Supplement of

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

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):

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

where $\mathbf{P}_x^{prior} = \left(\boldsymbol{x}_1^{'}, \boldsymbol{x}_2^{'}, \dots, \boldsymbol{x}_N^{'}\right)$ is an ensemble of prior perturbations, $\boldsymbol{x}_j^{'} = \boldsymbol{x}_j - \boldsymbol{x}_a, j = 1,2,\dots,N$, $\boldsymbol{x}_j$ is the j th perturbation, and N is the number of prior perturbations. According to Evensen (2009), the ensemble of posterior perturbations after assimilation is calculated as follows:

$$\mathbf{P}_x^{post} = \mathbf{P}_x^{prior}\mathbf{V}_2\sqrt{\mathbf{I} - \boldsymbol{\Sigma}_2^{\mathrm{T}}\boldsymbol{\Sigma}_2}\boldsymbol{\Phi}^{\mathrm{T}}, \tag{S3}$$

where

$$\mathbf{U}_2\boldsymbol{\Sigma}_2\mathbf{V}_2^{\mathrm{T}} = \mathbf{X}_2, \tag{S4}$$

$$\mathbf{X}_2 = \boldsymbol{\Lambda}^{-1/2}\mathbf{Z}^{\mathrm{T}}\mathbf{P}_y, \tag{S5}$$

$$\mathbf{Z}\boldsymbol{\Lambda}^{-1}\mathbf{Z}^{\mathrm{T}} = \left[\left(\mathbf{P}_y\right)\left(\mathbf{P}_y\right)^{\mathrm{T}} + (N-1)\mathbf{R}\right]^{-1}. \tag{S6}$$

and $\boldsymbol{\Phi}$ is a random orthogonal matrix, $\mathbf{P}_y = h\left(\mathbf{P}_x^{prior}\right) - h(\boldsymbol{x}_a)$. Then, the prior ($\mathbf{B}^{prior}$) 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). The prior perturbations of the scaling factor in the first inversion window were obtained through historical sampling, and prior perturbations in the following windows were generated through ensemble updating (Tian et al., 2020).

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):

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

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

where we assume that the flux uncertainties are time independent.

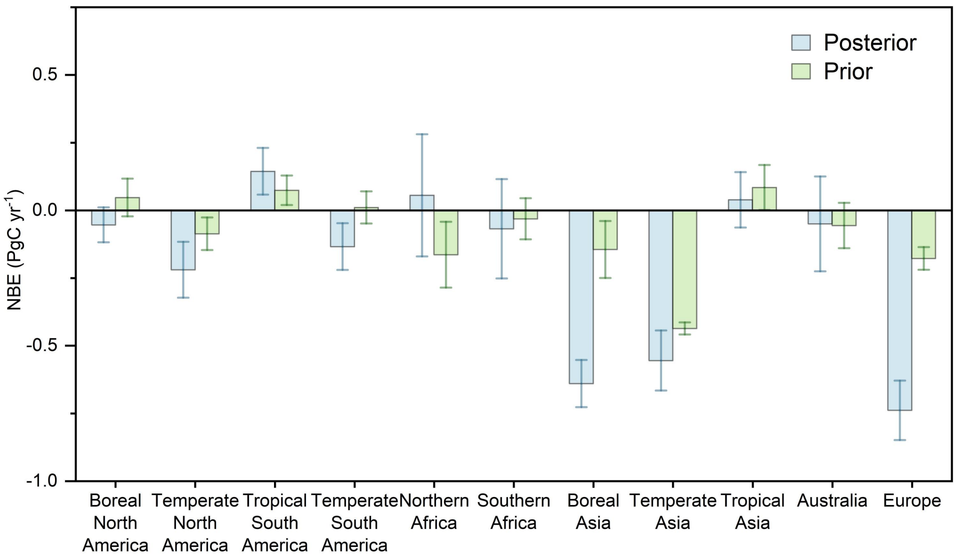

Figure S1. Annual mean (2015–2022) NBE at 11 TransCom land regions from GONGGA prior and posterior estimates. Error bar of NBE represents multi-year standard deviation.

