

Multi-Season Evaluation of CO₂ Weather in OCO-2 MIP Models

Li Zhang¹, Kenneth J. Davis^{1,2}, Andrew E. Schuh³, Andrew R. Jacobson⁴, Sandip Pal⁵, Yu Yan Cui¹, David Baker³, Sean Crowell⁶, Frederic Chevallier⁷, Marine Remaud⁷, Junjie Liu⁸, Brad Weir^{9,10}, Sajeer Philip¹¹, Matthew S. Johnson¹², Feng Deng¹³, Sourish Basu^{10,14}

¹Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA

²Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA

³Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA

⁴University of Colorado Boulder and NOAA Earth System Research Laboratory, Boulder, CO, USA;

⁵Department of Geosciences, Atmospheric Science Division, Texas Tech University, Lubbock, TX, USA

⁶School of Meteorology, University of Oklahoma, Norman, OK, USA

⁷Laboratoire des Sciences du Climat et de L'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ,

Université Paris-Saclay, Gif-sur-Yvette, France

⁸Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

⁹Universities Space Research Association, Columbia, MD, USA

¹⁰NASA Goddard Space Flight Center, Greenbelt, MD, USA

¹¹NASA Academic Mission Services by Universities Space Research Association at NASA Ames Research Center, Mountain View, CA, USA

¹²NASA Ames Research Center, Moffett Field, CA, USA

¹³Department of Physics, University of Toronto, Toronto, ON, Canada

¹⁴Earth System Science Interdisciplinary Center, University of Maryland, MD, USA

Correspondence to Li Zhang (alex.zhang@psu.edu)

Key Points:

- Global inversion systems are able to simulate observed CO₂ frontal differences but with varying degrees of success.
- Most global inversion systems underestimate dormant-season frontal and vertical CO₂ differences.

- Inversion framework differences appear to explain more of the model-data differences in CO₂ weather metrics than CO₂ data sources.

Abstract

The ability of current global models to simulate the transport of CO₂ by mid-latitude, synoptic-scale weather systems (i.e. CO₂ weather) is important for inverse estimates of regional and global carbon budgets but remains unclear without comparisons to targeted measurements. Here, we evaluate ten models that participated in the Orbiting Carbon Observatory-2 model intercomparison project (OCO-2 MIP version 9) with intensive aircraft measurements collected from the Atmospheric Carbon Transport (ACT)-America mission. We quantify model-data differences in the spatial variability of CO₂ mole fractions, mean winds, and boundary layer depths in 27 mid-latitude cyclones spanning four seasons over the central and eastern United States. We find that the OCO-2 MIP models are able to simulate observed CO₂ frontal differences with varying degrees of success in summer and spring, and most underestimate frontal differences in winter and autumn. The models may underestimate the observed boundary layer-to-free troposphere CO₂ differences in spring and autumn due to model errors in boundary layer height. Attribution of the causes of model biases in other seasons remains elusive. Transport errors, prior fluxes, and/or inversion algorithms appear to be the primary cause of these biases since model performance is not highly sensitive to the CO₂ data used in the inversion. The metrics presented here provide new benchmarks regarding the ability of atmospheric inversion systems to reproduce the CO₂ structure of mid-latitude weather systems. Controlled experiments are needed to link these metrics more directly to the accuracy of regional or global flux estimates.

Plain Language Summary

Global flux estimate systems use CO₂ observations, atmospheric transport models, CO₂ flux models (emissions and absorption), and mathematical optimization methods to estimate biosphere-atmosphere CO₂ exchange. Accurate representation of atmospheric transport is important for a reliable optimization of fluxes in these systems. We use intensive aircraft measurements of wind speed, boundary layer height, and horizontal and vertical differences of CO₂ concentrations within 27 mid-latitude cyclones collected by the Atmospheric Carbon Transport (ACT)-America mission to evaluate the performance of ten global flux estimate systems from the Orbiting Carbon Observatory-2 model intercomparison project (OCO-2 MIP). We find the models can simulate observed horizontal CO₂ differences between the warm and cold parts of cyclones with different degrees of success in summer and spring, but often underestimate the observed cross-frontal and vertical differences in CO₂ in winter and autumn. The models may underestimate the CO₂ differences between the boundary layer and the free troposphere due to model errors in boundary layer height and surface fluxes. These weather-oriented CO₂ metrics provide benchmarks for testing simulations of the CO₂ structure within cyclones. Future efforts are needed to link these metrics more directly to the accuracy of CO₂ flux estimates.

1. Introduction

Synoptic weather systems in the mid-latitudes are important for modulating the distribution and variability of atmospheric CO_2 (e.g., Hurwitz et al., 2004; Parazoo et al., 2008; Barnes et al., 2016; Pal et al., 2020a). The latitudinal transport and vertical mixing resulting from synoptic weather events are an important part of the global greenhouse gas (GHG) transport system. CO_2 inversion modelling uses atmospheric transport models, measurements of CO_2 mole fractions (henceforth $[\text{CO}_2]$), and a prior surface flux estimate to solve for surface fluxes. Misrepresentation of synoptic-scale weather systems in atmospheric transport models (e.g., errors in horizontal advection and vertical mixing) results in biases in regional and global flux inversions (Wang et al., 2007; Patra et al., 2008; Parazoo et al., 2011). Some attempts have been made to simulate $[\text{CO}_2]$ variability in weather systems with global models and to evaluate model performance with tower measurements, but tower data are spatially sparse and insufficient to describe the $[\text{CO}_2]$ variability within weather systems (e.g., Wang et al., 2007; Parazoo et al., 2011; Agustí-Panareda et al., 2019). Without comparison to intensive targeted measurements of $[\text{CO}_2]$ within weather systems, the ability of current-generation global inversion systems to simulate the $[\text{CO}_2]$ transport by weather systems at northern mid-latitudes remains unclear. It is therefore of great importance to conduct weather-scale evaluation of current state-of-the-art global models to identify possible biases in transport and consequently reduce the uncertainty in estimated fluxes.

Previous intercomparisons of fluxes inferred by inverse systems revealed large divergences at continental to subcontinental scales (e.g., Peylin et al., 2013; Houweling et al., 2015; Peiro et al., 2021). Peylin et al. (2013) showed notable uncertainties (0.17–0.49 PgC/yr) in annual CO_2 flux estimates over North America (N.A.) in a large model ensemble. The first Orbiting Carbon Observatory-2 model intercomparison project (OCO-2 MIP version 7) suggested that these subcontinental uncertainties do not appear to be decreasing (e.g., $>0.4 \text{ PgC/yr}$ in N.A.), even though the observations are becoming more abundant and the inversion systems are getting more sophisticated (Crowell et al., 2019). The difference in simulated atmospheric transport appears to be an important and persistent source of these flux uncertainties in global inversion systems (e.g., Parazoo et al., 2011; Parazoo et al., 2012; Schuh et al., 2019). Parazoo et al. (2012) suggested that mid-latitudes are an important location for differences in transport and the differences in moist frontal transport processes and associated vertical transport can lead to systematic differences in moist poleward and dry equatorward $[\text{CO}_2]$ transport. Recently, Schuh et al. (2019) showed differences between two commonly-used global transport models (GEOS-Chem and TM5) which are due at least in part to differences in vertical transport in the mid-latitudes. The transport differences result in space-time differences in simulated $[\text{CO}_2]$ and in annual zonal-mean inverse CO_2 flux estimates that differ by up to 1.7 PgC/yr . These findings motivate the need for using targeted observations to evaluate mid-latitude transport in these inversion systems.

Previous studies have evaluated the ability of global models to simulate $[\text{CO}_2]$ spatial-temporal variability. Denning et al. (1995) focused on annually- and zonally-averaged latitudinal gradients, revealing important biases due to the covariance between vertical mixing and fluxes. Stephens et al. (2007) studied seasonally- to annually-averaged $[\text{CO}_2]$ vertical profiles, revealing model-data differences in vertical mixing that were correlated with differences in the latitudinal distribution of the inverse flux estimates. Agustí-Panareda et al. (2019) evaluated the performance of a high-resolution global $[\text{CO}_2]$ simulation with point-based comparisons aggregated across many global observation sites. Parazoo et al. (2011) studied transport processes at a small number of tower-based observation sites, and Wang et al. (2007) studied how one weather system led to variability in atmospheric $[\text{CO}_2]$ across a small number of widely-spaced tower observational sites with a regional coupled model. Only Wang et al. (2007) examined the coherent structures in $[\text{CO}_2]$ linked to weather systems. None of these studies benefitted from observations that can resolve the $[\text{CO}_2]$ variability within a given weather system.

Current long-term $[\text{CO}_2]$ measurement systems are spatially sparse and insufficient to resolve the $[\text{CO}_2]$ structures within mid-latitude weather systems. Tower measurements (e.g., Andrews et al., 2014), although continuous in time, are limited in number and confined to the planetary boundary layer (PBL), making it difficult to use them to describe the horizontal and vertical structures of $[\text{CO}_2]$ within mid-latitude cyclones (e.g., Hurwitz et al., 2004; Lee et al., 2012). The long-term Greenhouse Gases Reference Network Aircraft program led by the NOAA Global Monitoring Laboratory (Sweeney et al., 2015) spans a large portion of North America. However, its observation frequency (weekly to monthly) and preference for sampling clear-weather days limits the ability of this network to resolve the dynamics within weather systems. Satellite column CO_2 measurements (χ_{co_2}) encompass most of the globe and are continuous along the satellite tracks, but are relatively sparse in space and time, and limited to measurements of the column. Further, cloudy air associated with synoptic weather systems tends to inhibit satellite retrievals near these frontal systems (Corbin and Denning, 2006). Airborne measurement programs, e.g., the Atmospheric Tomography Mission (ATom) and HIAPER Pole-to-Pole Observations program (HIPPO) focusing on intercontinental or pole-to-pole variations of atmospheric trace gases (Wofsy, 2011; Wofsy et al., 2018), although broad in spatial and vertical coverage, did not target weather systems.

