Response to reviewer:

Review2:

[General comments]

1 This reviewer doubts whether the manuscript is the final version that authors have revised, because the current manuscript has not been updated completely although authors replied that it was fixed or revised (For example, the localization scale is not fixed in the manuscript yet, and any additional brief description of OCO data aggregation cannot be found).

We are sorry for this oversight. The authors made sure that the current version is the most updated one with all revisions incorporated. For the specific comment on OCO data aggression, we added the following description (Page 7, Lines 26- Page 8 Line 7): :

" The OCO-2 retrievals used to obtain averages are based on the NASA Atmospheric CO2 Observations from Space XCO2 retrieval Algorithm version7r (O'Dell et al., 2012), as archived at https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_7r/summary (last access: 23 March 2017).
A two-step averaging method has been used in order to avoid the final average to be disproportionately weighted to one part of the averaging bin (track) with more good quality retrievals. In the first step, the "good quality" retrievals defined as WL<=15 and xco2_quality_flag=0 (another quality indicator of the data) are averaged over 1-second bins, with weights inversely proportional to the square of each retrievals posterior uncertainty. In the second step, all the 1-second bins, with at least one valid retrieval, are averaged over a 10-second interval to create 10-second averaged data. The OCO-2 averaging kernels are similarly averaged to create 10-second mean averaging kernels. This averaging method had been used for similar purpose in the recent study by Basu et al. (2018)."

2. This reviewer strongly suggests authors to check if the uploaded manuscript was the final one. If you add or modify some, it would be good to point out where you made changes (page and lines) in the author's response.

We apologize for the inconvenience caused. A copy of the revised manuscript with highlighted changes is also attached with this revision. As suggested, Page/Line numbers are also included in the reply to reviewers comments.

3. In addition, this reviewer mentioned that it needs to assimilate "additional" observation dataset such as GV+ and GOSAT to see whether long OW is necessary when there are "more" observations to constrain the surface carbon fluxes. However, authors seem to provide the results from the experiment assimilating only GV+. This is somewhat distant from what this reviewer asked authors to prove. More observational data, GV+ in addition to OCO, may give somewhat different sensitivity of the length of OW. If Figure 10 results from the experiment assimilating both GV+ and OCO, please revise the context appropriately. If Figure 10 results from the experiment assimilating only GV+, this reviewer strongly suggests authors to try the corresponding experiment again. This reviewer agrees that it is worth finding that assimilating each datasets required consistent length of OW (8d-15d) though. Before the publication, authors need to revise the manuscript carefully again.

We thank the reviewer for this extremely valuable suggestion. We have updated the experiments by assimilation both GV+ and OCO-2 data. We have updated Figure 10 and corresponding discussion. The discussion in Page19 Line 30 to Page 20 Line 22 is provided below.

" Our working hypothesis was that the optimal OW for the estimation of SCF could be reduced with more observations. We examined this hypothesis by using simulated OCO-2 observations and Global View Plus (GV+) observations. Similar to the OCO-2 pseudo-

observation, the GV+ pseudo-observations were also generated based on the actual location, time and corresponding error scale of the GV+ flask observations. The results show that the AW/OW lengths of 1day /8 day is also optimal with both the OCO-2 and GV+ observation characteristics. We estimated the SCF using the OCO-2 and GV+ pseudo-observations with the identical experiment settings as the OCO-2 experiments, except that the experiment with very long OW of 30 days is replaced with an experiment with a short OW of 4 days to better evaluate the impact from short OWs. Thus the current experiment sets are using OW of 2, 4, 8, 15 days.

The results from these experiments show that the AW/OW lengths of 1 day /8 day is still optimal for both the OCO-2 and GV+ observation characteristics (Figure 10). Generally, the time-mean RMSE of estimated SCF with OCO-2 and GV+ (Figure 10) are smaller than the corresponding estimates for OCO-2 only (Figure 5). The short OW of 2 days performs worse than the moderate OWs of 4 days, 8 days and 15 days. The time-averaged global mean RMSE is 0.046 $\frac{kgC}{m^2 yr}$ *for experiments with OW of 2 days (Figure 10a). The time-averaged global mean RMSE is only 0.040, 0.037 and 0.039 for experiments with OW of 4 days, 8 days and 30 days, respectively (Figure10 b, c and d). We see only a slight impact of observation the optimal OW length by assimilating additional GV+ observations. The best OW appears to be 8~15 days which produce the smallest RMSE when only OCO-2 observations only are assimilated. The smallest RMSE in the experiment is obtained in the experiment with OW of 8 days, when both OCO-2 and GV+ observations are assimilated into the system."*

[Specific comments]

1. p. 7, line 1: Why did the authors drop the citation of Liu et al. (2017) from the previous manuscript?

We thank the reviewer for pointing this out. The paper by Liu et al. (2017), is on hold because the first author took up job at a new location and due to his job commitments/obligations there. This is the reason for removing it from the reference list.

2. p. 10, line 22: Typo of localization scale (150km) has not been corrected yet, although it is important experimental setting.

We apologize for this oversight, it is corrected now. Page 11/Line 3

3. Figures 3, 5, and 10: It needs to address what period authors made temporal average for. It seems to me that authors may drop the spinup period from Jan. to the end of Feb., 2015. Please clarify it.

We thank the reviewer for this comment. The spin-up period is from Jan. to the end of Feb, 2015. We have now clearly described the period of simulation on Page 12 /Line 23-24 and Figure 3 caption Page 29.

"The average period is from March 1 2015 to the end of February 2016."

" Figure 3 The spatial pattern of the annual mean RMSE of estimated SCF from the experiments with different AW (EXP1-4) for the average period from 1 March 2015 to the end of February 2016. (January and February 2015 are treated as spinup period for our experiments)."

4. p.13, lines 19-22: The statement leads to a question, what would happen if the experiments start from APR. It would be great to add discussion about it, although authors do not need to show those results.

We thank the reviewer for this comment. We did some testing on this concern. The related discussion is added in the text: Page Page 15/Line 6-11

"The same results can be obtained from the same experiments with different initial time, indicating the robustness of our findings [figure not shown]. The convergence of estimated SCFs from the experiments starting from months with big SCF variation, such as April, is slightly slower than the experiments from the time with small SCF variation, such as January. While the estimated SCFs converges in a few analysis cycles (a few days) in our system (Figure 2), the small difference of convergence does not make any significant impact on the quality of estimated SCFs. Moreover, the calculation of RMSE of estimated SCFs has excluded the spinup period of first two months to remove the potential impact of initial condition and initial time. "

[Technical corrections]
1. p. 7, line 1: unbalanced parenthesis.

We really appreciate your careful reading of the manuscript. We have corrected this in the revised manuscript.

2. p. 8, line 31: unbalanced parenthesis

We thank the reviewer for spotting this. We have corrected this in the revised manuscript.
Now at Page 9/Line 12

3. p. 9, line 3: unbalanced parenthesis

We really appreciate your careful examination of the manuscript. We have corrected this in the revised manuscript.
Now at Page 9/Line 15

4. p. 16, lines 4-7: Text miscalls experiment #. EXP7 should be changed to EXP6, and EXP8 should be changed to EXP7. Please check them carefully again.

We thank the reviewer for finding this typo. We have corrected these in the revised manuscript.
Now at Page 17/Line 1-3

5. Figure 7: Color bars need to be rescaled to analyze figures better. For example, authors need to include large concentration better for the first four figures. Current color bar does not show what is going on over 409 ppmv (authors should add the unit "ppmv" in the figure too) although Figure 7f shows a difference greater than 5 ppmv. Due to the current color bar, Figure 7a and Figure 7c look almost identical (Figure 7b and Figure 7d, too). Please rescale the color bar, or add more colors to specify more levels of CO_2 concentration.

Thank you very much for your valuable suggestions. We have made these changes in the revised manuscript.

6. p. 16, line 31 – p. 17, lines 1-2: Please rephrase the sentence to clarify what authors would like to point out precisely.

Thank you very much for your suggestion. We changed this sentence as (Page 17/Line 28-29)
" *The analysis of CO_2 concentrations matches the "nature" run well. The error pattern also matches the CO_2 seasonal cycle and the error pattern of estimated SCF*"

7. p. 18, lines 5-6: Current sentence is not a grammatically complete. Please rephrase it. The sentence in the previous manuscript was better.

Thank you very much for your advice. We have made this correction for the sentence to read as (Page 19/Line 1-2)
"*In summary, we found that the OSSE experiments using long observation windows and short assimilation windows resulted in the best estimates of SCF*"

8. Figure 10: Resolution or shading style of the Figure 10 looks different from Figure 5. Is there any reason? If not, please be consistent.

Thank you very much. We have updated the experiments and figure 10. Now it is consistent.

9. p. 20, lines 4-5: Please clarify what you mean by "hybrid ensemble data assimilation system", since there are so many similar systems recently.

We thank the reviewer for this suggestions. We have made the following change in the text as (Page 21/Line 9-16)
" *The 4D-Var and the 4D-LETKF have been shown (Bonavita et al. 2015; Hamrud et al 2015) to have an essentially equivalent performance, and their hybrid blending the complete Kalman Gain matrices of the two systems in an EnKF framework was comparable to the hybrid ensemble data assimilation system currently operational at ECMWF, but with lower computational cost. The hybrid ensemble data assimilation system at ECMWF uses an ensemble of 4D-Var assimilation at reduced resolution to provide a flow-dependent estimate of background errors for use in 4D-Var assimilation (Bonavita et al. 2015).*"

10. p. 20, line 8: Could you please provide any examples?

We thank the reviewer for this comment. Unfortunately we did not have a chance to run comparison experiments yet. We have dropped this statement even though we are convinced it is correct.

Estimating Surface Carbon Fluxes Based on a Local Ensemble Transform Kalman Filter with a Short Assimilation Window and a Long Observation Window: an OSSE test in GEOS-Chem 10.1

[1,2] Yun Liu, [1]Eugenia Kalnay*, [1]Ning Zeng*, [3] Ghassem Asrar, [4]Zhaohui Chen, [5]Binghao Jia

1 Dept. of Atmospheric and Oceanic Science, University of Maryland – College Park
2 Dept. of Oceanography, Texas A & M university, College Station, TX
3 Joint Global Change Research Institute/PNNL, College Park, MD
4 School of Environmental Science, University of East Anglia, Norwich, UK
5 State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

*Corresponding authors: ekalnay@umd.edu, zeng@umd.edu

Abstract

We developed a Carbon data assimilation system to estimate the surface carbon fluxes using the Local Ensemble Transform Kalman Filter and atmospheric transfer model GEOS-Chem driven by the MERRA-1 reanalysis of the meteorological field based on the Goddard Earth Observing System Model, Version 5 (GEOS-5). This assimilation system is inspired by the method of Kang et al. [2011, 2012], who estimated the surface carbon fluxes in an Observing System Simulation Experiment (OSSE) mode, as evolving parameters in the assimilation of the atmospheric CO2, using a short assimilation window of 6 hours. They included the assimilation of the standard meteorological variables, so that the ensemble provided a measure of the uncertainty in the CO2 transport. After introducing new techniques such as "variable localization", and increased observation weights near the surface, they obtained accurate surface carbon fluxes at grid point resolution. We developed a new version of the LETKF related to the "Running-in-Place" (RIP) method used to accelerate the spin-up of EnKF data assimilation [Kalnay and Yang, 2010; Wang et al., 2013, Yang et al., 2014]. Like RIP, the new assimilation system uses the "no-cost smoothing" algorithm for the LETKF [Kalnay et al., 2007b], which allows shifting at no cost the Kalman Filter solution forward or backward within an assimilation window. In the new scheme a long "observation window" (e.g., 7-days or longer) is used to create an LETKF ensemble at 7-days. Then, the RIP smoother is used to obtain an accurate final analysis at 1-day. This new approach has the advantage of being based on a short assimilation window, which makes it more accurate, and of having been exposed to the future 7-days observations, which accelerates the spin up. The assimilation and observation windows are then shifted forward by one day, and the process is repeated. This reduces significantly the analysis error, suggesting that the newly developed assimilation method can be used with other Earth system models, especially in order to make greater use of observations in conjunction with models.

Key words: Carbon Data Assimilation, Surface Carbon Flux, LETKF

1. Introduction

The exchange of carbon among atmosphere, land and oceans contributes to changes in the Earth's climate, and is also sensitive to climate conditions. The CO2 concentration in the atmosphere is affected by both the natural variability of the Earth's planetary system, and anthropogenic emissions. The terrestrial and oceanic ecosystems absorb more than one-half of the anthropogenic CO2 emission [Le Quéré *et al.*, 2016]. One major scientific question is whether this rate of removal of CO2 from atmosphere will continue in future, and can it be enhanced? It is thus essential to better quantify the dynamics of earth surface carbon fluxes (SCF), and the variations of carbon sources and sinks, and their associated uncertainties.

