waliguan, Xining, Shangdianzi,
7 Beijing, Linan and Hangzhou, especially after the first 10 days, which can be
8 considered the spin-up period. Meanwhile, the improvements at Longfengshan and
9 Haerbin were limited due to the absence of observation data at those locations (see
10 Fig. 3a). Nevertheless, in general, the substantial benefits to the CO₂ concentrations in
11 the surface layer of assimilating GOSAT X_{CO_2} with EnKF are clear. All the results
12 illustrated that CFI-CMAQ can provide a convincing CO₂ initial analysis fields for
13 CO₂ flux inversion.

14 The impacts of assimilating $X_{CO_2}^p$ on surface CO₂ fluxes were also highly
15 impressive by CFI-CMAQ. On the whole, the prescribed CO₂ surface fluxes F_t^*
16 were much larger than the true surface CO₂ fluxes F_t^p in February 2010, especially
17 in the east and south of China. The monthly mean difference between F_t^* and F_t^p
18 reached 5 $\mu\text{mole m}^{-2} \text{s}^{-1}$ in Jing-Jin-Ji, the Yangtze River Delta, and Pearl River Delta
19 Urban Circle because of the strong local anthropogenic CO₂ emissions (see Figs. 7a,
20 7b and 7d). After assimilating $X_{CO_2}^p$, the ensemble mean of the assimilated surface
21 CO₂ fluxes $\overline{F_t^a}$ decreased sharply. Thus, the monthly mean values of $\overline{F_t^a}$ were
22 much smaller than F_t^* in most of the model domain in February 2010. The pattern of

1 the difference between $\overline{F_t^a}$ and F_t^* was similar to that of the difference between
2 F_t^p and F_t^* (see Fig. 7d). The ensemble mean of the assimilated surface CO₂ fluxes
3 $\overline{F_t^a}$ were also compared to the artificial true fluxes F_t^p , revealing that $\overline{F_t^a}$ was
4 equivalent to F_t^p in most of the model domain. The monthly mean difference
5 between $\overline{F_t^a}$ and F_t^p ranged from -0.1 to $0.1 \mu\text{mole m}^{-2} \text{s}^{-1}$ only (see Fig. 7e). In
6 addition, the root-mean-square errors (RMSEs) of the assimilated flux members were
7 analyzed. As shown in Fig. 8, the monthly mean RMSE was less than $0.5 \mu\text{mole m}^{-2}$
8 s^{-1} in most of the model domain, except in areas near to large cities such as Beijing,
9 Shanghai and Guangzhou, indicating that the assimilated CO₂ fluxes were reliable.

10 In order to evaluate the ability of CFI-CMAQ to optimize the surface CO₂ fluxes
11 comprehensively, the ratios of the monthly mean F_t^* to the monthly mean F_t^p were
12 analyzed. In actual implementation, we only analyzed the ratios where the absolute
13 values of the monthly mean F_t^p were larger than 0.1, to avoid random noise. As
14 shown in Fig. 9a, the ratios of the monthly mean F_t^* to the monthly mean F_t^p are
15 about 1.8 in most of China, except in the Qinghai–Tibet Plateau, where the absolute
16 values of the monthly mean F_t^p in February were very small and we did not analyze.
17 In addition, the ratios of the monthly mean $\overline{F_t^a}$ to the monthly mean F_t^p are shown
18 in Fig. 9b. This figure demonstrates that the impact of the assimilation of $X_{CO_2}^p$ by
19 CFI-CMAQ on CO₂ fluxes was great in the east and south of China in general, but the
20 influence was negligible in Northeast China due to the lack of observation data.

21 Time series of daily mean surface CO₂ fluxes extracted from F_t^* and $\overline{F_t^a}$ were
22 also compared with that from F_t^p at four national background stations in China and

1 their nearest large cities, similar to the CO₂ concentration assimilation. The results are
2 shown in Fig.10. The background time series were much larger than the artificial true
3 time series, especially at Haerbin, Shangdianzi, Beijing, Linan and Hangzhou, which
4 are strongly influenced by local anthropogenic CO₂ emissions. After
5 assimilating $X_{CO_2}^p$, the assimilated time series were near to the true time series with
6 acceptable bias, as expected, at Waliguan, Xining, Shangdianzi, Linan and Hangzhou
7 after the 10-day spin-up period. However, the improvements at Longfengshan and
8 Haerbin were negligible because of a lack of observations at these locations. Also, this
9 inversion system failed to show improvements at Beijing. One of the possible reasons
10 was that the values of the ensemble spread of $\lambda_{i,t|t-1}^a$ in Beijing area are too large (see
11 Fig. 11c). Beijing was located in Jing-Jin-Ji Urban Circle, which had strong local
12 anthropogenic CO₂ emissions during January to March. So the values of the ensemble
13 spread of $C_{i,t}^f$ in Beijing area at model-level 1 could be much larger than those in
14 other areas, which had weak local anthropogenic CO₂ emissions (see Fig. 11a). As a
15 result, the values of the ensemble spread of $\lambda_{i,t|t-1}^p$ before inflating in Beijing area are
16 much larger than those in other areas with small local anthropogenic CO₂ emissions
17 (see Fig. 11b). After inflating, the ensemble spread of $\lambda_{i,t|t-1}^p$ in Beijing area could be
18 too large, compare to those in other areas with small local anthropogenic CO₂
19 emissions (see Fig. 11c), which lead to the failure to reproduce the true fluxes in
20 Beijing area. Later, CFI-CMAQ will be improved by optimizing the covariance
21 inflation method.

22 Since the impact of assimilation $X_{CO_2}^p$ by CFI-CMAQ on CO₂ fluxes was in

1 general greater in the east and south of China than other model areas (see Figs.7 and
2 9), the time series of the daily mean CO₂ fluxes in that area averaged from $\overline{F_t^a}$ was
3 compared with those from F_t^* and F_t^p (see Fig. 12). This figure indicates that
4 CFI-CMAQ could in general reproduce the true fluxes with acceptable bias.

5 As stated in the above section, β was a newly introduced parameter. The prior
6 scaling factors should have been inflated indirectly through the inflated CO₂
7 concentration forecast. However, the values of the ensemble spread of $\lambda_{i,t|t-1}^p$ before
8 inflating were very small (ranging from 0 to 0.08 in most area at model-level 1, see
9 Fig. 11b), though the values of the ensemble spread of $C_{i,t}^f$ after inflating could
10 reach 1 to 14 ppmv in most area at model-level 1 (see Fig. 11a). So we had to inflate
11 them again before using them into Eq. (2) . Fig. 11c showed the distribution of the
12 ensemble spread of $\lambda_{i,t|t-1}^a$ at model-level 1 at 00 UT on 1 March 2010 when $\beta = 70$. It
13 showed that the values of the ensemble spread of $\lambda_{i,t|t-1}^a$ ranged from 0.1 to 0.8 in
14 most area. In order to investigate the sensitivity of the inflation factor of the scaling
15 factors β , a series of numerical experiments were conducted. As shown in Fig. 12,
16 CFI-CMAQ worked rather well for $\beta = 60, 70, 75, 80$. However, if β was much
17 smaller than 50 (e.g. $\beta = 10$), the impact of assimilation was small due to the small
18 ensemble spread; or if β was much larger than 80 (e.g. $\beta = 100$), the assimilated
19 CO₂ fluxes deviated markedly from the “true” CO₂ fluxes. In other words, the
20 performance of CFI-CMAQ greatly relies on the choice of β .

21 From the perspective of the lag-window, the differences among the four
22 assimilation sensitivity experiments with lag-windows of 3, 6, 9 and 12 days were

1 very small (see Fig. 13). Although Peters et al. (2007) indicated that the lag-window
 2 should be more than five weeks, it seemed that the smoother window had a slight
 3 influence on the assimilated results for CFI-CMAQ. It was clear that the assimilated
 4 results with a larger lag-window were better than those with a smaller lag-window;
 5 however, CFI-CMAQ performed very well even with a short lag-window (e.g. 3
 6 days).