The Atmospheric Carbon Transport (ACT)-America project was carried out during multiple seasons in 2016–2019 over a large portion of the U.S. to investigate the spatial and temporal variations of $[\text{CO}_2]$ within synoptic weather systems (Davis et al., 2021). These unprecedented and unique pictures of 4D spatial variations of $[\text{CO}_2]$ collected by ACT-America fill the gaps in regular in-situ and aircraft measurement networks and serve as a valuable dataset for studying the skills of global inversion systems in reproducing observed CO_2 weather. Analyses of the first ACT-America campaign in summer 2016 unveiled several quantitative characteristics of spatial variations of $[\text{CO}_2]$ in mid-latitude

cyclones including significant enhancements (5–30 ppm) in $[\text{CO}_2]$ across frontal boundaries and positive vertical $[\text{CO}_2]$ gradients (up to 20 ppm) between the PBL and free troposphere (FT) (Pal et al., 2020a). These intensive, weather-oriented measurements, alongside multi-model simulations from the most recent OCO-2 MIP (version 9), provide us with a new opportunity to better diagnose and quantify the impacts of atmospheric transport uncertainties on spatial distributions of $[\text{CO}_2]$ in global models and on estimated fluxes. Recently, Gaudet et al. (2020) compared OCO-2 MIP inversions constrained with in situ measurements (IS inversions, version 7) against vertical $[\text{CO}_2]$ profiles collected during the 2016 ACT-America campaign. They suggested that the OCO-2 MIP members underestimate the differences between the warm and cold sectors of cyclones and have a large model spread in sector-mean $[\text{CO}_2]$ (>5 ppm) which might be due to differences in inversion systems and/or prior fluxes among models. However, the study only focused on the vertical profiles of $[\text{CO}_2]$ in summer; cross-frontal transects, multi-seasonal, and meteorological evaluation of the OCO-2 MIP models remained unresolved. A more comprehensive examination of the transport of $[\text{CO}_2]$ from inversions constrained by multiple data sources (e.g., in situ measurements and satellite CO_2 retrievals) and meteorological variables (e.g., wind speed and boundary layer height) in different seasons is also needed.

In this study, ten modelling systems with different transport and inversion methods constrained by multiple suites of measurements (i.e., multiple observing modes of OCO-2 and in situ measurement networks) that participated in the OCO-2 MIP (version 9) are evaluated against aircraft measurements collected within mid-latitude cyclones during four seasons. Through a model-data comparison of both $[\text{CO}_2]$ fields and meteorological variables, we aim to shed light on the ability of global models to simulate carbon weather over the central and eastern U.S. We describe the OCO-2 MIP simulations, ACT-America campaigns, and ACT observational metrics that we employ (frontal and PBL-to-FT $[\text{CO}_2]$ differences) in section 2. We evaluate modelled $[\text{CO}_2]$ against ACT-America measurements in multiple seasons (section 3.1), present comparisons between observed and simulated $[\text{CO}_2]$ in the cold and warm sectors of frontal systems (section 3.2), and examine the model performance of ACT observational metrics (section 4). In Section 5, we examine two frontal cases, one in summer and one in winter, to explore the causes of model discrepancies in $[\text{CO}_2]$ weather. Finally, section 6 presents the conclusions of this study.

2. OCO-2 MIP and ACT-America

2.1. OCO-2 MIP

The OCO-2 MIP is an international collaboration of modelling efforts using global model systems to study the impact of varying atmospheric inversion systems and data sources on the estimated surface fluxes and atmospheric distributions of CO_2 (e.g., Crowell et al., 2019). In the present study, we use simulations from ten models that participated in the OCO-2 MIP (version 9). Characteristics of each of these models are listed in Table 1 while further de-

tails are available in Crowell et al. (2019) and Peiro et al. (2021). As outlined in Table 1, each inversion system utilizes one of five transport models (TM5, GEOS-Chem, GEOS-5, PCTM, and LMDz) and one of four meteorological re-analyses (ERA-Interim, MERRA2, GEOS-FP, and ERA5). Resolution of the transport model and inversion method vary considerably across the inversion systems. Note that CT is run at $1^\circ \times 1^\circ$ over N.A. and $2^\circ \times 3^\circ$ globally (referred to as CT_Global).

Five suites of experiments are performed with the OCO-2 MIP model systems to simulate global $[\text{CO}_2]$: (1) PRIOR, in which models are driven with prior fluxes and no inversion is performed; (2) IS, the inversion experiment assimilating in situ $[\text{CO}_2]$ measurements from the Cooperative Global Atmospheric Data Integration Project (<https://www.esrl.noaa.gov/gmd/ccgg/obspack/>); (3) LNLG, inversions that assimilate 10s averaged Land Nadir and Land Glint retrievals (version 9) observed by OCO-2 (Eldering et al., 2017); (4) OG, in which models assimilate 10s averaged OCO-2 Ocean Glint retrievals; and (5) LNLGOGIS, where inversions are constrained with in situ measurements and all OCO-2 χ_{CO_2} retrievals over both land and ocean surfaces. Modelled $[\text{CO}_2]$ and meteorological variables are sampled from hourly or 3-hourly 4D gridded simulations at the measurement times and locations using the same method as Zhang et al. (2020).

2.2. ACT-America aircraft measurements

The ACT-America project, a NASA Earth Venture Suborbital-2 (EVS-2) mission, is designed to advance the current understanding of surface fluxes of CO_2 and CH_4 at regional scales over North American mid-latitudes (Davis et al., 2021). The project consists of five field campaigns in four seasons spanning from 2016 to 2019 with each campaign lasting ~6 weeks (July 18–August 28, 2016, February 1–March 10, 2017, October 3–November 10, 2017, Apr 12–May 20, 2018, and June 17–July 27, 2019). The study area covers the eastern half of the U.S which is characterized by considerable spatial and seasonal variabilities of biogenic CO_2 fluxes (e.g., Huntzinger et al., 2012; Sweeney et al., 2015) and dynamic weather (e.g., Whittaker and Horn, 1984). In this study, we use the aircraft measurements during the first four campaigns, i.e. summer 2016 (July 15–August 28), winter 2017 (January 30–March 10), autumn 2017 (Oct 3–Nov 13), and spring 2018 (April 12–May 20). Flight-by-flight details are available in Pal and Davis (2021).

Two airborne platforms (NASA Wallops C130 and NASA Langley B200) equipped with in-situ sensors were employed during the campaigns to measure greenhouse gases (e.g., $[\text{CO}_2]$ and CH_4) and other trace gases (e.g., CO , C_2H_6 , and O_3) as well as meteorological variables (e.g., horizontal wind, temperature, pressure, and water vapor). The C130 also included lidars measuring planetary boundary layer height (PBLH) and partial χ_{CO_2} (Campbell et al., 2020; Wei et al., 2021). In total, 182 research flights (RFs) were flown in four seasons and the tracks of RFs in each season are shown in Figure S1. Each RF spanned from 4 to 6 hours during daytime with a spatial extent of 600 km or more

(Davis et al., 2021; Wei et al., 2021). There are 2–4 vertical levels during each flight extending from the PBL (~300 m above ground level (a.g.l.)) to the upper free troposphere (4000–9000 m above sea level (a.s.l.)) with the maximum altitude depending on weather and flight plan, and typically 8–12 vertical profiles on each flight day). On average, ~50% of the flight hours were carried out in the PBL. RFs on frontal days were designed to fly across the frontal boundaries collecting observations in both cold and warm sectors at different vertical levels.

In addition to $[\text{CO}_2]$ observations, airborne in-situ measurements of wind speed and lidar measurements of PBLH made during the campaigns are used in our model evaluation. In-situ wind speeds were measured on both airborne platforms (Wei et al., 2021). PBLHs were determined by the aerosol backscatter profiles measured with the Goddard Cloud Physics Laboratory lidar aboard on C130 using a wavelet method (Davis et al., 2000; Pal et al., 2010). A detailed description of the PBLH datasets for all four deployments can be found in Pal et al. (2020b).

In this work, we use 5-second averaged measurements of $[\text{CO}_2]$ and wind speed and 8-second averaged PBLH measurements. Measurements of $[\text{CO}_2]$ and wind speed are classified into three vertical layers: planetary boundary layer (PBL), lower free troposphere (LFT) which defined as the layer between the top of PBL and 4 km a.s.l.), and upper free troposphere (UFT) which extends from 4 km a.s.l. to the maximum altitude of aircraft measurement (typically 8 km a.s.l.) following the method in Pal et al. (2020a). PBLH varied, but PBL flight legs were conducted at about 300m above ground level (AGL), an altitude typically in the mid- to lower-PBL, and measurements were not conducted if the aircraft could not fly within the PBL.

2.3. ACT metrics of frontal $[\text{CO}_2]$ features

To investigate the skills of OCO-2 MIP models in simulating the horizontal and vertical structures of $[\text{CO}_2]$ across frontal boundaries (i.e., in the warm and cold sectors at different altitudes), we employ two metrics developed by the ACT-America project: (1) frontal $[\text{CO}_2]$ difference between the warm and cold sectors and (2) PBL-to-FT $[\text{CO}_2]$ difference within the warm and cold sectors of frontal systems (Pal et al., 2020a). These two metrics provide a quantitative understanding of the horizontal and vertical gradients of $[\text{CO}_2]$ associated with mixing processes within weather systems at a synoptic scale. The frontal difference is calculated by comparing the average $[\text{CO}_2]$ in the warm and cold sectors of frontal systems within the PBL, LFT, and UFT. These averages are computed along flight tracks that are a few to several hundred kilometres in length. Determination of warm and cold sectors of frontal systems (as well as pre- and post-frontal fair weather RFs) was based on surface synoptic map analyses, satellite images, and in-situ measurements of meteorological variables (Pal et al., 2020a). Note that $[\text{CO}_2]$ measurements during take-offs and landings are excluded to leave out the influences of small-scale anthropogenic CO_2 fluxes at airports. The frontal differences are calculated for each frontal event and sea-

sonal frontal metric statistics were then computed for both measurements and simulations.

The vertical PBL-to-FT $[\text{CO}_2]$ difference is defined as the difference between the layer-averaged $[\text{CO}_2]$ within the PBL and in the entire free troposphere (including LFT and UFT). Different from the method by Pal et al. (2020a) that only uses profile data, here we use both level leg and profile measurements for computing this ACT metric. We calculate observed and simulated PBL-to-FT $[\text{CO}_2]$ differences in both warm and cold sectors for each frontal research flight as well as for pre- and post-frontal flights. Pre-frontal flights are included in the warm sectors and post-frontal flights the cold sectors, as discussed in Pal et al. (2020a). Seasonal statistics of the PBL-to-FT differences are calculated using all frontal, pre-, and post-frontal flights in each season.

