Supplement of Earth Syst. Sci. Data, 16, 2857–2876, 2024
https://doi.org/10.5194/essd-16-2857-2024-supplement

Supplement of

A global surface CO_2 flux dataset (2015–2022) inferred from OCO-2 retrievals using the GONGGA inversion system

Zhe Jin et al.

Correspondence to: Xiangjun Tian (tianxj@itpcas.ac.cn) and Yilong Wang (wangyilong@itpcas.ac.cn)

Text S1: Method for calculating prior and posterior uncertainties

For the flux optimization, the optimized variable is the scaling factor. The posterior flux is the product of the posterior scaling factor and the prior flux:

$$F_{t,i,j}^{post} = \lambda_{t,i,j}^{post} \times F_{t,i,j}^{prior}, \tag{S1}$$

where $F_{t,i,j}^{post}$ is the posterior carbon flux, $\lambda_{t,i,j}^{post}$ is the posterior scaling factor, and $F_{t,i,j}^{prior}$ is the prior carbon flux, t denotes the current t th window, i denotes the i th grid in longitude, and j denotes the j th grid in latitude. Both of fluxes and scaling factors are gridded variables with the same horizontal resolution as the transport model. To characterize the prior uncertainty of NEE and ocean carbon fluxes, the NLS-4DVar method applies an ensemble to approximate the prior error covariance matrix (Tian et al., 2018):

$$\mathbf{B} = \frac{\left(\mathbf{P}_x^{prior}\right)\left(\mathbf{P}_x^{prior}\right)^{\mathrm{T}}}{N-1}. \tag{S2}$$

where $\mathbf{P}_x^{prior} = (\boldsymbol{x}_1', \boldsymbol{x}_2', \ldots, \boldsymbol{x}_N')$ is an ensemble of prior perturbations, $\boldsymbol{x}_j' = \boldsymbol{x}_j - \boldsymbol{x}_a, j = 1,2,\ldots,N$, $\boldsymbol{x}_j'$ is the j th perturbation, $\boldsymbol{x}_j$ is the j th sample, and N is the number of prior perturbations. In this study, N equals to 36. The prior perturbations of the scaling factors in the first inversion window were obtained through historical sampling of fluxes. We first created 108 samples from historical fluxes, which consists of the monthly mean fluxes from January 1, 2011 to December 31, 2019. Then they divided the monthly mean flux in September 2014 and subtracted 1 to form the ensemble of perturbations of flux scaling factors. Subsequently, 36 samples that could represent the key spatial patterns of the large ensemble were extracted using Random State Variable (RSV) method (Zhang et al., 2020), forming the prior perturbations for the first inversion window. After the inversion of the first window, the prior perturbations of the next window were updated (Tian et al., 2020):

$$\mathbf{P}_x^{prior,w+1} = \mathbf{P}_x^{prior,w}\mathbf{V}_2\Phi^{\mathrm{T}} \tag{S3}$$

Where $\mathbf{P}_x^{prior,w+1}$ is the ensemble of the prior perturbations for the next window, and $\mathbf{P}_x^{prior,w}$ is the ensemble of the prior perturbations for the current window. The matrix $\mathbf{V_2}$ can be calculated by Eq. (S5-S7) detailed below, and Φ^{T} is a random orthogonal matrix. The procedure was repeated through all inversion windows. Both NEE and ocean-atmosphere fluxes applied this sample generation method. The historical NEE were from ORHIDEE-MICT simulations (Guimberteau et al., 2018), and historical ocean-atmosphere fluxes were from Takahashi climatology results (Takahashi et al., 2009). As a result, the total uncertainty of our prior land and ocean fluxes at a global scale and for a full year, before assimilating XCO_2 observations, amount to an average of 4.7 Pg C yr^{-1} and 0.28 Pg C yr^{-1}, respectively. According to Evensen (2009), the ensemble of posterior perturbations after assimilation is calculated as follows:

$$\mathbf{P}_x^{post} = \mathbf{P}_x^{prior}\mathbf{V}_2\sqrt{\mathbf{I} - \mathbf{\Sigma}_2^{\mathrm{T}}\mathbf{\Sigma}_2}\Phi^{\mathrm{T}}, \tag{S4}$$

where

$$\mathbf{U}_2\mathbf{\Sigma}_2\mathbf{V}_2^{\mathrm{T}} = \mathbf{X}_2, \tag{S5}$$

$$\mathbf{X}_2 = \mathbf{\Lambda}^{-1/2}\mathbf{Z}^{\mathrm{T}}\mathbf{P}_y, \tag{S6}$$

$$\mathbf{Z}\mathbf{\Lambda}^{-1}\mathbf{Z}^{\mathrm{T}} = \left[\left(\mathbf{P}_y\right)\left(\mathbf{P}_y\right)^{\mathrm{T}} + (N-1)\mathbf{R}\right]^{-1}. \tag{S7}$$

and Φ is a random orthogonal matrix, $\mathbf{P}_y = h\left(\mathbf{P}_x^{prior}\right) - h(\boldsymbol{x}_a)$. Then, the prior ($\mathbf{B}$) and posterior ($\mathbf{B}^{post}$) error covariance matrices can be calculated using $\mathbf{P}_x^{prior}$ and $\mathbf{P}_x^{post}$, respectively, according to Eq. (S2).