A common approach for estimating SCF from atmospheric CO2 measurements and atmospheric transport models is referred to as a "top-down" approach. The "top-down" methods estimate SCF through techniques such as Bayesian synthesis approach [Rödenbeck et al., 2003; Gurney et al., 2004; Enting, 2002; Bousquet et al., 1999], different types of ensemble Kalman filters (EnKF) [e.g. Peters et al., 2005, 2007; Feng et al., 2009; Zupanski et al. 2007; Lokupitiya et al., 2008], or variational data assimilation method [e.g., Baker et al., 2006, 2010; Chevallier et al., 2009].

Kang et al. [2011, 2012] developed a "top-down" carbon data assimilation system by coupling an atmospheric general circulation model (AGCM), including atmospheric CO2 concentrations, with the Local Ensemble Transform Kalman Filter (LETKF) [Hunt et al., 2007]. The meteorological variables (wind, temperature, humidity, surface pressure) and CO2 concentrations were assimilated simultaneously in order to account for the uncertainties of the meteorological field, and their impact on the transport of atmospheric CO2. They carried out Observing System Simulation Experiments (OSSEs), and their carbon assimilation system achieved for the first time an accurate estimation of the evolving SCF at the model grid resolution, without requiring any *a priori* information. The surface carbon fluxes were considered as "unobserved evolving parameters", by augmenting the state vector at each column with a surface carbon flux (SCF). The Local Ensemble Transform Kalman Filter (LETKF) then estimated this evolving parameter from the error covariance between the low level atmospheric CO2 and the estimated SCF, and after a spin-up of about one month, the LETKF accurately recovered the nature run seasonal surface carbon fluxes.

Kang et al., [2011, 2012] used a short 6-hour assimilation window for both atmospheric and CO2 observations because atmospheric observations are usually assimilated at this frequency, and because most Ensemble Kalman Filter methods require short windows to ensure that the forecast perturbations growth remains linear. Such a short data assimilation window, required by the LETKF, also protects the system from becoming ill conditioned [Enting, 2002, Fig. 1.3], and as a result it does not require additional *a priori* information. We note further that the use of such a short assimilation window differs very much from most other "top-down" approaches for estimating SCF that use long assimilation windows varying from a few weeks to months [e.g., Baker et al., 2006, 2010; Peters et al., 2005, 2007; Michalak, 2008; Feng et al., 2009].

Although the Kang et al. methodology was successful, it is computationally expensive, requiring ensemble forecasts and data assimilation not only for the carbon variables, but also for the standard atmospheric variables, in order to estimate the uncertainties of the CO2 atmospheric transport process. In this study, we used an improved version of LETKF data assimilation system with a state-of-the-art atmospheric transport model, the GEOS-Chem [Bey et al., 2001; Nassar et al., 2013], which is driven by the MERRA-1 reanalysis of the Goddard Earth Observing System Model, Version 5 (GEOS5). The improved data assimilation system, unlike Kang et al [2011, 2012], does not include an estimation of transport uncertainties related to the meteorological field.

The ultimate goal of our LETKF_C system is to estimate the grid-point SCFs, which, as in Kang et al. [2011, 2012], are treated as time-evolving parameters in the system. As mentioned before, an Ensemble Kalman Filter requires a short assimilation window in order to have the ensemble perturbations evolve linearly and remain Gaussian. On the other hand, it is well known that the training needed to estimate evolving parameters through data assimilation could be quite long, so that it benefits from having many observations. Therefore, a short assimilation window would shorten the training period needed for the estimation of the SCF error covariance, hence lengthen the spin-up time.

To address this problem, we developed a new version of the LETKF using the "Running-in-Place" (RIP) method to accelerate the spin-up of EnKF data assimilation [Kalnay and Yang, 2010; Wang et al., 2013; Yang et al., 2012]. Like RIP, the new assimilation system uses the "no-cost smoothing" algorithm [Kalnay et al., 2007b] that allows shifting at a negligible cost of the Kalman Filter solution forward or backward within a given assimilation window. Briefly, the new scheme works like this: a long "observation window" (e.g., 7-days, containing all the observations within 7 days) is used to create a temporary LETKF ensemble analysis at 7-days. Then the RIP smoother is used to obtain a final analysis at 1-day. This analysis has the advantage of being based on a short assimilation window, which makes it more accurate, and of having been exposed to the 7-days of observations, which accelerates the spin up time. The assimilation and observation windows are then shifted forward by one day, and the process is repeated. We have tested this new method (short assimilation, long observation window) achieving a significant reduction of analysis errors, and we believe that this method could be useful in other data assimilation problems.

This paper is organized as follows: Section 2 briefly describes the new system used for CO2 data assimilation (LETKF_C). Section 3 explores the effect of combining assimilation and observation windows in an OSSE framework. Section 4 presents results of the proposed methodology applied to CO2 data. A summary and discussion are presented in section 5.

2. LETKF_C data assimilation system

A data assimilation system includes a forecast model, observations, and a data assimilation method that optimally combines them. In the proposed LETKF_C data assimilation system we use the GEOS-Chem as the forecast model and LETKF as the data assimilation method. The pseudo-observations for our OSSE experiments are created at the locations of the real carbon observations from Orbiting Carbon Observatory-2 (OCO-2) satellite [Crisp et al., 2004].

2.1 **GEOS-Chem model and the "nature" run**

GEOS-Chem is a global 3-D atmospheric Chemical transport model driven by the NASA reanalysis (MERRA-1) meteorological fields from the Goddard Earth Observing System data assimilation System Version 5, by the NASA Global Modeling and Assimilation Office [Bosilovich et al., 2015]. This model has been applied worldwide to a wide range of atmospheric composition and transport studies. The GEOS-Chem model used in this study is the version v10-01 with a resolution of 4° x 5° (latitude x longitude), and 47 hybrid pressure-sigma vertical levels for CO2 simulation [Nassar et al., 2013]. GEOS-Chem is driven by the MERRA-1 reanalysis with 72 hybrid vertical levels, extending from the surface up to 0.01 hPa. The data used in this study was provided by the GEOS-Chem support team, based at the Harvard and Dalhousie Universities with support from the NASA Earth Science Division and the Canadian National and Engineering Research Council, who re-gridded the original data of spatial resolution of 0.25° x 0.3125° into the resolution of 4° x 5°.

GEOS-Chem requires the SCFs as a set of parameters at each grid point in order to simulate the CO2 concentration in the atmosphere. It is not possible to observe the global SCFs directly. Therefore, the SCFs are created from a "bottom-up" approach (considered as "truth" in our experiments) and used for the simulation of atmospheric CO2 concentration with GEOS-Chem. The "bottom-up" SCFs used in this study include the three components shown in Equation (1): 1) terrestrial carbon fluxes (F_{TA}); 2) air-sea carbon fluxes (F_{OA}); 3) anthropogenic fossil fuel emissions (F_{fe}).

$$SCF = F_{TA} + F_{OA} + F_{fe} \quad (1)$$

The F_{TA} values are derived from the VEgetation Global Atmosphere Soils (VEGAS) model [Zeng et al., 2004; Zeng et al., 2005], forced by the real evolving weather, obtained from the GEOS-Chem. The F_{OA} values are from Takahashi et al. [2002], a climatological seasonal cycle estimated for the 1990s, and the F_{fe} values are from Fossil Fuel Data Assimilation System (FFDAS) for the year 2012 [Asefi-Najafabady et al., 2014]. The air-sea carbon flux and F_{fe} values were scaled using the global carbon budget data of Le Quéré et al. [2015], in order to include interannual variations. A nature run for atmospheric CO2 concentration simulation is driven by the SCFs in units of ($\frac{kgC}{m^2 yr}$) based on all three datasets.

In OSSEs, the nature run serves as the "truth". We assume that the true "bottom-up" carbon fluxes are not known in our data assimilation experiments, and they will be estimated using the atmospheric pseudo-observations derived from the "truth", as described in more detail below. The nature run obtained by coupling GEOS-Chem with VEGAS is fairly realistic [~~(~~figure not shown], so we use it to create the pseudo OCO-2 observations for the period of January 2015- March 2016.

2.2 Pseudo-Observations

The ultimate goal of this model-data assimilation system is to estimate the SCFs at every grid point using real observations such as the conventional surface CO2 measurements of GlobalViewplus (GV+) flask network provided by Cooperative Global Atmospheric Data Integration Project [2016], and the observations from satellites such as the Greenhouse Gases Observing Satellite (GOSAT) [Yokota et al., 2004], and the Orbiting Carbon Observatory-2 (OCO-2) [Crisp et al., 2004]. Therefore, it is very beneficial to choose a realistic observation network to generate the pseudo-observation for testing the proposed data assimilation system. In this study, we developed the pseudo-observations for the OSSE assimilation experiments based on a realistic OCO-2 observation product.

The OCO-2 observations are the CO2 column-averaged dry air mole fractions over the entire OCO-2 pixel (defined as Xco2). The synthetic observations cover the entire globe once every 14 days with very high spatial resolution. It includes 24 samples per second along the satellite track within ~ 7 km span. The observations are expected to be highly correlated over a short length scale. Furthermore, the observation quality is greatly affected by conditions such as cloud cover, surface type and the solar zenith angle at the time of measurement. The OCO-2 retrieval algorithm uses a warning level (WL) between 0 and 19 to indicate the quality of measurements, where WL=0 means "most likely good", and WL=19 means "least likely good" observations. To avoid highly correlated measurements being treated as independent measurements and to bring the spatial resolution in line with the resolution of atmosphere transfer model, David Baker provided an OCO-2 observation dataset which averaged the synthetic Xco2 in 10-second time window using the "good quality" observations retrieval defined by WL <= 15 (personal communication).

The OCO-2 retrievals used to obtain averages are based on the NASA Atmospheric CO2 Observations from Space XCO2 retrieval Algorithm version7r (O'Dell et al., 2012), as archived at https://disc.gsfc.nasa.gov/datasets/OCO2_L2_Lite_FP_7r/summary (last access: 23 March 2017). A two-step averaging method has been used in order to avoid the final average to be disproportionately weighted to one part of the averaging bin (track) with more good quality retrievals. In the first step, the “good quality” retrievals defined as WL<=15 and xco2_quality_flag=0 (another quality indicator of the data) are averaged over 1-second bins, with weights inversely proportional to the square of each retrievals posterior uncertainty. In the second step, all the 1-second bins, with at least one valid retrieval, are averaged over a 10-second interval to create 10-second averaged data. The OCO-2 averaging kernels are similarly averaged to create 10-second mean averaging kernels. This averaging method had been used for similar purpose in the recent study by Basu et al. (2018). ~~The details of the averaging method is described by Basu et al (2018).~~ In this study, we further aggregated the observations from David Baker at the nearest GEOS-Chem output time of the 0, 6, 12, 18 UTC for each model day. The typical one-day coverage of observation of OCO-2 is shown in Figure 1. The values of Xco2 in the winter are significantly larger than those in summer of the Northern hemisphere and the OCO-2 observations are missing in the winter, for middle and high latitude regions (latitude > ~30). We used the actual location, time and error scales of the OCO-2 observations to create the pseudo-observations for our experiment. The pseudo-observations are created by obtaining the “true” CO2 from the “nature” run using the location and time of the valid observation, then adding random errors with due consideration to the scales of the corresponding real observations. These derived pseudo-observations used in this study are based on the real observations associated error scales , thus are more realistic than the GOSAT observations also used in Kang et al. [2012], because they are anchored, for example, to the real OCO-2 observations and to their quality, and their statistical representation.

2.3 The LETKF data assimilation system

The ensemble Kalman filter (EnKF) is a powerful tool for data assimilation that was first introduced by Evensen [1994]. The key attribute of this method is to derive the forecast uncertainties from an ensemble of integrated model simulations. A variety of ensemble Kalman filter assimilation methods have been proposed [Burgers et al., 1998; Houtekamer and Mitchell, 1998; Anderson, 2001, 2003; Bishop et al., 2001; Whitaker and

Hamill, 2002; Tippett et al., 2003; Ott et al., 2004, Hunt et al., 2004]. The Local Ensemble Transform Kalman Filter (LETKF) introduced by Hunt et al. [2007] is chosen for this study.

The LETKF is an extension of the Local Ensemble Kalman Filter [Ott et al., 2004] with the implementation of the ensemble transform filter [Bishop et al., 2001; Wang and Bishop, 2003]. It is widely used for data assimilation, including several operational centers, and was also used for carbon data assimilations by Kang et al. [2011, 2012].

