Quantification of regional terrestrial biosphere CO$_2$ flux errors in v10 OCO-2 MIP models using airborne measurements

Jeongmin Yun[1*], Junjie Liu[1*], Brendan Byrne[1], Brad Weir[2], Lesley E. Ott[2], Kathryn McKain[3], Bianca Baier[3], Luciana V. Gatti[4]

[1]Jet Propulsion Laboratory, California institute of Technology, Pasadena, CA 91109, USA
[2]NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[3]NOAA Global Monitoring Laboratory, Boulder, Colorado USA
[4]General Coordination of Earth Science (CGCT), National Institute for Space Research (INPE), São José dos Campos, Brazil

Correspondence to: Jeongmin Yun (jeongmin.yun@jpl.nasa.gov) and Junjie Liu (junjie.liu@jpl.nasa.gov)

Abstract. Multi-inverse modeling inter-comparison projects (MIPs) provide a chance to assess the uncertainties in inversion estimates arising from various sources such as atmospheric CO$_2$ observations, transport models, and prior fluxes. However, accurately quantifying ensemble CO$_2$ flux errors remains challenging, often relying on the ensemble spread as a surrogate. This study proposes a method to quantify the errors of regional terrestrial biosphere CO$_2$ flux estimates from 10 inverse models within the Orbiting Carbon Observatory-2 (OCO-2) MIP by using independent airborne CO$_2$ measurements for the period 2015–2017. We first calculate the root-mean-square error (RMSE) between the ensemble mean of posterior CO$_2$ concentration estimates and airborne observations and then isolate the CO$_2$ concentration error caused solely by the ensemble mean of posterior terrestrial biosphere CO$_2$ flux estimates by subtracting the errors of observation and transport in seven regions. Our analysis reveals significant regional variations in the average monthly RMSE over three years, ranging from 0.90 to 2.04 ppm. The ensemble flux error projected into CO$_2$ space is a major component that accounts for 58-84% of the mean RMSE. We further show that in five regions, the observation-based error estimates exceed the atmospheric CO$_2$ errors computed from the ensemble spread of posterior CO$_2$ flux estimates by 1.37-1.89 times, implying an underestimation of the actual ensemble flux error, while their magnitudes are comparable in two regions. By identifying the most sensitive areas to airborne measurements through adjoint sensitivity analysis, we find that the underestimation of flux errors is prominent in eastern parts of Australia and East Asia, western parts of Europe and Southeast Asia, and midlatitude North America, suggesting the presence of systematic biases related to anthropogenic CO$_2$ emissions in inversion estimates. The regions with no underestimation were southeastern Alaska and northeastern South America. Our study emphasizes the value of independent airborne measurements not only for the overall evaluation of inversion performance but also for quantifying regional errors in ensemble terrestrial biosphere flux estimates.

1 Introduction

Understanding the sources and sinks of atmospheric CO_2 is essential for developing efficient strategies for climate change mitigation and accurate climate predictions. Terrestrial ecosystems have acted as major carbon sinks by absorbing around 30% of anthropogenic fossil and land-use CO_2 emissions over the past few decades (Friedlingstein et al., 2023). However,

35 the absorbed amount is highly variable both spatially and temporally as carbon exchange processes are sensitive to environment and climate change (Liu et al., 2017; Bastos et al., 2019; Piao et al., 2020). To track changes in the terrestrial carbon sinks, accurate estimates of the regional terrestrial carbon fluxes and their uncertainties are crucial.

Atmospheric CO_2 inverse modeling is one of the widely employed approaches to estimate terrestrial and air-sea

40 CO_2 fluxes from observed atmospheric CO_2 concentrations by using a transport model and data assimilation techniques. Most inverse modeling approaches are based on the Bayesian theory, wherein posterior flux estimates are estimated from prior knowledge and atmospheric CO_2 observations weighted by their uncertainties. This approach estimates a posterior probability distribution that can be represented as a maximum a posterori solution (referred to as $\hat{x}$), following the notation of Rodgers (2000). Theoretically, since atmospheric CO_2 observations generally have lower uncertainty than prior terrestrial

45 flux estimates, greater observation data leads to posterior fluxes approaching true values (Liu et al., 2014).

However, concerns have been raised that the inverse modeling results are sensitive to systematic errors in transport models and inversion setups that are not accounted for in the Bayesian framework (Basu et al., 2018; Philip et al., 2019; Schuh et al., 2019). In order to obtain more robust terrestrial flux estimates and assess their uncertainties resulting from

50 various sources (e.g., atmospheric transport and assimilation techniques), inverse modeling intercomparison projects (MIPs) have been conducted. These projects include the TransCom project (Gurney et al., 2004; Houweling et al., 2015), which was first initiated in 1998, as well as subsequent projects such as the Global Carbon Project (GCP; Friedlingstein et al., 2023), the REgional Carbon Cycle Assessment and Processes (RECCAP; Ciais et al., 2022), and the Orbiting Carbon Observatory-2 (OCO-2) MIP (Crowell et al., 2019; Peiro et al., 2022; Byrne et al., 2023). These MIPs involve different inverse modeling

55 groups using state-of-the-art transport modeling and assimilation techniques that assimilate in situ and satellite CO_2 data. Through these MIPs, researchers have analyzed differences in the maximum posteriori solution across models. OCO-2 MIP revealed a general agreement on global flux estimates among ensemble models, but significant discrepancies in regional fluxes, regardless of whether in-situ and/or satellite data are assimilated (Crowell et al., 2019; Peiro et al., 2022).

60 Realistic error quantification of posterior fluxes from atmospheric flux inversions is essential to understand how well the regional fluxes are constrained by current CO_2 observing network and identify regions with high uncertainty, allowing us to prioritize efforts to mitigate the error. The Bayesian formulation provides a method for calculating uncertainties on posterior fluxes based on prior uncertainties and uncertainties in the assimilated data. This can be calculated

65 analytically or approximated using a Monte Carlo method for variational methods (Chevallier et al., 2007; Feng et al., 2009; Liu et al., 2014), however, this is often computationally prohibitive for many inversion systems. This Bayesian posterior uncertainty accounts for random errors in the prior fluxes and observations but does not incorporate systematic errors, thus providing a potential underestimate of the total posterior error.

70 Errors in posterior fluxes are also commonly characterized through comparisons between independent atmospheric CO_2 measurements and posterior atmospheric CO_2 (Houweling et al., 2015; Crowell et al., 2019; Byrne et al., 2023). This approach can provide insights into the biases of current inverse modeling at the global, latitudinal, or site-specific scales. However, as atmospheric CO_2 concentrations are influenced by both local and remote sources, it is difficult to identify regions where the observation-model comparison results are representative. Furthermore, these comparisons do not indicate only posterior flux errors, as the differences between simulated and observed atmospheric CO_2 can arise from errors of not

75 only posterior fluxes, but also transport, representation, and measurement. Because of these limitations, regional posterior flux errors of the ensemble mean have been generally defined as the ensemble spread among ensemble posterior fluxes, but this method lacks an observational and theoretical basis and may not reflect actual errors (Byrne et al., 2023).

80 Our study aims to develop a framework to quantify the error in regional terrestrial biosphere CO_2 fluxes (from now on referred to as terrestrial CO_2 fluxes) estimated from an ensemble of inverse models by using airborne CO_2 measurements, transport modeling, and adjoint sensitivity analysis. Our target ensemble results are derived from 10 ensemble members in v10 OCO-2 MIP for the period 2015–2017, which provide both posterior CO_2 fluxes and posterior CO_2 concentrations sampled at observation sites and times. The ensemble assimilates OCO-2 column-averaged dry-air mole fraction (XCO$_2$) retrievals (ACOS v10; O'Dell et al., 2018) and in situ CO_2 measurements (Tohjima et al., 2005; Nara et al., 2017; Schuldt et

85 al., 2022). We first analyze the spatiotemporal variations and relative magnitudes of various error components that necessitate consideration when comparing airborne CO_2 observations with the ensemble estimates of posterior atmospheric CO_2. Next, we calculate the CO_2 concentration errors caused only by posterior CO_2 fluxes by subtracting the errors of observations and transport from the observation-model differences with long-term observations in seven regions. Then, we identify the most sensitive areas represented by our evaluation results and quantify errors of the ensemble mean annual

90 terrestrial CO_2 flux estimates in these areas.

2 Data and methodology

The aim of this study is to quantify the true errors of the ensemble terrestrial CO_2 fluxes generated by v10 OCO-2 MIP with airborne observations. Here, "error" refers to the magnitude of the differences between the observed and estimated values, without considering the sign. To achieve this, we employ three steps of analysis as described in Figure 1. First, we define

95 two quantities: 1) the root mean square errors (RMSE) between the mean of ensemble posterior CO_2 concentrations and

observed CO_2 concentration, and 2) an approximation of RMSE (ERR_{TOT}) defined as the sum of observation errors and errors in both flux and transport computed from an ensemble spread of posterior CO_2 concentrations (Section 2.2). The ratio between RMSE and ERR_{TOT} is then used to evaluate whether the spread of ensemble posterior fluxes overestimates or underestimates the true errors in the ensemble mean fluxes. Next, we calculate CO_2 uncertainties due to only the spread of

100 OCO-2 MIP ensemble posterior fluxes (Section 2.3). With the estimated flux error projected onto atmospheric CO_2, ERR_{TOT}, and RMSE, we derive the true errors of ensemble posterior fluxes in CO_2 space over seven regions. Then, we identify the areas that these airborne observations are most sensitive to using an adjoint sensitivity analysis and calculate the estimated posterior flux errors over the area in flux space. Assuming a linear observation operator, the study finally computes the true errors of the ensemble mean posterior fluxes over the identified sensitive areas by applying the ratio between the true

105 ensemble posterior errors and the estimated ensemble posterior errors in CO_2 space to the spread of ensemble posterior fluxes.
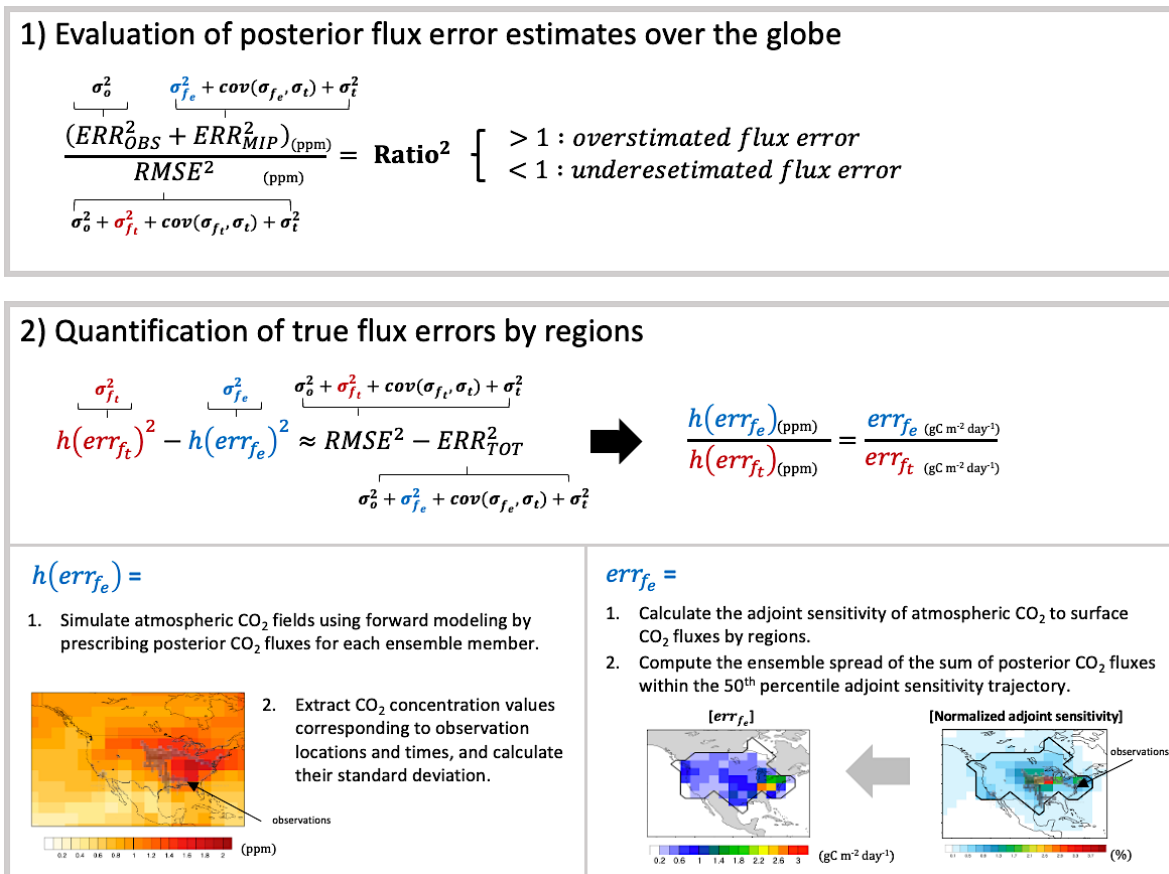
Figure 1: Flow chart summarizing the process of evaluating and quantifying errors in regional ensemble posterior fluxes.

