Response to Reviewer #2

We thank the reviewer#2 for the insightful and detailed comments and suggestions, which helped to significantly improve the manuscript. The reviewer's comments are shown in *blue italics* with the author responses in black.

General comments:

This study introduces the top down inversion of the natural biosphere carbon fluxes over China with a high horizontal resolution of about 64 km by joint optimization of initial CO_2 condition and biosphere carbon fluxes using GOSAT satellite observations. The magnitude of the estimated annual biosphere sink in China was consistent with most previous studies. In addition, the provincial biosphere carbon flux over China was also reestimated. Generally speaking, the paper is well written and scientific sound.

Main comments:

- *It is unclear how the uncertainties of the background carbon fluxes are used in the data assimilation. Since the uncertainties of the background carbon fluxes are critical for the inversion, please clarify it more detail.*

Thank the reviewer for the comment. In CMAQ simulation, the prior prescribed CO_2 emissions come from both anthropogenic sources and natural sources, including fossil-fuel emission, terrestrial ecosystem flux, oceanic flux, and biomass burning emissions. In the assimilation, the natural flux (i.e., biosphere–atmosphere exchange and ocean–atmosphere exchange) were assimilated, while the fossil-fuel and biomass-burning fluxes were fixed based on bottom-up estimates, which follows previous inversion work and reflects our faith in inventory-based emissions for fossil fuels (Peters et al., 2007, 2010; Tian et al., 2014; Wang et al., 2019; Wang et al., 2020).

Considering the high level of uncertainty in simulated bioflux in current terrestrial biosphere models, those *a priori* biospheric fluxes were interpolated from the widely recognized CT2019B products, which is a global inverse model of atmospheric CO_2 to produce quantitative estimates of atmospheric carbon uptake and release. CO_2 fluxes $F(x, y, t)$ in CT2019B are parameterized according to

$$F(x,y,t) = \lambda(x,y,t)\big(F_{bio}(x,y,t) + F_{ocean}(x,y,t)\big) + F_{ff}(x,y,t) + F_{fire}(x,y,t)$$

where $F_{bio}, F_{ocean}, F_{ff}$, and F_{fire} are prior flux model predictions for land biosphere, ocean, fossil fuel and biomass burning emissions respectively, and λ represents a set of unknown multiplicative scaling factors applied to the fluxes, to be estimated in the assimilation. These scaling factors are the final product of CT2019B optimized fluxes.

In CarbonTracker, the flux dynamical model is applied to the ensemble-mean parameter values λ as:

$$\lambda(\ t) = (\lambda_0 + \lambda(t-1) + \lambda(t-2))/3$$

where $\lambda(\ t)$ is the prior value of the scaling factors for timestep t, λ_0 is the initial prior vector with all elements set to 1.0, and $\lambda(t-1)$ and $\lambda(t-2)$ refers to the posterior scaling factors for the timestep $t-1$ and $t-2$ repectively. This model describes that parameter values λ for a new time step are chosen as a combination of optimized values from the two previous time steps and a fixed overall prior value of 1.0.

In this study, the Equation (1) describes the flux forecast model in JDAS by taking the a priori flux, the analysis flux from the previous assimilation cycle, and the forecast concentration as independent variables. We can see that *M* is used for linking the assimilated fluxes from the previous assimilating cycle, and *M* was set to 3 in CarbonTracker. In JDAS real practice, *M* was set to 4 days at the same time on each day to represent the average state of the biospheric diurnal variation at a certain seasonality level, as a result of several sensitivity tests which are not present here.

Measured CO_2 concentrations are the result of upstream surface fluxes and atmospheric transport process. Generally speaking, the longer in the past a flux event occurred, the smaller its impact will be on a given sample of air. Therefore, we choose an “assimilation window” to represents how far back in time we expect to be able to pinpoint a given flux signal from available measurements. CT2019B have designed the assimilation window length as 12 weeks. This helps to resolve fluxes in regions of the world with less dense observational coverage.

Similar to CarbonTracker which uses transport model as a forward operator in an ensemble fixed-lag Kalman smoother, JDAS is also extended to incorporate the ensemble Kalman smoother (EnKS) feature along with EnSRF. The EnKS allows for a sequential processing of the measurements in time and is used to assimilate the concentrations and update the fluxes. Thus, EnKS that can take into future observations into account is used to assimilate the concentrations and update the fluxes. The smoothing window of EnKS (i.e. denoted as assimilation window hereafter) was set to 24 h in this study. In an assimilation cycle, the fluxes for the 24-h smoothing window have been designed to be optimized hour by hour successively.