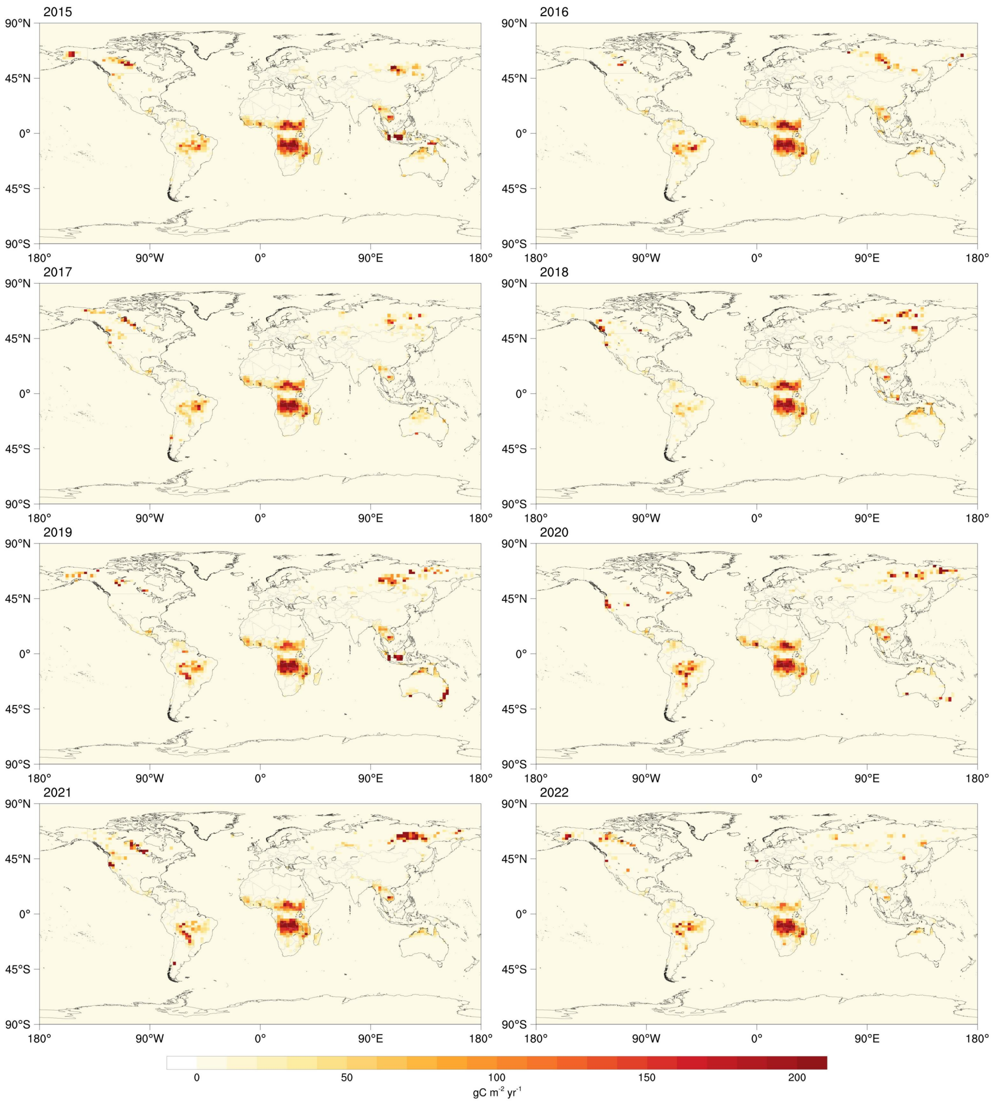

Figure S2. The spatial distribution of biomass burning emissions from GFED4.1s estimate during 2015–2022 period.

