Examining CO2 model observation residuals and their implications for carbon fluxes and transport using ACT-America observations

Tobias Gerken[1], Sha Feng[2], Klaus Keller[3], Thomas Lauvaux[4], Joshua P. Digangi[5], Yonghoon Choi[5], Bianca Baier[6], and Kenneth J Davis[3]

[1]James Madison University
[2]Pacific Northwest National Lab
[3]The Pennsylvania State University
[4]LSCE IPSL
[5]NASA Langley Research Center
[6]University of Colorado-Boulder

November 21, 2022

Abstract

Atmospheric CO2 inversion typically relies on the specification of prior flux and atmospheric model transport errors, which have large uncertainties. Here, we use ACT-America 30 airborne observations to compare total CO 2 model-observation mismatch in the eastern U.S. and during four climatological seasons for the mesoscale WRF(-Chem) and global scale CarbonTracker/TM5 (CT) models. Models used identical surface carbon fluxes, and CT was used as CO 2 boundary condition for WRF. Both models show reasonable agreement with observations, and CO 2 residuals follow near symmetric peaked (i.e. non-Gaussian) distribution with near zero bias of both models (CT: -0.34 +/- 3.12 ppm; WRF: 0.82 +/- 4.37 ppm). We also encountered large magnitude residuals at the tails of the distribution that contribute considerably to overall bias. Atmospheric boundary-layer biases (1-10 ppm) were much larger than free tropospheric biases (0.5-1 ppm) and were of same magnitude as model-model differences, whereas free tropospheric biases were mostly governed by CO2 background conditions. Results revealed systematic differences in atmospheric transport, most pronounced in the warm and cold sectors of synoptic systems, highlighting the importance of transport for CO2 residuals. While CT could reproduce the principal CO2 dynamics associated with synoptic systems, WRF showed a clearer distinction for CO2 differences across fronts. Variograms were used to quantify spatial coherence of residuals and showed characteristic residual length scales of approximately 100 km to 300 km. Our findings suggest that inclusion of synoptic weather-dependent and non-Gaussian error structure may benefit inversion systems.

Examining CO_2 model observation residuals and their implications for carbon fluxes and transport using ACT-America observations

Tobias Gerken[1,*], Sha Feng[1,], Klaus Keller[2,3], Thomas Lauvaux[4], Joshua P. DiGangi[5], Yonghoon Choi[5,6], Bianca Baier[7,8], Kenneth J. Davis[1,3]**

[1]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA 16802, USA

[2]Department of Geosciences, The Pennsylvania State University, University Park, PA 16802, USA

[3]Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA 16802, USA

[4]Laboratoire des Sciences du Climat et de l'Environnement, CEA, CNRS, UVSQ/IPSL, Université Paris-Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette CEDEX, France

[5]NASA Langley Research Center, Hampton, VA 2368, USA

[6]Science Systems and Applications Inc., Hampton, VA 23681, USA

[7]Cooperative Institute for Research in Environmental Sciences, University of Colorado-Boulder, Boulder, CO 80309, USA

[8]NOAA Global Monitoring Laboratory, Boulder, CO 80305, USA

[*]Now at: School of Integrated Sciences, James Madison University, Harrisonburg, VA 22807, USA

[**]Now at: Atmospheric Sciences & Global Change Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA

Key Points:

- CO_2 observed by aircraft in Eastern U.S. compares well to models, but residuals are strongly non-Gaussian.
- Model biases affect representation of cross-frontal CO_2 gradients governing CO_2 transport in storms.
- Inversion models may benefit from model-data mismatch errors dependent upon synoptic sector.

Corresponding author: Tobias Gerken, `gerkentx@jmu.edu`

Abstract

Atmospheric CO_2 inversion typically relies on the specification of prior flux and atmospheric model transport errors, which have large uncertainties. Here, we use ACT-America airborne observations to compare total CO_2 model-observation mismatch in the eastern U.S. and during four climatological seasons for the mesoscale WRF(-Chem) and global scale CarbonTracker/TM5 (CT) models. Models used identical surface carbon fluxes, and CT was used as CO_2 boundary condition for WRF. Both models show reasonable agreement with observations, and CO_2 residuals follow near symmetric peaked (i.e. non-Gaussian) distribution with near zero bias of both models (CT: -0.34 ± 3.12 ppm; WRF: 0.82 ± 4.37 ppm). We also encountered large magnitude residuals at the tails of the distribution that contribute considerably to overall bias. Atmospheric boundary-layer biases (1–10 ppm) were much larger than free tropospheric biases (0.5–1 ppm) and were of same magnitude as model-model differences, whereas free tropospheric biases were mostly governed by CO_2 background conditions. Results revealed systematic differences in atmospheric transport, most pronounced in the warm and cold sectors of synoptic systems, highlighting the importance of transport for CO_2 residuals. While CT could reproduce the principal CO_2 dynamics associated with synoptic systems, WRF showed a clearer distinction for CO_2 differences across fronts. Variograms were used to quantify spatial coherence of residuals and showed characteristic residual length scales of approximately 100 km to 300 km. Our findings suggest that inclusion of synoptic weather-dependent and non-Gaussian error structure may benefit inversion systems.