3. Model-data evaluation of $[\text{CO}_2]$

3.1 Seasonal observed and simulated total $[\text{CO}_2]$

We first evaluate the OCO-2 MIP simulations of the seasonal mean $[\text{CO}_2]$ in three vertical layers (PBL, LFT, and UFT) over the eastern U.S. (Figure 1). ACT measurements show higher averaged $[\text{CO}_2]$ in dormant seasons ($\sim 408\text{--}415$ ppm) due to a shallow PBL and ecosystem respiration, and lower mean in growing seasons ($\sim 396\text{--}409$ ppm) as a result of deep mixing and net photosynthesis fluxes (known as the CO_2 “rectifier” effect; Denning et al., 1999). The observed seasonal $[\text{CO}_2]$ difference during the ACT measurement periods (here defined as the seasonal peak-to-trough amplitude; $\text{AMP}_{\text{P-T}}$) ranges from ~ 10 to 19 ppm in the three vertical layers with the mean seasonal cycle most notable in the PBL ($\text{AMP}_{\text{P-T}} = 19$ ppm). Slower vertical mixing and a larger mixing volume attenuate and delay the seasonal cycle in the FT ($\text{AMP}_{\text{P-T}} = 10\text{--}13$ ppm).

The OCO-2 MIP ensembles driven by different suites of fluxes reproduce the observed seasonal variation of $[\text{CO}_2]$ (i.e. relatively high in winter and spring while low in summer and autumn) but have distinct and diverse performances. The ensembles simulate the observed $\text{AMP}_{\text{P-T}}$ in the FT to within 0–1 ppm, but under-predict the amplitude in the PBL by 2–3 ppm due to overestimates in seasonal mean $[\text{CO}_2]$ in summer and underestimates in spring. These biases in the PBL might result from model uncertainties in mixing or/and biases in biological fluxes. Model predictions using prior fluxes (i.e. the PRIOR experiment) tend to overestimate observed $[\text{CO}_2]$ in all four ACT-America seasons (ensemble biases are up to 4 ppm in summer and 1 ppm in winter) likely due to overestimates of the continental background. Posterior inversions with optimized surface fluxes have improved performance in total $[\text{CO}_2]$. The OG and LNLGOGIS ensembles are unbiased within the PBL in summer. However, the unbiased mean PBL- $[\text{CO}_2]$ in these two experiments is a result of compensating biases across the sectors of weather systems (see section 3.2). Meanwhile, the OG ensemble underestimates the observed mean FT $[\text{CO}_2]$ in summer by ~ 2 ppm and overestimates observed $[\text{CO}_2]$ in the LFT in winter by ~ 1 ppm. Mean biases in IS and LNLG ensembles are larger in the PBL (e.g., 0.2–1.1 ppm in

summer and -0.8–0.5 ppm in winter) than that in the FT (e.g., 0.1–0.7 ppm in summer and 0.1–0.5 ppm in winter).

The performance of individual OCO-2 MIP models in simulating observed seasonal average $[\text{CO}_2]$ is strongly model and season dependent, and there is no close correlation between model resolution and model biases in seasonal $[\text{CO}_2]$. A large model spread among OCO-2 MIP models exists in the PBL (e.g., 2–7 ppm). In contrast, models are closer to each other in the UFT (model spread = 1–3 ppm) where the model discrepancies are less influenced by differences in surface fluxes relative to the PBL. CT is close to measurements in summer but biased low in autumn (mean bias = 2–3 ppm) and spring (mean bias = 2–3 ppm) in the PBL. TM5-4DVAR overestimates PBL- $[\text{CO}_2]$ by 0.5–2 ppm in summer but agrees well with observations in winter, autumn, and spring. LoFI, with a relatively high resolution, matches well with observation in summer but is biased high by 1.2 ppm in winter and low by 1.5 ppm in spring in the PBL. The GEOS-Chem group overall overestimates PBL- $[\text{CO}_2]$ in summer but underestimates observations in winter.

3.2 Observed and simulated sector-mean $[\text{CO}_2]$

Assessments of $[\text{CO}_2]$ as a function of synoptic sectors in weather systems show that observed $[\text{CO}_2]$ in warm sectors of weather systems are distinctly different from that in cold sectors in all four seasons (Figure 2). In summer, $[\text{CO}_2]$ in warm sectors is considerably higher than that in cold sectors throughout the troposphere (e.g., 402 ppm vs. 392 ppm in the PBL; Figure 2a). While in winter, the warm-sector mean $[\text{CO}_2]$ is lower than the cold-sector mean by 0.8–2.0 ppm (e.g., 414 ppm vs. 412 ppm in the PBL; Figure 2b). Similar to winter, observed $[\text{CO}_2]$ is higher in cold sectors in autumn and spring within both the PBL and FT. These warm-cold sector contrasts in $[\text{CO}_2]$ are likely to be attributed to a combined effect of hemispheric gradient and differences in surface fluxes between the source regions of warm and cold air. In summer, observed $[\text{CO}_2]$ is nearly constant throughout the troposphere in the warm sectors, but $[\text{CO}_2]$ increases with altitude in the cold sectors (393 ppm in PBL vs. 402 ppm in UFT), most likely due to net biological uptake at the surface. In cold seasons, observed $[\text{CO}_2]$ decreases with altitude in both warm and cold sectors (Figure 2b–d). The PBL-FT $[\text{CO}_2]$ difference is the strongest in winter.

The agreement between some models/experiments and measurements in terms of the seasonal mean $[\text{CO}_2]$ breaks down when we conduct the evaluation as a function of synoptic sectors (Figure 2 and Figures S2–S9). In summer, OG underestimates $[\text{CO}_2]$ in warm sectors by ~3 ppm in the PBL but overestimates observed $[\text{CO}_2]$ in cold sectors by a similar magnitude. Biases in warm sectors are then offset by those in cold sectors, resulting in an unbiased ensemble mean in the PBL. In the FT, the ensemble persistently underestimates $[\text{CO}_2]$. We find similar results as for the LNLGOGIS experiment. In comparison, the IS and LNLG ensembles are closer to the observations than other ensembles in both the warm and cold sectors, which is different from the results when averaging the simulations over different sectors in weather systems. This implies that the

two experiments may have more realistic spatial distributions of surface fluxes as compared to the other experiments. The PRIOR simulations overestimate observed $[\text{CO}_2]$ in both warm and cold sectors in all seasons.

Large model-data discrepancies exist among different inversion systems, and discrepancies are overall smaller in warm sectors compared to cold sectors and smaller in FT compared to PBL (Figure 2 and Figures S2–S9). Model spread is smaller in the warm sectors than the cold sectors especially in summer and winter (3–7 ppm vs. 8–10 ppm). PCTM, CT, TM5-4DVAR, and LoFI better capture observations than GEOS-Chem models (CSU, Ames, CMS-Flux, and UT) in cold sectors within the PBL in summer with lower mean biases (0–3 ppm vs. 3–6 ppm; Figure S2) and higher model-data correlations ($R^2=0.6\text{--}0.8$ vs $R^2<0.6$; Figure S6). The high resolution of LoFI likely aids in its relatively good performance in model-data correlation and normalized standard deviation especially in summer, but the model does not surpass the others in winter and spring (Figures S7–S8). All models/model ensembles capture a higher warm-sector mean than the cold-sector in summer and the opposite in winter. But models overall have better performances in warm sectors compared to cold sectors with higher model-data correlations and lower normalized standard deviations in warm sectors (Figures S6–S9). Models are closer to each other in the FT than the PBL with much smaller model spreads in all seasons.

4 Horizontal and vertical variations of $[\text{CO}_2]$ in frontal systems

4.1 Observed and simulated frontal $[\text{CO}_2]$ differences

Observed summer frontal $[\text{CO}_2]$ differences are distinctly different from the other seasons (Figures 3 and Table 2). In summer, large positive frontal differences (i.e. $[\text{CO}_2]$ in warm sectors minus $[\text{CO}_2]$ in cold sectors) were observed with seasonal means of 9.2 ppm in the PBL and 2.5 ppm in the FT (Figure 3a). In winter, autumn, and spring, observed frontal differences became negative throughout the troposphere (except in the UFT in winter) with seasonal averages ranging from -3.2 to -2.1 ppm in the PBL and from -1.1 to -0.3 in the FT (Figures 3b–d) due to a change in the latitudinal $[\text{CO}_2]$ gradient compared to summertime. These observed features broadly reflect the hemispheric meridional gradient in fluxes with a south-to-north decrease in growing seasons and a south-to-north increase in dormant seasons, consistent with our understanding of continental biological fluxes and atmospheric transport (e.g., Sweeney et al., 2015).

All models can simulate positive frontal differences in summer but with varying degrees of success, and most of them can simulate negative frontal differences in other seasons (Figure 3 and Table 2). In summer, LoFI, CT, TM5-4DVAR, and PCTM simulate the observed frontal difference much better than the other MIP members in both PBL and FT (e.g., 6.7–9.3 ppm vs. 1.7–3.3 ppm in the PBL; Table 2) likely because the four models simulate the cold sector $[\text{CO}_2]$ more accurately (Figure 2a). The GEOS-Chem group significantly underestimates the horizontal differences in the lower troposphere by a factor of 4–5 mostly due to the overestimates in $[\text{CO}_2]$ within cold sectors (section 3.2). OU and CAMS

capture some of the observations and perform better than the GEOS-Chem models (Table 2). In the UFT, all models are much closer to observations and have smaller model discrepancies than in the PBL, likely due to comparable large-scale background $[\text{CO}_2]$ among the OCO-2 MIP members (section 3). In dormant seasons (winter and autumn), the models tend to simulate weaker negative frontal differences than observed ones in the PBL (Figures 3b–c and Table 2) with LoFI, CMAS, PCTM, and the TM5 group (CT, TM5-4DVAR, and OU) match the observations better than the GEOS-Chem models. In spring, models have divergent performance: PCTM, CAMS, and TM5-4DVAR better simulate the observed negative frontal differences than the other MIP members (Figure 3d). While other models, including CT, OU, and GEOS-Chem models, simulate weaker negative (or even positive) frontal differences compared to observations in the PBL and LFT. In the UFT, however, all models simulate stronger negative frontal differences than observations and have small model discrepancies.