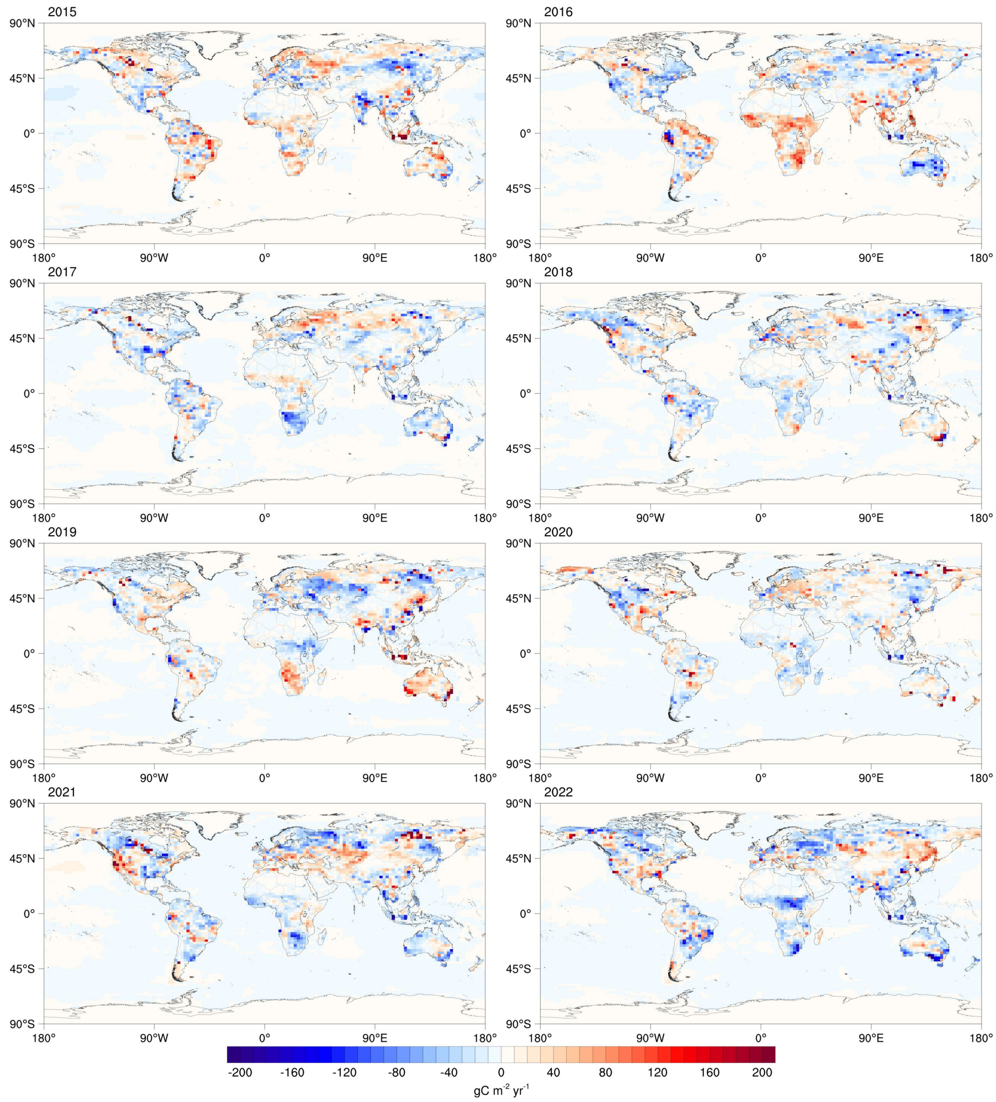

Figure S3. The annual NBE and ocean flux anomalies (annual value minus 8-year mean) during 2015–2022 period.