1 Introduction

To understand ongoing and future global climate change, it is necessary to improve our understanding of the terrestrial carbon cycle. Increasing atmospheric CO_2 concentrations from the combustion of fossil fuels and land-use change are partially balanced by carbon uptake in the terrestrial biosphere (Myhre et al., 2013). While the global carbon budget is constrained to a reasonable degree (Ciais et al., 2013), regional sources and sinks (Peylin et al., 2013; Crowell et al., 2019) as well as future trends (Friedlingstein et al., 2014) are much less well understood and cannot be easily diagnosed from terrestrial biosphere models, as they disagree substantially in temporal dynamics and the sign of carbon uptake(Huntzinger et al., 2012).

Atmospheric inversion systems provide a *top-down* approach to estimating terrestrial carbon fluxes and a complementary perspective to ecosystem models (Gurney et al., 2002; Bousquet et al., 1999). Most inversion models rely on both prior estimates of ecosystem carbon fluxes and atmospheric transport models to optimize fluxes with respect to observed atmospheric CO_2 mole fractions ([CO_2]). They are subject to uncertainties arising from limited observations of atmospheric [CO_2], atmospheric model transport errors, and uncertain prior flux estimates. Model transport errors in particular are widely considered to be a major source of uncertainty for atmospheric inversion systems (Peylin et al., 2005; Baker et al., 2006; Stephens et al., 2007; Gerbig et al., 2008; Chevallier et al., 2010; Lauvaux & Davis, 2014; Díaz-Isaac et al., 2014; Schuh et al., 2019). For example, Stephens et al. (2007) demonstrated the far reaching effects of the atmospheric transport model choice by showing that substantial biases in atmospheric CO_2 gradients (i.e. vertical mixing) resulted in considerable differences in estimated regional fluxes and Peylin et al. (2013) found a large uncertainty in North American terrestrial carbon sink ($0.75 \pm 0.45\,\mathrm{PgC\,y^{-1}}$) in a comparison of atmospheric inversion systems, highlighting the role of transport uncertainty for atmospheric inversion. Uncertainty attributed to transport models appears to be independent of regional sampling density, such that tropical and extratropical regions exhibit similar transport uncertainties (Basu et al., 2018). While additional atmospheric CO_2 observations in the tropics are crucially needed to constrain regional carbon balances, quantification and reduction of transport uncertainty is a priority for improving flux estimates in North America.

With respect to regional inversion systems, if was found that different atmospheric boundary-layer (ABL) parameterizations can cause substantial changes in regional inverse flux estimates (Lauvaux & Davis, 2014) due to differences in ABL depth and vertical mixing strength. Also, all physical parameterizations within one numerical weather model lead to considerable variability in ABL CO_2 (Díaz-Isaac et al., 2018). The impact of atmospheric mixing strength on inversion results is exacerbated by the fact that the CO_2 mass balance in inversion models must be maintained, which then leads to erroneous latitudinal transport of CO_2. Transport uncertainty clearly manifests itself in ABL CO_2 mole fractions, and large differences have been found within global and regional atmospheric models (e.g. Díaz-Isaac et al., 2018; Chen, Zhang, Lauvaux, et al., 2019; Schuh et al., 2019). At the same time, Gaubert et al. (2019) recently challenged the notion that vertical CO_2 gradients were the dominant cause of uncertainty in the North American carbon sink for current global inversions, and suggested that uncertainties in the fossil fuel prior were responsible.

Feng, Lauvaux, Davis, et al. (2019) showed that both fossil fuel fluxes and continental boundary conditions play important roles in the uncertainty in ABL CO_2 in addition to atmospheric transport, but concluded that biogenic fluxes, the typical objective of atmospheric inverse analyses, are the largest source of uncertainty.

While atmospheric inversions have been crucial for estimating global to continental scale carbon sources and sinks, limited progress has been made in constraining regional carbon fluxes on seasonal scales. The coarse resolution of transport models in global inversion systems (typically $1^\circ \times 1^\circ$ or coarser) may limit their ability to resolve finer scale atmospheric transport in weather systems and complex terrain (Geels et al., 2007). Regional inversions with higher model resolutions, such as CarbonTracker-Lagrange (Hu et al., 2019), have been successfully applied to constrain ecosystem carbon fluxes at regional (Lauvaux, Schuh, Bocquet, et al., 2012; Lauvaux, Schuh, Uliasz, et al., 2012; Schuh et al., 2013) and continental (Hu et al., 2019) scales, but rely on high density CO_2 observations as well as the model's ability to reproduce boundary layer processes and synoptic weather systems. Synoptic systems in the northern mid-latitudes are responsible for up to 70% of CO_2 variability through advection and are the dominant mechanism of day to day CO_2 variability in the ABL, and synoptic scale fronts create large contrasts in near surface CO_2 (Parazoo et al., 2008, 2011). Parazoo et al. (2012) highlighted that CO_2 flux estimates were highly sensitive to such synoptic scale gradients.

