Quantification of regional net CO_2 flux errors in the v10 OCO-2 MIP ensemble 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], Sebastien C. Biraud[5]

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

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. However, accurately quantifying ensemble CO_2 flux errors remains challenging, often relying on the ensemble spread. This study proposes a method to quantify the errors of regional net surface-atmosphere CO_2 flux estimates from the v10 Orbiting Carbon Observatory-2 (OCO-2) MIP models 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 concentrations and airborne observations and then isolate the CO_2 concentration errors caused solely by the ensemble mean of posterior net fluxes by subtracting the observation, representation, and transport errors in seven regions. Our analysis reveals that the flux errors projected into CO_2 space account for 55-85% of the regional average RMSE over the three years, ranging from 0.88 to 1.91 ppm. In five regions, the error estimates based on observations exceed those computed from the ensemble spread of posterior fluxes by 1.33-1.93 times, implying an underestimation of the actual flux errors, while their magnitudes are comparable in two regions. The adjoint sensitivity analysis identifies the underestimation of flux errors is prominent where the magnitudes of fossil fuel emissions exceed those of terrestrial biosphere fluxes by 3-31 times over the three years. This suggests the presence of systematic biases in the inversion estimates associated with errors in the prescribed fossil fuel emissions common to all models. Our study emphasizes the value of airborne measurements for quantifying regional errors in ensemble net CO_2 flux estimates.

1 Introduction

Atmospheric CO_2 inverse modeling is a widely employed approaches to estimate net surface-atmosphere CO_2 fluxes by assimilating observed atmospheric CO_2 concentrations. Most inverse modeling approaches are based on the Bayesian theory, wherein posterior flux is 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 $\hat{x}$) and an error covariance matrix, following the notation of Rodgers (2000). Theoretically, since atmospheric CO_2 observations generally have lower uncertainty than prior flux estimates, more observations lead to posterior fluxes approaching true values (Liu et al., 2014).

However, concerns have been raised that the inverse modeling results are sensitive to the selection of transport models, prior flux datasets, and data assimilation techniques 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 flux estimates and assess their uncertainties resulting from 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 1990s, as well as subsequent projects such as the Global Carbon Project (GCP; Friedlingstein et al., 2023; 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 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. The OCO-2 MIP has 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).

Realistic error quantification of posterior fluxes from atmospheric flux inversions is essential for understanding 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 uncertainties in prior fluxes and assimilated data. This can be calculated 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 explicitly incorporate systematic errors, thus providing a potential underestimate of the total posterior error.

Errors in the maximum a posteriori 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 include not only posterior flux errors, but also errors arising from 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 does not have an observational and theoretical basis and may not reflect actual errors (Byrne et al., 2023).

This study aims to develop a framework to quantify the errors in regional net surface-atmosphere CO_2 fluxes (terrestrial biosphere fluxes + fossil fuel emissions) 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 the 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 al., 2021a; 2021b). This study uses more than 833,000 airborne CO_2 measurements collected at 1-5 km altitude above ground level (AGL) from 20 different measurement projects (e.g., Baier et al., 2021; Miller et al., 2021; NOAA Carbon Cycle Group ObsPack Team, 2018; Schuldt et al., 2021a; 2021b). These data have broader spatial coverage and are less influenced by local sources compared to surface CO_2 data, thus capturing signals from regional surface CO_2 fluxes. We quantify the errors in ensemble mean estimates of posterior atmospheric CO_2 by comparing them with the airborne CO_2 data. We then estimate the contributions of various error components (e.g., representation, observation, transport, and flux errors) to the observation-model difference in atmospheric CO_2 and isolate the contribution of flux errors. Next, we identify the areas to which these airborne CO_2 are most sensitive to and quantify the annual net flux errors in these areas.

2 Data and methodology

The aim of this study is to quantify the true errors of the ensemble net surface-atmosphere CO_2 fluxes generated by the v10 OCO-2 MIP using airborne observations. Here, "error" refers to the magnitude of the differences between the true and estimated flux values, without considering the sign. To achieve this, we employ three steps of analysis as described in Figure 1. First, we define two quantities: 1) the root mean square errors (*RMSE*) between the ensemble mean of posterior CO_2 concentrations and observed CO_2 concentrations, and 2) ERR_{TOT} (Section 2.3). $RMSE^2$ represents the true errors in OCO-2 MIP ensemble mean of CO_2 concentrations including representation errors (σ_r^2), observation errors (σ_o^2), true flux errors projected onto CO_2 concentration ($\sigma_{f_t}^2$), transport errors (σ_t^2), and error covariances between the preceding two terms ($cov(\sigma_{f_t}, \sigma_t)$). ERR_{TOT}^2 is the sum of the estimated error components, defined as the sum of ERR_{REP}^2, ERR_{OBS}^2 and ERR_{MIP}^2.

ERR_{REP}^2 and ERR_{OBS}^2 indicate representation errors (σ_r^2) and observation errors (σ_o^2), respectively. ERR_{MIP}^2 is the sum of estimated flux errors projected onto CO_2 space (σ_{fe}^2) and transport errors (σ_t^2), and their error covariances ($cov(\sigma_{\mathrm{fe}}, \sigma_{\mathrm{t}})$), computed from an ensemble spread of posterior CO_2 concentrations. Here we separate representation errors from transport errors for computational purpose. The ratio between ERR_{TOT} and $RMSE$ is then used to evaluate whether the estimated flux errors, computed from the ensemble spread of posterior fluxes, overestimate or underestimate the true errors in the ensemble mean fluxes. Next, we calculate the estimated flux errors projected onto atmospheric CO_2 ($h(err_{fe})$) through atmospheric transport simulations (Section 2.4). With $h(err_{fe})$, ERR_{TOT}, and $RMSE$, we derive the true errors in ensemble mean of posterior fluxes projected onto CO_2 space ($h(err_{ft})$). Then, we identify the areas where these airborne observations are most sensitive to using an adjoint sensitivity analysis and calculate the estimated posterior flux errors over these regions (err_{fe}). Assuming a linear observation operator, the study finally computes the true errors of the ensemble mean posterior fluxes over the identified sensitive areas (err_{ft}) by applying the ratio between $h(err_{ft})$ and $h(err_{fe})$ to err_{fe}.

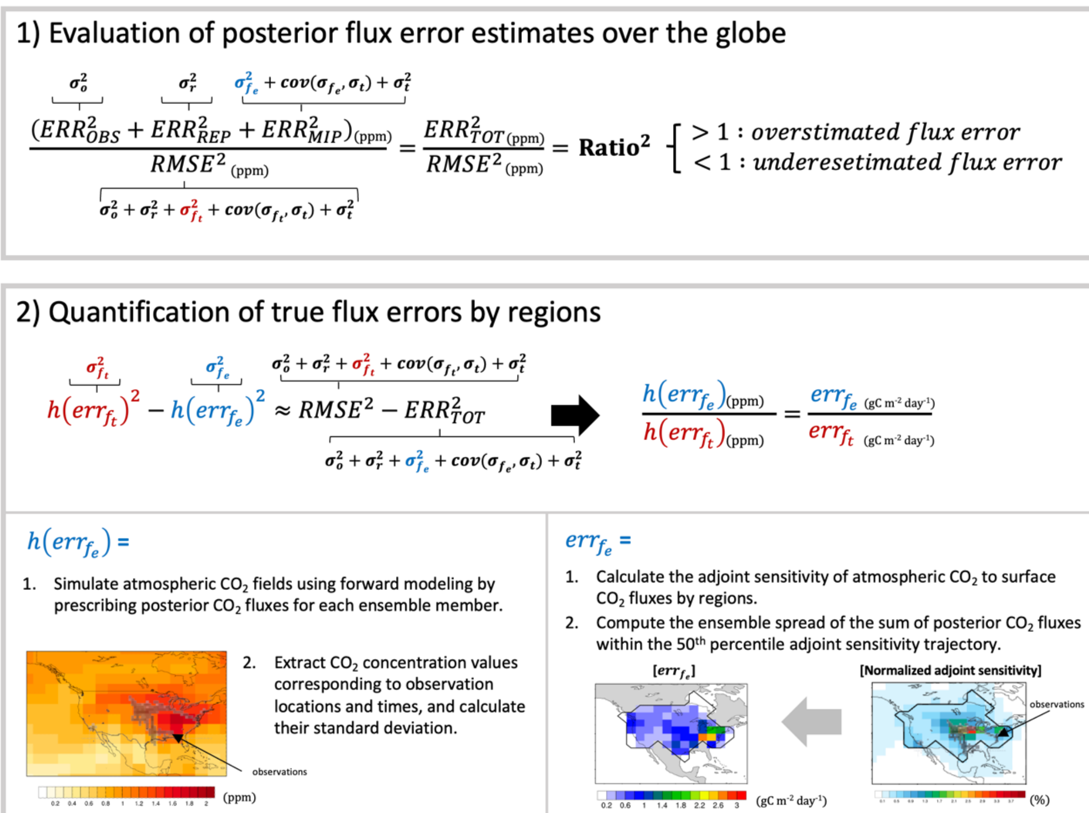

Figure 1: Flow chart summarizing the process of evaluating and quantifying errors in ensemble mean of regional posterior fluxes. RMSE² is the mean square errors between the ensemble mean of posterior CO_2 concentrations and observed CO_2 concentrations. ERR_{REP}^2 and ERR_{OBS}^2 denote estimates of observation errors and representation errors, respectively. ERR_{MIP}^2 is an ensemble spread of posterior CO_2 concentrations. ERR_{TOT}^2 is defined as the sum of ERR_{REP}^2, ERR_{OBS}^2, and ERR_{MIP}^2. err_{f_e} and err_{f_t} are estimates of flux errors, defined as an ensemble spread of posterior fluxes, and their true values. $h(err_{f_e})$ and $h(err_{f_t})$ are estimates of flux errors projected onto CO_2 concentrations and their true values. σ_o^2, σ_r^2, $\sigma_{f_t}^2$ ($\sigma_{f_e}^2$), σ_t^2, and $cov(\sigma_{\mathrm{f_t}}, \sigma_{\mathrm{t}})$ indicate the types of errors

represented by the error statics, namely observation errors, representation errors, true (estimated) flux errors projected onto CO_2 concentration, transport errors, and error covariances between the preceding two terms, respectively.

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 surface CO_2 flux estimates interpolated to 1°×1° 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 estimates based on Open-source Data Inventory for Anthropogenic CO_2 (ODIAC) dataset (Basu & Nassar, 2021), 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 use 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 exclude surface measurements and only consider airborne measurements made between 1 and 5 km AGL. In addition, only airborne measurement data that were not assimilated in the LNLGIS experiment are used for analysis.

2.2 Airborne CO_2 measurement data

Figure 2a shows the spatial distribution of the total number of airborne CO_2 measurements used in this study within each 1°x1° grid cell. The dataset includes two airborne measurement campaigns over the ocean (Atmospheric Tomography Mission; ATom; Thompson et al. 2022 and O_2/N_2 Ratio and CO_2 Airborne Southern Ocean Study; ORCAS; Stephens et al. 2018), as well as 18 campaigns over land. Specific airborne campaigns and their references are elaborated in Table 1. The majority of the datasets used in the study are from North America, accounting for 37% of the total number of observations for the period of 2015-2017, followed by East Asia with 35% and Alaska with 7%. The duration and extent of the airborne observations vary across different regions and time periods. Figure 2b illustrates the number of 1°×1° grid points in each of the seven regions where more than 10 observations are available per month. For Alaska, observations were concentrated during the Arctic-Boreal Vulnerability Experiment (ABoVE) campaign in 2017 (Sweeney et al. 2022). North America had observations for most of the analysis period, including observations from the Atmospheric Carbon and Transport – America

(ACT–America) campaign covering the eastern United States (Davis et al., 2021). The Long-term Comprehensive Observation Network for TRace gases by AIrLiner (CONTRAIL; Machida et al., 2008) project provides sparse observation in Europe and continuous observation in East and Southeast Asia from 2015 to 2017, as well as for Australia during 2015–2016. In South America, measurements were conducted at six different sites in 2017: the majority of these observations come from five flask measurement sites provided by the National Institute for Space Research (INPE), which likely have a low bias in measured flask sample CO_2 mole fractions of ~1 ppm or greater when ambient water vapor mole fractions are above ~1.5%. These biases in some aircraft flask CO_2 measurements have been noted in previous literature (Baier et al., 2020; Gatti et al., 2023) and impacted data have been removed from all other aircraft flask datasets. Despite the potential limitation of these South American observations, our analysis, aimed at introducing a method for quantifying flux errors, incorporates these data to offer guidance for future studies leveraging bias-corrected observations from this region. As discussed in more detail below, readers should keep in mind that our results from South America may have lower reliability compared to those from other regions.