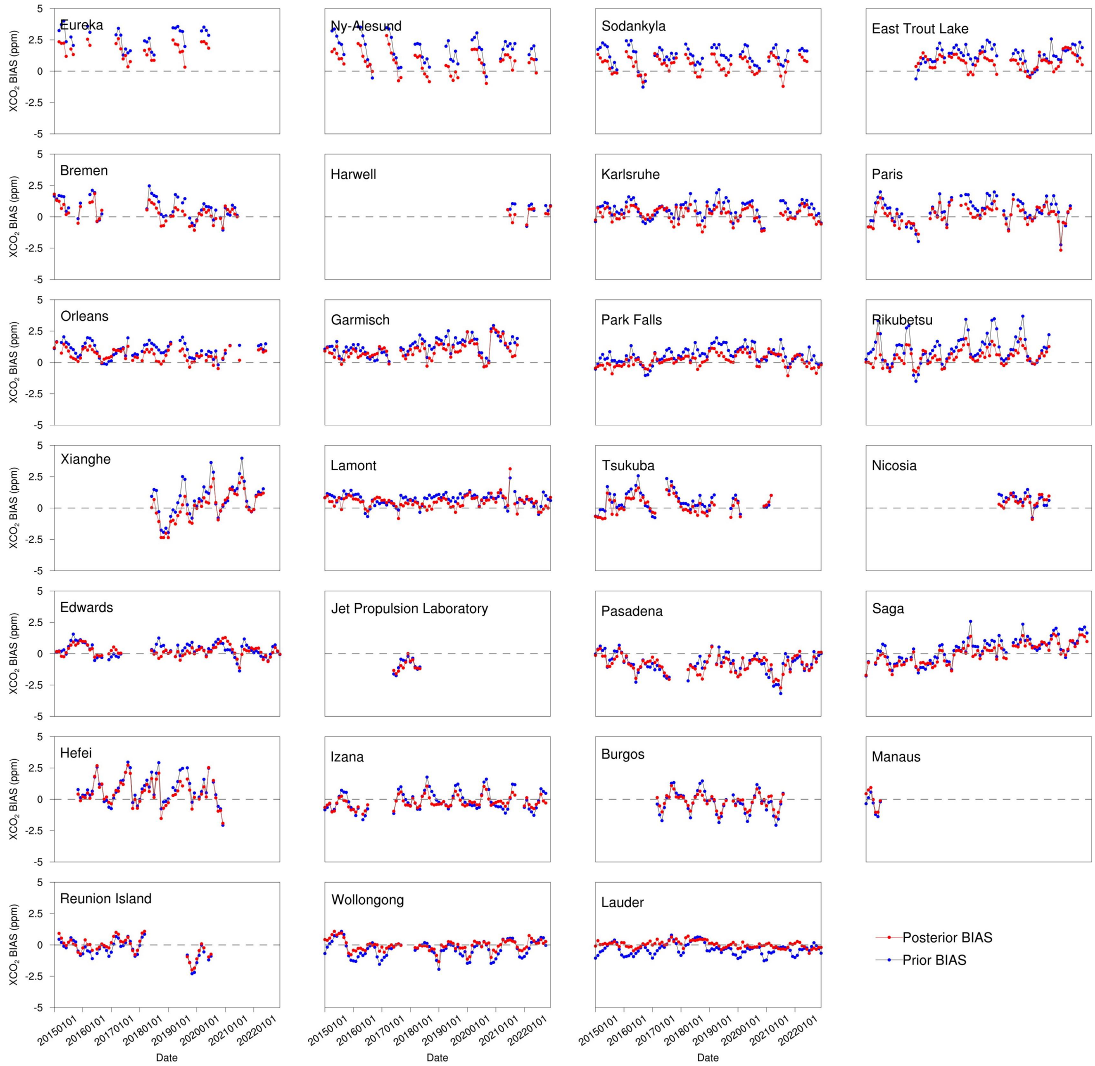

Figure S4. Time series of monthly averaged prior (blue) and posterior (red) simulated XCO_2 BIAS at each TCCON site (prior/posterior simulation – observation).