After obtaining the prior and posterior uncertainties of the scaling factors, the prior and posterior total flux uncertainties (σ_{total}^{prior} and σ_{total}^{post}) can be calculated according to the correlation between fluxes and scaling factors (Niwa and Fujii, 2020):

$$\sigma_{total}^{prior} = \sqrt{(\boldsymbol{F}^{prior})^{\mathrm{T}}\mathbf{B}(\boldsymbol{F}^{prior})}, \tag{S7}$$

$$\sigma_{total}^{post} = \sqrt{(\boldsymbol{F}^{prior})^{\mathrm{T}}\mathbf{B}^{post}(\boldsymbol{F}^{prior})}. \tag{S8}$$

where we assume that the flux uncertainties are time independent.

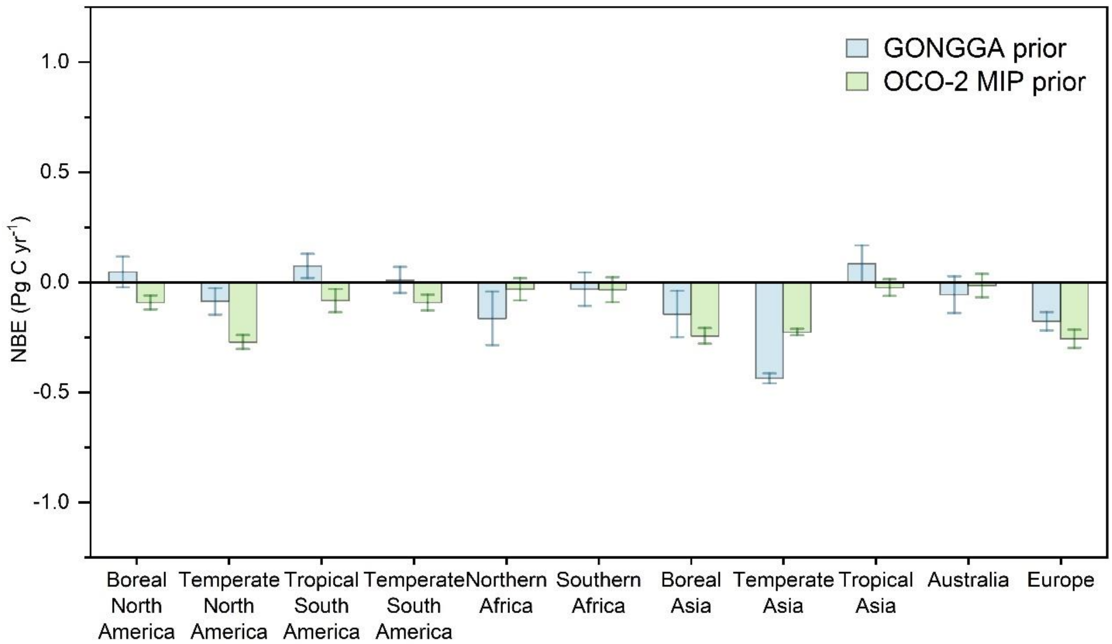

Figure S1. Annual mean (2015–2022) NBE at 11 TransCom land regions from GONGGA prior and OCO-2 MIP prior estimates. Error bar of NBE represents multi-year standard deviation.