As discussed earlier, we follow Kang et al., [2011] in estimating the SCFs as evolving parameters, augmenting the state vector C (the prognostic variable of atmospheric CO2) with the parameter SCF, i.e., $\mathrm{X} = [\mathrm{C}, \mathrm{SCF}]^{\mathrm{T}}$. The analysis mean $\bar{\mathrm{X}}^{\mathrm{a}}$ and its ensemble perturbations X^{a} are determined by Equations (2.1, 2.2) at every grid point, and the ensemble analysis is used as initial conditions for the ensemble forecast in the next cycle.

$$\bar{X}^a = \bar{X}^b + X^b \tilde{K}(y^o - \bar{y}^b) \qquad (2.1)$$

$$X^a = X^b[(K-1)\tilde{P}^a]^{1/2} \qquad (2.2)$$

Here $\bar{X}^b$ is the mean of the forecast (background) ensemble members; X^b is a matrix whose columns are the background perturbations of $X_k^b - \bar{X}^b$ for each ensemble member X_k^b (k=1,…,K), where K is the ensemble size); y^o is a vector of all the observations; $\bar{y}^b$ is the background ensemble mean in observation space ($\bar{y}^b = H(\bar{X}^b)$), where H is the observation forward operator that transforms values in the model space to those in the observation space); $\tilde{P}^a = \left[(Y^b)^T R^{-1} Y^b + \frac{(K-1)I}{r}\right]^{-1}$ is the analysis error covariance matrix in ensemble space, which is a function of $Y^b = HXb$, the matrix of background ensemble perturbations in the observation space, R, the observation error covariance (e.g., measurement error, aggregation error, representativeness error), and of r , a multiplicative inflation parameter; and $\tilde{K} = \tilde{P}^a Y^b R^{-1}$. LETKF assimilates simultaneously all observations within a certain distance at each analysis grid point, which defines the localization scale. Hunt et al. [2004] introduced a 4-dimensional version, and Hunt et al. [2007] provide a detailed documentation of the 4D-LETKF which we are using.

2.4 Choosing the long observation window (OW) and the short assimilation window (AW)

Like other data assimilation methods, LETKF proceeds in analysis cycles that consist of two steps, a forecast step and an analysis step. In the analysis step, the model forecast (also called prior or background) and the observations are optimally combined to produce the analysis (also called the posterior), which is the best estimate of the current state of the system under study. In the forecast step, the model is then advanced in time with the analysis as the initial condition and its result becomes the forecast for the next analysis cycle. All observations within the assimilation time window are used to constrain the state at the end of the assimilation window.

The focus of this study was on the estimation of SCFs that are time varying parameters in GEOS-Chem. As discussed earlier, a preliminary LETKF analysis, which provides the weights for each ensemble perturbation, is performed over a longer window (e.g., 7 days with observations starting at time t). Then, the "No-Cost" smoothing [Kalnay et al, 2007b, Kalnay and Yang, 2010] is applied, using the same analysis weights obtained at the end of the long observation window (e.g., 7 days) for each ensemble member, but combining the ensemble perturbations at the end of the corresponding short assimilation window (e.g., 1-day). This creates the final 1-day analysis (at time t+AW), which benefits from the information from all the observations made throughout the long OW (7 days), and from the linearity of the perturbations in the short AW of 1 day, which is required for accuracy. At this time the procedure is repeated starting at t+AW, one day later.

In this new approach, we have the flexibility to combine a short assimilation window (AW) of length m (e.g., m=1 day), with a long observation window (OW) of length n (e.g., n=7 days), to improve the estimation of SCF. In the forecast step, the model is integrated from t to t+n, to produce the forecast corresponding to the observations within the OW. In the analysis step, the observations and corresponding forecasts within the OW are used by the LETKF to estimate optimal weights for the ensemble members. The "No-Cost" smoother applies these optimal weights to determine the analysis of the model state and the SCF parameter at $t+m$. The resulting analysis is then used as the initial conditions for the next analysis cycle starting from time $t+m$.

2.5 Experimental setup

In our experiments we used an ensemble size of 20 members, which was reasonable since ~~our~~ the data assimilation ~~of~~ include only one state variable (CO2 concentration) and one parameter variable (SCF). ~~The~~ A similar experiment but with 80-member ensemble size showed only slight improvement of assimilation quality (figure not shown) but dramatically increased the computational cost. The initial ensemble is created by random selection of the state and flux values from the model-based "nature" run for both SCF and atmospheric CO2 concentration. Therefore, the initial uncertainties of fluxes and CO2 values are equivalent to their "natural" variability. Based on a sensitivity analysis, we found a horizontal localization radius of 15000 km is optimal for our system. Following Kang el al. [2012], a vertical localization is also applied by assigning a larger weight to the CO2 updating on surface layers to reflect the expected dominance of layers near the ground in the change of the total column CO2 measured by OCO-2.

2.6 Additive Inflation Method

The inflation is very important for our LETKF_C data assimilation system. The LETKF uses the forecast ensemble spread to represent forecast uncertainties. All EnKFs tend to underestimate the uncertainty in their state estimate because of nonlinearities and limited number of ensemble members (Whitaker and Hamill, 2002). Underestimating the uncertainty (ensemble spread) leads to overconfidence in the background state estimate, and less confidence in the observations, which will eventually lead the EnKF to ignore the observations and result in filter divergence. This is also true for our carbon-LETKF data assimilation system. The ensemble spread of CO2 in GEOS-Chem model decreases during model integration when the ensemble members are using the same meteorological forcing and SCF values, which is very different from the system with prognostic meteorological fields where the ensemble spread of model state increases during model integration (not shown). The ensemble spread of SCFs also does not increase during model integration because the SCFs are predicted using persistence, and the LETKF decreases the ensemble spreads for both SCFs and CO2 during analysis steps. Therefore, without inflation, the ensemble spread of the CO2 and SCFs would be continuously decreasing during data assimilation, and soon would become too small for LETKF to accept any observations, and hence, cause filter divergence.

There are different types of inflation methods that address the problem of overconfidence, such as multiplicative inflation, relaxation to prior, and additive inflation [e.g. Anderson and Anderson, 1999; Mitchell and Houtekamer, 2000; Zhang et al., 2004; Whitaker et al., 2008; Miyoshi, 2011]. For this study, we chose additive inflation, which adds random fields to the analysis before the ensemble forecast of the next analysis cycle. Additive inflation has some advantages compared to multiplicative inflation because it prevents the effective ensemble dimension from collapsing toward the dominant directions of error growth [Whitaker et al., 2008; Kalnay et al., 2007a]. We applied additive inflation to the ensemble of atmospheric CO2 and SCF to increase perturbations in the initial conditions, for the next time step. It is important for an additive inflation method to minimize the impact of model imbalance and initial shocks generated by adding the random fields into a model. Following Kang et al [2012], the added ~~random~~ fields are selected randomly from the model nature run. Pairs of atmospheric CO2 and surface CO2 flux fields are chosen randomly from model nature run within one year before the analysis time, their ensemble mean is removed and their difference are scaled to a magnitude corresponding to 30% of model seasonal variance to create the ensemble of random fields for additive inflation. Therefore, each selected random field~~s~~ is balanced, and when it is added into model, the balance will be essentially maintained.

3. Sensitivity analysis for AW and OW length

We tested the new version of the LETKF with short AW and long OW, described in previous sections by conducting two sets of experiments using the LETKF_C system in an OSSE framework with OCO-2 like observations. The first set of experiments used the regular 4D-LETKF settings (with a single window length AW=OW) to investigate the effect of the length of AW for estimating SCF. In the second set of experiments, we investigated the optimal OW length after choosing the best AW from the first set of experiments. The assimilation period for all experiments was 1 January 2015 to 1 March 2016. The annual mean RMSEs differences are calculated from the simulation results by removing the spin-up period of first two months (January and February 2015). The average period is from March 1 2015 to the end of February 2016. The details of experimental settings are shown in Table 1.

Table 1. Lengths of Assimilation Window (AW), and Observation Window (OW), and the

resulting time-averaged global mean RMSEs for different experiments. The first four experiments use regular 4D-LETKF, with AW=OW. The last four experiments use AW=1 day, found to be optimal, and different OWs.

	EXP1	EXP2	EXP3	EXP4	EXP5	EXP6	EXP7	EXP8
AW	6 hours	1 day	3 days	7 days	1 day	1 day	1 day	1 day
OW	6 hours	1 day	3 days	7 days	2 days	8 days	15 days	30 days
RMSE $(\frac{kgC}{m^2yr})$	0.077	**0.059**	0.068	0.074	0.053	**0.041**	**0.040**	0.050

3.1 Sensitivity analysis for different assimilation windows

The sensitivity of SCF estimates to the length of AW was investigated based on the first set of experiments (EXP1-EXP4) with regular 4D-LETKF settings, where the length of OW is the same as that of the AW. All experiments used the same observations and initial conditions. Since the temporal coverage of the OCO-2 observation network is too sparse for our LETKF_C assimilation system to estimate the SCF signal in a short time scales, we focus on the estimation of SCF for the seasonal and longer time scales.

Figure 2 shows the estimated global total surface fluxes from the first set of experiments. The “true” global total surface fluxes show a clear seasonal cycle with very large carbon uptake during the growing season of Northern Hemisphere (NH), from May to August, and carbon release during other seasons with the peak release during November. All experiments reproduced fairly well the seasonal cycle of SCF.

When the AW is very short (6 hours), there are large magnitude and high frequency noise overlaying the seasonal cycle. The magnitude of high frequency errors of SCF estimation in EXP1 is comparable with the seasonal variability of SCF (Figure 2a). When the AW=7 days, the high frequency errors of estimation decay, but the long assimilation window increases the analysis RMSE (EXP4). The EXP2 with AW= 1 day produced the best estimation of SCF among all four experiments with equal observation and assimilation windows (Figure 2).

The advantage of AW=1 day (EXP2) is clearly seen from the smaller average global root mean square error (RMSE) (Figure 2c). The RMSE of surface carbon flux is calculated as

$$RMSE(t) = \sqrt{E_x((F^a(x,t) - F^n(x,t))^2)} \quad (3)$$

where x and t are space and time location; F^a and F^n indicate the analysis and the "true" SCF from nature run, respectively. E_x is spatial average. The estimations from experiments with long AW (3 days and 7 days) have a smaller RMSE for the first three months (January to March), when the "truth" had very little variation because the long AWs enhances the signal and smoothes the high-frequency noise. The experiments with long AW could miss the fine-scale signals of SCF variation and fail to catch its variation with time. Therefore, the estimations with long AW showed large RMSE during the period when SCF had larger variations. The estimation with AW of 6 hour showed very large RMSE because of the overwhelming high frequency noise. The estimation with AW of 1 day had the smallest RMSE among all the experiments with regular 4D-LETKF.

The time-averaged RMSEs of SCFs is calculated as

$$RMSE(x) = \sqrt{E_t((F^a(x,t) - F^n(x,t))^2)} \quad (4)$$

which shows very similar spatial patterns, but different amplitudes for different experiments (Figure 3). The large RMSEs of SCF estimation located in Southeast American, Southeast of China and Russia, and resembled that of the SCF variance (not shown). The regions of higher variance indicate more information is needed to resolve such large variance by observations, which is hard to achieve. As expected, the SCF RMSE of 0.059 from EXP2 with AW of 1 day is significantly smaller than the RMSE from EXP1 with a short AW of 6 hour (0.077 $\frac{kgC}{m^2yr}$), and EXP3 and EXP4 with longer AWs of 3 days (0.068 $\frac{kgC}{m^2yr}$) and 7 days (0.074 $\frac{kgC}{m^2yr}$) respectively.

Our results suggest that the preferred AW for estimating SCF is 1 day. This is distinctly different from previously published studies that indicate either a very short AW (6 hours) [Kang et al 2011, 2012], or a very long AW (longer than a few weeks) [e.g., Baker et al., 2006, 2010; Peters et al., 2005, 2007; Michalak, 2008; Feng et al., 2009] is optimum. A short AW can better constrain the model state and therefore produce a better parameter estimation. It is worth mentioning that a very short AW of 6 hours can degrade the SCF estimation with high frequency noise in our LETKF-C system. We postulate that the high frequency noise is related to the sampling errors in the CO_2-SCF covariance that has smaller signal-noise ratio compared to those in experiments with longer AWs.

The same results can be obtained from the same experiments with different initial time, indicating the robustness of our findings [figure not shown]. The convergence of estimated SCFs from the experiments starting from months with big SCF variation, such as April, is slightly slower than the experiments from the time with small SCF variation, such as January. While the estimated SCFs converges in a few analysis cycles (a few days) in our system (Figure 2), the small difference of convergence does not make any significant impact on the quality of estimated SCFs. Moreover, the calculation of RMSE of estimated SCFs has excluded the spinup period of first two months to remove the potential impact of initial condition and initial time.