2.1 v10 OCO-2 MIP datasets

The v10 OCO-2 MIP provides multiple results from inverse models that assimilate different combinations of atmospheric CO_2 measurements for 2015–2020. Our study focused on the results from "LNLGIS" experiment, which assimilates the most observations except OCO-2 ocean glint XCO_2 retrievals that cause significant biases on inversion results (Byrne et al., 2023). The "LNLGIS" experiment incorporates v10 OCO-2 land nadir (LN) and glint (LG) XCO_2 retrievals, along with global in situ (IS) data (including surface, ship-based, and airborne measurements) included in the obspack_co2_1_OCO2MIP_v3.2.1_2021-09-14. Ten different inverse modeling groups provided monthly posterior flux estimates interpolated to 1°x1° horizontal resolution and co-sampled posterior atmospheric CO_2 data at the time and location of all types of observations. All of the inversion groups used the same fossil fuel emission dataset, but they independently chose their transport models, assimilation techniques, and prior flux estimates. These details are provided in Table S1, and more detailed explanations for each inverse modeling approach can be found in Byrne et al. (2023). Although the OCO-2 MIP provides data for the period 2015–2020, we used data for the first three years due to the limited number of airborne measurements available during the later years. To minimize the influence of local sources and maximize the influence of regional fluxes, we excluded surface measurements and only considered airborne measurements made between 1 and 5 km altitude.

Figure 2a shows the spatial distribution of the total number of airborne observations used in this study on a 1°x1° grid. The dataset includes two airborne measurement campaigns (Atmospheric Tomography Mission; ATom and O_2/N_2 Ratio and CO_2 Airborne Southern Ocean Study; ORCAS) over the ocean, as well as 19 campaigns over land. Specific airborne campaigns and their references are elaborated in Table 1. The majority of the datasets used in the study were from North America, accounting for 37%, followed by East Asia with 35% and Alaska with 7% for the period of 2015–2017. The duration and extent of the airborne observations vary across different regions. Figure 2b illustrates the number of grid points in each of the seven regions where more than 10 observations per month were available. For Alaska, observations were concentrated during the Arctic-Boreal Vulnerability Experiment (ABoVE) campaign in 2017. North America had observations for most of the analysis period, but observation data was collected over a wider area during the Atmospheric Carbon and Transport – America (ACT–America) campaign. In Europe, sparse observation data was made after intensive observations through the Civil Aircraft for the Regular Investigation of the atmosphere Based on an Instrument Container (IAGOS-CARIBIC) campaign in 2015. The Long-term Comprehensive Observation Network for TRace gases by AIrLiner (CONTRAIL) project provides continuous observation data in East and Southeast Asia from 2015 to 2017, as well as for Australia during 2015–2016. In South America, the largest area of information was obtained when measurements were conducted at six different sites in 2017.
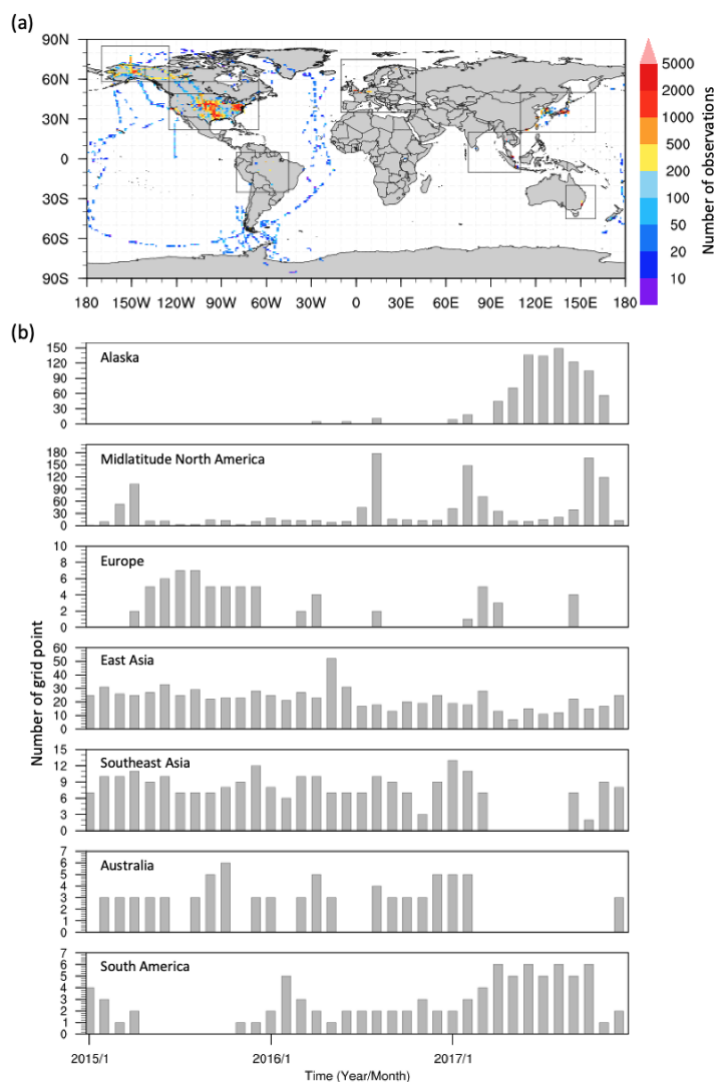
Figure 2: Total number of airborne measurements used in this study and monthly variations in the number of 1°x1° grid-points where observations were made for each region.

145   **Table 1. Data description for each airborne measurement campaign.**

Site code	Site name	Measurement campaign name	Measurement type	Data provider	Dataset identifier	Reference
ACG	Alaska Coast Guard	NOAA/GML Aircraft Program	In situ	National Oceanic and Atmospheric Administration (NOAA) Global Monitoring Laboratory (GML)		

ACT	Atmospheric Carbon and Transport – America (ACT-America)	ACT-America	In situ and flask	NASA Langley Research Center (NASA-LaRC), NOAA/GML	https://doi.org/10.3334/ORNLDAAC/1593	Baier et al. (2020) Wei et al. (2021)
AirCore NOAA	NOAA AirCore	NOAA AirCore Program	Balloon	NOAA/GML	https://doi.org/10.15138/6AV0-MY81	Karion et al. (2010)
ALF	Alta Floresta		Flask	National Institute for Space Research (INPE)	https://doi.org/10.1594/PANGAEA.926834 https://doi.org/10.1594/PANGAEA.949643	Gatti et al. (2023)
AOA	Aircraft Observation of Atmospheric trace gases by JMA		Flask	Japan Meteorological Agency (JMA)		Tsuboi et al. (2013)
CON	Comprehensive Observation Network for TRace gases by AIrLiner (CONTRAIL)		In situ and flask	National Institute for Environmental Studies (NIES), Meteorological Research Institute (MRI)	https://doi.org/10.17595/20190828.001 https://doi.org/10.17595/20190828.002 https://doi.org/10.17595/20180208.001	Machida et al. (2008) https://www.cger.nies.go.jp/contrail/index.html
CRV	Carbon in Arctic Reservoirs Vulnerability Experiment (CARVE)	Arctic-Boreal Vulnerability Experiment (ABoVE)	In situ	NOAA/GML	https://doi.org/10.3334/ORNLDAAC/1582	
DND	Dahlen, North Dakota		Flask	NOAA/GML		Sweeney et al. (2015)
GSFC	NASA Goddard Space Flight Center Aircraft Campaign		In situ	NASA Goddard Space Flight Center (NASA-GSFC)		Kawa et al. (2018)
IAGOS	In-service Aircraft for a Global Observing System	Civil Aircraft for the Regular Investigation of the atmosphere Based on an Instrument Container (IAGOS-CARIBIC)	In situ and flask	Karlsruhe Institute of Technology (IMK-ASF), Institute for Atmospheric and Environmental Sciences (IAU), Max Planck Institute for Biogeochemistry (MPI-BGC)		Filges et al. (2015) https://www.caribic-atmospheric.com/
LARC	LARC - NASA Langley Research Center Aircraft Campaign	Korea-United States Air Quality Study	In situ	NASA-LaRC		
MAN	Manaus		In situ	NOAA/GML		
MRC	Marcellus Pennsylvania		Flask	NOAA/GML		Barkley et al. (2017)
ORC	O2/N2 Ratio and CO2 Airborne Southern Ocean Study (ORCAS)		In situ	National Center for Atmospheric Research (NCAR)	https://doi.org/10.5065/D6SB445X	Stephens et al. (2018)
PAN	Pantanal, Mato Grosso do Sul		Flask	INPE		
RBA-B	Rio Branco		Flask	INPE	https://doi.org/10.1594/PANGAEA.926834 https://doi.org/10.1594/PANGAEA.949643	Gatti et al. (2023)
SAN	Santarem		Flask	INPE	https://doi.org/10.1594/PANGAEA.926834 https://doi.org/10.1594/PANGAEA.949643	Gatti et al. (2023)
SONGNEX2015	Shale Oil and Natural Gas Nexus 2015 (air campaign)		In situ	NOAA Chemical Sciences Laboratory (CSL)		

| **TEF** | Tefe | Flask | INPE | [https://doi.org/10.159 4/PANGAEA.926834](https://doi.org/10.1594/PANGAEA.926834) https://doi.org/10.159 4/PANGAEA.949643 | Gatti et al. (2023) |
| **TOM** | Atmospheric Tomography Mission (ATom) | In situ | NOAA/GML, Harvard University | https://doi.org/10.333 4/ORNLDAAC/1581 | |

2.2 Evaluation of ensemble posterior fluxes

We first employed the two matrixes defined in Eq. (1) and (2) to evaluate ensemble posterior flux errors proposed by Liu et al. (2021). One is RMSE between the ensemble mean of simulated atmospheric CO_2 and the observed one from airborne,
150 which can be written as:

$$RMSE^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\overline{h_\iota(\hat{x})} - y_{o,i}\right]\left[\overline{h_\iota(\hat{x})} - y_{o,i}\right]^T, \text{ where } \overline{h_\iota(\hat{x})} = \frac{1}{M}\sum_{j=1}^{M}h_{i,j}(\hat{x}_j) \qquad (1)$$

$\overline{h_\iota(\hat{x})}$ is the ensemble mean of posterior atmospheric CO_2 sampled at the time and location of the ith airborne observation $y_{o,i}$, within each month. N is the number of airborne measurements within 1°x1° grid-cell and M is the number of ensemble members (i.e., 10). The RMSE indicates the magnitude of the actual CO_2 errors in the ensemble estimates, which is also a
155 quantity broadly used to evaluate the accuracy of posterior fluxes (Crowell et al., 2019; Peiro et al., 2021; Byrne et al., 2023). However, as illustrated in Figure 1 and shown by Liu et al. (2021), $RMSE^2$ includes not only the projection of true flux errors on CO_2 concentration ($\sigma_{f_t}^2$), but also transport errors (σ_t^2), error covariances between the preceding two terms ($cov(\sigma_{f_t},\sigma_t)$), and airborne observation errors (σ_o^2) including representation and measurement errors. Different from Liu et al. (2021), $\overline{h_\iota(\hat{x})}$ is the ensemble mean generated by multiple types of transport models in OCO-2 MIP, accounting for transport
160 errors from meteorological forcing and dynamics. Note, in our analysis, outliers with more than 20 ppm differences between observation and model estimates were excluded from the analysis to obtain robust error estimates.

