



**HAL**  
open science

## Multi-Season Evaluation of CO<sub>2</sub> Weather in OCO-2 MIP Models

Li Zhang, Kenneth J Davis, Andrew E Schuh, Andrew R Jacobson, Sandip Pal, Yu Yan Cui, David Baker, Sean Crowell, Frederic Chevallier, Marine Remaud, et al.

► **To cite this version:**

Li Zhang, Kenneth J Davis, Andrew E Schuh, Andrew R Jacobson, Sandip Pal, et al.. Multi-Season Evaluation of CO<sub>2</sub> Weather in OCO-2 MIP Models. *Journal of Geophysical Research: Atmospheres*, 2022, 127 (2), pp.e2021JD035457. 10.1029/2021jd035457 . hal-03566530

**HAL Id: hal-03566530**

**<https://hal.science/hal-03566530>**

Submitted on 11 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# JGR Atmospheres

## RESEARCH ARTICLE

10.1029/2021JD035457

### Special Section:

Carbon Weather: Toward the next generation of regional greenhouse gas inversion systems

### Key Points:

- Global inversion systems are able to simulate observed CO<sub>2</sub> frontal differences but with varying degrees of success
- Most global inversion systems underestimate dormant-season frontal and vertical CO<sub>2</sub> differences
- Inversion systems appear to explain more of the model-data differences in CO<sub>2</sub> weather metrics than CO<sub>2</sub> data sources

### Supporting Information:

Supporting Information may be found in the online version of this article.

### Correspondence to:

L. Zhang,  
[alex.zhang@psu.edu](mailto:alex.zhang@psu.edu)

### Citation:

Zhang, L., Davis, K. J., Schuh, A. E., Jacobson, A. R., Pal, S., Cui, Y. Y., et al. (2022). Multi-season evaluation of CO<sub>2</sub> weather in OCO-2 MIP models. *Journal of Geophysical Research: Atmospheres*, 127, e2021JD035457. <https://doi.org/10.1029/2021JD035457>















Received 8 JUL 2021

Accepted 29 DEC 2021

### Author Contributions:

**Conceptualization:** Kenneth J. Davis  
**Data curation:** Li Zhang, Andrew E. Schuh, Andrew R. Jacobson, Yu Yan Cui, David Baker, Sean Crowell, Frederic Chevallier, Marine Remaud, Junjie Liu, Brad Weir, Sajeev Philip, Feng Deng, Sourish Basu  
**Formal analysis:** Li Zhang  
**Funding acquisition:** Kenneth J. Davis  
**Investigation:** Li Zhang, Kenneth J. Davis  
**Methodology:** Li Zhang, Kenneth J. Davis, Andrew E. Schuh, Andrew R. Jacobson, Sandip Pal, David Baker, Sean

## Multi-Season Evaluation of CO<sub>2</sub> Weather in OCO-2 MIP Models

Li Zhang<sup>1</sup> , Kenneth J. Davis<sup>1,2</sup> , Andrew E. Schuh<sup>3</sup> , Andrew R. Jacobson<sup>4</sup> , Sandip Pal<sup>5</sup> , Yu Yan Cui<sup>1</sup> , David Baker<sup>3</sup> , Sean Crowell<sup>6</sup> , Frederic Chevallier<sup>7</sup> , Marine Remaud<sup>7</sup>, Junjie Liu<sup>8,9</sup> , Brad Weir<sup>10,11,12</sup> , Sajeev Philip<sup>13,14</sup>, Matthew S. Johnson<sup>15</sup> , Feng Deng<sup>16</sup> , and Sourish Basu<sup>11,17</sup> 

<sup>1</sup>Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA, <sup>2</sup>Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA, <sup>3</sup>Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA, <sup>4</sup>University of Colorado Boulder and NOAA Earth System Research Laboratory, Boulder, CO, USA, <sup>5</sup>Department of Geosciences, Atmospheric Science Division, Texas Tech University, Lubbock, TX, USA, <sup>6</sup>School of Meteorology, University of Oklahoma, Norman, OK, USA, <sup>7</sup>Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France, <sup>8</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, <sup>9</sup>Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA, <sup>10</sup>Universities Space Research Association, Columbia, MD, USA, <sup>11</sup>NASA Goddard Space Flight Center, Greenbelt, MD, USA, <sup>12</sup>Now at Morgan State University, Baltimore, MD, USA, <sup>13</sup>NASA Academic Mission Services by Universities Space Research Association at NASA Ames Research Center, Mountain View, CA, USA, <sup>14</sup>Now at Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi, India, <sup>15</sup>NASA Ames Research Center, Moffett Field, CA, USA, <sup>16</sup>Department of Physics, University of Toronto, Toronto, ON, Canada, <sup>17</sup>Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

**Abstract** The ability of current global models to simulate the transport of CO<sub>2</sub> by mid-latitude, synoptic-scale weather systems (i.e., CO<sub>2</sub> 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 CO<sub>2</sub> 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 CO<sub>2</sub> 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 CO<sub>2</sub> 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 CO<sub>2</sub> data used in the inversion. The metrics presented here provide new benchmarks regarding the ability of atmospheric inversion systems to reproduce the CO<sub>2</sub> 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 CO<sub>2</sub> observations, atmospheric transport models, CO<sub>2</sub> flux models (emissions and absorption), and mathematical optimization methods to estimate biosphere-atmosphere CO<sub>2</sub> 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 CO<sub>2</sub> 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 CO<sub>2</sub> 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 CO<sub>2</sub> in winter and autumn. The models may underestimate the CO<sub>2</sub> differences between the boundary layer and the free troposphere due to model errors in boundary layer height and surface fluxes. These weather-oriented CO<sub>2</sub> metrics provide benchmarks for testing simulations of the CO<sub>2</sub> structure within cyclones. Future efforts are needed to link these metrics more directly to the accuracy of CO<sub>2</sub> flux estimates.

Crowell, Frederic Chevallier, Marine Remaud, Junjie Liu, Brad Weir, Sajeev Philip, Feng Deng, Sourish Basu

**Project Administration:** Kenneth J. Davis

**Resources:** Li Zhang, Andrew E. Schuh, Andrew R. Jacobson, Yu Yan Cui, David Baker, Sean Crowell, Frederic Chevallier, Marine Remaud, Junjie Liu, Brad Weir, Sajeev Philip, Feng Deng, Sourish Basu

**Software:** Li Zhang, Andrew E. Schuh, Andrew R. Jacobson, David Baker, Sean Crowell, Frederic Chevallier, Marine Remaud, Junjie Liu, Brad Weir, Sajeev Philip, Feng Deng, Sourish Basu

**Supervision:** Kenneth J. Davis

**Validation:** Li Zhang, Kenneth J. Davis, Sandip Pal

**Visualization:** Li Zhang

**Writing – original draft:** Li Zhang

**Writing – review & editing:** Li Zhang, Kenneth J. Davis, Andrew E. Schuh, Andrew R. Jacobson, Sandip Pal, Yu Yan Cui, David Baker, Sean Crowell, Frederic Chevallier, Marine Remaud, Junjie Liu, Brad Weir, Sajeev Philip, Feng Deng, Sourish Basu

## 1. Introduction

Synoptic weather systems in the mid-latitudes modulate the distribution and variability of atmospheric CO<sub>2</sub> through horizontal advection and vertical mixing and thus are an important part of the global greenhouse gas (GHG) transport system (e.g., Barnes et al., 2016; Hurwitz et al., 2004; Pal, Davis, Lauvaux, et al., 2020; Parazoo et al., 2008). CO<sub>2</sub> inversion modeling uses atmospheric transport models, measurements of CO<sub>2</sub> mole fractions (henceforth [CO<sub>2</sub>]), and a prior surface flux estimate to solve for surface fluxes. Misrepresentation of synoptic-scale weather systems in atmospheric transport models can result in biases in regional and global flux inversions (Parazoo et al., 2011; Patra et al., 2008; Wang et al., 2007). However, the ability of current-generation global inversion systems to simulate the [CO<sub>2</sub>] 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 discrepancies at continental to subcontinental scales (e.g., Houweling et al., 2015; Peiro et al., 2021; Peylin et al., 2013). Peylin et al. (2013) showed notable uncertainties (0.17–0.49 PgC/yr) in annual CO<sub>2</sub> 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 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 [CO<sub>2</sub>] 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 [CO<sub>2</sub>] and in annual zonal-mean inverse CO<sub>2</sub> 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.

Some attempts have been made to evaluate model performance of [CO<sub>2</sub>] variability in weather systems with limited tower measurements (e.g., Agustí-Panareda et al., 2019; Parazoo et al., 2011; Wang et al., 2007). Agustí-Panareda et al. (2019) evaluated the performance of a high-resolution global [CO<sub>2</sub>] simulation with point-based comparisons aggregated across tower 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 [CO<sub>2</sub>] across a small number of widely spaced tower observational sites with a regional coupled model. However, none of these studies benefitted from intensive targeted measurements that can resolve the [CO<sub>2</sub>] variability within a given weather system. This is because regular [CO<sub>2</sub>] measurement systems are spatially sparse and insufficient to resolve the [CO<sub>2</sub>] 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) (e.g., Hurwitz et al., 2004; Lee et al., 2012). The long-term NOAA Greenhouse Gases Reference Network Aircraft program (Sweeney et al., 2015), although spanning a large portion of North America, only conducts measurements weekly to monthly and prefers sampling clear-weather days. Satellite column CO<sub>2</sub> measurements ( $\chi_{\text{CO}_2}$ ) 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 & Denning, 2006). Airborne measurement programs, for example, 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 [CO<sub>2</sub>] within synoptic weather systems (Davis et al., 2021). These unprecedented and unique pictures of 4D spatial variations of [CO<sub>2</sub>] fill the gaps in regular in-situ and aircraft measurement networks and serve as a valuable data set for studying the skills of global inversion systems in reproducing observed CO<sub>2</sub> 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, Davis, Lauvaux, et al., 2020). 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. (2021) compared OCO-2 MIP inversions constrained with in situ measurements (IS inversions, version 7) against vertical  $[\text{CO}_2]$  profiles in 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.

In this study, ten modeling 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  $[\text{CO}_2]$  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 observational metrics in Section 2. In Section 3, we present process-based analyses of observed and simulated  $[\text{CO}_2]$  variability in two frontal cases (one in summer and one in winter). We then conduct seasonal statistical analyses of modeled  $[\text{CO}_2]$  in the cold and warm sectors of frontal systems (Section 4) and examine the model performance of ACT observational metrics in four different seasons (Section 5). Finally, we present our discussion and conclusions in Section 6.

## 2. OCO-2 MIP and ACT-America

### 2.1. OCO-2 MIP

The OCO-2 MIP is an international collaboration of modeling 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  $\text{CO}_2$  (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 details 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 reanalyses (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  $[\text{CO}_2]$ : (a) PRIOR, in which models are driven with prior fluxes and no inversion is performed; (b) 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/>); (c) LNLG, inversions that assimilate 10s averaged Land Nadir and Land Glint retrievals (version 9) observed by OCO-2 (Eldering et al., 2017); (d) OG, in which models assimilate 10s averaged OCO-2 Ocean Glint retrievals; and (e) LNLGOGIS, where inversions are constrained with in situ measurements and all OCO-2  $\chi_{\text{CO}_2}$  retrievals over both land and ocean surfaces. Note that the PRIOR experiment with CT, UT, and CAMS and the LNLGOGIS experiment with CSU are not conducted (Table 1). LoFI uses a bias-correction method to adjust a diagnostic collection of surface fluxes to match inventory data and in situ observations, which is different from other OCO-2 MIP inversion systems (Weir et al., 2021). The simulation with LoFI is considered as an IS experiment and other experiments with it are not available. Modeled  $[\text{CO}_2]$  and meteorological variables are sampled from hourly or 3-hourly 4D gridded simulations at the measurement times and locations using linear interpolations in time (from hourly/3-hourly to 5-s/8-s) and bilinear (in horizontal direction) and log-linear (in vertical direction) interpolations in space (from model grids to specific locations).

