Quantification of regional net CO$_2$ flux errors in the Orbiting Carbon Observatory-2 (OCO-2) v10 model intercomparison project (MIP) ensemble using airborne measurements

Jeongmin Yun[1], Junjie Liu[1], Brendan Byrne[1], Brad Weir[2,3], Lesley E. Ott[2], Kathryn McKain[4], Bianca C. Baier[4], Luciana V. Gatti[5], and Sebastien C. Biraud[6]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
[2]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[3]Goddard Earth Sciences Technology and Research II, Morgan State University, Baltimore, Maryland, USA
[4]NOAA Global Monitoring Laboratory, Boulder, Colorado, USA
[5]General Coordination of Earth Science (CGCT), National Institute for Space Research (INPE),
São José dos Campos, Brazil
[6]Lawrence Berkeley National Laboratory, Berkeley, California, USA

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

Abstract. Inverse model intercomparison 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 and often relies on the ensemble spread. This study proposes a method for quantifying the errors in regional net surface–atmosphere CO$_2$ flux estimates from models taken from the Orbiting Carbon Observatory-2 (OCO-2) v10 MIP by using independent airborne CO$_2$ measurements for the period 2015–2017. We first calculate the root mean square error (RMSE) between the ensemble mean of posterior CO$_2$ 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 from seven regions. Our analysis reveals that the flux errors projected onto CO$_2$ space account for 55 %–85 % of the regional average RMSE over the 3 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 a factor of 1.33–1.93, implying an underestimation of the actual flux errors, while their magnitudes are comparable in two regions. The adjoint sensitivity analysis identifies that the underestimation of flux errors is prominent where the magnitudes of fossil fuel emissions exceed those of terrestrial-biosphere fluxes by a factor of 3–31 over the 3 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 approach for estimating net surface–atmosphere CO_2 fluxes by assimilating observed atmospheric CO_2 concentrations. Most inverse-modeling approaches are based on 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 posteriori 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 a lower uncertainty than prior flux estimates, incorporating more observations causes the posterior fluxes to approach the true values (Liu et al., 2014).

However, concerns have been raised that 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 model 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 the 1990s, as well as subsequent projects, such as the Global Carbon Project (GCP; Ciais et al., 2022; Friedlingstein et al., 2023) 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 a posteriori solution across models. The OCO-2 MIP has revealed general agreement with regard to 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 regional fluxes are constrained by the current CO_2 observing network and for identifying regions with high uncertainty, allowing us to prioritize efforts to mitigate the error. The Bayesian formulation provides a method for calculating uncertainties in posterior fluxes based on uncertainties in prior fluxes and assimilated data. These uncertainties 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 underestimation 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 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 in the ensemble mean have generally been defined as the ensemble spread among ensemble posterior fluxes. However, 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 for quantifying 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 an adjoint sensitivity analysis. Our target ensemble results are derived from 10 ensemble members of the OCO-2 v10 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 (v10; O'Dell et al., 2018) and in situ CO_2 measurements (Tohjima et al., 2005; Nara et al., 2017; Schuldt et al., 2021a, b). This study uses more than $833\,000$ airborne CO_2 measurements collected at altitudes of 1–5 km above ground level (a.g.l.) from 20 different measurement projects (e.g., NOAA Carbon Cycle Group ObsPack Team, 2018; Baier et al., 2021; Miller et al., 2021; Schuldt et al., 2021a, b). These data have a broader spatial coverage and are less influenced by local sources compared to surface CO_2 data, thereby 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 measurements are most sensitive 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 in the ensemble net surface–atmosphere CO$_2$ fluxes generated by the OCO-2 v10 MIP using airborne observations. Here, the term "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 Fig. 1. First, we define two quantities: (1) the root mean square error (RMSE) between the ensemble mean of posterior CO$_2$ concentrations and observed CO$_2$ concentrations and (2) ERR$_{TOT}$ (Sect. 2.3). RMSE^2 represents the true error in the OCO-2 MIP ensemble mean of CO$_2$ concentrations, including the representation error (σ_r^2), observation error (σ_o^2), true flux error projected onto CO$_2$ concentrations ($\sigma_{f_t}^2$), transport error (σ_t^2), and error covariance between the preceding two terms (cov(σ_{f_t}, σ_t)). ERR$_{TOT}$ 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$_{OBS}^2$ denote the representation error (σ_r^2) and observation error (σ_o^2), respectively. ERR$_{MIP}^2$ represents the sum of the estimated flux error projected onto CO$_2$ space ($\sigma_{f_e}^2$) and the transport error (σ_t^2), along with the corresponding error covariance (cov(σ_{f_e}, σ_t)), computed from an ensemble spread of posterior CO$_2$ concentrations. Here, we separate representation errors from transport errors for computational purposes. The ratio between ERR$_{TOT}$ and the 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 error projected onto atmospheric CO$_2$ ($h(\text{err}_{f_e})$) through atmospheric transport simulations (Sect. 2.4). With $h(\text{err}_{f_e})$, ERR$_{TOT}$, and the RMSE, we derive the true error in the ensemble mean of posterior fluxes projected onto CO$_2$ space ($h(\text{err}_{f_t})$). Then, we identify the areas to which these airborne observations are most sensitive using an adjoint sensitivity analysis and calculate the estimated posterior flux error (err_{f_e}) over these regions. Assuming a linear observation operator, the study finally computes the true error in the ensemble mean posterior fluxes over the identified sensitive area (err_{f_t}) by applying the ratio between $h(\text{err}_{f_t})$ and $h(\text{err}_{f_e})$ to err_{f_e}.

2.1 The OCO-2 v10 MIP datasets

The OCO-2 v10 MIP provides multiple results from inverse models that assimilate different combinations of atmospheric CO$_2$ measurements for the period 2015–2020. Our study focuses on the results from the "LNLGIS" experiment, which assimilates the most observations except with respect to the OCO-2 ocean glint XCO$_2$ retrievals, which cause significant biases in inversion results (Byrne et al., 2023). The LNLGIS experiment incorporates OCO-2 v10 land nadir (LN) and land glint (LG) XCO$_2$ retrievals, along with global in situ (IS) data (including sur-

face, ship-based, and airborne measurements) taken from obspack_co2_1_OCO2MIP_v3.2.1_2021-09-14. A total of 10 different inverse-modeling groups provided monthly posterior surface CO$_2$ flux estimates interpolated to a $1° \times 1°$ horizontal resolution and co-sampled posterior atmospheric CO$_2$ data corresponding to the times and locations of all types of observations. All of the inversion groups used the same fossil fuel emission estimates, based on the Open-source Data Inventory for Anthropogenic CO$_2$ (ODIAC) dataset (Basu and Nassar, 2021), but they independently chose their transport models, assimilation techniques, and prior flux estimates. Further details are provided in Table S1 in the Supplement, 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 from the first 3 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 taken between 1 and 5 km a.g.l. 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 $1° \times 1°$ grid cell. The dataset includes data from two airborne measurement campaigns conducted over the ocean – the Atmospheric Tomography Mission (ATom; Thompson et al., 2022) and the O$_2$/N$_2$ Ratio and CO$_2$ Airborne Southern Ocean (ORCAS) study (Stephens et al., 2018) – as well as 18 campaigns conducted over land. Specific airborne campaigns and their references are detailed in Table 1. The majority of the datasets used in the study are from North America, accounting for 37 % of the total observations for the period 2015–2017, followed by East Asia (35 %) and Alaska (7 %). The duration and extent of the airborne observations vary across different regions and time periods. Figure 2b illustrates the number of $1° \times 1°$ 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 (CON-TRAIL; Machida et al., 2008) project provides sparse observations for Europe and continuous observations for East and Southeast Asia for the period 2015 to 2017, as well as for Australia for 2015–2016. In South America, measurements were conducted at six different sites in 2017, with the major-

1) Evaluation of posterior flux error estimates over the globe

$$
\frac{\overbrace{\sigma_o^2}\quad\overbrace{\sigma_r^2}\quad\overbrace{\sigma_{f_e}^2 + cov(\sigma_{f_e}, \sigma_t) + \sigma_t^2}}{(ERR_{OBS}^2 + ERR_{REP}^2 + ERR_{MIP}^2)_{(ppm)}}{RMSE^2\,_{(ppm)}} = \frac{ERR_{TOT\,(ppm)}^2}{RMSE^2\,_{(ppm)}} = \mathbf{Ratio^2}
\begin{cases} > 1 : overestimated\ flux\ error \\ < 1 : underestimated\ flux\ error \end{cases}
$$

$$\underbrace{\phantom{(ERR_{OBS}^2 + ERR_{REP}^2}}_{\sigma_o^2 + \sigma_r^2 + \sigma_{f_t}^2 + cov(\sigma_{f_t}, \sigma_t) + \sigma_t^2}$$

2) Quantification of true flux errors by region

$$
h\big(err_{f_t}\big)^2 \overset{\sigma_{f_t}^2}{} - h\big(err_{f_e}\big)^2 \overset{\sigma_{f_e}^2}{} \approx \overset{\sigma_o^2 + \sigma_r^2 + \sigma_{f_t}^2 + cov(\sigma_{f_t}, \sigma_t) + \sigma_t^2}{RMSE^2} - \underset{\sigma_o^2 + \sigma_r^2 + \sigma_{f_e}^2 + cov(\sigma_{f_e}, \sigma_t) + \sigma_t^2}{ERR_{TOT}^2} \quad \Longrightarrow \quad \frac{h\big(err_{f_e}\big)_{(ppm)}}{h\big(err_{f_t}\big)_{(ppm)}} = \frac{err_{f_e\ (gC\,m^{-2}\,day^{-1})}}{err_{f_t\ (gC\,m^{-2}\,day^{-1})}}
$$

$h\big(err_{f_e}\big)$ =

1. Simulate atmospheric CO₂ fields using forward modeling by prescribing posterior CO₂ fluxes for each ensemble member.

2. Extract CO₂ concentration values corresponding to observation locations and times and calculate their standard deviation.

err_{f_e} =

1. Calculate the adjoint sensitivity of atmospheric CO₂ to surface CO₂ fluxes by region.
2. Compute the ensemble spread of the sum of posterior CO₂ fluxes within the 50th percentile adjoint sensitivity trajectory.