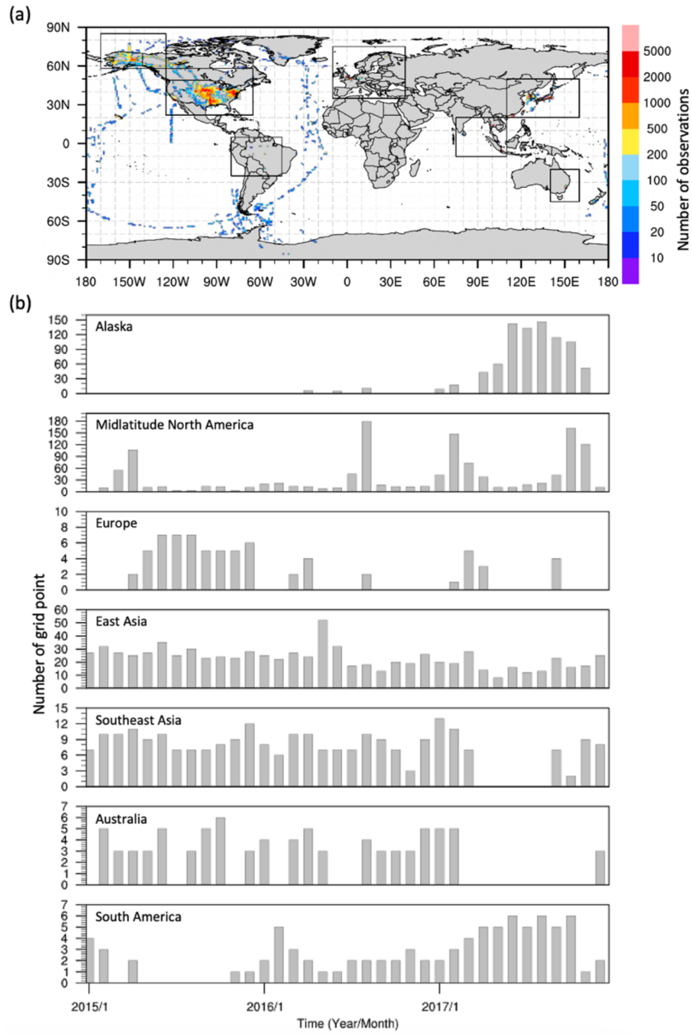

Figure 2: (a) Total number of airborne measurement data used in this study at each 1°×1° grid point and (b) the number of 1°×1° grid-points, where more than 10 data is available, within each region and each month for the period 2015–2017.

Table 1. Data description for each airborne measurement campaign.

Site code	Site name	Measurement campaign name	Measurement type	Data provider	ObsPack (*original*) dataset identifier	Reference
ACG	Alaska Coast Guard, Alaska, USA	NOAA/GML Aircraft Program	In situ	National Oceanic and Atmospheric Administration (NOAA) Global Monitoring Laboratory (GML)	http://doi.org/10.25925/20201204[a]	Karion et al. (2013)
ACT	Atmospheric Carbon and Transport – America (ACT-America), USA	ACT-America	In situ and flask	National Aeronautics and Space Administration Langley Research Center (NASA-LaRC), NOAA/GML	http://doi.org/10.25925/20201204[a] *https://doi.org/10.3334/ORNLDAAC/1593*	Baier et al. (2020) DiGangi et al. (2021) Wei et al. (2021)

AirCore NOAA	NOAA AirCore Program	NOAA AirCore Program	Balloon air sampler	NOAA/GML	No Obspack DOI[b] *https://doi.org/10.15138/6AV0-MY81*	Karion et al. (2010)
ALF	Alta Floresta, Brazil		Flask	National Institute for Space Research (INPE)	http://dx.doi.org/10.25925/20181030[c] https://doi.org/10.1594/PANGAEA.926834	Gatti et al. (2023)
CAR	Briggsdale, Colorado		Flask	NOAA/GML	http://doi.org/10.25925/20210517[d]	Sweeney et al. (2015)
CON	Comprehensive Observation Network for TRace gases by AIrLiner (CONTRAIL)		In situ	National Institute for Environmental Studies (NIES), Meteorological Research Institute (MRI)	http://doi.org/10.25925/20201204[a] *https://doi.org/10.17595/20180208.001*	Machida et al. (2008)
CRV	Carbon in Arctic Reservoirs Vulnerability Experiment (CARVE), Alaska	Arctic-Boreal Vulnerability Experiment (ABoVE)	In situ	NOAA/GML	http://doi.org/10.25925/20201204[a] *https://doi.org/10.3334/ORNLDAAC/1582*	Sweeney et al. (2022)
GSFC	Active Sensing of CO_2 Emissions over Nights, Days and Seasons (ASCENDS), USA	ASCENDS	In situ	NASA Goddard Space Flight Center (NASA-GSFC)	http://doi.org/10.25925/20201204[a]	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	Karlsruhe Institute of Technology (IMK-ASF), Institute for Atmospheric and Environmental Sciences (IAU), Max Planck Institute for Biogeochemistry (MPI-BGC)	http://doi.org/10.25925/20201204[a]	Filges et al. (2015)
KORUS	The Korea-United States Air Quality (KORUS-AQ) field study		In situ	NASA-LaRC	http://doi.org/10.25925/20201204[a] *https://doi.org/10.5067/ASDC/SUBORBITAL/KORUSAQ_TraceGas_AircraftInSitu_DC8_Data_1*	Vay et al., (2009)
MAN	Manaus, Brazil	NOAA/GML Aircraft Program	In situ	NOAA/GML	https://doi.org/10.25925/20210519[e]	
ORC	O2/N2 Ratio and CO2 Airborne Southern Ocean Study (ORCAS)		In situ	National Center for Atmospheric Research (NCAR)	http://doi.org/10.25925/20201204[a] *https://doi.org/10.5065/D6SB445X*	Stephens et al. (2018)
PAN	Pantanal, Mato Grosso do Sul, Brazil		Flask	INPE	http://dx.doi.org/10.25925/20181030[c]	
PFA	Poker Flat, Alaska	NOAA/GML Aircraft Program	Flask	NOAA/GML	http://doi.org/10.25925/20210517[d]	Sweeney et al. (2015)
RBA-B	Rio Branco, Brazil		Flask	INPE	http://dx.doi.org/10.25925/20181030[c] *https://doi.org/10.1594/PANGAEA.926834*	Gatti et al. (2023)
SAN	Santarém, Brazil		Flask	INPE	http://dx.doi.org/10.25925/20181030[c] *https://doi.org/10.1594/PANGAEA.926834*	Gatti et al. (2023)
SGP	Southern Great Plains, Oklahoma, USA	NOAA/GML Aircraft Program	Flask	The US Department of Energy (DOE)/Lawrence Berkeley National Laboratory (LBNL), NOAA/GML	http://doi.org/10.25925/20210517[d]	Biraud et al. (2013) Sweeney et al. (2015)
SONGNEX2015	Shale Oil and Natural Gas Nexus 2015 (air campaign), USA	Shale Oil and Natural Gas Nexus 2015 (air campaign)	In situ	NOAA Chemical Sciences Laboratory (CSL)	http://doi.org/10.25925/20201204[a]	
TEF	Tefé, Brazil		Flask	INPE	http://dx.doi.org/10.25925/20181030[c] *https://doi.org/10.1594/PANGAEA.926834*	Gatti et al. (2023)

TOM	Atmospheric Tomography Mission (ATom)	Atmospheric Tomography Mission (ATom)	In situ	NOAA/GML, Harvard University	http://doi.org/10.25925/20201204[a] *https://doi.org/10.3334/ORNLDAAC/1581*	Thompson et al. (2022)

[a]: obspack_co2_1_GLOBALVIEWplus_v6.1_2021-03-01 (Schuldt et al., 2021b)
[b]: obspack_co2_1_AirCore_v4.0_2020-12-28
[c]: obspack_co2_1_INPE_RESTRICTED_v2.0_2018-11-13 (NOAA Carbon Cycle Group ObsPack Team, 2018)
[d]: obspack_co2_1_NRT_v6.1.1_2021-05-17 (Schuldt et al., 2021a)
[e]: obspack_multi-species_1_manaus_profiles_v1.0_2021-05-20 (Miller et al., 2021)

2.3 Evaluation of ensemble posterior CO_2 fluxes

We first employ the two matrixes defined in Eq. (1) and (2) below to evaluate ensemble posterior net flux errors proposed by Liu et al. (2021). One is RMSE between the ensemble mean of posterior atmospheric CO_2 from OCO-2 MIP models and the atmospheric CO_2 from airborne measurements, which can be written as:

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

$\overline{h_\iota(\hat{x})}$ is the ensemble mean of posterior atmospheric CO_2 sampled at the time and location of the i^{th} airborne observation $y_{o,i}$, within each 1°×1° grid-cell in each month. *N* is the monthly total number of sampled data at each grid-cell. *M* is the number of ensemble members (i.e., 10). A single monthly *RMSE* value is computed using *N* measurement data at each grid-cell. The number of *RMSE* values is calculated per month within each region corresponds to the number of grid-cells shown in Figure 2b. The *RMSE* indicates the magnitude of the actual CO_2 errors in the ensemble estimates, which is also a quantity broadly used to evaluate the accuracy of posterior fluxes (Crowell et al., 2019; Peiro et al., 2022; Byrne et al., 2023). As illustrated in Figure 1 and as described in Appendix A (Eq. A3), $RMSE^2$ includes not only the projection of true flux errors on CO_2 concentration (σ_{ft}^2), but also transport errors (σ_t^2), their error covariances ($cov(\sigma_{ft},\sigma_t)$), representation errors (σ_r^2), and airborne observation errors (σ_o^2). Both transport errors and representation errors stem from transport models. Transport errors include the errors in model structures and meteorological fields, while representation errors arise from a mismatch in resolution between model simulations and observations.

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 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 errors (σ_{fe}^2) and transport errors (σ_t^2) and covariance terms between them ($cov(\sigma_{fe},\sigma_t)$) are replaced by the spread of ensemble (i.e., variance) posterior atmospheric CO_2 concentrations (ERR_{MIP}^2) defined as:

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

Different from Liu et al. (2021) which used only one transport model, ERR^2_{MIP} accounts transport errors because posterior atmospheric CO_2 were generated by multiple types of transport models in OCO-2 MIP driven by different meteorology fields. Thus, ERR^2_{MIP} term accounts for transport errors, but not representation errors due to the coarse spatial resolution of these transport models with the highest spatial resolution being 2°×2.5°.

To obtain representation errors and observation errors not captured by ERR^2_{MIP}, we additionally calculate ERR^2_{REP} and ERR^2_{obs}, respectively. ERR^2_{REP} indicates the representation errors (σ_r^2) in *RMSE*^{2} as shown in Figure 1 and is defined as a spatial variability of atmospheric CO_2 within a 2°×2.5° grid cell written as:

$$ERR^2_{REP} = \frac{1}{N}\sum_{i=1}^{N} VAR_{CO_2,i} \quad (3)$$

With the high-resolution (0.5°x0.625°) 3-hourly GEOS-5 simulation results for 2018 from NASA Goddard Space Flight Center (Weir et al., 2021), we calculate the variance of atmospheric CO_2 concentration within each 2°×2.5° grid cell at every 3-hour interval. Then, we sample the CO_2 variance value ($VAR_{CO_2,i}$) at the grid cell containing the ith observation and the time closest to the observation. Subsequently, the monthly mean values of the *N* co-sampled variances are derived (ERR^2_{REP}). We assume that the variances do not vary significantly across years, given relatively lower monthly variability of ERR_{REP} compared to that of RMSE and ERR_{MIP} (to be shown in Section 3.2). The reason for calculating CO_2 variance value within 2°×2.5° is because it is the finest resolution among the OCO-2 MIP models. We evaluate whether the representation errors, derived from simulated atmospheric CO_2 fields, represent the actual spatial variability of CO_2 concentration by comparing simulated CO_2 variance with the spatial variance of airborne measurement data from ACT-America project (Supplement Text and Fig. S1). The evaluation results support our approach.

