



# Assimilating the dynamic spatial gradient of a bottom-up carbon flux estimation as a unique observation in COLA (v2.0)

Zhiqiang Liu<sup>1,2</sup>, Ning Zeng<sup>3,4,1</sup>, Yun Liu<sup>5</sup>, Eugenia Kalnay<sup>3</sup>, Ghassem Asrar<sup>6</sup>, Qixiang Cai<sup>1</sup>, Pengfei Han<sup>7,1</sup>

<sup>1</sup>Laboratory of Numerical Modeling for Atmospheric Sciences & Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing, China

<sup>3</sup>Dept. of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA

<sup>4</sup>Earth System Science Interdisciplinary Center, College Park, Maryland, USA

<sup>5</sup>Dept. of Oceanography, Texas A & M University, College Station, TX, USA

<sup>6</sup>Universities Space Research Association, Columbia, Maryland, USA

<sup>7</sup>Carbon Neutrality Research Center, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

*Corresponding authors:* Zhiqiang Liu (liuzhiqiang@mail.iap.ac.cn) and Yun Liu (liu6@tamu.edu)

**Abstract.** Atmospheric inversion of high spatiotemporal surface CO<sub>2</sub> flux without dynamic constraints and sufficient observations is an ill-posed problem, and a priori flux from a "bottom-up" estimation is commonly used in "top-down" inversion systems for regularization purposes. Ensemble Kalman filter-based inversion algorithms usually weigh a priori flux to the background or directly replace the background with the a priori flux. However, the "bottom-up" flux estimations, especially the simulated terrestrial-atmosphere CO<sub>2</sub> exchange, are usually systematically biased at different spatiotemporal scales because of the deficiencies in understanding of some underlying processes. Here, we introduced a novel regularization algorithm into the Carbon in Ocean–Land–Atmosphere (COLA) data assimilation system, which assimilates a priori information as a unique observation (AAPO). The a priori information is not limited to "bottom-up" flux estimation. With the comprehensive assimilation regularization approach, COLA can apply the spatial gradient of the "bottom-up" flux estimation as a priori information to reduce the bias impact and enhance the dynamic information concerning the a priori "bottom-up" flux estimation. Benefiting from the enhanced signal-to-noise ratio in the spatial gradient, the global, regional, and grided flux estimations using the AAPO algorithm are significantly better than those obtained by the traditional regularization approach, especially over highly uncertain tropical regions in the context of observing simulation system experiments (OSSEs). We suggest that the AAPO algorithm can be applied to other greenhouse gas (e.g., CH<sub>4</sub>, NO<sub>2</sub>) and pollutant data assimilation studies.

## 1 Introduction

Climate change forced by the increasing atmospheric carbon dioxide (CO<sub>2</sub>) concentrations threatens the health of the environment, living systems, and the economy. Therefore, it is essential to accurately estimate earth surface carbon fluxes (SCFs) and their variations for both scientific purposes and policy-making, including supporting the Paris Agreement (Byrne et al., 2022; Chevallier, 2021; Friedlingstein et al., 2022; Jiang et al., 2022; Weir et al., 2022; Deng et al., 2022). The SCFs



can be inferred from atmospheric CO<sub>2</sub> measurements using "top-down" (hereafter quotation marks will be omitted) techniques  
35 of the Bayesian synthesis (e.g., Rodenbeck et al., 2003; Zammit-Mangion et al., 2022; Cho et al., 2022) and data assimilation  
(DA) techniques (e.g., Peters et al., 2007; Feng et al., 2009; Chevallier et al., 2010; J. Liu et al., 2014; Z. Liu et al., 2022).  
However, the top-down estimation could be ill-posed because of the sparseness feature of atmospheric CO<sub>2</sub> observations and  
systematic errors of the transport model and satellite retrieval (Basu et al., 2018; O'Dell et al., 2018; Yu et al., 2018; Schuh et  
al., 2019). To regularize the ill-posed problem, a priori SCFs precalculated from "bottom-up" (hereafter quotation marks will  
40 be omitted) terrestrial models and ocean biogeochemical models are commonly applied (Baker et al., 2006; Peters et al., 2007).  
However, there are significant differences and biases among bottom-up SCFs due to the systematic deficiencies in current  
bottom-up models, which could contaminate the top-down SCF estimation (Philip et al., 2019; Fu et al., 2021).

Kang et al. (2011, 2012) developed a carbon data assimilation system by coupling an atmospheric general circulation model  
45 with a local ensemble transform Kalman filter (LETKF) (Hunt et al., 2007). This system accurately estimated SCF at model  
grid resolution for the first time without directly applying a priori information by assimilating meteorology observations and  
CO<sub>2</sub> concentration observations simultaneously with a short assimilation window in observing system simulation experiments  
(OSSEs). Following this track, we developed the Carbon in Ocean–Land–Atmosphere (COLA) system (Liu et al., 2019; Liu  
et al., 2022), which maintains the LETKF with a short assimilation window and long observation window setting but replaces  
50 the general circulation model with an atmospheric transport model, GEOS-Chem (Nassar et al., 2013). The COLA system also  
applies a constrained ensemble Kalman filter (CEnKF) technique to retain global mass conservation (Liu et al., 2022). As a  
result, COLA can provide good estimates of SCFs at model grid resolution along with atmospheric CO<sub>2</sub> concentration analyses.

On the other hand, even though a priori information includes biases, it could be used to further improve the SCF estimation in  
55 COLA because it includes important dynamic information generated by terrestrial models, which is missing in the top-down  
inversion system. Therefore, it is worth exploring an appropriate a priori regularization method. The purpose of this paper is  
to introduce the development of the COLA (v2.0) system with a novel a priori regularization for SCF estimation, where a priori  
information is treated as a unique observation to be assimilated. This new approach is inspired by the CEnKF applied in COLA  
(Liu et al., 2022). With the comprehensive assimilation regularization, we are able to apply the spatial gradient of bottom-up  
60 SCF estimations as a priori information to improve SCF estimation. The paper is organized as follows: section 2 introduces  
the COLA system and presents the new approach of a priori regularization for SCF estimation; section 3 compares a priori  
flux and its spatial gradient in the context of the signal/noise ratio and demonstrates and validates the new approach in the  
context of OSSEs; and section 4 provides the conclusion and discussion.

## 2 Data and Methods

### 65 2.1 Carbon in Ocean–Land–Atmosphere (COLA) DA system



The COLA DA system consists of an atmosphere transport model of GEOS-Chem, a local ensemble transform Kalman filter (LETKF) algorithm, a constrained EnKF, and the assimilated observations. Currently, we are using GEOS-Chem of version 13.0.2 driven by the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) meteorology reanalysis (Gelaro et al., 2017; The International GEOS-Chem User Community, 2021). GEOS-Chem requires the SCFs as boundary forcings to simulate the atmospheric CO<sub>2</sub> concentration. The SCFs are usually generated from bottom-up SCF estimations, including terrestrial carbon fluxes (F<sub>TA</sub>), terrestrial fire fluxes (F<sub>IR</sub>), air-sea carbon fluxes (F<sub>OA</sub>), and anthropogenic fossil fuel emissions (F<sub>FE</sub>).

The DA algorithm used in the COLA system is LETKF, which was created by Hunt et al. (2007). LETKF is a powerful, efficient deterministic EnKF variation. It is widely used for DA, including several operational centers, and was first used for carbon data assimilations by (Kang et al., 2011, 2012). LETKF estimates SCF as evolving parameters by augmenting it with the state vector CO<sub>2</sub>. Similar to the other EnKF, the LETKF prefers a short assimilation window to produce accurate model state analysis, which reduces noise within the background for parameter estimation. On the other hand, parameter estimation requires a long training period to enhance the model response to the estimated parameter (the signal). Therefore, COLA implements a new version of LETKF with a unique feature of a short assimilation window (1 day) and a long observation window (7 days) to enhance the SCF estimation (Liu et al., 2019).