**Table 1**  
Key Model Information of Each OCO-2 MIP Member

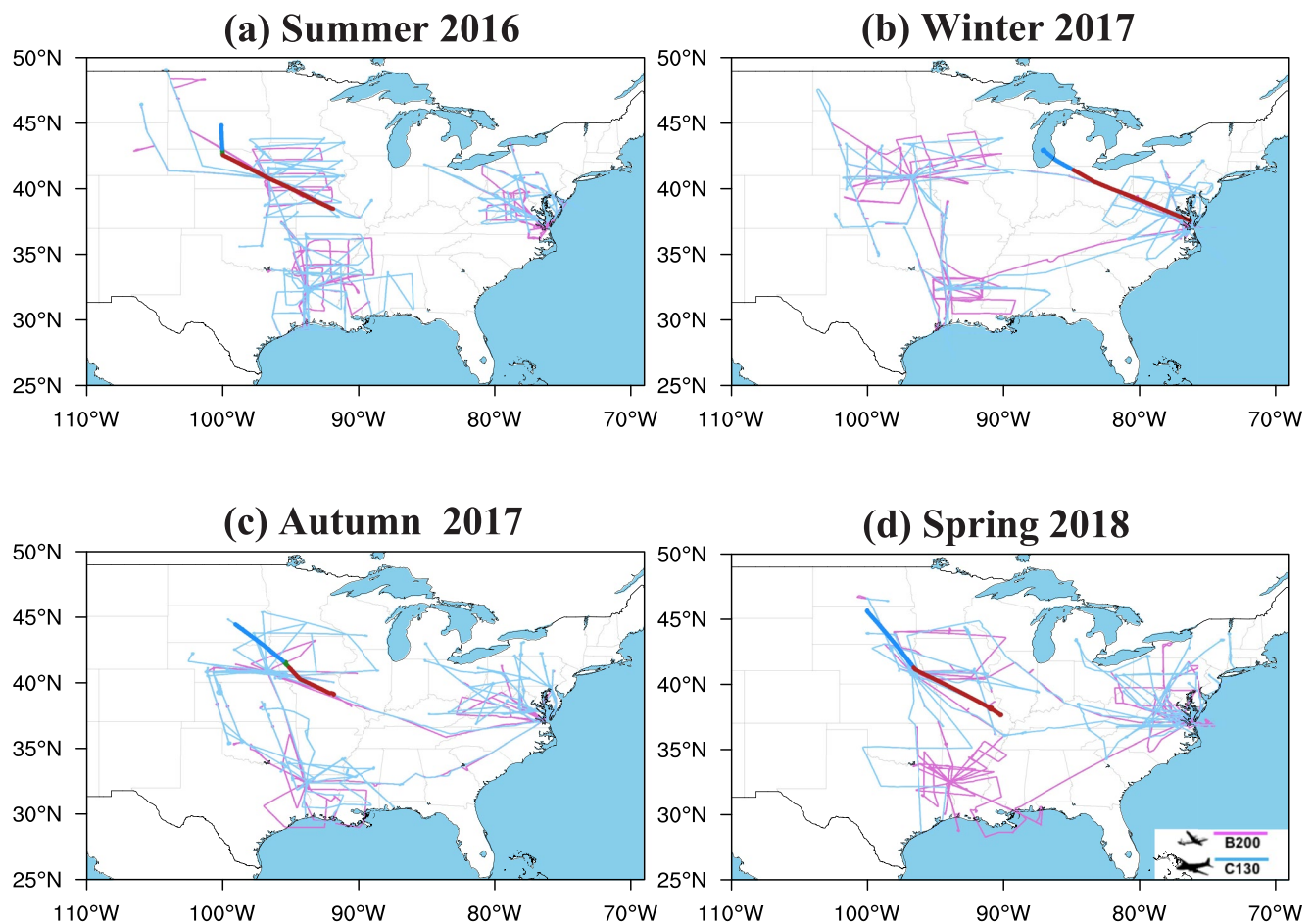
OCO-2 model	Transport model	Meteorology	Horizontal resolution of transport model	Inverse method	Experiments	References
CT2019	TM5	ERA-Interim	2° × 3° global; 1° × 1° in N.A.	EnKF	IS, LNLG, OG, LNLGOGIS	Peters et al. (2007)
OU	TM5	ERA-Interim	4° × 6°	4D-Var	PRIOR, IS, LNLG, OG, LNLGOGIS	Crowell et al. (2018)
TM5-4DVAR	TM5	ERA-Interim	2° × 3°	4D-Var	PRIOR, IS, LNLG, OG, LNLGOGIS	Basu et al. (2013)
CSU	GEOS-Chem	MERRA-2	4° × 5°	Bayesian synthesis	PRIOR, IS, LNLG, OG,	Schuh et al. (2010)
Ames	GEOS-Chem	MERRA-2	4° × 5°	4D-Var	PRIOR, IS, LNLG, OG, LNLGOGIS,	Philip et al. (2019)
CMS-Flux	GEOS-Chem	GEOS-FP	4° × 5°	4D-Var	PRIOR, IS, LNLG, OG, LNLGOGIS,	Liu et al. (2017)
UT	GEOS-Chem	GEOS-FP	4° × 5°	4D-Var	IS, LNLG, OG, LNLGOGIS	Deng and Chen (2011)
PCTM	PCTM	MERRA-2	2° × 2.5°	4D-Var	PRIOR, IS, LNLG, OG, LNLGOGIS,	Baker et al. (2010)
LoFI	GEOS-5	MERRA-2	0.5° × 0.625°	–	IS	Weir et al. (2021)
CAMS	LMDz	ERA5	1.875° × 3.75°	Variational	IS, LNLG, OG, LNLGOGIS	Chevallier et al. (2005); Remaud et al. (2018)

Note. ERA-Interim, European Centre for Medium-Range Weather Forecasts Reanalysis-Interim (Dee et al., 2011); ERA5, European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (Hersbach et al., 2020); MERRA-2, Modern-Era Retrospective Analysis for Research and Applications, Version 2 (Gelaro et al., 2017); GEOS-FP, Goddard Earth Observing System-Forward Processing ([https://gmao.gsfc.nasa.gov/GMAO\\_products/NRT\\_products.php](https://gmao.gsfc.nasa.gov/GMAO_products/NRT_products.php)).

## 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<sub>2</sub> and CH<sub>4</sub> 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 (18 July–28 August 2016, 1 February–10 March 2017, 3 October–10 November 2017, 12 April–20 May 2018, and 17 June–27 July 2019). The study area covers the eastern half of the U.S which is characterized by considerable spatial and seasonal variations of biogenic CO<sub>2</sub> fluxes (e.g., Huntzinger et al., 2012; Sweeney et al., 2015) and dynamic weather (e.g., Whittaker & Horn, 1984). All ACT-America in situ aircraft data is publicly available at <https://daac.ornl.gov/actamerica> (Davis et al., 2018). In this study, we use the aircraft measurements during the first four campaigns, that is, summer 2016 (15 July–28 August), winter 2017 (30 January–10 March), autumn 2017 (3 October–13 November), and spring 2018 (12 April–20 May). 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., [CO<sub>2</sub>] and CH<sub>4</sub>) and other trace gases (e.g., CO, C<sub>2</sub>H<sub>6</sub>, and O<sub>3</sub>), 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  $\chi_{\text{CO}_2}$  (Campbell et al., 2020; Wei et al., 2021). In total, 182 research flights were flown in four seasons and the tracks of flights in each season are shown in Figure 1. Each research flight spanned from 4 to 6 hr 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 (4,000–9,000 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. Research flights on frontal days were designed to fly across the frontal boundaries collecting observations in both cold and warm sectors at different vertical levels.



**Figure 1.** Flight tracks of B200 (purple) and C130 (light blue) in summer 2016, winter 2017, autumn 2017, and spring 2018. Thick lines on maps represent the tracks of four selected frontal flights, that is, 4 August 2016, 7 March 2017, 26 October 2017, and 2 May 2018, that are analyzed in detail in this study (flight tracks in warm sectors of the four frontal systems are colored in dark red while cold sectors dark blue).

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, Davis, Pauly, et al. (2020).

In this work, we use 5-s averaged measurements of  $[\text{CO}_2]$  and wind speed and 8-s 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, Davis, Lauvaux, et al. (2020). The entire free troposphere (FT) includes both LFT and UFT. 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: (a) frontal  $[\text{CO}_2]$  difference between the warm and cold sectors and (b) PBL-to-FT

[CO<sub>2</sub>] difference within the warm and cold sectors of frontal systems (Pal, Davis, Lauvaux, et al., 2020). These two metrics provide a quantitative understanding of the horizontal and vertical gradients of [CO<sub>2</sub>] associated with mixing processes within weather systems at a synoptic scale. The frontal difference is calculated by comparing the average [CO<sub>2</sub>] 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 flights) was based on surface synoptic map analyses, satellite images, and in-situ measurements of meteorological variables (Pal, Davis, Lauvaux, et al., 2020). Note that [CO<sub>2</sub>] measurements during take-offs and landings are excluded to leave out the influences of small-scale anthropogenic CO<sub>2</sub> fluxes at airports. The frontal differences are calculated for each frontal event and seasonal frontal metric statistics were then computed for both measurements and simulations.

The vertical PBL-to-FT [CO<sub>2</sub>] difference is defined as the difference between the layer-averaged [CO<sub>2</sub>] within the PBL and FT. Given the relatively coarse resolutions of global models, here we use both level leg and profile measurements for computing this ACT metric to conduct a synoptic-scale model-data comparison. This is slightly different from the method by Pal, Davis, Lauvaux, et al. (2020) that only uses profile data. We calculate observed and simulated PBL-to-FT [CO<sub>2</sub>] 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, Davis, Lauvaux, et al. (2020). Seasonal statistics of the PBL-to-FT differences are calculated using all frontal, pre-, and post-frontal flights in each season.