ERR^2_{OBS} represents the observation errors (σ_o^2) in *RMSE*^{2} as shown in Figure 1. 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. 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 observation error (ERR_{OBS}) for all airborne observations at 0.1 ppm. However, in reality, systematic errors could be present in airborne observation stemming from instrument or setup biases, calibration offsets, and other factors. Especially, CO_2 measurements in South America from INPE might exhibit a higher measurement error compared to other regions because of unresolved water vapor contamination issues in those flask measurements, which could result in both a low bias (~1-3 ppm at 3% absolute humidity, respectively) and spurious variability (Baier et al., 2020). The potential effects of these systematic errors on our findings will be addressed in Section 4. This study only employs ERR^2_{OBS} for calculating ERR^2_{TOT} and does not compare it with other error quantities in Section 3.

Therefore, ERR_{TOT}, the approximation for $RMSE$, is defined as:

$$ERR_{TOT}^2 = ERR_{OBS}^2 + ERR_{REP}^2 + ERR_{MIP}^2 \tag{4}$$

By applying 1000 bootstrap resampling to the monthly grid-based error statistics (e.g., $RMSE$, ERR_{MIP}, ERR_{REP}, and ERR_{TOT}) within each region, we obtain regional mean values of these error statistics, along with their corresponding 95% confidence intervals.

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

$$Ratio^2 = \frac{ERR_{TOT}^2}{RMSE^2} \tag{5}$$

Given that ERR_{REP}^2 reasonably depict actual representation errors, $Ratio^2$ can indicate whether posterior flux and transport errors computed from the ensemble spread is an overestimation or underestimation of true flux and transport errors. In this study, we assume that the estimated transport errors from the ensemble spread among transport models used in OCO-2 MIP represent the true transport errors and the difference between $RMSE^2$ and ERR_{TOT}^2 mainly arises from the difference in the flux error variances ($\sigma_{f_t}^2$ and $\sigma_{f_e}^2$). Thus, 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 errors in the ensemble mean fluxes. A ratio greater than 1 means that the posterior flux errors are overestimated, and vice versa. However, our assumption regarding transport errors may be a strong assumption given that the transport errors are derived from 10 ensemble members, covering four different transport models, which might not fully capture the actual transport errors. We discuss how this assumption affects our key results in Section 4.

2.4 Quantification of the uncertainties of ensemble mean of posterior CO_2 fluxes

In addition to the qualitative evaluations of posterior flux errors using the ratios between ERR_{TOT} and RMSE, 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:

$$h\left(err_{f_e}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{k=1}^{M}\frac{1}{M}\sum_{j=1}^{M}\left[h_k\left(\hat{x}_{k,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right]\left[h_k\left(\hat{x}_{k,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right]^T \tag{6}$$

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 approximate this error term using the GEOS-Chem model as depicted:

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

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

To get $h_{GC}\left(err_{fe}\right)^2$, we conduct a set of forward simulations using the GEOS-Chem transport model (within the GEOS-Chem Adjoint model v8.2j; Henze et al., 2007). In all ten experiments, consistent meteorology and emission forcing data are used from the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2; Gelaro et al., 2017) and Open-source Data Inventory for Anthropogenic CO_2 (ODIAC; Oda and Maksyutov, 2015); identical monthly balanced hourly terrestrial biosphere fluxes from SiB4 (Haynes et al., 2021) are also employed. However, in each experiment, the prescribed monthly fluxes of terrestrial ecosystems and oceans are based on the posterior fluxes from the respective ten OCO-2 MIP ensemble members. All experiments are performed at 2°×2.5° 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°×1° grid-cell, we derive $h_{GC}\left(err_{fe}\right)^2$.

Because we assume that the spread of ensemble transport models used in OCO-2 MIP represents the true transport errors included in $RMSE^2$, the transport errors along with observation errors and representation errors would cancel out when we calculate the difference between monthly $RMSE^2$ and ERR_{TOT}^2. Consequently, the difference between monthly $RMSE^2$ and ERR_{TOT}^2 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:

$$h\left(err_{f_t}\right)^2 - h\left(err_{f_e}\right)^2 = RMSE^2 - ERR_{TOT}^2 \qquad (8)$$

From Eq. (8), we can derive the true errors of the ensemble mean fluxes in CO_2 space, $h\left(err_{f_t}\right)^2$. Out of 181 cases, representing the total months of observations across all seven regions, $h\left(err_{f_t}\right)$ can be derived using this equation in 158 cases. In 23 cases (13% of total cases), $h\left(err_{f_t}\right)$ cannot be derived when ERR_{TOT} and/or $h\left(err_{f_e}\right)$ values fell outside the applicable range. Around 40% of the exception cases occur in South America where observation cover only one to six 1°×1° grid cells per month, suggesting that observations are insufficient to quantify the monthly flux errors in this region.

In order to link those terms with flux errors in flux space, we first identify the areas sensitive to airborne CO_2 measurements by conducting sensitivity experiments using the GEOS-Chem Adjoint model. Seven sets of adjoint sensitivity experiments are conducted to examine the sensitivity of airborne measurements in each region (defined in Figure 2a) to surface CO_2 fluxes for the month of observations. The sensitivity experiments use the same meteorology and CO_2 emission datasets as the forward simulations, along with the ensemble mean of posterior terrestrial biosphere and ocean 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:

$$J = \sum_{i=1}^{N} h_i(\hat{x}) \qquad (9)$$

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:

$$\gamma_{l,t} = \frac{\partial J}{\partial \hat{x}_{l,t}} \tag{10}$$

Monthly cumulative sensitivity (β) with respect to surface fluxes is determined by integrating $\gamma_{l,t}$ from the measurement time (t_0) to the initial time (t_{-T}) for each month:

$$\beta_l = \sum_{t=t_0}^{t-T} \gamma_{l,\,t} \tag{11}$$

In order to find the most sensitive areas to the airborne observations, we select the areas accounting for 50% of the global total values of β for each region and month. Areas with sensitivity values lower than 0.1% (0.15% for Alaska, Australia, and Southeast Asia) of the total value of β are excluded due to occasional cases where observations are 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% are included in the effective area. We then compute the estimated posterior flux errors in flux space ($err_{fe}^2 = \sigma_{fe}^2$) by calculating the ensemble spread of the total posterior flux values (and area-averaged mean values) over the effective area for each month for the period 2015–2017, as illustrated in Figure 1. The estimated mean posterior flux errors (err_{fe}) over the selected areas in each month exhibits a significant correlation ($p \leq 0.05$) with the monthly $h\left(err_{fe}\right)$ in all regions, except for Australia where the observational campaign was conducted in specific months (Fig. S2). While the observed atmospheric CO_2 concentration is influenced by both land and ocean sources, a comparison of the magnitudes of err_{fe} between ocean and land within the effective areas reveals that, on average, the land flux errors contribute more than 95% to the total flux errors in all regions (Fig. S3). This result indicates that our evaluation results based on atmospheric CO_2 can be applied to deriving the actual errors of posterior net land CO_2 fluxes within the selected area in flux space.

This study provides both monthly and three-year mean values of regional flux error statistics for the period 2015–2017. Technically, it is possible to derive the monthly true errors in the ensemble mean of net land CO_2 fluxes using the monthly error statistics. However, to obtain more robust results, we compute the true errors of annual total fluxes over the analysis period. To identify the areas contributing most to the computed mean error statistics, we calculate the number of months selected as the effective areas for monthly airborne observations. Those grid cells, at 2°×2.5° resolution, corresponding to the effective areas are assigned a value of 1, while the remaining cells are assigned a value of 0 for each month. We then calculate composite values for each grid cell over the three years. A higher number of months indicates more information in those grid cells was utilized in calculating the three-year regional mean error statistics. We define that our three-year mean error statistics mostly represent the areas where the composite values exceed eight, corresponding to 20% of the total analysis months (i.e., 36).

The observation operator, which converts surface CO_2 fluxes to atmospheric CO_2, is generally assumed linear. Therefore, we can obtain the true errors in the ensemble annual total net land fluxes in those areas, err_{f_t} ($= \sigma_{f_t}$), by multiplying the ratio between three-year mean values of $h(err_{f_t})$ and $h(err_{f_e})$ by the ensemble spread of the annual total net land flux estimates (err_{f_e}) within the effective areas. The equation can be written as:

$$err_{f_t} = \frac{h(err_{f_t})}{h(err_{f_e})} \times err_{f_e} \quad (12)$$

One thing readers should keep in mind is that the err_{f_e} is identical to the ensemble spread of posterior terrestrial biosphere fluxes because all OCO-2 MIP models used uniform fossil fuel emission estimates and assumed them to be perfectly known. Lastly, to explore characteristics of regions where average annual total err_{f_t} is significantly underestimated, we compute the ensemble mean of average annual posterior terrestrial biosphere CO_2 fluxes and fossil fuel CO_2 emissions (from ODIAC) in the effective area.

3 Results

3.1 Spatiotemporal variations of the ensemble posterior CO_2 concentration errors and other major error components

Because the magnitude of land-atmosphere CO_2 fluxes is generally over 10 times greater than ocean-atmosphere CO_2 fluxes, the observed 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 AGL and then calculate the mean error statistics for the entire campaign period. Comparisons to observations from ATom and ORCAS campaigns reveal a 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 1.5 ppm over North America (Fig. S4), suggesting significant contributions of errors in land fluxes to the differences between observed and simulated atmospheric CO_2. Additionally, consistently elevated *RMSE* values (>1.5 ppm) commonly appear over the west coast of Africa throughout the seasons.

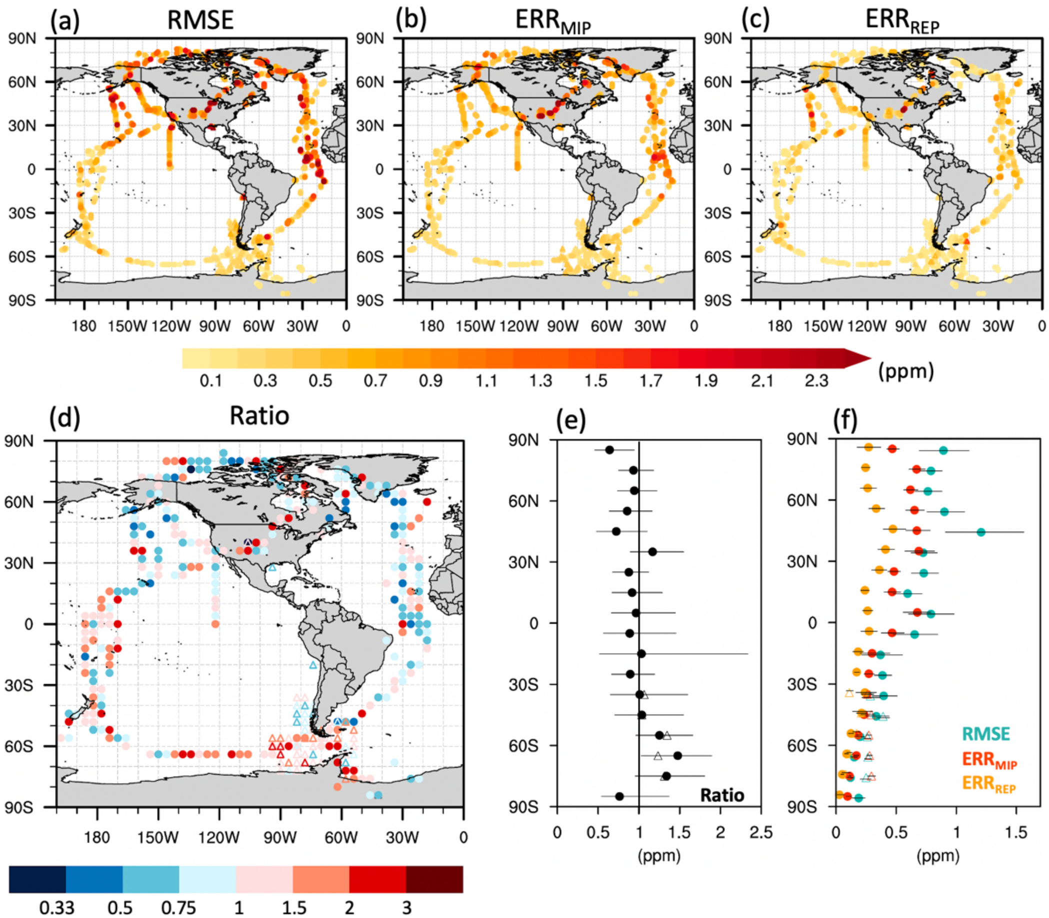

Figure 3: Spatial distributions of (a) *RMSE*, (b) ERR_{MIP}, (c) ERR_{REP}, and (d) *Ratio* (= $\sqrt{(ERR_{OBS}\,(=0.1\,ppm))^2 + ERR_{REP}^2 + ERR_{MIP}^2}\ /\ RMSE$) where ATom (circle) and ORCAS (triangle) airborne measurements were taken and (e and f) their latitudinal distributions smoothed by 10° moving average with 95% confidence intervals derived from 1000 bootstrap samples of datasets (error bar).