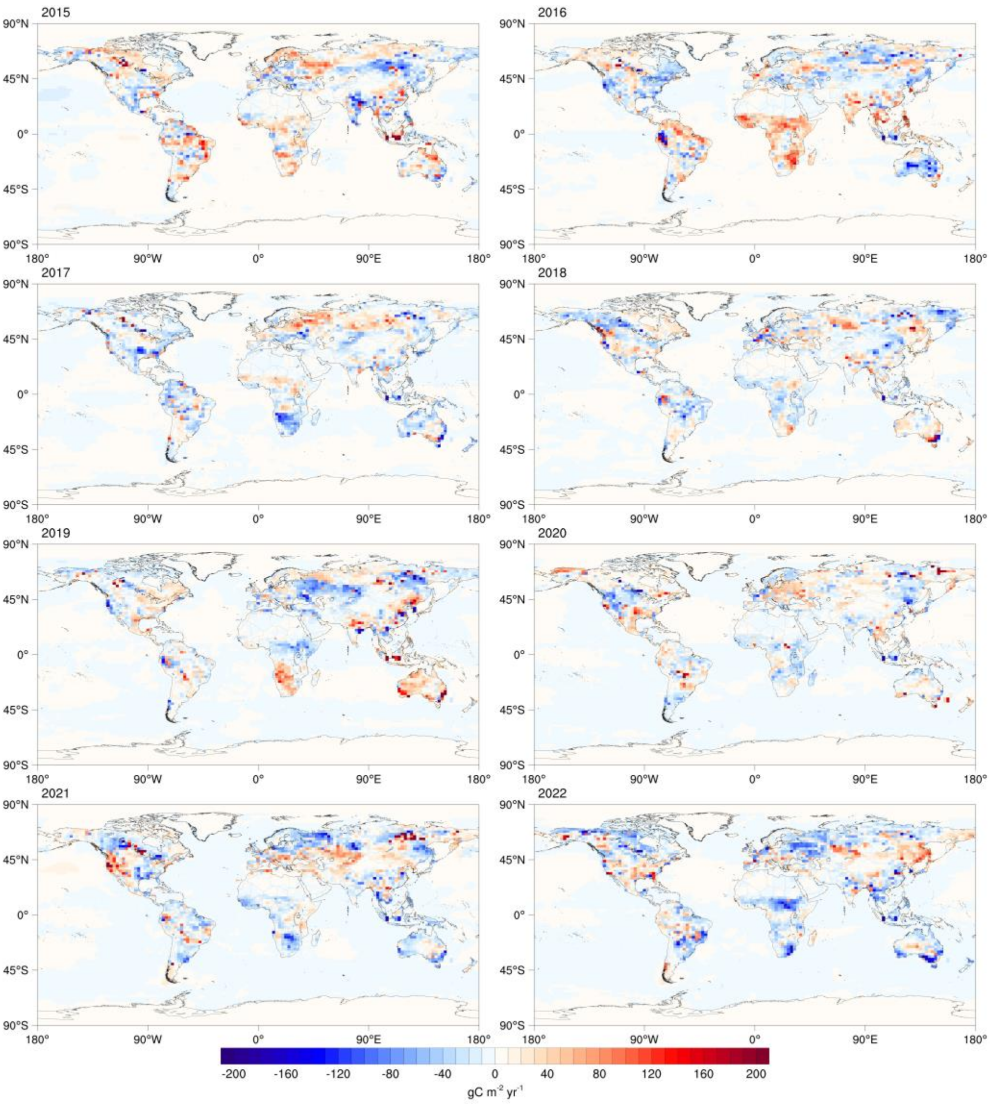

Figure S2. The annual NBE and ocean flux anomalies (annual value minus 8-year mean) during 2015–2022 period.