3.2 Sensitivity analysis for different observation windows (OW)

The results presented earlier and associated discussion suggest that parameter estimation through data assimilation benefits from long training time and having sufficient number of observations, implying that the length of OW is critical for the estimation of desired parameter(s). We investigated the effect of such sensitivity to find out the suitable length of OW for estimating SCF in the second set of experiments (EXP5-EXP8), all based on the optimum AW=1 day that was identified from the first set of experiments, with different OW lengths.

The estimated global total SCFs in the second set of experiments show a clear seasonal cycle matching the “truth” (Figure 4a). Compared with EXP2 (OW=1) shown with the green line in Figure 2a), EXP5 (OW=2days) reduced the high frequency noise significantly when the OW length was increased from 1 day to 2 days. There is still some high frequency noise in the SCF estimation for EXP5, because the observations for 2 days are not sufficient to smooth out the high frequency~~t~~ noise introduced into the estimation through data assimilation. The estimated global total SCFs for EXP6 (OW=8days), EXP7 (OW=15), EXP8 (OW=30) are much smoother than that of EXP5 (OW=1day), because of their longer OW. However, the estimation for OW of 30 days shows a clear time shift compared with “truth”, especially during the transient period when the majority of ecosystems /plants switching from dormant phase in the winter to the growing phase in the spring. The surface carbon fluxes change rapidly during this period. The time shift can also be seen in the estimations for these experiments with OW of 15 days, but it is less pronounced. In the proposed LETKF technique, most of observations in a long OW are introduced at a time later than the assimilation time. Since the SCFs are temporally evolving parameters, the information (variation) of future surface fluxes is brought into the estimation of current time when the future observations are included in the OW. Therefore, the estimated SCF with a very long OW tend to shift towards its future value. The estimated SCF with moderate OW=8 days and 15 days (EXP6 and EXP7) are more accurate than those with a short OW of 2 days (EXP5) and very long OW of 30 days (EXP8), by avoiding the significant high frequent noise observed in EXP5 (OW=2 days) and the significant time shift present in EXP8 (OW=30 days). The global mean RMSEs of estimated SCF from OW=8 and 15 days (EXP6 and EXP7) are significantly smaller than those from OW=2 and 30 days, i.e., EXP5 and EXP8 (Figure 4c).

The spatial pattern of time-average RMSE of SCF for EXP5 (OW=2 days; Figure 5) is similar to those in the first set of experiments, which had short AW=OW (Figure 3). The regions with large RMSE in EXP5 (OW=2 days) disappear with OW=7 and 15 days in EXP6 and EXP7, because the long OWs enhance the signals for SFC estimation. The large RMSE in SCF estimates for EXP8 (OW=30 days) are primarily in the Northern Hemisphere mid-latitudes, because of the time shift in estimations with OW=30 days. The mean RMSEs of experiments with moderate OWs of 8 and 15 days are 0.041 $\frac{kgC}{m^2yr}$ and 0.040 $\frac{kgC}{m^2yr}$, respectively, which is significantly smaller than those from experiments with OWs of 2 days (0.053 $\frac{kgC}{m^2yr}$) and 30 days (0.050 $\frac{kgC}{m^2yr}$).

A longer OW requires a longer forecast period for each forecast step, which results in additional computational time/cost. For example, EXP7 with OW of 8 days used 8-time more computational time compared to EXP2. Furthermore, the length of OW is also constrained by the time scale of estimation parameters. A long OW tends to generate a time shift for its estimation. For seasonal and longer time scales, OW(s) in moderate range of 8~15 days appear to be most suitable for the LETKF_C estimates of the SCF. EXP6~~7~~ and EXP7~~8~~ show almost the same quality of SCF estimation, but EXP6~~7~~ has higher computational efficiency. The best configuration thus appears to be EXP6~~7~~ with an OW of 8 days and AW of 1 day, referred as the "benchmark" experiment hereafter.

We note that the high frequency noise in EXP1 with a short AW of 6 hours can be smoothed out by a long OW (i.e. 8-15 days). We postulate that an experiment with AW of 6 hours and OW 8 days will produce similarly realistic estimations as the "benchmark" experiment; however, it would require much more computational time.

4 Evaluating estimated fluxes from the "benchmark" experiment

With the moderate long observation and short assimilation windows, we obtained best estimates of surface carbon fluxes, and their seasonal cycle. This section describes the SCF estimates from the "benchmark" experiment. Figure 6 shows a comparison of surface carbon fluxes based on the "benchmark" assimilation experiment and nature ("truth") run for Northern Hemisphere Summer (June, July and August) and Winter seasons (December, January, and February). The "bottom-up" carbon fluxes used in the "nature" run show a very strong seasonal cycle over the continents, except Antarctica. The North Hemisphere mid-latitude areas are very large carbon sinks in the Summer, and carbon sources in the Winter, as expected. The strong seasonal cycle of surface fluxes mainly related to the variability of terrestrial ecosystems that absorbs large amount of CO2 during the growing season (Spring and Summer) and release carbon back to the atmosphere during dormant seasons (Fall and Winter). The estimated surface fluxes in the seasonal time scale follow closely the "truth". The benchmark assimilation experiment closely reproduces the spatial pattern of surface fluxes globally, for different seasons. The difference between the benchmark estimation and "truth" shown in Figures 6 e & f are very small. There are some positive carbon flux differences over Northern Hemisphere mid-latitudes in the Winter, thus a positive bias in estimated atmospheric CO2 concentration is expected.

The analysis of CO2 concentrations matches the "nature" run well. The error pattern also matches the CO2 seasonal cycle and the error pattern of estimated SCF. ~~A successful estimation of surface fluxes requires a good assimilation of atmospheric CO2, and a good estimation of surface flux parameters to enable the model-assimilation system to produce a good analysis of atmosphere CO2.~~ Figure 7 shows the comparison of surface atmospheric CO2 concentrations between the benchmark assimilation experiment and nature ("truth") run, for the Northern Hemisphere Summer and Winter. The spatial pattern of assimilated CO2 matches the "truth" very well. The analysis successfully reproduced the seasonal cycle of CO2 over Northern Hemisphere mid-latitudes, with low CO2 concentration in Summer (Figures 7a-c) and high CO2 in Winter (Figures 7b-d), consistent with seasonal cycle of CO2 absorption and release by terrestrial ecosystems. There are positive CO2 concentrations located at high latitudes of North America and far East Asia regions during Winter 2016 (Figure 7f), due to the positive bias in estimated SCF (Figure 6f).

The consistency of annual mean estimated SCF for both benchmark experiment and "truth" is a very important feature for our LETKF_C assimilation system (Figure 8a). In EnKF assimilation the ensemble spread is considered as a good representation of uncertainties associated with both parameters and model state [e.g., Evensen 2007, Liu et al. 2014]. The surface carbon fluxes are special parameters that vary with time and it is very hard to quantify their uncertainty during assimilation. When the ensemble spread of parameters are too small to drive model with a robust response, the estimation fails. The additive inflation with 30% of nature variability is used to maintain the amplitude of parameters ensemble spread. Although the ensemble spread of the global total surface flux, in our experiments, is bigger than its error (Figure 8a), we were still able to estimate very well the global total surface CO2 fluxes (ensemble mean), and their seasonal variability. This is consistent with findings of Liu el al [2014], that parameter estimation can tolerate some inconsistency between parameter ensemble spread and parameter error.

The global mean RMSE of SCF decrease from an initial value of ~0.1 $kg\ C\ m^{-2}y^{-1}$ to ~ 0.04 $kg\ C\ m^{-2}y^{-1}$ in just a few analysis cycles (Figure 8b). It does not further decrease during following assimilation cycles because the SCF values vary temporally. The signals added by observations are mainly used to reproduce the temporal variation of SCF.

It is very important for a SCF estimation to reproduce the spatial distribution of the annual mean of the SCF, since it identifies the carbon sources and sinks in the Earth System. Though the amplitude of annual mean SCF is much smaller than the seasonal cycle of SCF, the estimated spatial pattern of annual mean SCF in the benchmark experiment (Eq. 5) is generally consistent with the "truth" (Figure 9).

$$\Delta F(x) = E_t\left(F^a(x,t)\right) - E_t\left(F^n(x,t)\right) \quad (5)$$

In summary, we found that the OSSE experiments using long observation windows and short assimilation windows resulted in the best estimates of SCF~~to be the best choice for estimating SCF~~.

5 Summary and Discussion

We have developed a LETKF-GEOS-Chem carbon data assimilation (LETKF_C) system for estimating the surface carbon fluxes (SCF). The GEOS-Chem atmospheric transport model is driven by the single realization of meteorology fields from MERRA reanalysis. The proposed system captured the "true" SCF spatial and temporal variability. The system performed best with a choice of short assimilation and long observation windows.

The LETKF requires a short assimilation window to avoid an ill-posed condition caused by the nonlinear processes in the forecast model with a long forecast time. The parameter estimation favors a long training period and many observations. Based on these features, we developed a new method to accurately estimate the SCF. The new scheme separates original assimilation time window into observation (OW) and assimilation (AW) windows, allowing the flexibility to apply an OW that is different than the AW. Like RIP, the new technique takes advantage of the "no-cost smoothing" algorithm developed for the LETKF by Kalnay et al. [2007b] that allows to transport the Kalman Filter solution forward or backward within the observation window.

The new method was applied to the LETKF_C system in the OSSE mode using a dataset developed based on the OCO-2 observation characteristics. The sensitivity experiments for this model-assimilation system demonstrated that the new technique, i.e. with a short AW and long OW, significantly improves the SCF estimation as compared to regular 4D-LETKF with identical observation and assimilation windows. The best AW for SCF estimation is 1 day, which is different from the typical AW of 6 hours used in the meteorological assimilations. An OW in the range of 8-15 days is required to estimate the surface carbon fluxes for seasonal and longer time scales. The benchmark experiment with AW of 1 day and the OW of 8 days successfully reproduced the mean seasonal and annual SCF.

Our working hypothesis was that that the optimal OW for the estimation of SCF could be reduced with more observations. We examined this hypothesis by using simulated OCO-2 observations and Global View Plus (GV+) observations. Similar to the OCO-2 pseudo-observation, the GV+ pseudo-observations were also generated based on the actual location, time and corresponding error scale of the GV+ flask observations. The results show that the AW/OW lengths of 1day /8 day is also optimal with both the OCO-2 and GV+ observation characteristics. We estimated the SCF using the OCO-2 and GV+ pseudo-observations with the identical experiment settings as the OCO-2 experiments, except we replace the experiment with very long OW of 30 days with an experiment with a short OW of 4 days to better evaluate the impact from short OWs. Thus the current experiments settings are using OW of 2, 4, 8, 15 days.

The results from these experiments show that the AW/OW lengths of 1day /8 day is still optimal for both the OCO-2 and GV+ observation characteristics (Figure 10). ~~We also tested the estimation of SCF using simulated Global View Plus (GV+) observations. The results show that the AW/OW lengths of 1day /8 day is also optimal for the GV+ observation characteristics. Similar to the OCO-2 pseudo-observation, the GV+ pseudo-observations were generated based on the real location, time and corresponding error scale of the GV+ flask observations. We estimated the SCF using the GV+ pseudo-observations with the identical experiment settings as the OCO-2 experiments (Figure 10).~~ Generally, the time-mean RMSE of estimated SCF with OCO-2 and GV+ (Figure 10) are ~~larger~~ smaller than the corresponding estimates for OCO-2 only (Figure 5)~~because~~. The short OW of 2 days performs worse than the moderate OWs of 4 days, 8 days and 15 days. The time-averaged global mean RMSE is 0.046 $\frac{kgC}{m^2yr}$ for experiments with OW of 2 days (Figure 10a). The time-averaged global mean RMSE is only 0.040, 0.037 and 0.039 $\frac{kgC}{m^2yr}$ for experiments with OW of 4 days, 8 days and 30 days, respectively (Figure10 b, c and d). We only see a slight impact of observation coverage on the optimal OW length. The best OW appears to be 8~15 days which produce the smallest RMSE when only OCO-2 observations only are assimilated. The smallest RMSE in the experiment is obtained in the experiment with the best OW of 8 days, when both OCO-2 and GV+ observations are assimilated into the system.