Another unique feature of COLA is that it applies a constrained ensemble Kalman filter (CEnKF) to retain the global mass conservation of the system (Liu et al., 2022). The standard LETKF step updates the model state based on statistical information, which could lead to model dynamic imbalance and loss of mass (Pan and Wood, 2006; Zeng et al., 2017; Janjić and Zeng, 2021). With CEnKF, an additional assimilation step is applied to the first guess of the global CO<sub>2</sub> mass as the observation with zero uncertainty to adjust CO<sub>2</sub> at each model grid point, ensuring the consistency between the analysis and the first guess after the standard LETKF procedure.

## 2.2 Assimilating a priori information as unique observations (AAPO)

Conceptually, COLA can further improve the SCF estimation with a priori regularization using the a priori generated from an independent bottom-up estimation. COLA treats SCFs as stationary parameters, where the SCFs are only updated by the LETKF statistically. There is no dynamic constraint for SCF estimation, and any gap in data could be filled with a priori information from a bottom-up estimation that includes the important dynamic information about the parameter of interest. There are two widely used a priori regularization approaches for EnKF-based carbon inversion systems. Feng et al. (2009) treated the a priori information as a background for EnKF calculation,

$$\bar{f}_t^b = f_t^p, \quad (1)$$



where  $\bar{f}$  is the SCF; the bar denotes the ensemble mean; and the subscripts b, p, and t represent the background, a priori information of bottom-up estimation, and time, respectively. This approach omits useful information on the temporal dependency of SCFs, which is very important for resolving the subseasonal variation in SCFs, for example, in the COLA system with a short assimilation window of 1 day. Peters et al. (2007) treated the a priori information as anchoring values and averaged it with SCF analysis as,

$$\bar{f}_t^b = \alpha \cdot \bar{f}_{t-2}^a + \alpha \cdot \bar{f}_{t-1}^a + (1 - 2\alpha) \cdot f_t^p, \quad (2)$$

where  $\alpha$  denotes the average weight of the analysis ranging from 0 to 1, and the subscript a represents the analysis. The averaging coefficient of  $\alpha$  is ad hoc because it lacks a statistical foundation.

Here, we propose a new a priori regularization method to better follow the DA principle. The new approach treats the a priori information as unique observations to be assimilated into the COLA system in the LETKF analysis step (Fig. 1). Thus, the data include the atmospheric CO<sub>2</sub> observation and the a priori information,

$$\mathbf{y}^o = [\mathbf{y}_c^o, \mathbf{y}_{ap}^o], \quad (3)$$

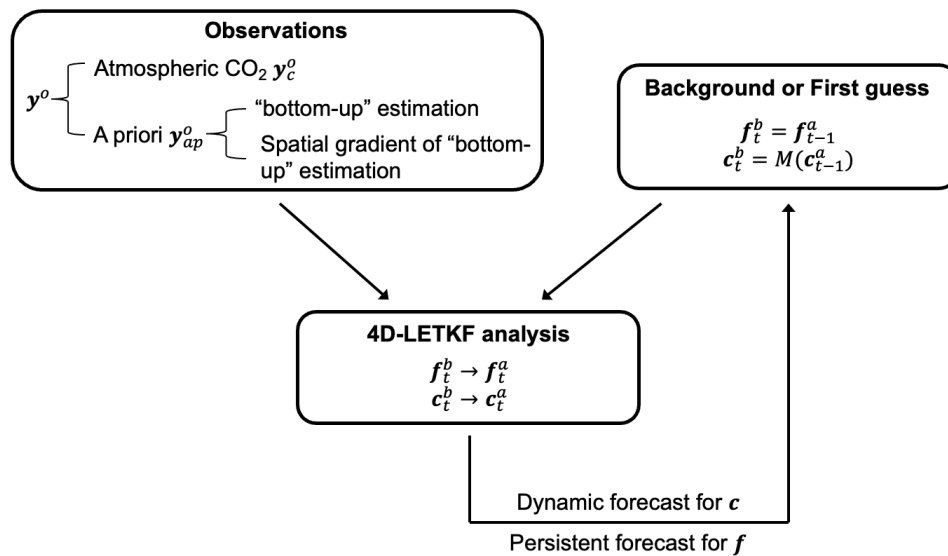
where  $\mathbf{y}^o$  is the observation; the subscripts c and ap denote the CO<sub>2</sub> observation and the a priori information, respectively. The new regularization algorithm follows a similar perspective as Peters et al. (2007) but with a comprehensive EnKF approach. The a priori information used for regularization is not limited to SCF estimations. EnKF updates the parameter using the covariance between the parameter and observation variables. Therefore, a priori information can be any estimation related to bottom-up SCFs. As discussed before, a bottom-up estimation of SCFs includes important dynamic information as well as significant biases. It will induce signal as well as noise when assimilated into the system. The impact of the signal must be larger than the impact of noise to obtain improved SCF estimation. Therefore, the estimation will benefit from the enhancement of the signal/noise ratio within the a priori information.

In COLA, the main purpose of applying a priori regularization is to introduce the dynamic constraint for SCF estimation. The spatial gradient of a bottom-up estimation of SCF could be a better choice than the SCF estimation itself as the a priori information. We also speculate that the signal-to-noise ratio within the spatial gradient is larger than the SCF estimation itself. For example, the SCFs in a terrestrial model can systematically drift in a regionally spatially coherent manner because they are affected by the same model dynamic/physiological deficiency with a similar terrestrial eco-environment. The spatial gradient calculation can reduce the systematic drift. On the other hand, the SCF spatial gradient reflects the dynamic response of the terrestrial and oceanic systems to large-scale climate forcing, which is the information we seek. The signal-to-noise ratio of SCF compared to its spatial gradient will be discussed in section 3.2.

Numerically, we define the SCF spatial gradient at a given grid as the SCF difference with its surrounding grids divided by the distance between them. Fig. 1 illustrates the assimilation steps of the COLA system. The atmospheric CO<sub>2</sub> observations and



the a priori SCF spatial gradient are assimilated at different LETKF steps to separate their impact, although both are treated as observations. In addition to the large-scale weather/climate forcing, the spatial gradient of the bottom-up SCF is also dependent on the local terrain, vegetation type, population, and urbanization. Therefore, the localization radius for the a priori SCF spatial gradient is set to a small value, and the SCF spatial gradient only constrains the local SCF.



**Figure 1: The assimilation cycle of the COLA system, illustrating how and where the a priori is assimilated.**

### 2.3 SCF ensemble spread versus a priori uncertainty

COLA is a flow-dependent ensemble-based DA using the ensemble spread to represent the forecast and analysis uncertainties. The SCFs are temporally varying parameters to be estimated by the COLA. With a default setting, an estimated parameter is set as stationary during model integration (forecast step) and only updated by the DA algorithm in the analysis steps. As a result, the parameter ensemble spread remains unchanged during the forecast step and is reduced in the analysis step. This reduction leads to a progressively decreasing parameter ensemble spread, which eventually could result in a too-small parameter ensemble spread and filter divergence for parameter estimation. Furthermore, this parameter estimation procedure totally ignores the uncertainties related to the temporal variations in SCFs, thus further deteriorating the negative ensemble spread issue. Following Kang et al. (2012), we apply an additive inflation method to the SCFs to overcome this problem. This inflation method adds to each ensemble member with an anomalous SCF field chosen randomly from the bottom-up estimation of CASA within 30 days centered at the analysis time (Kang et al., 2012; Liu et al., 2019, 2022).