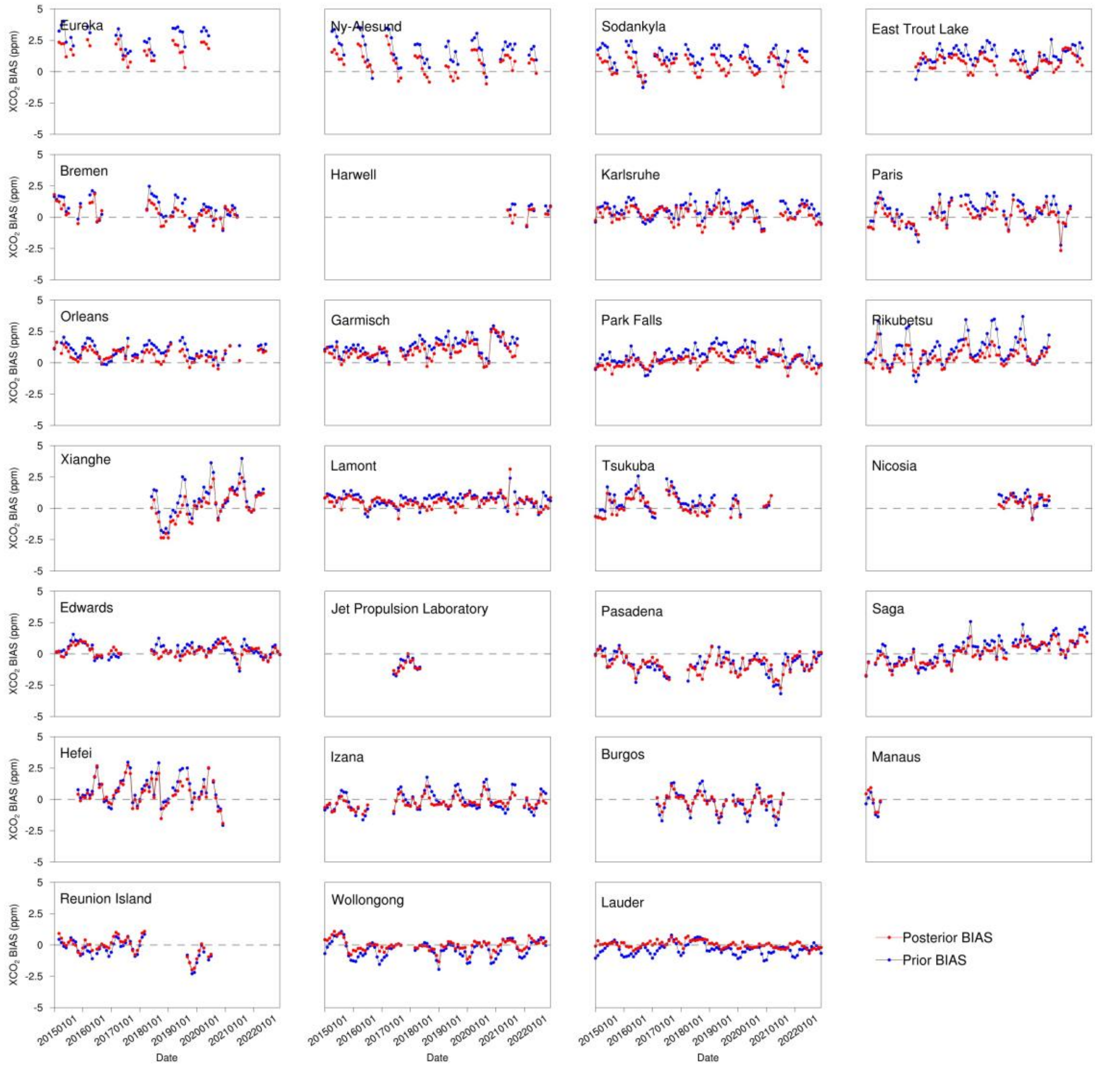

Figure S3. Time series of monthly averaged prior (blue) and posterior (red) simulated XCO_2 bias at each TCCON site (prior/posterior simulation – observation).

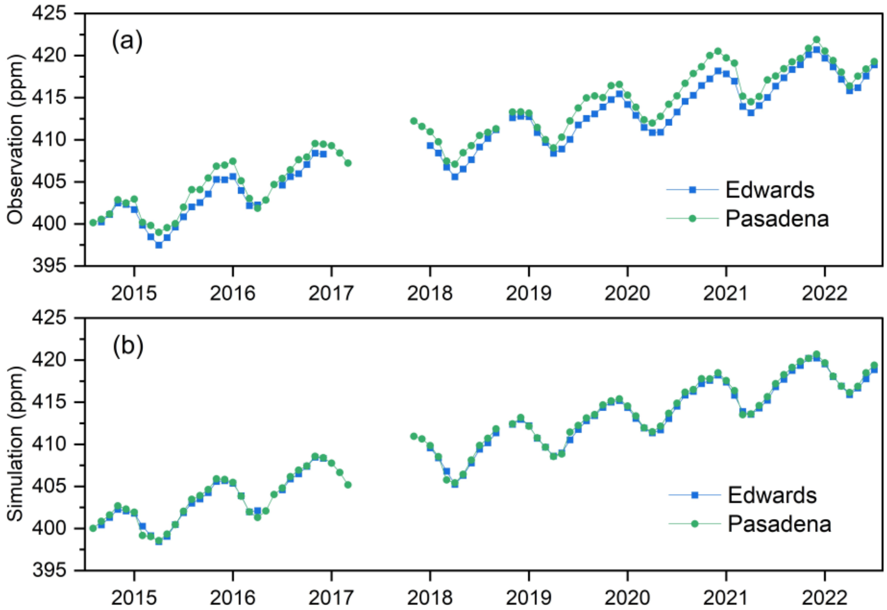

Figure S4. Time series of monthly (a) TCCON observations and (b) corresponding posterior simulations at Edwards (blue) and Pasadena (green) during 2015−2021 period.

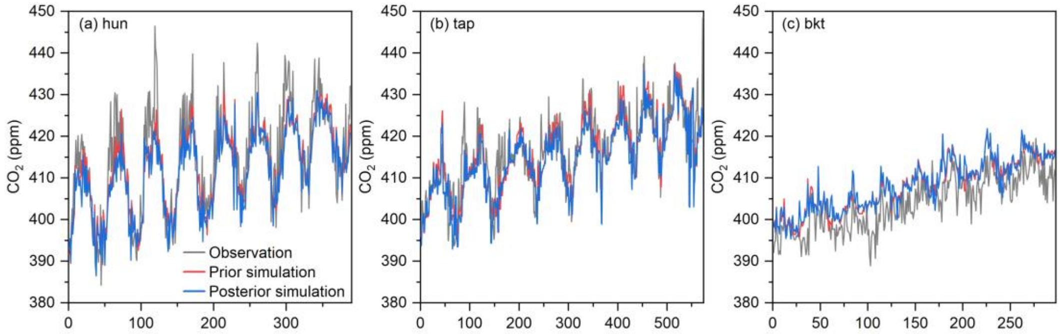

Figure S5. Time series of ObsPack surface flask observations as well as corresponding prior and posterior simulations at three sites that posterior RMSE exceed 4.0 ppm.