Both ERR_{MIP} and ERR_{REP} exhibit similar spatial distributions as *RMSE* (Figure 3a-c, f). However, ERR_{MIP} has a stronger positive correlation with *RMSE* (r = 0.57 and 0.58 for ATom and ORCAS, respectively) compared to ERR_{REP} (r = 0.35 and 0.32), with an average greater magnitude (0.49 and 0.32 ppm) than ERR_{REP} (0.27 and 0.20 ppm) globally for the whole campaign periods. Particularly, ERR_{MIP} and ERR_{REP} account for 75% and 37% of the anomalous high *RMSE* values (1.5 ppm) in Northern America (32-50N and 85-124W), and 75% and 30% of the *RMSE* values (1.2 ppm) along the west 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 compare *RMSE* with the sum of ERR_{MIP}, ERR_{REP}, 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. S5). Airborne CO_2 measurements in this area are predominantly influenced by ocean fluxes due to the limited land extent and the significant distance from land (Yun et al., 2022), suggesting the true posterior ocean 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 biosphere fluxes in tropical Africa (Liu et al., 2021), our results imply that true errors of the ensemble mean terrestrial biosphere fluxes in this region may be larger than the estimated errors based on the OCO-2 MIP ensemble spread. These findings agree with Gaubert et al. (2023), which shows that most inverse models assimilating OCO-2 XCO_2 retrievals tend to overestimate the net carbon sources in this region because of potential positive biases in OCO-2 retrievals.

In the northern mid-to-high latitudes, characterized by significant land 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 flux errors based on ensemble spread can differ depending on regions and seasons, emphasizing the need for a more detailed evaluation of flux errors at a regional level based on long-term independent observation.

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

In this section, we calculate 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 values falling within the 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=0.46), mid-latitude North America (r=0.63), Europe (r=0.60) and East Asia (r=0.60). Furthermore, the correlation coefficient is greater than or comparable to that with ERR_{REP}. 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* does not exhibit a significant correlation with either ERR_{MIP} or ERR_{REP} in

Southeast Asia, Australia, and South 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.

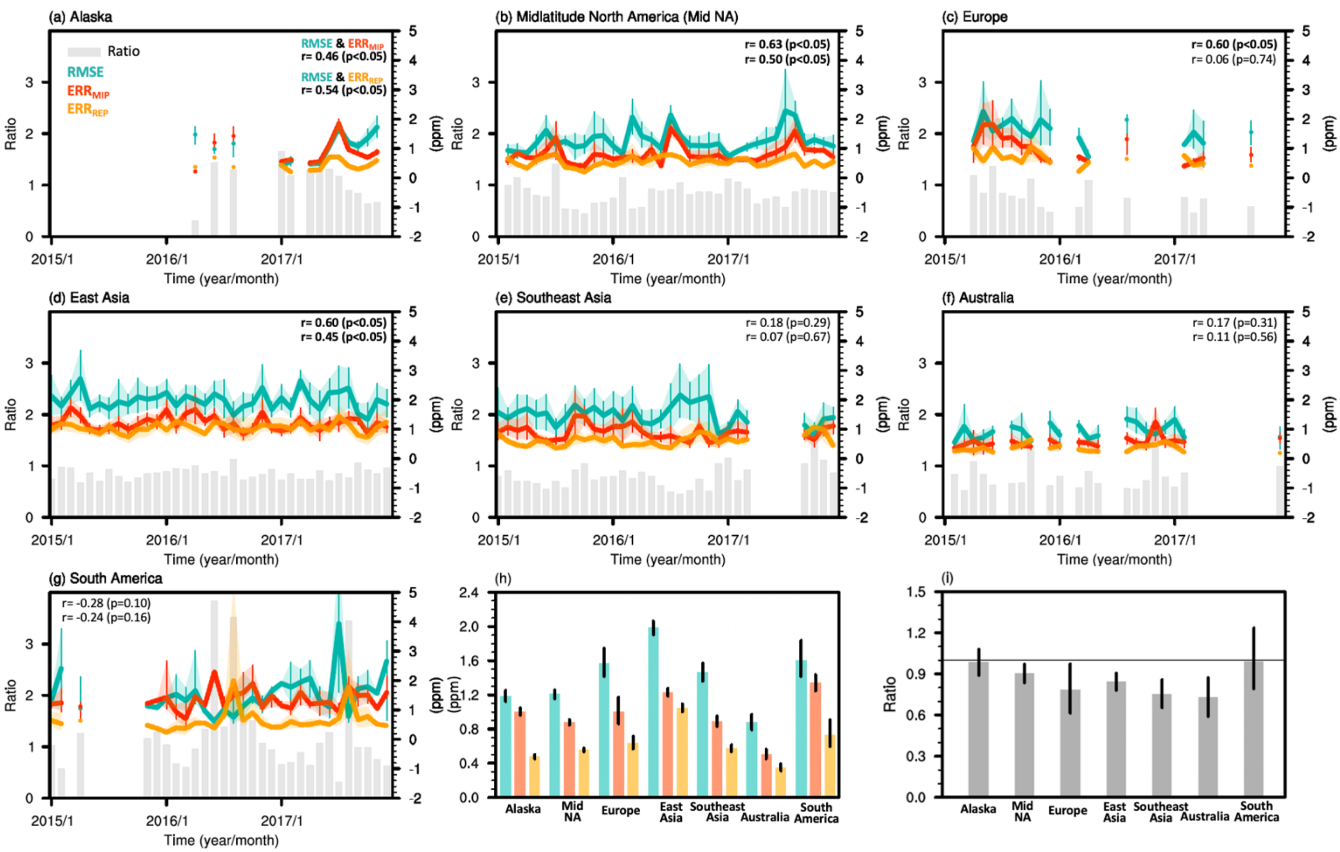

Figure 4: (a-g) Monthly values of *RMSE, ERR_{MIP}, ERR_{REP}*, and *Ratio* for each region and (h, i) their mean values for the period 2015–2017. The upper right number in (a-g) indicates the correlation coefficients between *RMSE* and *ERR_{MIP}* and *ERR_{REP}*. The shaded areas and error bars represent 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 three-year average *RMSE* is the largest in East Asia (1.98 [1.90, 2.06] ppm: mean [95% confidence intervals]), followed by Europe (1.57 [1.41, 1.74] ppm) and the lowest in Australia (0.88 [0.79, 0.97] ppm), followed by Alaska (1.19 [1.12, 1.25] ppm). ERR_{MIP} is the primary error component for *RMSE*, accounting for 58-83% of the *RMSE*, surpassing the ERR_{REP} in all the regions by 1.2-2.1 time. In East Asia, the difference between ERR_{MIP} and ERR_{REP} 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.

The ratio between ERR_{TOT} and $RMSE$ also show significant variability across regions. Our results indicate that, on average, the estimated flux errors in Alaska and South America closely match the true flux errors with ratios of 0.98 [0.89, 1.08] and 0.99 [0.79, 1.24], respectively, while mid-latitude North America, Europe, East Asia, Southeast Asia, and Australia show significant underestimation at a 95% confidence level with ratios of 0.90 [0.83, 0.97], 0.79 [0.61, 0.97], 0.84 [0.78, 0.91], 0.75 [0.65, 0.86], and 0.73 [0.59, 0.87], respectively, throughout the analysis period. Furthermore, the monthly variabilities (i.e., standard deviation) of the ratios are much greater in regions with diverse campaign durations and routes, such as South America (0.87), than in East Asia (0.21), characterized by a consistent three-year observation campaign along the same paths. This suggests that the spatial variability in the degree of flux error underestimation or overestimation may exceed the temporal variability.

3.3 Error quantification of v10 OCO-2 MIP ensemble posterior net CO_2 fluxes by regions

Next, by incorporating the monthly $RMSE$, ERR_{TOT}, and $\boldsymbol{h(err_{f_e})}$, we derive monthly true posterior flux errors in CO_2 space (i.e., $\boldsymbol{h(err_{f_t})}$) for each region during the period 2015–2017 (Figure 5). Regionally averaged $\boldsymbol{h(err_{f_t})}$ exhibits different seasonal and monthly variability compared to $\boldsymbol{h(err_{f_e})}$. In the northern mid-latitude regions, $\boldsymbol{h(err_{f_e})}$ shows clear seasonal cycles for the entire analysis period, despite different observation routes in each month. For example, in mid-latitude North America and East Asia, the growing season (May to October; 0.6 and 0.9 ppm, respectively) experiences higher $\boldsymbol{h(err_{f_e})}$ than the non-growing season (November to April; 0.4 and 0.7 ppm). The seasonal variations are also observed in $\boldsymbol{h(err_{f_t})}$ in East Asia and partially in mid-latitude North America for 2017, but they are not discernible in Alaska and Europe. In addition, monthly $\boldsymbol{h(err_{f_t})}$ does not exhibit a significant correlation ($p<0.05$) with monthly $\boldsymbol{h(err_{f_e})}$ in Alaska, midlatitude North America, Southeast Asia, and South America. $\boldsymbol{h(err_{f_t})}$ displays greater monthly variability than $\boldsymbol{h(err_{f_e})}$. For example, in mid-latitude North America and East Asia, the standard deviation of monthly $\boldsymbol{h(err_{f_t})}$ is 1.8 and 2.3 times greater than that of monthly $\boldsymbol{h(err_{f_e})}$.