7 At the end of this subsection, the assimilation results of the reference experiment
 8 in which $\lambda_{i,t-1}^p$ were set to 1 will be addressed briefly. The impact of assimilation
 9 $X_{CO_2}^p$ on CO₂ fluxes was disordered. The monthly mean values of the difference
 10 between the prior true surface CO₂ fluxes and the ensemble mean values of the
 11 assimilated surface CO₂ fluxes were irregular noise (see Fig. 14). The main reason is
 12 that all the elements of the scaling factors to be optimized in the smoother window are
 13 only random numbers. As stated in the above section, only $\lambda_{i,1|0}^a$ needed to be
 14 optimized in the first assimilation cycle. However, $\lambda_{i,1|0}^a$ were rand fields (in other
 15 words, all the elements of $\lambda_{i,1|0}^a$ are only random numbers) because they could not
 16 generated by other ways at the first time. So their spatial correlations were too small.
 17 The correlations between the scaling factors and the observations were also too small.
 18 Therefore it was impossible to systematically change the values of $\lambda_{i,1|0}^a$ in large areas
 19 where the observations located after assimilating observations at $t=1$. Thus the
 20 signal-to-noise problem arose. So the elements of $\lambda_{i,1|1}^a$ are only random numbers too.
 21 Though $\lambda_{i,2|1}^a$ could be generated automatically by the smoothing operator when all
 22 $\lambda_{i,2|1}^p$ were set to 1, the elements of $\lambda_{i,2|1}^a$ are random numbers too since the smoothing

1 operator is only a linear operator. Similarly, it was impossible to systematically
2 change the values of $\lambda_{i,1\parallel}^a$ and $\lambda_{i,2\parallel}^a$ in large areas after assimilating observations at
3 $t = 2$. As this inversion system continued assimilating observations, all future scaling
4 factors could be created by the smoothing operator and then updated. But this
5 inversion system could not ingest the observations effectively because all the elements
6 of the scaling factors were always random numbers. **Though the 9 days lag-window in**
7 **the reference experiment is too short compared to the 5 weeks lag-window**
8 **recommended by Peters et al(2007), this reference experiment could illustrate the**
9 **limitation by only using the smoothing operator as the persistence dynamical model.**
10 **If the lag-window was around 5 weeks, we could get better results because there were**
11 **more observations in every assimilation cycle. However, the results could not be**
12 **better than those obtained by CFI-CMAQ because most grids have no observations**
13 **(refer to Fig. 3a) and the signal-to-noise problem still remained.**

14

15 **4 Summary and conclusions**

16 A regional surface CO₂ flux inversion system, CFI-CMAQ, has been developed
17 to optimize CO₂ fluxes at grid scales. It operates under a joint data assimilation
18 framework by applying EnKF to constrain the CO₂ concentrations and applying EnKS
19 to optimize the surface CO₂ flux, which is similar to Kang et al. (2011, 2012) and
20 Tian et al. (2013). The persistence dynamical model, which was first introduced by
21 Peters et al. (2007) by applying the smoothing operator to transport the useful
22 observed information onto the next assimilation cycle, is further developed. We

1 associated the smoothing operator with the atmospheric transport model to constitute
2 the persistence dynamical model to forecast the surface CO₂ flux scaling factors for
3 the purpose of resolving the ‘signal-to-noise’ problem, as well as transporting the
4 useful observed information onto the next assimilation cycle. In this application, the
5 scaling factors to be optimized in the flux inversion system can be forecast at the grid
6 scale without random noise. The OSSEs showed that the performance of CFI-CMAQ
7 is effective and promising. In general, it could reproduce the true fluxes at the grid
8 scale with acceptable bias.

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

15

16 *Acknowledgments.* This work was supported by the National Natural Science Foundation of China (Grant No. 41130528),
17 the Strategic Priority Research Program–Climate Change: Carbon Budget and Relevant Issues (XDA05040404), the National
18 High Technology Research and Development Program of China (2013AA122002). CarbonTracker results used to generate the
19 initial condition are provided by NOAA ESRL, Boulder, Colorado, USA from the website at <http://carbontracker.noaa.gov>. The
20 numerical calculations in this paper have been done on the IBM Blade cluster system in the High Performance Computing Center
21 (HPCC) of Nanjing University.

22

23 **References**

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

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

3 Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO₂,
4 Tellus B, 58, 359–365, 2006.

5 Boden, T. A., Marland, G., and Andres, R. J.: Global, regional, and national fossil-fuel CO₂
6 emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory,
7 U.S. Department of Energy, Oak Ridge, Tenn., U.S.A, doi:10.3334/CDIAC/00001_V2011,
8 2011.

9 Chevallier, F. M. F., Peylin, P., Bousquet, S. S. P., Br éon, F.-M., Ch édin, A., and Ciais, P.:
10 Inferring CO₂ sources and sinks from satellite observations: Method and application to TOVS
11 data, J. Geophys. Res., 110, D24309, doi:10.1029/2005JD006390, 2005.

12 Chevallier, F., Br éon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory
13 to the estimation of CO₂ sources and sinks: Theoretical study in a variational data
14 assimilation framework, J. Geophys. Res., 112, D09307, doi:10.1029/2006JD007375, 2007a.

15 Chevallier, F.: Impact of correlated observation errors on inverted CO₂ surface fluxes from OCO
16 measurements, Geophys. Res. Lett., 34, L24804, doi:10.1029/2007GL030463, 2007b.

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

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

27 Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R.,
28 Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A.,
29 Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V.,
30 Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R.,

1 Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O.,
2 Wunch, D., and Yung, Y. L.: The ACOS CO₂ retrieval algorithm – Part II: Global X_{CO₂} data
3 characterization, *Atmos. Meas. Tech.*, 5, 687-707, doi:10.5194/amt-5-687-2012, 2012.

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

7 Engelen, R. J., Serrar, S., and Chevallier, F.: Four-dimensional data assimilation of atmospheric
8 CO₂ using AIRS observations, *J. Geophys. Res.*, 114, D03303, doi:10.1029/2008JD010739,
9 2009.

10 Feng, L., Palmer, P. I., Bösch, H., and Dance, S.: Estimating surface CO₂ fluxes from space-borne
11 CO₂ dry air mole fraction observations using an ensemble Kalman Filter, *Atmos. Chem.*
12 *Phys.*, 9, 2619–2633, 2009.

13 Feng, L., Palmer, P. I., Yang, Y., Yantosca, R. M., Kawa, S. R., Paris, J.-D., Matsueda, H., and
14 Machida, T.: Evaluating a 3-D transport model of atmospheric CO₂ using ground-based,
15 aircraft, and space-borne data, *Atmos. Chem. Phys.*, 11, 2789–2803, doi:
16 10.5194/acp-11-2789-2011, 2011.

17 Gurney, K. R., Law, R. L., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L.,
18 Chen, Y. H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J.,
19 Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J.,
20 Sarmiento, J., Taguchi, S., Takahashi, T., Yuen, C. W.: Towards robust regional estimates of
21 CO₂ sources and sinks using atmospheric transport models, *Nature*, 415, 626–630, 2002

22 Gurney, K. R., Mendoza, D. L., Zhou, Y. Y., Fischer, M. L., Miller, C. C., Geethakumar, S. and
23 Du Can, S. D.: High resolution fossil fuel combustion CO₂ emission fluxes for the United
24 States, *Environ. Sci. & Technol.*, 43, 5535-5541. doi:10.1021/es900806c, 2009.

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

27 Huang, Z. K., Peng, Z., Liu, H. N., Zhang, M. G.: Development of CMAQ for East Asia CO₂ data
28 assimilation under an EnKF framework: a first result, *Chinese Science Bulletin*, 59:
29 3200-3208, doi: 10.1007/s11434-014-0348-9, 2014.

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

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

3 Kang, J.-S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T., and Ide, K.: “Variable localization” in an
4 ensemble Kalman filter: application to the carbon cycle data assimilation, J. Geophys. Res.,
5 116, D09110, doi:10.1029/2010JD014673, 2011.

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

9 Kou, X., Zhang, M., Peng, Z.: Numerical Simulation of CO₂ Concentrations in East Asia with
10 RAMS-CMAQ, Atmos, Oceanic Sci Lett, 6, 179-184, 2013.

11 Kretschmer, R., Gerbig, C., Karstens, U., Koch, F.-T.: Error characterization of CO₂ vertical
12 mixing in the atmospheric transport model WRF-VPRM, Atmos. Chem. Phys., 12,
13 2441–2458, 2012.

14 Liu, J., Fung, I., Kalnay, E., Kang, J.: CO₂ transport uncertainties from the uncertainties in
15 meteorological fields, Geophys. Res. Lett., 38, L12808, doi: 10.1029/2011GL047213,
16 2011.

17 Liu, J., Fung, I., Kalnay, E., Kang, J.-S., Olsen, E. T., and Chen, L.: Simultaneous assimilation of
18 AIRS X_{co2} and meteorological observations in a carbon climate model with an ensemble
19 Kalman filter, J. Geophys. Res., 117, D05309, doi: 10.1029/2011JD016642, 2012.

20 Liu Z., Bambha, R. P., Pinto, J. P.: Toward verifying fossil fuel CO₂ emissions with the
21 Community Multi-scale Air Quality (CMAQ) model: motivation, model description and
22 initial simulation. J. Air & Waste Management Assoc., 64, 419-435, doi:
23 10.1080/10962247.2013.816642, 2013.

24 Marland, G.: Uncertainties in accounting for CO₂ from fossil fuels, J. of Indust. Ecol. 12, 136-139,
25 doi:10.1111/j.1530-9290.2008.00014.x, 2008.

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

29 O’Dell, C. W., Connor, B., Bösch, H., O’Brien, D., Frankenberg, C., Castano, R., Christi, M.,
30 Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F.,

1 Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The
2 ACOS CO₂ retrieval algorithm – Part 1: Description and validation against synthetic
3 observations, *Atmos. Meas. Tech.*, 5, 99–121, doi:10.5194/amt-5-99-2012, 2012.

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

6 Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D.,
7 Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO₂ surface
8 fluxes from atmospheric trace gas observations, *J. Geophys. Res.*, 110, D24304, doi:
9 10.1029/2005JD006157, 2005.