In practice, the true flux errors are often approximated by the spread of ensemble fluxes, so the projection of true flux errors to CO_2 concentrations and transport errors are approximated by the ensemble spread of the simulated CO_2
165 concentrations in OCO-2 MIP as shown in Appendix A. To evaluate whether this approximation represents the true errors in the ensemble mean fluxes and mean simulated CO_2 concentrations, we define another quantity ERR_{TOT}^2 (Figure 1). Different from RMSE, the variance terms of flux error ($\sigma_{f_e}^2$) and transport error (σ_t^2) and covariance terms between them ($cov(\sigma_{f_e},\sigma_t)$) are replaced by the spread of ensemble (i.e., variance) posterior atmospheric CO_2 concentrations (ERR_{MIP}^2) defined as:

$$ERR_{MIP}^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}\left[h_{i,j}(\hat{x}_j) - \overline{h_\iota(\hat{x})}\right]\left[h_{i,j}(\hat{x}_j) - \overline{h_\iota(\hat{x})}\right]^T \qquad (2)$$

170

In addition, to obtain error information not captured by ERR_{MIP}^2, we additionally calculated ERR_{obs}^2 by combining representation error and measurement error:

$$ERR_{obs}^2 = ERR_{o\_r}^2 + ERR_{o\_m}^2 \qquad (3)$$

It corresponds to the σ_o^2 included in RMSE. $ERR_{o\_r}^2$ indicates the representation error and is defined as within-grid cell variances of estimated atmospheric CO_2 in the models written as:

$$ERR_{o\_r}^2 = \frac{1}{N}\sum_{i=1}^{N} VAR_{CO_2,i} \qquad (4)$$

Using high-resolution (0.5º x 0.625º) 3-hourly GEOS-Chem simulation results for 2018 from NASA Goddard Space Flight Center (Weir et al., 2021), we calculated the variances of atmospheric CO_2 within 2°x2.5° grid cell at a resolution of 0.5º x 0.625º and sampled the values at the corresponding i^{th} airborne measurement times and locations ($VAR_{CO_2,i}$). Subsequently, the monthly mean values of the N co-sampled variances are derived. It is assumed that the variances do not vary significantly across years. The reason for choosing the 2°x2.5°, which is the finest resolution among the OCO-2 MIP members used, is that ERR_{MIP} contains information about the difference in representation errors within the ensemble members. $ERR_{o\_m}^2$ is the observation measurement error. Unfortunately, this information is missing from many of the airborne measurement datasets included in the given OCO-2 MIP ObsPack format, even though uncertainties may be included in the original datasets (e.g., 0.06 ppm for ACT-America; Baier et al., 2020). The World Meteorological Organization (WMO) community has established network compatibility objectives for the precision of atmospheric CO_2 measurements: 0.1 ppm in the Northern Hemisphere and 0.05 ppm in the Southern Hemisphere. Assuming an ideal situation without systematic bias, we set the measurement error (ERR_{o_m}) for all airborne observations at 0.1 ppm. However, in reality, systematic errors could be present in airborne observation data stemming from instrument or setup biases, calibration offsets, and other factors. The potential effects of these systematic errors on our findings will be addressed in Section 4. Therefore, ERR_{TOT}, the approximation for RMSE, is defined as:

$$ERR_{TOT}^2 = ERR_{OBS}^2 + ERR_{MIP}^2 \qquad (5)$$

To evaluate whether the spread of ensemble CO_2 fluxes from OCO-2 MIP represents the true flux errors in the ensemble mean, we calculated the ratio between monthly RMSE and ERR_{TOT}:

$$Ratio^2 = \frac{ERR_{TOT}^2}{RMSE^2}, \qquad (6)$$

If we could assume that the difference between $RMSE^2$ and ERR_{TOT}^2 mainly arises from the difference in the flux error variances ($\sigma_{f_t}^2$ and $\sigma_{f_e}^2$), the ratio can tell us whether the estimated posterior flux errors computed from ensemble spread is an overestimation or underestimation of true errors in the ensemble mean. A ratio close to 1 indicates that the estimated posterior flux errors derived from the ensemble model spread are close to the true posterior flux error in the ensemble mean fluxes. A ratio greater than 1 means that the posterior flux errors are overestimated, and vice versa. Finally, by applying 1000

bootstrap resampling to the monthly grid-based matrices for each region, the regional average of error matrices is obtained, along with its corresponding 95% confidence intervals.

2.3 Quantification of the uncertainties of ensemble posterior fluxes

205   In addition to the qualitative evaluations of posterior flux errors using the ratios between RMSE and ERR$_{TOT}$, we propose a method to quantitatively assess the ensemble posterior flux errors (i.e., variance of flux errors) in both CO_2 space and flux space. To do this, we first need to calculate the variance of atmospheric CO_2 errors due to only the ensemble spread of posterior fluxes from OCO-2 MIP ($h\left(err_{f_e}\right)^2$). As shown in the Appendix A, this term can be written as:

$$h\left(err_{f_e}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{k=1}^{M}\frac{1}{M}\sum_{j=1}^{M}\left[h_k\left(\hat{x}_{k,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right]\left[h_k\left(\hat{x}_{k,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right]^T \tag{7}$$

210   Using all transport models engaged in the OCO-2 MIP would be ideal to derive $h\left(err_{f_e}\right)^2$, but, in this study, we approximated this error term using the GEOS-Chem model as depicted:

$$h\left(err_{f_e}\right)^2 \approx h_{GC}\left(err_{f_e}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}\left[\overline{h_{GC}(\hat{x}_i)} - h_{GC}\left(\hat{x}_{j,i}\right)\right]\left[\overline{h_{GC}(\hat{x}_i)} - h_{GC}\left(\hat{x}_{j,i}\right)\right]^T, \tag{8}$$

where $\overline{h_{GC}(\hat{x}_i)} = \frac{1}{M}\sum_{j=1}^{M}h_{GC}\left(\hat{x}_{j,i}\right)$

215   To get $h_{GC}\left(err_{f_e}\right)^2$, we conducted a set of forward simulations using the GEOS-Chem Adjoint model (v8.2j; Henze et al., 2007). In all ten experiments, consistent meteorology and emission forcing data were used from the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2; Gelaro et al., 2017), Open-source Data Inventory for Anthropogenic CO_2 (ODIAC; Oda and Maksyutov, 2015), and Global Fire Emissions Database version 4 (GFEDv4; Randerson et al., 2015); identical annually balanced 3-hourly terrestrial fluxes from SiB3 (Baker et al., 2010)

220   were also employed. However, in each experiment, the prescribed monthly fluxes of terrestrial ecosystems and oceans were based on the posterior fluxes from the respective ten OCO-2 MIP ensemble members. All experiments were performed at 2°x2.5° horizontal resolution and 47 vertical levels for the period 2015–2017. By calculating the mean of variances of simulated CO_2 concentrations among the ten experiments at i[th] airborne observations within each 1°x1° grid-cell, we derived the $h_{GC}\left(err_{f_e}\right)^2$.

225

    Because we assume that the difference between monthly $RMSE^2$ and ERR_{TOT}^2 mainly arises from the difference in the flux error variances ($\sigma_{f_t}^2$ and $\sigma_{f_e}^2$), the difference between monthly true flux errors ($h\left(err_{f_t}\right)^2$) and estimated flux errors ($h\left(err_{f_e}\right)^2 \approx h_{GC}\left(err_{f_e}\right)^2$) projected onto CO_2 space can be derived from the difference between $RMSE^2$ and ERR_{TOT}^2 as shown:

230    $h\left(err_{f_t}\right)^2 - h\left(err_{f_e}\right)^2 = RMSE^2 - ERR_{TOT}^2$ (9)

From Eq. (9), we can derive the true errors of the ensemble mean fluxes in CO_2 space, $h\left(err_{f_t}\right)^2$. In order to link this term with flux errors in flux space, we first identify the areas that these CO_2 concentrations are sensitive to by conducting sensitivity experiments using the GEOS-Chem Adjoint model. Seven sets of adjoint sensitivity experiments are

235 conducted to examine the sensitivity of airborne measurements in each region (defined in Figure 2a) to terrestrial and ocean CO_2 fluxes for the month of observation. The sensitivity experiments used the same meteorology and CO_2 emission datasets as the forward simulations, along with the ensemble mean of posterior terrestrial and air-sea flux values. The following explanation of the sensitivity analysis uses the same notation as Liu et al. (2015). The cost function (J) is defined as the sum of simulated CO_2 concentrations where airborne observations were made within each region and month:

240 $$J = \sum_{i=1}^{N} h_i(\hat{x}),\qquad(10)$$

The sensitivity of observations to surface fluxes at l^{th} grid-cell and t^{th} time is derived from the partial derivative of J with respect to surface fluxes ($\hat{x}_{l,k}$) written as:

$$\gamma_{l,t} = \frac{\partial J}{\partial \hat{x}_{l,t}},\qquad(11)$$

Monthly cumulative sensitivity (β) with respect to surface fluxes is determined by integrating $\gamma_{l,t}$ from the measurement

245 time (t_0) to the initial time (t_{-T}) for each month:

$$\beta_l = \sum_{t=t_0}^{t-T} \gamma_{l,t},\qquad(12)$$

In order to find the most sensitive areas to the observation, we selected the areas accounting for 50% of the global total values of β. We then calculated the estimated posterior flux errors in flux space ($err_{f_e}^2 = \sigma_{f_e}^2$) by calculating the

250 ensemble spread of the average (and sum) of posterior flux values from OCO-2 MIP within these areas. Regions with sensitivity values lower than 0.1% (0.15% for Alaska, Australia, and Southeast Asia) of the total value of β were excluded due to occasional cases where observations were influenced uniformly across too wide regions as a result of active atmospheric mixing. Additionally, to avoid excessive consideration of localized effects due to a large number of observations occurring in a single location, regions with sensitivity values greater than 1% were included in the effective

255 area.

The observation operator, which converts surface fluxes to atmospheric CO_2, is generally assumed linear, therefore, by multiplying the ratio between $h\left(err_{f_t}\right)$ and $h\left(err_{f_e}\right)$ with err_{f_e}, we obtain the true errors in the ensemble mean fluxes, err_{f_t} (= σ_{f_t}). The equation can be written as:

260 $$err_{f_t} = \frac{h\left(err_{f_t}\right)}{h\left(err_{f_e}\right)} \times err_{f_e}\qquad(13)$$

3 Results

3.1 Spatiotemporal variations in ensemble posterior CO$_2$ concentration errors and major error components

Because the magnitude of terrestrial CO$_2$ fluxes is generally over 10 times greater than air-sea fluxes, the observed

265   atmospheric CO$_2$ over the oceans carries signals from nearby land fluxes. The four ATom campaigns spanning four seasons and the ORCAS campaign during austral summer spanned wide latitudinal ranges, primarily over the oceans, providing a unique opportunity to analyze the latitudinal distributions of inverse modeling errors and contributions of main error sources. We compare the ensemble posterior CO$_2$ to airborne CO$_2$ measurements taken between 1-5 km and then calculate the mean error statistics for the entire campaign period. Comparisons to observations from ATom and ORCAS campaigns reveal a

270   general increase in RMSE values towards the northern high latitudes, reaching 1.2 ppm at 40°N (Figure 3a, f). The latitudinal gradient becomes particularly evident during the summer season, with RMSE values exceeding 2 ppm over North America (Fig. S1), suggesting significant contributions of errors in posterior terrestrial fluxes to the differences between observed and simulated atmospheric CO$_2$. Additionally, consistently elevated RMSE values (>2 ppm) commonly appeared over the west coast of Africa throughout the seasons.
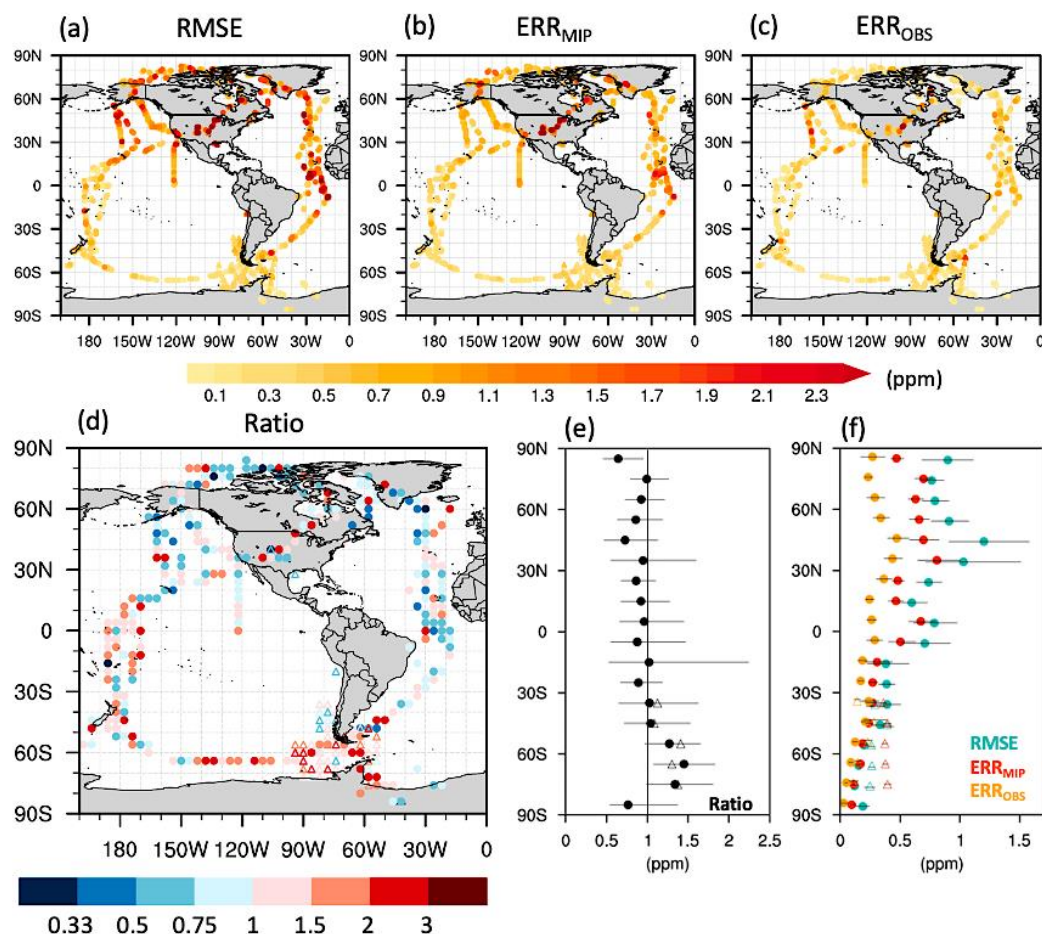
Figure 3: Spatial distributions of (a) RMSE, (b) ERR$_{MIP}$, (c) ERR$_{OBS}$, and (d) Ratio (= $\sqrt{ERR_{OBS}^2 + ERR_{MIP}^2}$ / RMSE) where ATom (circle) and ORCAS (triangle) airborne measurements were taken and (e and f) their latitudinal distributions smoothed by 10° moving average with 95% confidence intervals derived from 1000 bootstrap samples of datasets (error bar).