It is therefore desirable that transport models are capable of producing relevant frontal processes such as (i) advection of upstream CO_2 gradients (e.g. Keppel-Aleks et al., 2011, 2012), (ii) moist convective lifting of ABL air and (Schuh et al., 2019) (iii) modification of ecosystem CO_2 exchange due to weather effects (e.g. Chan et al., 2004). Comparing global inversion system's ABL dynamics, vertical mixing, and convection at frontal boundaries were also identified as priorities for improving CO_2 flux estimates in the northern mid-latitudes (Schuh et al., 2019).

The CarbonTracker (Peters et al., 2007) global inversion modeling system uses the Transport Model Version 5 (TM5) atmospheric model (Krol et al., 2005) with ECMWF (European Centre for Medium Range Weather Forecasting) ERA-Interim reanalysis meteorological drivers to estimate surface fluxes of CO_2. TM5's spatial resolution above North

America is $1° \times 1°$. CarbonTracker ingests a variety of global CO_2 data sources including daily flask observations, hourly surface time series data, and aircraft observations (Andrews et al., 2014; Sweeney et al., 2015) and and can be used as a reference point for inversion systems.

The NASA funded Atmospheric Carbon and Transport (ACT) -America Earth Venture Suborbital Mission was designed to observe atmospheric CO_2 and CH_4 mole fractions in the central and eastern United States, the dominant region for North American ecosystem CO_2 fluxes and atmospheric [CO_2] variability, and provide the observational basis for improving regional flux inversions in this region and across the midlatitudes. The ecosystem fluxes, atmospheric CO_2 mole fractions (Sweeney et al., 2015) and weather patterns all exhibit strong seasonal variability (e.g. Merrill & Moody, 1996). ACT-America sampled atmospheric CO_2 and CH_4 and associated weather variables across (i) multiple altitudes, (ii) fair weather and frontal conditions (including cross-frontal differences), (iii) multiple regions, and (iv) all four meteorological seasons within the scope of five, six-week flight campaigns. ACT-America provides an ideal test-bed for exploring the ability of atmospheric models to simulate atmospheric CO_2 across weather systems typical of the central and eastern United States, and thus shed light on both global and regional atmospheric inversion system behavior.

In this work, we compare atmospheric CO_2 model-observation differences between ACT-America data using both the global CarbonTracker inversion system and the mesoscale Weather Research and Forecasting model (Skamarock et al., 2008) coupled with chemistry, commonly known as WRF-Chem, which was run for the ACT-America study domain using CabonTracker surface carbon fluxes and lateral boundary conditions. For simplicity, we use WRF throughout this paper, when referring to WRF-Chem and CT when referring to the specific CarbonTracker-data used in this work (see Methods). CarbonTracker is used when we refer to the overall inversion system. This experiment thus focuses on how these two different transport systems represent atmospheric CO_2 with respect to the ACT observations given the same fluxes.

We analyze the properties of CO_2 model-observation differences along flight tracks and establish a baseline and general approach for comparing mesoscale (WRF) and continental scale (CarbonTracker) model errors, which can be further extended to other atmospheric inversion (e.g. CarbonTracker-Lagrange) or regional modeling systems. Model-data residuals are investigated as a function of region, altitude, climatological season, and airmass associated with frontal structure. These analyses – and the frontal analysis in particular – enable a comparison of the mesoscale and continental scale models for atmospheric conditions that are important to CO_2 transport. At the same time, these synoptically active conditions are often avoided in airborne networks such as the NOAA CCGG (Carbon Cycle and Greenhouse Gases) Aircraft Program (Sweeney et al., 2015) and partially hidden from satellite remote sensing due to cloud interference (e.g. Parazoo et al., 2008).

This paper investigates total model-data mismatch; our results are intended to guide future diagnostic studies that will separate flux and transport errors.