COLA assimilates the a priori SCF spatial gradients into the system, which needs to define the a priori uncertainty. In this study, we simply set the a priori uncertainty proportional to the uncertainty of the analysis ensemble uncertainty,



$$\sigma_{i,t}^{\text{ap}} = \alpha \cdot \sigma_{i,t}^{\text{a}}, \quad (4)$$

where  $\sigma_{i,t}^{\text{ap}}$  is the a priori uncertainty of the bottom-up estimation or the spatial gradient of the bottom-up estimation, and  $\alpha$  is the scaling factor. This approach is a simple adjustment for the a priori uncertainty. We find an optimal scaling factor of 5 based on several tests. In reality, a bottom-up SCF estimation product may come with its uncertainty estimation. We may derive the uncertainty of the SCF spatial gradient from it. The importance and impact of those uncertainties and whether their accuracies are good enough for DA application remain to be further explored in the future.

### 3 Observing system simulation experiments (OSSEs)

#### 3.1 Experimental setup

In this section, we evaluated the new algorithm in the context of OSSEs. The experimental period spans from 1 September 2014 to the end of 2015, with a spin-up period in 2014. GEOS-Chem is running with a horizontal resolution of  $4^\circ \times 5^\circ$  and 47 hybrid pressure-sigma vertical levels for  $\text{CO}_2$  simulation. We assimilate both the surface and satellite data as described in Liu et al. (2022). The surface data are obtained from the  $\text{CO}_2$  GLOBALVIEWplus v8.0 ObsPack. For satellite data, we used a 10-second averaged ACOS v10 level two retrieval of land-nadir and land-glint from the Orbiting Carbon Observatory-2 (OCO-2) (Crisp et al., 2017; O'Dell et al., 2018; Baker et al., 2021; Cox et al., 2022). Following Liu et al. (2022), the observation networks and their error scales are used to create the pseudo observations in the OSSEs. We set the  $\text{CO}_2$  observation localization radius to 4000 kilometers.

A nature run is driven by the  $F_{\text{OA}}$  from Rödenbeck et al. (2014),  $F_{\text{FE}}$  from the Open-source Data Inventory of Anthropogenic  $\text{CO}_2$  emissions (ODIAC) (Oda et al., 2018), and the  $F_{\text{IR}}$  and  $F_{\text{TA}}$  generated from the terrestrial model of the VEGAS model (Zeng et al., 2005), which serves as the "truth" for the OSSE. The  $F_{\text{TA}}$  is chosen as the flux to be estimated. The other fluxes of  $F_{\text{IR}}$ ,  $F_{\text{FE}}$ , and  $F_{\text{OA}}$  are set to be identical in the nature run. Another terrestrial model of CASA (Potter & Klooster, 1997) is used to provide an independent bottom-up  $F_{\text{TA}}$  estimation for the a priori regularization in our experiments. The difference between the  $F_{\text{TA}}$  from VEGAS and CASA represents the bias among bottom-up estimations related to terrestrial model deficiencies.

#### 3.2 Signal-to-noise ratio (SNR) analysis

In this section, we mimic the typical bias of a priori information and derive its signal-to-noise ratio (SNR) based on two different terrestrial model estimations of the VEGAS and CASA. The estimations of VEGAS and CASA are treated as the truth and the a priori information, respectively. The variations in the truth are regarded as the signal. The differences between the "truth" and the a priori information are regarded as the noise. We define the SNR of the spatial gradient and the  $F_{\text{TA}}$  as,



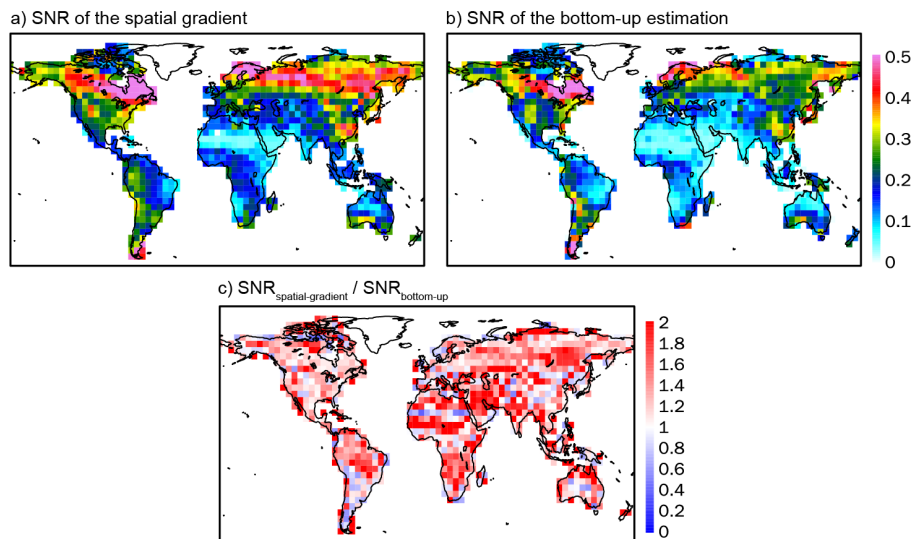
$$\text{SNR}^{\text{SG}}(x, t) = \frac{|\overline{f^{\text{TSG}}(x, t)} - \overline{f^{\text{TSG}}(x, t-1)}|}{|\overline{f^{\text{PSG}}(x, t)} - \overline{f^{\text{TSG}}(x, t)}|}, \quad (5)$$

$$\text{SNR}^{\text{S}}(x, t) = \frac{|f^{\text{T}}(x, t) - f^{\text{T}}(x, t-1)|}{|f^{\text{P}}(x, t) - f^{\text{T}}(x, t)|}, \quad (6)$$

where the superscripts PSG and P denote the a priori of SCF spatial gradient and SCF, respectively; the superscripts TSG and T denote the truth of SCF spatial gradient and  $F_{\text{TA}}$ , respectively; and  $x$  denotes the grid point location at time  $t$ .

Both  $\text{SNR}^{\text{SG}}$  and  $\text{SNR}^{\text{S}}$  over the Northern Hemisphere are larger than those over the tropics and Southern Hemisphere (Fig. 2a, b), which is consistent with the larger  $F_{\text{TA}}$  seasonal cycle over the Northern Hemisphere than over the tropics and Southern Hemisphere. The  $\text{SNR}^{\text{SG}}$  is significantly greater than  $\text{SNR}^{\text{S}}$  nearly everywhere (Fig. 2c), indicating that the spatial gradient is better than SCF itself in terms of SNR. The  $\text{SNR}^{\text{SG}}$  is almost double the corresponding  $\text{SNR}^{\text{S}}$  in the Sahel, southern Africa, tropical South America, and eastern Siberia, where the noise of the a priori  $F_{\text{TA}}$  is significantly larger than that in the other areas (Fig. 4a).

Here, we use VEGAS and CASA estimations as examples to reveal that the spatial gradient has a larger SNR than the  $F_{\text{TA}}$  itself. The same comparison can be made based on other terrestrial model estimations. Conceptually, DA induces both signal and noise into a system by assimilating observations with errors. A DA application expects the impact of the signal to dominate and offset the impact of noise. Therefore, it tends to achieve better analysis by increasing the SNR of assimilated observations. Thus, we speculate that the spatial gradients are better than the flux itself as a priori information. Since the current understanding of the carbon cycle processes is highly uncertain in the tropics and Eurasia (O'Sullivan et al., 2022), the two models we chose differ significantly in these areas, which is a good proxy for the real-world scenario.



**Figure 2:** a) The annual mean signal-to-noise ratio pattern of the a priori  $F_{\text{TA}}$  spatial gradient. b) The same as a) but

for the a priori  $F_{TA}$ . c) Annual mean signal-to-noise ratio of the spatial gradient divided by the signal-to-noise ratio of the  $F_{TA}$ .