The distribution of ensemble spread of CO_2 flux in January 2016 is provided in Figure R1. It shows that the values of the ensemble spread ranges from 0.2 to 0.8 in most areas, which are consistent with our previous studies (Peng et al., 2015 in Figure 11c and Peng et al., 2023).

We are sorry there is some confusion about the smoothing window in flux forecast model and EnKS in this manuscript. To avoid the confusion, we have modified the relevant parts in Section 2.1.2 (forecast model of ensemble fluxes) and Section 2.2.2 (EnKS assimilation scheme) in the revised manuscript.

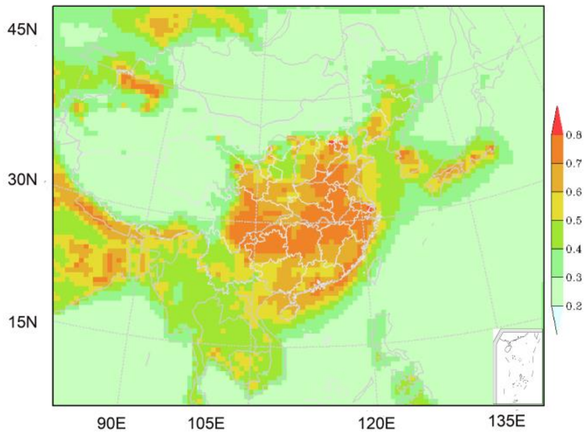

Figure R1. The ensemble spread of $\boldsymbol{\lambda}_{i,t}^{a}$ at model level 1 in January 2016, when β=80.

- *How the uncertainties of the boundary concentrations are considered in the study? How often the boundary and initial concentrations are imported from the CT2019B, and are the boundary concentrations are also optimized?*

Thank the reviewer for the comment. As the initial and lateral boundary atmospheric CO_2 concentrations, the global 4D CO_2 data were created using the optimized surface fluxes and simulated atmospheric transport of CarbonTracker, version CT2019B, from the National Oceanic and

Atmospheric Administration (NOAA), with a spatial resolution of 3° × 2°, 25 vertical levels, and a temporal resolution of 3 h, which represent the optimum estimate of the distribution of atmospheric CO_2 (Jacobson et al., 2020).

In each EnKS analysis step, CMAQ integrated and generated a 3D CO_2 concentration ensemble derived by the N ensemble fluxes with perturbed CO_2 initial and boundary conditions. The ensemble assimilation was performed for the period 0000 UTC 25 December 2015 to 2300 UTC 31 December 2016 using the perturbed initial conditions and boundary conditions by adding Gaussian random noise with a standard deviation of 5%. In an assimilation cycle, the fluxes for the 24-h assimilation window have been designed to be optimized hour by hour successively. Accordingly, the fluxes have been adjusted 24 times before generating posterior fluxes. In this way, both the initial and boundary concentrations are optimized every hour.

We have modified the relevant parts in the revised manuscripts (Line 298−329). The detailed description of EnKS-based assimilation system configuration can be referred to Section 2.2 (JDAS CO_2 assimilation framework) and Section 2.4 (Experimental design and evaluation method) in the manuscript.

- *The a priori fluxes from CT2019B are at a 3-h intervals, how was the hour-by-hour assimilation conducted? Are the initial conditions are also optimized every hour?*

In an assimilation cycle, the fluxes for the 24-h assimilation window have been designed to be optimized hour by hour successively. Accordingly, the fluxes have been adjusted 24 times before generating posterior fluxes. Actually, the NOAA operational EnKF system, which is an EnSRF and modified with the EnKS feature, is further extended to jointly assimilate the CO_2 initial conditions and fluxes to update the flux and concentration fields, respectively. The EnKS allows for a sequential processing of the measurements in time, which updates the ensemble at prior times every time new observations are available. Thus, EnKS that can take into future observations into account is used to assimilate the concentrations and update the fluxes.

In this study, the state vector $\mathbf{x}$ includes the mass concentration $\mathbf{C}$ and the emission $\boldsymbol{E}$, i.e.