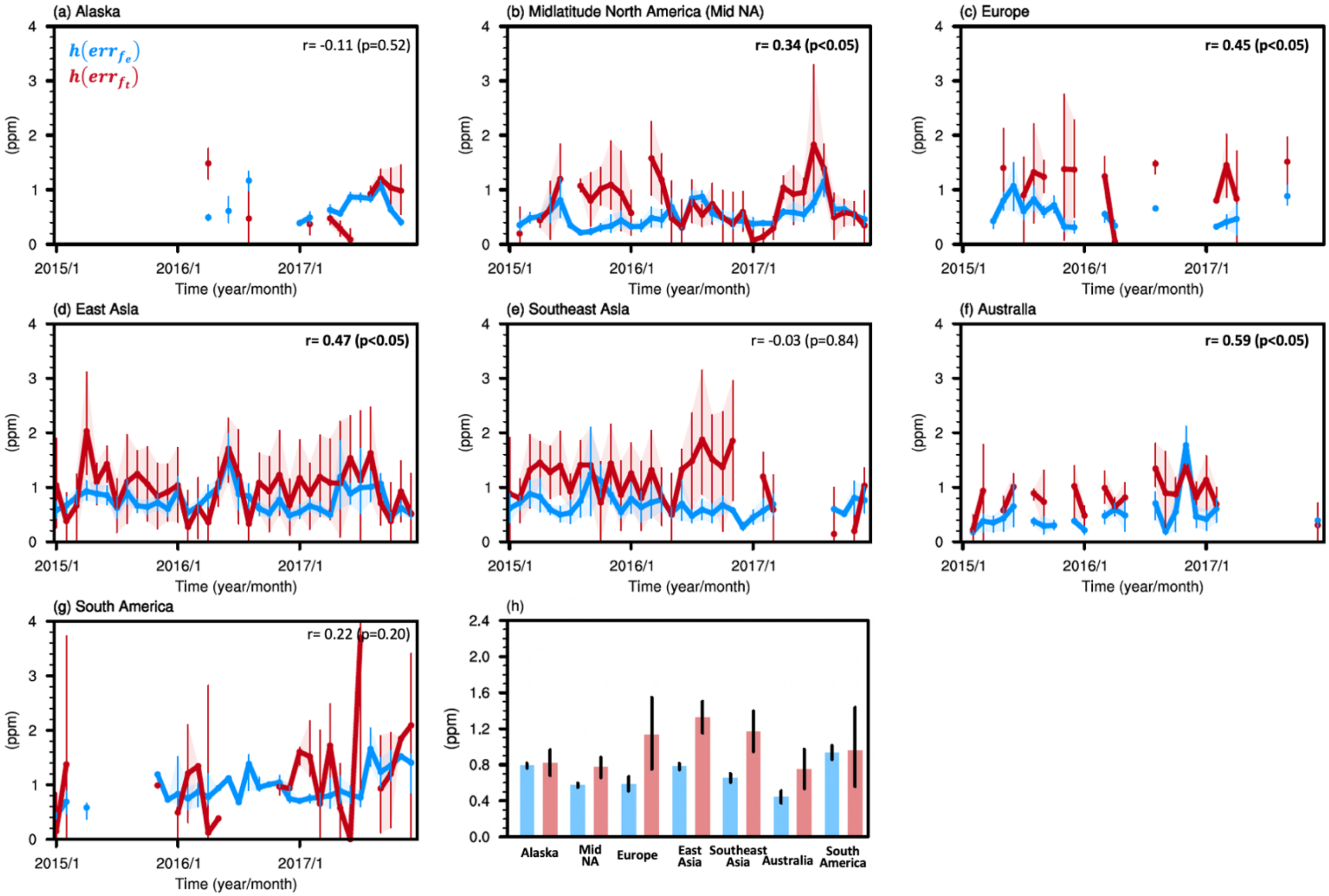

Figure 5: (a-g) Monthly values of $h(err_{f_e})$ and $h(err_{f_t})$ for each region and (h) their mean values for the period 2015–2017. The upper right number indicates the correlation coefficient between them. The shaded areas and error bars represent the 95% confidence intervals derived from 1000 bootstrap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

The comparison between the three-year average $h(err_{f_t})$ and $RMSE$ highlights the substantial contributions of posterior flux errors to the differences between airborne observations and simulated atmospheric CO_2 from OCO-2 MIP ensemble models. The $h(err_{f_t})$ tends to be larger in regions with higher $RMSE$, peaking in East Asia ($h(err_{f_t})$=1.32 ppm and $RMSE$=1.98 ppm) and reaching a minimum in Australia ($h(err_{f_t})$=0.75 ppm and $RMSE$=0.88 ppm) (Figures 4h and 5h). The $h(err_{f_t})$ accounts for up to 85% of the $RMSE$ in Australia, followed by Southeast Asia (80%) and a minimum of 60% of the $RMSE$ in South America, followed by mid-latitude North America (64%). This indicates dominant contributions of posterior flux errors to $RMSE$, surpassing representation and transport errors in the first two regions.

The regional mean ratios between $h(err_{f_e})$ and $h(err_{f_t})$ throughout the analysis period indicate significant underestimations at a 95% confidence level of true posterior flux errors in mid-latitude North America, Europe, East Asia, Southeast Asia, and Australia by a factor of 0.74 [0.61, 0.88], 0.52 [0.27, 0.78], 0.59 [0.48, 0.70], 0.56 [0.41, 0.72], and 0.59 [0.34, 0.87], respectively (Figure 5h). In contrast, Alaska and South America exhibit comparable estimates of true flux errors by factors of 0.96 [0.76, 1.17] and 0.97 [0.49, 1.54], respectively. The regions with significant underestimation align with those identified in the previous analysis based on ratios between ERR_{TOT} and *RMSE* (Section 3.2), but the $h(err_{f_e})$ to $h(err_{f_t})$ ratios imply weaker underestimation of true flux errors. The ratios have larger uncertainty range in the regions where observations conducted over limited times and locations, such as those in Europe, Australia, and South America than in the mid-latitude North America and East Asia where observations cover wider areas and occur more frequently.

Finally, by using the three-year regional mean ratios between $h(err_{f_e})$ and $h(err_{f_t})$, we compute the true errors in the annual net land fluxes over the effective areas averaged for the period 2015–2017 (Figure 6). We find that the actual flux errors are underestimated, particularly in regions where annual CO_2 emissions from fossil fuel combustion exceed annual terrestrial biosphere fluxes by 3-31 times. The airborne measurements carried out in mid-latitude North America, East Asia, and Southeast Asia are influenced by a broad region encompassing the United States, the eastern part of East Asia, and the western part of Southeast Asia where fossil fuel CO_2 emissions are 1,341, 2,443, and 815 Tg C year^{-1}, respectively. The first two regions are estimated as significant terrestrial biosphere CO_2 sinks, with estimated fluxes of −414 ± 279 (ensemble mean ± 1σ) and −561 ± 380 Tg C year^{-1}, in contrast to Southeast Asia (26 ± 118 Tg C year^{-1}). However, the CO_2 sinks are more than 3 and 4 times smaller than the fossil fuel CO_2 emissions, respectively. The recalculated net land flux errors in these regions exceed the ensemble spread with values of 374, 643, and 211 Tg C year^{-1}. Observations in Europe and Australia, conducted over limited periods and specific locations, mainly represent certain areas in the western Europe and the southeastern part of Australia, where fossil fuel emissions (234 and 53 Tg C year^{-1}, respectively) are around four and five times greater than terrestrial biosphere sinks (−51 ± 34 and −10 ± 67 Tg C year^{-1}). The recalculated net land flux errors in these regions are also larger than the ensemble spread, estimated at 65 and 114 Tg C year^{-1}, respectively. On the contrary, the most influential areas for the observation in Alaska and South America, encompassing the southeastern region of Alaska and the northern part of Brazil, characterized as a terrestrial biosphere sinks of −8 ± 11 Tg C year^{-1} and sources of 625 ± 387 Tg C year^{-1}, respectively, which are comparable to or more than 10 times greater than fossil fuel emissions (10 and 38 Tg C year^{-1}). The observation-based estimates of true net land flux errors are almost identical to the ensemble spread in both regions with values of 11 and 398 Tg C year^{-1}, respectively.

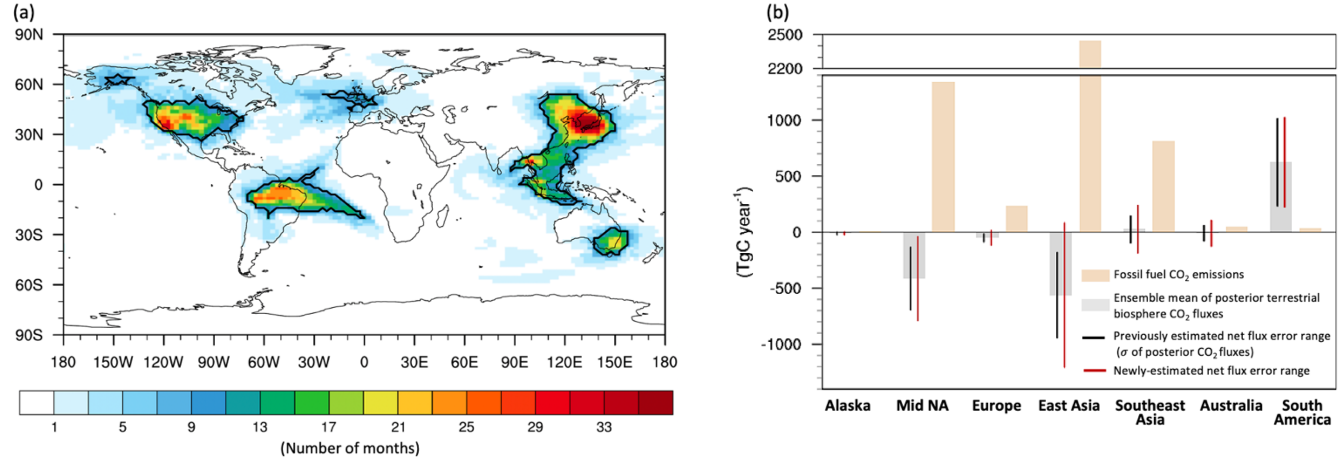

Figure 6: (a) Number of months selected as the effective area for airborne measurements. The outlined area represents selected areas for more than eight months or equal. (b) Annual total terrestrial biosphere CO_2 fluxes obtained from the ensemble mean of ten OCO-2 MIP models and annual total fossil fuel CO_2 emissions estimated from ODIAC data for each outlined area averaged over the period 2015–2017. The black error bars denote ± one standard deviation of the posterior net land fluxes, identical to those of the posterior terrestrial biosphere fluxes. The error bars in red indicate the newly-estimated range of errors in the posterior net land fluxes from this study.

4. Discussion and conclusions

Our results show that the errors in the posterior net land CO_2 fluxes is a major factor contributing to the RMSE between posterior simulated CO_2 and airborne observations for the period 2015–2017. Our findings reaffirm the feasibility of evaluating inversion performance on land flux estimates through a direct comparison between airborne observations and model data (Houweling et al., 2015; Chevallier et al., 2019; Crowell et al., 2019; Byrne et al., 2023). However, when evaluating inversion estimates at regional scales, the significance of representation and transport errors become pronounced. Our results show that regional variations in representation errors, along with the sum of transport errors and their covariances with flux errors (inferred from the difference between ERR_{MIP} and $h(err_{f_e})$; Fig. S6), exceed those in true flux errors projected into CO_2 space, indicating that regional differences in *RMSE* do not directly correspond to differences in flux errors. For example, although the three-year mean errors in representation and transport in East Asia exceed those in Southeast Asia by 0.5 and 0.3 ppm, respectively, the disparity in projected mean true flux errors onto CO_2 space between the two regions is only 0.2 ppm. This result is supported by previous studies highlighting that the spatial distributions of simulated CO_2 concentrations can vary significantly depending on the transport model (Schuh et al., 2023) and their spatial resolution (Stanevich et al., 2020). Therefore, when utilizing airborne CO_2 measurements (and potentially other CO_2 observation) 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 representation and transport errors.

Our analysis reveals that the true errors in the ensemble mean of posterior net CO_2 flux estimates is significantly greater than the ensemble spread of flux estimates in five out of seven regions with higher fossil fuel emissions compared to terrestrial biosphere fluxes. Possible explanation for this result is the presence of errors in the prescribed fossil fuel emissions common to all OCO-2 MIP models. OCO-2 MIP models treated fossil fuel emissions as perfectly known values and adjusted terrestrial biosphere and ocean CO_2 fluxes to minimize the difference between the simulated and observed CO_2 concentrations. Thus, if there are errors in the prescribed fossil fuel emission estimates, these errors propagate into the posterior natural flux estimates. The assumption used in the OCO-2 MIP models is, in fact, the one often applied in conventional global atmospheric inverse models as it is considered that the errors in fossil fuel emission estimates are relatively lower than those in natural flux estimates at national scales (4-20%; Andres et al., 2014). However, the emission errors become substantial when considering spatial distribution at model grid scale and temporal variability within a year (Zhang et al., 2016; Gurney et al., 2021). Oda et al. (2023) showed significant impacts of differences in fossil fuel emission estimates on posterior terrestrial biosphere flux estimates near the source regions. OCO-2 MIP models used identical fossil fuel emission estimates and thus their posterior net flux estimates share common biases induced by the errors in the fossil fuel emission estimates. Because these systematic biases are not captured by the ensemble spread of flux estimates, true flux errors exceed the errors computed from the ensemble spread in the main source regions. In addition to this, 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 biosphere 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 in current ensemble inverse modeling estimates.