10 Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J.
11 B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G.
12 R., Randerson, J. T., Wennberg, P. O., Krol, M. C., Tans, P. P.: An atmospheric perspective
13 on North American carbon dioxide exchange: CarbonTracker, *P. Natl. Acad. Sci. USA*, 104,
14 18925–18930, 2007.

15 Peters, W., KROL, M. C., Van Der WERF, G. R., Houwling, S., Jones, C. D., Hughes, J.,
16 Schaefer, K., Masarie, K. A., Jacobson, A. R., Miller, J. B., Cho, C. H., Ramonet, M.,
17 Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., Di Sarra, A. G.,
18 Piachentina, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H. A. J.,
19 Van Der Laan, S., Neubert, R. E. M., Jordan, A., Rodo, X., Morgui, J. –A., Vermeulen, A. T.,
20 Popa, E., Rozanski, M., Manning, A. C., Leuenberger, M., Uglietti, C., Dolman, A. J., Ciais,
21 P., Heimann, M., Tans, P. P.: Seven years of recent European net terrestrial carbon dioxide
22 exchange constrained by atmospheric observations, *Global Change Biology*, 16(4),
23 1365-2486, 2009.

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

28 Pillai, D., Gerbig, C., Ahmadov, R., Rödenbeck, C., Kretschmer, R., Koch, T., Thompson, R.,
29 Neininger, B., and Lavrić J. V.: High-resolution simulations of atmospheric CO₂ over
30 complex terrain – representing the Ochsenkopf mountain tall tower, *Atmos. Chem. Phys.*, 11,

1 7445-7464, doi:10.5194/acp-11-7445-2011, 2011.

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

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

7 Tian, X., Xie, Z., Sun, Q.: A POD-based ensemble four-dimensional variational assimilation
8 method, *Tellus A*, 63, 805-816, 2011.

9 Tian, X., Xie, Z., Liu, Y., Cai, Z., Fu, Y., Zhang, H., and Feng, L.: A joint data assimilation
10 system (Tan-Tracker) to simultaneously estimate surface CO₂ fluxes and 3-D atmospheric
11 CO₂ concentrations from observations, *Atmos. Chem. Phys. Discuss.*, 13, 24755-24784,
12 2013.

13 van der Werf, G. R., Randerson, J. T.: Global fire emissions and the contribution of deforestation,
14 savanna, forest, agricultural, and peat fires (1997–2009). *Atmos. Chem. Phys.*, 10,
15 11707–11735, doi: 10.5194/acp-10-11707-2010, 2010.

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

19 Zhang, H. F., B. Z. Chen, I. T. van der Laan-Luijkx, J. Chen, G. Xu, J. W. Yan, L. X. Zhou, Y.
20 Fukuyama, P. P. Tans, and W. Peters, Net terrestrial CO₂ exchange over China during
21 2001–2010 estimated with an ensemble data assimilation system for atmospheric CO₂, *J.*
22 *Geophys. Res. Atmos.*, 119, 3500–3515, doi:10.1002/2013JD021297, 2014.

23 Zhang, H. F., Chen, B. Z., Machida, T., Matsueda, H., Sawa, Y., Fukuyama, Y., Langenfelds, R.,
24 van der Schoot, M., Xu, G., Yan, J. W., Cheng, M. L., Zhou, L. X., Tans, P. P., and Peters, W.:
25 Estimating Asian terrestrial carbon fluxes from CONTRAIL aircraft and surface CO₂
26 observations for the period 2006–2010, *Atmos. Chem. Phys.*, 14, 5807-5824,
27 doi:10.5194/acp-14-5807-2014, 2014.

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

1 Zhang, M., Uno, I., Carmichael, G. R., Akimoto, H., Wang, Z., Tang, Y., Woo, J., Streets, D. G.,
2 Sachse, G. W., Avery, M. A., Weber, R. J., Talbot, R. W.: Large-scale structure of trace gas
3 and aerosol distributions over the western Pacific Ocean during the Transport and Chemical
4 Evolution Over the Pacific (TRACE-P) experiment, *J. Geophys. Res.*, 108(D21), 8820, doi:
5 10.1029/2002JD002946, 2003.

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

8

9

10

11

1 **List of Figures**

2 Fig. 1. Schematic diagram of the smoother window.

3 $(\lambda_{i,t-1-M|t-1}^a, \lambda_{i,t-M|t-1}^a, \lambda_{i,t-M+1|t-1}^a, \dots, \lambda_{i,j|t-1}^a, \dots, \lambda_{i,t-1|t-1}^a)$ are the optimized scaling factors in
4 the smoother window and $C_{i,t-1}^a$ are the assimilated CO₂ concentrations fields at time

5 $t-1$ in the previous assimilation cycle $t-1-M \sim t-1$.

6 $(\lambda_{i,t-M|t-1}^a, \lambda_{i,t-M+1|t-1}^a, \dots, \lambda_{i,j|t-1}^a, \dots, \lambda_{i,t-1|t-1}^a, \lambda_{i,t|t-1}^a)$ are the scaling factors in the smoother
7 window and $C_{i,t}^f$ are the forecast CO₂ concentrations fields at time t which need to

8 be optimized in the current assimilation cycle $t-M \sim t$.

9

10 Fig. 2. Flowchart of the CFI-CMAQ system used to optimize surface CO₂ fluxes at

11 each assimilation cycle. The system includes the following four parts in turn: (1)

12 forecasting of the linear scaling factors $\lambda_{i,t|t-1}^a$ (red arrows); (2) optimization of the

13 scaling factors in the smoother window by EnKS (see Fig. 1) (blue arrows); (3)

14 updating of the flux in the smoother window (green arrows); and (4) assimilation of

15 the CO₂ concentration fields at time t by EnKF (black arrows).

16

17 Fig. 3. (a) Total number of observations in February 2010 in the model grid. Each

18 symbol indicates the total number of all GOSAT X_{CO2} measurements in the

19 corresponding model grid. Monthly mean values in February 2010 of (b) $X_{CO_2}^p$,

20 column mixing ratio of C_i^p ; (c) $X_{CO_2}^f$, column mixing ratio of C_i^f ; (d) $\overline{X_{CO_2}^a}$,

21 column mixing ratio of $\overline{C_i^a}$; (e) $X_{CO_2}^p - X_{CO_2}^f$; and (f) $X_{CO_2}^p - \overline{X_{CO_2}^a}$. All column

22 mixing ratios are column-averaged with real GOSAT X_{CO2} averaging kernels at

23 GOSAT X_{CO2} locations. Each symbol indicates the monthly average value of all X_{CO2}

1 estimates in the model grid. \overline{C}_t^a are the ensemble mean values of the assimilated
2 CO₂ concentrations fields of a CFI-CMAQ OSSE, in which the lag-window was 9
3 days and β was 70. And they are the same OSSE in Fig. 3 to Fig. 6.

4

5 Fig. 4. Monthly mean values of (a) C_t^p , the artificial true simulations driven by the
6 prior surface CO₂ fluxes F_t^p ; (b) C_t^f , the background simulations driven by
7 magnified surface CO₂ fluxes $F_t^* = (1.8 + \delta(x, y, z, t))F_t^p$; (c) \overline{C}_t^a , the ensemble
8 mean values of the assimilated CO₂ concentrations fields; (d) $C_t^p - C_t^f$; (e) $C_t^p - \overline{C}_t^a$;
9 and (f) $100 * (C_t^p - \overline{C}_t^a) / C_t^p$ at model-level 1 in February 2010. Black lines EF and
10 GH indicate the positions of the cross sections shown in Fig. 5.

11

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

15

16 Fig. 6. Daily mean time series of CO₂ concentrations at national background stations
17 in China and their nearest large cities from 1 Jan. to 20 Mar. 2010 extracted from the
18 artificial true simulations C_t^p (black), background simulations C_t^f (red), and the
19 ensemble mean values of the assimilated CO₂ concentrations fields \overline{C}_t^a (blue). All
20 time series were interpolated to the observation locations by the spatial bilinear
21 interpolator method. The sites used are (a) Waliguan (36.28 N, 100.91 E), (b) Xining
22 (36.56 N, 101.74 E), (c) Longfengshan (44.73 N, 127.6 E), (d) Haerbin (45.75 N,
23 126.63 E), (e) Shangdianzi (40.65 N, 117.12 E), (f) Beijing (39.92 N, 116.46 E), (g)

1 Linan (30.3 N, 119.73 E), and (h) Hangzhou (30.3 N,120.2 E).

2

3 Fig. 7. Monthly mean values in February 2010 of (a) F_t^p , the prior true surface CO₂
4 fluxes; (b) F_t^* , the prescribed CO₂ surface fluxes, $F_t^* = (1.8 + \delta(x, y, z, t))F_t^p$; (c)
5 $\overline{F_t^a}$, the ensemble mean values of the assimilated surface CO₂ fluxes; (d) $F_t^p - F_t^*$;
6 and (e) $F_t^p - \overline{F_t^a}$ (units: $\mu\text{mole m}^{-2} \text{ s}^{-1}$). $\overline{F_t^a}$ are the assimilated results of an
7 CFI-CMAQ OSSE, in which the lag-window was 9 days and β was 70. And they
8 are the same in Fig. 7 to Fig. 10.