### 3. Case Study of Frontal Events

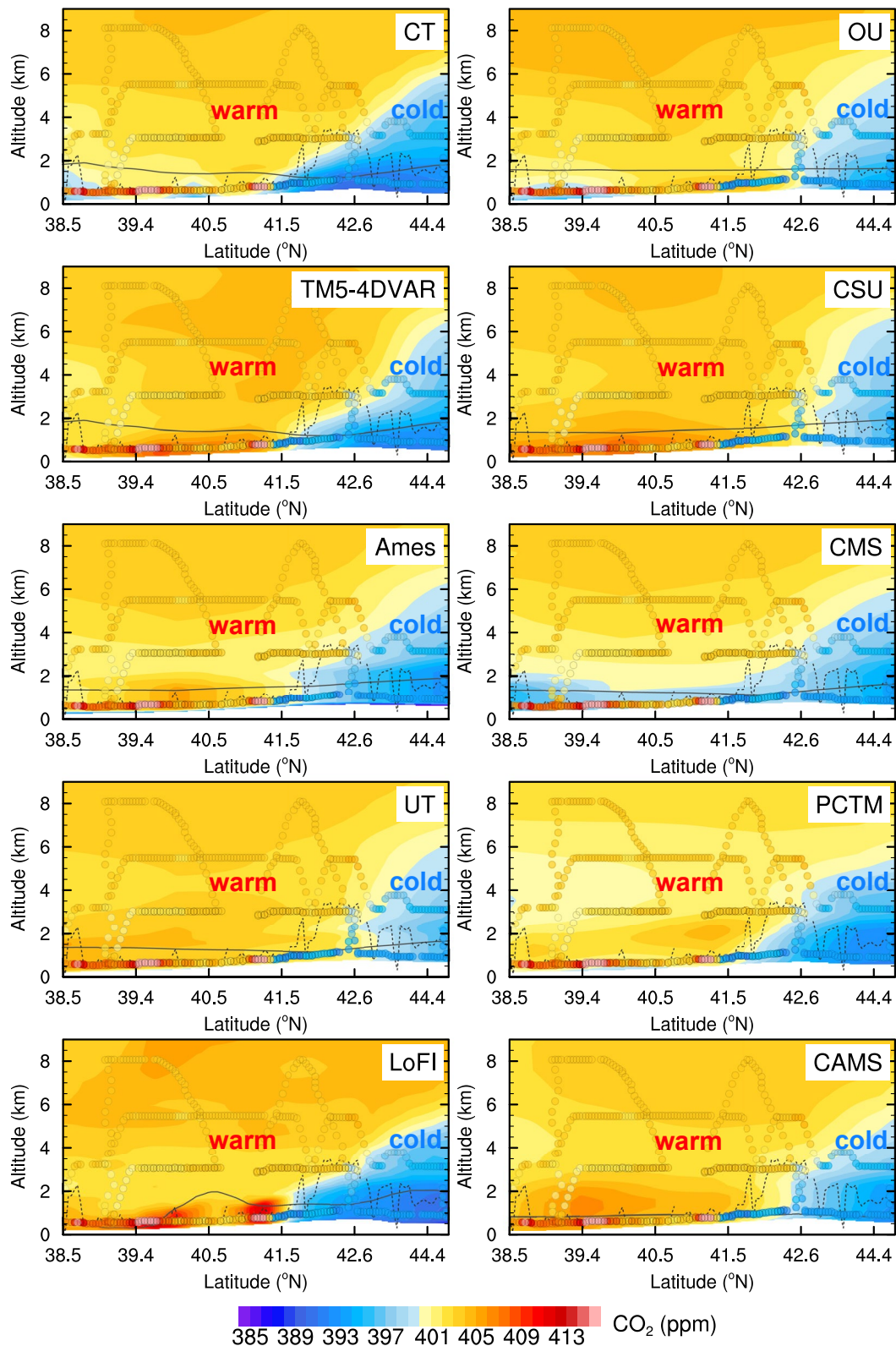
We start with process-oriented investigations of typical frontal events to investigate the spatial structures of observed [CO<sub>2</sub>] and to assess the model skills in simulating the observed [CO<sub>2</sub>] variability in details. Here, we select four frontal events during ACT campaigns and focus on examining a summer frontal case over the mid-west U.S. (Section 3.1) and a winter case in the northeast U.S. (Section 3.2), and we only show the simulations in the IS experiment which is conducted with all models. Model results in other inversion experiments (i.e., PRIOR, LNLG, OG, and LNLGOGIS) are shown in Figures S1 and S2 in Supporting Information S1. Results of the autumn and spring frontal events can be found in Figures S3–S8 in Supporting Information S1.

#### 3.1. A Summer Event on 4 August 2016

On 4 August 2016, the ACT-America team flew across a strong cold front in the mid-west region (Figure 1a). The cold front moved from northwest to southeast during the afternoon of 4 August. 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 & Davis, 2021). During this flight, aircraft flew over Missouri, Kansas, and Nebraska within the warm sector of the frontal system and over South Dakota in the cold sector (Figure 1a).

Latitude-altitude curtains of observed [CO<sub>2</sub>] over the mid-west (Figure 2) clearly show the cold [CO<sub>2</sub>]-poor air encounters warm [CO<sub>2</sub>]-rich air at 41.5°N at the surface and lifts the warm air. Enhanced [CO<sub>2</sub>] 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-[CO<sub>2</sub>] 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<sub>2</sub> (Pal, Davis, Lauvaux, et al., 2020). In comparison, in the cold sector (41.5°–45°N), observed [CO<sub>2</sub>] is lower within the PBL (385–395 ppm) than the FT (395–400 ppm), resulting from net biological uptake by upwind boreal forests in Canada and croplands in the mid-west on the passage of air masses (e.g., Miles et al., 2012). Observed frontal and PBL-to-FT [CO<sub>2</sub>] features in this event are consistent with the seasonal averages in summer (see Section 5).

It is encouraging that most of the models reproduce much of the observed frontal structure in this case with equatorward flows of cold low-[CO<sub>2</sub>] air from the north and poleward flows of warm high-[CO<sub>2</sub>] air (Figure 2). All models can capture the fact that the cold front lifts the air in the warm sector near the surface. However, the models underestimate [CO<sub>2</sub>] 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<sub>2</sub> uptake in these models. TM5-4DVAR, LoFI, CSU, and CAMS relatively better simulate high PBL-[CO<sub>2</sub>] 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 S5 in Supporting



**Figure 2.** Latitude-height curtains of model-simulated [CO<sub>2</sub>] (color shading) from the OCO-2 MIP in the IS experiment along the flight track crossing the mid-west on 4 August 2016 (15:00–18:00 UTC; thick line in Figure 1). The color-coded circles denote measurements of [CO<sub>2</sub>] by two aircraft (C130 and B200), the dashed line represents the boundary layer height derived from measurements by the airborne lidar on C130 and the solid gray lines denote model-simulated boundary layer height. All models are sampled for the same time period.



Information S1) 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. 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-based models (CSU, Ames, CMS-Flux, and UT) and CAMS in the PBL but is reasonably well-simulated in the FT. GEOS-Chem-based models overestimate PBL- $[\text{CO}_2]$  by up to 10 ppm near the surface frontal boundary (at around  $42^\circ\text{N}$ ; Figure 2). These models are biased low in wind speed in cold sectors (see Section 5) and simulate weak sinks (or even net sources) in the northern US (South Dakota, North Dakota, Montana, and Wyoming) and southern Canada (Figure S5 in Supporting Information S1) 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-based models 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-Flux, and CAMS; Figure S1 in Supporting Information S1) which have similar spatial patterns but different magnitudes in surface fluxes over N.A. (Figure S5 in Supporting Information S1). The results suggest the frontal  $[\text{CO}_2]$  difference likely to be strongly related to the spatial distribution of fluxes. The OG experiment (Figure S1d in Supporting Information S1) 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 S5d in Supporting Information S1).

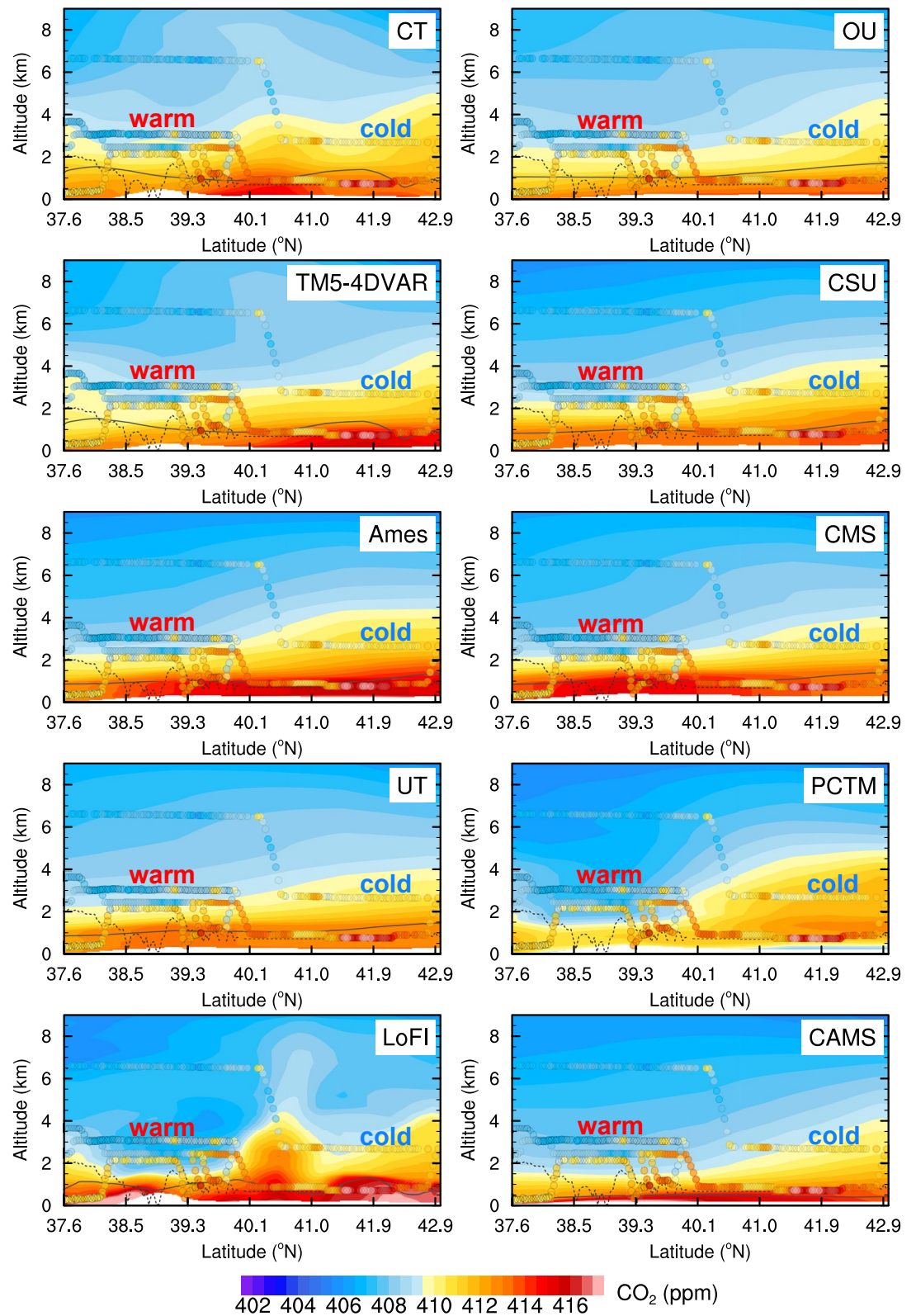
### 3.2. A Winter Event on 7 March 2017

On 7 March 2017, C130 and B200 were flying in the northeast across a strong cold front (Figure 1b). 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 & Davis, 2021). The flight tracks crossed the warm sector in Virginia and Ohio ( $37^\circ$ – $41^\circ\text{N}$ ) and the cold sector in Indiana, Michigan, and Wisconsin ( $41^\circ$ – $43^\circ\text{N}$ ; Figure 1b).