Differences in modelled wind speed may partially explain the model discrepancies in frontal difference during summer. TM5 (CT, TM5-4DVAR, and OU), PCTM, LoFI, and CAMS reproduce observed wind speeds within frontal systems more accurately than the GEOS-Chem models (Figure 4a). GEOS-Chem winds are biased low, by 0.5–3 m/s in the PBL-LFT and by 2–4 m/s in the UFT within cold sectors. These biases in winds may lead to weaker advection of low- $[\text{CO}_2]$ air masses from the north in cold sectors during summer than other models, contributing to an overestimate in $[\text{CO}_2]$ within cold sectors and subsequent biases in frontal difference in GEOS-Chem. But it is also possible that biases in surface fluxes in the mid-west and southern Canada where cold-sector air originates might be responsible for the overestimate of $[\text{CO}_2]$ in cold sectors. In non-summer seasons, all models reproduce observed wind speeds in both warm and cold sectors without significant biases (Figures 4b–d).

Differences in modelled PBLH do not readily explain the diverging model performance in frontal $[\text{CO}_2]$ differences. In summer, PBLH in GEOS-Chem is comparable to that in TM5 in cold sectors (Figure 5a), which does not necessarily result in comparable model performance in $[\text{CO}_2]$ within the PBL and frontal $[\text{CO}_2]$ differences. While LoFI reproduces the observed frontal differences in summer, it is, however, biased high by ~400–1000 m in PBLH. In winter (Figure 5b), both LoFI and CAMS reproduce the observed PBLH but they differ in frontal $[\text{CO}_2]$ differences. TM5 (TM5-4DVAR, CT, and OU) is biased low in both warm and cold sectors in winter. However, TM5-4DVAR outperforms CT and OU in summer frontal differences, which may suggest more reasonable surface fluxes in TM5-4DVAR. Most models overestimate PBLH in autumn within both warm and cold sectors (Figure 5c), which may partially explain the weaker frontal $[\text{CO}_2]$ differences in models. In spring, the models slightly underestimate PBLH in both warm and cold sectors (e.g., biased low by 64–230 m in warm sectors and by 97–387 m in cold sectors; Figure 5d). These results suggest that differences in horizontal transport and surface fluxes may play an important role in the model discrepancies in frontal $[\text{CO}_2]$ difference.

There is no strong relationship between model resolution and performance in frontal $[\text{CO}_2]$ differences. Comparison between simulations of CT at $1^\circ \times 1^\circ$ and $2^\circ \times 3^\circ$ (CT_Global) suggests that model resolution is likely to have a moderate impact (up to 20%) on the simulated horizontal difference in summer (6.7 ± 0.5 vs. 5.5 ± 0.3 ppm in the PBL and 2.6 ± 1.3 vs. 2.3 ± 1.0 ppm in the FT) but has negligible impacts in other seasons (Figure 3 and Table 2). LoFI, at a fine resolution of $0.5^\circ \times 0.625^\circ$, reproduces observed frontal differences in summer and winter but does not perform better than other models in autumn and spring. This may suggest surface fluxes are reasonably represented in summer and winter but are biased in autumn and spring in LoFI. Models with a resolution of $2^\circ \times 3^\circ$ or less (e.g., TM5-4DVAR, LoFI, and PCTM) tend to simulate observed frontal $[\text{CO}_2]$ differences better than models at $4^\circ \times 5^\circ$ in summer, winter, and autumn within the PBL. However, OU, at a coarse resolution of $4^\circ \times 6^\circ$, better simulates the frontal difference in summer than GEOS-Chem at $4^\circ \times 5^\circ$.

Within a single inverse modelling system, varying the data source makes a relatively small difference in the simulated frontal $[\text{CO}_2]$ difference (Figure 3). Comparisons among PRIOR, IS, LNLG, and LNLGOGIS experiments (OG is excluded since it tends to degrade model performance) show that the models have limited spread (5–25%, represented as the normalized standard deviation among the four experiments with the same model) among the simulated frontal differences in the PBL. In the FT, the models have even smaller discrepancies among different data source experiments. These results suggest that differences in flux estimates may be caused primarily by elements of the inversion frameworks (e.g., prior flux errors, coherence lengths for flux errors; model-data mismatch errors that underweight the observations) than by different observational $[\text{CO}_2]$ data sources.

4.2 Observed and simulated PBL-to-FT $[\text{CO}_2]$ differences

Observed PBL-to-FT in $[\text{CO}_2]$ difference in summer is also distinctly different from those in other seasons (Figure 6 and Table 2). In summer, observations show a large mean negative PBL-to-FT $[\text{CO}_2]$ difference in cold sectors (-4.1 ± 6.2 ppm, the uncertainty representing the range of the PBL-to-FT difference across different frontal cases) and a weak positive vertical difference in warm sectors with large case-to-case variabilities (0.6 ± 3.7 ppm). These results based on all level-leg and profile measurements are different from the values derived from profile measurements only (ranging from -6 to 20 ppm in warm sectors and -24 to -2 ppm in cold sectors, Pal et al., 2020a). During winter and autumn (Figure 6b–c), large positive PBL-to-FT differences existed in both warm and cold sectors (3.9–5.7 ppm). The enhanced $[\text{CO}_2]$ in the PBL relative to the FT during cold seasons results from biological respiration at mid-latitudes and accumulation of fossil fuel emissions within the shallow PBL. In spring, the observed vertical PBL-to-FT difference wanes as photosynthesis turns up (2.4–2.6 ppm; Figures 6d).

The models often simulate weaker (or even the opposite) PBL-to-FT $[\text{CO}_2]$ differences compared to observations in all seasons. In summer, all models tend

to simulate a negative PBL-to-FT $[\text{CO}_2]$ difference in warm sectors in contrast to the observed positive value (Figure 6a) possibly due to model biases in surface fluxes in warm sectors. PBLH simulated with TM5 is close to observations in warm sectors but does not help the TM5 group (CT, TM5-4DVAR, and OU) better resolve this metric than other models (Figure 5a). In cold sectors, PCTM, LoFI, and CT better simulate the PBL-to-FT $[\text{CO}_2]$ difference while other models simulate much weaker ones than observations. The overall overestimates in PBLH in OU, TM5-4DVAR, MERRA-2-driven GEOS-Chem models, and CAMS may explain the weaker PBL-to-FT difference but cannot explain the stronger PBL-to-FT difference in LoFI and CT in cold sectors. Biases in surface fluxes and convection mixing thus may be responsible for some models' failures in capturing the PBL-to-FT metric in summer. Cloudy conditions within weather systems may tend to hide CO_2 signals from satellites and hinder regular aircraft measurements, which may introduce biases in flux inversions (e.g., Parazoo et al., 2012). Parameterized convection in LoFI and PCTM may be too weak in summer, leading to too stronger PBL-to-FT differences in cold sectors. In winter, the OCO-2 MIP models have large model discrepancies in the PBL-to-FT metric. CAMS and LoFI capture observed PBLH better than other models in both warm and cold sectors but do not better resolve the PBL-to-FT differences. TM5 and GEOS-Chem overall underestimate PBLH in winter. These model biases in PBLH, however, cannot explain the model-data mismatch in PBL-to-FT $[\text{CO}_2]$ difference.

In autumn and spring, the models tend to simulate weaker PBL-to-FT $[\text{CO}_2]$ differences than observed in both warm and cold sectors (Figures 6c-d), which may be linked to the high-biased model PBLH in autumn and low-biased PBLH in spring (Figures 5c-d). In autumn, overestimated PBL mixing in models can lead to an overestimate of the dilution of positive CO_2 surface fluxes, causing smaller PBL-to-FT $[\text{CO}_2]$ differences in models relative to observations (Figure 6c); In late spring when photosynthesis begins, the underestimates in PBL mixing in models lead to unrealistically low $[\text{CO}_2]$ within the PBL (Figure 2d), resulting in weaker positive (or sometimes even negative) PBL-to-FT $[\text{CO}_2]$ difference compared to observations (Figure 6d).

The PBL-to-FT differences simulated with OCO-2 MIP models appear to be independent of transport model resolution. For example, the fine resolution of $0.5^\circ \times 0.625^\circ$ doesn't help LoFI better resolve PBL-to-FT difference than other models in most seasons, suggesting bias in vertical mixing processes and/or biases in fluxes in the model. OU ($6^\circ \times 4^\circ$) better resolves the PBL-to-FT differences than models like CT in warm sectors in all seasons. This may be because convection parameterization in TM5 can be more effective in venting the boundary layer at a coarse transport model resolution (e.g., $6^\circ \times 4^\circ$) than at a fine resolution (e.g., $1^\circ \times 1^\circ$) (Krol et al., 2005). Meanwhile, comparisons between the NA ($1^\circ \times 1^\circ$) and global ($2^\circ \times 3^\circ$) simulations with CT show negligible impacts of transport model resolution on PBL-to-FT $[\text{CO}_2]$ differences in all seasons (Figure 6).

5 Case study of frontal events

To further investigate the model skills in simulating frontal structures of $[\text{CO}_2]$ and identify model discrepancies in surface CO_2 fluxes in more detail, it is instructive to have a closer look at typical frontal events. Here, we select four frontal events during ACT campaigns and focus on examining a summer frontal case over the mid-west U.S. (section 5.1) and a winter case in the northeast U.S. (section 5.2). Analyses of the cases in autumn and spring can be found in Figures S10–S13 in the supporting materials.

5.1 A summer event on August 4, 2016

On August 4, 2016, the ACT-America team flew across a strong cold front in the mid-west region. The cold front moved from northwest to southeast during the afternoon of August 4. The surface front was moving slowly southward throughout the mid-west. Areas ahead of the cold front, such as Nebraska and Minnesota, had showers and storms (Pal and Davis, 2021). During this RF, aircraft flew over Missouri, Kansas, and Nebraska within the warm sector of the frontal system and over South Dakota in the cold sector (Figure S1a).

Observed frontal and PBL-to-FT $[\text{CO}_2]$ features in this event are consistent with the seasonal averages in summer. Latitude-altitude curtains of observed $[\text{CO}_2]$ over the mid-west (Figure 7) clearly show the cold $[\text{CO}_2]$ -poor air encounters warm $[\text{CO}_2]$ -rich air at 41.5°N at the surface and lifts the warm air. Enhanced $[\text{CO}_2]$ was observed by B200 over Missouri and Kansas in the warm sector near the surface (405–415 ppm) relative to the FT (400–405 ppm). The enhanced PBL- $[\text{CO}_2]$ might be due to positive biological fluxes caused by respiration associated with lower radiative forcing under cloud covers in the warm sector and the accumulation of fossil-fuel CO_2 (Pal et al., 2020a). In comparison, in the cold sector (41.5°N – 45°N), observed $[\text{CO}_2]$ is lower within the PBL (385–395 ppm) than the FT (395–400 ppm), resulting from net biological uptake by up-wind boreal forests in Canada and croplands in the mid-west on the passage of air masses (e.g., Miles et al., 2012).