9

10 Fig. 8. Monthly mean RMSEs of $\overline{F_t^a}$ in February 2010 (units: $\mu\text{mole m}^{-2} \text{ s}^{-1}$).

11

12 Fig. 9. (a) Ratios of monthly mean F_t^* to monthly mean F_t^p ; and (b) ratios of
13 monthly mean $\overline{F_t^a}$ to monthly mean F_t^p in Feb. 2010. The white part indicates the
14 ratios where the absolute values of monthly mean F_t^p are larger than 0.1, not
15 analyzed in this study. The black square labeled I indicates the domain where surface
16 CO₂ fluxes were used for the results presented in Fig. 12 and 13.

17

18 Fig. 10. Daily mean time series of CO₂ fluxes at national background stations in
19 China and their nearest large cities from 1 Jan to 20 Mar. 2010 extracted from the
20 prior true surface CO₂ fluxes F_t^p (black), the prescribed CO₂ surface fluxes F_t^*
21 (red), and the assimilated CO₂ fluxes $\overline{F_t^a}$ (blue). All time series were interpolated to
22 the observation locations by the spatial bilinear interpolator method. The sites used
23 are (a) Waliguan, (b) Xining, (c) Longfengshan, (d) Haerbin, (e) Shangdianzi, (f)

1 Beijing, (g) Linan, and (h) Hangzhou.

2

3 Fig. 11. (a) Ensemble spread of $C_{i,t}^f$ after inflating; (b) ensemble spread of $\lambda_{i,t|t-1}^p$
4 before inflating; (c) ensemble spread of $\lambda_{i,t|t-1}^a$ at model-level 1 at 00 UT on 1 March
5 2010 when $\beta = 70$.

6

7 Fig. 12. Time series of daily mean CO₂ fluxes averaged in domain I (shown in Fig. 9a)
8 from 1 Jan. to 20 Mar. 2010 with the inflation factor of scaling factors $\beta = 50, 60, 70,$
9 75 and 80 . The black dashed line is the time series averaged from F_t^* and the black
10 solid line is the time series averaged from F_t^p .

11

12 Fig. 13. Time series of daily mean CO₂ fluxes averaged in domain I (shown in Fig. 9a)
13 from 1 Jan. to 20 Mar 2010 with different smoother windows (3, 6, 9 and 12 days).
14 The black dashed line is the time series averaged from F_t^* and the black solid line is
15 the time series averaged from F_t^p .

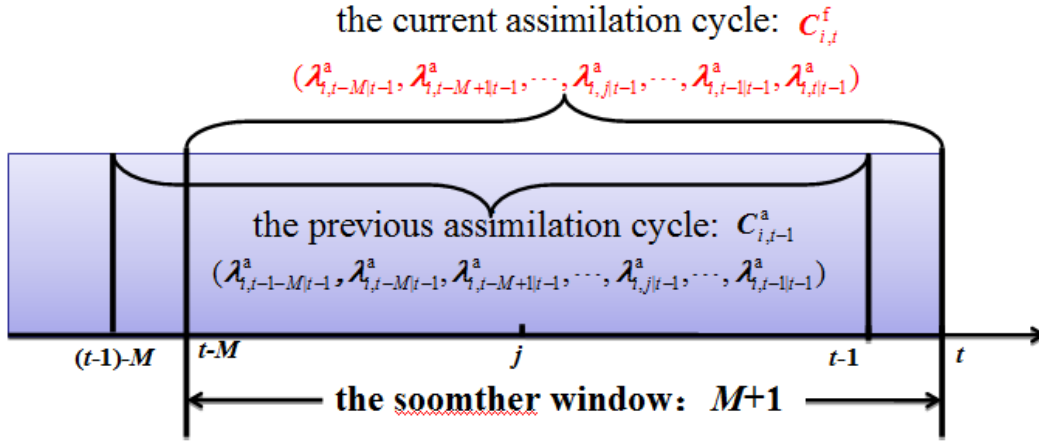
16

17 Fig. 14. Monthly mean values of the difference between the prior true surface CO₂
18 fluxes and the ensemble mean values of the assimilated surface CO₂ fluxes (units:
19 $\mu\text{mole m}^{-2} \text{s}^{-1}$) of the reference experiment in which $\lambda_{i,t|t-1}^p$ were set to 1.

20

21

1



2

3

4

5

Fig. 1. Schematic diagram of the smoother window.

6

$(\lambda_{i,t-1-M|t-1}^a, \lambda_{i,t-M|t-1}^a, \lambda_{i,t-M+1|t-1}^a, \dots, \lambda_{i,j|t-1}^a, \dots, \lambda_{i,t-1|t-1}^a)$ are the optimized scaling factors in the

7

smoother window and $C_{i,t-1}^a$ are the assimilated CO₂ concentrations fields at time $t-1$ in the

8

previous assimilation cycle $t-1-M \sim t-1$. $(\lambda_{i,t-M|t-1}^a, \lambda_{i,t-M+1|t-1}^a, \dots, \lambda_{i,j|t-1}^a, \dots, \lambda_{i,t-1|t-1}^a, \lambda_{i,t|t-1}^a)$ are

9

the scaling factors in the smoother window and $C_{i,t}^f$ are the forecast CO₂ concentrations fields at

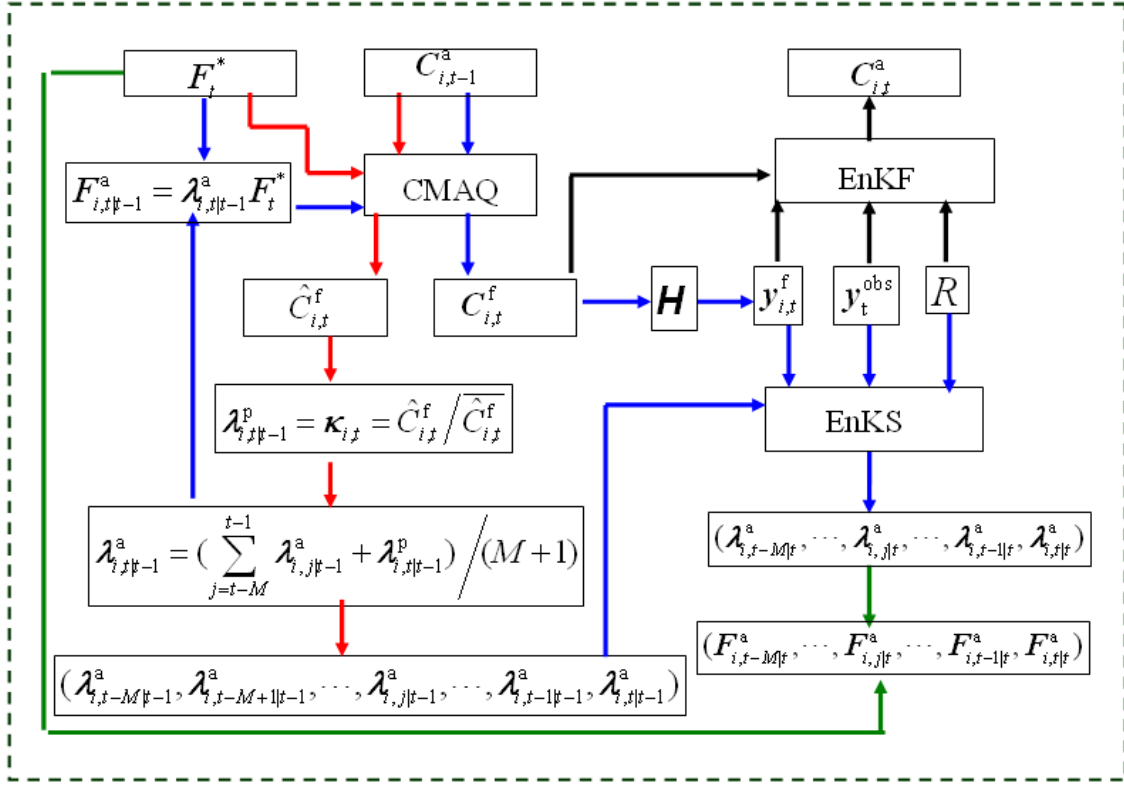
10

time t which need to be optimized in the current assimilation cycle $t-M \sim t$.