2 Materials and Methods

2.1 ACT-America aircraft observations

This work uses 5 s averaged aircraft CO_2 dry mole fractions measured using a PICARRO G2401-m cavity ring down spectrometer and $[CO_2]$ calibration is traceable to X2007-scale. Data are published as part of the *ACT-America: L3 Merged In Situ Atmospheric Trace Gases and Flask Data, Eastern USA* dataset (Update: 2019-03-04) (Davis et al., 2018), which is freely available from the Oak Ridge National Lab Distributed Archive Center (ORNL DAAC) (Wei et al., in review). The NASA Langley Beechcraft B-200 King Air and the NASA Goddard Space Flight Center's C-130H aircraft were used to collect high quality insitu and remote sensing measurements across the Eastern United States. Given the average speed of the aircraft (100 and 120 $\mathrm{m\,s^{-1}}$, respectively), the 5 s averaged aircraft observations have a spatial resolution of 500-600 m (Chen, Zhang, Lauvaux, et al., 2019). Data used in this work were collected during four intensive observation period flight campaigns aligning approximately with climatological seasons. We use these campaigns as proxies for seasonal greenhouse gas behavior.

During each of the flight campaigns aircraft were operated from 3 different bases (Wallops/Norfolk, Virginia; Lincoln, Nebraska; Shreveport, Louisiana), which approximately correspond to study domains (Table 1) referred to as NorthEast Mid-Atlantic (NEMA), Mid-West (MW), and South Central (SC) U.S. We are using geographic coordinates of individual measurement locations to delineate flight regions. The South Central U.S. are defined as flights as Texas, Oklahoma, and the area south of latitude N 37.00°

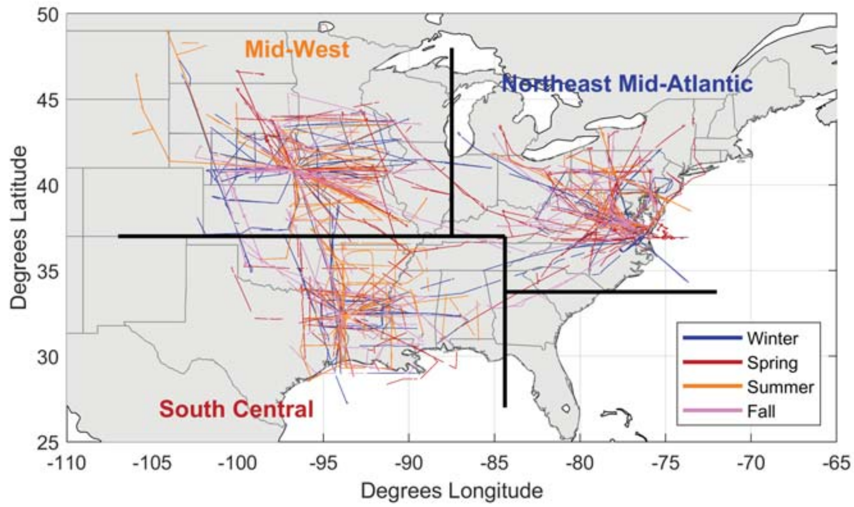

Figure 1. Overview of ACT-America observation data considered in this work. Colored lines indicate level-leg flight tracks by campaign. Study sub-regions as outlined in text are also indicated.

(Latitude of the Oklahoma-Kansas border) as well as west of longitude E 84.39° (Longitude of the city of Atlanta). The Mid-West U.S. region is defined as the area north of N 37.00° and west of E 87.5° (Longitude of the Illinois-Indiana border) and extending south to N 33.75° (Latitude of Atlanta), but excluding the area previously defined as South Central. The geographic distribution of flight observations used in this work is displayed in Figure 1.

We divide aircraft data into three altitude classes which roughly correspond to the atmospheric boundary-layer (<1.5 km; all altitudes in above ground level), the lower free troposphere that is frequently affected by convective clouds and mixing (≥1.5 to <4.0 km; LFT), and higher free troposphere which is less often affected by convection and thus might be akin to background conditions (≥4 km) (Sweeney et al., 2015; Baier et al., 2020). Flight planning during the ACT-America campaign was cognizant of these altitude classes. For example, despite large diurnal and seasonal variability of ABL-heights, ACT-America flight legs below 1.5 km altitude attempted to stay within the ABL by maintaining, whenever possible, a flight altitude of 330 m AGL. The next level is altitude was specifically selected to be above the ABL, depending on forecasts and ABL depths observed in flight. To reduce mis-classification of vertical levels, we confined our analysis to flight segments that were classified as level-legs, meaning without considerable (>500 m) flight altitude changes indicative of either vertical profiling or maneuvers to evade clouds during visual flight rules, as defined in the ACT-America maneuver flag product produced by the *ACT_ManeuverFlags* algorithm Version 1.0 (Gerken, 2019).