~~the spatial and temporals coverage of GV+ are smaller than those of OCO-2, though GV+ directly observes the surface CO2 concentration. The short OW of 2 days and long OW of 30 days perform much worse than the moderate OWs of 8 days and 15 days. The time-averaged global mean RMSE is 0.215~~ $\cancel{\frac{kgC}{m^2yr}}$ ~~and 0.117~~ $\cancel{\frac{kgC}{m^2yr}}$ ~~for experiments with OW of 2 days and 30 days, respectively (Figure10a, d). The time-averaged global mean RMSE is only 0.083 when the optimal OW of 8 days was used (Figure 10 b).~~

Two ~~very~~ different ~~types~~ sets experiments ~~of observations~~ (OCO-2 vs OCO-2 and GV+) ~~leading to~~suggesting the same optimal OW of 8 days indicates that the observation coverage and observation type is not the major factor in deciding the length of optimal OW. We speculate that the optimal OW is mainly decided by the time-scale of model response to the SCF uncertainties because LETKF constrains parameters (SCF) based on the mapping function of parameter-state covariance, hence, only the model response to the parameter uncertainties provide the signal for parameter estimation.

It is worth noting that our approach works best for estimating parameters that vary slowly over moderate time scales. It ~~is~~ may not be optimum for estimating SCF variation for short time-scales such as sub-daily to daily because the variations shorter than OWs are filtered out. Furthermore, we used a coarse spatial resolution (4° x 5°) GEOS-Chem in our study. We postulate that the optimal AW/OW could be different when a higher spatial resolution version of GEOS-Chem is used with the proposed assimilation system, because models with different resolutions response to the SCF may be different. This issue also merits further~~,~~ ~~which is worth~~ exploring in the future.

Our new developed short AW and long OW technique is different from the standard 4D-variational method and the 4D-LETKF. The 4D-Var and the 4D-LETKF have been shown (Bonavita et al. 2015; Hamrud et al 2015) to have an essentially equivalent performance, and their hybrid blending the complete Kalman Gain matrices of the two systems in an EnKF framework was comparable to the hybrid ensemble data assimilation system currently operational at ECMWF, but with lower computational cost. The hybrid ensemble data assimilation system at ECMWF uses an ensemble of 4D-Var assimilation at reduced resolution to provide a flow-dependent estimate of background errors for use in 4D-Var assimilation (Bonavita et al. 2015). ~~Our new approach uses short AW and long OW in order to extract more information from the observations, and could be better than~~

~~both 4D-LETKF and 4D-Var using equal assimilation windows for problems where there are longer time scales in the observations.~~ The short AW and long OW approach can be used with other Earth system models for parameter estimation, when the parameters have slow and smooth variations in time and space, and the observations are too limited to constrain the parameters well.

6 Code and data availability

This study focused on developing a new methodology for estimating carbon flux based on a carbon cycle model/data assimilation system. It does not generate any new dataset. The related code for GEOS-Chem and LETKF can be accessed from http://acmg.seas.harvard.edu/geos/doc/man/chapter_2.html#DownCode and https://github.com/takemasa-miyoshi/letkf, respectively.

References:

Anderson, JL.: An ensemble adjustment Kalman filter for data assimilation. Mon. Wea. Rev., 129, 2884–2903, 2001.

Anderson JL.: A local least squares framework for ensemble filtering. Mon. Wea. Rev., 131, 634–642, 2003.

Anderson, J. L., and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127, 2741–2758, doi:10.1175/15200493(1999)127<2741:AMCIOT>2.0.CO;2, 1999.

Asefi-Najafabady, S., Rayner, P. J., Gurney, K. R., McRobert, A., Song, Y., Coltin, K., Huang, J., Elvidge, C., and Baugh, K.: A multiyear, global gridded fossil fuel CO2 emission data product: Evaluation and analysis of results, J. Geophys. Res. Atmos., 119, 10,213–10,231, doi:10.1002/2013JD021296, 2014.

Banks, H.T.: Control and estimation in distributed parameter systems. In: H.T. Banks, Editor, Frontiers in Applied Mathematics vol. 11, SIAM, Philadelphia, pp 227, 1992a.

Banks, H.T.: Computational issues in parameter estimation and feedback control problems for partial differential equation systems. Physica D 60, 226-238, 1992b.

Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO2, Tellus, Ser. B, 58, 359–365, doi:10.1111/j.1600-0889.2006.00218.x, 2006.

Baker, D. F., Bösch, H., Doney, S. C., O'Brien, D., and Schimel D. S.: Carbon source/sink information provided by column CO2 measurements from the Orbiting Carbon Observatory, Atmos. Chem. Phys.,10, 4145–4165, doi:10.5194/acp-10-4145-2010, 2010.

Basu, S., Baker, D. F., Chevallier, F., Patra, P. K., Liu, J., and Miller, J. B.: The impact of transport model differences on CO2surface flux estimates from OCO-2 retrievals of column average CO2, Atmos. Chem. Phys., 18, 7189-7215, https://doi.org/10.5194/acp-18-7189-2018, 2018.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B., Fiore, A. M., Li, Q., Liu, H., Mickley, L. J., and Schultz M.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, J. Geophys. Res., 106, 23,073-23,096, 2001.

Nassar, R., Napier-Linton, L., Gurney, K.R., Andres, R.J., Oda, T., Vogel, F.R., and Deng, F.: Improving the temporal and spatial distribution of CO2 emissions from global fossil fuel emission data sets, J. Geophys. Res. Atmos., 118, 917-933, doi:10.1029/2012JD018196, 2013.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive sampling with the ensemble transformation kalman filter. Part i: theoretical aspects. Mon. Wea. Rev., 129, 420–436, 2001.

Bosilovich, M. G., Akella, S., Coy, L. et al.: MERRA-2: Initial evaluation of the climate. Series on Global Modeling and Data Assimilation, NASA/TM, 104606, 2015.

Bonavita M. G., Hamrud, M., and Isaksen, L.: EnKF and hybrid gain ensemble data assimilation. Part II: EnKF and hybrid gain results. Mon. Wea. Rev., 143, 4865–4882, doi:10.1175/MWR-D-15-0071.1, 2015.

Burgers, G., Van Leeuwen, P., Evensen, G.: Analysis scheme in the ensemble Kalman filter. Mon. Wea. Rev., 126, 1719–1724, 1998.

Chevallier, F., Engelen, R. J., Carouge, C., Conway, T. J., Peylin, P., Pickett-Heaps, C., Ramonet, M., Rayner, P. J., and Xueref-Remy I.: AIRS-based versus flask-based estimation of carbon surface fluxes, J. Geophys. Res., 114, D20303, doi:10.1029/2009JD012311, 2009.

Cooperative Global Atmospheric Data Integration Project: Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2015; obspack_co2_1_GLOBALVIEWplus_v2.1_2016_09_02; NOAA Earth System Research Laboratory, Global Monitoring Division. http://dx.doi.org/10.15138/G3059Z, 2016.

Crisp, D., Randerson, J. T., Wennberg, P. O., Yung, Y. L., and Kuang, Z.: The Orbiting Carbon Observatory (OCO) mission, Adv. Space Res., 34, 700–709, doi:10.1016/j.asr.2003.08.062, 2004.

Evensen G.: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res., 99(C5), 10143–10162, 1994.

Enting, I. G.: Inverse Problems in Atmospheric Constituent Transport, Cambridge Univ. Press, New York, doi:10.1017/CBO9780511535741, 2002.

Feng, L., Palmer, P. I., Bösch, H., and Dance S.: Estimating surface CO2 fluxes from space-borne CO2 dry air mole fraction observations using an ensemble Kalman filter, Atmos. Chem. Phys., 9, 2619–2633, doi:10.5194/acp-9-2619-2009, 2009.

Hamrud M., Bonavita M., and Isaksen, L.: EnKF and Hybrid Gain Ensemble Data Assimilation. Part I: EnKF Implementation. Mon Wea Rev, DOI: 10.1175/MWR-D-14-00333.1, 2015.

Harlim, J. and Hunt, B. R.: Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model. Tellus A, 59: 731–748. doi:10.1111/j.1600-0870.2007.00255.x, 2007.

Houtekamer, P. L., Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev., 126, 796–811., 1998.

Hunt, B. R., Kostelich, E., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D, 230, 112–126, doi:10.1016/j.physd.2006.11.008, 2007.

Liu Y., Liu, Z., Zhang, S., Jacob, R., Lu, F., Rong, X., Wu, S.: Ensemble-Based Parameter Estimation in a Coupled General Circulation Model. Journal of climate, 27, 7151–7162, 2014.

Le Quéré, C., Moriarty, R., Andrew, R. M. et al.: Global carbon budget 2014, Earth Syst. Sci. Data, 7, 47-85, doi:10.5194/essd-7-47-2015, 2015.

Le Quéré C., Andrew, R. M., Canadell, J. G. et al.: Global Carbon Budget 2016, Earth

Syst. Sci. Data, 8, 605-649, doi:10.5194/essd-8-605-2016, 2016.

Lokupitiya, R. S., Zupanski, D., Denning, A. S., Kawa, S. R., Gurney, K. R., and Zupanski M.: Estimation of global CO2 fluxes at regional scale using the maximum likelihood ensemble filter, J. Geophys. Res., 113, D20110, doi:10.1029/2007JD009679, 2008.

Kalnay E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J.: 4-D-Var or ensemble Kalman filter?. Tellus, Ser. A, 59, 758–773, doi:10.1111/j.1600-0870.2007.00261.x, 2007a.

Kalnay E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J.: Response to the discussion on “4-D-Var or EnKF?” by Nils Gustafsson. Tellus, Ser. A, 59, 778-780, doi: 10.1111/j.1600-0870.2007.00263.x, 2007b.

Kalnay, E. and Yang, S.-C.: Accelerating the spin-up of Ensemble Kalman Filtering. Q.J.R. Meteorol. Soc., 136: 1644–1651. doi:10.1002/qj.652, 2010.

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., Fung, I.: Estimation of surface carbon fluxes with an advanced data assimilation methodology: SURFACE CO 2 FLUX ESTIMATION. Journal of geophysical research, 117., doi:10.1029/2012JD018259, 2012.

Mitchell, H. L., and Houtekamer, P. L.: An adaptive ensemble Kalman filter. Mon. Wea. Rev., 128, 416–433, 2000.

Michalak, A. M.: Adapting a fixed-lag Kalman smoother to a geostatistical atmospheric inversion framework, Atmos. Chem. Phys., 8, 6789–6799, 2008.

Miyoshi, T.: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. Mon. Wea. Rev., 139, 1519–1535, doi:10.1175/2010MWR3570.1, 2011.

O'Dell, C. W., Connor, B., Bösch, 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

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 CO2 surface fluxes from atmospheric trace gas observations, J. Geophys. Res., 110, D24304, doi:10.1029/2005JD006157, 2005.

Peters, W., Jacobson, A.R., Sweeney, C. et al.: An atmospheric perspective on North American carbon dioxide exchange: Carbon tracker, Proc. Natl. Acad. Sci. U. S. A., 104, 18,925–18,930, doi:10.1073/pnas.0708986104, 2007.

Tippett, M., Anderson, J. L., Bishop, C. H., Hamill, T. M., Whitaker, J. S.: Ensemble square root filters. Mon. Wea. Rev., 131, 1485–1490, 2003.

Wang, S., Xue, M., Schenkman, A. D., and Min, J.: An iterative ensemble square root filter and tests with simulated radar data for storm scale data assimilation. Quart. J. Roy. Meteor. Soc., 13**9**, 1888-1903, 2013.

Whitaker, J. S., and Hamill, T. M.: Ensemble data assimilation without perturbed observations. Mon. Wea. Rev., 130, 1913–1924., 2002.

Whitaker J. S., Wei, X., Song, Y., and Toth, Z.: Ensemble data assimilation with the NCEP global forecast system. Mon. Wea. Rev., 136, 463–482, 2008.

Yang, S., Kalnay, E., and Miyoshi T.: Accelerating the EnKF Spinup for Typhoon Assimilation and Prediction. Wea. Forecasting, 27, 878–897, https://doi.org/10.1175/WAF-D-11-00153.1, 2012.

Yokota, T., Oguma, H., Morino, I., and Inoue, G.: A nadir looking SWIR FTS to monitor CO2 column density for Japanese GOSAT project, in Proceedings of the Twenty-fourth International Symposium on Space Technology and Science (Selected Papers), pp. 887–889, Jpn. Soc. for Aeronaut. and Space Sci., Tokyo, 2004.

Zeng, N., Mariotti, A., and Wetzel, P.: Terrestrial mechanisms of interannual CO2 variability, Global Biogeochemical Cycles, 19, GB1016, doi:10.1029/2004GB002273, 2005

Zeng, N., Qian, H., Munoz, E., and Iacono, R.: How strong is carbon cycle-climate feedback under global warming? Geophys. Res. Lett., 31 L20203, doi:10.1029/2004GL020904, 2004.