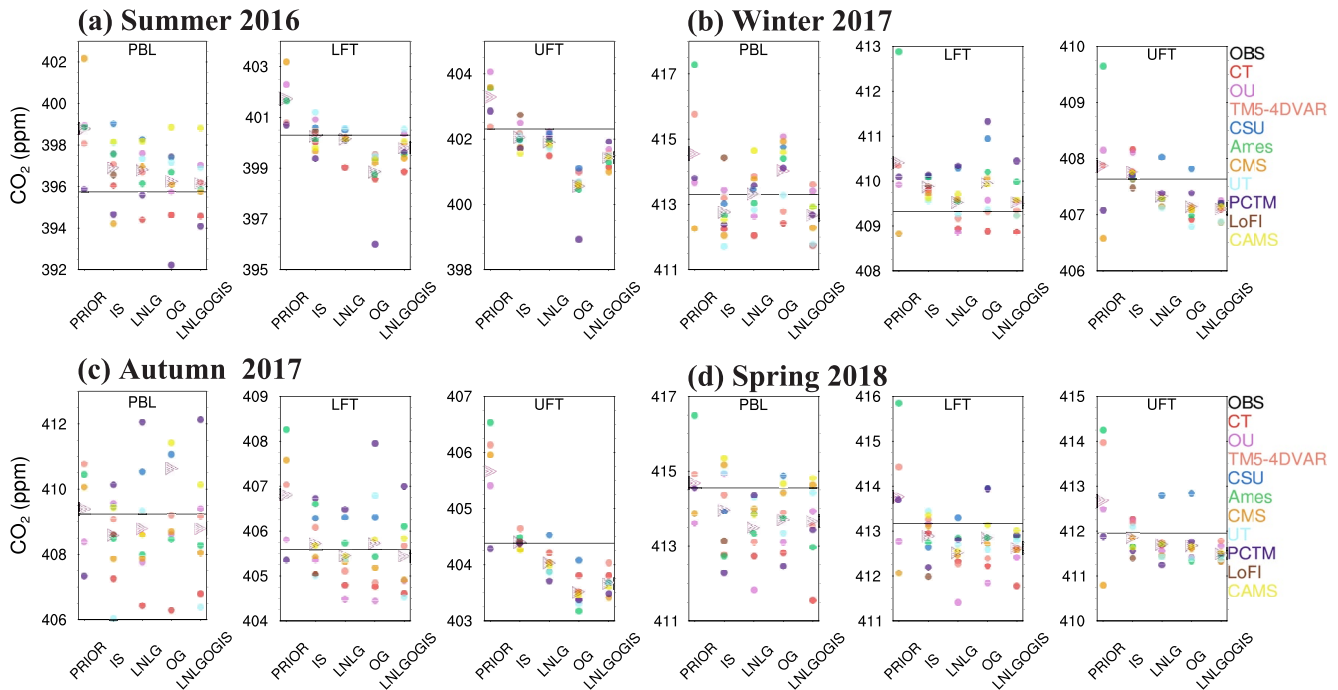
Latitude-altitude curtains of measurements (Figure 3) show enhanced  $[\text{CO}_2]$  within the PBL (402–416 ppm) compared to the FT (385–405 ppm) in both warm and cold sectors as the northern mid-latitudes are a net source of  $\text{CO}_2$  due to ecosystem respiration (Figure S6 in Supporting Information S1) 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 S3–S4 in Supporting Information S1). Observed frontal and PBL-to-FT  $[\text{CO}_2]$  differences during this winter frontal event are also similar to the seasonal statistics (see Section 5).

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 3). This is because the models simulate positive fluxes over the majority of the U.S. in winter (Figure S6 in Supporting Information S1). However, the models tend to underestimate  $[\text{CO}_2]$  in cold sectors ( $41^\circ$ – $43^\circ\text{N}$ ) and overestimate  $[\text{CO}_2]$  in warm sectors in the PBL ( $37^\circ$ – $41^\circ\text{N}$ ), leading to underestimates in frontal differences. Large model discrepancies are found within the PBL in both warm and cold sectors. In the warm sector, the underestimates in PBLH may contribute to the overestimates in PBL- $[\text{CO}_2]$  in most models. In the cold sector (Indiana, Michigan, and Wisconsin), LoFI and CAMS are closer to measurements within the PBL compared to other models, contributing to their better performance in simulating the warm-cold contrast. TM5-4DVAR has the best performance among the TM5-based models in the cold sector and might be benefiting from its wintertime surface fluxes. Ames is closer to measurements in the cold sector than other GEOS-Chem models possibly because of its more realistic surface fluxes in winter.

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, LoFI, and IS experiment of OU; Figure S6 in Supporting Information S1) 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 higher than  $300 \text{ g C m}^{-2} \text{ y}^{-1}$  in winter 2017. IS fluxes in many models (e.g., CT, OU, TM5-4DVAR, Ames, UT, and CAMS) are distinct from prior fluxes while are similar to LNLGGIS fluxes



**Figure 3.** Same as Figure 2, but for 7 March 2017 along the flight track crossing the northeast U.S. (16:00–19:00 UTC; thick line in Figure 1).



**Figure 4.** Mean CO<sub>2</sub> mole fraction (henceforth, [CO<sub>2</sub>]) within the planetary boundary layer (PBL), lower free-troposphere (LFT), and upper free-troposphere (UFT) as observed with aircraft (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 S6 in Supporting Information S1). Simulated [CO<sub>2</sub>] frontal structures in these models with posterior fluxes, however, are not improved compared to their PRIOR experiments. In CMS-Flux, different sets of flux inversions don't noticeably change the simulated frontal structures of [CO<sub>2</sub>] in (Figure S2 in Supporting Information S1) since experiments with different [CO<sub>2</sub>] data sources have similar spatial patterns of fluxes (Figure S6 in Supporting Information S1). In other models (e.g., CSU and PCTM), spatial patterns of flux noticeably change across experiments constrained by different CO<sub>2</sub> data, leading to diverging model performance in [CO<sub>2</sub>] frontal structure in winter.

## 4. Model-Data Evaluation of [CO<sub>2</sub>]

### 4.1. Seasonal Observed and Simulated Total [CO<sub>2</sub>]

We now evaluate the OCO-2 MIP simulations of the seasonal mean [CO<sub>2</sub>] in three vertical layers (PBL, LFT, and UFT) over the eastern U.S. (Figure 4). ACT measurements show higher averaged [CO<sub>2</sub>] in winter and spring (~408–415 ppm in the PBL) due to a shallow PBL and ecosystem respiration, and lower mean in summer and autumn (~396–409 ppm in the PBL) as a result of deep mixing and net photosynthesis fluxes (known as the CO<sub>2</sub> “rectifier” effect; Denning et al., 1999). The observed seasonal [CO<sub>2</sub>] difference during the ACT measurement periods (here defined as the seasonal peak-to-trough amplitude; AMP<sub>P-T</sub>) ranges from ~10 to 19 ppm in the three vertical layers with the mean seasonal cycle most notable in the PBL (AMP<sub>P-T</sub> = 19 ppm). Slower vertical mixing and a larger mixing volume attenuate and delay the seasonal cycle in the FT (AMP<sub>P-T</sub> = 10–13 ppm).

The OCO-2 MIP ensembles driven by different suites of fluxes reproduce the observed seasonal variation of [CO<sub>2</sub>] (i.e., relatively high in winter and spring while low in summer and autumn) but have distinct and diverse performance. The ensembles simulate the observed AMP<sub>P-T</sub> 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 [CO<sub>2</sub>] 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 [CO<sub>2</sub>] 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 4.2). Meanwhile, the OG ensemble has a much larger mean bias (1.5–1.8 ppm) in FT- $[\text{CO}_2]$  than other ensembles in summer (Table S1 in Supporting Information S1). 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).

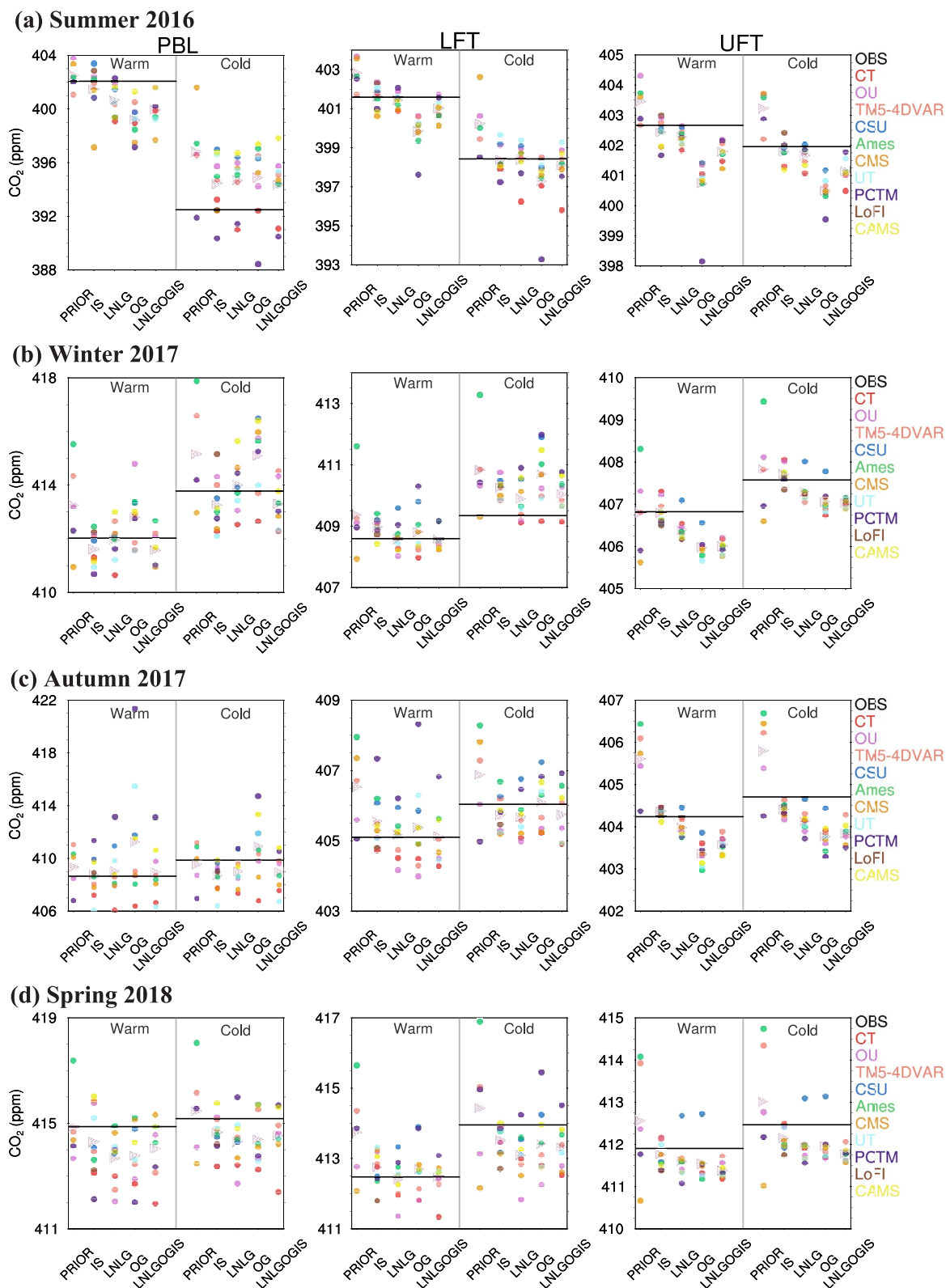
The performance of individual OCO-2 MIP models in simulating observed seasonal average  $[\text{CO}_2]$  is strongly model and season dependent, and there is no close correlation between model resolution and model biases in seasonal  $[\text{CO}_2]$ . A large model spread among OCO-2 MIP models exists 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-based models (CSU, Ames, CMS-Flux, and UT) overall overestimates PBL- $[\text{CO}_2]$  in summer but underestimates observations in winter.