### 3.3 OSSE results

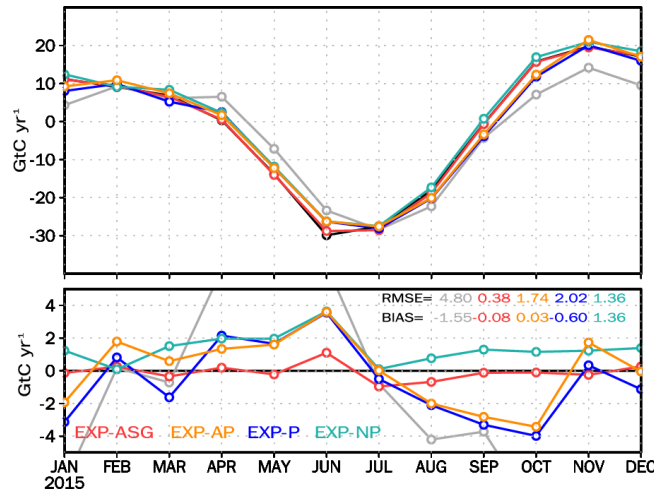
Four experiments are performed with COLA in the context of OSSE to evaluate the proposed a priori regularization approach. The first experiment estimates the  $F_{TA}$  by only assimilating the synthetic observations of the atmospheric  $CO_2$  concentration (EXP-NP). Based on the first experiment, experiment two (EXP-P) uses the approach described in Eq. (2) with a priori  $F_{TA}$  from CASA and  $\alpha$  of 0.2; experiment three (EXP-ASG) applies the proposed assimilation approach with a priori  $F_{TA}$  spatial gradient in the CASA estimation; experiment four (EXP-AP) assimilates the a priori  $F_{TA}$  of the CASA estimation. We compare EXP-ASG with the other three experiments to illustrate the improvement of SCF estimation. The evaluation metrics of root-mean-square error (RMSE) and mean bias are defined as,

$$RMSE^A(x) = \sqrt{\frac{\sum_{t=t_1}^{t_2} (f^A(x,t) - f^T(x,t))^2}{t_2 - t_1}}, \quad (7)$$

$$RMSE^{ASG}(x) = \sqrt{\frac{\sum_{t=t_1}^{t_2} (|f^{ASG}(x,t) - f^{TSG}(x,t-1)|)^2}{t_2 - t_1}}, \quad (8)$$

$$MB^A(i) = \frac{\sum_{t=t_1}^{t_2} (f^A(i,t) - f^T(i,t))}{t_2 - t_1}, \quad (9)$$

where  $x$  denotes a grid point; the superscripts A and T denote the analysis of  $F_{TA}$  and truth, respectively; ASG denotes the analysis of the spatial gradient of  $F_{TA}$ ; all are calculated for the time from  $t_1$  and  $t_2$ .



**Figure 3:** The top figure is the global total seasonal cycle of the truth (black), the a priori  $F_{TA}$  (gray), and the four assimilation experiments. The bottom figure is the global total difference of the four assimilation experiments compared to the truth. The annual mean RMSE and mean biases of the a priori  $F_{TA}$  and the four experiments are denoted at the upper right corner of the bottom figure.

First, we focus on the global total  $F_{TA}$  (Fig. 3). The estimated errors are shown in annual mean flux biases and the RMSE of the seasonal cycle. The flux biases of EXP-P and EXP-AP are  $-0.60 \text{ GtC yr}^{-1}$  and  $0.03 \text{ GtC yr}^{-1}$ , respectively, which are



significantly smaller than those of EXP-NP ( $1.36 \text{ GtC yr}^{-1}$ ). The improvement is mainly due to the negative bias of  $F_{TA}$  from CASA compromising the positive bias of the estimates in EXP-NP without a priori regularization. The EXP-ASG removes most of this bias with only  $0.08 \text{ GtC yr}^{-1}$  remaining.

230 The RMSE of  $F_{TA}$  in EXP-ASG is significantly smaller than those with other experiments. Instead of improving the estimation, the a priori regularization with the  $F_{TA}$  from CASA degrades the seasonal cycle estimation of the global  $F_{TA}$ . Both EXP-P and EXP-AP show an increase in RMSE of  $F_{TA}$  concerning EXP-NP because of the very large seasonal cycle errors within the a priori  $F_{TA}$  used for regularization (Fig. 3). The new AAPO regularization is clearly better than the traditional method even with the same biases for the a priori  $F_{TA}$  from CASA. The bias and RMSE for EXP-AP are reduced by 66% and 24% compared to  
 235 EXP-P, respectively. The AAPO regularization with an improved a priori  $F_{TA}$  spatial gradient further improved the estimation. The bias and RMSE for EXP-ASG are reduced by 87% and 81% compared to EXP-P, respectively.

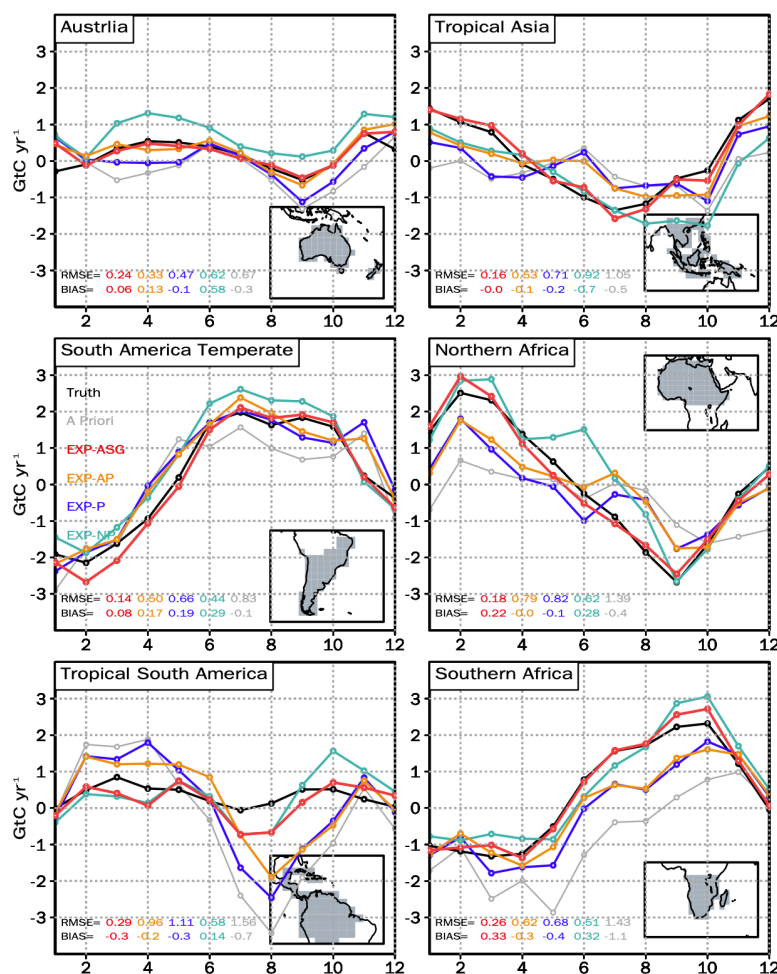


Figure 4: The seasonal cycles of the regional  $F_{TA}$  for truth (black), a priori (gray), EXP-ASG (red), EXP-AP (orange),



240 **EXP-P (blue), and EXP-NP (green) in the tropics and Australia. The values in the bottom-left corner of all figures are the RMSE, the mean bias of the four experiments, and the a priori  $F_{TA}$  compared to the truth  $F_{TA}$ . The regions are defined based on the OCO2MIP mask (Crowell et al., 2019).**