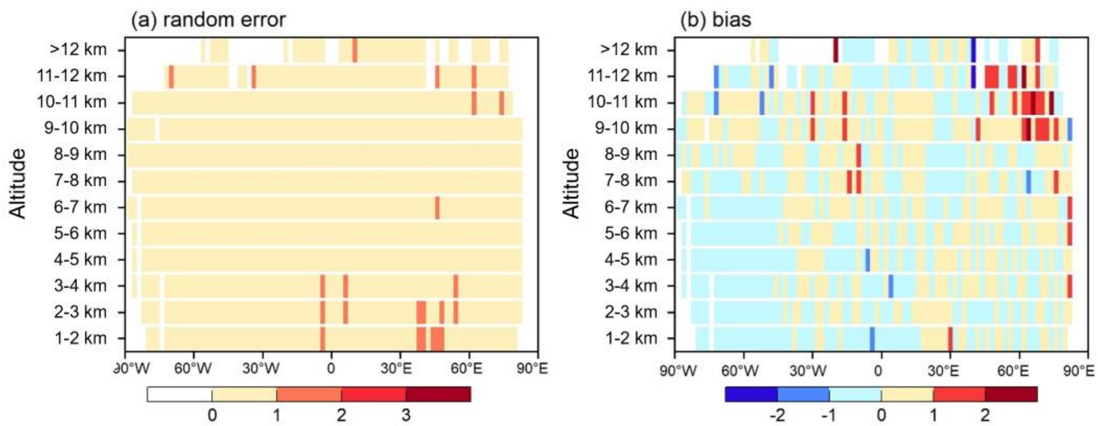

Figure S6. (a) random error and (b) bias between posterior CO_2 simulations and aircraft observations as a function of latitude and altitude (posterior simulations minus observations; unit: ppm). The altitudes are binned every kilometer from 1 km to 12 km, and for altitudes above 12 km.