#### 4.2. 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 5). In summer,  $[\text{CO}_2]$  in warm sectors is considerably higher than that in cold sectors throughout the troposphere (e.g., 402 vs. 392 ppm in the PBL; Figure 5a). 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 vs. 412 ppm in the PBL; Figure 5b). 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 (Figures 5b–5d). 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 5 and Figures S9–S16 in Supporting Information S1). In summer, OG underestimates  $[\text{CO}_2]$  in warm sectors by  $\sim 3$  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 5 and Figures S9–S16 in Supporting Information S1). In summer, model spread is smaller in warm sectors than cold sectors (3–7 ppm vs. 8–10 ppm). The models (except LoFI) often underestimate observations in warm sectors in summer. In cold sectors, CT and PCTM are biased low while other models overestimate observations within the PBL (Figure S9 in Supporting Information S1). In winter, model biases are much smaller compared to summer with CT and UT having relatively smaller biases than other models (Figure S10 in Supporting Information S1). In autumn, OU, TM5-4VAR, and LoFI have smaller biases than other models (Figure S11 in Supporting Information S1). In spring, model biases in warm sectors are comparable to cold sectors and the models with posterior fluxes often underestimate the observations in both warm and cold sectors (Figure 5 and Figure S12 in Supporting Information S1). In all four seasons, the models overall have much smaller model mean biases in the FT than the PBL within both warm and cold sectors (Figures S9–S12 in Supporting Information S1).



**Figure 5.** Mean [CO<sub>2</sub>] 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 aircraft (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.

The models overall have better performance in warm sectors than cold sectors in autumn and spring in terms of model-data correlations and normalized standard deviations but have diverse performance in summer and winter (Figures S13–S16 in Supporting Information S1). In summer, the models agree better with observations in the FT with higher correlations compared to the PBL (Figure S13 in Supporting Information S1). The high resolutions of CT and LoFI likely aid in their relatively good performance in model-data correlation and normalized standard deviation in summer, but these two models do not surpass the others in winter and spring. In winter, the TM5-based models (CT, OU, and TM5-4DVAR) are often better correlated with observations than the other models (Figure S14 in Supporting Information S1). In autumn, the models much better capture the observed  $[\text{CO}_2]$  variation in warm sectors compared to cold sectors and all models have difficulties reproducing  $[\text{CO}_2]$  variation in cold sectors within the UFT ( $R^2 < 0.5$ ; Figure S15 in Supporting Information S1). In spring, the models (except PCTM) reasonably capture the variations in  $[\text{CO}_2]$  in warm sectors ( $R^2 = 0.5\text{--}0.8$ ) but are poorly correlated with observations in cold sectors within the PBL and LFT ( $R^2 < 0.5$ ; Figure S16 in Supporting Information S1). More descriptions of the model-data correlation and normalized standard deviation can be found in the in Supporting Information S1.

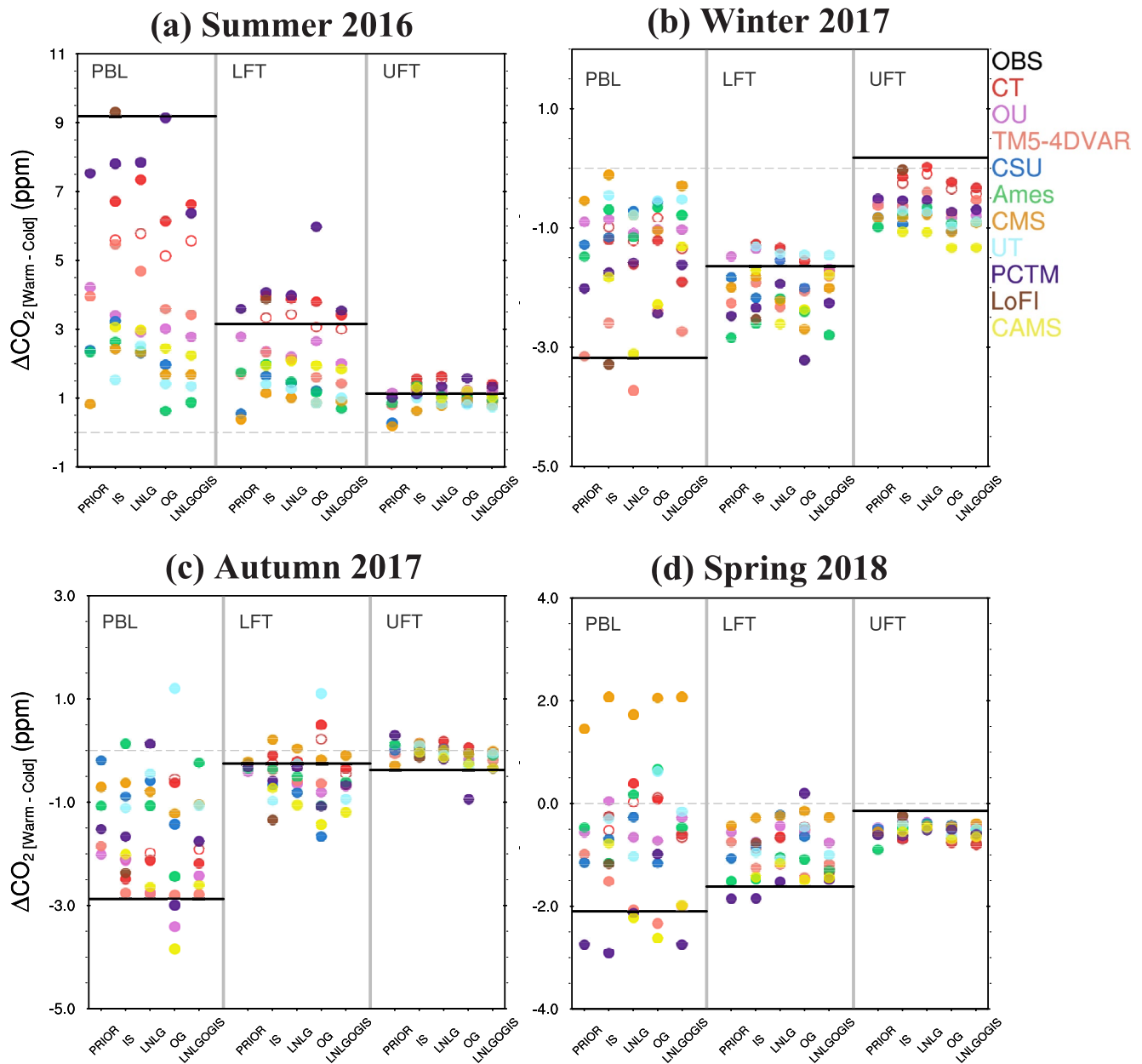
## 5. Horizontal and Vertical Variations of $[\text{CO}_2]$ in Frontal Systems

### 5.1. Observed and Simulated Frontal $[\text{CO}_2]$ Differences

Observed summer frontal  $[\text{CO}_2]$  differences are distinctly different from the other seasons (Figure 6 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 6a). 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 6b–6d) due to a change in the latitudinal  $[\text{CO}_2]$  gradient compared to summertime. The variability in observed frontal differences (i.e., the ranges across different frontal cases) is often comparable in magnitude to the mean frontal differences (Table 2). 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 6 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 5a). The GEOS-Chem-based models (CSU, Ames, CMS-Flux, and UT) significantly underestimate 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. OU and CAMS capture some of the observations and perform better than the GEOS-Chem-based 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. In dormant seasons (winter and autumn), the models tend to simulate weaker negative frontal differences than observed ones in the PBL (Figures 6b and 6c and Table 2) with LoFI, CAMS, PCTM, and the TM5-based models (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 6d). While other models, including CT, OU, and GEOS-Chem-based 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. The models also often simulate weaker variability in frontal  $[\text{CO}_2]$  difference across different frontal cases due to coarse model resolutions (Table 2).

Within a single inverse modeling system, varying the data source makes a relatively small difference in the simulated frontal  $[\text{CO}_2]$  difference (Figure 6). Comparisons among PRIOR, IS, LNLG, and LNLGOGIS experiments (OG is excluded since it tends to degrade model performance) show that the models have a 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



**Figure 6.** Frontal  $[CO_2]$  difference ( $CO_2$  mole fraction in the warm sector minus that in the cold sector) within the PBL, LFT, and UFT as observed with aircraft (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.

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  $[CO_2]$  data sources.

Differences in modeled wind speed may partially explain the model discrepancies in frontal difference during summer. The TM5-based models (CT, TM5-4DVAR, and OU), PCTM, LoFI, and CAMS reproduce observed wind speeds within frontal systems more accurately than the GEOS-Chem-based models (CSU, Ames, CMS-Flux, and UT) (Figure 7a). 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- $[CO_2]$  air masses from the north in cold sectors during summer than other models, contributing to an overestimate in  $[CO_2]$  within cold sectors and subsequent biases in frontal difference in GEOS-Chem-based models. 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

**Table 2**

Statistics of Frontal CO<sub>2</sub> Differences in the PBL and FT and PBL-to-FT CO<sub>2</sub> Differences in the Warm and Cold Sectors as Observed With Aircraft and Simulated With OCO-2 MIP Models (Unit: ppm)