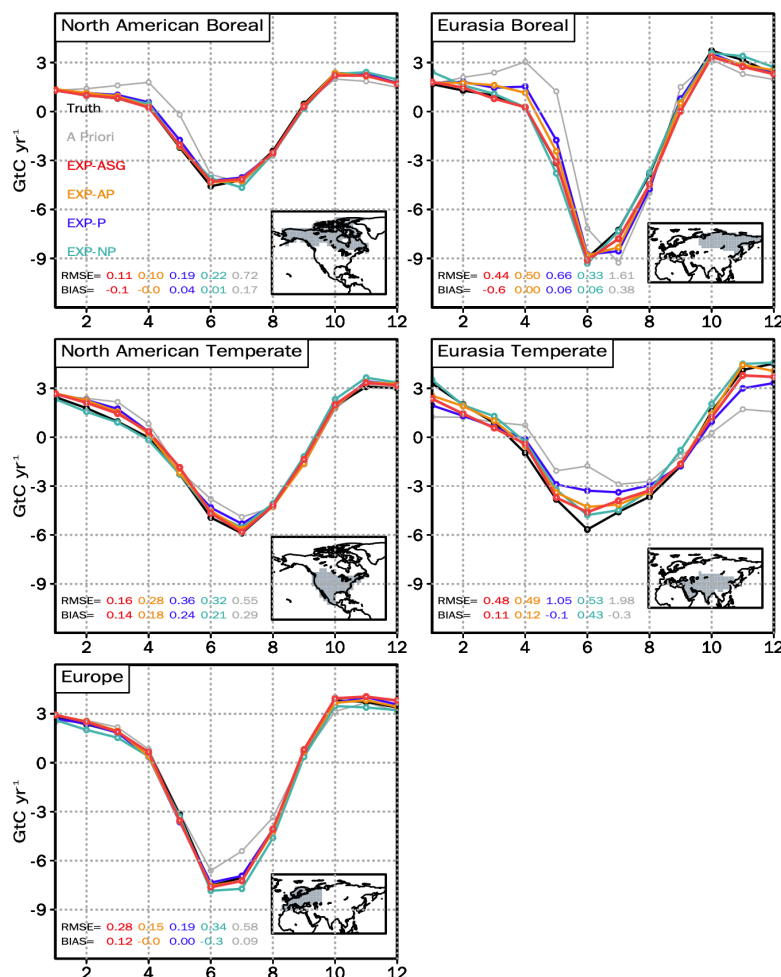
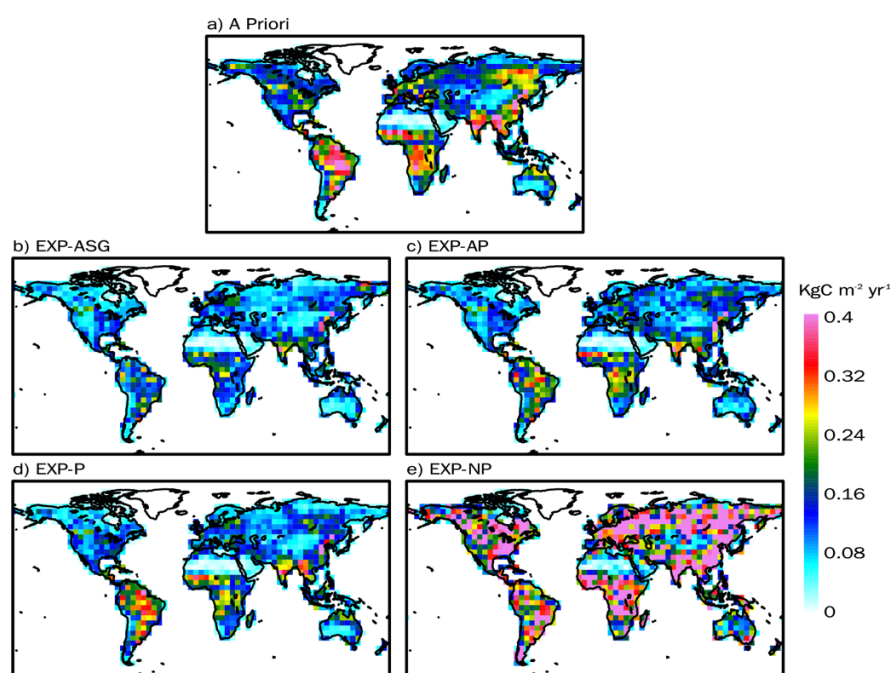


Figure 5: Same as in Fig. 4, but for the Northern Hemisphere regions.

At the regional scale, defined based on the OCO2MIP mask (Crowell et al., 2019), all four experiments reproduce the Northern Hemisphere seasonal cycle accurately, which is mainly because of the better observation coverage and less biased a priori  $F_{TA}$  as compared to the Southern Hemisphere regions (Fig. 5). In the Southern Hemisphere, significant differences exist between the "truth"  $F_{TA}$  and the a priori  $F_{TA}$  from CASA. The differences are larger than the RMSE of the estimated  $F_{TA}$  of EXP-NP for most regions, especially South America Temperate, Tropic South America, Northern Africa, and Southern Africa, where a priori  $F_{TA}$  degrades the  $F_{TA}$  estimation with the RMSEs increasing for EXP-P and EXP-AP compared to EXP-NP. The seasonal estimation of EXP-ASG is significantly better than the other experiments in terms of RMSE (Fig. 4). The regional seasonal estimation of EXP-AP is better than that of EXP-P, which indicates that the AAPO regularization is better than the traditional



method, and this is consistent with the global seasonal results presented earlier. In tropical South America, the seasonal amplitude of the a priori  $F_{TA}$  is significantly larger than the truth of  $F_{TA}$ , which greatly contaminates the estimation in EXP-AP and EXP-P. A similar phenomenon also appears in some other regions (e.g., northern and southern Africa) where the a priori and the truth of  $F_{TA}$  show a large discrepancy. However, the a priori  $F_{TA}$  spatial gradient can help separate the biased information of the  $F_{TA}$  and significantly improve the seasonal estimation in regions where the a priori  $F_{TA}$  is not good and observations are rare.

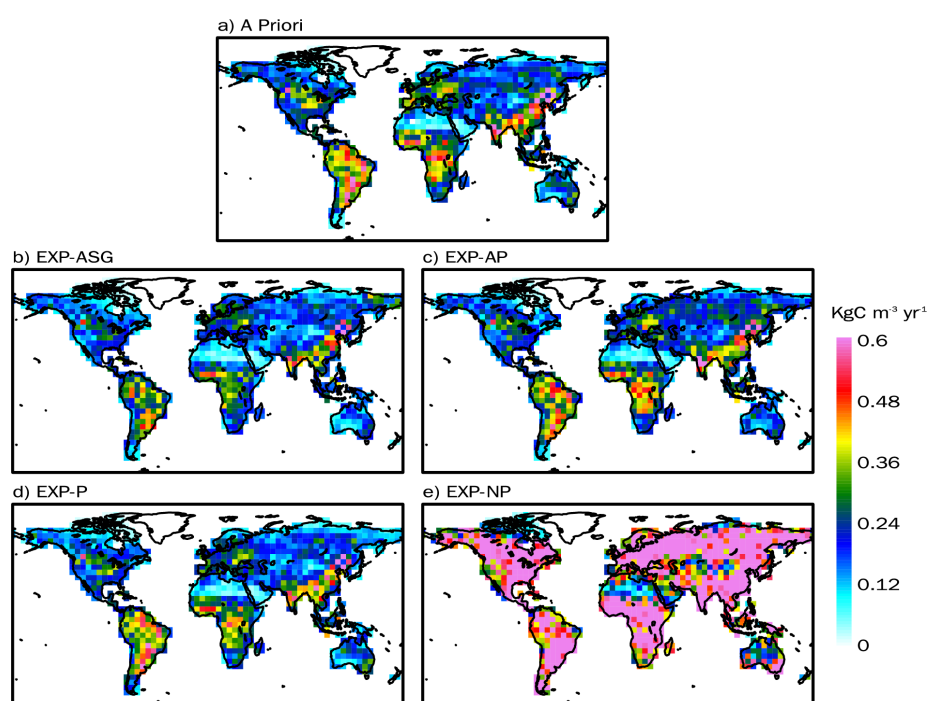


**Figure 6: The RMSE of the  $F_{TA}$  in (a) the a priori, (b) EXP-ASG, (c) EXP-AP, (d) EXP-P, and (e) EXP-NP calculated based on Equation 7 from January to December 2015.**