The reliability of our observation-based regional flux error estimates is based upon the data availability of airborne measurements. Although our approach is generally effective in estimating a regional mean of monthly $h\left(err_{f_t}\right)$, it is not applicable in 15% of our total cases (shown in Figure 5), when measurements were mostly made in local areas covering one to six 1°×1° grid cells within each region. 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 locally. Extending the calculation period to several months or longer (e.g., Figure 5h) is a suitable strategy for mitigating the impact of outliers and obtaining more robust results. In fact, the ratios of three-year mean $h\left(err_{f_e}\right)$ to $h\left(err_{f_t}\right)$, which are key metrics for quantifying regional flux errors (Figure 5h), have a smaller uncertainty in mid-latitude North America and East Asia where wide and consistent airborne data are available, than over Europe and South America, where airborne observations are sparse and only have intermittent data coverage. In addition, it is noteworthy that the $h\left(err_{f_e}\right)$ to $h\left(err_{f_t}\right)$ ratios derived from continuous observations enable the computation of unbiased true errors in the ensemble mean of annual posterior net fluxes averaged for the analysis period, compared to those from limited observation periods (e.g., in Alaska). These results highlight the importance of having frequent airborne measurements with extensive spatial coverage for the reliable error quantification of regional net flux estimates derived from inverse models.

The performance of inverse models in simulating atmospheric CO_2 may vary by season. However, airborne measurements were not uniformly conducted across all seasons in most analyzed regions. Among the seven regions analyzed, the CONTRAIL program in East Asia has continuously conducted CO_2 measurements over three years with routes repeated throughout all seasons. This has resulted in the most sensitive area to the measurements exhibiting similar spatial patterns in the NH vegetation growing season (from May to October) and non-growing season, encompassing the northeast part of China, Korea, and Japan (Fig. S7). The airborne measurements in East Asia offer a unique opportunity to explore the seasonal variations of regional error statistics. For the period of 2015–2017, the regional averages of both *RMSE* and ERR_{TOT} exhibit, on average, 12% and 11% higher values during the non-growing season compared to the growing season (Fig. S8). In contrast, the regional averages of $h(err_{f_e})$ and $h(err_{f_t})$ have greater values during the growing season, 0.90 [0.84, 0.97] and 1.37 [1.13, 1.62] ppm respectively, compared to the non-growing season (0.66 [0.62, 0.70] and 1.30 [1.06, 1.54] ppm) because of the tendency for CO_2 errors to increase proportionally with the magnitude of flux values. Consequently, the ratio of $h(err_{f_e})$ to $h(err_{f_t})$ is slightly lower during the non-growing season with 0.51 [0.39, 0.64] compared to the growing season with 0.66 [0.50, 0.83], indicating a relatively greater underestimation of true flux errors when the terrestrial biosphere CO_2 sinks are relatively smaller. This result aligns with our finding that the true net land flux errors are significantly underestimated where fossil fuel emissions have larger magnitude than terrestrial biosphere fluxes. Furthermore, the consistent ratio of $h(err_{f_e})$ to $h(err_{f_t})$ below 1, without significant seasonal variations in East Asia, suggests that our conclusions, drawn from the analysis of seven regions, may not be seasonally dependent.

To capture the signals from regional surface CO_2 fluxes, we used atmospheric CO_2 data observed and simulated within the 1-5 km AGL altitude range. The choice of this altitude range may influence regional error statistics, as the performance of inverse models could vary with altitude. To gauge this sensitivity, we compared error statistics derived from atmospheric CO_2 data with two altitude ranges: 1-3 km AGL and 1-5 km AGL. Among the seven analyzed regions, Australia and South America were excluded in this additional analysis because the airborne observation in these two regions cover fewer than 100 grid cells for the analysis period and narrowing the altitude range resulted in the loss of over 30% of the grid cells. The areas sensitive to airborne CO_2 measurements within the two altitude ranges exhibit nearly identical spatial patterns in Alaska, mid-latitude North America, Europe, East Asia, and Southeast Asia, indicating that observations at lower altitudes are more sensitive to surface CO_2 fluxes (Fig. S9). Because of the higher sensitivity, error statistics in all regions have larger values when calculated using data from the 1-3 km AGL altitude range compared to the 1-5 km AGL altitude range (Fig. S10). For example, in mid-latitude North America, the regional averages of *RMSE*, ERR_{TOT}, $h(err_{f_e})$, and

$h(err_{f_t})$ are 1.42 [1.36, 1.49], 1.34 [1.30, 1.39], 0.72 [0.69, 0.76], and 0.86 [0.72, 1.01] ppm when calculated using data within the 1-3 km AGL altitude range. In comparison, when computed from the data within the 1-5 km AGL altitude range, these values are 1.21 [1.15, 1.26], 1.09 [1.06, 1.13], 0.57 [0.55, 0.60], and 0.77 [0.66, 0.88] ppm. However, the ratio of three-year mean $h(err_{f_e})$ to $h(err_{f_t})$ does not show significant differences based on the altitude ranges, with the difference being between 0.02 and 0.11. Again, these results suggest that our observation-based regional flux error estimates are not sensitive to the choice of altitude range for longer time periods.

Our study computes true flux errors for the ensemble mean estimates by comparing $RMSE^2$ and ERR_{TOT}^2. However, discrepancies between true and estimated values of observation, representation, and transport errors, as well as covariances between flux errors and transport errors, could contribute to variations in $RMSE^2$ and ERR_{TOT}^2. Due to a lack of information for all datasets, we set observation errors under ideal conditions (i.e., 0.1 ppm). In reality, inadequate quality control can result in significant systematic biases for specific regions and time periods (Masarie et al., 2011; Baier et al., 2020), impacting our results, especially in South America. For instance, if the average measurement error is 0.5 ppm instead of the assumed 0.1 ppm during the analysis period, the calculated true flux error would decrease from 398 to 334 Tg C year^{-1} for South America and from 374 to 260 Tg C year^{-1} for mid-latitude North America.

Representation errors and $h(err_{f_e})$ are derived using the GEOS-5 and GEOS-Chem models but these values depend on the transport model and meteorological fields 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. Employing all transport models also 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 errors including both diagonal and off-diagonal terms. In addition, 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. We further investigate how our main results would be affected if the estimated transport errors deviate from actual errors by 20% and 40% of the difference between $RMSE^2$ and ERR_{TOT}^2. The ratio of regional mean of $h(err_{f_e})$ to $h(err_{f_t})$ increases by, on average, only up to 0.04 and 0.09 in the seven regions throughout the analysis period, respectively (Fig. S11). In both cases, the estimated flux errors in mid-latitude North America, Europe, East Asia, and Southeast Asia still show significant underestimation at a 95% confidence level, while not in Alaska and South America. In Australia, characterized by a wide uncertainty range, significant underestimation is also observed in the 20% cases, supporting the robustness of our findings. In the future OCO-2 MIP, the participation of inverse modeling groups using other transport models or meteorological forcing data might contribute to estimating transport errors closer to actual values.

This study uses monthly posterior flux estimates for the calculation of monthly $h(err_{f_e})$. However, posterior flux estimates from each OCO-2 MIP model have different sub-monthly patterns, which could modify the sub-monthly variations in posterior atmospheric CO_2 concentrations and, in turn, affect their ensemble spread. To examine their potential impact on the results, we conduct an analysis with different publicly available hourly (or 3-hourly) terrestrial biosphere fluxes (Chevallier et al., 2019; Jacobson et al., 2020; Ott et al., 2020; Haynes et al., 2021; Liu and Bowman, 2024) which are from seven OCO-2 MIP prior flux models (Ames, Baker, CAMS, CMS-Flux, CT, OU, and WOMBAT; Table S1). By incorporating the monthly-balanced hourly flux estimates into the monthly posterior fluxes, we generate hourly posterior terrestrial biosphere flux estimates for these seven models. Since the assimilation window for each OCO-2 MIP model ranges from one week to one month, the weekly variations in posterior fluxes may differ from those in the prior fluxes. Nonetheless, with only the monthly posterior flux estimates being publicly available, this approach offers valuable insights into how different sub-monthly patterns of posterior fluxes could affect our main results. Our analysis shows that the regional averages of $h(err_{f_e})$ derived from the monthly posterior flux estimates from the seven models are, on average, within ±10% of the values originally obtained using flux estimates from 10 models for the period 2015–2017, except for Europe (13% lower) (Fig. S12a). When accounting for different sub-monthly patterns of posterior fluxes across models, the regional averages of $h(err_{f_e})$ increase by 10-22% (0.06-0.14 ppm) across six regions, with a 45% (0.23 ppm) increase in Europe. These results suggest that our earlier calculation, assuming identical sub-monthly flux variations, underestimates $h(err_{f_e})$. We further investigate whether our main finding remains robust even if we adjust the original values of $h(err_{f_e})$ using the potential underestimation rate. After making the correction, we found that the ratios of the regional average $h(err_{f_e})$ to $h(err_{f_t})$ increase the most in Europe by 0.14 and only up to 0.07 in the other six regions, as $h(err_{f_t})$ also rises with $h(err_{f_e})$ according to Eq. (8) (Fig. S12b). Moreover, the $h(err_{f_t})$ still exhibits significant underestimation ($p<0.05$) in mid-latitude North America, Europe, East Asia, Southeast Asia, and Australia. This indicates that our main results are robust to the inclusion or exclusion of sub-monthly flux patterns in the calculation of $h(err_{f_e})$.

In summary, our study provides an observation-based method for quantifying errors in the ensemble mean of regional net CO_2 flux estimates which can be widely applied in inverse modeling inter-comparison projects like 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 CO_2 emissions. This result provides observation-based evidence supporting previous studies (Oda et al., 2023; Wang et al., 2020) that emphasized the impact of fossil fuel emission errors on global atmospheric CO_2 inversions. This finding offers important insights into understanding the sources of errors in current inverse modeling and highlights the need for improving fossil fuel emission estimates and developing inversion methods that optimize both fossil fuel emissions and natural fluxes. Airborne observations provide a broader footprint compared to ground-based observations. Leveraging this advantage, our study evaluates 19% of the total global land cover (excluding Antarctica and Greenland) but data scarcity limits the evaluation of the remaining 81%. In addition to the ongoing airborne measurement programs including CONTRAIL, IAGOS-CARIBIC, and various airborne programs under INPE, NASA, and 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 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 surface CO_2 fluxes.

Appendix A

Following Eq. (1) in the main text,

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

where $\overline{h(\hat{x}_i)}$ denotes ensemble mean of posterior CO_2 concentrations in OCO-2 MIP models corresponding to i^{th} airborne observation ($y_{o,i}$) within each 1°×1° grid-cell in each month. N is the total number of airborne measurement data sampled at each grid-cell monthly. M is the ensemble size (i.e., 10 members).

The Eq. (A1) can be rewritten as,

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

$$= \frac{1}{N}\sum_{i=1}^{N}[y_{o,i} - h_t(\hat{x}_{t,i})]\,[y_{o,i} - h_t(\hat{x}_{t,i})]^T - 2\left(y_o - h_t(\hat{x}_{t,i})\right) * \left(\overline{h(\hat{x}_i)} - h_t(\hat{x}_{t,i})\right)$$
$$+ [\overline{h(\hat{x}_i)} - h_t(\hat{x}_{t,i})][\overline{h(\hat{x}_i)} - h_t(\hat{x}_{t,i})]^T, \tag{A3}$$

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 and representation errors, (ii) covariances between errors of observation and representation and errors of flux and transport, and (iii) variances of flux and transport errors in the ensemble estimates, respectively. Assuming the independence of observation and representation errors from transport and flux errors, Eq. (A3) can be simplified to:

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

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:

$$\frac{1}{N}\sum_{i=1}^{N}\,[\overline{h(\hat{x}_i)} - h_t(\hat{x}_{t,i})][\overline{h(\hat{x}_i)} - h_t(\hat{x}_{t,i})]^T$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right) - \left(h_t(\hat{x}_{t,i}) - \overline{h(\hat{x}_{t,i})}\right)\right]\left[\left(\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right) - \left(h_t(\hat{x}_{t,i}) - \overline{h(\hat{x}_{t,i})}\right)\right]^T \tag{A5}$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left[\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right]\left[\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right]^T - 2\left(\overline{h(\hat{x}_i)} - \overline{h(\hat{x}_{t,i})}\right)\left(h_t(\hat{x}_{t,i}) - \overline{h(\hat{x}_{t,i})}\right)$$

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

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

In OCO-2 MIP, by approximating the ensemble spread of the posterior fluxes as true errors in the mean fluxes, it assumes that the values of the first and second terms on the right-hand side of Eq. (A4) can be written as the sum of observation errors (ERR_{OBS}^2), representation errors (ERR_{REP}^2), and the ensemble spread of posterior CO_2 concentrations across OCO-2 MIP models (ERR_{MIP}^2), respectively:

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

We assume that the observation errors are independent of the representation errors.