Figure 1. Flow chart summarizing the process of evaluating and quantifying errors in the ensemble mean of regional posterior fluxes. $RMSE^2$ represents the mean square error between the ensemble mean of posterior CO₂ concentrations and observed CO₂ concentrations. ERR_{REP}^2 and ERR_{OBS}^2 denote estimates of representation errors and observation errors, respectively. ERR_{MIP}^2 represents the ensemble spread of posterior CO₂ concentrations. ERR_{TOT}^2 is defined as the sum of ERR_{REP}^2, ERR_{OBS}^2, and ERR_{MIP}^2, while err_{f_e} and err_{f_t} represent estimates of flux errors, defined as the ensemble spread of posterior fluxes, and their true values, respectively. Moreover, $h(err_{f_e})$ and $h(err_{f_t})$ denote estimates of flux errors projected onto CO₂ concentrations and their true values, respectively, while σ_o^2, σ_r^2, $\sigma_{f_t}^2$ ($\sigma_{f_e}^2$), σ_t^2, and $cov(\sigma_{f_t}, \sigma_t)(cov(\sigma_{f_e}, \sigma_t))$ TS1 refer to the types of errors represented by the error statistics – namely, observation errors, representation errors, true (estimated) flux errors projected onto CO₂ concentrations, transport errors, and error covariances between the preceding two terms, respectively.

ity of these observations coming from five flask measurement sites provided by the National Institute for Space Research (INPE), which likely exhibit a low bias in measured flask sample CO₂ mole fractions (~ 1 ppm or greater) when ambient water vapor mole fractions are above $\sim 1.5\,\%$. These biases have been noted in some aircraft flask CO₂ measurements in the 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 limitations 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 exhibit lower reliability compared to those from other regions.

2.3 Evaluation of the ensemble posterior CO₂ fluxes

We first employ the two matrixes defined in Eqs. (1) and (2) to evaluate the ensemble posterior net flux errors proposed by Liu et al. (2021). One matrix is the RMSE between the ensemble mean of posterior atmospheric CO₂ from OCO-2 MIP models and the atmospheric CO₂ from airborne mea-

Table 1. Data descriptions for the airborne measurement campaigns.

Site code	Site name	Measurement campaign name	Measurement type	Data provider	ObsPack (original) dataset identifier	Reference
ACG	Alaska Coast Guard, Alaska, USA	National Oceanic and Atmospheric Administration (NOAA) Global Monitoring Laboratory (GML) Aircraft Program	In situ	NOAA GML	https://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	https://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)
AirCoreNOAA	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)	https://doi.org/10.25925/20181030[c] https://doi.org/10.1594/PANGAEA.926834	Gatti et al. (2023)
CAR	Briggsdale, Colorado, USA		Flask	NOAA GML	https://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)	https://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, USA	Arctic–Boreal Vulnerability Experiment (ABoVE)	In situ	NOAA GML	https://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)	https://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)	https://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	https://doi.org/10.25925/20201204[a] https://doi.org/10.5067/ASDC/SUBORBITAL/KORUSAQ_TraceGas_AircraftInSitu_DC8_Data_1	Vay et al. (2009)

Site code	Site name	Measurement campaign name	Measurement type	Data provider	ObsPack (original) dataset identifier	Reference
MAN	Manaus, Brazil	NOAA GML Aircraft Program	In situ	NOAA GML	https://doi.org/10.25925/20210519[e]	
ORC	O$_2$/N$_2$ Ratio and CO$_2$ Airborne Southern Ocean (ORCAS) study		In situ	National Center for Atmospheric Research (NCAR)	https://doi.org/10.25925/20181030[c] https://doi.org/10.5065/D6SB445X	Stephens et al. (2018)
PAN	Pantanal, Mato Grosso do Sul, Brazil		Flask	INPE	https://doi.org/10.25925/20181030[c] https://doi.org/10.25925/20181030[c]	
PFA	Poker Flat, Alaska, USA	NOAA GML Aircraft Program	Flask	NOAA GML	https://doi.org/10.25925/20210517[d]	Sweeney et al. (2015)
RBA-B	Rio Branco, Brazil		Flask	INPE	https://doi.org/10.25925/20181030[c] https://doi.org/10.1594/PANGAEA.926834	Gatti et al. (2023)
SAN	Santarém, Brazil		Flask	INPE	https://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	https://doi.org/10.25925/20210517[d] https://doi.org/10.1594/PANGAEA.926834	Biraud et al. (2013), Sweeney et al. (2015)
SONGNEX 2015	The 2015 Shale Oil and Natural Gas Nexus air campaign, USA	The 2015 Shale Oil and Natural Gas Nexus air campaign	In situ	NOAA Chemical Sciences Laboratory (CSL)	https://doi.org/10.25925/20201204[a]	
TEF	Tefé, Brazil		Flask	INPE	https://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	https://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).
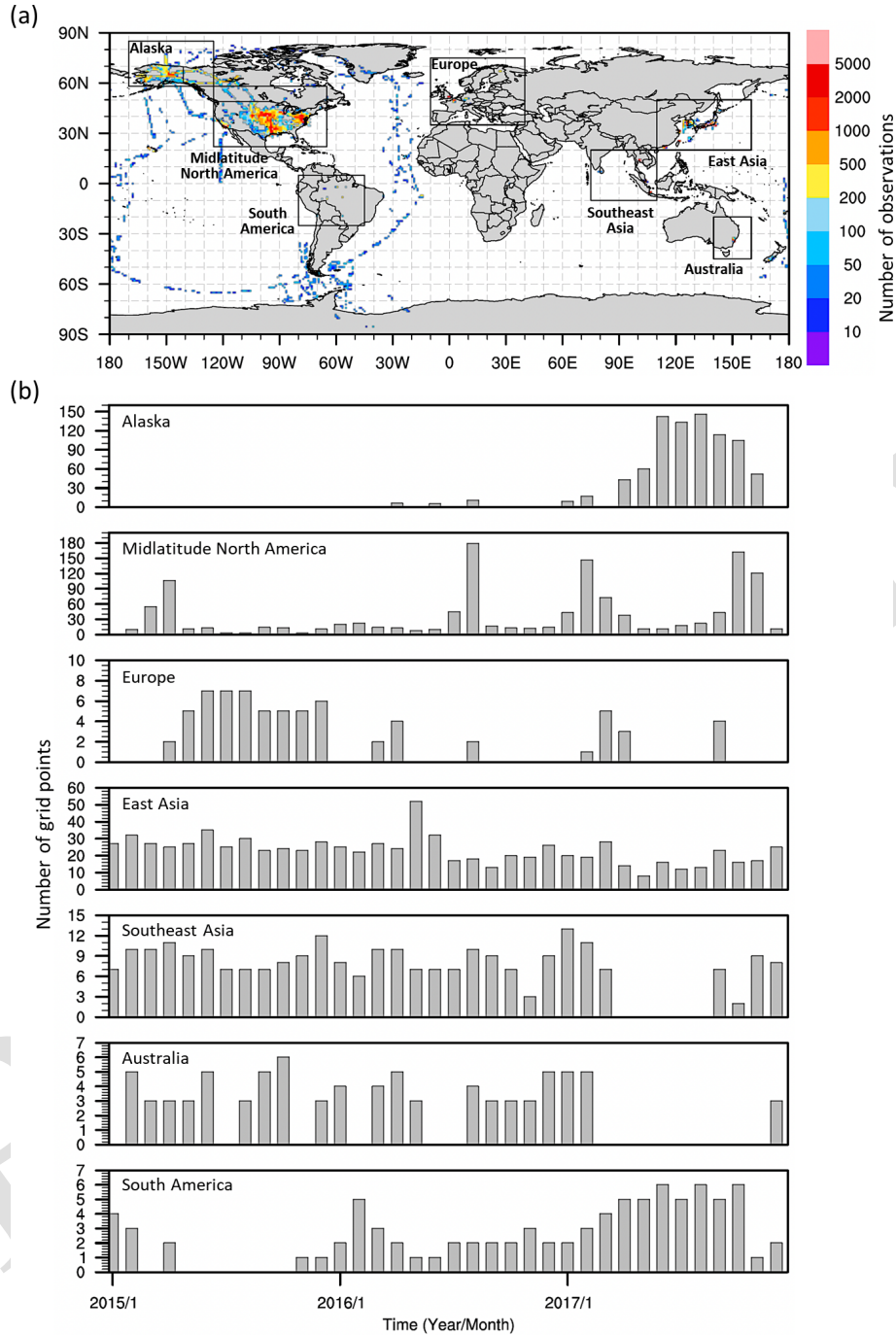
Figure 2. (a) The total number of airborne measurement data points used in this study at each $1° \times 1°$ grid point and **(b)** the number of $1° \times 1°$ grid points with more than 10 data points available within each region and for each month during the period 2015–2017.

surements, which can be written as

$$\text{RMSE}^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\overline{h_i\left(\hat{x}\right)} - y_{\text{o},i}\right]\left[\overline{h_i\left(\hat{x}\right)} - y_{\text{o},i}\right]^T,$$

$$\text{where } \overline{h_i\left(\hat{x}\right)} = \frac{1}{M}\sum_{j=1}^{M}h_{i,j}\left(\hat{x}_j\right). \quad (1)$$

Here, $\overline{h_i(\hat{x})}$ is the ensemble mean of posterior atmospheric CO$_2$ sampled at the time and location of the ith airborne observation, $y_{\text{o},i}$, within each $1° \times 1°$ grid cell for each month. N is the monthly total number of sampled data points in each grid cell. M is the number of ensemble members (i.e., 10). A single monthly RMSE value is computed using N measurement data points for each grid cell. The total number

of RMSE values computed per month within each region corresponds to the number of grid cells shown in Fig. 2b. The RMSE indicates the magnitude of the actual CO_2 errors in the ensemble estimates, which is also a metric broadly used to evaluate the accuracy of posterior fluxes (Crowell et al., 2019; Peiro et al., 2022; Byrne et al., 2023). As illustrated in Fig. 1 and described in Appendix A (Eq. A3), $RMSE^2$ includes not only the projection of the true flux error onto CO_2 concentrations ($\sigma_{f_t}^2$) but also the transport error (σ_t^2), the error covariance between these two terms ($cov(\sigma_{f_t}, \sigma_t)$), the representation error (σ_r^2), and the airborne observation error (σ_o^2). Both transport errors and representation errors stem from transport models. Transport errors include errors in model structures and meteorological fields, while representation errors arise from a mismatch in resolution between model simulations and observations.