11

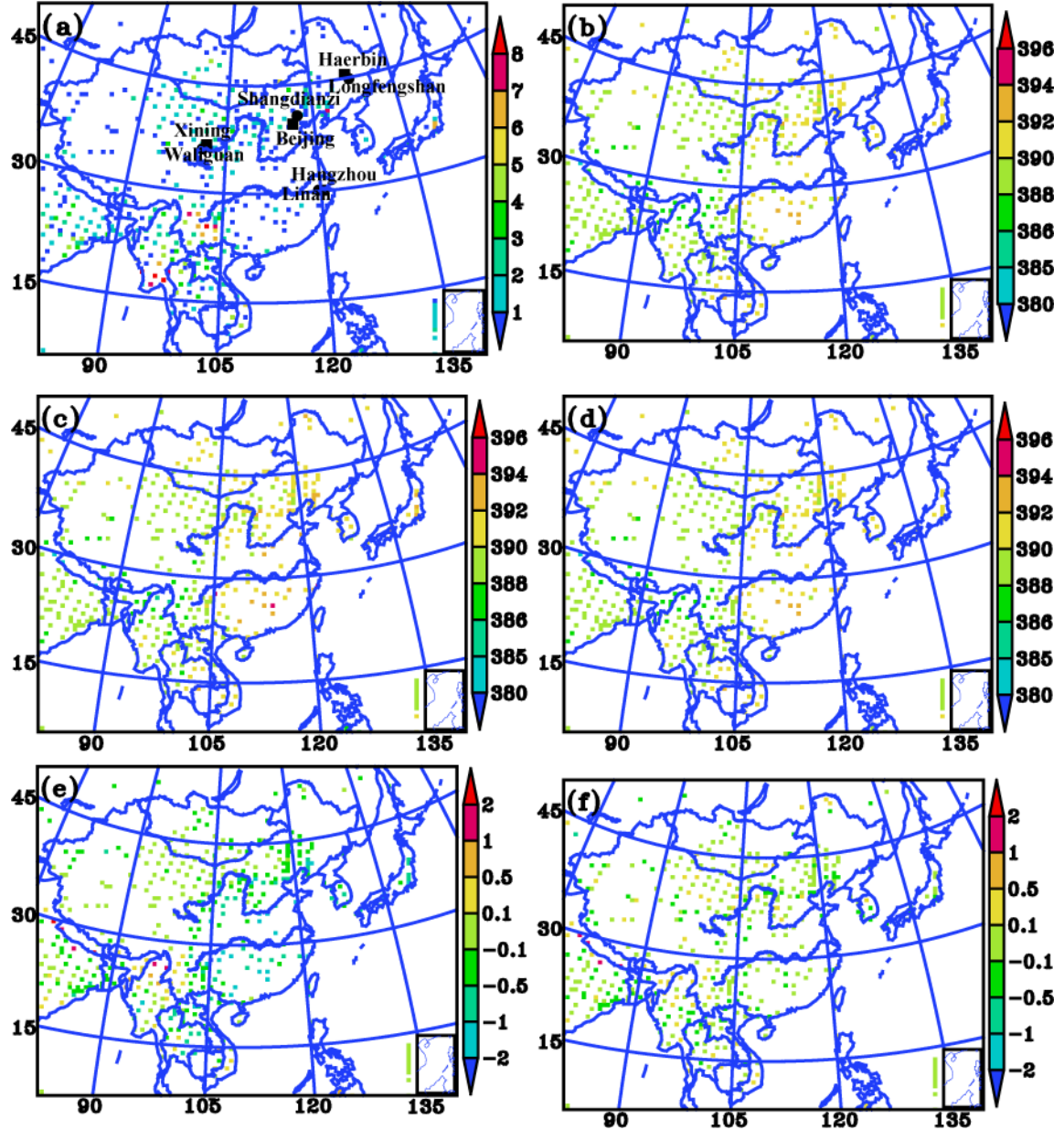
12

1



2

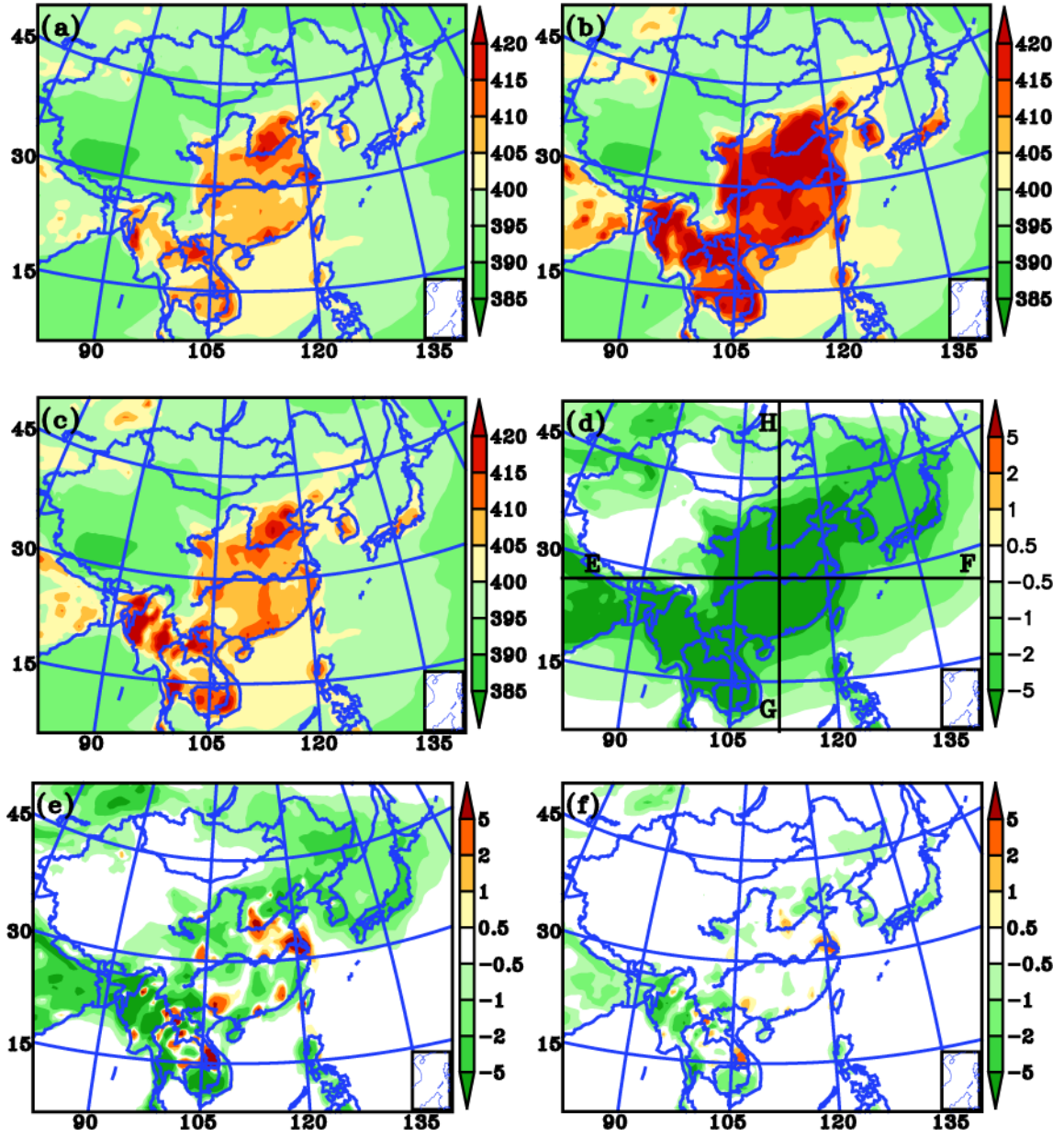
3 Fig. 2. Flowchart of the CFI-CMAQ system used to optimize surface CO₂ fluxes at each
 4 assimilation cycle. The system includes the following four parts in turn: (1) forecasting of the
 5 linear scaling factors $\lambda_{i,t|t-1}^a$ (red arrows); (2) optimization of the scaling factors in the smoother
 6 window by EnKS (see Fig. 1) (blue arrows); (3) updating of the flux in the smoother window
 7 (green arrows); and (4) assimilation of the CO₂ concentration fields at time t by EnKF (black
 8 arrows).
 9



1
2 Fig. 3. (a) Total number of observations in February 2010 in the model grid. Each symbol
3 indicates the total number of all GOSAT X_{CO_2} measurements in the corresponding model grid.
4 Monthly mean values in February 2010 of (b) $X_{CO_2}^p$, column mixing ratio of C_t^p ; (c) $X_{CO_2}^f$,
5 column mixing ratio of C_t^f ; (d) $\overline{X_{CO_2}^a}$, column mixing ratio of $\overline{C_t^a}$; (e) $X_{CO_2}^p - X_{CO_2}^f$; and (f)
6 $X_{CO_2}^p - \overline{X_{CO_2}^a}$. All column mixing ratios are column-averaged with real GOSAT X_{CO_2} averaging
7 kernels at GOSAT X_{CO_2} locations. Each symbol indicates the monthly average value of all X_{CO_2}
8 estimates in the model grid. $\overline{C_t^a}$ are the ensemble mean values of the assimilated CO_2
9 concentrations fields of a CFI-CMAQ OSSE, in which the lag-window was 9 days and
10 β was 70. And they are the same OSSE in Fig. 3 to Fig. 6.