Both ERR$_{MIP}$ and ERR$_{OBS}$ exhibit similar spatial distributions as RMSE (Figure 3a-c, f). However, ERR$_{MIP}$ has a stronger positive correlation with RMSE (r = 0.60 and 0.81 for ATom and ORCAS, respectively) compared to ERR$_{OBS}$ (r = 0.35 and 0.46), with an average greater magnitude (0.51 and 0.33 ppm) than ERR$_{OBS}$ (0.37 and 0.32 ppm) globally for the whole campaign periods. Particularly, ERR$_{MIP}$ and ERR$_{OBS}$ account for 67% and 33% of the anomalous high RMSE values (2.2 ppm) in Northern America (32-50N and 85-124W), and 84% and 38% of the RMSE values (1.2 ppm) along the east coast of Africa. These findings indicate that ERR$_{MIP}$ which represents errors in posterior fluxes and transport is the most significant factor in explaining RMSE.

Next, in order to assess the proximity of the estimated posterior flux errors, based on the spread of OCO-2 MIP ensemble fluxes, to the true posterior flux errors of the ensemble mean, we compared the root mean square error (RMSE) with the sum of ERR_{MIP} and ERR_{OBS} (referred to as ERR_{TOT}). The ratio of ERR_{TOT} to RMSE exceeds one over the tropical Pacific and the Southern Ocean (Figure 3d, e), indicating that the ensemble spread of posterior fluxes overestimates true flux errors over the regions sensitive to these observations. This overestimation pattern consistently appears for both the ATom and ORCAS campaigns across all seasons (Fig. S2). Airborne CO_2 measurements in this area are predominantly influenced by air-sea fluxes due to the limited land extent and the significant distance from land (Yun et al., 2022), suggesting the true posterior air-sea flux errors may be smaller than the spread of the ensemble posterior flux estimates. In contrast, a ratio of ERR_{TOT} to RMSE less than one was observed along the African coast during the ATom campaigns, with the exception of the 2018 spring campaign conducted in a relatively distant region from Africa. Considering that these airborne observations are known to be sensitive to terrestrial fluxes in tropical Africa (Liu et al., 2022), our results imply that true errors of the ensemble mean terrestrial fluxes in this region may be larger than the estimated errors based on the OCO-2 MIP ensemble spread. These findings agree with Liu et al. (2022), which demonstrated an underestimation of posterior flux errors in CMS-Flux inverse model, indicating most inverse models have common significant errors for this region.

In the northern mid-to-high latitudes, characterized by significant terrestrial CO_2 flux impacts on atmospheric CO_2 variations (Yun et al., 2022), the ratio of ERR_{TOT} to RMSE exhibits substantial variation across space and time. The ratio between ERR_{TOT} to RMSE is greater than one within the North American continent during summer and autumn. However, in other areas, there is a mixed pattern with ratios both below and above one, although the majority of the areas exhibit ratios less than one during winter. These findings highlight that the degree of underestimation or overestimation of true terrestrial flux errors based on ensemble spread can differ depending on regions and seasons, emphasizing the need for a more detailed evaluation of terrestrial flux errors at a regional level based on long-term independent observation data.

3.2 Evaluation of v10 OCO-2 MIP ensemble posterior CO_2 flux errors by regions

In this section, we calculated the regionally averaged monthly error statistics by comparing the ensemble posterior CO_2 to airborne measurements over seven regions for 2015–2017. RMSE values in all these regions exhibit significant monthly variations with a range of 1-3 ppm, with no clear seasonality possibly due to variations in observation routes (Figure 4). Consistent with the results shown in Section 3.1, ERR_{MIP} is the most significant factor explaining the variations of RMSE. Among the seven regions, significant positive correlations ($p<0.05$) between monthly RMSE and ERR_{MIP} exist in Alaska ($r=0.53$), mid-latitude North America ($r=0.45$), Europe ($r=0.68$) and East Asia ($r=0.61$). Furthermore, the correlation coefficient is greater than or comparable to that with ERR_{OBS}. This suggests that in these regions, temporal variations of the errors in posterior fluxes and transport are the major contributors to the temporal variations of RMSE. On the other hand, RMSE did not exhibit a significant correlation with either ERR_{MIP} or ERR_{OBS} in Southeast Asia, Australia, and South

320 America. This implies that the estimated posterior flux errors based on ensemble spread may not represent the temporal variations in true flux errors in those regions.
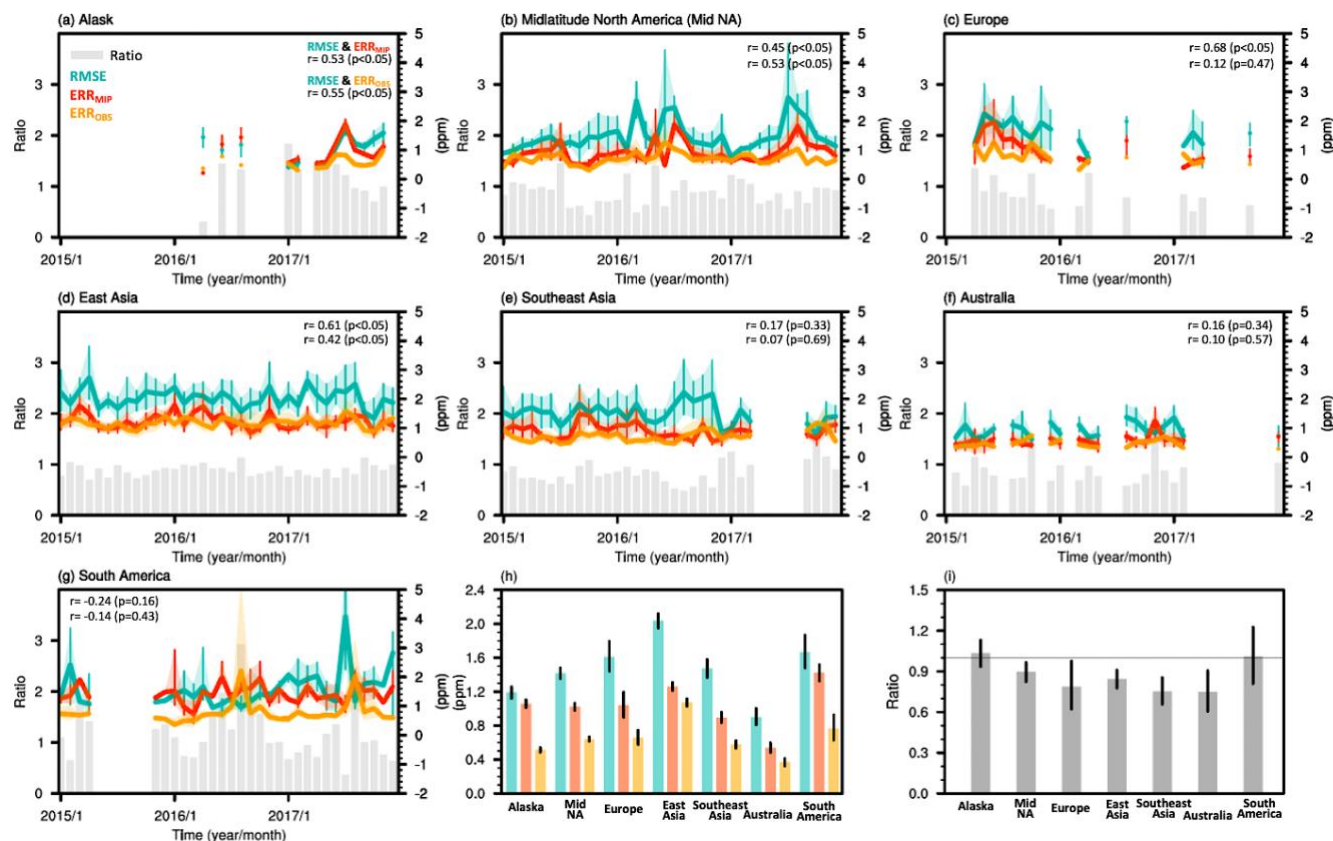
Figure 4: (a-g) Monthly variations of RMSE, ERR$_{MIP}$, ERR$_{OBS}$, and Ratio for each region and (h, i) their mean values for the period
325 **2015–2017. The upper right number in (a-g) indicates the correlation coefficient between RMSE and ERR$_{MIP}$ and ERR$_{OBS}$. The shaded areas and error bars represent the 95% confidence intervals derived from 1000 bootstrap samples of datasets.**

RMSE values exhibit significant variability not only over time but also across regions. The average monthly RMSE is the largest in East Asia (2.04 [1.95, 2.13] ppm: mean [95% confidence intervals]), followed by South America (1.66 [1.46,
330 1.88] ppm) and the lowest in Australia (0.90 [0.81, 0.99] ppm), followed by Alaska (1.19 [1.12, 1.26] ppm). ERR$_{MIP}$ is the primary error component for RMSE, accounting for 59-89% of the RMSE, surpassing the ERR$_{OBS}$ in all the regions by 1.2-2.1 times. In East Asia, the difference between ERR$_{MIP}$ and ERR$_{OBS}$ is relatively small compared to other regions. This could be attributed to the presence of numerous significant carbon sources, particularly along the coastal areas, resulting in increased spatial variability of CO_2 within the coarse grid cell of OCO-2 MIP inverse modeling.
335

The ratio between ERR$_{TOT}$ and RMSE serves as an indicator of how closely the estimated errors based on the ensemble spread of OCO-2 MIP fluxes align with true errors in the ensemble mean fluxes in each region. Our results show that, on average, Alaska and South America exhibited slight overestimation of true flux error with ratios of 1.04 [0.93, 1.14] and 1.01 [0.81, 1.25], respectively, while mid-latitude North America, Europe, East Asia, Southeast Asia, and Australia

340   showed underestimation with ratios of 0.90 [0.83, 0.97], 0.79 [0.62, 0.98], 0.84 [0.78, 0.91], 0.75 [0.66, 0.85], and 0.76 [0.61, 0.90], respectively, throughout the analysis period. Moreover, we observed significant monthly variability in the ratios of ERR$_{TOT}$ to RMSE in regions with diverse campaign durations and routes, such as South America. In contrast, East Asia, characterized by a consistent three-year observation campaign along the same paths, exhibits relatively lower variability. This suggests that the spatial variability in the degree of flux error underestimation or overestimation may be greater than the

345   temporal variability.

3.3 Error quantification of v10 OCO-2 MIP ensemble posterior terrestrial CO$_2$ fluxes by regions

Next, to quantitatively calculate the true regional posterior CO$_2$ flux errors, we computed monthly mean atmospheric CO$_2$ errors due to the estimated posterior flux errors (i.e., $h\left(err_{f_e}\right)$) by conducting forward transport model simulations. The average contributions of monthly $h\left(err_{f_e}\right)^2$ to ERR^2_{MIP} are the highest in Australia (76%), followed by Alaska (65%), and

350   lowest in Europe (34%), followed by East Asia (40%), indicating notable contributions of transport error and/or their covariances with flux error to the ERR^2_{MIP} in the latter regions (Fig. S3). By incorporating the monthly RMSE, ERR$_{TOT}$, and $h\left(err_{f_e}\right)$, we calculated monthly true posterior flux error in CO$_2$ space (i.e., $h\left(err_{f_t}\right)$) for each region during the period 2015–2017 (Figure 5). The $h\left(err_{f_t}\right)$ calculation was applicable in the majority of cases (159 out of 183, 87% of the total), with exceptions for a few instances where the ERR$_{TOT}$ or $h\left(err_{f_e}\right)$ values fell outside the applicable range, especially in