Zhang, F., Snyder, C., and Sun, J.: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. Mon. Wea. Rev., 132, 1238–1253, 2004.

Zupanski, D., Denning, A. S., Uliasz, M., Zupanski, M., Schuh, A. E., Rayner, P. J., Peters, W., and Corbin, K. D.: Carbon flux bias estimation employing Maximum Likelihood

Ensemble Filter (MLEF), J. Geophys. Res., 112, D17107, doi:10.1029/2006JD008371. 2007.

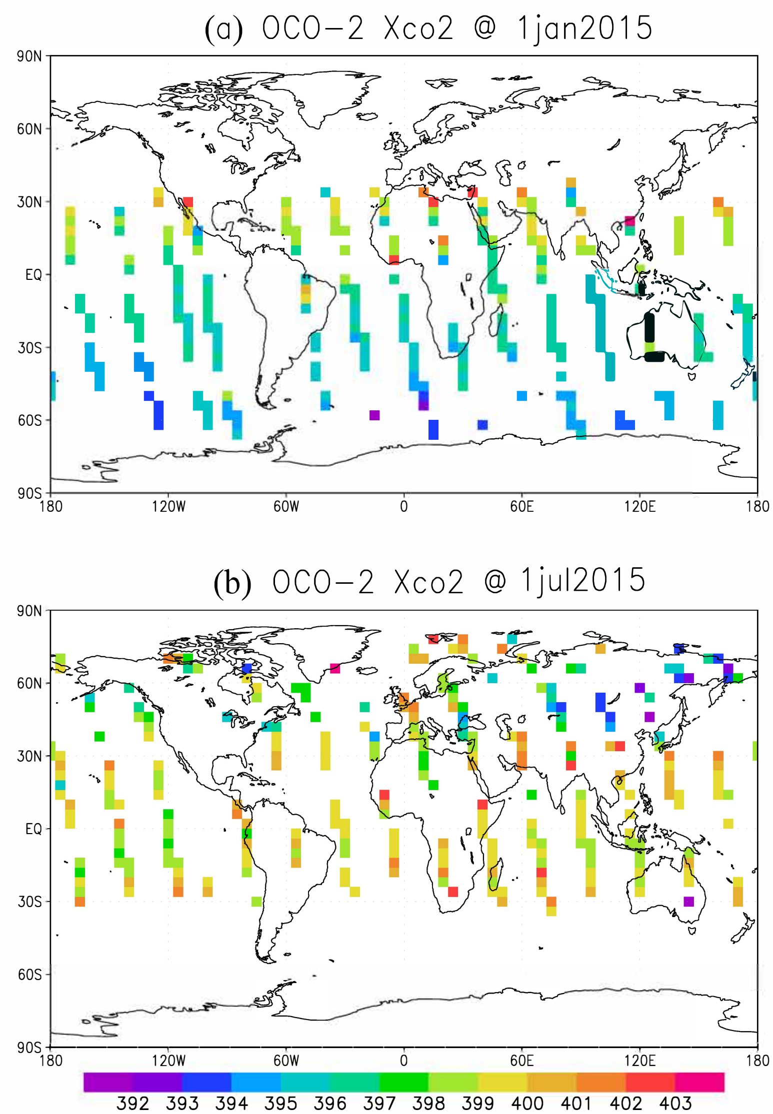


Figure 1. The 10-seconds average of good quality OCO-2 Xco2 observations (Warning Level <=15), obtained from David Baker for (a) 1 January 2015 and (b) 1 July 2015.

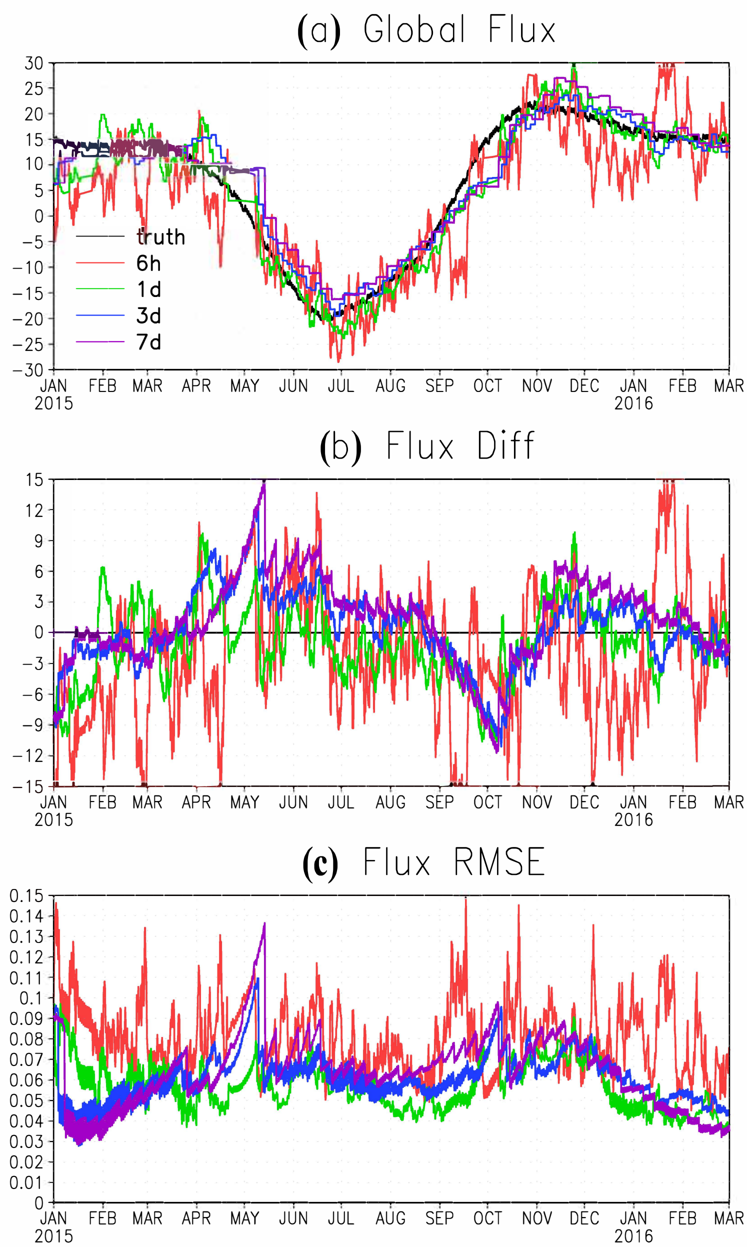


Figure 2. (a) the global total SCF from nature run ("truth", black line) and from the estimations of the first set of experiments with different AW. (b) the difference of global total SCF between the estimations from the experiments with different AW and the nature run ("truth"). (c) the global average RMSE of the estimated SCFs from the experiments with different AW.

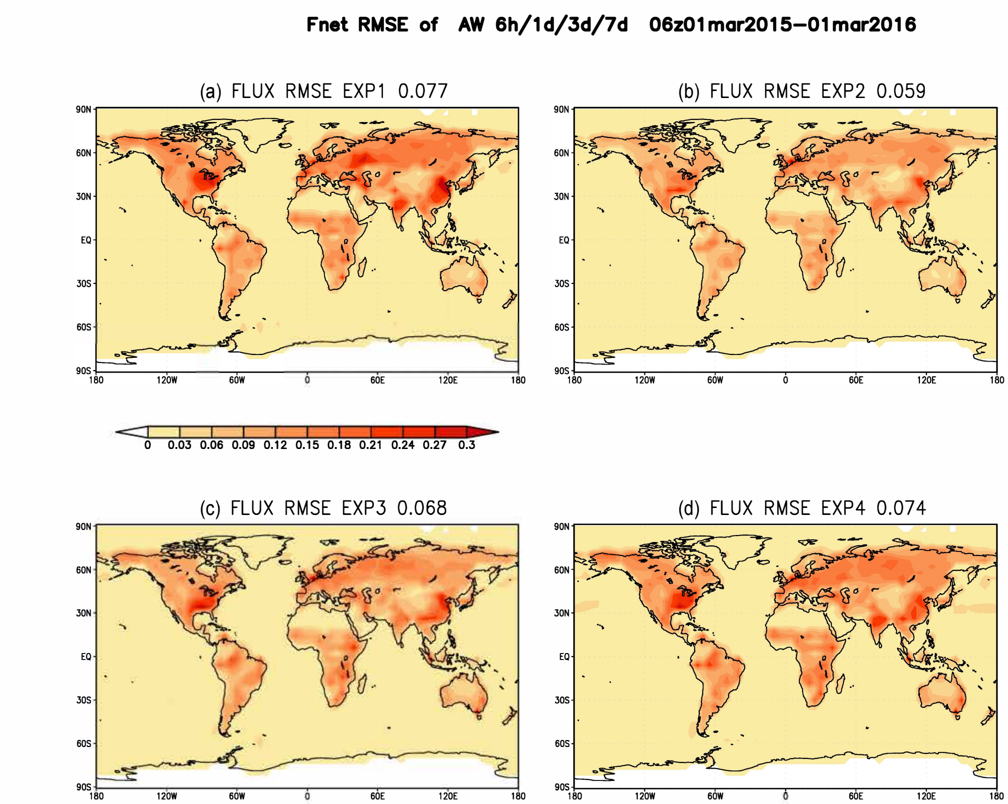


Figure 3. The spatial pattern of the annual mean RMSE of estimated SCF from the experiments with different AW (EXP1-4) for the average period from 1 March 2015 to the end of February 2016. (January and February 2015 are treated as spinup period for our experiments).

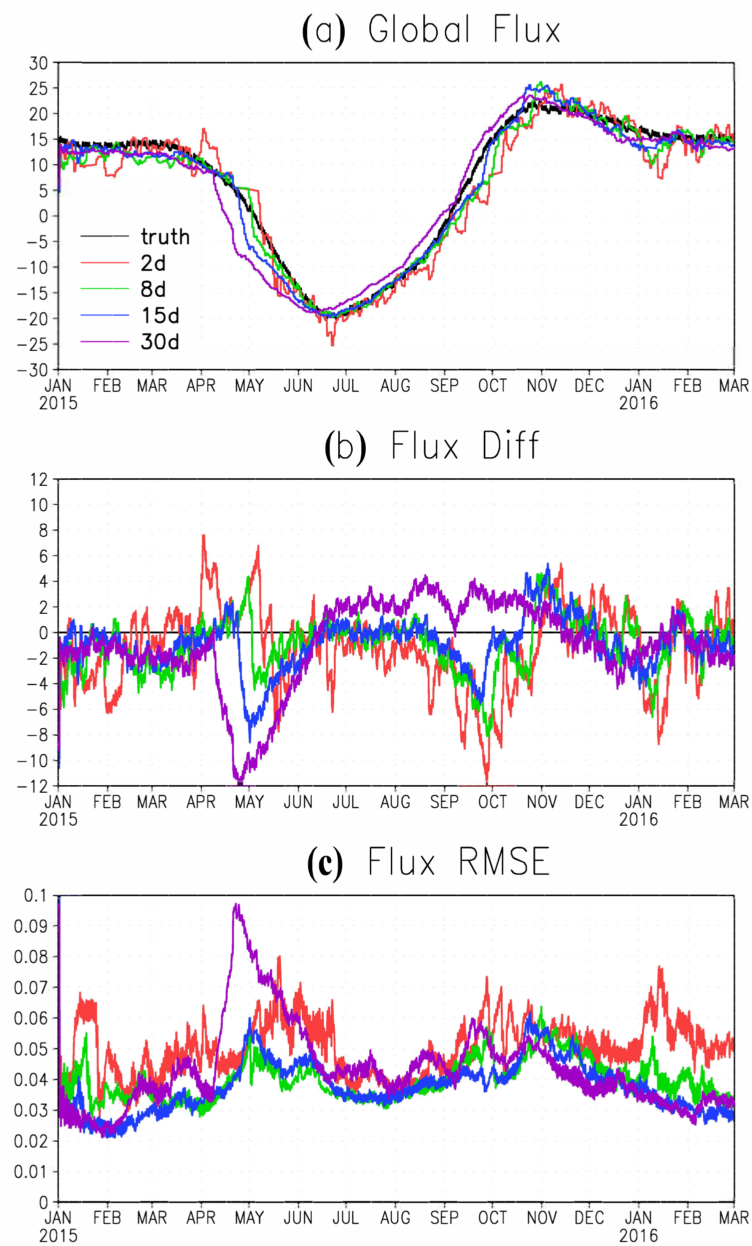


Figure 4. Same as Figure 2, except for the second set of experiments with different OW, but same AW of 1 day.