It is encouraging that most of the models reproduce much of the observed frontal structure in this case with equatorward flows of cold low- $[\text{CO}_2]$ air from the north and poleward flows of warm high- $[\text{CO}_2]$ air (Figure 7). All models can capture the fact that the cold front lifts the air in the warm sector near the surface. However, the models underestimate $[\text{CO}_2]$ observed in the warm sector air especially in the PBL, suggesting that biological fluxes in the warm sector are biased toward too much net CO_2 uptake in these models. LoFI, CSU, and CAMS relatively better simulate high PBL- $[\text{CO}_2]$ in the warm sector than other models. These three models all show positive surface fluxes over the southern states such as New Mexico and Texas (Figure 8) where warm air passes according to back-trajectory analysis (figures not shown), whereas models with large negative fluxes in the warm sector (e.g., CT and CMS-Flux) underestimate observed $[\text{CO}_2]$.

Large discrepancies in simulated PBL- $[\text{CO}_2]$ exist in the cold sector and near the

frontal boundary, consistent with the findings in section 4. CT, TM5-4DVAR, PCTM, and LoFI better capture depleted $[\text{CO}_2]$ in the air masses transporting equatorward from the north in the cold sector relative to other models. The boundary between high- and low- $[\text{CO}_2]$ air masses is misplaced in GEOS-Chem models (CSU, Ames, UT, and CMS-Flux) and CAMS in the PBL but is reasonably well-simulated in the FT. GEOS-Chem models overestimate PBL- $[\text{CO}_2]$ by up to 10 ppm near the surface frontal boundary (at around 42°N ; Figure 7). These models are biased low in wind speed in cold sectors (Figure 4) and simulate weak sinks (or even net sources) in the northern US (South Dakota, North Dakota, Montana, and Wyoming) and southern Canada (Figure 8) where cold air originated during this frontal event according to back-trajectory analysis (figures not shown). The misplacement of the frontal boundary in GEOS-Chem may be due to horizontal transport biases in the cold sector as a result of model underestimates in wind speed and/or a blurry boundary as a result of low model resolutions.

The simulated spatial structures of $[\text{CO}_2]$ in different data source experiments (except OG) do not noticeably differ in several models (OU, CMS, and CAMS) which have similar spatial patterns but different magnitudes in surface fluxes over N.A. (Figure 8). The results suggest the frontal $[\text{CO}_2]$ difference likely to be strongly related to the spatial distribution of fluxes. The OG experiment with most models fails to simulate the $[\text{CO}_2]$ frontal structures due to $[\text{CO}_2]$ underestimates in the warm sector (biased low by up to 20 ppm) which is likely caused by unrealistically large sinks over the eastern U.S. (Figure 8d).

5.2 A winter event on March 7, 2017

On March 7, 2017, C130 and B200 were flying in the northeast across a strong cold front (Figure S1b). The surface front was over Ohio and migrating eastward. Regions behind the front had relatively clear weather conditions, whereas strong southerly flow and heavy rain ahead of the cold front were observed along the flight path over Ohio within the warm sector (Pal and Davis, 2021). The flight tracks crossed the warm sector in Virginia and Ohio (37°N – 41°N) and the cold sector in Indiana, Michigan, and Wisconsin (41°N – 43°N ; Figure S1b).

Observed frontal and PBL-to-FT $[\text{CO}_2]$ differences during this winter frontal event are also similar to the seasonal statistics. Measurements show enhanced $[\text{CO}_2]$ within the PBL (402–416 ppm) compared to the FT (385–405 ppm) in both warm and cold sectors (Figure 9) as the northern mid-latitudes are a net source of CO_2 due to ecosystem respiration (Figure 10) and anthropogenic emissions in boreal winter. $[\text{CO}_2]$ in cold sectors is much higher than in the warm sector in both the PBL and FT (e.g., 410–416 ppm vs. 400–405 ppm in the PBL). Similar frontal $[\text{CO}_2]$ structures were also observed in autumn and spring (Figures S10–S13).

The models simulate similar locations of the frontal boundary and overall capture the observed features of higher $[\text{CO}_2]$ in the cold sector relative to the warm sector and higher $[\text{CO}_2]$ in the PBL relative to the FT (Figure 9). This

is because the models simulate positive fluxes over the majority of the U.S. in winter (Figure 10). However, the models tend to underestimate $[\text{CO}_2]$ in cold sectors (41 °N–43 °N) and overestimate $[\text{CO}_2]$ in warm sectors in the PBL (37 °N–41 °N), leading to underestimates in frontal differences. Large model discrepancies are found within the PBL in both warm and cold sectors. LoFI and CMAS are closer to measurements in the cold sector PBL compared to other models, contributing to their better performance in simulating the warm-cold contrast. The underestimates in PBLH in the warm sector may contribute to the overestimates in PBL- $[\text{CO}_2]$ in models. In the cold sector (Indiana, Michigan, and Wisconsin), the higher PBLHs in the TM5 group (CT, TM5-4DVAR, and OU) result in much lower $[\text{CO}_2]$ than the GEOS-Chem models even though they have larger CO_2 sources over the region (Figure 10).

Models with large positive fluxes over the warm sector in the eastern U.S. in winter (e.g., $>300 \text{ g C m}^{-2} \text{ y}^{-1}$ in TM5-4DVAR and LoFI) better capture observed frontal $[\text{CO}_2]$ differences in this case and seasonal averaged frontal differences. We, therefore, suggest fluxes over Virginia and Ohio to be lower than $300 \text{ g C m}^{-2} \text{ y}^{-1}$ in winter 2017. In CMS and CAMS, experiments with different $[\text{CO}_2]$ data sources have similar spatial patterns of fluxes (Figure 9). Thus, different sets of flux inversions don’t noticeably change the simulated frontal structures of $[\text{CO}_2]$ in these models. In other models (e.g., CSU and Ames), spatial patterns of flux noticeably change across experiments constrained by different CO_2 data, leading to diverging model performances in $[\text{CO}_2]$ frontal structure in winter. For example, LNLG and OG experiments with CSU, which have large positive fluxes over the eastern U.S., overestimate warm-sector $[\text{CO}_2]$ by up to 10 ppm, whereas the IS experiment with weak CO_2 sources (or sinks) over the eastern U.S. much better simulate the observed $[\text{CO}_2]$ with the warm sector.

6. Discussion and Conclusions

Our sector-dependent evaluation of simulated $[\text{CO}_2]$ reveals that averaging the model simulations within an entire weather system may cancel out model biases in different sectors of the weather system (i.e., warm and cold sectors) and consequently hide possible model errors in CO_2 transport or/and fluxes. In OG and LNLGOGIS experiments, the unbiased PBL- $[\text{CO}_2]$ is a result of cancelling out the model underestimates in warm sectors with the overestimates in cold sectors of weather systems, which hides possible surface flux biases in the two experiments. In comparison, although the IS and LNLG ensembles are biased in $[\text{CO}_2]$ across weather systems, they better capture observed $[\text{CO}_2]$ in both warm and cold sectors, suggesting the two experiments may have more reasonable flux estimates than the other data experiments. The findings indicate the benefits of using weather-aware metrics to evaluate the performance of inversion models via their posterior $[\text{CO}_2]$ estimates.

We find that frontal $[\text{CO}_2]$ structures exist in all OCO-2 MIP members and the models generally simulate the correct signs of observed frontal and PBL-to-FT $[\text{CO}_2]$ differences (i.e. positive frontal differences in summer and negative in the other seasons; weak positive/negative PBL-to-FT differences in summer

and positive in the other seasons). However, we see frequent underestimates of frontal differences and consistent underestimates of vertical differences in the OCO-2 MIP models. Some transport biases (e.g., biased PBLH and wind speed) are identified, but no simple and direct connection can be made to many of the model-measurement mismatches in frontal $[\text{CO}_2]$ structures.

Several explanations coexist for the model biases in frontal $[\text{CO}_2]$ structures. First, posterior fluxes are possibly biased in the models. Recent studies (Cui et al., 2021a; Cui et al., 2021b) suggested that the seasonality of net ecosystem CO_2 exchange (NEE) might be systematically underestimated in most OCO-2 MIP models. That is, NEE is possibly overestimated in summer (not enough net uptake) whereas underestimated in winter (not enough net respiration). These possible surface flux biases can lead to weaker spatial $[\text{CO}_2]$ differences compared to observations in all seasons. This, however, might not be true for all of the global inversions.

Second, the models may persistently overestimate vertical mixing, leading to weaker vertical $[\text{CO}_2]$ differences in most models in the four seasons. The identified model biases in PBLH (e.g., high-biased PBLH in summer and autumn) may explain some of the underestimates in vertical differences in summer and autumn, but have difficulties in explaining model biases in winter and spring when PBLH are underestimated in most models. Parameterized vertical mixing, attempting to represent mixing due to convective clouds, might be overestimated in most of the transport models. Too vigorous vertical mixing can dilute the impact of NEE on PBL- $[\text{CO}_2]$, causing underestimates of vertical $[\text{CO}_2]$ differences and contribute to the underestimates of cross-frontal $[\text{CO}_2]$ differences in the PBL. Sensitivity studies running global models at multiple resolutions, and comparisons with cloud-resolving regional models will be helpful to guide improvements in parameterizations of PBL and cloud-convective mixing.

Third, horizontal mixing is biased in some models. In GEOS-Chem-based models, horizontal winds appear to be underestimated within weather systems. This biased horizontal mixing contributes to the underestimated cross-frontal $[\text{CO}_2]$ differences in GEOS-Chem models. By comparing zonal-averaged transport differences in summer and winter seasons between TM5 and GEOS-Chem driven by the same surface CO_2 fluxes, Schuh et al (2019) found that GEOS-Chem traps flux signals close to the surface, while TM5 has more vigorous mixing vertically. These large-scale seasonal transport differences appear to be different from our findings that TM5-based systems better simulate the cross-frontal differences within individual weather systems at a subcontinental scale. However, the underestimated frontal differences in GEOS-Chem documented here may be related to a tendency of GEOS-Chem to keep flux impacts near the surface and mix $[\text{CO}_2]$ less across latitudes.