355   South America.
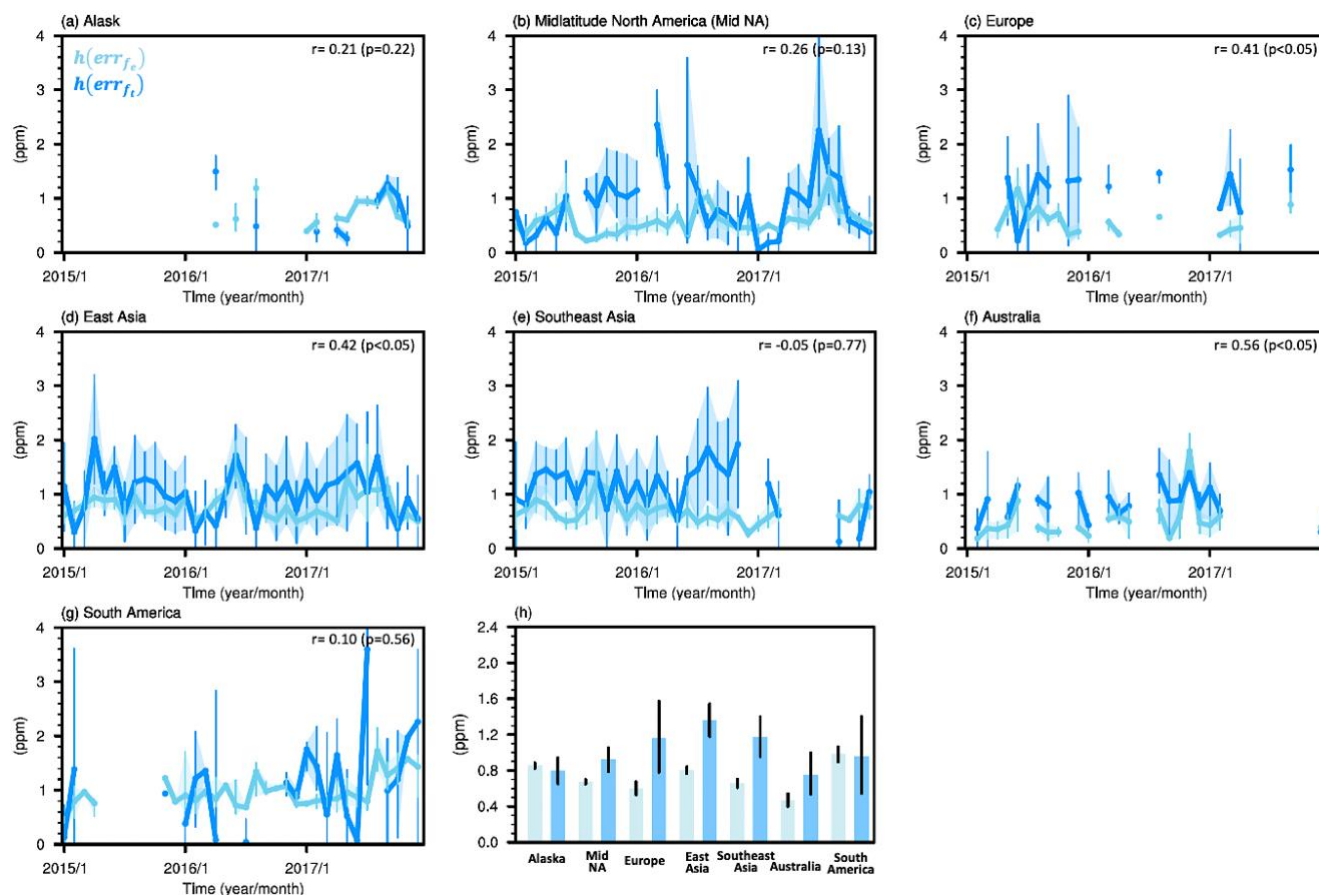
Figure 5: (a-g) Monthly variations of $h(err_{f_e})$ and $h(err_{f_t})$ for each region and (h) their mean values for the period 2015–2017. The upper right number indicates the correlation coefficient between them. The shaded areas and error bars represent the 95% confidence intervals derived from 1000 bootstrap samples of datasets.

360

A significant positive correlation ($p<0.05$) between the regionally averaged monthly true flux errors in CO_2 space and the estimated CO_2 errors is observed in three of seven regions (Figure 5a-g). However, the true flux errors in CO_2 space $h(err_{f_t})$ exhibit distinct patterns of variability compared to the estimated CO_2 errors, $h(err_{f_e})$. In the northern mid-latitude regions, $h(err_{f_e})$ shows clear seasonal cycles for the entire analysis period, despite different observation routes in each

365 month. The seasonal variations are also observed in $h(err_{f_t})$ in Alaska and partially in mid-latitude North America for 2017, but they are not discernible in Europe and East Asia. For example, in East Asia, the growing season (May to October; 1.1 ppm) experiences higher $h(err_{f_e})$ than the non-growing season (November to April; 0.8 ppm), but the difference in $h(err_{f_t})$ between these seasons is only 0.1 ppm. Moreover, $h(err_{f_t})$ displays greater monthly variability than $h(err_{f_e})$. These findings suggest that while the ensemble model spread is capable of capturing general temporal variations of true posterior

370 flux errors in some regions, it may not fully manifest its specific characteristics of seasonal and monthly variability.

The regional average values of the ratios between $h\left(err_{f_e}\right)$ and $h\left(err_{f_t}\right)$ throughout the analysis period show underestimations of true posterior flux errors in mid-latitude North America, Europe, East Asia, Southeast Asia, and Australia by a factor of 0.73 [0.59, 0.86], 0.52 [0.29, 0.77], 0.59 [0.48, 0.70], 0.56 [0.42, 0.71], and 0.61 [0.35, 0.91],

375 respectively (Figure 5h). In contrast, Alaska and South America exhibit a slight overestimation of true flux errors by factors of 1.07 [0.84, 1.31] and 1.03 [0.49, 1.59], respectively, with a particularly large uncertainty range in South America. The regions with underestimation and overestimation align with those identified in the previous analysis based on ratios between ERR_{TOT} and RMSE, but the $h\left(err_{f_e}\right)$ to $h\left(err_{f_t}\right)$ ratios are smaller for underestimation and larger for overestimation. The regional average of $h\left(err_{f_t}\right)$ has a tendency to be larger in regions with higher RMSE values with a maximum in East Asia

380 (1.36 ppm) and a minimum in Australia (0.75 ppm). The $h\left(err_{f_t}\right)$ account for up to 84% of the RMSE in Australia and a minimum of 58% of the RMSE in South America and exceed the observation errors in all regions. The results highlight the substantial contributions of posterior flux errors to the differences between airborne observations and simulated atmospheric CO_2 from OCO-2 MIP ensemble models.

385 Finally, to find the specific areas represented by our evaluation results, we determined the most influential areas for observed atmospheric CO_2 within the seven regions for each month (details on Section 2). The ensemble spread of mean posterior fluxes over the selected areas in each month (err_{f_e}), exhibits a significant correlation (p<0.05) with the monthly $h\left(err_{f_e}\right)$ in all regions, except for Australia where the observational campaign was conducted in specific months (Fig. S4). While the observed atmospheric CO_2 concentration is influenced by both terrestrial and oceanic sources, a comparison of the

390 magnitudes of err_{f_e} between ocean and land within the effective areas reveals that, on average, the terrestrial flux error contributes to more than 95% of the total posterior flux errors in all regions (Fig. S5). This result indicates that our evaluation results based on atmospheric CO_2 can be applied to deriving the actual errors of posterior terrestrial flux within the selected area in flux space. Figure 6a illustrates the selected number of months as effective areas for airborne CO_2 measurements from 2015 to 2017. A higher number of months indicates a greater amount of information from that region

395 was utilized in calculating the regional mean error statistics for the entire analysis period, indicating higher representativeness. To identify primary areas that contribute to the computed mean error statistics, we considered regions that were selected as effective areas for at least eight months or more, accounting for a minimum of 20% of the total analysis period. We calculated the ensemble mean annual terrestrial fluxes for the analysis period within these regions and computed their errors by multiplying the ratio between $h\left(err_{f_e}\right)$ and $h\left(err_{f_t}\right)$ with ensemble spread of regional terrestrial flux

400 estimates (Figure 6b).
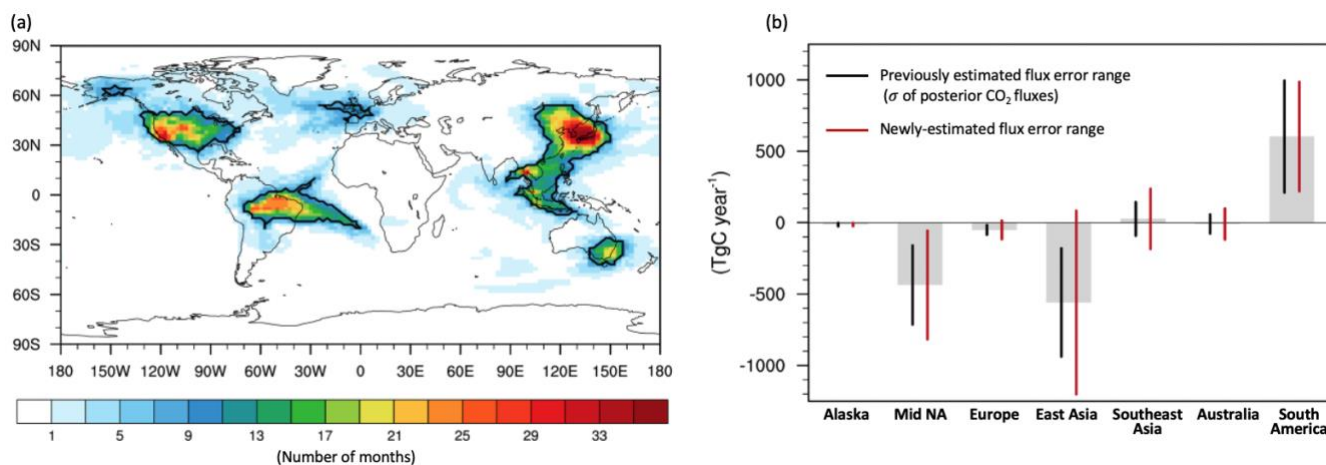
Figure 6: (a) Number of months selected as the effective area for airborne measurements. The outlined area represents selected areas for more than eight months or equal. (b) Annual total terrestrial CO₂ flux obtained from the ensemble mean of ten inversion estimates for each outlined area averaged over the period 2015–2017. The error bars in black and red indicate the one standard deviation of the inversion estimates and the newly estimated error range from this study, respectively.

The airborne observations carried out in mid-latitude North America and East Asia are influenced by a broad region encompassing the United States and eastern parts of East Asia, known for significant terrestrial carbon sinks with estimated fluxes of -436 ± 277 (ensemble mean $\pm$ 1σ) and -559 ± 379 TgC year^{-1}. The recalculated terrestrial flux errors in these regions exceed the ensemble spread with values of 380 and 642 TgC year^{-1}. In Australia, the airborne observations are primarily sensitive to the southeastern part, characterized as a terrestrial carbon sink of -9 ± 67 TgC year^{-1}. The recalculated flux error in this region is also larger than the ensemble spread, estimated at 108 TgC year^{-1}. Observations in Alaska and Europe, conducted over limited periods and specific locations, mainly represent the southeastern region of Alaska and certain areas in the western Europe, with ecosystems serving as carbon sinks (-13 ± 12 and -51 ± 34 TgC year^{-1}). The estimated true flux error is larger than the ensemble spread in Europe with a value of 65 TgC year^{-1}, while it is comparatively smaller in Alaska with a value of 11 TgC year^{-1}. On the contrary, the most influential areas for the observation in Southeast Asia and South America, encompassing the western part of Southeast Asia and the northern part of Brazil, are identified as carbon sources (27 ± 118 and 603 ± 392 TgC year^{-1}, respectively). The true flux error is greater than the ensemble spread in Southeast Asia (210 TgC year^{-1}), while it is slightly lower in South America (382 TgC year^{-1}). The results indicate that the actual errors of ensemble terrestrial flux from v10 OCO-2 MIP are underestimated, particularly in regions with high anthropogenic carbon emissions.

4. Discussion and conclusions

OCO-2 MIP ensembles have given us an opportunity to understand uncertainties of inversion estimates from various error
425   sources. As a vital evaluation tool, airborne measurements have been widely utilized in evaluating the top-down CO_2
inversion results (Houweling et al., 2015; Chevallier et al., 2019; Crowell et al., 2019; Byrne et al., 2023). Further from
conventional qualitative evaluation, which compares airborne observations to posterior CO_2 concentrations, we
quantitatively derive regional posterior flux errors that contribute to the observations and model differences (i.e., RMSE).
Our results show that the true error of the ensemble mean of terrestrial posterior fluxes is a major factor explaining RMSE
430   for the period 2015–2017. Our findings reaffirm the feasibility of evaluating inversion performance on terrestrial flux
estimates through a direct comparison between airborne observations and model data. However, when evaluating inversion
estimates at regional scales, the significance of representation and transport errors become pronounced. Our results show that
regional variations in observation errors, along with the sum of transport errors and their covariances with flux errors, exceed
those in true flux errors projected into CO_2 space, indicating that regional differences in RMSE do not directly correspond to
435   differences in flux errors. For example, on average greater observation and transport errors of around 1.5 times in East Asia
than those in Southeast Asia result in a significant divergence of over 0.8 ppm in RMSE between the two regions, despite a
mere 0.2 ppm disparity in true posterior flux errors in CO_2 space. This result could be a natural consequence as observation
and transport errors arise not only from terrestrial CO_2 fluxes but also from anthropogenic CO_2 sources and atmospheric
circulation (Schuch et al., 2019; Stanevich et al., 2020). Therefore, when utilizing airborne CO_2 measurements (and
440   potentially other CO_2 observational data) to analyze the detailed characteristics of ensemble posterior flux estimates at a
regional (or latitudinal) level, it is crucial to account for the contributions of observation and transport errors.