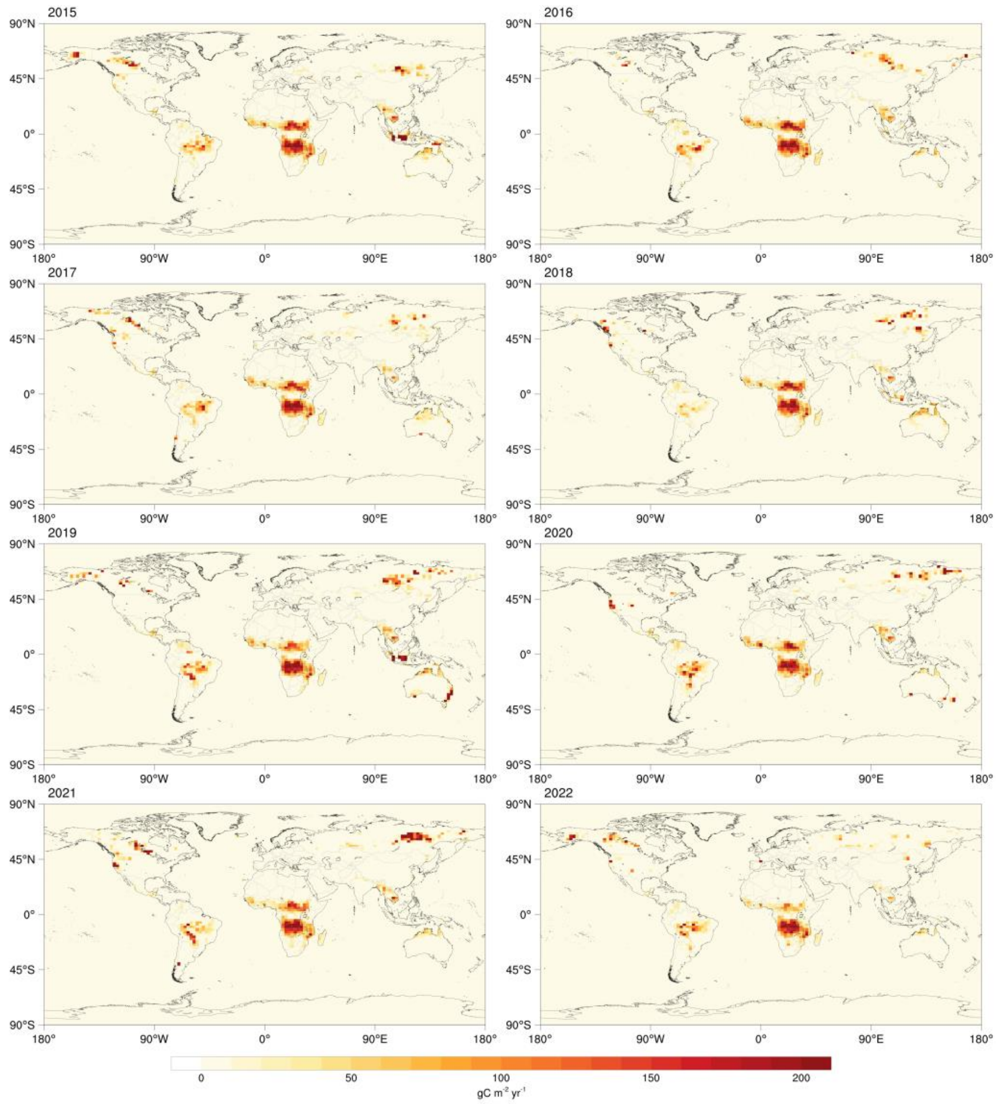

Figure S7. The spatial distribution of biomass burning emissions from GFED4.1s estimate during 2015–2022 period.

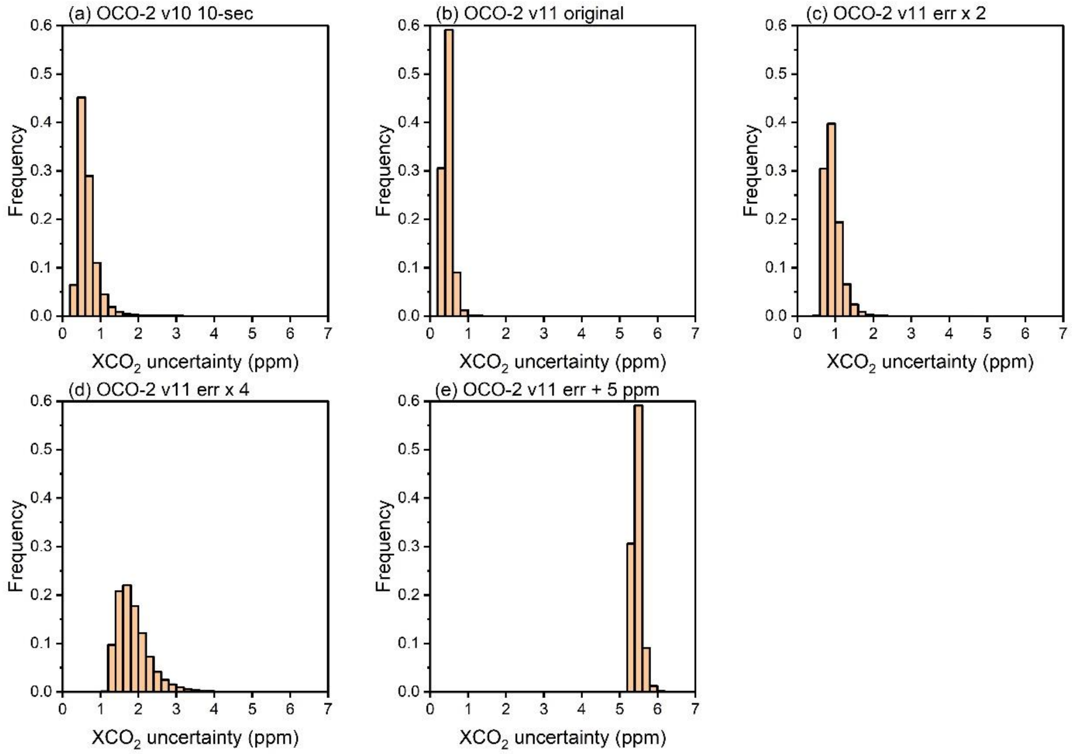

Figure S8. The distribution of (a) OCO-2 v10r 10 s averaged XCO_2 uncertainties, (b) default OCO-2 v11r XCO_2 uncertainties, (c) OCO-2 v11r XCO_2 uncertainties doubled, (d) OCO-2 v11r XCO_2 uncertainties quadrupled, and (d) OCO-2 v11r XCO_2 uncertainties added by 5 ppm.