Additionally, ACT observations are classified by airmass conditions. Flights were planned to sample synoptic systems by flying cross-frontal transects through cold and warm sectors of the system. Similarly, fair weather flights were planned to sample fair weather conditions as well as pre-frontal warm airmasses and post-frontal cold airmasses. During days when no frontal crossings were flown, all data were either attributed to cold/ warm airmasses or fair conditions depending flight location with respect to the synoptic systems as indicated by National Weather Service surface analysis maps. During flights when fronts (typically cold fronts) were crossed, data flags were manually assigned to separate flights into warm and cold airmasses based on equivalent potential temperature (θ_e), wind, and trace gas changes across fronts. Airmass flags and flight type flags are published on the ONRL DAAC as part of the ACT L3 merged data set (Davis et al., 2018). ACT research flights were typically conducted from local time mid-morning – i.e. after the development of a sufficiently deep convective ABL for aircraft operation within the ABL – to late afternoon, corresponding to range of the C-130 aircraft and to avoid night-time conditions and collapsed ABLs.

2.1.1 CarbonTracker

We use total posterior atmospheric [CO_2] from NOAA's CarbonTracker (Peters et al., 2007, with updates documented at http://carbontracker.noaa.gov) available from the NOAA Global Monitoring Laboratory. Give that our ACT-America research period spans the years from 2016 to 2018, the CT2017 release is used for the summer 2016 campaign, while other seasons use the CT-NRT.v2019-2 (CarbonTracker – Near-Real Time). CT-NRT, designed to extend CarbonTracker between official releases, employs the same TM5 atmospheric model, while assimilating a smaller subset of [CO_2] observations. Similarly, real-time meteorology and a simplified terrestrial ecosystem carbon flux prior are being used for CT-NRT. A recent two-season comparison between CT-NRT and ACT aircraft observations found overall reasonable agreement between modeled and observed [CO_2],

Table 1. ACT-America Aircraft Campaigns

Campaign	Region	Start & End Dates	# Flight Days[a]
Summer 2016	Northeast Mid-Atlantic	June 18–27	7
	Mid-West	Aug 01–14	10
	South Central	Aug 16–28	9
Winter 2017	South Central	Jan 30 – Feb 12	8
	Mid-West	Feb 13–26	9
	Northeast Mid-Atlantic	Feb 27 – Mar 10	9
Fall 2017	Northeast Mid-Atlantic	Oct 03–14	7
	Mid-West	Oct 16–27	8
	South Central	Oct 30 – Nov 10	7
Spring 2018	South Central	Apr 12–22	9
	Mid-West	Apr 23 – May 02	8
	Northeast Mid-Atlantic	May 04 – May 20	9

[a] Transit flights between regions are attributed to their destination region

but substantial differences in bias between region and season (Chen, Zhang, Zhang, et al., 2019). ABL heights for CarbonTracker are obtained using NOAA's Observation Package (OBSPACK, Masarie et al., 2014) for CT2017 and CT-NRT-2019.2. We find that CT-NRT-2019.2, CT2019, and CT2017 have very similar ABL heights along ACT flight tracks.

CT2017 assimilates CO_2 observations from 254 sites to estimate a weekly set of biome-specific scaling factors for North America that are applied to prior biospheric [CO_2] flux model estimates. The scaling factors adjust the fluxes in order to minimize the difference between modeled and observed atmospheric [CO_2]. These biome-specific scaling factors are estimated independently for each of the 19 potential biomes within each TransCom regions (Gurney et al., 2002). Prior flux estimates for fossil fuel and wildfire CO_2 fluxes are not optimized. To estimate the impact of biases in prior fluxes, CT2017 uses two sets of priors (two each for terrestrial, ocean, fossil-fuel and wildfire carbon fluxes) and the final inversion result is the mean flux of the two inversions.

Two versions of the CASA model (Carnegie-Ames Stanford Approach Potter et al., 1993, 2003) are used for the terrestrial biospheric prior and originate from the GFED (Global Fire Emission Database) project (van der Werf et al., 2006; Giglio et al., 2009, 2013). Monthly net ecosystem carbon exchange from CASA as used in GFED 4.1s and GFED_CMS are scaled to 3-hourly fluxes similar to Olsen and Randerson (2004), while ensuring smooth month to month transitions following Rasmussen (1991). GFED 4.1 and GFED_CMS are also used as priors for wild-fire fluxes and rely on MODIS (MODerate resoltion Imaging Spectrometer) fire counts and CASA to estimate wildfire carbon loss.

As prior for fossil fuel emissions the ODIAC2016 and *Miller* datasets are used in CT2017. The *Miller* dataset uses estimated total global fossil fuel CO_2 emissions from the Carbon Dioxide Information and Analysis Center (CDIAC, Boden et al., 2016), which are spatially mapped to a $1°\times1°$ grid using the spatial patterns of the EDGAR4.2 inventory (Comission, 2019) and temporal distribution of Blasing et al. (2005). ODIAC (Oda & Maksyutov, 2011) emissions are also based on CDIAC, but differs in the spatial mapping of fluxes, which is based on proxy data such as power-plant locations, night-light images, and aviation tracks. Because of ODIAC's yearly temporal resolution, seasonal changes were derived using CDIAC monthly fossil fuel emission inventories (Andres et al., 2011). Diurnal and day of the week fossil fuel cycles are imposed on monthly emissions using scaling factors (Nassar et al., 2013).