Forth, analyses of individual front events suggest that blurry boundaries due to relatively low model resolution may contribute to the underestimates in cross-frontal differences in GEOS-Chem-based models. The blurry boundaries may also contribute to the underestimates in wind speed within frontal systems in

GEOS-Chem. Finer model resolutions appear to aid in resolving the observed frontal $[\text{CO}_2]$ differences in summer, winter, and autumn, generally consistent with the findings in previous studies of GHG transport (e.g., Agustí-Panareda et al., 2019; Stanevich et al., 2020). However, they have negligible impacts on the PBL-to-FT $[\text{CO}_2]$ difference in OCO-2 MIP models. We, therefore, argue that horizontal resolution may only explain some of the underestimates in cross-frontal differences in GEOS-Chem models and is not a dominant source of model biases in frontal $[\text{CO}_2]$ structures.

Comparison among OCO-2 MIP models constrained by different suites of surface CO_2 fluxes suggest that data sources in different experiments are unlikely to explain most of the model biases in frontal $[\text{CO}_2]$ structures. The varying data source makes a relatively small difference in the simulated frontal difference and is not the dominant source of biases and model discrepancies in frontal $[\text{CO}_2]$ structures. Among the OCO-2 MIP models, inversion methods, prior fluxes, data sources, and transport models vary considerably. These factors may intertwine and all influence the estimated $[\text{CO}_2]$ and fluxes. More work (e.g., controlled experiments in which multiple transport models constrained with the same fluxes or one common transport model driven by different suites of fluxes) are needed to isolate the impacts of prior fluxes, inversion methods, and transport models on the biases in frontal $[\text{CO}_2]$ structures in the OCO-2 MIP models found in this study.

There are several caveats to our analyses that need further efforts in the future. First, the varying definitions of the top of PBL in OCO-2 MIP models may complicate our assessment of the impacts of PBLH on PBL-to-FT $[\text{CO}_2]$ differences. LMDz (used in CAMS) and TM5 (used in CT, OU, and TM5-4DVAR) define PBLH as the first height where the bulk Richardson number (Ri_b) exceeds a critical value (Holtslag and Boville, 1993). The critical value of Ri_b (Ri_{bc}) is set to be 0.4 in LMDz (Locatelli et al., 2013) but 0.3 in TM5 (Koffi et al., 2016). The higher Ri_{bc} in LMDz might lead to an overestimate in PBLH (Locatelli et al., 2015) relative to TM5 even though the two models are driven by similar re-analysis fields. Meanwhile, all GEOS-Chem models participated in the OCO-2 MIP use the default full-mixing PBL scheme (Bey et al., 2001), which directly takes PBLH values from MERRA-2 or GEOS-FP (calculated based on the total eddy diffusion coefficient of heat) and assumes instantaneous vertical mixing from the surface to the top of the PBL. This full-mixing scheme tends to overestimate vertical mixing except at places where the PBL is extremely unstable (Lin and McElroy, 2010). Further studies of the extent to which these different definitions of boundary layer top will influence the PBLH in OCO-2 MIP models are needed. Second, ACT-America was carried out in three key regions: South, Midwest, and Mid-Atlantic (Davis et al., 2021), which are characterized by different ecosystems and fluxes. Evaluation of regional differences in the OCO-2 MIP model performance might lead to more details of model biases in transport and surface fluxes in the three separate regions. Finally, the analyses of typical front events suggest some possible errors in surface CO_2 fluxes, but whether these flux errors are directly linked to the transport biases in models

remains unknown. More research is needed to quantify the link between accurate simulation of weather systems and accurate estimates of surface fluxes at regional and global scales in inversion systems.

Our multi-model intercomparison constrained by intensive aircraft measurements unveils large model discrepancies in simulating CO₂ weather and points out several possible transport issues in current global models. The biased transport imposes challenges to achieving accurate flux estimates with global inversion systems. Large carbon sinks were reported recently in the southwest and northeast provinces of China (Wang et al., 2020). These estimates, however, may be subject to uncertainties due to possible transport biases in the inversion system used in that study (Schuh et al., 2020). Similar targeted aircraft measurements and multi-model evaluation as presented in our study are needed to improve the current understanding of transport uncertainties in inversion systems over other mid-latitude regions such as China. These future efforts will narrow the uncertainties in estimated CO₂ fluxes at a subcontinental scale and foster our confidence in global flux inversions.

Acknowledgments

This study is funded by the National Aeronautics and Space Administration (NASA) under awards NNX15AG76G (Penn State), NNX15AJ07G (Colorado State), 80NSSC19K0730 (Texas Tech). Co-author SP is also supported by a faculty start-up grant from Texas Tech. The ACT-American data is publicly available at the NASA Langley Research Centre (LaRC) airborne science data archive (<https://www-air.larc.nasa.gov/missions/ACT-America>) and the Oak Ridge National Laboratory (ORNL) database (<https://actamerica.ornl.gov>). OCO-2 MIP model results are available at <https://www.esrl.noaa.gov/gmd/ccgg/OCO2/>. We thank NASA Headquarters and NASA’s Airborne Sciences Program and Earth System Science Pathfinder Program Office for their support of the ACT-America mission. We are also grateful to the ACT data management team at NASA LaRC and ORNL for their work, and to the ACT flight and instrument crews for their extensive field work. The authors declare that they have no conflict of interest. The statements, findings, and conclusions are those of the author(s) and should not be construed as the views of the agencies.

References

- <https://doi.org/10.5194/acp-19-7347-2019>
- <https://doi.org/10.5194/amt-7-647-2014>
- <https://doi.org/10.5194/acp-10-4145-2010>
- <https://doi.org/10.1002/2016JD025109>
- <https://doi.org/10.5194/acp-13-8695-2013>
- <https://doi.org/10.1029/2001JD000807>
- <https://doi.org/10.1029/2019EA000847>

<https://doi.org/10.1029/2005jd006390>
<https://doi.org/10.1029/2006GL025910>
<https://doi.org/10.5194/acp-19-9797-2019>
<https://doi.org/10.1002/2017jd027836>
<https://doi.org/https://doi.org/10.1029/2020JD034406>
<https://doi.org/10.1175/bams-d-20-0300.1>
[https://doi.org/10.1175/1520-0426\(2000\)017](https://doi.org/10.1175/1520-0426(2000)017)
<https://doi.org/10.5194/bg-8-3263-2011>
<https://doi.org/10.1038/376240a0>
<https://doi.org/10.1034/j.1600-0889.1999.t01-1-00010.x>
<https://doi.org/10.5194/amt-10-549-2017>
[https://doi.org/10.1175/1520-0442\(1993\)006](https://doi.org/10.1175/1520-0442(1993)006)
<https://doi.org/10.1002/2014JD022962>
<https://doi.org/10.1016/j.ecolmodel.2012.02.004>
[https://doi.org/10.1175/1520-0469\(2004\)061](https://doi.org/10.1175/1520-0469(2004)061)
<https://doi.org/10.5194/gmd-9-3137-2016>
<https://doi.org/10.5194/acp-5-417-2005>
<https://doi.org/10.1016/j.atmosenv.2011.09.068>
<https://doi.org/10.1016/j.atmosenv.2010.02.009>
<https://doi.org/10.1126/science.aam5690>
<https://doi.org/10.5194/acp-13-9917-2013>
<https://doi.org/10.5194/acp-15-9765-2015>
<https://doi.org/10.1029/2011JG001781>
<https://doi.org/10.5194/angeo-28-825-2010>
<https://doi.org/10.3334/ORNLDAAAC/1862>
<https://doi.org/10.1029/2019jd030526>
<https://doi.org/10.3334/ORNLDAAAC/1825>
<https://doi.org/10.1029/2011GL047238>
<https://doi.org/10.5194/acp-8-7239-2008>
<https://doi.org/10.5194/acp-12-6405-2012>
<https://doi.org/10.1029/2007GB003081>

<https://doi.org/10.1073/pnas.0708986104>
<https://doi.org/10.5194/bg-10-6699-2013>
<https://doi.org/10.5194/acp-19-13267-2019>
<https://doi.org/10.5194/gmd-11-4489-2018>
<https://doi.org/10.5194/bg-7-1625-2010>
<https://doi.org/10.1029/2018gb006086>
<https://doi.org/10.5194/gmd-13-3839-2020>
<https://doi.org/10.1126/science.1137004>
<https://doi.org/10.1002/2014jd022591>
<https://doi.org/10.1038/s41586-020-2849-9>
<https://doi.org/10.1029/2006JD007410>
<https://doi.org/10.1029/2020EA001634>
<https://doi.org/10.5194/acp-2020-496>
<https://doi.org/10.1002/joc.3370040307>
<https://doi.org/10.1098/rsta.2010.0313>
<https://doi.org/10.5194/gmd-11-305-2018>
<https://doi.org/10.5194/acp-20-10379-2020>