11

1



2

3 Fig. 4. Monthly mean values of (a) C_t^p , the artificial true simulations driven by the prior surface

4 CO₂ fluxes F_t^p ; (b) C_t^f , the background simulations driven by magnified surface CO₂ fluxes

5 $F_t^* = (1.8 + \delta(x, y, z, t))F_t^p$; (c) $\overline{C_t^a}$, the ensemble mean values of the assimilated CO₂

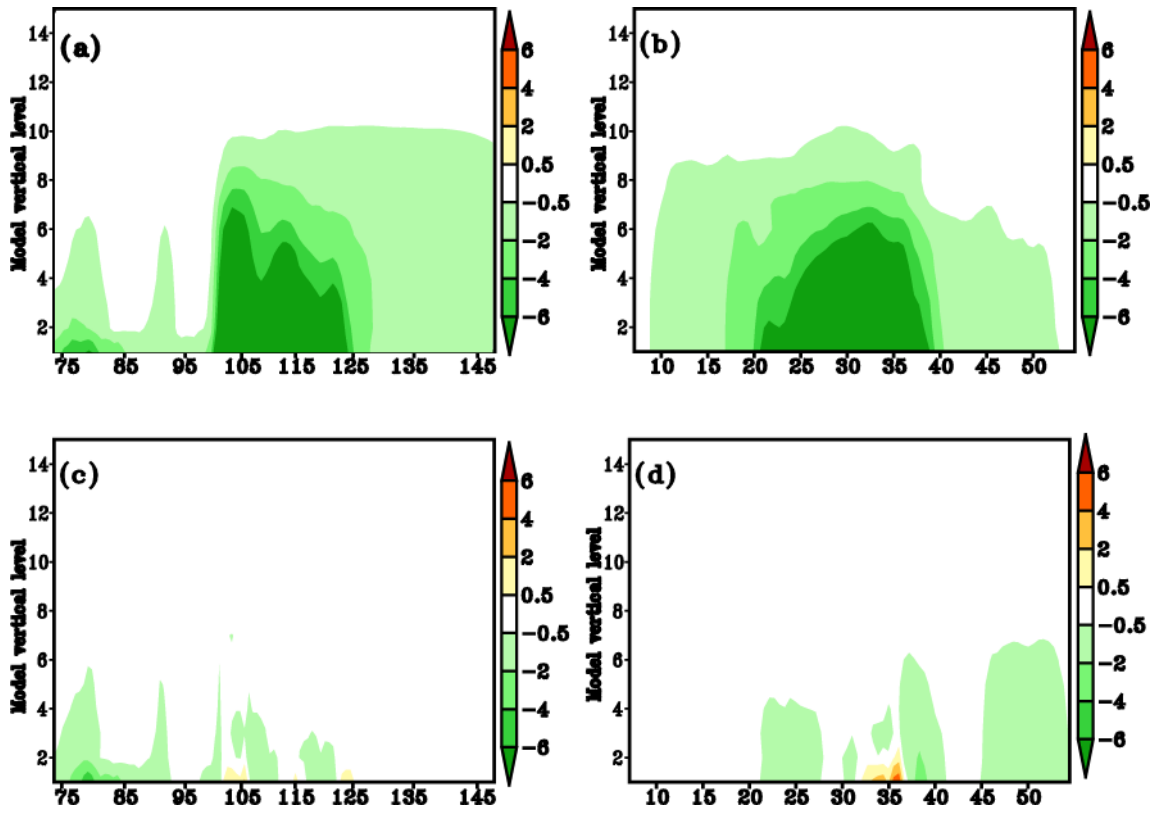
6 concentrations fields; (d) $C_t^p - C_t^f$; (e) $C_t^p - \overline{C_t^a}$; and (f) $100 * (C_t^p - \overline{C_t^a}) / C_t^p$ at

7 model-level 1 in February 2010. Black lines EF and GH indicate the positions of the cross sections

8 shown in Fig. 5.

9

1



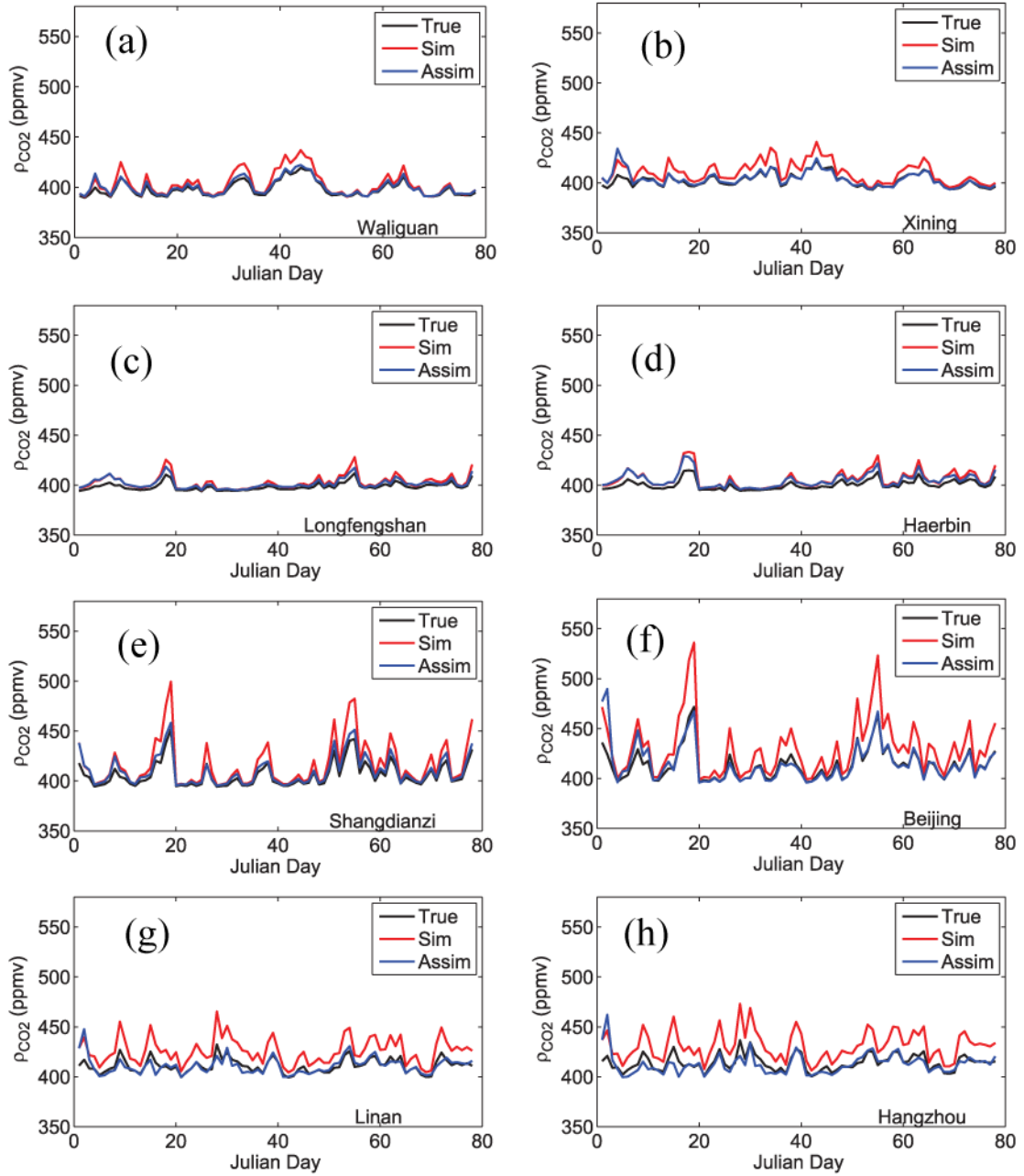
2

3 Fig. 5. Monthly mean cross sections of $C_t^p - C_t^f$ along line (a) EF and (b) GH, and monthly

4 mean cross sections of $C_t^p - \overline{C_t^a}$ along line (c) EF and (d) GH (cross section lines shown in Fig.

5 4d) in February 2010.

6



1

2 Fig. 6. Daily mean time series of CO₂ concentrations at national background stations in China and
 3 their nearest large cities from 1 Jan. to 20 Mar. 2010 extracted from the artificial true simulations

4 C_t^p (black), background simulations C_t^f (red), and the ensemble mean values of the
 5 assimilated CO₂ concentrations fields $\overline{C_t^a}$ (blue). All time series were interpolated to the

6 observation locations by the spatial bilinear interpolator method. The sites used are (a) Waliguan