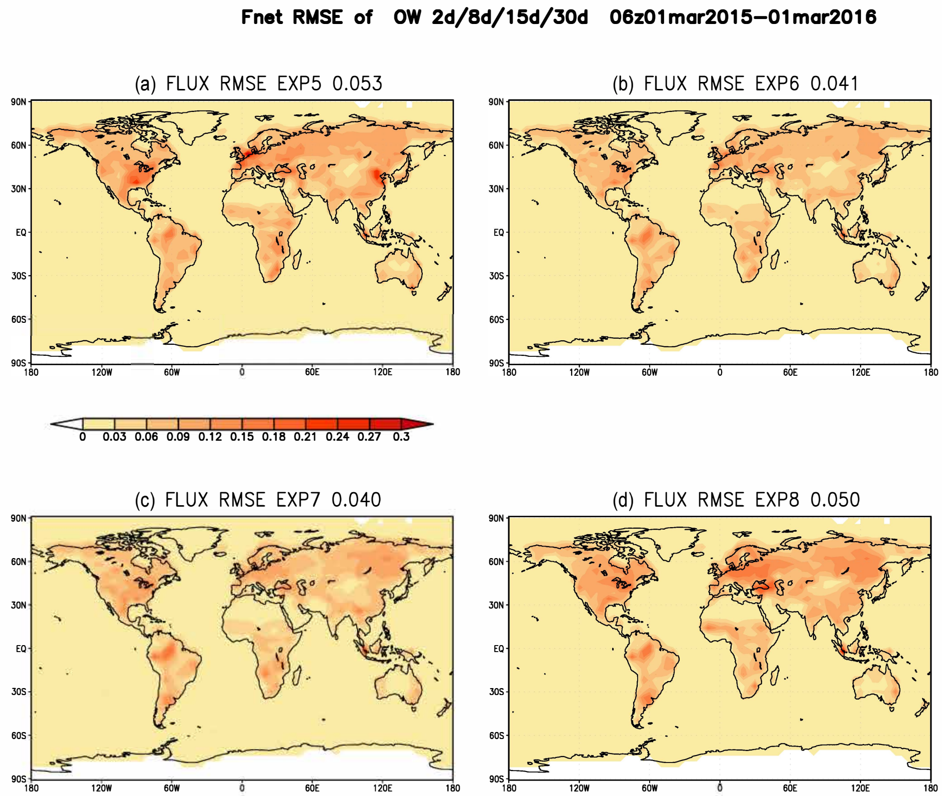


Figure 5. Same as Figure 3, except for the second set of experiments with different OW, but similar AW of 1 day.

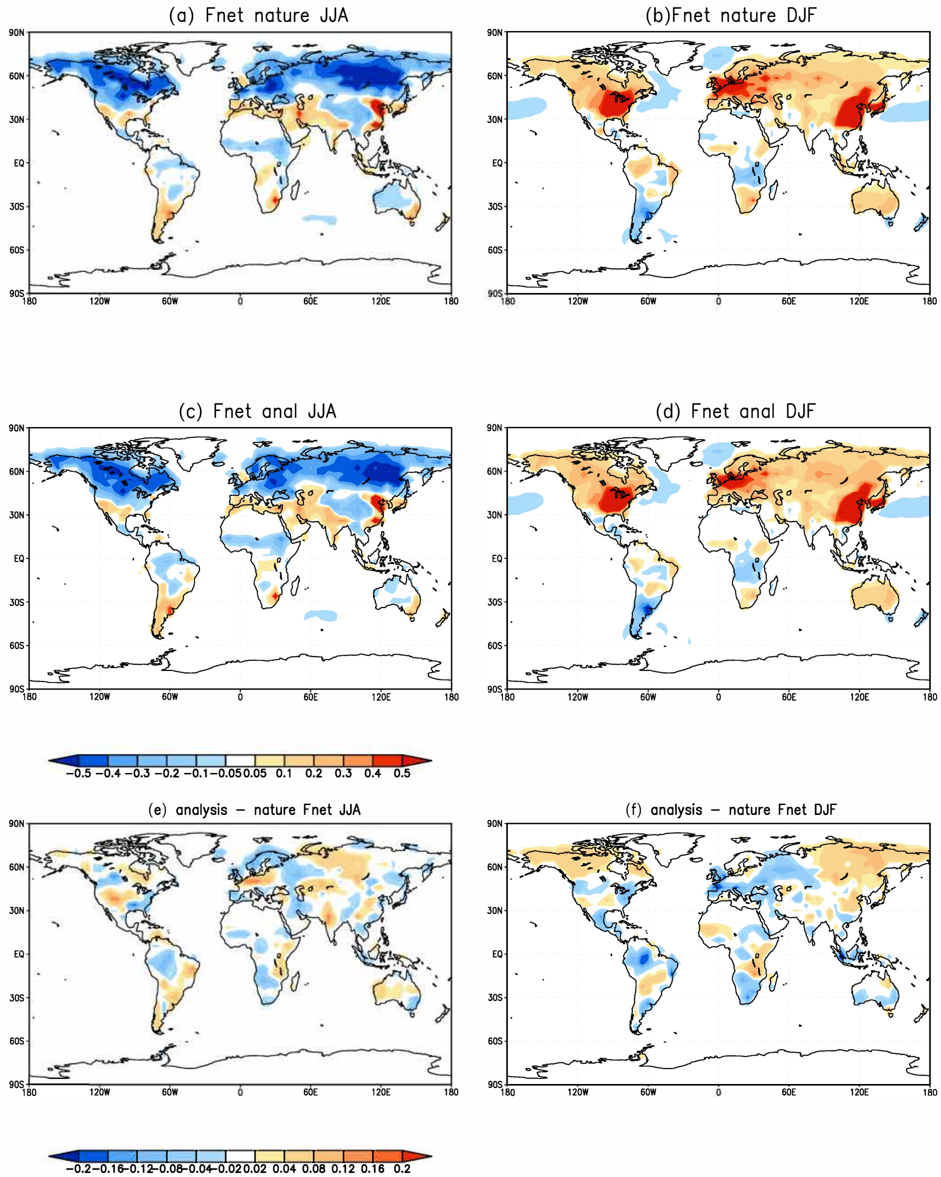
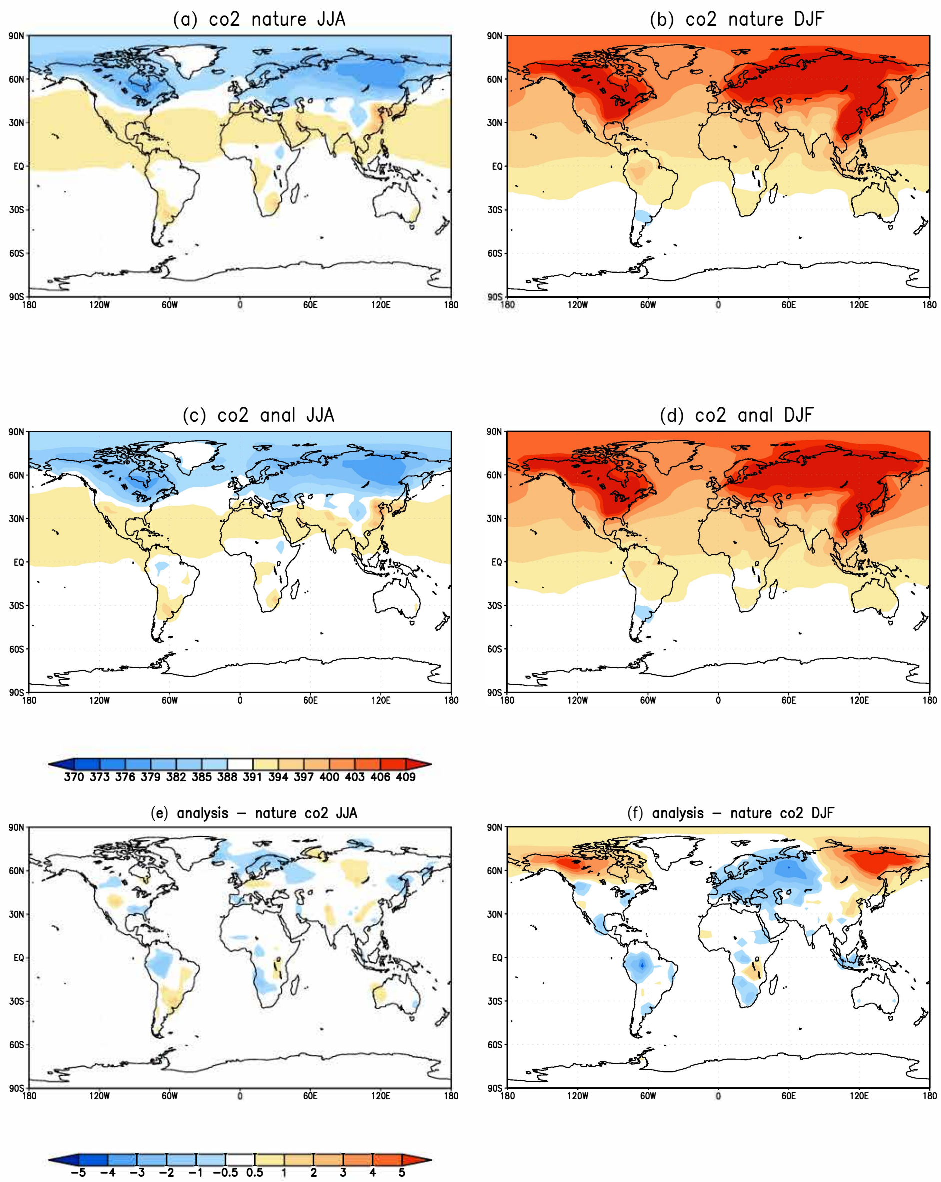


Figure 6.~~:~~ The SCF of "nature" run and estimation from benchmark experiment for Northern Hemisphere Summer (a, c and e), and Winter (b, d, and f).The a and b are the "truth" from the "nature" run; the c and d are the estimates from benchmark experiment; and the e and f are the difference between estimation and "truth".


(a) co2 nature JJA
(b) co2 nature DJF
(c) co2 anal JJA
(d) co2 anal DJF
370 373 376 379 382 385 388 391 394 397 400 403 406 409
(e) analysis − nature co2 JJA
(f) analysis − nature co2 DJF
−5 −4 −3 −2 −1 −0.5 0.5 1 2 3 4 5

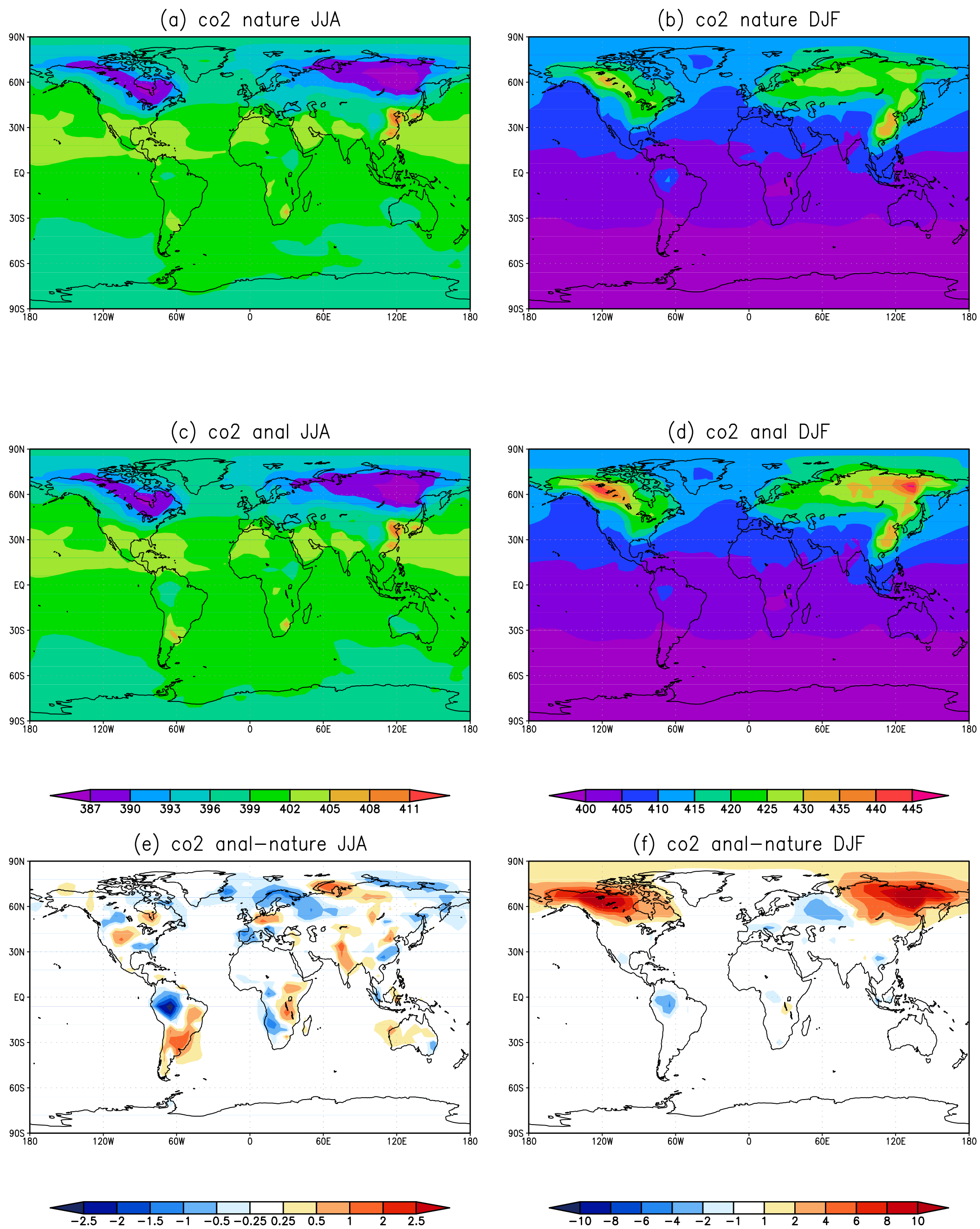


Figure 7. Same as Figure 6, except for surface concentrations of CO2. Where (a) and (c) share the upper left colorbar; (b) and (d) use the upper right colorbar.