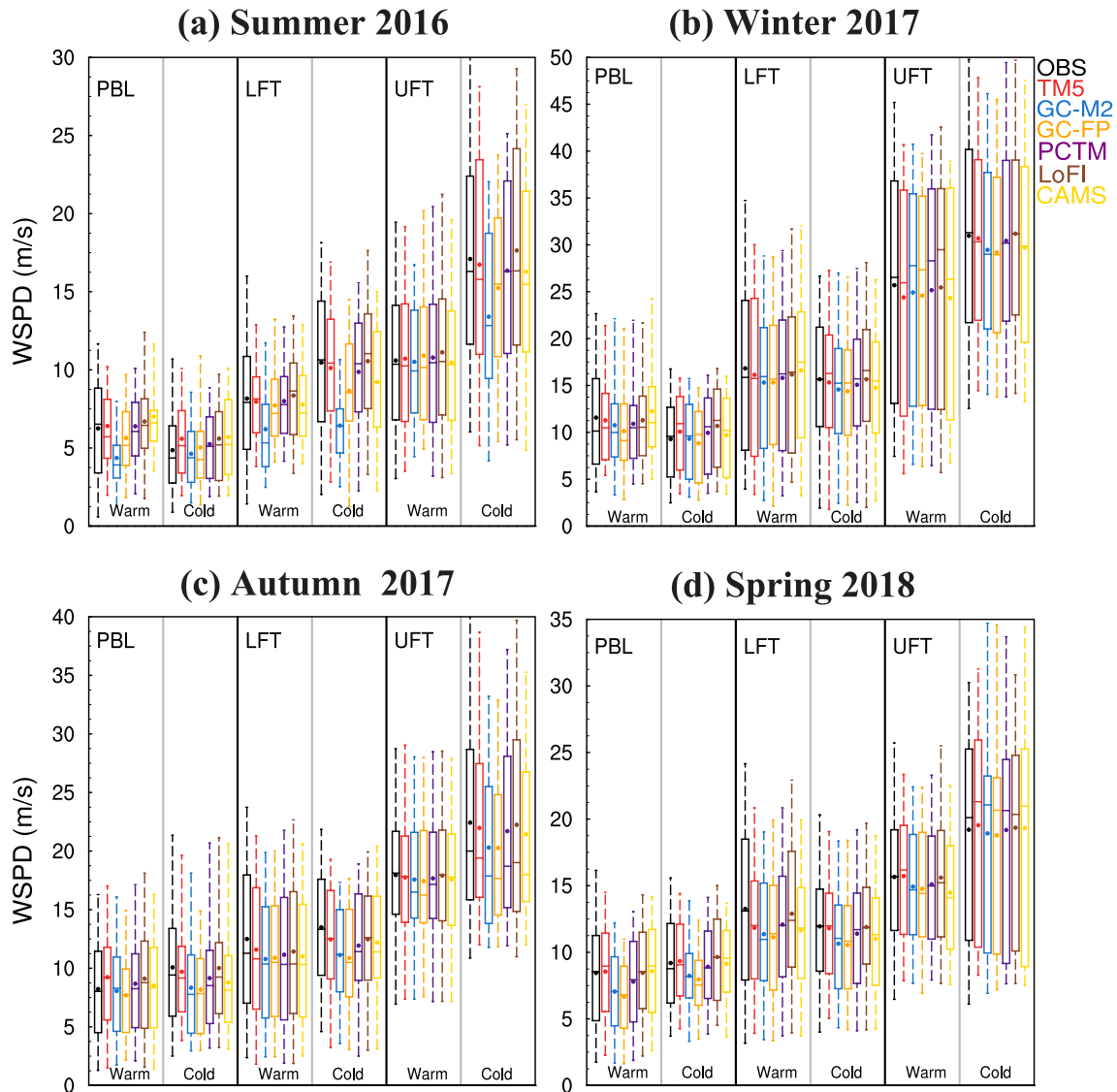
	OBS	CT (CT_Global)	OU	TM5-4DVAR	CSU	Ames	CMS-flux	UT	PCTM	LoFI	CAMS
Frontal difference (PBL)											
Summer	9.2 ± 8.5	6.7 ± 4.2 (5.5 ± 3.9)	3.3 ± 2.3	4.2 ± 2.9	2.5 ± 1.6	1.8 ± 2.7	1.7 ± 2.2	1.7 ± 2.7	7.7 ± 5.5	9.3 ± 5.8	2.7 ± 3.5
Winter	-3.2 ± 5.1	-1.5 ± 3.2 (-1.1 ± 2.5)	-1.0 ± 3.7	-2.9 ± 5.0	-0.9 ± 2.3	-1.0 ± 3.6	-0.6 ± 3.4	-0.6 ± 2.8	-1.9 ± 4.5	-3.3 ± 3.8	-2.1 ± 3.4
Autumn	-2.9 ± 3.5	-1.9 ± 2.1 (-1.6 ± 2.0)	-2.5 ± 2.6	-2.6 ± 2.5	-0.8 ± 3.8	-0.9 ± 2.4	-0.9 ± 2.8	-0.4 ± 3.3	-1.6 ± 3.2	-2.4 ± 3.3	-2.8 ± 2.5
Spring	-2.1 ± 3.7	-0.1 ± 2.5 (-0.3 ± 2.4)	-0.4 ± 1.7	-1.8 ± 3.1	-0.8 ± 1.8	-0.2 ± 3.0	1.9 ± 2.9	-0.2 ± 1.7	-2.3 ± 3.6	-1.2 ± 4.7	-1.9 ± 1.9
Frontal difference (FT)											
Summer	2.5 ± 3.7	2.6 ± 1.9 (2.3 ± 1.5)	1.8 ± 1.8	1.4 ± 1.3	1.0 ± 1.2	1.2 ± 1.6	0.8 ± 1.5	1.0 ± 1.5	2.8 ± 2.4	2.6 ± 2.3	1.5 ± 1.6
Winter	-1.1 ± 1.7	-0.8 ± 1.0 (-0.9 ± 0.9)	-1.1 ± 1.0	-1.3 ± 1.5	-1.4 ± 0.8	-1.7 ± 1.0	-1.5 ± 0.9	-1.1 ± 0.8	-1.5 ± 1.3	-1.3 ± 1.6	-1.7 ± 1.6
Autumn	-0.3 ± 2.6	0.0 ± 1.3 (-0.1 ± 1.1)	-0.3 ± 0.4	-0.3 ± 0.9	-0.5 ± 0.3	-0.3 ± 0.8	0.0 ± 1.1	-0.2 ± 0.6	-0.4 ± 0.6	-0.7 ± 1.3	-0.6 ± 0.6
Spring	-1.1 ± 2.8	-0.8 ± 1.5 (-0.7 ± 1.2)	-0.5 ± 0.8	-0.9 ± 1.3	-0.6 ± 0.6	-0.9 ± 1.3	-0.4 ± 1.3	-0.7 ± 0.7	-0.9 ± 1.5	-0.5 ± 2.1	-1.0 ± 1.5
PBL-to-FT difference (warm sector)											
Summer	0.6 ± 3.7	-1.5 ± 1.9 (-1.4 ± 1.7)	-0.7 ± 1.5	-1.4 ± 2.4	-0.4 ± 1.8	-1.5 ± 3.1	-2.2 ± 3.5	-2.1 ± 3.8	-1.0 ± 2.4	0 ± 2.1	0.1 ± 2.9
Winter	3.9 ± 3.0	3.2 ± 2.2 (3.3 ± 2.1)	4.7 ± 2.9	4.7 ± 3.4	3.7 ± 1.2	4.7 ± 2.1	4.1 ± 1.8	3.5 ± 1.3	3.2 ± 1.6	4.5 ± 2.3	4.8 ± 1.9
Autumn	4.7 ± 3.5	2.5 ± 2.7 (2.4 ± 2.7)	4.2 ± 4.3	3.9 ± 3.9	4.6 ± 4.1	3.3 ± 3.2	3.3 ± 3.2	4.3 ± 3.0	6.3 ± 5.0	4.1 ± 3.4	5.3 ± 3.6
Spring	2.4 ± 2.3	1.0 ± 2.0 (1.0 ± 1.9)	1.3 ± 2.3	0.9 ± 2.4	1.2 ± 1.4	2.1 ± 2.5	2.3 ± 3.7	1.6 ± 1.9	0.5 ± 3.8	1.7 ± 3.3	2.3 ± 1.4
PBL-to-FT difference (cold sector)											
Summer	-4.1 ± 6.2	-3.9 ± 4.4 (-3.5 ± 3.6)	-1.9 ± 3.5	-2.9 ± 3.9	-1.7 ± 2.9	-1.6 ± 4.0	-2.6 ± 5.3	-2.5 ± 3.7	-4.3 ± 4.0	-4.7 ± 5.0	-0.4 ± 2.8
Winter	5.7 ± 3.0	4.0 ± 1.9 (3.7 ± 1.7)	5.2 ± 2.3	6.0 ± 3.1	4.2 ± 2.3	4.8 ± 2.3	4.4 ± 2.8	3.9 ± 2.4	4.6 ± 3.3	6.3 ± 3.1	5.8 ± 2.1
Autumn	5.0 ± 5.0	2.5 ± 1.3 (2.3 ± 1.2)	3.8 ± 2.0	4.1 ± 2.1	3.7 ± 2.2	2.7 ± 1.9	2.6 ± 2.5	3.1 ± 2.3	4.3 ± 3.4	3.4 ± 1.9	5.1 ± 3.4
Spring	2.6 ± 2.6	0.3 ± 1.9 (0.4 ± 1.9)	1.5 ± 1.3	1.5 ± 1.7	1.5 ± 1.3	1.3 ± 1.7	1.0 ± 2.6	1.4 ± 1.4	1.8 ± 2.9	1.6 ± 1.8	2.4 ± 1.6

Note. Model results are averaged across five suites of experiments constrained with different data sources. Uncertainties represent the ranges of ACT metrics across all frontal cases.

for the overestimate of [CO<sub>2</sub>] in cold sectors. In non-summer seasons, all models reproduce observed wind speeds in both warm and cold sectors without significant biases (Figures 7b–7d).

Differences in modeled PBLH do not readily explain the diverging model performance in frontal [CO<sub>2</sub>] differences. Note that we focus on comparing the mean observed and modeled PBLH values. In summer, PBLH in GEOS-Chem-based models (CSU, Ames, CMS-Flux, and UT) is comparable to that in TM5-based models (CT, TM5-4DVAR, and OU) in cold sectors (Figure 8a), which does not necessarily result in comparable model performance in [CO<sub>2</sub>] within the PBL and frontal [CO<sub>2</sub>] differences. While LoFI reproduces the observed frontal differences in summer, it is, however, biased high by ~400–1,000 m in mean PBLH. In winter (Figure 8b), both





**Figure 7.** Box-whisker plots of wind speed in warm and cold sectors of frontal systems as observed with aircraft (OBS) and simulated with TM5 (used in CT, OU, and TM5-4DVAR), GEOS-Chem driven by MERRA2 (GC-M2; used in CSU and Ames), GEOS-Chem driven by GEOS-FP (GC-FP; used in CMS-Flux and UT), LoFI (GEOS-5), and CAMS (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.

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 mean PBLH in autumn within both warm and cold sectors (Figure 8c), 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 8d). These results suggest that differences in horizontal transport, vertical transport (both PBL and convective mixing), 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 vs. 5.5 ppm in the PBL and 2.6 vs. 2.3 ppm in the FT) but has negligible impacts in other seasons (Figure 6 and Table 2). LoFI, at a