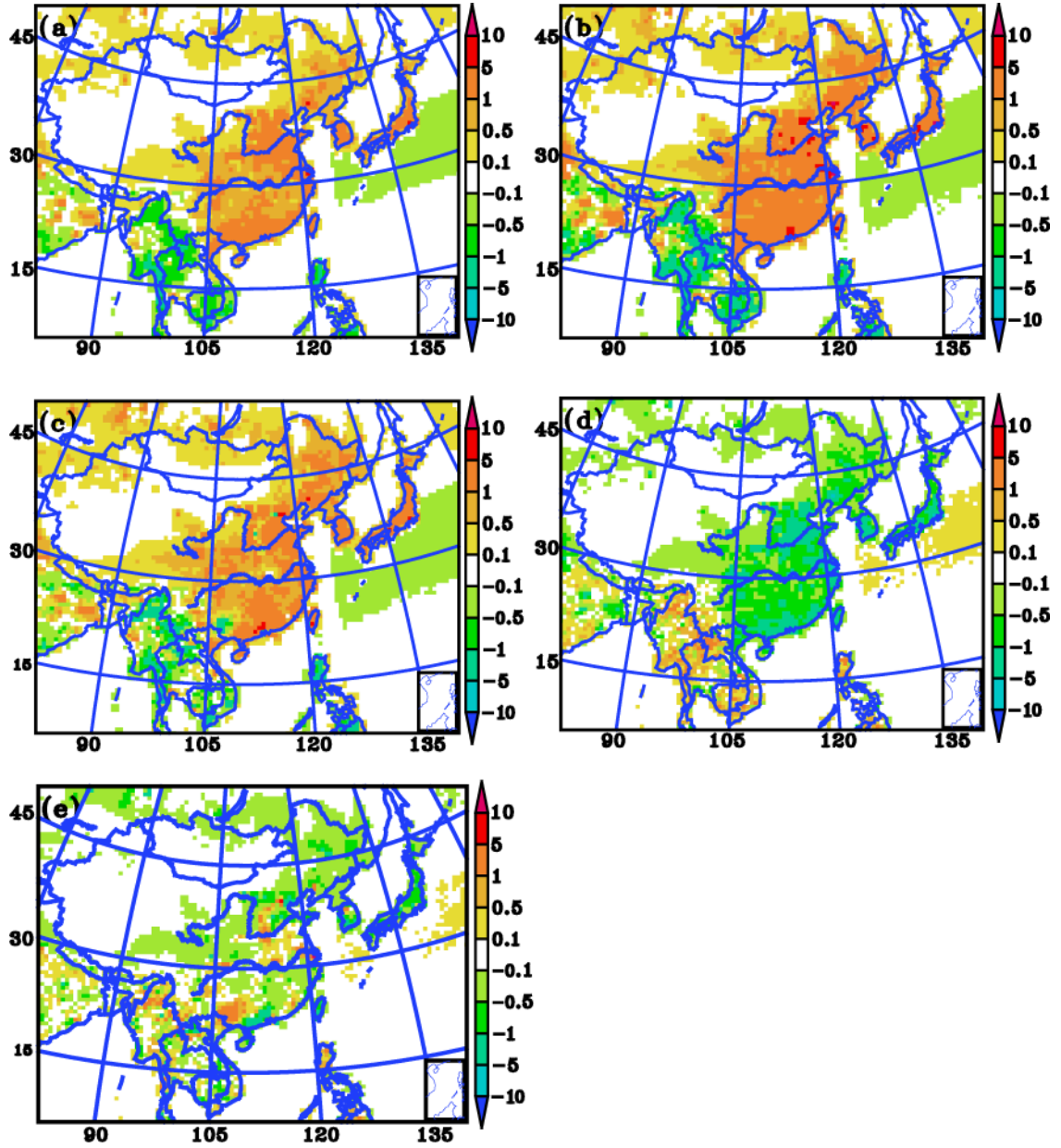
7 (36.28 N, 100.91 E), (b) Xining (36.56 N, 101.74 E), (c) Longfengshan (44.73 N, 127.6 E), (d)

8 Haerbin (45.75 N, 126.63 E), (e) Shangdianzi (40.65 N, 117.12 E), (f) Beijing (39.92 N,

9 116.46 E), (g) Linan (30.3 N, 119.73 E), and (h) Hangzhou (30.3 N, 120.2 E).

10

1



2

3 Fig. 7. Monthly mean values in February 2010 of (a) F_t^p , the prior true surface CO₂ fluxes; (b)

4 F_t^* , the prescribed CO₂ surface fluxes, $F_t^* = (1.8 + \delta(x, y, z, t))F_t^p$; (c) $\overline{F_t^a}$, the

5 ensemble mean values of the assimilated surface CO₂ fluxes; (d) $F_t^p - F_t^*$; and (e)

6 $F_t^p - \overline{F_t^a}$ (units: $\mu\text{mole m}^{-2} \text{s}^{-1}$). $\overline{F_t^a}$ are the assimilated results of an CFI-CMAQ OSSE, in

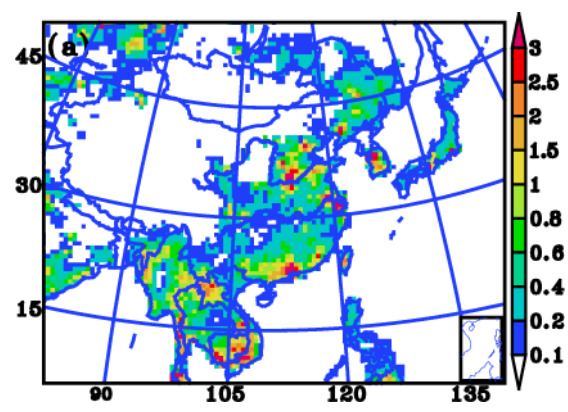
7 which the lag-window was 9 days and β was 70. And they are the same in Fig. 7 to

8

Fig. 10.

9

1



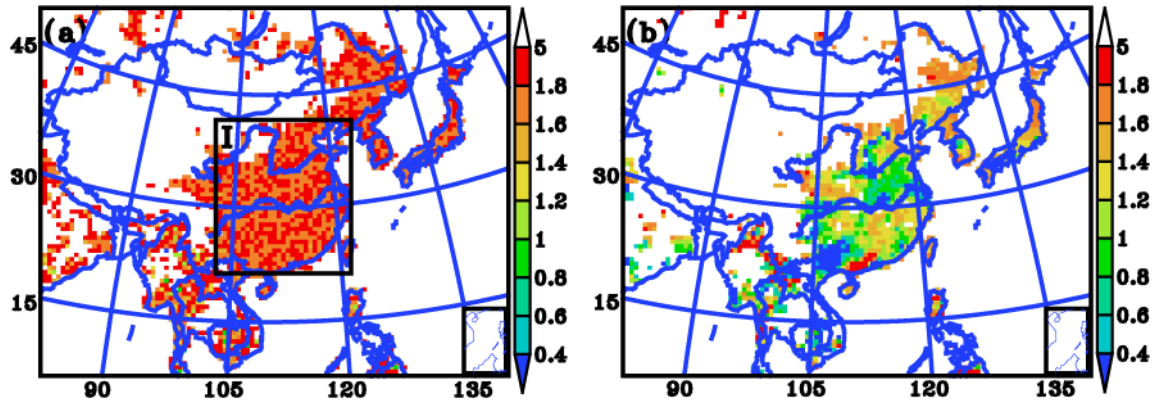
2

3

4 Fig. 8. Monthly mean RMSEs of $\overline{F_t^a}$ in February 2010 (units: $\mu\text{mole m}^{-2} \text{s}^{-1}$).

5

6



1

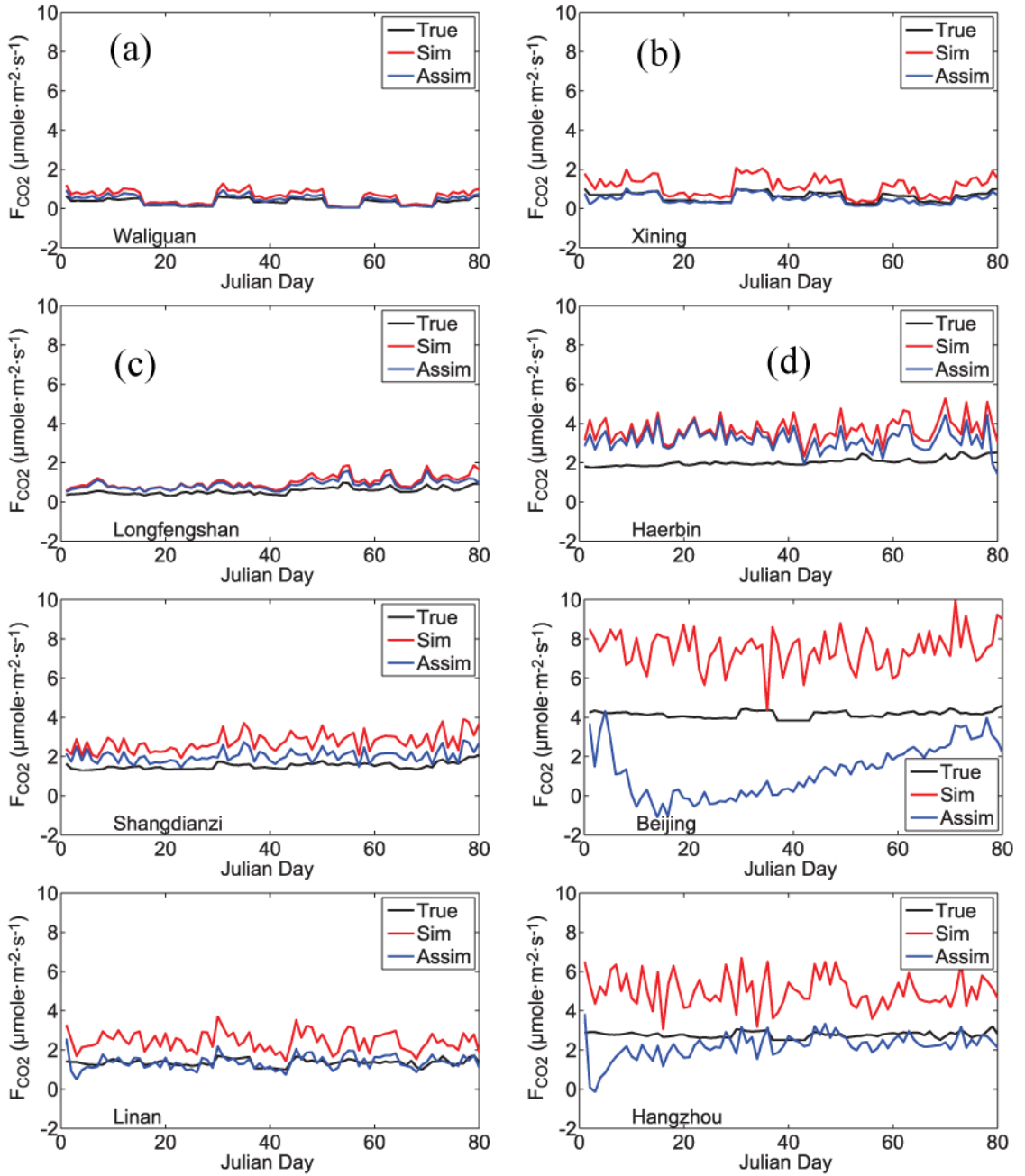
2 Fig. 9. (a) Ratios of monthly mean F_t^* to monthly mean F_t^P ; and (b) ratios of monthly mean