For ocean basins, oceanic, instead of biospheric, CO_2 fluxes are optimized. Both ocean priors – the Ocean Inversion Flux prior (OIF, Jacobson et al., 2007) and pCO2-Clim (Takahashi et al., 2009) – are based on estimates of air-water differences in CO_2 partial pressure from either ocean inversions (OIF) or direct observations (pCO2-Clim).

Consequently, CT2017 provides a complete set of carbon surface fluxes from the terrestrial biosphere, oceans, fossil fuels and wildfires as well as atmospheric CO_2 mole fractions, which are available at 3-hourly temporal resolution and $1° \times 1°$ spatial resolutions over North America. CO_2 mole fractions are reported on TM5's 25 model layers (Krol et al., 2005), which include 6 layers below 1.5 km and 15 layers below 10 km. CarbonTracker has unrealistically large differences between the first (25 m) and second (103 m) atmospheric layer in well-mixed conditions (Díaz-Isaac et al., 2014). However, these model levels are considerably below the typical ABL level-leg flight altitude of $\sim$ 330 m AGL. CT2017 includes parameterized convective CO_2 mass-flux.

2.1.2 WRF-Chem

The mesoscale model is WRF-Chem v3.6.1 (Powers et al., 2017; Skamarock et al., 2008; Grell et al., 2005; Fast et al., 2006) with the modification to transport greenhouse gases as passive tracers described in Lauvaux, Schuh, Uliasz, et al. (2012). Trace gas boundary conditions are provided from CarbonTracker at 3-hourly interval posterior CO_2 mole fractions and surface fluxes introduced in the last subsection. An extra step is taken to assure the conservation of mass when ingesting CarbonTracker CO_2 mole fractions into the WRF-Chem domain. More details of the mass conservation of CO_2 can be found in Butler et al. (2020).

The domain of interest contains most of North America at 27 km horizontal resolution. The model has 50 levels up to 50 hPa with 20 levels in the lowest 1 km. The model meteorology is initialized every 5 days and driven with ERA5 reanalysis every 6 hours at 25 km horizontal resolution. The WRF-Chem dynamic is relaxed to ERA5 meteorology every 6 hours using grid nudging. Each meteorological re-initialization is started at a 12-hour setback from the end of the previous 5-day run. The first twelve hours of every 5-day simulation are considered spin-up and discarded from the final analysis. We also update sea surface temperature every 6 hours at 12-km resolution. Choices of the model physics parameterizations used in this experiment are documented as the baseline setup in Feng, Lauvaux, Davis, et al. (2019) and Feng, Lauvaux, Keller, et al. (2019) and model output for all ACT campaigns is archived and publicly available at the Pennsylvania State University DataCommons (Feng et al., 2020).

CO_2 fluxes in WRF are taken from CarbonTracker as described above and remain separate tracers in the model simulations. For analyses requiring total atmospheric CO_2 mole fractions, the surface flux tracers are summed and added to the boundary condition CO_2 tracer.

2.2 Analysis of CO_2 residuals

Differences between modeled and observed CO_2 are calculated by subtracting [CO_2] observed along the aircraft flight from modeled [CO_2] using the nearest neighbor in space and time. Chen, Zhang, Zhang, et al. (2019) found while comparing CT-NRT v2017 to ACT observations that temporal and spatial interpolation impacted calculated RMDSs of typically less than 0.4 ppm in the ABL, which is considerably smaller (order 10% or less) than RMSDs calculated in this work. The resulting residuals thus include both errors from model transport and surface fluxes. Given that CT and WRF use the same flux dataset, differences in residual should be a representation of differences in atmospheric transport including model resolution.

We calculate statistical measures – including bias, median deviation, root mean square deviation, and mean absolute deviation – for the entire data set as well as separated by region, season, and meteorological airmass. Confidence intervals for the above statistical measures are calculated using a block-bootstrap, which accounts for temporal autocorrelation using an *optimal block-length* approach (Politis & White, 2004; Patton et al., 2009). For each subset of the data, we also separate the dataset by vertical flight level. These divisions enable us to gain more understanding of the causes for model-data differences such as the impact of biological fluxes from different regions, and the impact of vertical mixing on continental background [CO_2].

We adopt the following notation for all quantities: The observed arithmetic mean and standard deviation of a quantity x are presented as $\overline{x} \pm \sigma$.