In practice, true flux errors are often approximated by the spread of ensemble fluxes, so the sum of projection of true flux errors to CO_2 concentrations and transport errors is approximated by the ensemble spread of simulated CO_2 concentrations in the 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 (Fig. 1). Unlike the RMSE, the variance terms of the flux error ($\sigma_{f_e}^2$) and transport error (σ_t^2), as well as the covariance term between these errors ($cov(\sigma_{f_e}, \sigma_t)$), are replaced by the spread of the ensemble (i.e., variance) of posterior atmospheric CO_2 concentrations (ERR_{MIP}^2), defined as

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

Unlike in Liu et al. (2021), which used only one transport model, ERR_{MIP}^2 accounts for transport errors because posterior atmospheric CO_2 was generated by multiple types of transport models in the OCO-2 MIP, driven by different meteorology fields. Thus, the ERR_{MIP}^2 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 $2° \times 2.5°$.

To obtain representation errors and observation errors not captured by ERR_{MIP}^2, we additionally calculate ERR_{REP}^2 and ERR_{obs}^2, respectively. ERR_{REP}^2 denotes the representation error (σ_r^2) in $RMSE^2$, as shown in Fig. 1. It is defined as the spatial variability in atmospheric CO_2 within a $2° \times 2.5°$ grid cell and is written as

$$ERR_{REP}^2 = \frac{1}{N} \sum_{i=1}^{N} VAR_{CO_2, i}. \tag{3}$$

Using high-resolution ($0.5° \times 0.625°$) 3-hourly GEOS-5 simulation results for 2018 from the NASA Goddard Space Flight Center (Weir et al., 2021), we calculate the variance of atmospheric CO_2 concentrations within each $2° \times 2.5°$ grid cell at each 3 h interval. Then, we sample the CO_2 variance ($VAR_{CO_2, i}$) value corresponding to the grid cell containing the ith observation and at the time closest to that of the observation. Subsequently, the monthly mean value of the N co-sampled variances is derived (ERR_{REP}^2). We assume that the variances do not vary significantly across years, given the lower monthly variability in ERR_{REP} compared to that in the RMSE and ERR_{MIP} (as will be shown in Sect. 3.2). The reason for calculating the CO_2 variance within a $2° \times 2.5°$ grid is that this grid represents 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 in CO_2 concentrations by comparing the simulated CO_2 variance with the spatial variance of airborne measurement data from the ACT-America project (see the Supplement and Fig. S1). The evaluation results support our approach.

ERR_{OBS}^2 represents the observation error (σ_o^2) in $RMSE^2$, as shown in Fig. 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 for the Northern Hemisphere and 0.05 ppm for the Southern Hemisphere. Assuming an ideal situation without any systematic bias, we set the observation error (ERR_{OBS}) for all airborne observations to 0.1 ppm. However, in reality, systematic errors could be present in airborne observations, stemming from instrument or setup biases, calibration offsets, and other factors. In particular, CO_2 measurements from INPE taken in South America might exhibit higher measurement errors compared to those taken in other regions due to unresolved water vapor contamination issues in the flask measurements, which may result in both a low bias ($\sim$ 1–3 ppm at 3 % absolute humidity) and spurious variability (Baier et al., 2020). The potential effects of these systematic errors on our findings will be addressed in Sect. 4. This study employs only ERR_{OBS}^2 for calculating ERR_{TOT}^2 and does not compare it with other error quantities from Sect. 3.

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

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

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

To evaluate whether the spread of ensemble CO_2 fluxes from the OCO-2 MIP represents the true flux errors in the ensemble mean, we calculate the ratio between monthly

ERR_{TOT} and RMSE values as follows:

$$Ratio^2 = \frac{ERR_{TOT}^2}{RMSE^2}. \qquad (5)$$

Given that ERR_{REP}^2 reasonably depicts actual representation errors, $Ratio^2$ can indicate whether posterior flux and transport errors computed from the ensemble spread overestimate or underestimate the true flux and transport errors. In this study, we assume that the estimated transport errors from the ensemble spread among the transport models used in the OCO-2 MIP represent the true transport errors and that the difference between $RMSE^2$ and ERR_{TOT}^2 mainly arises from the difference in the flux error variances ($\sigma_{f_t}^2$ and $\sigma_{f_e}^2$). 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 the transport errors may be a strong one given that they 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 Sect. 4.

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

In addition to qualitative evaluations of posterior flux errors using the ratios between ERR_{TOT} and the RMSE, we propose a method for quantitatively assessing the ensemble posterior flux errors (i.e., the 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 solely to the ensemble spread of posterior fluxes from the OCO-2 MIP ($h(err_{f_e})^2$). As shown in Appendix A, this term can be written as

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

All transport models engaged in the OCO-2 MIP would ideally be used to derive $h(err_{f_e})^2$. However, in this study, we approximate this error term using the GEOS-Chem model, as depicted in the following:

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

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

To obtain $h_{GC}(err_{f_e})^2$, we conduct a set of forward simulations using the GEOS-Chem transport model (within the GEOS-Chem adjoint model (v35j); Henze et al., 2007).

In all 10 experiments, consistent meteorology and emission forcing data from version 2 of the Modern-Era Retrospective analysis for Research and Applications (MERRA-2; Gelaro et al., 2017) and the Open-source Data Inventory for Anthropogenic CO_2 (ODIAC; Oda and Maksyutov, 2015) are used. 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 posterior fluxes from the respective 10 OCO-2 MIP ensemble members. All experiments are performed at a $2° \times 2.5°$ horizontal resolution and with 47 vertical levels for the period 2015–2017. By calculating the mean of the variances of simulated CO_2 concentrations among the 10 experiments for the ith airborne observation within each $1° \times 1°$ grid cell, we derive $h_{GC}(err_{f_e})^2$.

Because we assume that the spread of ensemble transport models used in the OCO-2 MIP represents the true transport errors included in $RMSE^2$, the transport errors, along with the observation errors and representation errors, cancel out when we calculate the difference between monthly $RMSE^2$ and ERR_{TOT}^2 values. Consequently, the difference between monthly $RMSE^2$ and ERR_{TOT}^2 values arises from the differences in the flux error variances ($\sigma_{f_t}^2$ and $\sigma_{f_e}^2$). The difference between the monthly true flux error ($h(err_{f_t})^2$) and the estimated flux error ($h(err_{f_e})^2 \approx h_{GC}(err_{f_e})^2$) projected onto CO_2 space can be derived from the difference between $RMSE^2$ and ERR_{TOT}^2 as follows:

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

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

In order to link these 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 (as defined in Fig. 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 flux 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 the simulated CO_2 concentrations at the locations of airborne observations within each region and for each month:

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

The sensitivity of observations to surface fluxes at the lth grid cell and tth time is derived from the partial derivative of J with respect to surface fluxes ($\hat{x}_{l,k}$), expressed as

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

The 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 identify the areas most sensitive 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 uniformly influenced across excessively wide regions as a result of active atmospheric mixing. Additionally, to avoid the excessive consideration of localized effects caused by 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 error in flux space ($\mathrm{err}_{f_e}^2 = \sigma_{f_e}^2$) by calculating the ensemble spread of the total posterior flux values (and area-averaged mean values) over the effective area for each month during the period 2015–2017, as illustrated in Fig. 1. The estimated mean posterior flux error (err_{f_e}) over the selected areas for each month exhibits a significant correlation ($p \le 0.05$) with monthly $h(\mathrm{err}_{f_e})$ values across all regions, except for Australia, where the observational campaign was conducted during 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_{f_e} between ocean and land within the effective areas reveals that, on average, 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 in posterior net land CO_2 fluxes within the selected areas in flux space.

This study provides both monthly and 3-year mean values of regional flux error statistics for the period 2015–2017. Technically, it is possible to derive monthly true errors in the ensemble mean of net land CO_2 fluxes using monthly error statistics. However, to obtain more robust results, we compute the true errors in 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 effective areas for monthly airborne observations. These grid cells, at a $2° \times 2.5°$ 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 3 years. A higher number of months indicates that more information from those grid cells is utilized in calculating the 3-year regional mean error statistics. We define our 3-year mean error statistics as mostly representing the areas where the composite values exceed 8, corresponding to 20 % of the total months analyzed (i.e., 36 months).