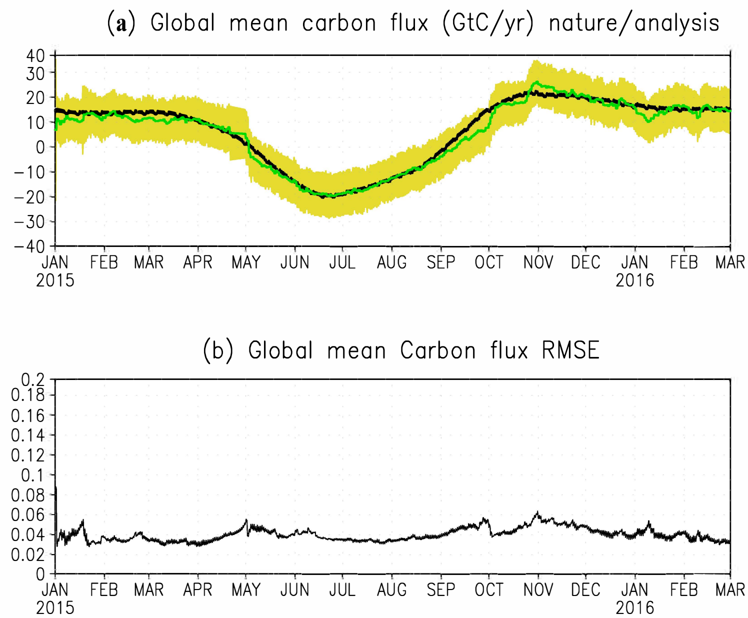


Figure 8. (a) The global total SCF of "truth" and estimation from the benchmark experiment: the black line is the truth, green line is the ensemble mean of the estimation, and yellow shading is the ensemble spread. (b) the global mean RMSE of the estimated SCF from the benchmark experiment.

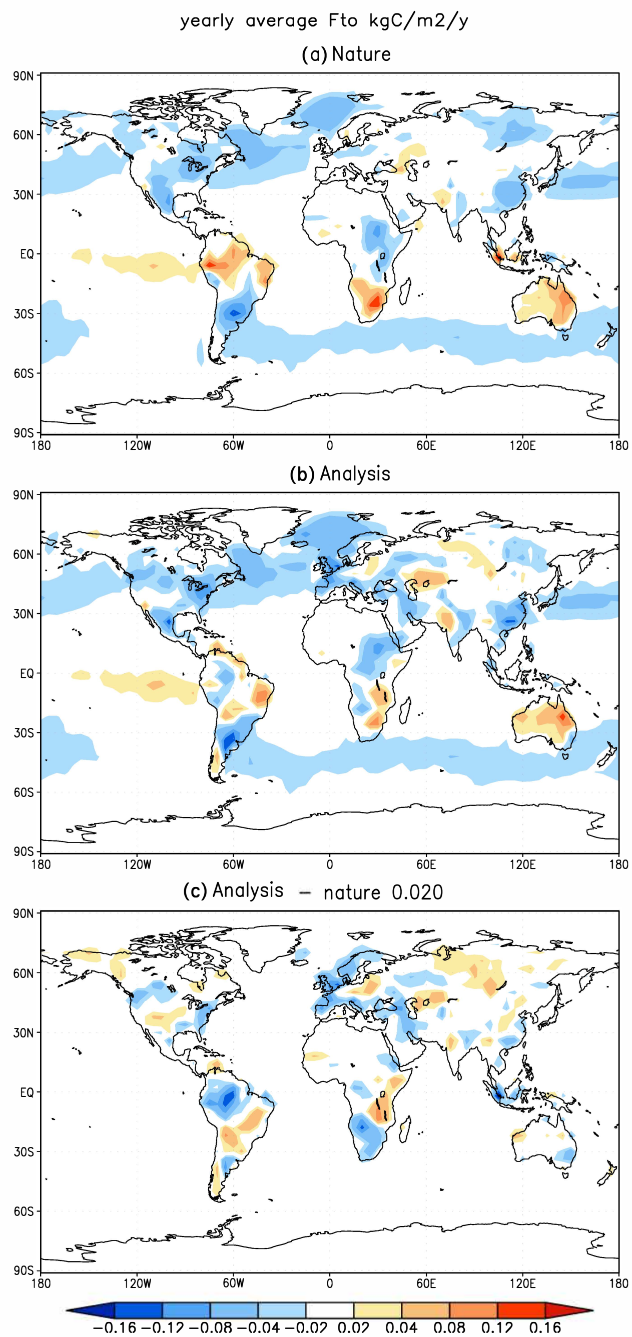
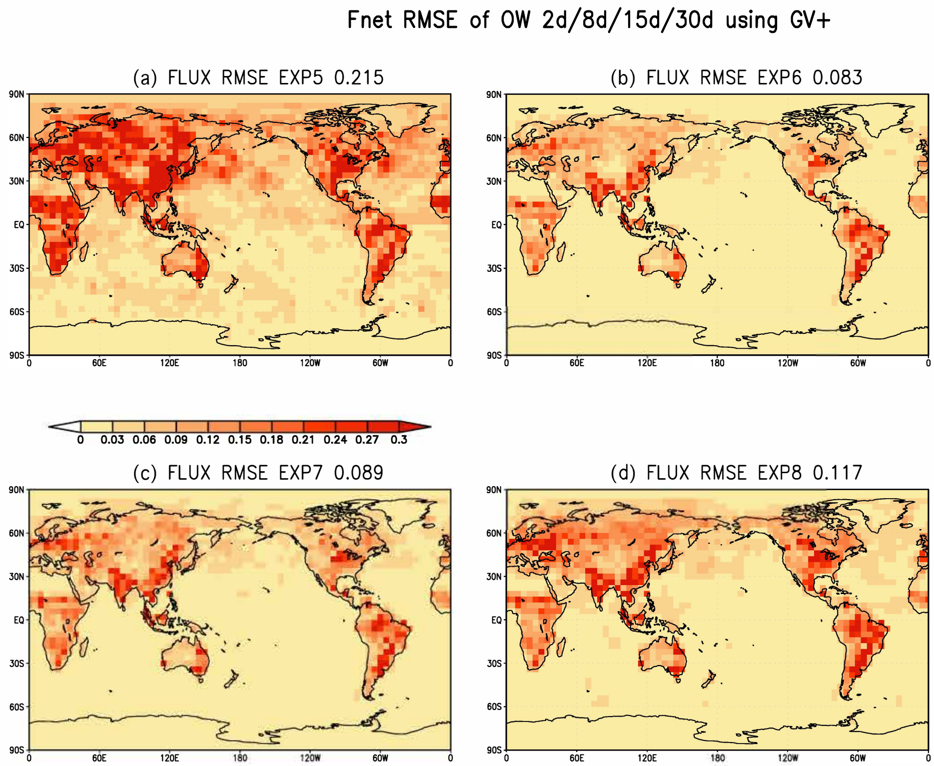


Figure 9. (a) the annual mean of SCF (with the FFE removed) for “nature” run; (b) the annual mean of estimated SCF (with the FFE removed) from benchmark experiment ; and (c) their differences.


Fnet RMSE of OW 2d/8d/15d/30d using GV+
(a) FLUX RMSE EXP5 0.215
(b) FLUX RMSE EXP6 0.083
(c) FLUX RMSE EXP7 0.089
(d) FLUX RMSE EXP8 0.117
0 0.03 0.06 0.09 0.12 0.15 0.18 0.21 0.24 0.27 0.3
90N 60N 30N EQ 30S 60S 90S
0 60E 120E 180 120W 60W 0

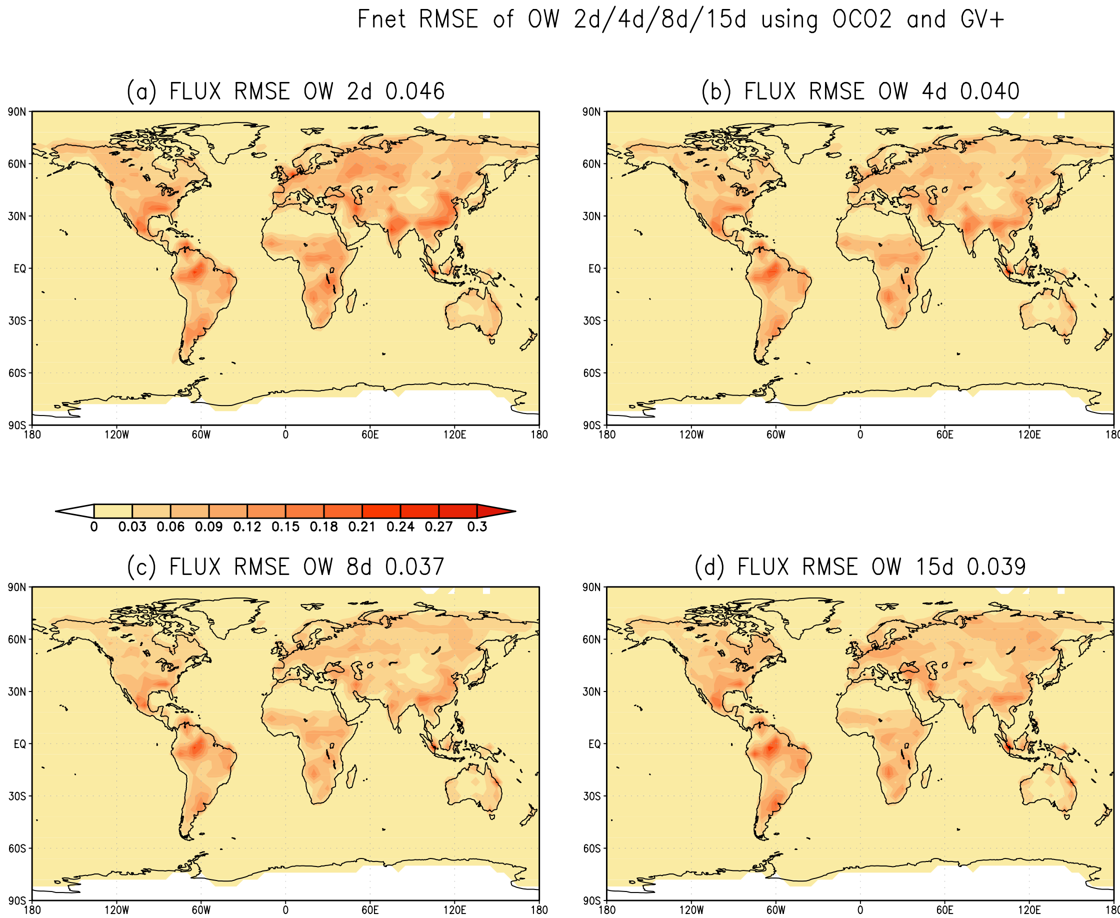


Figure 10. Same as Figure 5, except for ~~using~~ assimilating ~~the~~ both OCO-2 and GV+ Pseudo-Observations. The panel (a), (b), (c) and (d) show the results using OW of 2 days, 4 days, 8 days and 15 days respective.