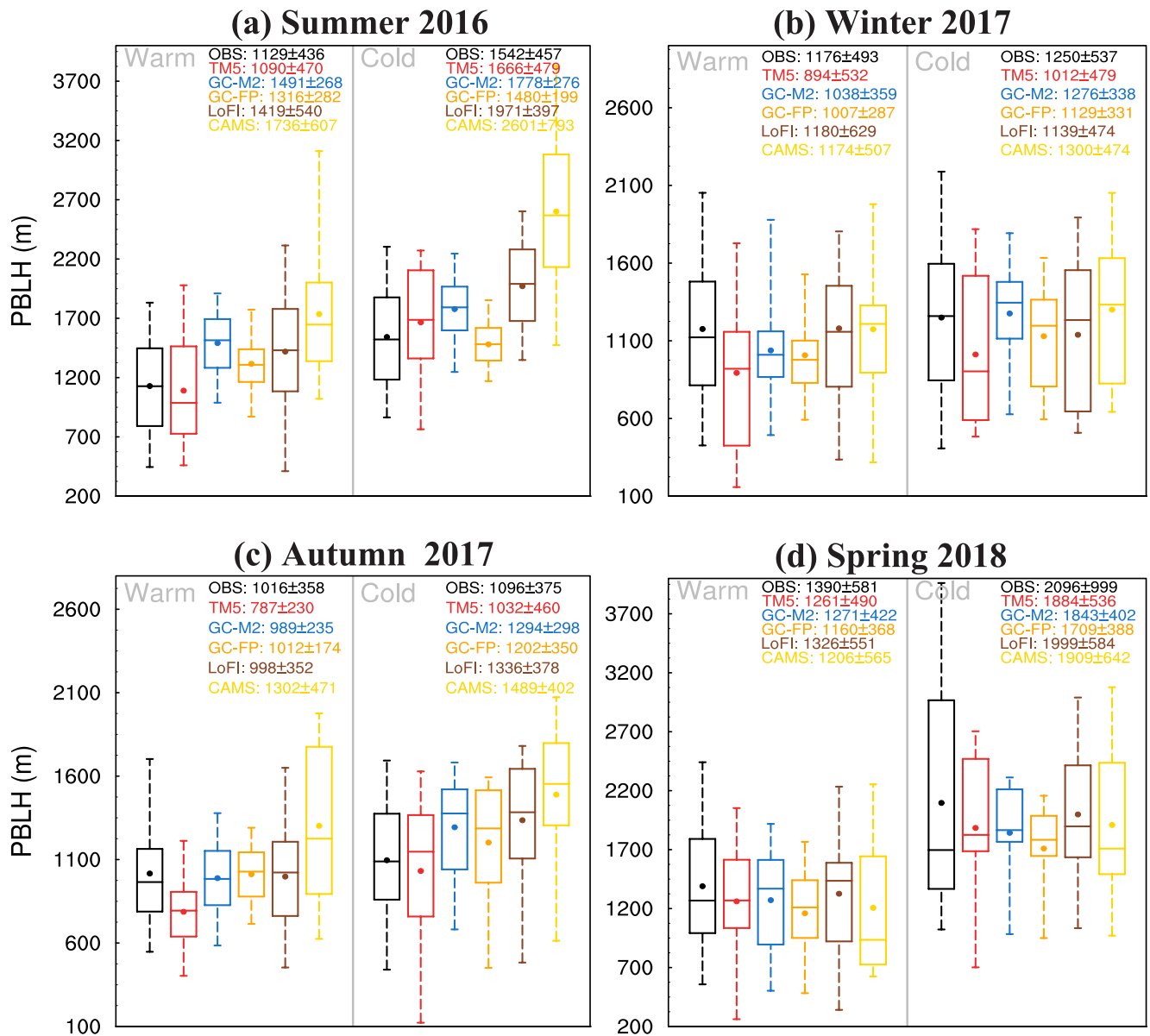
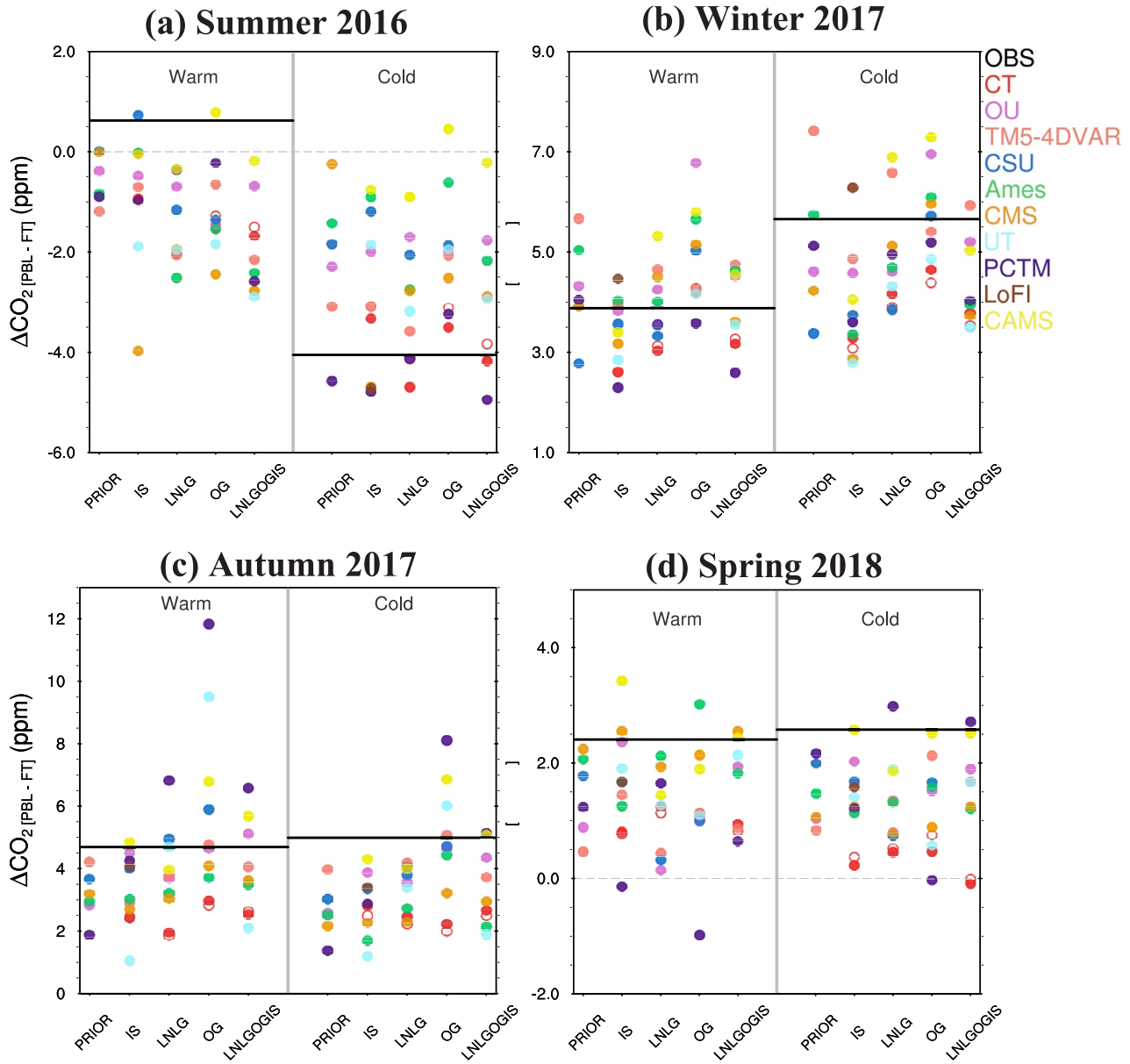


Figure 8. Same as Figure 7, but for planetary boundary layer depth (PBLH) in (a) summer 2016, (b) winter 2017, (c) autumn 2017, and (d) spring 2018.

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 the GEOS-Chem-based models (CSU, Ames, CMS-Flux, and UT) at  $4^\circ \times 5^\circ$ .

### 5.2. 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 9 and Table 2). In summer, observations show a large mean negative PBL-to-FT  $[\text{CO}_2]$  difference in cold sectors ( $-4.1 \pm 6.2$  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 variability ( $0.6 \pm 3.7$  ppm).



**Figure 9.** Same as Figure 6, but for PBL-to-FT [CO<sub>2</sub>] difference in warm and cold sectors of frontal systems in (a) summer 2016, (b) winter 2017, (c) autumn 2017, and (d) spring 2018.

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, Davis, Lauvaux, et al., 2020). During winter and autumn, large positive PBL-to-FT differences existed in both warm and cold sectors ( $3.9$ – $5.7$  ppm). The enhanced [CO<sub>2</sub>] 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 ( $2.4$ – $2.6$  ppm) wanes as photosynthesis turns up.

The models often simulate weaker (or even the opposite) PBL-to-FT [CO<sub>2</sub>] differences compared to observations but often capture some of the variability of PBL-to-FT differences across frontal cases in all seasons (Table 2). In summer, all models tend to simulate a negative PBL-to-FT [CO<sub>2</sub>] difference in warm sectors in contrast to the observed positive value (Figure 9a) 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-based group (CT, TM5-4DVAR, and OU) better resolve this metric than other models (Figure 8a). 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  $\text{CO}_2$  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 9c and 9d), which may be linked to the high-biased model PBLH in autumn and low-biased PBLH in spring (Figures 8c and 8d). In autumn, overestimated PBL mixing in models can lead to an overestimate of the dilution of positive  $\text{CO}_2$  surface fluxes, causing smaller PBL-to-FT  $[\text{CO}_2]$  differences in models relative to observations (Figure 9c); In late spring when photosynthesis begins, the underestimates in PBL mixing in models lead to unrealistically low  $[\text{CO}_2]$  within the PBL (Figure 5d), resulting in weaker positive (or sometimes even negative) PBL-to-FT  $[\text{CO}_2]$  difference compared to observations (Figure 9d).

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 9).

## 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  $\text{CO}_2$  transport or/and fluxes. In OG and LNLGOGIS experiments, the unbiased PBL- $[\text{CO}_2]$  is a result of canceling 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, Jacobson, et al., 2021; Cui, Zhang, et al., 2021) suggested that the seasonality of net ecosystem  $\text{CO}_2$  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.

Besides, the intra-monthly and diurnal variations of CO<sub>2</sub> fluxes are not captured in our simulations, which may affect the model ability to match the [CO<sub>2</sub>] observations.

Second, the models may persistently overestimate vertical mixing, leading to weaker vertical [CO<sub>2</sub>] 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-[CO<sub>2</sub>], causing underestimates of vertical [CO<sub>2</sub>] differences and contribute to the underestimates of cross-frontal [CO<sub>2</sub>] 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 (CSU, Ames, CMS-Flux, and UT), horizontal winds appear to be underestimated within weather systems. This biased horizontal mixing contributes to the underestimated cross-frontal [CO<sub>2</sub>] 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<sub>2</sub> 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 [CO<sub>2</sub>] less across latitudes.

Fourth, 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 [CO<sub>2</sub>] 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 [CO<sub>2</sub>] 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 [CO<sub>2</sub>] structures.

Comparison among OCO-2 MIP models constrained by different suites of surface CO<sub>2</sub> fluxes suggest that data sources in different experiments are unlikely to explain most of the model biases in frontal [CO<sub>2</sub>] 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 [CO<sub>2</sub>] 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 [CO<sub>2</sub>] 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 [CO<sub>2</sub>] 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 [CO<sub>2</sub>] 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 & 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 reanalysis 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 MER-RA-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 & 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<sub>2</sub> 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 CO<sub>2</sub> 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., 2019). 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 CO<sub>2</sub> fluxes at a subcontinental scale and foster our confidence in global flux inversions.

### Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

### Data Availability Statement

The ACT-American data is publicly available at the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (<https://daac.ornl.gov/actamerica>). OCO-2 MIP model results are available at <https://www.esrl.noaa.gov/gmd/ccgg/OCO2/>.

### Acknowledgments

This study was funded by the National Aeronautics and Space Administration (NASA) under the following awards: NNX15AG76G to Penn State (Davis); NNX15AJ07G to Colorado State (Baker and Schuh); 80NSSC19K0730 to Texas Tech and a Texas Tech faculty start-up grant (Pal); and 80NM0018D0004 to Jet Propulsion Laboratory, Caltech (Liu). Additional support for research was provided by NASA Grants NNX12AP90G (Davis) and NNX14AJ17G (Davis). Co-author Johnson acknowledges the internal funding from NASA's Earth Science Research and Analysis Program and co-author Philip acknowledges financial support of the NASA Academic Mission Services by Universities Space Research Association at NASA Ames Research Center. The ORNL DAAC is sponsored by the National Aeronautics and Space Administration under Interagency Agreement 80GSFC19T0039. ORNL participation in ACT-America was funded by Interagency Agreement NNL15AA10I. The statements, findings, and conclusions are those of the author(s) and should not be construed as the views of the agencies. 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.