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

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

$$= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}\frac{1}{M}\sum_{k=1}^{M}\left[\left(h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right) - \left(h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right)\right]$$
$$\left[\left(h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right) - \left(h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right)\right]^T \quad \text{(A9)}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{k=1}^{M}\frac{1}{M}\sum_{j=1}^{M}\left[h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right]\left[h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right]^T$$
$$-2\left(h_k(\hat{x}_{k,i}) - h_k(\hat{x}_{j,i})\right)\left(h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right)$$
$$+\left[h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right]\left[h_j(\hat{x}_{j,i}) - h_k(\hat{x}_{j,i})\right]^T \quad \text{(A10)}$$

Same as Eq. (A6), the three terms on the right-hand side of Eq. (A10) correspond to the approximated (i) variances of flux errors, (ii) covariances between flux errors and transport errors, and (iii) variances of transport errors, respectively. For the calculation of the first term, utilizing all participating transport models in the OCO-2 MIP would be ideal but, in this study, we approximate it using the GEOS-Chem model.

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 are conducted using the 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). ODIAC fossil fuel CO_2 emission data is available at 10.17595/20170411.001. Hourly (or 3-hourly) terrestrial biosphere carbon flux datasets from CASA-GFED3, CASA-GFED4.1s, SiB4, CARDAMOM, and ORCHIDEE, which are v10 OCO-2 MIP prior flux models, are available at doi.org/10.5067/VQPRALE26L20, https://gml.noaa.gov/aftp//products/carbontracker/co2/CT2019B/fluxes/, doi.org/10.3334/ORNLDAAC/1847, doi.org/10.5067/1XO0PZEZOR1H, and https://ads.atmosphere.copernicus.eu/datasets/cams-global-greenhouse-gas-inversion?tab=download (version v21r1).

Author contributions

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

We are grateful for the airborne CO_2 observation data provision by NOAA, NASA-LaRC, NASA-GSFC, INPE, NIES, MRI, the European research infrastructure IAGOS-CARIBIC, DOE/LBNL, and Harvard University. We thank to C. Sweeney, K. Davis, J. P. Digangi, M. Y. Martin, , J. B. Miller, E. Gloor, W. Peters, 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. Peischl, K. Aikin, S. Wofsy, B. Daube, R. Commane for contributing airborne CO_2 data. 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.

Financial support

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004) and has been supported by the National Aeronautics and Space Administration (grant no. 20-OCOST20-0012). South America airborne CO_2 measurements are supported by the European Research Council through ASICA (grant no. 649087). Observations collected at the SGP site were partly supported by the Office of Biological and Environmental Research of the US Department of Energy under contract no. DE-AC02-05CH11231 as part of the Atmospheric Radiation Measurement Program (ARM), ARM Aerial Facility, and Environmental System Science Program (ESS).

References

1. Andres, R. J., Boden, T. A., and Higdon, D.: A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission, Tellus B: Chem. Phys. Meteorol., 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, J. Geophys. Res. Atmos., 125, e2019JD031339, https://doi.org/10.1029/2019JD031339, 2020.
3. Baier, B., Sweeney, C., Tans, P., Newberger, T., Higgs, J., and Wolter, S.: NOAA AirCore atmospheric sampling system profiles (Version 20210813), NOAA Global Monitoring Laboratory [data set], https://doi.org/10.15138/6AV0-MY81, 2021.
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, Atmos. Chem. Phys., 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, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci., 380, 20210192, https://doi.org/10.1098/rsta.2021.0192, 2021.
6. Basu, S., & Nassar, R.: Fossil Fuel CO_2 Emissions for the OCO2 Model Intercomparison Project (MIP) (2020.1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4776925, 2021.
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, Atmos. Chem. Phys., 18, 7189–7215, https://doi.org/10.5194/acp-18-7189-2018, 2018.
8. Biraud, S. C., Torn, M. S., Smith, J. R., Sweeney, C., Riley, W. J., and Tans, P. P.: A multi-year record of airborne CO_2 observations in the US Southern Great Plains, Atmos. Meas. Tech., 6, 751–763, https://doi.org/10.5194/amt-6-751-2013, 2013.
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 Syst. Sci. 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, J. Geophys. Res. Atmos., 112, https://doi.org/10.1029/2006JD007375, 2007.
11. 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, Atmos. Chem. Phys., 19, 14233–14251, https://doi.org/10.5194/acp-19-14233-2019, 2019.
12. 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), Geosci. Model Dev., 15, 1289–1316, https://doi.org/10.5194/gmd-15-1289-2022, 2022.
13. 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, Atmos. Chem. Phys., 19, 9797–9831, https://doi.org/10.5194/acp-19-9797-2019, 2019.
14. Davis, K. J., Browell, E. V., Feng, S., Lauvaux, T., Obland, M. D., Pal, S., Baier, B. C., Baker, D. F., Baker, I. T., Barkley, Z. R., Bowman, K. W., Cui, Y. Y., Denning, A. S., DiGangi, J. P., Dobler, J. T., Fried, A., Gerken, T., Keller, K., Lin, B., Nehrir, A. R., Normile, C. P., O'Dell, C. W., Ott, L. E., Roiger, A., Schuh, A. E., Sweeney, C., Wei, Y., Weir, B., Xue, M., and Williams, C. A.: The Atmospheric Carbon and Transport (ACT)-America Mission, Bull. Am. Meterol. Soc., 102, E1714–E1734, https://doi.org/10.1175/BAMS-D-20-0300.1, 2021.
15. DiGangi, J. P., Choi, Y., Nowak, J. B., Halliday, H. S., Diskin, G. S., Feng, S., Barkley, Z. R., Lauvaux, T., Pal, S., Davis, K. J., Baier, B. C., and Sweeney, C.: Seasonal Variability in Local Carbon Dioxide Biomass Burning Sources Over Central and Eastern US Using Airborne In Situ Enhancement Ratios, J. Geophys. Res. Atmos., 126, e2020JD034525, https://doi.org/10.1029/2020JD034525, 2021.
16. 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, Atmos. Chem. Phys., 9, 2619–2633, https://doi.org/10.5194/acp-9-2619-2009, 2009.
17. Feng, S., Lauvaux, T., Davis, K. J., Keller, K., Zhou, Y., Williams, C., Schuh, A. E., Liu, J., and Baker, I.: Seasonal Characteristics of Model Uncertainties From Biogenic Fluxes, Transport, and Large-Scale Boundary Inflow in

Atmospheric CO_2 Simulations Over North America, J. Geophys. Res. Atmos., 124, 14325–14346, https://doi.org/10.1029/2019JD031165, 2019.

18. 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: Chem. Phys. Meteorol., 67, 27989, https://doi.org/10.3402/tellusb.v67.27989, 2015.
19. Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Barbero, L., Bates, N. R., Becker, M., Bellouin, N., Decharme, B., Bopp, L., Brasika, I. B. M., Cadule, P., Chamberlain, M. A., Chandra, N., Chau, T.-T.-T., Chevallier, F., Chini, L. P., Cronin, M., Dou, X., Enyo, K., Evans, W., Falk, S., Feely, R. A., Feng, L., Ford, D. J., Gasser, T., Ghattas, J., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Joos, F., Kato, E., Keeling, R. F., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot, N., McGuire, P. C., McKinley, G. A., Meyer, G., Morgan, E. J., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K. M., Olsen, A., Omar, A. M., Ono, T., Paulsen, M., Pierrot, D., Pocock, K., Poulter, B., Powis, C. M., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Séférian, R., et al.: Global Carbon Budget 2023, Earth Syst. Sci. Data, 15, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023, 2023.
20. Gatti, L. V., Cunha, C. L., Marani, L., Cassol, H. L. G., Messias, C. G., Arai, E., Denning, A. S., Soler, L. S., Almeida, 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.
21. Gaubert, B., Stephens, B. B., Baker, D. F., Basu, S., Bertolacci, M., Bowman, K. W., Buchholz, R., Chatterjee, A., Chevallier, F., Commane, R., Cressie, N., Deng, F., Jacobs, N., Johnson, M. S., Maksyutov, S. S., McKain, K., Liu, J., Liu, Z., Morgan, E., O'Dell, C., Philip, S., Ray, E., Schimel, D., Schuh, A., Taylor, T. E., Weir, B., van Wees, D., Wofsy, S. C., Zammit-Mangion, A., and Zeng, N.: Neutral Tropical African CO2 Exchange Estimated From Aircraft and Satellite Observations, Global Biogeochem. Cycles, 37, e2023GB007804, https://doi.org/10.1029/2023GB007804, 2023.
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, 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 Biogeochem. Cycles, 18, https://doi.org/10.1029/2003GB002111, 2004.
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. Haynes, K.D., Baker, I.T., and Denning, A.S.: SiB4 Modeled Global 0.5-Degree Monthly Carbon Fluxes and Pools, 2000-2018. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1848, 2021.
26. Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmos. Chem. Phys., 7, 2413–2433, https://doi.org/10.5194/acp-7-2413-2007, 2007.
27. Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., Deng, F., Dlugokencky, E. J., Feng, L., 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, J. Geophys. Res. Atmos., 120, 5253–5266, https://doi.org/10.1002/2014JD022962, 2015.
28. Jacobson, A. R., Schuldt, K. N., Miller, J. B., Oda, T., Tans, P., Andrews, A., Mund, J., Ott, L., Collatz, G. J., Aalto, T., Afshar, S., Aikin, K., Aoki, S., Apadula, F., Baier, B., Bergamaschi, P., Beyersdorf, A., Biraud, S. C., Bollenbacher, A., Bowling, D., Brailsford, G., Abshire, J. B., Chen, G., Chen, H., Chmura, L., Climadat, S., Colomb, A., Conil, S., Cox, A., Cristofanelli, P., Cuevas, E., Curcoll, R., Sloop, C. D., Davis, K., Wekker, S. D., Delmotte, M., DiGangi, J. P., Dlugokencky, E., Ehleringer, J., Elkins, J. W., Emmenegger, L., Fischer, M. L., Forster, G., Frumau, A., Galkowski, M., Gatti, L. V., Gloor, E., Griffis, T., Hammer, S., Haszpra, L., Hatakka, J., Heliasz, M., Hensen, A., Hermanssen, O., Hintsa, E., Holst, J., Jaffe, D., Karion, A., Kawa, S. R., Keeling, R., Keronen, P., Kolari, P., Kominkova, K., Kort, E., Krummel, P., Kubistin, D., Labuschagne, C., Langenfelds, R., Laurent, O., Laurila, T., Lauvaux, T., Law, B., Lee, J., Lehner, I., Leuenberger, M., Levin, I., Levula, J., Lin, J., Lindauer, M., Loh, Z., Lopez, M., Luijkx, I. T., Lund Myhre, C., Machida, T., Mammarella, I., Manca, G., Manning, A., Manning, A., Marek, M. V., Marklund, P., Martin, M. Y., Matsueda, H., McKain, K., Meijer, H., Meinhardt, F., Miles, N., Miller, C. E., Mölder, M., Montzka, S., Moore, F., Morgui, J.-A., Morimoto, S., Munger, B., Necki, J., Newman, S., Nichol, S., Niwa, Y., O'Doherty, S., Ottosson-Löfvenius, M., Paplawsky, B., Peischl, J., Peltola, O., Pichon, J.-M., Piper, S., Plass-Dölmer, C., Ramonet, M., Reyes-Sanchez, E., Richardson, S., Riris, H., Ryerson, T., Saito, K., Sargent, M., Sasakawa, M., Sawa, Y., Say, D., Scheeren, B., Schmidt, M., Schmidt, A., Schumacher, M., Shepson, P., Shook, M., Stanley, K., Steinbacher, M., Stephens, B., Sweeney, C., Thoning, K., Torn, M., Turnbull, J., Tørseth, K., Bulk, P. V. D., Laan-Luijkx, I. T. V. D., Dinther, D. V., Vermeulen, A., Viner, B., Vitkova, G., Walker, S., Weyrauch, D., Wofsy,

S., Worthy, D., Young, D., and Zimnoch, M.: Carbontracker CT2019B, Model published 2020 by NOAA Earth System Research Laboratory, Global Monitoring Division, https://doi.org/10.25925/20201008, 2020.