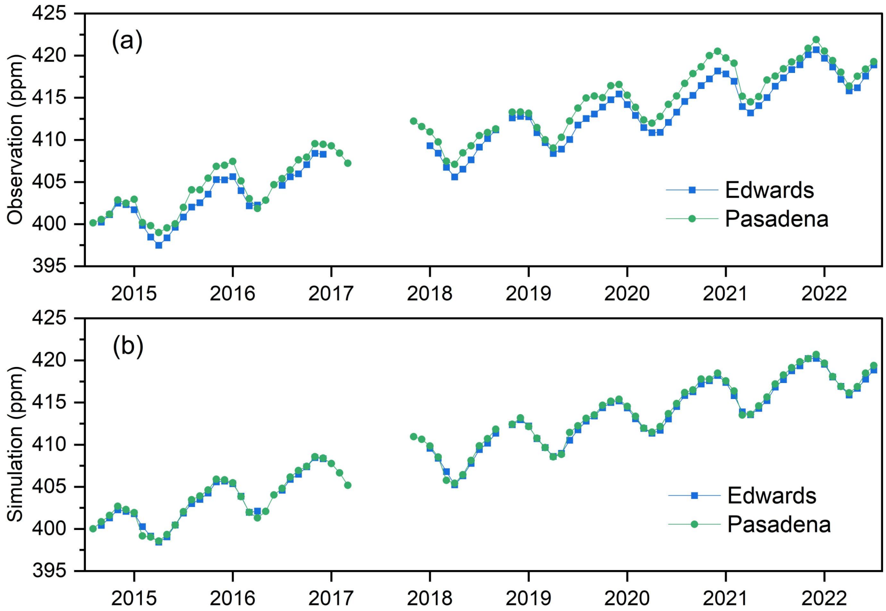

Figure S5. Time series of monthly (a) TCCON observations and (b) corresponding posterior simulations at Edwards (blue) and Pasadena (green) during 2015–2021 period.

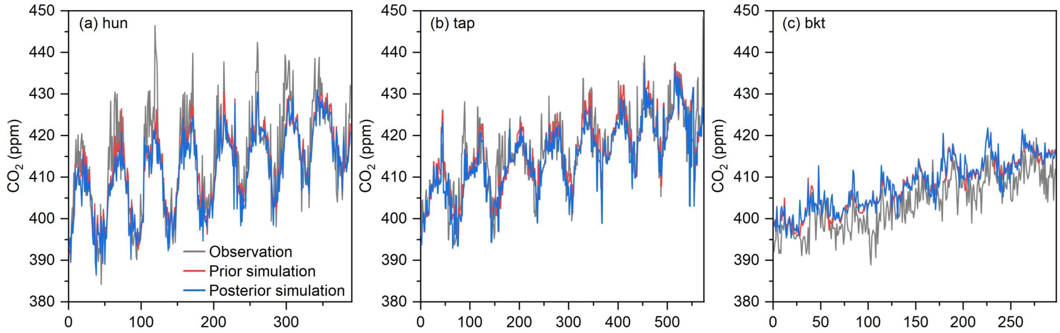

Figure S6. Time series of ObsPack surface flask observations as well as corresponding prior and posterior simulations at three sites that posterior RMSE exceed 4.0 ppm.

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, 2009. Data assimilation: The ensemble kalman filter. Springer-Verlag Berlin Heidelberg, 307 pp.

Niwa Y, Fujii Y. 2020. A conjugatebfgsmethod for accurate estimation of a posterior error covariance matrix in a linear inverse problem. Quarterly Journal of the Royal Meteorological Society, 146(732): 3118-3143

Tian X, Han R, Zhang H. 2020. An adjoint-free alternating direction method for four-dimensional variational data assimilation with multiple parameter tikhonov regularization. Earth and Space Science, 7(11)

Tian X, Zhang H, Feng X, Xie Y. 2018. Nonlinear least squares en4dvar to 4denvar methods for data assimilation: Formulation, analysis, and preliminary evaluation. Monthly Weather Review, 146(1): 77-93