The observation operator, which converts surface CO_2 fluxes to atmospheric CO_2, is generally assumed to be linear. Therefore, in these areas, we can obtain the true error in the ensemble annual total net land fluxes, $\mathrm{err}_{f_t} (= \sigma_{f_t})$, by multiplying the ratio between the 3-year mean values of $h(\mathrm{err}_{f_t})$ and $h(\mathrm{err}_{f_e})$ 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

$$\mathrm{err}_{f_t} = \frac{h(\mathrm{err}_{f_t})}{h(\mathrm{err}_{f_e})} \times \mathrm{err}_{f_e}. \tag{12}$$

One thing readers should keep in mind is that 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 the characteristics of regions where the average annual total err_{f_t} value is significantly underestimated, we compute the ensemble mean of the average annual posterior terrestrial-biosphere CO_2 fluxes and the fossil fuel CO_2 emissions (from the ODIAC dataset) for the effective area.

3 Results

3.1 Spatiotemporal variations in 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 that of 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, conducted during austral summer, covered wide latitudinal ranges, primarily over the oceans, providing a unique opportunity for analyzing the latitudinal distributions of inverse-modeling errors and the contributions of major error sources. We compare the ensemble posterior CO_2 to airborne CO_2 measurements taken between 1 and 5 km a.g.l. and then calculate the mean error statistics for the entire campaign period. Comparisons with observations from the ATom and ORCAS campaigns reveal a general increase in RMSE

values toward the northern high latitudes, reaching 1.2 ppm at 40° N (Fig. 3a and 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 the 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.

Both ERR_{MIP} and ERR_{REP} exhibit spatial distributions similar to that of the RMSE (Fig. 3a–c and f). However, ERR_{MIP} has a stronger positive correlation with the RMSE ($r = 0.57$ and 0.58 for the ATom and ORCAS campaigns, respectively) than ERR_{REP} ($r = 0.35$ and 0.32), with an average magnitude (0.49 and 0.32 ppm) greater than that of ERR_{REP} (0.27 and 0.20 ppm) globally across both campaign periods. In particular, ERR_{MIP} (ERR_{REP}) accounts for 75 % (37 %) of the anomalously high RMSE values (1.5 ppm) in North America (32–50° N and 85–124° W) and 75 % (30 %) of the RMSE values (1.2 ppm) along the west coast of Africa. These findings indicate that ERR_{MIP}, representing errors in posterior fluxes and transport, is the most significant factor in explaining the RMSE.

Next, in order to assess the proximity of the estimated posterior flux errors, based on the spread of the OCO-2 MIP ensemble fluxes, to the true posterior flux errors in the ensemble mean, we compare the RMSE with the sum of ERR_{MIP}, ERR_{REP}, and ERR_{OBS} (referred to as ERR_{TOT}). The ratio of ERR_{TOT} to the RMSE exceeds 1 over the tropical Pacific and the Southern Ocean (Fig. 3d and e), indicating that the ensemble spread of posterior fluxes overestimates true flux errors over 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 for 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 that the true posterior ocean flux errors may be smaller than the spread of the ensemble posterior flux estimates. In contrast, a ratio between ERR_{TOT} and the RMSE of less than 1 was observed along the African coast during the ATom campaigns, with the exception of the 2018 spring campaign, which was conducted in a region relatively far 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 the true errors in 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), who show that most inverse models assimilating OCO-2 XCO_2 retrievals tend to overestimate the net carbon sources in this region due to potential positive biases in the OCO-2 retrievals.

In the northern mid-to-high latitudes, characterized by significant land CO_2 flux impacts on atmospheric CO_2 varia-

tions (Yun et al., 2022), the ratio between ERR_{TOT} and the RMSE exhibits substantial variation across space and time. The ratio between ERR_{TOT} and the RMSE is greater than 1 within the North American continent during summer and fall. However, in other areas, there is a mixed pattern involving ratios both below and above 1, although the majority of areas exhibit ratios less than 1 during winter. These findings highlight that the degree of underestimation or overestimation of true flux errors, based on the ensemble spread, can vary depending on the region and season, emphasizing the need for a more detailed evaluation of flux errors at the regional level using long-term independent observations.

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

In this section, we calculate regionally averaged monthly error statistics by comparing the ensemble posterior CO_2 to airborne measurements taken over seven regions during 2015–2017. RMSE values across all these regions exhibit significant monthly variations, with values falling within the range of 1–3 ppm and no clear seasonality, possibly due to variations in observation routes (Fig. 4). Consistent with the results shown in Sect. 3.1, ERR_{MIP} is the most significant factor explaining the variations in the RMSE. Among the seven regions, significant positive correlations ($p < 0.05$) between monthly RMSE and ERR_{MIP} values are observed in Alaska ($r = 0.46$), midlatitude 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 concerning ERR_{REP}. This suggests that, in these regions, temporal variations in the errors in posterior fluxes and transport are the major contributors to temporal variations in the RMSE. On the other hand, the 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 the ensemble spread may not represent the temporal variations in true flux errors in these regions.

RMSE values exhibit significant variability not only over time but also across regions. The 3-year average RMSE is highest in East Asia (mean: 1.98 ppm; 95 % confidence intervals: [1.90, 2.06] ppm), followed by Europe (1.57, [1.41, 1.74] ppm), and lowest in Australia (0.88, [0.79, 0.97] ppm), with Alaska having a slightly higher RMSE (1.19, [1.12, 1.25] ppm). ERR_{MIP} is the primary error component of the RMSE, accounting for 58 %–83 % of the RMSE and surpassing the ERR_{REP} values for all the regions by a factor of 1.2–2.1. In East Asia, the difference between ERR_{MIP} and ERR_{REP} is relatively small compared to that observed in other regions. This could be attributed to the presence of numerous significant carbon sources, particularly along coastal areas, resulting in increased spatial variability in CO_2 within the coarse grid cells of the OCO-2 MIP inverse modeling.
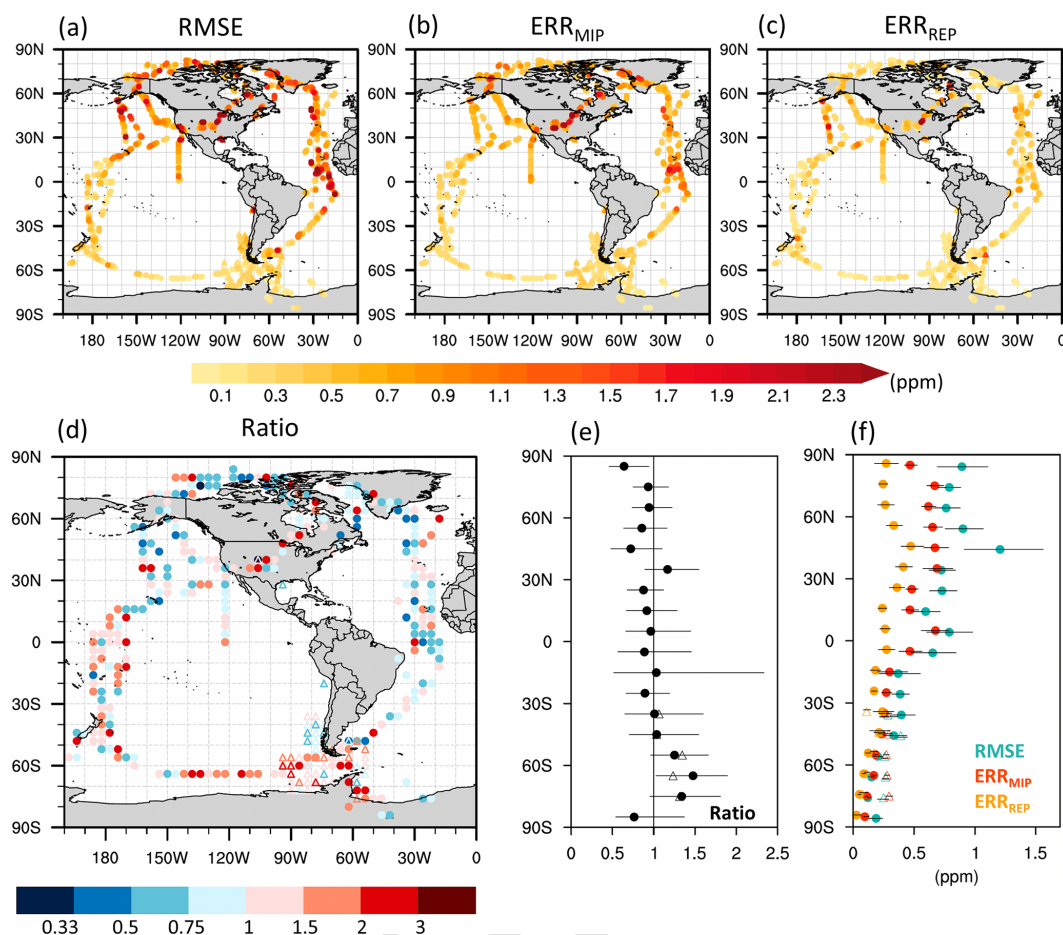
Figure 3. Spatial distributions of **(a)** RMSE, **(b)** ERR$_{MIP}$, **(c)** ERR$_{REP}$, and **(d)** Ratio $\left(\sqrt{(ERR_{OBS})^2 + ERR_{REP}^2 + ERR_{MIP}^2 / RMSE}\right)$ values with respect to where ATom airborne measurements (circles) and ORCAS airborne measurements (triangles) were taken and **(e, f)** their latitudinal distributions, smoothed by a 10° moving average with 95 % confidence intervals derived from 1000 bootstrap samples of datasets (error bars).