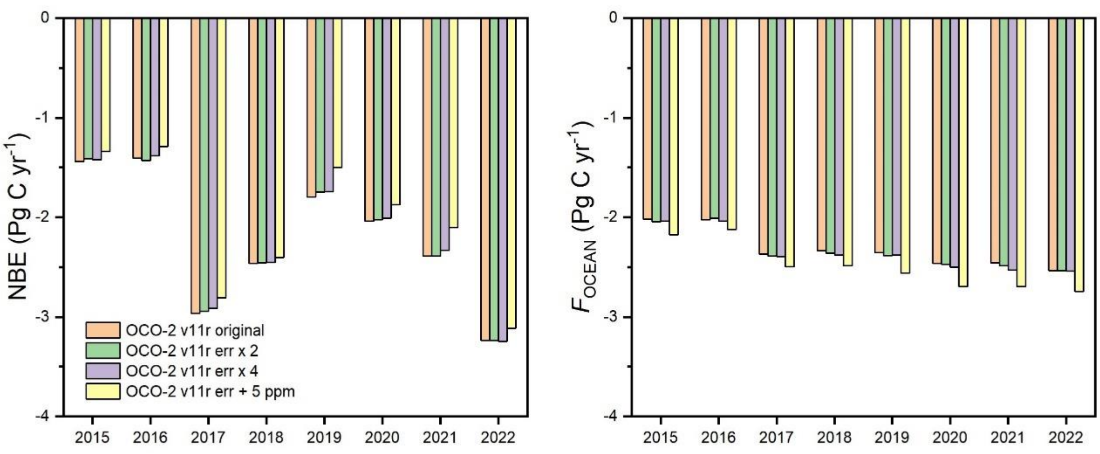

Figure S9. The global annual NBE and F_{OCEAN} from GONGGA posterior estimates with default OCO-2 v11r XCO_2 uncertainties (orange), doubled OCO-2 v11r original XCO_2 uncertainties (green), quadrupled OCO-2 v11r original XCO_2 uncertainties (purple), and OCO-2 v11r original XCO_2 uncertainties added by 5 ppm (yellow).

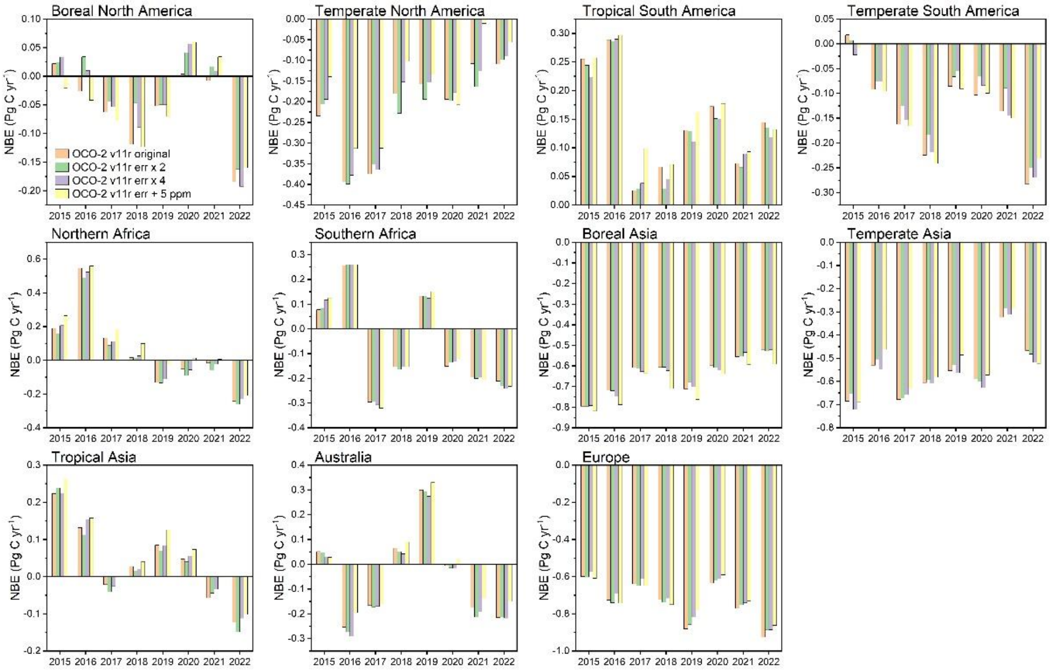

Figure S10. NBE in 11 TransCom land regions from GONGGA posterior estimates with default OCO-2 v11r XCO_2 uncertainties (orange), doubled OCO-2 v11r original XCO_2 uncertainties (green), quadrupled OCO-2 v11r XCO_2 uncertainties (purple), and OCO-2 v11r XCO_2 uncertainties added by 5 ppm (yellow).

Table S1. OCO-2 MIP v10 participants and model details.