3 $\overline{F_t^a}$ to monthly mean F_t^P in Feb. 2010. The white part indicates the ratios where the absolute

4 values of monthly mean F_t^P are larger than 0.1, not analyzed in this study. The black square

5 labeled I indicates the domain where surface CO₂ fluxes were used for the results presented in Fig.

6 12 and 13.



1
2 Fig. 10. Daily mean time series of CO₂ fluxes at national background stations in China and their
3 nearest large cities from 1 Jan to 20 Mar. 2010 extracted from the prior true surface CO₂ fluxes
4 F_t^p (black), the prescribed CO₂ surface fluxes F_t^* (red), and the assimilated CO₂ fluxes $\overline{F_t^a}$
5 (blue). All time series were interpolated to the observation locations by the spatial bilinear
6 interpolator method. The sites used are (a) Waliguan, (b) Xining, (c) Longfengshan, (d) Haerbin,
7 (e) Shangdianzi, (f) Beijing, (g) Linan, and (h) Hangzhou.

1

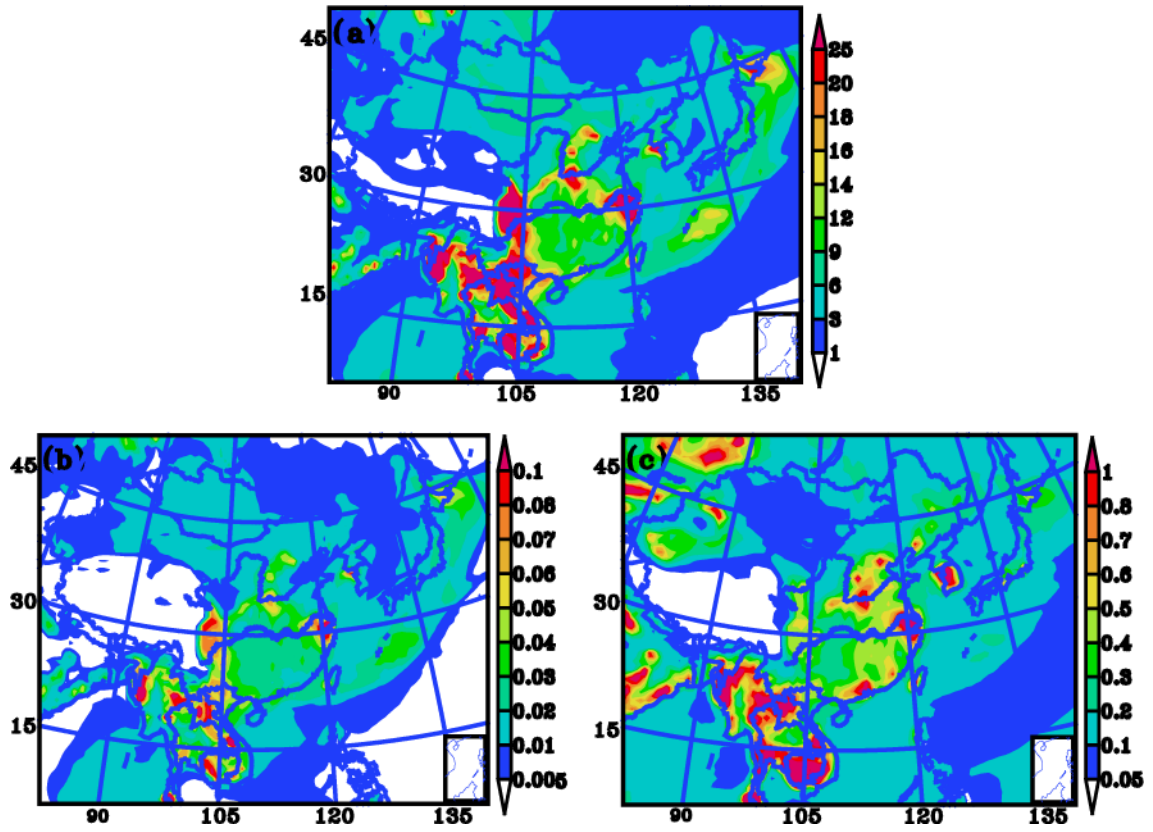
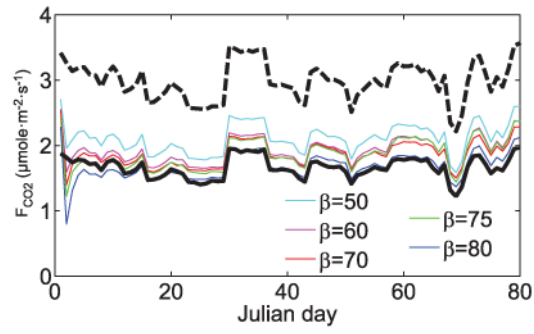


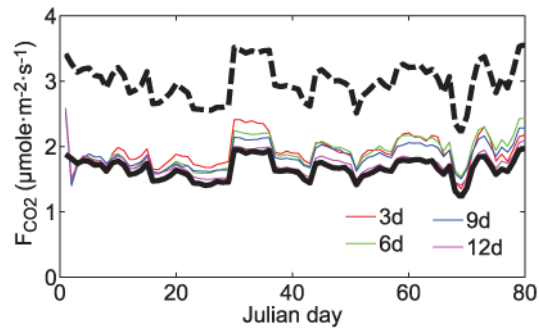
Fig. 11. (a) Ensemble spread of $C_{i,t}^f$ after inflating; (b) ensemble spread of $\lambda_{i,t|t-1}^p$ before inflating; (c) ensemble spread of $\lambda_{i,t|t-1}^a$ at model-level 1 at 00 UT on 1 March 2010 when $\beta = 70$.



1
 2 Fig. 12. Time series of daily mean CO₂ fluxes averaged in domain I (shown in Fig. 9a) from 1 Jan.
 3 to 20 Mar. 2010 with the inflation factor of scaling factors $\beta = 50, 60, 70, 75$ and 80 . The black
 4 dashed line is the time series averaged from \mathbf{F}_t^* and the black solid line is the time series
 5 averaged from \mathbf{F}_t^p .

6
 7

1



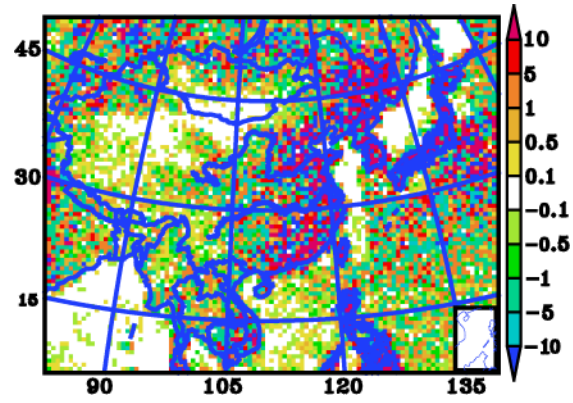
2

3 Fig. 13. Time series of daily mean CO_2 fluxes averaged in domain I (shown in Fig. 9a) from 1 Jan.
4 to 20 Mar 2010 with different smoother windows (3, 6, 9 and 12 days). The black dashed line is

5 the time series averaged from F_t^* and the black solid line is the time series averaged from F_t^P .

6

1



2

3 Fig. 14. Monthly mean values of the difference between the prior true surface CO₂ fluxes and the

4 ensemble mean values of the assimilated surface CO₂ fluxes (units: $\mu\text{mole m}^{-2} \text{s}^{-1}$) of the reference

5 experiment in which $\lambda_{i,t|t-1}^p$ were set to 1.

6