At the grid-point scale, EXP-NP produces a very noisy  $F_{TA}$  with abnormally large RMSEs of the estimated  $F_{TA}$  because LETKF overfits the gridded  $F_{TA}$  with limited observations (Fig. 6e). Therefore, a priori regularization is expected to enhance the results for COLA. The RMSEs in EXP-P, EXP-AP, and EXP-ASG are significantly reduced compared to EXP-NP. Large errors exist in the a priori  $F_{TA}$  over the South American, African, and Indian regions (Fig. 6a). When the a priori  $F_{TA}$  is used for regularization, those errors are also introduced into the system resulting in the final  $F_{TA}$  estimation with a large RMSE shown in the same locations, especially for EXP-P (Fig. 6b, c). Those errors are less severe within EXP-AP with the AAPO regularization applied concerning EXP-P. The use of AAPO with an improved a priori  $F_{TA}$  spatial gradient avoids those large errors within the a priori  $F_{TA}$  and produces much better estimation in EXP-ASG. At the model grid scale, EXP-ASG significantly outperformed EXP-AP and EXP-P for most grid points.



In addition to the RMSE of  $F_{TA}$ , we also analyzed the RMSE of the  $F_{TA}$  spatial gradient. For EXP-AP and EXP-P, which assimilate or add the a priori  $F_{TA}$ , the spatial gradient information is digested simultaneously within the original  $F_{TA}$ . Similar to the RMSE of  $F_{TA}$ , EXP-NP also produced very noisy spatial gradient estimation (Fig. 7e). With the help of the two types of a priori information, the other three experiments significantly reduced the RMSE concerning EXP-NP (Fig. 7b-d). However, for EXP-ASG, which directly assimilates the spatial gradient information, its RMSE is smaller than that of EXP-AP and EXP-P. Moreover, the RMSEs of EXP-AP and EXP-P are almost identical to the RMSE of the a priori  $F_{TA}$  spatial gradient at most grid points, indicating that the second-order spatial gradient information is not effectively digested in the two experiments and that the first-order biases in the original a priori  $F_{TA}$  dominate the resulting information.



**Figure 7: The RMSE of the spatial gradient of  $F_{TA}$  in (a) the a priori, (b) EXP-ASG, (c) EXP-AP, (d) EXP-P, and (e) EXP-NP calculated based on Equation 8 from January to December 2015.**

#### 4 Summary and Discussion

In this study, we developed a novel algorithm for the ensemble-based  $\text{CO}_2$  inversion system, in which the spatial gradient of a bottom-up model estimation is dynamically assimilated as a unique observation (AAPO). The AAPO algorithm that uses the spatial gradient of a bottom-up model estimation as the a priori information aims at separating the first-order systematic biases in the bottom-up model estimations out of the inversion framework from a comprehensive DA perspective. In the context of OSSEs that assimilate in-situ and OCO-2 land-nadir and land-glint observations, it significantly overperformed the traditional



scheme that directly adds the a priori flux to the first guess. As a result, the spatial gradient consideration helps improve the accuracy of regional flux estimation, especially over the highly uncertain tropics.

EnKF is widely used in state data assimilation studies, such as weather forecasting, for its flexibility, efficiency, and error transport feature. However, the advantage of error transport is partly sacrificed or abandoned by introducing the a priori flux information to the background in most of the EnKF-based CO<sub>2</sub> inversion methods (Peters et al., 2007; Feng et al., 2009). This is because of the loss of a dynamic model to provide the background and the background covariance estimations. Different from most EnKF-based systems, COLA maintains the mean and error transport advantages of the EnKF by including the dynamic information constraints of the a priori flux spatial gradient and using an additive covariance inflation method (Liu et al., 2022).

In addition to the CO<sub>2</sub> inversion problem investigated in this study, the proposed new assimilation algorithm can also be applied to ensemble-based source/sink inversion studies of methane, nitrogen dioxide, and other chemical species. Based on the concept of applying constraints as unique observations, the spatial gradient constraint can be introduced to variational methods. With the increasing number and spatial coverage of CO<sub>2</sub> observations and the improving accuracy of atmospheric transport and satellite retrieval algorithms, the dependency on the a priori estimation is expected to be further reduced. The proposed new approach in this study offers a unique strategy and a new approach for improving the estimation of geophysical parameters and greenhouse gas fluxes, especially for observation- and understanding-limited regions of the world.

*Code and data availability.* The code for the new AAPO regularization scheme can be accessed at <https://doi.org/10.5281/zenodo.7592827>.

*Author contributions.* ZL and YL conceived the a priori regularization scheme. ZL, YL, and NZ designed the OSSEs. ZL wrote the code and ran the OSSEs. ZL, NZ, YL, and EK developed the system. QC supplied the VEGAS model output. ZL, YL, NZ, and GA wrote the paper. All authors contributed to the preparation of this paper.

*Acknowledgments.* We thank the OCO-2 science team and the NOAA greenhouse gas team for providing the CO<sub>2</sub> observations. Special thanks to Dr. David Baker for processing the 10-second averaged OCO-2 retrievals and Drs. Sourish Basu and Tom Oda for providing the hourly ODIAC data. We are also grateful to the OCO-2 Flux MIP colleagues for their valuable feedback and discussion of COLA and CEnKF.

*Financial support.* This work was supported by the National Key R&D Program of China (No. 2017YFB0504000).



## References

- Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO<sub>2</sub>, *Tellus B Chem. Phys. Meteorol.*, 58, 359–365, <https://doi.org/10.1111/j.1600-0889.2006.00218.x>, 2006.
- 325 Baker, D. F., Bell, E., Davis, K. J., Campbell, J. F., Lin, B., and Dobler, J.: A new exponentially-decaying error correlation model for assimilating OCO-2 column-average CO<sub>2</sub> data, using a length scale computed from airborne lidar measurements, *Geosci Model Dev Discuss*, 29, <https://doi.org/10.5194/gmd-2020-444>, 2021.
- Basu, S., Baker, D. F., Chevallier, F., Patra, P. K., Liu, J., and Miller, J. B.: The impact of transport model differences on CO<sub>2</sub> surface flux estimates from OCO-2 retrievals of column average CO<sub>2</sub>, *Atmospheric Chem. Phys.*, 18, 7189–7215, <https://doi.org/10.5194/acp-18-7189-2018>, 2018.
- 330 Byrne, B., Baker, D. F., Basu, S., Bertolacci, M., Bowman, K. W., Carroll, D., Chatterjee, A., Chevallier, F., Ciais, P., Cressie, N., Crisp, D., Crowell, S., Deng, F., Deng, Z., Deutscher, N. M., Dubey, M., Feng, S., García, O., Griffith, D. W. T., Herkommer, B., Hu, L., Jacobson, A. R., Janardanan, R., Jeong, S., Johnson, M. S., Jones, D. B. A., Kivi, R., Liu, J., Liu, Z., Maksyutov, S., Miller, J. B., Miller, S. M., Morino, I., Notholt, J., Oda, T., O'Dell, C. W., Oh, Y.-S., Ohyama, H., Patra, P. K., Peiro, H., Petri, C., Philip, S., Pollard, D. F., Poulter, B., Remaud, M., Schuh, A., Sha, M. K., Shiomi, K., Strong, K., Sweeney, C., Té, Y., Tian, H., Velasco, V. A., Vrekoussis, M., Warneke, T., Worden, J. R., Wunch, D., Yao, Y., Yun, J., Zammit-Mangion, A., and Zeng, N.: National CO<sub>2</sub> budgets (2015–2020) inferred from atmospheric CO<sub>2</sub> observations in support of the Global Stocktake, *Earth Syst. Sci. Data Discuss*, <https://doi.org/10.5194/essd-2022-213>, 2022.
- Chevallier, F.: Fluxes of Carbon Dioxide From Managed Ecosystems Estimated by National Inventories Compared to  
 340 Atmospheric Inverse Modeling, *Geophys. Res. Lett.*, 48, <https://doi.org/10.1029/2021GL093565>, 2021.
- Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morgui, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO<sub>2</sub> surface fluxes at grid point scale  
 345 estimated from a global 21 year reanalysis of atmospheric measurements, *J. Geophys. Res.*, 115, D21307, <https://doi.org/10.1029/2010JD013887>, 2010.
- Cho, T., Chung, J., Miller, S. M., and Saibaba, A. K.: Computationally efficient methods for large-scale atmospheric inverse modeling, *Geosci. Model Dev.*, 15, 5547–5565, <https://doi.org/10.5194/gmd-15-5547-2022>, 2022.
- Cox, A., Di Sarra, A. G., Vermeulen, A., Manning, A., Beyersdorf, A., Zahn, A., Manning, A., Watson, A., Karion, A.,  
 350 Hoheisel, A., Leskinen, A., Hensen, A., Arlyn Andrews, Jordan, A., Frumau, A., Colomb, A., Scheeren, B., Law, B., Baier, B., Munger, B., Paplawsky, B., Viner, B., Stephens, B., Daube, B., Labuschagne, C., Myhre, C. L., Couret, C., Hanson, C., Miller, C. E., Lunder, C. R., Plass-Duelmer, C., Plass-Duelmer, C., Gerbig, C., Sloop, C. D., Sweeney, C., Kubistin, D., Goto, D., Jaffe, D., Heltai, D., Van Dinter, D., Bowling, D., Lam, D. H. Y., Munro, D., Dickon Young, Worthy, D., Dlugokencky, E., Kozlova, E., Gloor, E., Cuevas, E., Reyes-Sanchez, E., Hintsa, E., Kort, E., Morgan, E., Obersteiner, F.,