29. Karion, A., Sweeney, C., Tans, P., and Newberger, T.: AirCore: An Innovative Atmospheric Sampling System, J. Atmos. Oceanic Tech., 27, 1839–1853, https://doi.org/10.1175/2010JTECHA1448.1, 2010.
30. Karion, A., Sweeney, C., Wolter, S., Newberger, T., Chen, H., Andrews, A., Kofler, J., Neff, D., and Tans, P.: Long-term greenhouse gas measurements from aircraft, Atmos. Meas. Tech., 6, 511–526, https://doi.org/10.5194/amt-6-511-2013, 2013.
31. 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: 20190000855, NASA/TP–2018-219034, GSFC-E-DAA-TN64573, 2018.
32. 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., 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.
33. 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, Atmos. Chem. Phys., 19, 12007–12024, https://doi.org/10.5194/acp-19-12007-2019, 2019.
34. Liu, J., Bowman, K. W., Lee, M., Henze, D. K., Bousserez, N., Brix, H., James Collatz, G., Menemenlis, D., Ott, L., Pawson, S., Jones, D., and Nassar, R.: Carbon monitoring system flux estimation and attribution: impact of ACOS-GOSAT XCO_2 sampling on the inference of terrestrial biospheric sources and sinks, Tellus B: Chem. Phys. Meteorol., 66, 22486, https://doi.org/10.3402/tellusb.v66.22486, 2014.
35. Liu, J., Bowman, K. W., and Henze, D. K.: Source-receptor relationships of column-average CO_2 and implications for the impact of observations on flux inversions, J. Geophys. Res. Atmos., 120, 5214–5236, https://doi.org/10.1002/2014JD022914, 2015.
36. 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 carbon cycle responses of the tropical continents to the 2015–2016 El Niño, Science, 358, eaam5690, https://doi.org/10.1126/science.aam5690, 2017.
37. 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 Syst. Sci. Data, 13, 299–330, https://doi.org/10.5194/essd-13-299-2021, 2021.

38. Liu, J. and Bowman, K.: Carbon Monitoring System Carbon Flux Land Prior L4 V3, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Last access: 17 June 2024, 10.5067/1XO0PZEZOR1H, 2024.
39. 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 CO_2 and Other Trace Gas Species Using Commercial Airlines, J. Atmos. Oceanic Tech., 25, 1744–1754, https://doi.org/10.1175/2008JTECHA1082.1, 2008.
40. 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 CO_2 measurement bias on CarbonTracker surface flux estimates. J. Geophys. Res. Atmos., 116, D17, https://doi.org/10.1029/2011JD016270, 2011.
41. Miller, J. B., Martins, G. A., de Souza, R. A., and Schuldt, K. N.: Manaus Aircraft profile data for the period 2017-2020; obspack_multi-species_1_manaus_profiles_v1.0_202105-19, NOAA Global Monitoring Laboratory [data set], https://doi.org/10.25925/20210519, 2021.
42. Nara, H., Tanimoto, H., Tohjima, Y., Mukai, H., Nojiri, Y., and Machida, T.: Emission factors of CO_2, CO and CH_4 from Sumatran peatland fires in 2013 based on shipboard measurements, Tellus B: Chem. Phys. Meteorol., 69, 1399047, https://doi.org/10.1080/16000889.2017.1399047, 2017.
43. 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, Atmos. Chem. Phys., 23, 2293–2314, https://doi.org/10.5194/acp-23-2293-2023, 2023.
44. NOAA Carbon Cycle Group ObsPack Team: INPE atmospheric carbon dioxide data for the period 2015-2017, obspack_co2_1_INPE_RESTRICTED_v2.0_2018-11-13, NOAA Earth System Research Laboratory, Global Monitoring Division, http://dx.doi.org/10.25925/20181030, 2018.
45. Oda and Maksyutov, 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, 2015.
46. Oda, T., Feng, L., Palmer, P. I., Baker, D. F., and Ott, L. E.: Assumptions about prior fossil fuel inventories impact our ability to estimate posterior net CO_2 fluxes that are needed for verifying national inventories. *Environ. Res. Lett.*, *18*(12), 124030, https://doi.org/10.1088/1748-9326/ad059b, 2023.
47. 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,

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.

48. Ott, L.: GEOS-Carb CASA-GFED 3-hourly Ecosystem Exchange Fluxes 0.5 degree×0.625 degree V3, Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, MD, USA, https://doi.org/10.5067/VQPRALE26L20, 2020.

49. 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, Atmos. Chem. Phys., 22, 1097–1130, https://doi.org/10.5194/acp-22-1097-2022, 2022.

50. 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, Atmos. Chem. Phys., 19, 13267–13287, https://doi.org/10.5194/acp-19-13267-2019, 2019.

51. Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific, River Edge, USA, ISBN 978-9810227401, 2000.

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

53. Schuldt, K. N., Jacobson, A. R., Aalto, T., Andrews, A., Bakwin, P., Bergamaschi, P., Biermann, T., Biraud, S. C., Chen, H., Colomb, A., Conil, S., Cristofanelli, P., De Mazière, M., De Wekker, S., Delmotte, M., Dlugokencky, E., Emmenegger, L., Fischer, M. L., Hatakka, J., Heliasz, M., Hermanssen, O., Holst, J., Jaffe, D., Karion, A., Kazan, V., Keronen, P., Kominkova, K., Kubistin, D., Laurent, O., Laurila, T., Lee, J., Lehner, I., Leuenberger, M., Lindauer, M., Lopez, M., Mammarella, I., Manca, G., Marek, M. V., McKain, K., Miller, J. B., Miller, C. E., Myhre, C. L., Mölder, M., Müller-Williams, J., Piacentino, S., Pichon, J. M., Plass-Duelmer, C., Ramonet, M., Scheeren, B., Schumacher, M., Sha, M. K., Sloop, C. D., Smith, P., Steinbacher, M., Sweeney, C., Tans, P., Thoning, K., Trisolino, P., Tørseth, K., Viner, B., Vitkova, G., and di Sarra, A. G.: Multi-laboratory compilation of atmospheric carbon dioxide data for the years 2020–2021; obspack_co2_1_NRT_v6.1.1_2021-05-17, NOAA Earth System Research Laboratory, Global Monitoring Laboratory [data set], https://doi.org/10.25925/20210517, 2021a.

54. 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., 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, 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., 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, 2021b.

55. 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, Geosci. Model Dev., 13, 3839–3862, https://doi.org/10.5194/gmd-13-3839-2020, 2020.

56. 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, Bull. Am. Meteorol. Soc., 99, 381–402, https://doi.org/10.1175/BAMS-D-16-0206.1, 2018.

57. 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, J. Geophys. Res. Atmos., 120, 5155–5190, https://doi.org/10.1002/2014JD022591, 2015.

58. Sweeney, C., Chatterjee, A., Wolter, S., McKain, K., Bogue, R., Conley, S., Newberger, T., Hu, L., Ott, L., Poulter, B., Schiferl, L., Weir, B., Zhang, Z., and Miller, C. E.: Using atmospheric trace gas vertical profiles to evaluate model fluxes: a case study of Arctic-CAP observations and GEOS simulations for the ABoVE domain, Atmos. Chem. Phys., 22, 6347–6364, https://doi.org/10.5194/acp-22-6347-2022, 2022.

59. Thompson, C. R., Wofsy, S. C., Prather, M. J., Newman, P. A., Hanisco, T. F., Ryerson, T. B., Fahey, D. W., Apel, E. C., Brock, C. A., Brune, W. H., Froyd, K., Katich, J. M., Nicely, J. M., Peischl, J., Ray, E., Veres, P. R., Wang, S., Allen, H. M., Asher, E., Bian, H., Blake, D., Bourgeois, I., Budney, J., Bui, T. P., Butler, A., Campuzano-Jost, P., Chang, C., Chin, M., Commane, R., Correa, G., Crounse, J. D., Daube, B., Dibb, J. E., DiGangi, J. P., Diskin, G. S., Dollner, M., Elkins, J. W., Fiore, A. M., Flynn, C. M., Guo, H., Hall, S. R., Hannun, R. A., Hills, A., Hintsa, E. J., Hodzic, A., Hornbrook, R. S., Huey, L. G., Jimenez, J. L., Keeling, R. F., Kim, M. J., Kupc, A., Lacey, F., Lait, L. R., Lamarque, J.-F., Liu, J., McKain, K., Meinardi, S., Miller, D. O., Montzka, S. A., Moore, F. L., Morgan, E. J., Murphy, D. M., Murray, L. T., Nault, B. A., Neuman, J. A., Nguyen, L., Gonzalez, Y., Rollins, A., Rosenlof, K., Sargent, M., Schill, G., Schwarz, J. P., Clair, J. M. S., Steenrod, S. D., Stephens, B. B., Strahan, S. E., Strode, S. A., Sweeney, C., Thames, A. B., Ullmann, K., Wagner, N., Weber, R., Weinzierl, B., Wennberg, P. O., Williamson, C. J., Wolfe, G. M., and Zeng, L.: The NASA Atmospheric Tomography (ATom) Mission: Imaging the Chemistry of the Global Atmosphere, Bull. Am. Meteorol. Soc., 103, E761–E790, https://doi.org/10.1175/BAMS-D-20-0315.1, 2022.

60. 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.

61. 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)

62. Vay, S. A., Tyler, S. C., Choi, Y., Blake, D. R., Blake, N. J., Sachse, G. W., Diskin, G. S., and Singh, H. B.: Sources and transport of $\Delta^{14}C$ in CO_2 within the Mexico City Basin and vicinity, Atmos. Chem. Phys., 9, 4973–4985, https://doi.org/10.5194/acp-9-4973-2009, 2009.

63. Wang, J. S., Oda, T., Kawa, S. R., Strode, S. A., Baker, D. F., Ott, L. E., and Pawson, S.: The impacts of fossil fuel emission uncertainties and accounting for 3-D chemical CO_2 production on inverse natural carbon flux estimates from satellite and in situ data, *Environ. Res. Lett.*, *15*(8), 085002, https://doi.org/10.1088/1748-9326/ab9795, 2020.

64. Wei, Y., R. Shrestha, S. Pal, T. Gerken, S. Feng, J. McNelis, D. Singh, M. M. Thornton, A. G. Boyer, M. A. Shook, G. Chen, B. C. Baier, Z. R. Barkley, J. D. Barrick, J. R. Bennett, E. V. Browell, J. F. Campbell, L. J. Campbell, Y. Choi, J. Collins, J. Dobler, M. Eckl, A. Fiehn, A. Fried, J. P. Digangi, R. Barton-Grimley, H. Halliday, T. Klausner, S. Kooi, J. Kostinek, T. Lauvaux, B. Lin, M. J. McGill, B. Meadows, N. L. Miles, A. R. Nehrir, J. B. Nowak, M. Obland, C. O'Dell, R. M. P. Fao, S. J. Richardson, D. Richter, A. Roiger, C. Sweeney, J. Walega, P. Weibring, C. A.

Williams, M. M. Yang, Y. Zhou, and K. J. Davis, 2021, Atmospheric Carbon and Transport – America (ACT-America) Datasets: Description, Management, and Delivery, Earth Space Sci, 8, e2020EA001634, https://doi.org/10.1029/2020EA001634, 2021.

65. 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, Sci. Adv., 7, eabf9415, https://doi.org/10.1126/sciadv.abf9415, 2021.
66. Yun, J., Jeong, S., Gruber, N., Gregor, L., Ho, C.-H., Piao, S., Ciais, P., Schimel, D., and Kwon, E. Y.: Enhance seasonal amplitude of atmospheric CO_2 by the changing Southern Ocean carbon sink, Sci. Adv., 8, eabq0220, https://doi.org/10.1126/sciadv.abq0220, 2022.
67. 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.