2.3 Variograms

To assess spatial statistics of CO_2 residuals, we compute empirical (semi-)variograms (Matheron, 1963) for each flight day:

$$\gamma(D) = \frac{1}{2|N(D)|} \sum_{N(D)} (R_i - R_j)^2, \tag{1}$$

where $N(D)$ is the set of all pairwise Euclidean distances $(i - j)$, $|N(D)|$ the number of distinct pairs, and R_i and R_j are the residuals at spatial locations i and j. Distance (on WGS84 ellipsoid) pair calculation and is performed separately for individual level-legs at each altitude level, to minimize the impact of atmospheric change. Vertical distances are not included in the variogram calculations as horizontal distances are much larger than altitude differences within the same level-leg. Subsequently, the empirical variograms for ABL, LFT, and HFT as well as WRF and CT are calculated using all distance pairs. Euclidian distance calculations are performed using *Experimental (Semi-)Variogram* version 1.4.0 (Schwanghart, 2013). Distances are binned into 36 classes using a geometric scaling between 1 and 750 km. To remove the disproportionate impact of outliers, including local CO_2 plumes (e.g. directly downwind of conventional power plants) that caused spikes of more than 100 ppm in [CO_2], on variance calculations, we only considered [CO_2] residuals between the 1^{st} and 99^{th} percentiles for the variogram.

To characterize spatial residual statistics, we fit an exponential variogram of form

$$\gamma(D) = c_0 + c_1 \left(1 - \exp\left(\frac{-D}{L}\right)\right), \tag{2}$$

to the observational data, where c_0 is the nugget (y-intercept of variogram), c_1 the sill (the limit of γ at infinite D) and L the characteristic length-scale of the variogram. As proposed by Schwanghart (2013), the range (distance at which the γ approximates the sill is assumed to be $3\,L$. The exponential fit is conducted with Matlab2018b's *lsqnonlin*-solver using weighted least squares using the inverse of the standard deviation of CO_2 residuals in each distance bin and a lower parameter bound of 0 is enforced for nugget, range, and sill.

3 Results and discussion

This study considers a total of 402,838 [CO_2] observations collected during four ACT campaigns which are compared to modeled [CO_2] from CT and WRF (Figure 2 and Supporting Table S1). The models appear to be capable of reproducing the multimodal shape of observed [CO_2], which is both caused by the seasonality of CO_2 fluxes and mixing, and the general increase of mean atmospheric CO_2 between 2016 and 2018 associated with anthropogenic carbon emissions. The resulting [CO_2] residuals for CT and WRF follow near symmetric, peaked distributions with high kurtosis ($\sim$ 59 and $\sim$42 for CT and WRF, respectively) and near zero mean (CT: -0.34$\pm$ 3.12 ppm; WRF: 0.82 $\pm$ 4.37 ppm for mean $\pm$ standard deviation). These residual distributions are clearly and significantly different (Figure 2 c) from normal distributions with identical means and standard deviations. Skewness is small compared to kurtosis (-2.1 and 2.7 for CT and WRF) but of opposite sign indicating skew towards negative bias for CT and positive bias for WRF. Note that the mode of the residual histogram is slightly positive (<0.5 ppm) for both models.

The skewness of residuals can be attributed to CT's apparent lack of modeled [CO_2] in excess of approximately 416 ppm, while WRF underpredicts [CO_2] at values below approximately 400 ppm (Figure 2 a+b). CT's more pronounced [CO_2] peak at approximately 412 ppm is attributed to the fact that CT exhibits a narrower range of modeled ABL [CO_2] during winter and spring compared to both ACT observations and WRF (Supporting

Figure S1). Consequently, CT's winter and spring [CO_2] in the ABL show much less overlap with fall and summer [CO_2] and the resulting PDF appears less smooth (Supporting Figure S1 b) compared to the corresponding PDFs of ACT observations and WRF. Furthermore, the too narrow peak in CT can be attributed mainly to the Northeast Mid-Atlantic region (Supporting Figure S2).