The ratio between ERR$_{TOT}$ and the RMSE also shows significant variability across regions. Our results indicate that, on average, the estimated flux errors for 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 underestimations at the 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 variability (i.e., standard deviation) of the ratios is 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 3-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 OCO-2 v10 MIP ensemble posterior net CO$_2$ fluxes by region

Next, by incorporating the monthly RMSE, ERR$_{TOT}$, and $h\left(err_{f_e}\right)$ values, we derive the monthly true posterior flux error in CO$_2$ space (i.e., $h\left(err_{f_t}\right)$) for each region with respect to the period 2015–2017 (Fig. 5). Regionally averaged $h\left(err_{f_t}\right)$ values exhibit different seasonal and monthly variability compared to $h\left(err_{f_e}\right)$ values. In the northern midlatitude regions, $h\left(err_{f_t}\right)$ shows clear seasonal cycles for the entire analysis period, despite there being different observation routes for each month. For example, in midlatitude North America and East Asia, the growing season (May to October) exhibits higher $h\left(err_{f_e}\right)$ values (0.6 and 0.9 ppm, respectively) than the non-growing season (November to April; 0.4 and 0.7 ppm, respectively). Seasonal variations in $h\left(err_{f_t}\right)$ are also observed in East Asia and partially observed in midlatitude North America with respect to 2017, but they are not discernible in Alaska and Europe. In addi-

tion, monthly $h\left(\mathrm{err_{f_t}}\right)$ values do not exhibit a significant correlation ($p < 0.05$) with monthly $h\left(\mathrm{err_{f_e}}\right)$ values in Alaska, Southeast Asia, and South America. Moreover, $h\left(\mathrm{err_{f_t}}\right)$ displays greater monthly variability than $h\left(\mathrm{err_{f_e}}\right)$. For example, in midlatitude North America and East Asia, the standard deviation of monthly $h\left(\mathrm{err_{f_t}}\right)$ values is 1.8 and 2.3 times greater than that of monthly $h\left(\mathrm{err_{f_e}}\right)$ values respectively.

The comparison between the 3-year average $h\left(\mathrm{err_{f_t}}\right)$ value and the 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 $h\left(\mathrm{err_{f_t}}\right)$ value tends to be larger in regions with a higher RMSE, peaking in East Asia ($h\left(\mathrm{err_{f_t}}\right) = 1.32$ ppm; RMSE = 1.98 ppm) and reaching a minimum in Australia ($h\left(\mathrm{err_{f_t}}\right) = 0.75$ ppm; RMSE = 0.88 ppm) (Figs. 4h and 5h). The $h\left(\mathrm{err_{f_t}}\right)$ value accounts for up to 85 % of the RMSE in Australia, followed by Southeast Asia (80 %), and accounts for a minimum of 60 % of the RMSE in South America, with the contribution in midlatitude North America being slightly higher (64 %). This indicates the dominant contributions of posterior flux errors to the RMSE, surpassing the contributions of representation and transport errors in the first two regions.

Throughout the analysis period, the regional mean ratios between $h\left(\mathrm{err_{f_e}}\right)$ and $h\left(\mathrm{err_{f_t}}\right)$ indicate significant underestimations, at a 95 % confidence level, of true posterior flux errors in midlatitude North America, Europe, East Asia, Southeast Asia, and Australia by factors 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 (Fig. 5h). In contrast, Alaska and South America exhibit comparable estimates of true flux errors, with factors of 0.96 [0.76, 1.17] and 0.97 [0.49, 1.54], respectively. The regions with significant underestimations align with those identified in the previous analysis based on the ratios between $\mathrm{ERR_{TOT}}$ and the RMSE (Sect. 3.2), but the ratios between $h\left(\mathrm{err_{f_e}}\right)$ and $h\left(\mathrm{err_{f_t}}\right)$ imply a stronger underestimation of the true flux errors. The ratios have larger uncertainty ranges in regions where observations were conducted over limited durations and locations, such as Europe, Australia, and South America, than in midlatitude North America and East Asia, where observations cover wider areas and occur more frequently.

Finally, using the 3-year regional mean ratios between $h\left(\mathrm{err_{f_e}}\right)$ and $h\left(\mathrm{err_{f_t}}\right)$, we compute the true errors in the annual net land fluxes over the effective areas, averaged for the period 2015–2017 (Fig. 6). We find that the actual flux errors are underestimated, particularly in regions where annual fossil fuel CO_2 emissions exceed annual terrestrial-biosphere fluxes by a factor of 3–31. The airborne measurements carried out in midlatitude 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 correspond to values of 1341, 2443, and 815 Tg C yr^{-1}, respectively. The first two areas

are estimated as significant terrestrial-biosphere CO_2 sinks, with estimated fluxes of -414 ± 279 (ensemble mean $\pm 1\sigma$) and -561 ± 380 Tg C yr^{-1}, in contrast to Southeast Asia (26 ± 118 Tg C yr^{-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 yr^{-1}, respectively. Observations across Europe and Australia, conducted over limited periods and at specific locations, mainly represent certain areas in western Europe and the southeastern part of Australia, where fossil fuel emissions (234 and 53 Tg C yr^{-1}, respectively) are around 4 and 5 times greater than terrestrial-biosphere sinks (-51 ± 34 and -10 ± 67 Tg C yr^{-1}, respectively). The recalculated net land flux errors in these regions are also larger than the ensemble spread, with estimates of 65 and 114 Tg C yr^{-1}, respectively. In contrast, the most influential areas for observations in Alaska and South America, encompassing the southeastern region of Alaska and the northern part of Brazil, are characterized as terrestrial-biosphere sinks of -8 ± 11 Tg C yr^{-1} and sources of 625 ± 387 Tg C yr^{-1}, respectively. These values are comparable to, or more than 10 times greater than, those of fossil fuel emissions (10 and 38 Tg C yr^{-1}, respectively). 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 yr^{-1}, respectively.