Our analysis revealed that the true errors of ensemble mean posterior fluxes is greater than the ensemble spread of
flux estimates in five out of seven regions. This underestimation can be attributed to the errors from the common
assumptions and observations applied to all ensemble members, which are not captured by the ensemble spread. OCO-2 MIP
445   models treat the fossil fuel emissions as true values and use the same dataset (i.e., ODIAC). The uncertainty of fossil fuel
emissions is relatively small at national and annual scales (4-20%; Andres et al., 2014), while it becomes substantial when
considering spatial distribution at model grid scale and temporal variability within a year (Zhang et al., 2016; Gurney et al.,
2021). The underestimation of true flux errors in main source regions, along with the absence of underestimation in Alaska
450   and South America, suggests the potential presence of systematic biases originating from errors in fossil fuel emission
estimates. Additionally, the regional and seasonal sampling biases of CO_2 measurements and satellite retrieval errors could
contribute to these systematic biases (Kulawik et al., 2019). Eight prior flux datasets also may not adequately represent the
errors of terrestrial fluxes, which exhibit significant variations among estimates (Feng et al., 2019). Therefore, further study
to uncover the causes of underestimation in true flux errors is required in order to understand uncertainty sources overlooked
455   in current ensemble inverse modeling estimates.

Calculated error statistics display substantial monthly variability, particularly in regions where observations are limited to specific areas. Although our approach is generally effective in estimating monthly true errors of ensemble mean posterior flux, it faces challenges in approximately 15% of cases. This limitation may be attributed to the application of a common method for calculating observation errors across all data points, which might not adequately identify specific outliers. Caution is required when applying our approach to monthly-scale analysis, especially when using observations made in local areas. Extending the calculation period to several months or longer (e.g., Figure 6h) is a suitable strategy for mitigating the impact of outliers and obtaining more robust results. In addition, it should be noted that the number of airborne observations used to calculate the true errors of annual ensemble terrestrial flux varied across regions. Airborne observations were concentrated in the United States and East Asia, covering a wide area with consistent monitoring. However, other regions, particularly Europe and Alaska, had sparse and intermittent data coverage. This implies that the reliability of the error estimation results in the latter case could be comparatively lower.

This study computes true flux errors for ensemble mean estimates by comparing RMSE and ERR$_{TOT}$. In this calculation, we made an assumption that the difference between $RMSE^2$ and ERR_{TOT}^2 comes from the difference between actual and estimated flux error variances. However, discrepancies between true and estimated values of not only observation and transport errors, as well as covariances between flux errors and transport errors, could contribute to variations in $RMSE^2$ and ERR_{TOT}^2. First, due to a lack of information, we set observation measurement errors under ideal conditions. In reality, inadequate quality control can result in significant systematic bias for specific regions and time periods (Masarie et al., 2011), impacting our results. For instance, if this led to an average measurement error of 0.5 ppm during the analysis period, the calculated true flux error would decrease from 380 to 303 TgC year^{-1} for mid-latitude North America and from 392 to 320 TgC year^{-1} for South America. Second, we derived observation representation error and $h(err_{f_e})$ using a GEOS-Chem model but these values depend on the transport model and meteorological field used. Employing our approach across all participating MIP models to compute these two error terms and subsequently averaging them would lead to a more realistic flux error quantification in future studies. In addition, employing all transport models would facilitate the calculation of variances of flux errors and their covariance with transport errors included in ERR$_{MIP}$ as shown in Appendix A, and subsequently enable the determination of the total true flux error including both diagonal and off-diagonal terms. Lastly, previous studies show that 8-10 different ensemble members are required for robust transport error estimates (Feng et al., 2019; Lauvaux et al., 2019). However, out of the 10 ensemble members in OCO-2 MIP, three employed TM5 and five utilized GEOS-Chem (Table S1). The ensemble size might not be enough to fully capture the range of true transport errors, therefore, the participation of inverse modeling groups using other transport models or meteorological forcing data might contribute to estimating transport errors closer to actual values.

In summary, our study provides an observation-based method for quantifying ensemble posterior errors of regional

490   terrestrial biosphere CO_2 fluxes which can be widely applied in inverse modeling inter-comparison projects, including the

OCO-2 MIP. The evaluation results of the OCO-2 MIP ensemble members reveal the true errors of ensemble posterior

fluxes are larger compared to the ensemble spread in regions with high anthropogenic emissions. This finding offers

important insights into understanding the sources of errors in current inverse modeling and guides future research aimed at

resolving the errors in terrestrial CO_2 fluxes. Airborne observations provide a broader footprint compared to ground-based

495   observations. Leveraging this advantage, our study evaluated 19% of the total land (excluding Antarctica and Greenland) but

data scarcity limited the evaluation of the remaining 81%. In addition to the ongoing airborne measurement programs

including CONTRAIL, IAGOS-CARIBIC, and various airborne programs under National Oceanic and Atmospheric

Administration (NOAA), airborne observations have been conducted in unexplored regions, including Siberia (e.g., Narbaud

et al., 2023), Africa (e.g., Barker et al., 2020), and Northern Europe (e.g., Barker et al., 2021). The sustained efforts to

500   maintain and expand airborne observations along with a collaborative data-sharing and management system (e.g., ObsPack)

will contribute to accurately estimating and reducing the uncertainties of regional terrestrial CO_2 fluxes.

Appendix A

Following Eq. (1) in the main text,

$$RMSE^2 = \frac{1}{N}\sum_{i=1}^{N}[y_{o,i} - \overline{h(\hat{x}_\iota)}]\,[y_{o,i} - \overline{h(\hat{x}_\iota)}]^T, \qquad \text{where } \overline{h(\hat{x}_\iota)} = \frac{1}{M}\sum_{j=1}^{M}h_j(\hat{x}_{j,i}) \qquad (A1)$$

505   where $\overline{h(\hat{x}_\iota)}$ denotes ensemble mean of posterior CO_2 concentrations in OCO-2 MIP models at i^{th} airborne observation ($y_{o,i}$).

N is the number of airborne measurements in each 1°x1° grid cell and M is the ensemble size (i.e., 10 members). The Eq.

(A1) can be rewritten as,

$$RMSE^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\left(y_{o,i} - h_t(\hat{x}_{t,i})\right) - \left(\overline{h(\hat{x}_\iota)} - h_t(\hat{x}_{t,i})\right)\right]\left[\left(y_o - h_t(\hat{x}_{t,i})\right) - \left(\overline{h(\hat{x}_\iota)} - h_t(\hat{x}_{t,i})\right)\right]^T \qquad (A2)$$

$$= \frac{1}{N}\sum_{i=1}^{N}[y_{o,i} - h_t(\hat{x}_{t,i})]\,[y_{o,i} - h_t(\hat{x}_{t,i})]^T - 2\left(y_o - h_t(\hat{x}_{t,i})\right) * \left(\overline{h(\hat{x}_\iota)} - h_t(\hat{x}_{t,i})\right)$$

510
$$+\left[\overline{h(\hat{x}_\iota)} - h_t(\hat{x}_{t,i})\right]\left[\overline{h(\hat{x}_\iota)} - h_t(\hat{x}_{t,i})\right]^T, \qquad (A3)$$

where $h_t(\hat{x}_t)$ denotes the estimated CO_2 concentration obtained from an error-free atmospheric transport model (h_t) and true

CO_2 fluxes ($\hat{x}_t$). The three terms on the right-hand side of Eq. (A3) indicate the (i) variances of observation error (ii)

covariance between observation error and errors of flux and transport, and (iii) variances of flux and transport errors in the

ensemble estimates, respectively. Assuming the independence of observation errors from transport and flux errors, Eq. (A3)

515   can be simplified to:

$$RMSE^2 = \frac{1}{N}\sum_{i=1}^{N}[y_{o,i} - h_t(\hat{x}_{t,i})]\,[y_{o,i} - h_t(\hat{x}_{t,i})]^T + \left[\overline{h(\hat{x}_\iota)} - h_t(\hat{x}_{t,i})\right]\left[\overline{h(\hat{x}_\iota)} - h_t(\hat{x}_{t,i})\right]^T \qquad (A4)$$

Further, the second term on the right-hand side of Eq. (A4) can be rewritten by separating the flux error and transport error terms as follows:

$$520 \quad \frac{1}{N}\sum_{i=1}^{N}\left[\overline{h(\hat{x}_i)} - h_t(\hat{x}_{t,i})\right]\left[\overline{h(\hat{x}_i)} - h_t(\hat{x}_{t,i})\right]^T$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right) - \left(h_t(\hat{x}_{t,i}) - \overline{h(\hat{x}_{t,i})}\right)\right]\left[\left(\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right) - \left(h_t(\hat{x}_{t,i}) - \overline{h(\hat{x}_{t,i})}\right)\right]^T \quad (A5)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right]\left[\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right]^T - 2\left(\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right)\left(h_t(\hat{x}_{t,i}) - \overline{h(\hat{x}_{t,i})}\right)$$

$$+ \left[h_t(\hat{x}_{t,i}) - \overline{h(\hat{x}_{t,i})}\right]\left[h_t(\hat{x}_{t,i}) - \overline{h(\hat{x}_{t,i})}\right]^T \quad (A6)$$

The three terms on the right-hand side of Eq. (A6) indicate the (i) variances of flux error in concentration space (ii)
525 covariances between flux error and transport error, and (iii) variances of transport error, respectively.

In this study, the values of the first and second terms on the right-hand side of Eq. (A4) are estimated by the sum of observation representative and measurement errors (ERR_{OBS}^2), and the ensemble spread of posterior CO_2 concentrations across OCO-2 MIP models (ERR_{MIP}^2), respectively. When the ensemble spread represents the actual ensemble mean of
530 transport and posterior flux errors in OCO-2 MIP models, Eq (A4) can be reformulated as follows:

$$RMSE^2 \approx ERR_{TOT}^2 = ERR_{OBS}^2 + ERR_{MIP}^2 \quad (A7)$$

ERR_{MIP}^2 can be also rewritten by separating flux error and transport error terms as follows:

$$ERR_{MIP}^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}\left[\overline{h(\hat{x}_i)} - h_j(\hat{x}_{j,i})\right]\left[\overline{h(\hat{x}_i)} - h_j(\hat{x}_{j,i})\right]^T \quad (A8)$$

$$535 \quad = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}\frac{1}{M}\sum_{k=1}^{M}\left[\left(h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right) - \left(h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right)\right]$$

$$\left[\left(h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right) - \left(h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right)\right]^T \quad (A9)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{k=1}^{M}\frac{1}{M}\sum_{j=1}^{M}\left[h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right]\left[h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right]^T$$

$$-2\left(h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right)\left(h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right)$$

$$+\left[h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right]\left[h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right]^T \quad (A10)$$

540 Same as Eq. (A6), the three terms on the right-hand side of Eq. (A10) correspond to the approximated (i) variances of flux error, (ii) covariances between flux error and transport error, and (iii) variances of transport error, respectively. This study assumes that the difference between $RMSE^2$ and ERR_{TOT}^2 mainly arises from the difference in the first term, the variance of flux error. For the calculation of this term, utilizing all participating transport models in the OCO-2 MIP would be ideal but, in this study, we approximated it using the GEOS-Chem model.

545 **Code and Data availability**

The inverse modelling results and airborne CO_2 measurement data involved in v10 OCO-2 MIP project are available at https://www.gml.noaa.gov/ccgg/OCO2_v10mip/download.php. The high-resolution global GEOS-Chem simulation results used to calculate representation error can be obtained from Brad Weir (brad.weir@nasa.gov) and Lesley Ott (lesley.e.ott@nasa.gov) upon request. The forward and adjoint sensitivity simulations for this work were conducted using the

550 publicly available GEOS-Chem Adjoint model. The model can be downloaded from http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_Adjoint (Henze et al., 2007; last accessed: 29 Jun 2023).

Author contributions

JY and JL designed this study and JY performed the analyses. JL, BB, BW, KM, and BB reviewed and provided input to the manuscript. BW and LEO provided high-resolution global GEOS-Chem simulation results. KM, BB, and LVG provided

555 airborne CO_2 observations. JY led the writing with input from all coauthors.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

This research has been supported by the National Aeronautics and Space Administration (grant no. 20-OCOST20-0012). We

560 are grateful for the airborne CO_2 observation data provision by NOAA, NASA-LaRC, NASA-GSFC, INPE, JMA, NIES, MRI, the European research infrastructure IAGOS-CARIBIC, and Harvard University. We thank to C. Sweeney, K. Mckain, K., Davis, J. P. Digangi, M. Y. Martin, B. Baier, L. V. Gatti, J. B. Miller, E. Gloor, W. Peters, K. Saito, T. Machida, H. Matsueda, Y. Sawa, T. Niwa, E. Dlugokencky, S. R. Kawa, J. B. Abshire, H. Riris, F. Obersteiner, H. Boenisch, T. Gehrlein, A. Zahn, C. Gerbig, T. Schuck, G. Chen, M. Shook, G. A. Martins, R. A. F. de Souza, B. Stephens, E. Kort, T. Ryerson, J.

565 Peischl, K. Aikin, S. Wofsy, B. Daube, R. Commane. We also appreciate all of the data providers of the datasets used in GEOS-Chem model simulations. Especially, we thank all research groups involved in OCO-2 MIP contributing to provide v10 OCO-2 and in-situ CO_2 datasets and inverse modeling outputs.