Model	Contact	Institution	Transport Model	Meteorology	Inverse Method
Ames	Matthew Johnson and Sajeev Philip	NASA Ames Research Center	GEOS-Chem	MERRA-2	4D-Var
CAMS	Frédéric Chevallier	LSCE France	LMDz	ERA-interim	4D-Var
COLA	Zhiqiang Liu	–	–	–	–
CMS-Flux	Junjie Liu	NASA JPL	GEOS-Chem	GEOS-FP	4D-Var
CSU	Andrew Schuh	Colorado State University	GEOS-Chem	MERRA-2	Bayesian synthesis
CT	Andy Jacobson	University of Colorado and NOAA GML	TM5	ERA-interim	EnKF
JHU	Scot Miller	–	–	–	–
LoFI	Brad Weir	–	–	–	–
NIES	Shamil Maksyuotov	–	–	–	–
OU	Sean Crowell	University of Oklahoma	TM5	ERA-interim	4D-Var
PCTM	David Baker	Colorado State University	PCTM	MERRA-2	4D-Var
TM5-4DVAR	Sourish Basu	University of Maryland and NASA GMAO	TM5	ERA-interim	4D-Var
UT	Feng Deng	University of Toronto	GEOS-Chem	GEOS-FP	4D-Var
WOMBAT	Michael Bertolacci, Andrew Zammit Mangion, Noel Cressie	University of Wollongong	GEOS-Chem	MERRA-2	MCMC

Table S2. Annual and six-year mean NBP at Boreal North America and Northern Africa from OCO-2 MIP v10 IS and LNLG experiments. Uncertainties are the one standard deviation spread in the inversion ensemble.

Region	Year	Experiment	NBE (PgC yr^{-1})	Experiment	NBE
Boreal North America	2015	IS	-0.28 ± 0.36	LNLG	-0.22 ± 0.56
	2016		-0.36 ± 0.37		-0.11 ± 0.49
	2017		-0.34 ± 0.38		-0.22 ± 0.53
	2018		-0.40 ± 0.33		-0.21 ± 0.53
	2019		-0.46 ± 0.37		-0.27 ± 0.59
	2020		-0.44 ± 0.48		-0.07 ± 0.53
	Mean		-0.38 ± 0.38		-0.18 ± 0.54
Northern Africa	2015	IS	0.23 ± 1.42	LNLG	0.87 ± 0.89
	2016		0.65 ± 1.42		1.24 ± 0.88
	2017		0.32 ± 1.40		0.90 ± 0.95
	2018		0.01 ± 1.26		0.70 ± 0.87
	2019		0.24 ± 1.15		0.73 ± 0.85
	2020		0.34 ± 1.18		0.64 ± 0.92
	Mean		0.30 ± 1.31		0.85 ± 0.90

Reference

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, 2, Springer-Verlag Berlin Heidelberg, 307 pp., https://doi.org/10.1007/978-3-642-03711-5, 2009.

Guimberteau, M., Zhu, D., Maignan, F., Huang, Y., Yue, C., Dantec-Nédélec, S., Ottlé, C., Jornet-Puig, A., Bastos, A., Laurent, P., Goll, D., Bowring, S., Chang, J., Guenet, B., Tifafi, M., Peng, S., Krinner, G., Ducharne, A., Wang, F., Wang, T., Wang, X., Wang, Y., Yin, Z., Lauerwald, R., Joetzjer, E., Qiu, C., Kim, H., and Ciais, P.: ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: model description and validation, Geosci. Model Dev., 11, 121-163, https://doi.org/10.5194/gmd-11-121-2018, 2018.

Niwa, Y. and Fujii, Y.: A conjugate BFGS method for accurate estimation of a posterior error covariance matrix in a linear inverse problem, Q. J. Roy. Meteor. Soc., 146, 3118-3143, https://doi.org/10.1002/qj.3838, 2020.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Koertzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R., and de Baar, H. J. W.: Climatological mean and decadal change in surface ocean pCO_2, and net sea-air CO_2 flux over the global oceans, Deep-Sea Res. Pt. II, 56, 554-577, https://doi.org/10.1016/j.dsr2.2008.12.009, 2009.

Tian, X., Zhang, H., Feng, X., and Xie, Y.: Nonlinear least squares En4DVar to 4DEnVar methods for data assimilation: Formulation, analysis, and preliminary evaluation, Mon. Weather Rev., 146, 77-93, https://doi.org/10.1175/mwr-d-17-0050.1, 2018.

Zhang, H., Tian, X., Cheng, W., and Jiang, L.: System of Multigrid Nonlinear Least-squares Four-dimensional Variational Data Assimilation for Numerical Weather Prediction (SNAP): System Formulation and Preliminary Evaluation, Adv. Atmos. Sci., 37, 1267-1284, https://doi.org/10.1007/s00376-020-9252-1, 2020.