Agustí-Panareda, A., Diamantakis, M., Massart, S., Chevallier, F., Muñoz-Sabater, J., Barré, J., et al. (2019), Modelling CO₂ weather – why horizontal resolution matters, *Atmos. Chem. Phys.*, *19*(11), 7347-7376. Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., et al. (2014), CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory’s Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, *Atmos. Meas. Tech.*, *7*(2), 647-687. Baker, D. F., Bösch, H., Doney, S. C., O’Brien, D., & Schimel, D. S. (2010), Carbon source/sink information provided by column CO₂ measurements from the Orbiting Carbon Observatory, *Atmos. Chem. Phys.*, *10*(9), 4145-4165. Barnes, E. A., Parazoo, N., Orbe, C., & Denning, A. S. (2016), Isentropic transport and the seasonal cycle amplitude of CO₂, *Journal of Geophysical Research: Atmospheres*, *121*(13), 8106-8124. Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., et al. (2013), Global CO₂ fluxes estimated from GOSAT retrievals of total column CO₂, *Atmos. Chem. Phys.*, *13*(17), 8695-8717. Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., et al. (2001), Global modeling of tropospheric chemistry with

assimilated meteorology: Model description and evaluation, *Journal of Geophysical Research: Atmospheres*, 106(D19), 23073-23095. Campbell, J. F., Lin, B., Dobler, J., Pal, S., Davis, K., Obland, M. D., et al. (2020), Field Evaluation of Column CO₂ Retrievals From Intensity-Modulated Continuous-Wave Differential Absorption Lidar Measurements During the ACT-America Campaign, *Earth and Space Science*, 7(12), e2019EA000847. Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., et al. (2005), Inferring CO₂ sources and sinks from satellite observations: Method and application to TOVS data, *Journal of Geophysical Research: Atmospheres*, 110(D24). Corbin, K. D., & Denning, A. S. (2006), Using continuous data to estimate clear-sky errors in inversions of satellite CO₂ measurements, *Geophysical Research Letters*, 33(12). Crowell, S., Baker, D., Schuh, A. E., Basu, S., Jacobson, A. R., Chevallier, F., et al. (2019), The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network, *Atmos. Chem. Phys.*, 19(15), 9797-9831. Crowell, S., Randolph Kawa, S., Browell, E. V., Hammerling, D. M., Moore, B., Schaefer, K., et al. (2018), On the Ability of Space-Based Passive and Active Remote Sensing Observations of CO₂ to Detect Flux Perturbations to the Carbon Cycle, *Journal of Geophysical Research: Atmospheres*, 123(2), 1460-1477. Cui, Y. Y., Jacobson, A. R., Feng, S., Wesloh, D., Barkley, Z. R., Zhang, L., et al. (2021a), Evaluation of CarbonTracker’s Inverse Estimates of North American Net Ecosystem Exchange of CO₂ From Different Observing Systems Using ACT-America Airborne Observations, *Journal of Geophysical Research: Atmospheres*, 126(12), e2020JD034406. Cui, Y. Y., Zhang, L., Schuh, A., Baker, D. F., Jacobson, A. R., Chevallier, F., et al. (2021b), Seasonal strength of terrestrial net ecosystem CO₂ exchange from the temperate North America are underestimated in most global inversion models, *In preparation*. Davis, K. J., Browell, E. V., Feng, S., Lauvaux, T., Obland, M. D., Pal, S., et al. (2021), The Atmospheric Carbon and Transport (ACT) – America Mission, *Bulletin of the American Meteorological Society*, 1-54. Davis, K. J., Gamage, N., Hagelberg, C. R., Kiemle, C., Lenschow, D. H., & Sullivan, P. P. (2000), An Objective Method for Deriving Atmospheric Structure from Airborne Lidar Observations, *Journal of Atmospheric and Oceanic Technology*, 17(11), 1455-1468. <1455:Aomfda>2.0.Co;2Deng, F., & Chen, J. M. (2011), Recent global CO₂ flux inferred from atmospheric CO₂ observations and its regional analyses, *Biogeosciences*, 8(11), 3263-3281. Denning, A. S., Fung, I. Y., & Randall, D. (1995), Latitudinal gradient of atmospheric CO₂ due to seasonal exchange with land biota, *Nature*, 376(6537), 240-243. Denning, A. S., Takahashi, T., & Friedlingstein, P. (1999), Can a strong atmospheric CO₂ rectifier effect be reconciled with a “reasonable” carbon budget?, *Tellus B*, 51(2), 249-253. Eldering, A., O’Dell, C. W., Wennberg, P. O., Crisp, D., Gunson, M. R., Viatte, C., et al. (2017), The Orbiting Carbon Observatory-2: first 18 months of science data products, *Atmos. Meas. Tech.*, 10(2), 549-563. Gaudet, B. J., Davis, K. J., Pal, S., Jacobson, A. R., Schuh, A., Lauvaux, T., et al. (2020), Regional-scale, sector-specific evaluation of global CO₂ inversion models using aircraft data from the ACT-America project, *Journal of Geophysical Research: Atmospheres*, in review. Holtslag, A. A. M., & Boville, B. A. (1993), Local Versus Nonlocal

Boundary-Layer Diffusion in a Global Climate Model, *Journal of Climate*, 6(10), 1825-1842. <1825:Lvnbl>2.0.Co;2Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., et al. (2015), An intercomparison of inverse models for estimating sources and sinks of CO₂ using GOSAT measurements, *Journal of Geophysical Research: Atmospheres*, 120(10), 5253-5266. Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., et al. (2012), North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison, *Ecological Modelling*, 232, 144-157. Hurwitz, M. D., Ricciuto, D. M., Bakwin, P. S., Davis, K. J., Wang, W., Yi, C., et al. (2004), Transport of Carbon Dioxide in the Presence of Storm Systems over a Northern Wisconsin Forest, *Journal of the Atmospheric Sciences*, 61(5), 607-618. <0607:Tocdit>2.0.Co;2Koffi, E. N., Bergamaschi, P., Karstens, U., Krol, M., Segers, A., Schmidt, M., et al. (2016), Evaluation of the boundary layer dynamics of the TM5 model over Europe, *Geosci. Model Dev.*, 9(9), 3137-3160. Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., et al. (2005), The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, *Atmos. Chem. Phys.*, 5(2), 417-432. Lee, T. R., De Wekker, S. F. J., Andrews, A. E., Kofler, J., & Williams, J. (2012), Carbon dioxide variability during cold front passages and fair weather days at a forested mountaintop site, *Atmospheric Environment*, 46, 405-416. Lin, J.-T., & McElroy, M. B. (2010), Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: Implications to satellite remote sensing, *Atmospheric Environment*, 44(14), 1726-1739. Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., et al. (2017), Contrasting carbon cycle responses of the tropical continents to the 2015-2016 El Niño, *Science*, 358(6360), eaam5690. Locatelli, R., Bousquet, P., Chevallier, F., Fortems-Cheney, A., Szopa, S., Saunois, M., et al. (2013), Impact of transport model errors on the global and regional methane emissions estimated by inverse modelling, *Atmos. Chem. Phys.*, 13(19), 9917-9937. Locatelli, R., Bousquet, P., Saunois, M., Chevallier, F., & Cressot, C. (2015), Sensitivity of the recent methane budget to LMDz sub-grid-scale physical parameterizations, *Atmos. Chem. Phys.*, 15(17), 9765-9780. Miles, N. L., Richardson, S. J., Davis, K. J., Lauvaux, T., Andrews, A. E., West, T. O., et al. (2012), Large amplitude spatial and temporal gradients in atmospheric boundary layer CO₂ mole fractions detected with a tower-based network in the U.S. upper Midwest, *Journal of Geophysical Research: Biogeosciences*, 117(G1). Pal, S., Behrendt, A., & Wulfmeyer, V. (2010), Elastic-backscatter-lidar-based characterization of the convective boundary layer and investigation of related statistics, *Ann. Geophys.*, 28(3), 825-847. Pal, S., & Davis, K. J. (2021), ACT-America Field Campaign Catalogue. ORNL DAAC, Oak Ridge, Tennessee, USA. Pal, S., Davis, K. J., Lauvaux, T., Browell, E. V., Gaudet, B. J., Stauffer, D. R., et al. (2020a), Observations of Greenhouse Gas Changes Across Summer Frontal Boundaries in the Eastern United States, *Journal of Geophysical Research: Atmospheres*, 125(5), e2019JD030526. Pal, S., Davis, K. J., Pauly, R. M., McGill, M. J., Campbell, L., Hoffman, K., et al. (2020b), ACT-America: CPL-derived Atmospheric Boundary Layer Top Height, Eastern US, 2016-2018. ORNL DAAC, Oak Ridge,

Tennessee, USA. Parazoo, N. C., Denning, A. S., Berry, J. A., Wolf, A., Randall, D. A., Kawa, S. R., et al. (2011), Moist synoptic transport of CO₂ along the mid-latitude storm track, *Geophysical Research Letters*, *38*(9). Parazoo, N. C., Denning, A. S., Kawa, S. R., Corbin, K. D., Lokupitiya, R. S., & Baker, I. T. (2008), Mechanisms for synoptic variations of atmospheric CO₂ in North America, South America and Europe, *Atmos. Chem. Phys.*, *8*(23), 7239-7254. Parazoo, N. C., Denning, A. S., Kawa, S. R., Pawson, S., & Lokupitiya, R. (2012), CO₂ flux estimation errors associated with moist atmospheric processes, *Atmos. Chem. Phys.*, *12*(14), 6405-6416. Patra, P. K., Law, R. M., Peters, W., Rödenbeck, C., Takigawa, M., Aulagnier, C., et al. (2008), TransCom model simulations of hourly atmospheric CO₂: Analysis of synoptic-scale variations for the period 2002–2003, *Global Biogeochemical Cycles*, *22*(4). Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., et al. (2021), Four years of global carbon cycle observed from OCO-2 version 9 and in situ data, and comparison to OCO-2 v7, *Atmos. Chem. Phys.*, *submitted*. Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., et al. (2007), An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proceedings of the National Academy of Sciences*, *104*(48), 18925-18930. Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., et al. (2013), Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions, *Biogeosciences*, *10*(10), 6699-6720. Philip, S., Johnson, M. S., Potter, C., Genovesse, V., Baker, D. F., Haynes, K. D., et al. (2019), Prior biosphere model impact on global terrestrial CO₂ fluxes estimated from OCO-2 retrievals, *Atmos. Chem. Phys.*, *19*(20), 13267-13287. Remaud, M., Chevallier, F., Cozic, A., Lin, X., & Bousquet, P. (2018), On the impact of recent developments of the LMDz atmospheric general circulation model on the simulation of CO₂ transport, *Geosci. Model Dev.*, *11*(11), 4489-4513. Schuh, A. E., Denning, A. S., Corbin, K. D., Baker, I. T., Uliasz, M., Parazoo, N., et al. (2010), A regional high-resolution carbon flux inversion of North America for 2004, *Biogeosciences*, *7*(5), 1625-1644. Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., et al. (2019), Quantifying the Impact of Atmospheric Transport Uncertainty on CO₂ Surface Flux Estimates, *Global Biogeochemical Cycles*, *33*(4), 484-500. Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., et al. (2020), Chinese Land Carbon Sink Obscured by Transport Model Uncertainty, *Matters Arising*, *submitted*. Stanevich, I., Jones, D. B. A., Strong, K., Parker, R. J., Boesch, H., Wunch, D., et al. (2020), 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*(9), 3839-3862. Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., et al. (2007), Weak Northern and Strong Tropical Land Carbon Uptake from Vertical Profiles of Atmospheric CO₂, *Science*, *316*(5832), 1732-1735. Sweeney, C., Karion, A., Wolter, S., Newberger, T., Guenther, D., Higgs, J. A., et al. (2015), Seasonal climatology of CO₂ across North America from aircraft measurements in the NOAA/ESRL Global Greenhouse Gas Reference Network, *Journal of Geophysical Research: Atmospheres*,