4 Discussion and conclusions

Our results show that errors in posterior net land CO_2 fluxes are a major contributor to the RMSE between simulated posterior 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 becomes 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 $\mathrm{ERR_{MIP}}$ and $h\left(\mathrm{err_{f_e}}\right)$; Fig. S6), exceed those in true flux errors projected onto CO_2 space, indicating that regional differences in the RMSE do not directly correspond to differences in flux errors. For example, although the 3-year mean errors in representation and transport for East Asia exceed those for Southeast Asia by 0.5 and 0.3 ppm, respectively, the disparity in the mean true flux errors projected onto CO_2 space between the two regions is only 0.2 ppm. This result is supported by previous studies, highlighting that spatial distributions of simulated CO_2 concentrations can vary significantly depending on the transport model (Schuh et al., 2019) and its spatial resolution (Stanevich et al., 2020). Therefore,
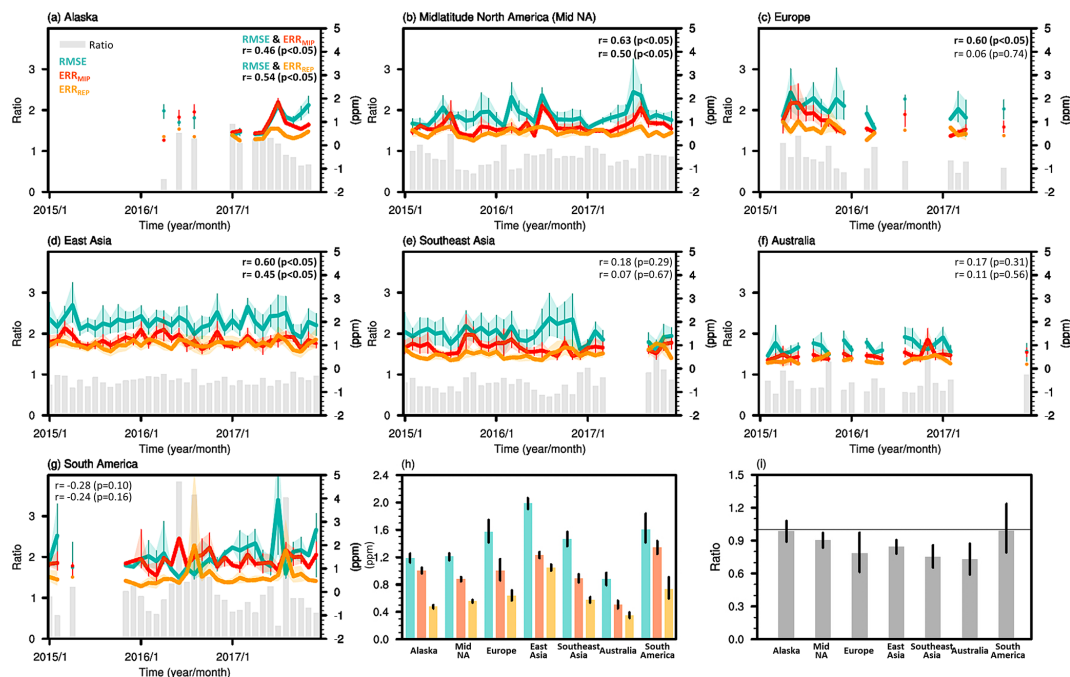
Figure 4. (a–g) Monthly RMSE, ERR_{MIP}, ERR_{REP}, and Ratio values for each region and **(h, i)** the corresponding mean values for the period 2015–2017. The upper-right numbers in panels **(a)**–**(g)** indicate the correlation coefficient between the RMSE and ERR_{MIP} and that between the RMSE and ERR_{REP}. The shaded areas and error bars represent 95 % confidence intervals derived from 1000 bootstrap samples of datasets.
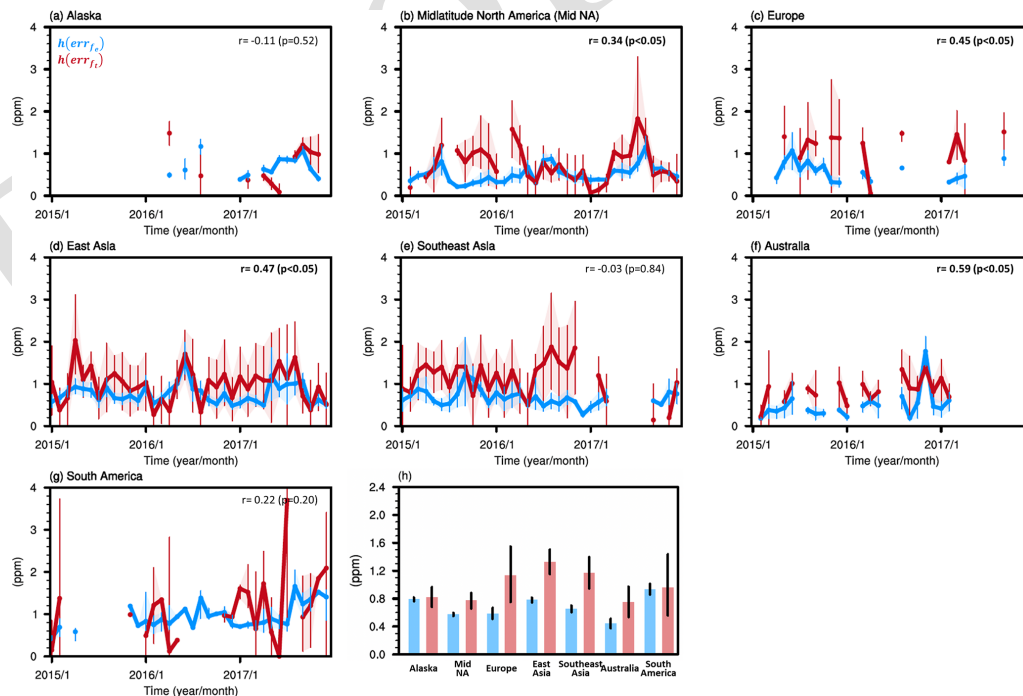
Figure 5. (a–g) Monthly values of $h\left(err_{f_e}\right)$ and $h\left(err_{f_t}\right)$ for each region and **(h)** the corresponding mean values for the period 2015–2017. In panels **(a–g)**, the upper-right corner displays a number indicating the correlation coefficient between $h\left(err_{f_e}\right)$ and $h\left(err_{f_t}\right)$. The shaded areas and error bars represent 95 % confidence intervals derived from 1000 bootstrap samples of datasets.
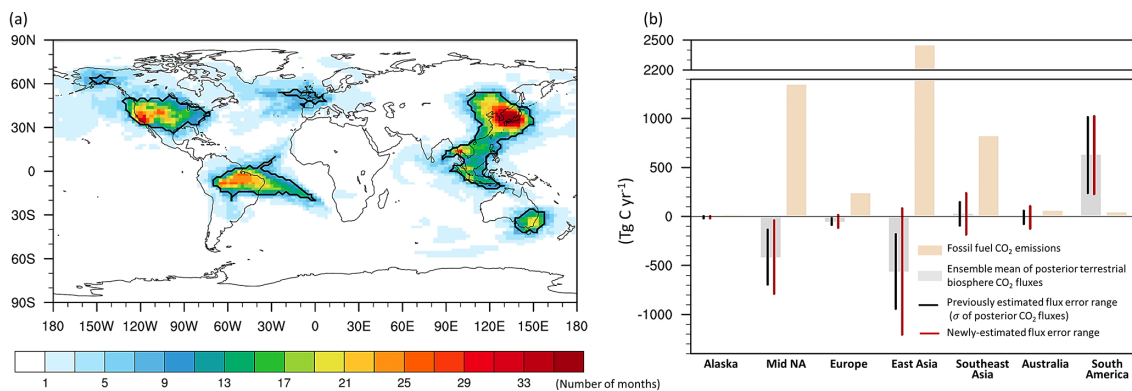
Figure 6. (a) Number of months selected as the effective areas for airborne measurements. The outlined areas represent selected areas for 8 months or more. **(b)** Annual total terrestrial-biosphere CO$_2$ fluxes obtained from the ensemble mean of 10 OCO-2 MIP models and annual total fossil fuel CO$_2$ emissions estimated from ODIAC data for each outlined area, averaged over the period 2015–2017. The black error bars each represent ± 1 standard deviation of the posterior net land fluxes, identical to the standard deviation of the posterior terrestrial-biosphere fluxes. The red error bars indicate newly estimated ranges of errors in the posterior net land fluxes from this study.

when utilizing airborne CO$_2$ measurements (and potentially other CO$_2$ observations) 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 true errors in the ensemble mean of the posterior net CO$_2$ flux estimates are significantly greater than the ensemble spread of the flux estimates in five out of the seven regions where fossil fuel emissions are higher than terrestrial-biosphere fluxes. A possible explanation for this result is the presence of errors in the prescribed fossil fuel emissions common to all OCO-2 MIP models. The 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, one that is often applied in conventional global atmospheric inverse models as errors in fossil fuel emission estimates are considered to be 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 distributions at model grid scales and temporal variability over a year (Zhang et al., 2016; Gurney et al., 2021). Oda et al. (2023) showed the significant impacts of differences in fossil fuel emission estimates on posterior terrestrial-biosphere flux estimates near the source regions. The OCO-2 MIP models used identical fossil fuel emission estimates, and thus their posterior net flux estimates share common biases, induced by errors in the fossil fuel emission estimates. Because these systematic biases are not captured by the ensemble spread of the flux estimates, the true flux errors exceed the errors computed from the ensemble spread

in the main source regions. In addition, the regional and seasonal sampling biases of CO$_2$ measurements and satellite retrieval errors may contribute to these systematic biases (Kulawik et al., 2019). Using eight prior flux datasets may also not adequately represent the errors in the terrestrial-biosphere fluxes, which exhibit significant variations across the estimates (Feng et al., 2019). Therefore, further studies are needed to uncover the causes of underestimation in true flux errors and thus to understand the uncertainty sources overlooked in current ensemble inverse-modeling estimates.

The reliability of our observation-based regional flux error estimates depends on the availability of airborne measurement data. Although our approach is generally effective for estimating regional means of monthly $h\left(\mathrm{err}_{f_t}\right)$ values, it is not applicable in [TS2] 15 % of our cases (shown in Fig. 5), where measurements were mostly taken in local areas covering one to six $1° \times 1°$ 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 analyses, especially when using observations made locally. Extending the calculation period to several months or longer (e.g., Fig. 5h) is a suitable strategy for mitigating the impact of outliers and obtaining more robust results. In fact, the ratios between the 3-year mean values of $h\left(\mathrm{err}_{f_e}\right)$ and $h\left(\mathrm{err}_{f_t}\right)$, which serve as key metrics for quantifying regional flux errors (Fig. 5h), exhibit smaller uncertainties over midlatitude North America and East Asia, where consistent airborne data covering broad areas are available, than over Europe and South America, where airborne observations are sparse and data coverage is intermittent. In addition, it is noteworthy that the $h\left(\mathrm{err}_{f_e}\right)$-to-$h\left(\mathrm{err}_{f_t}\right)$ ratios derived from continuous observations enable the computation of unbiased true errors in the ensemble mean of annual posterior net

fluxes, averaged over the analysis period, compared to those derived from limited observation periods (e.g., for Alaska). These results highlight the importance of frequent airborne measurements with extensive spatial coverage for reliably quantifying errors in 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, East Asia was monitored by the CONTRAIL program, which continuously conducted CO_2 measurements over 3 years, with routes repeated throughout all seasons. This resulted in the area most sensitive to the measurements exhibiting similar spatial patterns during the vegetation growing season (from May to October) and the non-growing season, covering the northeastern part of China, Korea, and Japan (Fig. S7). The airborne measurements taken in East Asia offer a unique opportunity to explore seasonal variations in regional error statistics. For the period 2015–2017, the regional averages of both the RMSE and ERR_{TOT} are, on average, 12 % and 11 % higher, respectively, during the non-growing season than during the growing season (Fig. S8). In contrast, the regional averages of $h\left(err_{f_e}\right)$ and $h\left(err_{f_t}\right)$ are higher during the growing season (0.90 [0.84, 0.97] and 1.37 [1.13, 1.62] ppm, respectively) than during the non-growing season (0.66 [0.62, 0.70] and 1.30 [1.06, 1.54] ppm, respectively) because CO_2 errors tend to increase proportionally with the magnitude of the flux values. Consequently, the ratio of $h\left(err_{f_e}\right)$ to $h\left(err_{f_t}\right)$ is slightly lower during the non-growing season, with a value of 0.51 [0.39, 0.64], than during the growing season, with a value of 0.66 [0.50, 0.83], indicating a greater underestimation of true flux errors when the terrestrial-biosphere CO_2 sinks are relatively smaller. This result aligns with our finding that true net land flux errors are significantly underestimated when fossil fuel emissions are larger in magnitude than terrestrial-biosphere fluxes. Furthermore, the consistent ratio of $h\left(err_{f_e}\right)$ to $h\left(err_{f_t}\right)$ 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 use atmospheric CO_2 data observed and simulated within the 1–5 km a.g.l. altitude range. The choice of this altitude range may have influenced the regional error statistics as the performance of the inverse models could vary with altitude. To gauge this sensitivity, we compare error statistics derived from atmospheric CO_2 data across two altitude ranges: 1–3 and 1–5 km a.g.l. Among the seven analyzed regions, Australia and South America are excluded from this additional analysis because the airborne observations for these two regions cover fewer than 100 grid cells with respect to 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 across Alaska, midlatitude North America, Europe, East Asia, and Southeast Asia, indicating that observations corresponding to lower altitudes are more sensitive to surface CO_2 fluxes (Fig. S9). Because of the higher sensitivity, error statistics for all regions have larger values when calculated using data from the 1–3 km a.g.l. altitude range than when determined using data from the 1–5 km a.g.l. altitude range (Fig. S10). For example, in midlatitude North America, the regional averages of the RMSE, ERR_{TOT}, $h\left(err_{f_e}\right)$, and $h\left(err_{f_t}\right)$ 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, respectively, when calculated using data from the 1–3 km a.g.l. altitude range. In comparison, when computed using data from the 1–5 km a.g.l. 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, respectively. However, the ratio between the 3-year mean values of $h\left(err_{f_e}\right)$ and $h\left(err_{f_t}\right)$ does not show significant differences based on the altitude ranges, with differences ranging from 0.02 to 0.11. Again, these results suggest that our observation-based regional flux error estimates are not sensitive to the choice of altitude range over longer time periods.