$\mathbf{x} = [\boldsymbol{C}, \boldsymbol{E}]^{\mathrm{T}}$. Here, the state variables of mass concentration $\mathbf{C}$ are the CO_2 concentrations. The ensemble forecast concentration fields of CO_2 are respectively used in calculating ensemble fluxes $\boldsymbol{E}_{i,t}^{f}$ as described in Section 2.2.1. The ensemble members of chemical fields $\mathbf{C}^{\mathrm{f}}$ are forecasted using CMAQ, forced by the forecast emissions $\mathbf{E}^{\mathrm{f}}$ whose initial conditions are previously analyzed concentration fields. Now, the background of the joint vector, $\mathbf{x}^{\mathrm{f}} = \left[\mathbf{C}^{\mathrm{f}}, \mathbf{E}^{\mathrm{f}}\right]^{\mathrm{T}}$, has been produced. Then, the analyzed state vector, $\mathbf{x}^{\mathrm{a}} = [\mathbf{C}^{\mathrm{a}}, \boldsymbol{E}^{\mathrm{a}}]^{\mathrm{T}}$, is optimized by applying the EnKS, respectively. The configurations of the EnKS were as follows: 1) ensemble size was set to 50; 2) the horizontal localization radius was 1280 km; 3) the covariance inflation factor β was set to 80; 4) the smoothing window (i.e. denoted as assimilation window hereafter) was set to 24 h, as sensitivity experiments about smoothing windows has been tested to find the optimum length in our previous study (Peng, et al., 2023). In addition, hour-by-hour assimilation was adopted attribute to the novel flux forecast model, fine-scale CMAQ forward hourly simulation output, as well as the hourly observations. Thus, the initial condition, boundary concentrations and flux are optimized every hour.

We have modified the relevant parts (Section 2.2) in the revised manuscripts, and we hope we can make the meaning clear now.

- *It is better to separate the results and discussion.*

Thank the review for the comment. And we have separated the results and discussion in the revised manuscript (Section 3 and Section 4).

Specific comments:

- *P9 How do you determine the values of the horizontal covariance localization radius and the inflation factor?*

The localization radius 1280 km follows our previous research including Peng et al., 2015, Peng et al., 2018, Peng et al., 2023, which localize the impact of observation and ameliorate spurious error correlations between observations and state variables. Thus, covariance localization (Houtekamer & Mitchell, 2001) with the Gaspari and Cohn (Gaspari & Cohn, 1999) function of 1280 km length scale, are utilized.

Moreover, the covariance inflation factor β was set to 80 to preserve the ensemble spread ranging to some extent. The distribution of ensemble spread of CO_2 flux in January 2016 is provided in Figure R1. It shows that the values of the ensemble spread ranges from 0.2 to 0.8 in most areas, which are consistent with our previous studies (Peng et al., 2015 in Figure 11c and Peng et al. 2023).

We have modified the relevant parts in the revised manuscript (Line 265–270).

Here are the above-mentioned references.

Peng, Z., Zhang, M. G., Kou, X. X., Tian, X. J., & Ma, X. G. (2015). A regional carbon flux data assimilation system and its preliminary evaluation in East Asia. *Atmospheric Chemistry and Physics*, 15, 1087–1104. https://doi.org/10.5194/acp-15-1087-2015.

Peng, Z., Lei, L. L., Liu, Z. Q., Sun, J. N., Ding, A, J., Ban, J. M., et al. (2018). The impact of multi-species surface chemical observation assimilation on air quality forecasts in China. *Atmospheric Chemistry and Physics*, 18, 17387–17404. https://doi.org/10.5194/acp-18-17387-2018

Peng, Z., Kou, X. X., Zhang, M. G., Lei, L. L., Miao, S. G., Wang, H. M., Jiang, F., Han, X., and Fang, S. X. (2023). CO_2 flux inversion with a regional joint data assimilation system based on CMAQ, EnKS, and surface observations. *Journal of Geophysical Research-Atmosphere*, 128, e2022JD037154. https://doi. org/10.1029/2022JD037154

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129, 123–137. https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Gaspari, G., & Cohn S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757. https://doi.org/10.1002/qj.49712555417

- *Why the Table 2 is firstly appeared in the main text?*

Thank the review for the comment. And we have adjusted the order of the tables.

Line 526 The horizontal resolution of the CMAQ model in the study is about 64 km, why the results cannot resolve the Shanghai?

The total area of Shanghai is 6340.5 km^2. The CMAQ configuration used here was 64 × 64 km^2 (i.e. 4096 km^2 each grid) fixed grid cells centered at 35 °N and 116 °E in a rotated polar stereographic map projection. This domain, having 105 (west–east) × 86 (south–north) grid points, covered the whole of mainland China and its surrounding regions (Fig. 1). Thus, owing to the insufficient grid resolution, Shang has been mixed with the neighbouring areas, especially Jiangsu and Zhejiang provinces. In addition, Hong Kong and Macao are not discussed, because the results cannot resolve these areas too.