References

1. Andres, R. J., Boden, T. A., and Higdon, D.: A new evaluation of the uncertainty associated with CDIAC estimates

570   of fossil fuel carbon dioxide emission, Tellus B: Chemical and Physical Meteorology, 66, 23616, https://doi.org/10.3402/tellusb.v66.23616, 2014.

2. Baier, B. C., Sweeney, C., Choi, Y., Davis, K. J., DiGangi, J. P., Feng, S., Fried, A., Halliday, H., Higgs, J., Lauvaux, T., Miller, B. R., Montzka, S. A., Newberger, T., Nowak, J. B., Patra, P., Richter, D., Walega, J., and Weibring, P.: Multispecies Assessment of Factors Influencing Regional CO_2 and CH_4 Enhancements During the Winter 2017 ACT-America Campaign, Journal of Geophysical Research: Atmospheres, 125, e2019JD031339, https://doi.org/10.1029/2019JD031339, 2020.

3. Baker, I. T., Denning, A. S., and Stöckli, R.: North American gross primary productivity: regional characterization and interannual variability, 62, 533–549, https://doi.org/10.1111/j.1600-0889.2010.00492.x, 2010.

4. Barker, P. A., Allen, G., Gallagher, M., Pitt, J. R., Fisher, R. E., Bannan, T., Nisbet, E. G., Bauguitte, S. J.-B., Pasternak, D., Cliff, S., Schimpf, M. B., Mehra, A., Bower, K. N., Lee, J. D., Coe, H., and Percival, C. J.: Airborne measurements of fire emission factors for African biomass burning sampled during the MOYA campaign, Atmospheric Chemistry and Physics, 20, 15443–15459, https://doi.org/10.5194/acp-20-15443-2020, 2020.

5. Barker, P. A., Allen, G., Pitt, J. R., Bauguitte, S. J.-B., Pasternak, D., Cliff, S., France, J. L., Fisher, R. E., Lee, J. D., Bower, K. N., and Nisbet, E. G.: Airborne quantification of net methane and carbon dioxide fluxes from European Arctic wetlands in Summer 2019, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 380, 20210192, https://doi.org/10.1098/rsta.2021.0192, 2021.

6. Bastos, A., Ciais, P., Chevallier, F., Rödenbeck, C., Ballantyne, A. P., Maignan, F., Yin, Y., Fernández-Martínez, M., Friedlingstein, P., Peñuelas, J., Piao, S. L., Sitch, S., Smith, W. K., Wang, X., Zhu, Z., Haverd, V., Kato, E., Jain, A. K., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Peylin, P., Poulter, B., and Zhu, D.: Contrasting effects of CO_2 fertilization, land-use change and warming on seasonal amplitude of Northern Hemisphere CO_2 exchange, Atmospheric Chemistry and Physics, 19, 12361–12375, https://doi.org/10.5194/acp-19-12361-2019, 2019.

7. Basu, S., Baker, D. F., Chevallier, F., Patra, P. K., Liu, J., and Miller, J. B.: The impact of transport model differences on CO_2 surface flux estimates from OCO-2 retrievals of column average CO_2, Atmospheric Chemistry and Physics, 18, 7189–7215, https://doi.org/10.5194/acp-18-7189-2018, 2018.

8. 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. K., Feng, S., García, O. E., 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., Velazco, V. A., Vrekoussis, M., Warneke, T., Worden, J. R., Wunch, D., Yao, Y., Yun, J., Zammit-Mangion, A., and Zeng, N.: Pilot top-down CO_2 Budget constrained by the v10 OCO-2 MIP Version 1, Committee on Earth Observing Satellites [data set], https://doi.org/10.48588/npf6-sw92, Version 1.0, 2022.

9. 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. K., Feng, S., García, O. E.,

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., Velazco, V. A., Vrekoussis, M., Warneke, T., Worden, J. R., Wunch, D., Yao, Y., Yun, J., Zammit-Mangion, A., and Zeng, N.: National CO_2 budgets (2015–2020) inferred from atmospheric CO_2 observations in support of the global stocktake, Earth System Science Data, 15, 963–1004, https://doi.org/10.5194/essd-15-963-2023, 2023.

10. Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2006JD007375, 2007.

11. Chevallier, F., Feng, L., Bösch, H., Palmer, P. I., and Rayner, P. J.: On the impact of transport model errors for the estimation of CO_2 surface fluxes from GOSAT observations, Geophysical Research Letters, 37, https://doi.org/10.1029/2010GL044652, 2010.

12. Chevallier, F., Remaud, M., O'Dell, C. W., Baker, D., Peylin, P., and Cozic, A.: Objective evaluation of surface- and satellite-driven carbon dioxide atmospheric inversions, Atmospheric Chemistry and Physics, 19, 14233–14251, https://doi.org/10.5194/acp-19-14233-2019, 2019.

13. Ciais, P., Bastos, A., Chevallier, F., Lauerwald, R., Poulter, B., Canadell, J. G., Hugelius, G., Jackson, R. B., Jain, A., Jones, M., Kondo, M., Luijkx, I. T., Patra, P. K., Peters, W., Pongratz, J., Petrescu, A. M. R., Piao, S., Qiu, C., Von Randow, C., Regnier, P., Saunois, M., Scholes, R., Shvidenko, A., Tian, H., Yang, H., Wang, X., and Zheng, B.: Definitions and methods to estimate regional land carbon fluxes for the second phase of the REgional Carbon Cycle Assessment and Processes Project (RECCAP-2), Geoscientific Model Development, 15, 1289–1316, https://doi.org/10.5194/gmd-15-1289-2022, 2022.

14. 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, Atmospheric Chemistry and Physics, 19, 9797–9831, https://doi.org/10.5194/acp-19-9797-2019, 2019.

15. Cui, Y. Y., Jacobson, A. R., Feng, S., Wesloh, D., Barkley, Z. R., Zhang, L., Gerken, T., Keller, K., Baker, D., and Davis, K. J.: Evaluation of CarbonTracker's Inverse Estimates of North American Net Ecosystem Exchange of CO_2 From Different Observing Systems Using ACT-America Airborne Observations, Journal of Geophysical Research: Atmospheres, 126, e2020JD034406, https://doi.org/10.1029/2020JD034406, 2021.

16. Deng, F., Jones, D. B. A., Henze, D. K., Bousserez, N., Bowman, K. W., Fisher, J. B., Nassar, R., O'Dell, C., Wunch, D., Wennberg, P. O., Kort, E. A., Wofsy, S. C., Blumenstock, T., Deutscher, N. M., Griffith, D. W. T., Hase, F., Heikkinen, P., Sherlock, V., Strong, K., Sussmann, R., and Warneke, T.: Inferring regional sources and sinks of

atmospheric CO_2 from GOSAT XCO_2 data, Atmospheric Chemistry and Physics, 14, 3703–3727,
640 https://doi.org/10.5194/acp-14-3703-2014, 2014.

17. Feng, L., Palmer, P. I., Bösch, H., and Dance, S.: Estimating surface CO_2 fluxes from space-borne CO_2 dry air mole
fraction observations using an ensemble Kalman Filter, Atmospheric Chemistry and Physics, 9, 2619–2633,
https://doi.org/10.5194/acp-9-2619-2009, 2009.

18. Feng, S., Lauvaux, T., Davis, K. J., Keller, K., Zhou, Y., Williams, C., Schuh, A. E., Liu, J., and Baker, I.: Seasonal
645 Characteristics of Model Uncertainties From Biogenic Fluxes, Transport, and Large-Scale Boundary Inflow in
Atmospheric CO_2 Simulations Over North America, Journal of Geophysical Research: Atmospheres, 124, 14325–
14346, https://doi.org/10.1029/2019JD031165, 2019.

19. Filges, A., Gerbig, C., Chen, H., Franke, H., Klaus, C., and Jordan, A.: The IAGOS-core greenhouse gas package: a
measurement system for continuous airborne observations of CO_2, CH_4, H_2O and CO, Tellus B: Chemical and
650 Physical Meteorology, 67, 27989, https://doi.org/10.3402/tellusb.v67.27989, 2015.

20. Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T.,
Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R.
B., Alin, S. R., Alkama, R., Arneth, A., Arora, V. K., Bates, N. R., Becker, M., Bellouin, N., Bittig, H. C., Bopp, L.,
Chevallier, F., Chini, L. P., Cronin, M., Evans, W., Falk, S., Feely, R. A., Gasser, T., Gehlen, M., Gkritzalis, T.,
655 Gloege, L., Grassi, G., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina,
T., Jain, A. K., Jersild, A., Kadono, K., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I.,
Landschützer, P., Lefèvre, N., Lindsay, K., Liu, J., Liu, Z., Marland, G., Mayot, N., McGrath, M. J., Metzl, N.,
Monacci, N. M., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pan, N., Pierrot, D.,
Pocock, K., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Rodriguez, C., Rosan, T. M., Schwinger, J.,
660 Séférian, R., Shutler, J. D., Skjelvan, I., Steinhoff, T., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tanhua, T.,
Tans, P. P., Tian, X., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., Walker, A. P., Wanninkhof,
R., Whitehead, C., Willstrand Wranne, A., et al.: Global Carbon Budget 2022, Earth System Science Data, 14, 4811–
4900, https://doi.org/10.5194/essd-14-4811-2022, 2022.

21. Gatti, L. V., Cunha, C. L., Marani, L., Cassol, H. L. G., Messias, C. G., Arai, E., Denning, A. S., Soler, L. S., Almeida,
665 C., Setzer, A., Domingues, L. G., Basso, L. S., Miller, J. B., Gloor, M., Correia, C. S. C., Tejada, G., Neves, R. A. L.,
Rajao, R., Nunes, F., Filho, B. S. S., Schmitt, J., Nobre, C., Corrêa, S. M., Sanches, A. H., Aragão, L. E. O. C.,
Anderson, L., Von Randow, C., Crispim, S. P., Silva, F. M., and Machado, G. B. M.: Increased Amazon carbon
emissions mainly from decline in law enforcement, Nature, 621, 318–323, https://doi.org/10.1038/s41586-023-
06390-0, 2023.

670 22. 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., 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. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1,

675   2017.

23.   Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Peylin, P., Prather, M., and Taguchi, S.: Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, Global Biogeochemical Cycles, 18, https://doi.org/10.1029/2003GB002111, 2004.

680   24.   Gurney, K. R., Liang, J., Roest, G., Song, Y., Mueller, K., and Lauvaux, T.: Under-reporting of greenhouse gas emissions in U.S. cities, Nat Commun, 12, 553, https://doi.org/10.1038/s41467-020-20871-0, 2021.

25.   Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmospheric Chemistry and Physics, 7, 2413–2433, https://doi.org/10.5194/acp-7-2413-2007, 2007.

26.   Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., Deng, F., Dlugokencky, E. J., Feng, L.,
685   Ganshin, A., Hasekamp, O., Jones, D., Maksyutov, S., Marshall, J., Oda, T., O'Dell, C. W., Oshchepkov, S., Palmer, P. I., Peylin, P., Poussi, Z., Reum, F., Takagi, H., Yoshida, Y., and Zhuravlev, R.: An intercomparison of inverse models for estimating sources and sinks of CO_2 using GOSAT measurements, Journal of Geophysical Research: Atmospheres, 120, 5253–5266, https://doi.org/10.1002/2014JD022962, 2015.

27.   Karion, A., Sweeney, C., Tans, P., and Newberger, T.: AirCore: An Innovative Atmospheric Sampling System,
690   Journal of Atmospheric and Oceanic Technology, 27, 1839–1853, https://doi.org/10.1175/2010JTECHA1448.1, 2010.

28.   Kawa, S. R., Abshire, J. B., Baker, D. F., Browell, E. V., Crisp, D., Crowell, S. M. R., Hyon, J. J., Jacob, J. C., Jucks, K. W., Lin, B., Menzies, R. T., and Ott, L. E., and Zaccheo, T. S.: Active Sensing of CO_2 Emissions over Nights, Days, and Seasons (ASCENDS): Final Report of the ASCENDS Ad Hoc Science Definition Team, Document ID:
695   20190000855, NASA/TP–2018-219034, GSFC-E-DAA-TN64573, 2018.

29.   Kulawik, S. S., Crowell, S., Baker, D., Liu, J., McKain, K., Sweeney, C., Biraud, S. C., Wofsy, S., O'Dell, C. W., Wennberg, P. O., Wunch, D., Roehl, C. M., Deutscher, N. M., Kiel, M., Griffith, D. W. T., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Mazière, M., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D. F., Morino, I., Uchino, O., Hase, F., Feist, D. G., Roche, S., Strong, K., Kivi, R., Iraci, L., Shiomi, K., Dubey, M. K., Sepulveda, E.,
700   Rodriguez, O. E. G., Té, Y., Jeseck, P., Heikkinen, P., Dlugokencky, E. J., Gunson, M. R., Eldering, A., Crisp, D., Fisher, B., and Osterman, G. B.: Characterization of OCO-2 and ACOS-GOSAT biases and errors for CO_2 flux estimates, Atmospheric Measurement Techniques Discussions, 1–61, https://doi.org/10.5194/amt-2019-257, 2019.