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

Representation errors and $h\left(err_{f_e}\right)$ 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 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 the OCO-2 MIP, 3 employed TM5 and 5 utilized GEOS-Chem

(Table S1). This ensemble size may not be sufficient for fully capturing the range of true transport errors. We further investigate how our main results would be affected if the estimated transport errors deviated by 20 % or 40 % from the actual errors, based on the difference between RMSE^2 and $\mathrm{ERR}_{\mathrm{TOT}}^2$. The ratio of the regional mean of $h\left(\mathrm{err}_{\mathrm{f_e}}\right)$ to that of $h\left(\mathrm{err}_{\mathrm{f_t}}\right)$ increases, on average, by only up to 0.04 and 0.09, respectively, in the seven regions throughout the analysis period (Fig. S11). In both cases, the estimated flux errors for midlatitude North America, Europe, East Asia, and Southeast Asia still show significant underestimations at a 95 % confidence level, whereas this is not observed for Alaska and South America. In Australia, characterized by a wide uncertainty range, significant underestimation is also observed in the "20 % deviation" cases, supporting the robustness of our findings. In future versions of the OCO-2 MIP, the participation of inverse-modeling groups using other transport models or meteorological forcing data might contribute to the estimation of transport errors closer to actual values.

This study uses monthly posterior flux estimates for the calculation of monthly $h\left(\mathrm{err}_{\mathrm{f_e}}\right)$ values. 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, 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 1 week to 1 month, the weekly variations in the 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 might affect our main results. Our analysis shows that the regional averages of $h\left(\mathrm{err}_{\mathrm{f_e}}\right)$ derived from the monthly posterior flux estimates from the 7 models are, on average, within ±10 % of the values originally obtained using flux estimates from all 10 models for the period 2015–2017, except with respect to Europe (13 % lower) (Fig. S12a). When accounting for different sub-monthly patterns of posterior fluxes across models, the regional averages of $h\left(\mathrm{err}_{\mathrm{f_e}}\right)$ increase by 10 %–22 % (0.06–0.14 ppm) across six regions, with a 45 % (0.23 ppm) increase for Europe. These results suggest that our earlier calculation, assuming identical sub-monthly flux variations, underestimates $h\left(\mathrm{err}_{\mathrm{f_e}}\right)$. We further investigated whether our main finding remains robust even when adjusting the original values of $h\left(\mathrm{err}_{\mathrm{f_e}}\right)$ using the potential-underestimation

rate. After making this correction, we found that the ratios between the regional average values of $h\left(\mathrm{err}_{\mathrm{f_e}}\right)$ and $h\left(\mathrm{err}_{\mathrm{f_t}}\right)$ increased the most in Europe (by 0.14) and by only up to 0.07 in the other six regions as $h\left(\mathrm{err}_{\mathrm{f_t}}\right)$ also increases with $h\left(\mathrm{err}_{\mathrm{f_e}}\right)$ according to Eq. (8) (Fig. S12b). Moreover, $h\left(\mathrm{err}_{\mathrm{f_t}}\right)$ still exhibits significant underestimation ($p < 0.05$) in midlatitude 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 $h\left(\mathrm{err}_{\mathrm{f_e}}\right)$.

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 model intercomparison projects, such as the OCO-2 MIP. The evaluation results of the OCO-2 MIP ensemble members reveal that the true errors in ensemble posterior fluxes are larger than the ensemble spread in regions with high anthropogenic CO_2 emissions. This result provides observation-based evidence supporting previous studies (Wang et al., 2020; Oda et al., 2023) 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 ongoing airborne measurement programs – including the CONTRAIL and IAGOS-CARIBIC programs, as well as various airborne programs hosted by 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). 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 uncertainties in regional surface CO_2 fluxes.

Appendix A

Following Eq. (1) in the main text,

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

Here, $\overline{h(\hat{x}_i)}$ denotes the ensemble mean of the posterior CO_2 concentrations from the OCO-2 MIP models corresponding to the ith airborne observation ($y_{\mathrm{o},i}$) within each $1° \times 1°$ grid

cell for each month. N is the total number of data points from airborne measurements sampled in each grid cell monthly. M is the ensemble size (i.e., 10 members).

Equation (A1) can be rewritten as TS3

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

$$= \frac{1}{N}\sum_{i=1}^{N}\left[y_{\mathrm{o},i} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]\left[y_{\mathrm{o},i} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]^T$$
$$- 2\left(y_{\mathrm{o}} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right) \times \left(\overline{h\left(\hat{x}_i\right)} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right)$$
$$+ \left[\overline{h\left(\hat{x}_i\right)} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]\left[\overline{h\left(\hat{x}_i\right)} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]^T, \quad \text{(A3)}$$

where $h_{\mathrm{t}}(\hat{x}_{\mathrm{t}})$ denotes the estimated CO_2 concentration obtained from an error-free atmospheric transport model (h_{t}) and a set of true CO_2 fluxes ($\hat{x}_{\mathrm{t}}$). The three terms on the right-hand side of Eq. (A3) represent (i) the variances of observation and representation errors, (ii) the covariances between errors in observation and representation and errors in flux and transport, and (iii) the variances of flux and transport errors in the ensemble estimates. Assuming the independence of observation and representation errors from transport and flux errors, Eq. (A3) can be simplified as

$$\mathrm{RMSE}^2 = \frac{1}{N}\sum_{i=1}^{N}\left[y_{\mathrm{o},i} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]\left[y_{\mathrm{o},i} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]^T$$
$$+ \left[\overline{h\left(\hat{x}_i\right)} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]\left[\overline{h\left(\hat{x}_i\right)} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]^T. \quad \text{(A4)}$$

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:

$$\frac{1}{N}\sum_{i=1}^{N}\left[\overline{h\left(\hat{x}_i\right)} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]\left[\overline{h\left(\hat{x}_i\right)} - h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right)\right]^T$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(\overline{h\left(\hat{x}_i\right)} - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right) - \left(h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right) - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right)\right]$$
$$\left[\left(\overline{h\left(\hat{x}_i\right)} - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right) - \left(h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right) - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right)\right]^T \quad \text{(A5)}$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left[\overline{h\left(\hat{x}_i\right)} - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right]\left[\overline{h\left(\hat{x}_i\right)} - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right]^T$$
$$- 2\left(\overline{h\left(\hat{x}_i\right)} - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right)\left(h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right) - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right)$$
$$+ \left[h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right) - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right]\left[h_{\mathrm{t}}\left(\hat{x}_{\mathrm{t},i}\right) - \overline{h\left(\hat{x}_{\mathrm{t},i}\right)}\right]^T. \quad \text{(A6)}$$

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

In the OCO-2 MIP, by approximating the ensemble spread of the posterior fluxes as true errors in the mean fluxes, it is assumed that the values of the first and second terms on the right-hand side of Eq. (A4) can be written as the sum of the

set of observation errors ($\mathrm{ERR}^2_{\mathrm{OBS}}$), the set of representation errors ($\mathrm{ERR}^2_{\mathrm{REP}}$), and the ensemble spread of posterior CO_2 concentrations across the OCO-2 MIP models ($\mathrm{ERR}^2_{\mathrm{MIP}}$):

$$\mathrm{RMSE}^2 \approx \mathrm{ERR}^2_{\mathrm{TOT}} = \mathrm{ERR}^2_{\mathrm{OBS}} + \mathrm{ERR}^2_{\mathrm{REP}} + \mathrm{ERR}^2_{\mathrm{MIP}}. \quad \text{(A7)}$$

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

$\mathrm{ERR}^2_{\mathrm{MIP}}$ can be also rewritten by separating the flux error and transport error terms as follows:

$$\mathrm{ERR}^2_{\mathrm{MIP}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}$$
$$\left[\overline{h\left(\hat{x}_i\right)} - h_j\left(\hat{x}_{j,i}\right)\right]\left[\overline{h\left(\hat{x}_i\right)} - h_j\left(\hat{x}_{j,i}\right)\right]^T \quad \text{(A8)}$$
$$= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}\frac{1}{M}\sum_{k=1}^{M}$$
$$\left[\left(h_k\left(\hat{x}_{k,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right) - \left(h_j\left(\hat{x}_{j,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right)\right]$$
$$\left[\left(h_k\left(\hat{x}_{k,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right) - \left(h_j\left(\hat{x}_{j,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right)\right]^T \quad \text{(A9)}$$
$$= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{k=1}^{M}\frac{1}{M}\sum_{j=1}^{M}$$
$$\left[h_k\left(\hat{x}_{k,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right]\left[h_k\left(\hat{x}_{k,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right]^T$$
$$- 2\left(h_k\left(\hat{x}_{k,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right)\left(h_j\left(\hat{x}_{j,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right)$$
$$+ \left[h_j\left(\hat{x}_{j,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right]\left[h_j\left(\hat{x}_{j,i}\right) - h_k\left(\hat{x}_{j,i}\right)\right]^T. \quad \text{(A10)}$$