- 355 Apadula, F., Francois Gheusi, Meinhardt, F., Moore, F., Vitkova, G., Chen, G., Bentz, G., Giordane A. Martins, Manca, G.,  
 Brailsford, G., Forster, G., Boenisch, H., Riris, H., Meijer, H., Moossen, H., Timas, H., Matsueda, H., Huilin Chen, Levin,  
 I., Lehner, I., Mammarella, I., Bartyzel, J., Abshire, J. B., Elkins, J. W., Levula, J., Jaroslaw Necki, Pichon, J. M., Peischl,  
 J., Müller-Williams, J., Turnbull, J., Miller, J. B., Lee, J., Lin, J., Jooil Kim, Josep-Anton Morgui, Pitt, J., DiGangi, J. P.,  
 Lavric, J., Hatakka, J., Coletta, J. D., Worsley, J., Holst, J., Lehtinen, K., Kominkova, K., McKain, K., et al.: Multi-laboratory  
 360 compilation of atmospheric carbon dioxide data for the period 1957-2021;  
 obspack\_co2\_1\_GLOBALVIEWplus\_v8.0\_2022-08-27, , <https://doi.org/10.25925/20220808>, 2022.
- Crisp, D., Pollock, H. R., Rosenberg, R., Chapsky, L., Lee, R. A. M., Oyafulo, F. A., Frankenberg, C., O'Dell, C. W., Bruegge,  
 C. J., Doran, G. B., Eldering, A., Fisher, B. M., Fu, D., Gunson, M. R., Mandrake, L., Osterman, G. B., Schwandner, F. M.,  
 Sun, K., Taylor, T. E., Wennberg, P. O., and Wunch, D.: The on-orbit performance of the Orbiting Carbon Observatory-2  
 365 (OCO-2) instrument and its radiometrically calibrated products, *Atmospheric Meas. Tech.*, 10, 59–81,  
<https://doi.org/10.5194/amt-10-59-2017>, 2017.
- Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., Liu, J., Deng, F., Feng, L., McKain, K., Chatterjee,  
 A., Miller, J. B., Stephens, B. B., Eldering, A., Crisp, D., Schimel, D., Nassar, R., O'Dell, C. W., Oda, T., Sweeney, C.,  
 Palmer, P. I., and Jones, D. B. A.: The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network,  
 370 *Atmospheric Chem. Phys.*, 19, 9797–9831, <https://doi.org/10.5194/acp-19-9797-2019>, 2019.
- Deng, Z., Ciais, P., Tzompa-Sosa, Z. A., Saunio, M., Qiu, C., Tan, C., Sun, T., Ke, P., Cui, Y., Tanaka, K., Lin, X., Thompson,  
 R. L., Tian, H., Yao, Y., Huang, Y., Lauerwald, R., Jain, A. K., Xu, X., Bastos, A., Sitch, S., Palmer, P. I., Lauvaux, T.,  
 d'Aspremont, A., Giron, C., Benoit, A., Poulter, B., Chang, J., Petrescu, A. M. R., Davis, S. J., Liu, Z., Grassi, G., Albergel,  
 C., Tubiello, F. N., Perugini, L., Peters, W., and Chevallier, F.: Comparing national greenhouse gas budgets reported in  
 375 UNFCCC inventories against atmospheric inversions, *Earth Syst. Sci. Data*, 14, 1639–1675, <https://doi.org/10.5194/essd-14-1639-2022>, 2022.
- Feng, L., Palmer, P. I., Bosch, H., and Dance, S.: Estimating surface CO<sub>2</sub> fluxes from space-borne CO<sub>2</sub> dry air mole fraction  
 observations using an ensemble Kalman Filter, *Atmospheric Chem. Phys.*, 9, 2619–2633, 2009.
- Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quéré, C., Peters, G. P.,  
 380 Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker,  
 M., Bellouin, N., Bopp, L., Chau, T. T. T., Chevallier, F., Chini, L. P., Cronin, M., Currie, K. I., Decharme, B., Djeutchouang,  
 L. M., Dou, X., Evans, W., Feely, R. A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber,  
 N., Gürses, Ö., Harris, I., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Lujikx, I. T., Jain, A., Jones, S. D., Kato, E.,  
 Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre,  
 385 N., Lienert, S., Liu, J., Marland, G., McGuire, P. C., Melton, J. R., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa,  
 Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J.,  
 Schwingshackl, C., Séférian, R., Sutton, A. J., Sweeney, C., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F., van  
 der Werf, G. R., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, C.,