120(10), 5155-5190. Wang, J., Feng, L., Palmer, P. I., Liu, Y., Fang, S., Bösch, H., et al. (2020), Large Chinese land carbon sink estimated from atmospheric carbon dioxide data, *Nature*, 586(7831), 720-723. Wang, J.-W., Denning, A. S., Lu, L., Baker, I. T., Corbin, K. D., & Davis, K. J. (2007), Observations and simulations of synoptic, regional, and local variations in atmospheric CO₂, *Journal of Geophysical Research: Atmospheres*, 112(D4). Wei, Y., Shrestha, R., Pal, S., Gerken, T., Feng, S., McNelis, J., et al. (2021), Atmospheric Carbon and Transport – America (ACT-America) Datasets: Description, Management, and Delivery, *Earth and Space Science*, in press, e2020EA001634. Weir, B., Ott, L. E., Collatz, G. J., Kawa, S. R., Poulter, B., Chatterjee, A., et al. (2020), Calibrating satellite-derived carbon fluxes for retrospective and near real-time assimilation systems, *Atmos. Chem. Phys. Discuss.*, 2020, 1-26. Whittaker, L. M., & Horn, L. H. (1984), Northern Hemisphere extratropical cyclone activity for four mid-season months, *Journal of Climatology*, 4(3), 297-310. Wofsy, S. C. (2011), HIPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1943), 2073-2086. Wofsy, S. C., Afshar, S., Allen, H. M., Apel, E. C., Asher, E. C., Barletta, B., et al. (2018), ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, edited, ORNL Distributed Active Archive Center, doi:10.3334/ORNLDAAAC/1581. Yu, K., Keller, C. A., Jacob, D. J., Molod, A. M., Eastham, S. D., & Long, M. S. (2018), Errors and improvements in the use of archived meteorological data for chemical transport modeling: an analysis using GEOS-Chem v11-01 driven by GEOS-5 meteorology, *Geosci. Model Dev.*, 11(1), 305-319. Zhang, L., Lin, M., Langford, A. O., Horowitz, L. W., Senff, C. J., Klovnski, E., et al. (2020), Characterizing sources of high surface ozone events in the southwestern US with intensive field measurements and two global models, *Atmos. Chem. Phys.*, 20(17), 10379-10400.

Tables and Figures

Table 1. Key model information of each OCO-2 MIP member.

OCO2 Model	Transport model	Meteorology	Horizontal resolution of transport model	Inv
CT2019	TM5	ERA-Interim	2°×3° global; 1°×1° in N.A.	En
OU	TM5	ERA-Interim	4°×6°	4D
TM5-4DVAR	TM5	ERA-Interim	2°×3°	4D
CSU	GEOS-Chem	MERRA-2	4°×5°	Ba
Ames	GEOS-Chem	MERRA-2	4°×5°	4D
CMS-Flux	GEOS-Chem	GEOS-FP	4°×5°	4D
UT	GEOS-Chem	GEOS-FP	4°×5°	4D
PCTM	PCTM	MERRA-2	2°×2.5°	4D
LoFI	GEOS-5	MERRA-2	0.5°×0.625°	--
CAMS	LMDz	ERA5	1.875°×3.75°	Va

Table 2. Statistics of frontal CO₂ differences in the PBL and FT and PBL-to-FT CO₂ differences in the warm and cold sectors as observed with aircrafts and simulated with OCO-2 MIP models (unit: ppm). Model results are averaged across five suits of experiments constrained with different data sources.

	OBS	CT	OU	TM5-CSU	Ames	CMS	UT	PCTM	LoFI	CAMS
	(CT_GlobalDVAR)									
Frontal difference (PBL)										
Summer	±8.5 (5.5±0.3)	±0.5	±0.6	±0.9	±0.5	±0.9	±0.6	±0.6	±1.0	±0.4
Winter	±5.1 (-1.1±0.2)	±0.3	±0.1	±0.5	±0.3	±0.4	±0.4	±0.1	±0.4	±0.8
Autumn	±3.5 (-1.6±0.7)	±0.8	±0.6	±0.4	±0.5	±1.0	±0.2	±1.1	±1.1	±0.8
Spring	±3.7 (-0.3±0.4)	±0.4	±0.3	±0.5	±0.4	±0.7	±0.3	±0.7	±0.8	±0.8
Frontal difference (FT)										
Summer	±3.7 (2.3±1.0)	±1.3	±0.7	±0.5	±0.5	±0.4	±0.3	±0.2	±1.7	±0.5
Winter	±1.7 (-0.9±0.7)	±0.7	±0.5	±0.8	±0.6	±0.9	±0.7	±0.3	±1.0	±0.6
Autumn	±2.6 (-0.1±0.2)	±0.3	±0.3	±0.3	±0.6	±0.4	±0.2	±0.7	±0.4	±0.5
Spring	±2.8 (-0.7±0.3)	±0.3	±0.1	±0.4	±0.3	±0.4	±0.1	±0.3	±0.7	±0.4

	OBS	CT	OU	TM5-CSU	Ames	CMS	UT	PCTM	LoFI	CAMS
	(CT_Global)			DVAR						
PBL-to-FT difference (warm sector)										
Summer	1.7 ± 0.7 (-1.4 \pm 0.5)	± 0.5	± 0.4	± 0.7	± 1.0	± 1.1	± 1.4	± 0.5	± 0.9	± 0.5
Winter	3.0 ± 3.3 (3.3 \pm 0.7)	± 0.7	± 1.2	± 0.6	± 1.0	± 0.7	± 0.8	± 0.5	± 0.7	± 1.0
Autumn	3.5 ± 2.4 (2.4 \pm 0.4)	± 0.4	± 0.9	± 0.7	± 1.0	± 0.3	± 0.5	± 3.8	± 3.7	± 1.2
Spring	2.3 ± 1.0 (1.0 \pm 0.2)	± 0.2	± 0.9	± 0.4	± 0.7	± 0.6	± 0.3	± 0.5	± 1.1	± 0.9
PBL-to-FT difference (cold sector)										
Summer	6.2 ± 3.5 (3.5 \pm 0.5)	± 0.6	± 0.2	± 0.5	± 0.4	± 0.9	± 1.6	± 0.7	± 0.7	± 0.6
Winter	3.0 ± 3.7 (3.7 \pm 0.6)	± 0.6	± 1.0	± 1.0	± 1.1	± 1.2	± 1.2	± 0.9	± 0.7	± 1.5
Autumn	5.0 ± 2.3 (2.3 \pm 0.2)	± 0.3	± 0.8	± 0.6	± 0.7	± 1.0	± 0.5	± 2.1	± 2.5	± 1.3
Spring	2.6 ± 0.4 (0.4 \pm 0.3)	± 0.3	± 0.5	± 0.5	± 0.5	± 0.2	± 0.2	± 0.6	± 1.2	± 0.3

Figure 1. Mean CO₂ mole fraction (henceforth, [CO₂]) within the planetary boundary layer (PBL), lower free-troposphere (LFT), and upper free-troposphere (UFT) as observed with aircrafts (black lines) and simulated with OCO-2 MIP members (colored dots) in five experiments (PRIOR, IS, LNLG, OG, and LNLGOGIS) in (a) summer 2016, (b) winter 2017, (c) autumn

2017, and (d) spring 2018. The triangles denote the ensemble mean of OCO-2 MIP models.

Figure 2. Mean $[\text{CO}_2]$ in cold and warm sectors of frontal systems within the planetary boundary layer (PBL), lower free-troposphere (LFT), and upper free-troposphere (UFT) as observed with aircrafts (black lines) and simulated with OCO-2 MIP members (colored dots) in five experiments (PRIOR, IS, LNLG, OG, and LNLGOGIS) in (a) summer 2016, (b) winter 2017, (c) autumn 2017, and (d) spring 2018. The triangles denote the ensemble mean of OCO-2 MIP models.

Figure 3. Frontal $[\text{CO}_2]$ difference (CO_2 mole fraction in the warm sector minus that in the cold sector) within the PBL, LFT, and UFT as observed with aircrafts (black lines) and simulated with OCO-2 MIP members (colored dots; red hollow circles denote CT_Global) in five experiments in (a) summer 2016, (b) winter 2017, (c) autumn 2017, and (d) spring 2018.

Figure 4. Box-whisker plots of wind speed in warm and cold sectors of frontal systems as observed with aircrafts and simulated with different transport models (TM5, GEOS-Chem driven by MERRA2 (GC-MERRA2), GEOS-Chem driven by GEOS-FP (GC-FP), GEOS5, and LMDz) within the PBL, LFT, and UFT in (a) summer 2016, (b) winter 2017, (c) autumn 2017, and (d) spring 2018. Lower and upper sides of the boxes denote lower and upper quartiles, respectively, and the dots and horizontal bars inside the boxes indicate the mean and median, respectively. The lengths of whiskers denote the 5%-95% range.

Figure 5. Same as Figure 4, but for planetary boundary layer depth (PBLH) in (a) summer 2016, (b) winter 2017, (c) autumn 2017, and (d) spring 2018.

Figure 6. Same as Figure 3, but for PBL-to-FT $[\text{CO}_2]$ difference in warm and cold sectors of frontal systems in (a) summer 2016, (b) winter 2017, (c) autumn 2017, and (d) spring 2018.

Figure 7. Latitude-height curtains of model-simulated $[\text{CO}_2]$ (color shading) from the OCO-2 MIP in (a) PRIOR, (b) IS, (c) LNLG, (d) OG, and (e) LNLGOGIS experiments along the flight track crossing the mid-west on August 4, 2016 (purple line in Figure 8). The color-coded circles denote measurements of $[\text{CO}_2]$ by two aircrafts (C130 and B200), the dashed line represents the boundary layer height derived from measurements by the airborne lidar on C130, the solid grey lines denote model-simulated boundary layer height, and the white lines are isentropic surfaces.

Figure 8. Maps of surface CO_2 fluxes (including net biosphere exchanges, air-sea net exchanges, and fire emissions) from the OCO-2 MIP in (a) PRIOR, (b) IS, (c) LNLG, (d) OG, and (e) LNLGOGIS experiments in summer 2016. The purple line represents selected frontal research flight tracks on August 4, 2016.

Figure 9. Same as Figure 7, but for March 7, 2017 along the flight track crossing the northeast U.S. (purple line in Figure 10).

Figure 10. Same as Figure 8, but for winter 2017.