### References

- Agustí-Panareda, A., Diamantakis, M., Massart, S., Chevallier, F., Muñoz-Sabater, J., Barré, J., et al. (2019). Modelling CO<sub>2</sub> weather – Why horizontal resolution matters. *Atmospheric Chemistry and Physics*, 19(11), 7347–7376. <https://doi.org/10.5194/acp-19-7347-2019>
- Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., et al. (2014). CO<sub>2</sub>, CO, and CH<sub>4</sub> 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. *Atmospheric Measurement Techniques*, 7(2), 647–687. <https://doi.org/10.5194/amt-7-647-2014>
- Baker, D. F., Bösch, H., Doney, S. C., O'Brien, D., & Schimel, D. S. (2010). Carbon source/sink information provided by column CO<sub>2</sub> measurements from the Orbiting Carbon Observatory. *Atmospheric Chemistry and Physics*, 10(9), 4145–4165. <https://doi.org/10.5194/acp-10-4145-2010>
- Barnes, E. A., Parazoo, N., Orbe, C., & Denning, A. S. (2016). Isentropic transport and the seasonal cycle amplitude of CO<sub>2</sub>. *Journal of Geophysical Research: Atmospheres*, 121(13), 8106–8124. <https://doi.org/10.1002/2016JD025109>
- Basu, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., et al. (2013). Global CO<sub>2</sub> fluxes estimated from GOSAT retrievals of total column CO<sub>2</sub>. *Atmospheric Chemistry and Physics*, 13(17), 8695–8717. <https://doi.org/10.5194/acp-13-8695-2013>
- 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. <https://doi.org/10.1029/2001JD000807>
- Campbell, J. F., Lin, B., Dobler, J., Pal, S., Davis, K., Obland, M. D., et al. (2020). Field evaluation of column CO<sub>2</sub> retrievals from intensity-modulated continuous-wave differential absorption lidar measurements during the ACT-America campaign. *Earth and Space Science*, 7(12), e2019EA000847. <https://doi.org/10.1029/2019EA000847>
- Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., et al. (2005). Inferring CO<sub>2</sub> sources and sinks from satellite observations: Method and application to TOVS data. *Journal of Geophysical Research: Atmospheres*, 110(D24). <https://doi.org/10.1029/2005jd006390>
- Corbin, K. D., & Denning, A. S. (2006). Using continuous data to estimate clear-sky errors in inversions of satellite CO<sub>2</sub> measurements. *Geophysical Research Letters*, 33(12). <https://doi.org/10.1029/2006GL025910>
- 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. *Atmospheric Chemistry and Physics*, 19(15), 9797–9831. <https://doi.org/10.5194/acp-19-9797-2019>
- 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 CO<sub>2</sub> to detect flux perturbations to the carbon cycle. *Journal of Geophysical Research: Atmospheres*, 123(2), 1460–1477. <https://doi.org/10.1002/2017jd027836>
- Cui, Y. Y., Jacobson, A. R., Feng, S., Wesloh, D., Barkley, Z. R., Zhang, L., et al. (2021). Evaluation of CarbonTracker's inverse estimates of North American net ecosystem exchange of CO<sub>2</sub> from different observing systems using ACT-America airborne observations. *Journal of Geophysical Research: Atmospheres*, 126(12), e2020JD034406. <https://doi.org/10.1029/2020JD034406>
- Cui, Y. Y., Zhang, L., Feng, S., Jacobson, A. R., Johnson, M. S., Philip, S., et al. (2021). Do global inversion systems systematically underestimate the seasonal amplitude of net ecosystem CO<sub>2</sub> flux? Paper presented at the AGU Fall Meeting 2021 abstracts.

- 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*, 102(9), E1714–E1734. <https://doi.org/10.1175/bams-d-20-0300.1>
- 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. [https://doi.org/10.1175/1520-0426\(2000\)017<1455:Aomfda>2.0.Co;2](https://doi.org/10.1175/1520-0426(2000)017<1455:Aomfda>2.0.Co;2)
- Davis, K. J., Obland, M. D., Lin, B., Lauvaux, T., O'Dell, C., Meadows, B., et al. (2018). ACT-America: L3 merged in situ atmospheric trace gases and flask data. ORNL DAAC. <https://doi.org/10.3334/ORNLDAAC/1593>
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. <https://doi.org/10.1002/qj.828>
- Deng, F., & Chen, J. M. (2011). Recent global CO<sub>2</sub> flux inferred from atmospheric CO<sub>2</sub> observations and its regional analyses. *Biogeosciences*, 8(11), 3263–3281. <https://doi.org/10.5194/bg-8-3263-2011>
- Denning, A. S., Takahashi, T., & Friedlingstein, P. (1999). Can a strong atmospheric CO<sub>2</sub> rectifier effect be reconciled with a “reasonable” carbon budget? *Tellus B: Chemical and Physical Meteorology*, 51(2), 249–253. <https://doi.org/10.3402/tellusb.v51i2.16277>
- 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. *Atmospheric Measurement Techniques*, 10(2), 549–563. <https://doi.org/10.5194/amt-10-549-2017>
- Gaudet, B. J., Davis, K. J., Pal, S., Jacobson, A. R., Schuh, A., Lauvaux, T., et al. (2021). Regional-scale, sector-specific evaluation of global CO<sub>2</sub> inversion models using aircraft data from the ACT-America project. *Journal of Geophysical Research: Atmospheres*, 126(4), e2020JD033623. <https://doi.org/10.1029/2020JD033623>
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454. <https://doi.org/10.1175/jcli-d-16-0758.1>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Holtstlag, 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. [https://doi.org/10.1175/1520-0442\(1993\)006<1825:Lvnblsd>2.0.Co;2](https://doi.org/10.1175/1520-0442(1993)006<1825:Lvnblsd>2.0.Co;2)
- Houweling, 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 CO<sub>2</sub> using GOSAT measurements. *Journal of Geophysical Research: Atmospheres*, 120(10), 5253–5266. <https://doi.org/10.1002/2014JD022962>
- 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. <https://doi.org/10.1016/j.ecolmodel.2012.02.004>
- 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. [https://doi.org/10.1175/1520-0469\(2004\)061<0607:Tocdit>2.0.Co;2](https://doi.org/10.1175/1520-0469(2004)061<0607:Tocdit>2.0.Co;2)
- Koffi, 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. *Geoscientific Model Development*, 9(9), 3137–3160. <https://doi.org/10.5194/gmd-9-3137-2016>
- 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. *Atmospheric Chemistry and Physics*, 5(2), 417–432. <https://doi.org/10.5194/acp-5-417-2005>
- 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. <https://doi.org/10.1016/j.atmosenv.2011.09.068>
- 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. <https://doi.org/10.1016/j.atmosenv.2010.02.009>
- 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. <https://doi.org/10.1126/science.aam5690>
- 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. *Atmospheric Chemistry and Physics*, 13(19), 9917–9937. <https://doi.org/10.5194/acp-13-9917-2013>
- Locatelli, R., Bousquet, P., Saunois, M., Chevallier, F., & Cressot, C. (2015). Sensitivity of the recent methane budget to LMDz sub-grid-scale physical parameterizations. *Atmospheric Chemistry and Physics*, 15(17), 9765–9780. <https://doi.org/10.5194/acp-15-9765-2015>
- 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 CO<sub>2</sub> mole fractions detected with a tower-based network in the U.S. upper Midwest. *Journal of Geophysical Research: Biogeosciences*, 117(G1). <https://doi.org/10.1029/2011JG001781>
- Pal, S., Behrendt, A., & Wulfmeyer, V. (2010). Elastic-backscatter-lidar-based characterization of the convective boundary layer and investigation of related statistics. *Annales Geophysicae*, 28(3), 825–847. <https://doi.org/10.5194/angeo-28-825-2010>
- Pal, S., & Davis, K. J. (2021). ACT-America field campaign catalogue. ORNL DAAC. <https://doi.org/10.3334/ORNLDAAC/1862>
- Pal, S., Davis, K. J., Lauvaux, T., Browell, E. V., Gaudet, B. J., Stauffer, D. R., et al. (2020). Observations of greenhouse gas changes across summer frontal boundaries in the Eastern United States. *Journal of Geophysical Research: Atmospheres*, 125(5), e2019JD030526. <https://doi.org/10.1029/2019jd030526>
- Pal, S., Davis, K. J., Pauly, R. M., McGill, M. J., Campbell, L., Hoffman, K., et al. (2020). ACT-America: CPL-derived atmospheric boundary layer top height, Eastern US, 2016–2018. ORNL DAAC. <https://doi.org/10.3334/ORNLDAAC/1825>
- Parazoo, N. C., Denning, A. S., Berry, J. A., Wolf, A., Randall, D. A., Kawa, S. R., et al. (2011). Moist synoptic transport of CO<sub>2</sub> along the mid-latitude storm track. *Geophysical Research Letters*, 38(9). <https://doi.org/10.1029/2011GL047238>
- 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 CO<sub>2</sub> in North America, South America and Europe. *Atmospheric Chemistry and Physics*, 8(23), 7239–7254. <https://doi.org/10.5194/acp-8-7239-2008>
- Parazoo, N. C., Denning, A. S., Kawa, S. R., Pawson, S., & Lokupitiya, R. (2012). CO<sub>2</sub> flux estimation errors associated with moist atmospheric processes. *Atmospheric Chemistry and Physics*, 12(14), 6405–6416. <https://doi.org/10.5194/acp-12-6405-2012>
- Patra, P. K., Law, R. M., Peters, W., Rödenbeck, C., Takigawa, M., Aulagnier, C., et al. (2008). TransCom model simulations of hourly atmospheric CO<sub>2</sub>: Analysis of synoptic-scale variations for the period 2002–2003. *Global Biogeochemical Cycles*, 22(4). <https://doi.org/10.1029/2007GB003081>

- 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. *Atmospheric Chemistry and Physics Discussion*. (in review). <https://doi.org/10.5194/acp-2021-373>
- 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 of the United States of America*, 104(48), 18925–18930. <https://doi.org/10.1073/pnas.0708986104>
- 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 CO<sub>2</sub> inversions. *Biogeosciences*, 10(10), 6699–6720. <https://doi.org/10.5194/bg-10-6699-2013>
- 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 CO<sub>2</sub> fluxes estimated from OCO-2 retrievals. *Atmospheric Chemistry and Physics*, 19(20), 13267–13287. <https://doi.org/10.5194/acp-19-13267-2019>
- 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 CO<sub>2</sub> transport. *Geoscientific Model Development*, 11(11), 4489–4513. <https://doi.org/10.5194/gmd-11-4489-2018>
- 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. <https://doi.org/10.5194/bg-7-1625-2010>
- 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 CO<sub>2</sub> surface flux estimates. *Global Biogeochemical Cycles*, 33(4), 484–500. <https://doi.org/10.1029/2018gb006086>
- 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. *Geoscientific Model Development*, 13(9), 3839–3862. <https://doi.org/10.5194/gmd-13-3839-2020>
- Sweeney, C., Karion, A., Wolter, S., Newberger, T., Guenther, D., Higgs, J. A., et al. (2015). Seasonal climatology of CO<sub>2</sub> across North America from aircraft measurements in the NOAA/ESRL global greenhouse gas reference network. *Journal of Geophysical Research: Atmospheres*, 120(10), 5155–5190. <https://doi.org/10.1002/2014jd022591>
- 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. <https://doi.org/10.1038/s41586-020-2849-9>
- 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 CO<sub>2</sub>. *Journal of Geophysical Research: Atmospheres*, 112(D4). <https://doi.org/10.1029/2006JD007410>
- 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*, 8, e2020EA001634. <https://doi.org/10.1029/2020EA001634>
- Weir, B., Ott, L. E., Collatz, G. J., Kawa, S. R., Poulter, B., Chatterjee, A., et al. (2021). Bias-correcting carbon fluxes derived from land-surface satellite data for retrospective and near-real-time assimilation systems. *Atmospheric Chemistry and Physics*, 21(12), 9609–9628. <https://doi.org/10.5194/acp-21-9609-2021>
- Whittaker, L. M., & Horn, L. H. (1984). Northern Hemisphere extratropical cyclone activity for four mid-season months. *Journal of Climatology*, 4(3), 297–310. <https://doi.org/10.1002/joc.3370040307>
- Wofsy, S. C. (2011). HIAPER 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. <https://doi.org/10.1098/rsta.2010.0313>
- 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*. ORNL Distributed Active Archive Center. <https://doi.org/10.3334/ORNLDAAC/1581>