- Yue, X., Zaehle, S., and Zeng, J.: Global Carbon Budget 2021, *Earth Syst. Sci. Data*, 14, 1917–2005,  
 390 <https://doi.org/10.5194/essd-14-1917-2022>, 2022.
- Fu, Y., Liao, H., Tian, X., Gao, H., Jia, B., and Han, R.: Impact of Prior Terrestrial Carbon Fluxes on Simulations of  
 Atmospheric CO<sub>2</sub> Concentrations, *J. Geophys. Res. Atmospheres*, 126, <https://doi.org/10.1029/2021JD034794>, 2021.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G.,  
 Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W.,  
 395 Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M.,  
 Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications,  
 Version 2 (MERRA-2), *J. Clim.*, 30, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>, 2017.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform  
 Kalman filter, *Phys. Nonlinear Phenom.*, 230, 112–126, <https://doi.org/10.1016/j.physd.2006.11.008>, 2007.
- 400 Janjić, T. and Zeng, Y.: Weakly Constrained LETKF for Estimation of Hydrometeor Variables in Convective-Scale Data  
 Assimilation, *Geophys. Res. Lett.*, 48, <https://doi.org/10.1029/2021GL094962>, 2021.
- Jiang, F., He, W., Ju, W., Wang, H., Wu, M., Wang, J., Feng, S., Zhang, L., and Chen, J. M.: The status of carbon neutrality  
 of the world's top 5 CO<sub>2</sub> emitters as seen by carbon satellites, *Fundam. Res.*, 2, 357–366,  
<https://doi.org/10.1016/j.fmre.2022.02.001>, 2022.
- 405 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, <https://doi.org/10.1029/2010JD014673>,  
 2011.
- Kang, J.-S., Kalnay, E., Miyoshi, T., Liu, J., and Fung, I.: Estimation of surface carbon fluxes with an advanced data  
 assimilation methodology, *J. Geophys. Res. Atmospheres*, 117, D24101, <https://doi.org/10.1029/2012JD018259>, 2012.
- 410 Liu, J., Bowman, K. W., Lee, M., Henze, D. K., Bousserez, N., Brix, H., James Collatz, G., Menemenlis, D., Ott, L., Pawson,  
 S., Jones, D., and Nassar, R.: Carbon monitoring system flux estimation and attribution: impact of ACOS-GOSAT XCO<sub>2</sub>  
 sampling on the inference of terrestrial biospheric sources and sinks, *Tellus B Chem. Phys. Meteorol.*, 66, 22486,  
<https://doi.org/10.3402/tellusb.v66.22486>, 2014.
- Liu, Y., Kalnay, E., Zeng, N., Asrar, G., Chen, Z., and Jia, B.: Estimating surface carbon fluxes based on a local ensemble  
 415 transform Kalman filter with a short assimilation window and a long observation window: an observing system simulation  
 experiment test in GEOS-Chem 10.1, *Geosci. Model Dev.*, 12, 2899–2914, <https://doi.org/10.5194/gmd-12-2899-2019>,  
 2019.
- Liu, Z., Zeng, N., Liu, Y., Kalnay, E., Asrar, G., Wu, B., Cai, Q., Liu, D., and Han, P.: Improving the joint estimation of CO<sub>2</sub>  
 and surface carbon fluxes using a constrained ensemble Kalman filter in COLA (v1.0), *Geosci Model Dev.*, 18,  
 420 <https://doi.org/10.5194/gmd-15-5511-2022>, 2022.
- 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 CO<sub>2</sub> emissions from global fossil fuel emission data sets, *J. Geophys. Res. Atmospheres*, 118, 917–



- 933, <https://doi.org/10.1029/2012JD018196>, 2013.
- Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic CO<sub>2</sub>, version 2016  
 425 (ODIAC2016): a global monthly fossil fuel CO<sub>2</sub> gridded emissions data product for tracer transport simulations and surface  
 flux inversions, *Earth Syst. Sci. Data*, 10, 87–107, <https://doi.org/10.5194/essd-10-87-2018>, 2018.
- O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H.,  
 Mandrake, L., Merrelli, A., Natraj, V., Nelson, R. R., Osterman, G. B., Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B.  
 430 J., Oyafuso, F., Chang, A., McDuffie, J., Smyth, M., Baker, D. F., Basu, S., Chevallier, F., Crowell, S. M. R., Feng, L.,  
 Palmer, P. I., Dubey, M., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Ohyama,  
 H., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Velazco, V. A.: Improved retrievals  
 of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, *Atmospheric Meas. Tech.*, 11,  
 6539–6576, <https://doi.org/10.5194/amt-11-6539-2018>, 2018.
- O'Sullivan, M., Friedlingstein, P., Sitch, S., Anthoni, P., Arneth, A., Arora, V. K., Bastrikov, V., Delire, C., Goll, D. S., Jain,  
 435 A., Kato, E., Kennedy, D., Knauer, J., Lienert, S., Lombardozzi, D., McGuire, P. C., Melton, J. R., Nabel, J. E. M. S.,  
 Pongratz, J., Poulter, B., Séférian, R., Tian, H., Vuichard, N., Walker, A. P., Yuan, W., Yue, X., and Zaehle, S.: Process-  
 oriented analysis of dominant sources of uncertainty in the land carbon sink, *Nat. Commun.*, 13, 4781,  
<https://doi.org/10.1038/s41467-022-32416-8>, 2022.
- Pan, M. and Wood, E. F.: Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble  
 440 Kalman Filter, *J. Hydrometeorol.*, 7, 534–547, <https://doi.org/10.1175/JHM495.1>, 2006.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P.,  
 Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans,  
 P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proc. Natl. Acad. Sci.*, 104,  
 18925–18930, <https://doi.org/10.1073/pnas.0708986104>, 2007.
- 445 Philip, S., Johnson, M. S., Potter, C., Genovesse, V., Baker, D. F., Haynes, K. D., Henze, D. K., Liu, J., and Poulter, B.: Prior  
 biosphere model impact on global terrestrial CO<sub>2</sub> fluxes estimated from OCO-2 retrievals, *Atmospheric Chem. Phys.*, 19,  
 13267–13287, <https://doi.org/10.5194/acp-19-13267-2019>, 2019.
- Rodenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO<sub>2</sub> flux history 1982–2001 inferred from atmospheric data  
 using a global inversion of atmospheric transport, *Atmospheric Chem. Phys.*, 3, 1919–1964, 2003.
- 450 Rödenbeck, C., Bakker, D. C. E., Metzl, N., Olsen, A., Sabine, C., Cassar, N., Reum, F., Keeling, R. F., and Heimann, M.:  
 Interannual sea–air CO<sub>2</sub> flux variability from an observation-driven ocean mixed-layer scheme, *Biogeosciences*, 11, 4599–  
 4613, <https://doi.org/10.5194/bg-11-4599-2014>, 2014.
- Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., Chevallier, F., Crowell, S., Davis, K. J., Deng, F.,  
 Denning, S., Feng, L., Jones, D., Liu, J., and Palmer, P. I.: Quantifying the Impact of Atmospheric Transport Uncertainty  
 455 on CO<sub>2</sub> Surface Flux Estimates, *Glob. Biogeochem. Cycles*, 33, 484–500, <https://doi.org/10.1029/2018GB006086>, 2019.
- The International GEOS-Chem User Community: [geoschem/GCClassic](https://geoschem/GCClassic): GEOS-Chem 13.0.2, ,



- <https://doi.org/10.5281/ZENODO.4681204>, 2021.
- Weir, B., Oda, T., Ott, L. E., and Schmidt, G. A.: Assessing progress toward the Paris Climate Agreement from Space, *Environ. Res. Lett.*, <https://doi.org/10.1088/1748-9326/ac998c>, 2022.
- 460 Yu, K., Keller, C. A., Jacob, D. J., Molod, A. M., Eastham, S. D., and Long, M. S.: Errors and improvements in the use of archived meteorological data for chemical transport modeling: an analysis using GEOS-Chem v11-01 driven by GEOS-5 meteorology, *Geosci. Model Dev.*, 11, 305–319, <https://doi.org/10.5194/gmd-11-305-2018>, 2018.
- Zammit-Mangion, A., Bertolacci, M., Fisher, J., Stavert, A., Rigby, M., Cao, Y., and Cressie, N.: WOMBAT v1.0: a fully Bayesian global flux-inversion framework, *Geosci. Model Dev.*, 15, 45–73, <https://doi.org/10.5194/gmd-15-45-2022>, 2022.
- 465 Zeng, N., Mariotti, A., and Wetzel, P.: Terrestrial mechanisms of interannual CO<sub>2</sub> variability, *Glob. Biogeochem. Cycles*, 19, <https://doi.org/10.1029/2004GB002273>, 2005.
- Zeng, Y., Janjić, T., Ruckstuhl, Y., and Verlaan, M.: Ensemble-type Kalman filter algorithm conserving mass, total energy and enstrophy: SQPEns Conserving Mass, Total Energy and Enstrophy, *Q. J. R. Meteorol. Soc.*, 143, 2902–2914, <https://doi.org/10.1002/qj.3142>, 2017.

470