30.   Lauvaux, T., Díaz-Isaac, L. I., Bocquet, M., and Bousserez, N.: Diagnosing spatial error structures in CO_2 mole fractions and XCO_2 column mole fractions from atmospheric transport, Atmospheric Chemistry and Physics, 19,
705   12007–12024, https://doi.org/10.5194/acp-19-12007-2019, 2019.

31. 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 XCO₂ sampling on the inference of terrestrial biospheric sources and sinks, Tellus B: Chemical and Physical Meteorology, 66, 22486, https://doi.org/10.3402/tellusb.v66.22486, 2014.

710 32. Liu, J., Bowman, K. W., and Henze, D. K.: Source-receptor relationships of column-average CO₂ and implications for the impact of observations on flux inversions, Journal of Geophysical Research: Atmospheres, 120, 5214–5236, https://doi.org/10.1002/2014JD022914, 2015.

33. Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., Bloom, A. A., Wunch, D., Frankenberg, C., Sun, Y., O'Dell, C. W., Gurney, K. R., Menemenlis, D., Gierach, M., Crisp, D., and Eldering, A.: Contrasting
715 carbon cycle responses of the tropical continents to the 2015–2016 El Niño, Science, 358, eaam5690, https://doi.org/10.1126/science.aam5690, 2017.

34. Liu, J., Baskaran, L., Bowman, K., Schimel, D., Bloom, A. A., Parazoo, N. C., Oda, T., Carroll, D., Menemenlis, D., Joiner, J., Commane, R., Daube, B., Gatti, L. V., McKain, K., Miller, J., Stephens, B. B., Sweeney, C., and Wofsy, S.: Carbon Monitoring System Flux Net Biosphere Exchange 2020 (CMS-Flux NBE 2020), Earth System Science
720 Data, 13, 299–330, https://doi.org/10.5194/essd-13-299-2021, 2021.

35. Machida, T., Matsueda, H., Sawa, Y., Nakagawa, Y., Hirotani, K., Kondo, N., Goto, K., Nakazawa, T., Ishikawa, K., and Ogawa, T.: Worldwide Measurements of Atmospheric CO₂ and Other Trace Gas Species Using Commercial Airlines, Journal of Atmospheric and Oceanic Technology, 25, 1744–1754, https://doi.org/10.1175/2008JTECHA1082.1, 2008.

725 36. Masarie, K. A., Pétron, G., Andrews, A., Bruhwiler, L., Conway, T. J., Jacobson, A. R., Miler, J.B., Tans, P. P., Worthy, D. E., and Peters, W.: Impact of CO₂ measurement bias on CarbonTracker surface flux estimates. Journal of Geophysical Research: Atmospheres, 116, D17, https://doi.org/10.1029/2011JD016270, 2011.

37. Nara, H., Tanimoto, H., Tohjima, Y., Mukai, H., Nojiri, Y., and Machida, T.: Emission factors of CO₂, CO and CH₄ from Sumatran peatland fires in 2013 based on shipboard measurements, Tellus B, 69,
730 1399047, https://doi.org/10.1080/16000889.2017.1399047, 2017.

38. Narbaud, C., Paris, J.-D., Wittig, S., Berchet, A., Saunois, M., Nédélec, P., Belan, B. D., Arshinov, M. Y., Belan, S. B., Davydov, D., Fofonov, A., and Kozlov, A.: Disentangling methane and carbon dioxide sources and transport across the Russian Arctic from aircraft measurements, Atmospheric Chemistry and Physics, 23, 2293–2314, https://doi.org/10.5194/acp-23-2293-2023, 2023.

735 39. 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. 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,

740   O., and Velazco, V. A.: Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version
            8 ACOS algorithm, Atmos. Meas. Tech., 11, 6539–6576, https://doi.org/10.5194/amt-11-6539-2018, 2018.

    40.   Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., Chevallier, F., Liu, J., Eldering, A., Crisp,
            D., Deng, F., Weir, B., Basu, S., Johnson, M. S., Philip, S., and Baker, I.: Four years of global carbon cycle observed
            from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7,
745   Atmospheric Chemistry and Physics, 22, 1097–1130, https://doi.org/10.5194/acp-22-1097-2022, 2022.

    41.   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_2 fluxes estimated from OCO-2 retrievals, Atmospheric
            Chemistry and Physics, 19, 13267–13287, https://doi.org/10.5194/acp-19-13267-2019, 2019.

    42.   Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., Ciais, P., Friedlingstein, P., and Sitch, S.: Interannual
750   variation of terrestrial carbon cycle: Issues and perspectives, Global Change Biology, 26, 300–318,
            https://doi.org/10.1111/gcb.14884, 2020.

    43.   Randerson, J. T., Van Der Werf, G. R., Giglio, L., Collatz, G. J., & Kasibhatla, P. S. (2015). Global Fire Emissions
            Database, Version 4, (GFEDv4), ORNL DAAC, Oak Ridge, Tennessee, USA.

    44.   Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., Chevallier, F., Crowell, S., Davis, K. J.,
755   Deng, F., Denning, S., Feng, L., Jones, D., Liu, J., and Palmer, P. I.: Quantifying the Impact of Atmospheric Transport
            Uncertainty   on CO_2 Surface   Flux   Estimates,   Global   Biogeochemical   Cycles,   33,   484–500,
            https://doi.org/10.1029/2018GB006086, 2019.

    45.   Schuldt, K. N., Mund, J., Luijkx, I. T., Aalto, T., Abshire, J. B., Aikin, K., Andrews, A., Aoki, S., Apadula, F., Baier,
            B., Bakwin, P., Bartyzel, J., Bentz, G., Bergamaschi, P., Beyersdorf, A., Biermann, T., Biraud, S. C., Boenisch, H.,
760   Bowling, D., Brailsford, G., Chen, G., Chen, H., Chmura, L., Clark, S., Climadat, S., Colomb, A., Commane, R.,
            Conil, S., Cox, A., Cristofanelli, P., Cuevas, E., Curcoll, R., Daube, B., Davis, K., Mazière, M. D., De Wekker, S.,
            Della Coletta, J., Delmotte, M., DiGangi, J. P., Dlugokencky, E., Elkins, J. W., Emmenegger, L., Fang, S., Fischer,
            M. L., Forster, G., Frumau, A., Galkowski, M., Gatti, L. V., Gehrlein, T., Gerbig, C., Gheusi, F., Gloor, E., Gomez-
            Trueba, V., Goto, D., Griffis, T., Hammer, S., Hanson, C., Haszpra, L., Hatakka, J., Heimann, M., Heliasz, M., Hensen,
765   A., Hermanssen, O., Hintsa, E., Holst, J., Jaffe, D., Joubert, W., Karion, A., Kawa, S. R., Kazan, V., Keeling, R.,
            Keronen, P., Kolari, P., Kominkova, K., Kort, E., Kozlova, E., Krummel, P., Kubistin, D., Labuschagne, C., Lam,
            D. H., Langenfelds, R., Laurent, O., Laurila, T., Lauvaux, T., Law, B., Lee, O. S., Lee, J., Lehner, I., Leppert, R.,
            Leuenberger, M., Levin, I., Levula, J., Lin, J., Lindauer, M., Loh, Z., Lopez, M., Machida, T., Mammarella, I., Manca,
            G., Manning, A., Manning, A., Marek, M. V., Martin, M. Y., Matsueda, H., McKain, K., Meijer, H., Meinhardt, F.,
770   Merchant, L., Mihalopoulos, N., Miles, N., Miller, C. E., Miller, J. B., Mitchell, L., Montzka, S., Moore, F., Morgan,
            E., Morgui, J.-A., Morimoto, S., Munger, B., Myhre, C. L., Mölder, M., Obersteiner, F., M uller-Williams, J., Necki,
            J., Newman, S., Nichol, S., Niwa, Y., O'Doherty, S., Paplawsky, B., Peischl, J., Peltola, O., Pichon, J. M., Piper, S.,
            Plass-Duelmer, C., Ramonet, M., Ramos, R., Reyes-Sanchez, E., Richardson, S., Riris, H., Rivas, P. P., Ryerson, T.,

Saito, K., Sargent, M., Sasakawa, M., Sawa, Y., Say, D., Scheeren, B., Schuck, T., Schumacher, M., Seifert, T., Sha, M. K., Shepson, P., Shook, M., Sloop, C. D., Smith, P., Steinbacher, M., Stephens, B., Sweeney, C., Tans, P., Thoning, K., Timas, H., Torn, M., Trisolino, P., Turnbull, J., Tørseth, K., Vermeulen, A., Viner, B., Vitkova, G., Walker, S., Watson, A., Wofsy, S., Worsey, J., Worthy, D., Young, D., Zahn, A., Zimnoch, M., van Dinther, D., and van den Bulk, P.: Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957–2019; obspack_co2_1_GLOBALVIEWplus_v6.1_2021-03-01, NOAA Earth System Research Laboratory, Global Monitoring Laboratory [data set], https://doi.org/10.25925/20201204, 2021.

46. Stanevich, I., Jones, D. B. A., Strong, K., Parker, R. J., Boesch, H., Wunch, D., Notholt, J., Petri, C., Warneke, T., Sussmann, R., Schneider, M., Hase, F., Kivi, R., Deutscher, N. M., Velazco, V. A., Walker, K. A., and Deng, F.: Characterizing model errors in chemical transport modeling of methane: impact of model resolution in versions v9-02 of GEOS-Chem and v35j of its adjoint model, Geoscientific Model Development, 13, 3839–3862, https://doi.org/10.5194/gmd-13-3839-2020, 2020.

47. Stephens, B. B., Long, M. C., Keeling, R. F., Kort, E. A., Sweeney, C., Apel, E. C., Atlas, E. L., Beaton, S., Bent, J. D., Blake, N. J., Bresch, J. F., Casey, J., Daube, B. C., Diao, M., Diaz, E., Dierssen, H., Donets, V., Gao, B.-C., Gierach, M., Green, R., Haag, J., Hayman, M., Hills, A. J., Hoecker-Martínez, M. S., Honomichl, S. B., Hornbrook, R. S., Jensen, J. B., Li, R.-R., McCubbin, I., McKain, K., Morgan, E. J., Nolte, S., Powers, J. G., Rainwater, B., Randolph, K., Reeves, M., Schauffler, S. M., Smith, K., Smith, M., Stith, J., Stossmeister, G., Toohey, D. W., and Watt, A. S.: The O_2/N_2 Ratio and CO_2 Airborne Southern Ocean Study, Bulletin of the American Meteorological Society, 99, 381–402, https://doi.org/10.1175/BAMS-D-16-0206.1, 2018.

48. Sweeney, C., Karion, A., Wolter, S., Newberger, T., Guenther, D., Higgs, J. A., Andrews, A. E., Lang, P. M., Neff, D., Dlugokencky, E., Miller, J. B., Montzka, S. A., Miller, B. R., Masarie, K. A., Biraud, S. C., Novelli, P. C., Crotwell, M., Crotwell, A. M., Thoning, K., and Tans, P. P.: Seasonal climatology of CO_2 across North America from aircraft measurements in the NOAA/ESRL Global Greenhouse Gas Reference Network, Journal of Geophysical Research: Atmospheres, 120, 5155–5190, https://doi.org/10.1002/2014JD022591, 2015.

49. Tohjima, Y., Mukai, H., Machida, T., Nojiri, Y., and Gloor, M.: First measurements of the latitudinal atmospheric O_2 and CO_2 distributions across the western Pacific, Geophys. Res. Lett., 32, L17805, https://doi.org/10.1029/2005GL023311, 2005.

50. Tomohiro Oda, Shamil Maksyutov (2015), ODIAC Fossil Fuel CO_2 Emissions Dataset (Version name: ODIAC2020b), Center for Global Environmental Research, National Institute for Environmental Studies, doi:10.17595/20170411.001. (Reference date: 2023/04/01)

51. Tsuboi, K., Matsueda, H., Sawa, Y., Niwa, Y., Nakamura, M., Kuboike, D., Saito, K., Ohmori, H., Iwatsubo, S., Nishi, H., Hanamiya, Y., Tsuji, K., and Baba, Y.: Evaluation of a new JMA aircraft flask sampling system and laboratory trace gas analysis system, Atmospheric Measurement Techniques, 6, 1257–1270, https://doi.org/10.5194/amt-6-1257-2013, 2013.

52. Weir, B., Crisp, D., O'Dell, C. W., Basu, S., Chatterjee, A., Kolassa, J., Oda, T., Pawson, S., Poulter, B., Zhang, Z., Ciais, P., Davis, S. J., Liu, Z., and Ott, L. E.: Regional impacts of COVID-19 on carbon dioxide detected worldwide from space, Science Advances, 7, eabf9415, https://doi.org/10.1126/sciadv.abf9415, 2021.

53. Zhang, X., Gurney, K. R., Rayner, P., Baker, D., and Liu, Y.-P.: Sensitivity of simulated CO_2 concentration to sub-annual variations in fossil fuel CO_2 emissions, Atmos. Chem. Phys., 16, 1907–1918, https://doi.org/10.5194/acp-16-1907-2016, 2016.