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

Code and data availability. The TS4 inverse-modeling results and airborne CO_2 measurement data from the OCO-2 v10 MIP are available on the official website of the NOAA/ESRL Global Monitoring Laboratory at https://www.gml.noaa.gov/ccgg/OCO2_v10mip/download.php (NOAA/ESRL Global Monitoring Laboratory, 2022). The high-resolution global GEOS-Chem simulation results used to calculate the representation errors can be obtained upon request from Brad Weir (brad.weir@nasa.gov) or Lesley Ott (lesley.e.ott@nasa.gov). The forward and adjoint sensitivity simulations for this work were conducted using the publicly available GEOS-Chem adjoint model. This model can be downloaded from http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_Adjoint (Henze et al., 2007). ODIAC fossil fuel CO_2 emission data are available at https://doi.org/10.17595/20170411.001 (Oda and Maksyutov, 2015). Hourly (or 3-hourly) terrestrial-biosphere carbon flux datasets from CASA-GFED3, CASA-GFED4.1s, SiB4, CARDAMOM, and ORCHIDEE – all OCO-2 v10 MIP prior flux models – are available at https://doi.org/10.5067/VQPRALE26L20 (Ott, 2020), https://doi.org/10.25925/20201008 (Jacobson et al.,

2020), https://doi.org/10.3334/ORNLDAAC/1847 (Haynes et al., 2021), https://doi.org/10.5067/1XO0PZEZOR1H (Liu and Bowman, 2024), and on the official website of the CAMS (https://ads.atmosphere.copernicus.eu/datasets/cams-global-greenhouse-gas-inversion?tab=download, Copernicus Atmosphere Monitoring Service, 2020).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/acp-25-1-2025-supplement.

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

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. We are grateful to NOAA, NASA LaRC, NASA GSFC, INPE, NIES, MRI, the European research infrastructure IAGOS-CARIBIC, the DOE's LBNL, and Harvard University for providing the airborne CO_2 observation data. We thank Colm Sweeney, Kenneth Davis, Joshua P. DiGangi, Melissa Y. Martin, John B. Miller, Emanuel Gloor, Wouter Peters, Toshinobu Machida, Hidekazu Matsueda, Yousuke Sawa, Yosuke Niwa, Ed Dlugokencky, Stephan R. Kawa, James B. Abshire, Haris Riris, Florian Obersteiner, Harald Boenisch, Torsten Gehrlein, Andreas Zahn, Christoph Gerbig, Tanja Schuck, Gao Chen, Michael Shook, Giordane A. Martins, Rodrigo A. de Souza, Britton Stephens, Eric Kort, Thomas Ryerson, Jeff Peischl, Ken Aikin, Steve Wofsy, Bruce Daube, and Roisin Commane for contributing airborne CO_2 data. We also appreciate all the providers of the datasets used in the GEOS-Chem model simulations. We especially thank all research groups involved in the OCO-2 MIP for their contributions to providing v10 OCO-2 and in situ CO_2 datasets, as well as 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). The South American airborne CO_2 measurements were supported by the European Research Council through the ASICA project (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 (ARM) program, the ARM Aerial Facility, and the Environmental System Science (ESS) program.

Review statement. This paper was edited by Christoph Gerbig and reviewed by two anonymous referees.

References

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, 66, 23616, https://doi.org/10.3402/tellusb.v66.23616, 2014.

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.

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.

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.

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. T. Roy. Soc. A, 380, 20210192, https://doi.org/10.1098/rsta.2021.0192, 2021.

Basu, S. and Nassar, R.: Fossil Fuel CO2 Emissions for the OCO2 Model Intercomparison Project (MIP), Version 2020.1, Zenodo [data set], https://doi.org/10.5281/zenodo.4776925, 2021.

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.

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.

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.

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, D09307, https://doi.org/10.1029/2006JD007375, 2007.

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.

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.

Copernicus Atmosphere Monitoring Service: CAMS global inversion-optimised greenhouse gas fluxes and concentrations, v21r1, Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store, https://ads.atmosphere.copernicus.eu/datasets/cams-global-greenhouse-gas-inversion?tab=download (last access: 1 October 2024), 2020.

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.

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, B. Am. Meterol. Soc., 102, E1714–E1734, https://doi.org/10.1175/BAMS-D-20-0300.1, 2021.

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.

Feng, L., Palmer, P. I., Bösch, H., and Dance, S.: Estimating surface CO2 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.

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.

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, 67, 27989, https://doi.org/10.3402/tellusb.v67.27989, 2015.

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., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tans, P. P., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., van Ooijen, E., Wanninkhof, R., Watanabe, M., Wimart-Rousseau, C., Yang, D., Yang, X., Yuan, W., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.: Global Carbon Budget 2023, Earth Syst. Sci. Data, 15, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023, 2023.

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.

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., John-

son, 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. Cy., 37, e2023GB007804, https://doi.org/10.1029/2023GB007804, 2023.

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.

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. Cy., 18, GB1010, https://doi.org/10.1029/2003GB002111, 2004.

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.

Haynes, K. D., Baker, I. T., and Denning, A. S.: SiB4 Modeled Global 0.5-Degree Hourly Carbon Fluxes and Productivity, 2000-2018, ORNL DAAC [data set], Oak Ridge, Tennessee, USA, https://doi.org/10.3334/ORNLDAAC/1847, 2021.

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 (data/code available at: http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_Adjoint, last access: 29 June 2023).

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 CO2 using GOSAT measurements, J. Geophys. Res.-Atmos., 120, 5253–5266, https://doi.org/10.1002/2014JD022962, 2015.

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.

Karion, A., Sweeney, C., Tans, P., and Newberger, T.: AirCore: An Innovative Atmospheric Sampling System, J. Atmos. Ocean. Tech., 27, 1839–1853, https://doi.org/10.1175/2010JTECHA1448.1, 2010.

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.

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 CO2 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.

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 CO2 flux estimates, Atmos. Meas. Tech. Discuss. [preprint], https://doi.org/10.5194/amt-2019-257, 2019.

Lauvaux, T., Díaz-Isaac, L. I., Bocquet, M., and Bousserez, N.: Diagnosing spatial error structures in CO2 mole fractions and XCO2 column mole fractions from atmospheric transport, Atmos. Chem. Phys., 19, 12007–12024, https://doi.org/10.5194/acp-19-12007-2019, 2019.

Liu, J. and Bowman, K.: Carbon Monitoring System Carbon Flux Land Prior L4 V3, Goddard Earth Sciences Data and Information Services Center (GES DISC) [data set], Greenbelt, MD, USA, https://doi.org/10.5067/1XO0PZEZOR1H, 2024.

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, 66, 22486, https://doi.org/10.3402/tellusb.v66.22486, 2014.

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.

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.

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. Ocean. Tech., 25, 1744–1754, https://doi.org/10.1175/2008JTECHA1082.1, 2008.

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, D17305, https://doi.org/10.1029/2011JD016270, 2011.

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.

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, 69, 1399047, https://doi.org/10.1080/16000889.2017.1399047, 2017.

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.

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 [data set], https://doi.org/10.25925/20181030, 2018.

NOAA/ESRL Global Monitoring Laboratory: The OCO-2 v10 MIP [data set], https://gml.noaa.gov/ccgg/OCO2_v10mip/download.php (last access: 22 January 2025), 2022.

Oda, T. and Maksyutov, S.: ODIAC Fossil Fuel CO2 Emissions Dataset, Version ODIAC2020b, Center for Global Environmental Research, NIES [data set], https://doi.org/10.17595/20170411.001, 2015.

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, 124030, https://doi.org/10.1088/1748-9326/ad059b, 2023.

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.

Ott, L.: GEOS-Carb CASA-GFED 3-hourly Ecosystem Exchange Fluxes 0.5 degree x 0.625 degree V3, Goddard Earth Sciences Data and Information Services Center (GES DISC) [data set], Greenbelt, MD, USA, https://doi.org/10.5067/VQPRALE26L20, 2020. TS5

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.

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.

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

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. Cy., 33, 484–500, https://doi.org/10.1029/2018GB006086, 2019.

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.

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.

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.

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

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.

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.

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, B. Am. Meteorol. Soc., 103, E761–E790, https://doi.org/10.1175/BAMS-D-20-0315.1, 2022.

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.

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.

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, 085002, https://doi.org/10.1088/1748-9326/ab9795, 2020.

Wei, Y., Shrestha, R., Pal, S., Gerken, T., Feng, S., McNelis, J., Singh, D., Thornton, M. M., Boyer, A. G., Shook, M. A., Chen, G., Baier, B. C., Barkley, Z. R., Barrick, J. D., Bennett, J.

R., Browell, E. V., Campbell, J. F., Campbell, L. J., Choi, Y., Collins, J., Dobler, J., Eckl, M., Fiehn, A., Fried, A., Digangi, J. P., Barton-Grimley, R., Halliday, H., Klausner, T., Kooi, S., Kostinek, J., Lauvaux, T., Lin, B., McGill, M. J., Meadows, B., Miles, N. L., Nehrir, A. R., Nowak, J. B., Obland, M., O'Dell, C., Fao, R. M. P., Richardson, S. J., Richter, D., Roiger, A., Sweeney, C., Walega, J., Weibring, P., Williams, C. A.,Yang, M. M., Zhou, Y., and Davis, K. J. : Atmospheric Carbon and Transport – America (ACT-America) Datasets: Description, Management, and Delivery, Earth Space Sci., 8, e2020EA001634, https://doi.org/10.1029/2020EA001634, 2021.

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.

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.

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.

Remarks from the typesetter