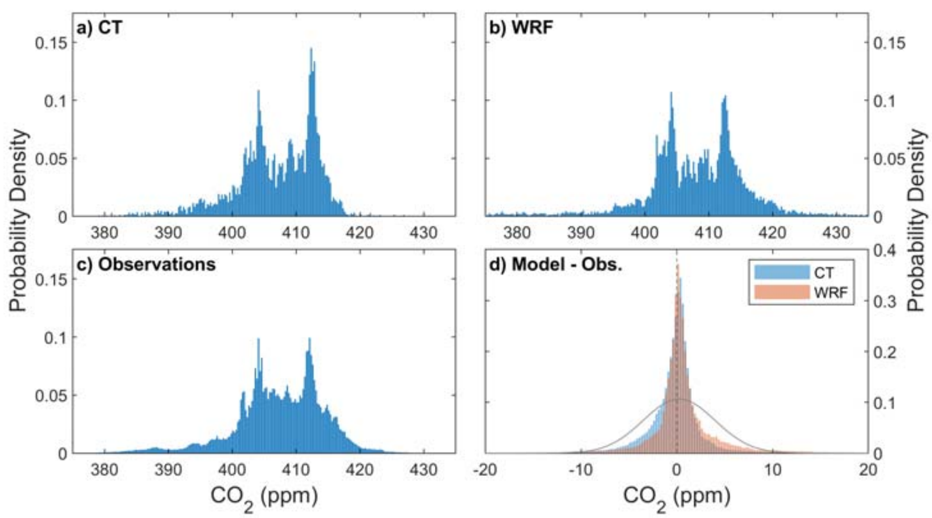

Figure 2. Overview of modeled and observed CO_2 mole fractions during four ACT campaigns 2016-2018. (a) CarbonTracker (CT); (b) WRF-Chem; (c) Aircraft observations; and (d) resulting CO_2 (Modeled – Observed CO_2) for CT and WRF. The grey line in (d) shows the normal distribution with similar mean and standard deviation to WRF residuals for reference.

3.1 Characterization of CO_2 residuals

For the remainder of the analysis, we focus on [CO_2] residuals and their spatio-temporal statistics. This limits the impact of increasing ambient [CO_2] due to fossil fuel emissions and seasonal CO_2 climatologies on our analysis. Division of residuals by altitude level (Figure 3) reveals that the total difference (Figure 2 d) in [CO_2] residual distribution between WRF and CT is primarily reflective of differences in the ABL. Here, CT exhibits a more peaked distribution with negative bias, while WRF's distribution is wider and with positive bias. The overall shape of CT and WRF residual distribution is non-Gaussian at all levels for CT and WRF and becomes markedly narrower and more peaked with increasing height, while the ABL [CO_2] exhibits pronounced *heavy tails*. Comparing the residual distributions between CT and WRF (Figure 3, right column) shows that the difference in residual PDFs in the ABL is not only due to the difference in mean residuals between CT and WRF, but also due to the opposite skewness of the underlying residual distributions. For free tropospheric levels (LFT, HFT), we find that that the difference in residual PDFs is primarily caused by a shift in the mean of the distribution (i.e. bias) rather than the shape of the distribution.

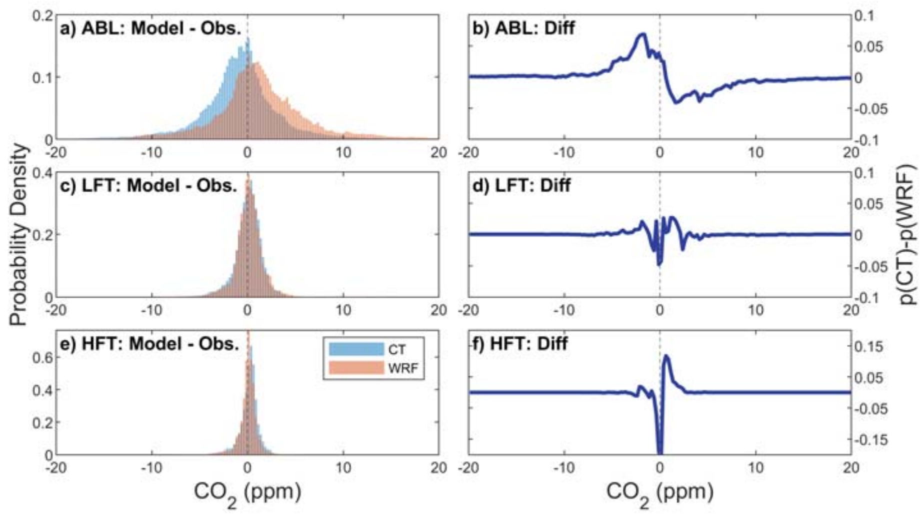

Figure 3. Probability density of model observation CO_2 residuals for CarbonTracker and WRF-Chem separated by vertical level: (a) atmospheric boundary-layer – ABL; (c) lower free troposphere – LFT; (e) higher free troposphere – HFT; and differences in their respective probability density functions (b,d,f).

Figure 3 a also reveals that while the majority of ABL [CO_2] residuals fall into a narrow range (Interquartile range of -2.76–1.07 ppm and -1.07–3.87 ppm, respectively; Supporting Table S1) compared to the entire range of residuals, residuals are heavy tailed. To characterize this larger range of residuals, we also calculated the 2.5^{th} and 97.5^{th} percentiles, which presents a compromise between representing the tail ends of the residual of the distribution, while not including outliers, which for example can result from CO_2 plumes in the vicinity of power plants.

We find that the general picture encountered for the residual PDFs (Figure 3, left column) holds generally true when residuals are separated by season, region, and airmass

(Figure 4 and Supporting Table S1). While we encounter that mean residual and IQR vary across cases (see discussion in the following section), ABL IQRs are within the range of ±5 ppm (see also Supporting Table S1). At the same time, the tails of the residual distribution are much larger in magnitude for both CT and WRF and can exceed -10 ppm and 15 ppm for the 2.5^{th} and 97.5^{th} percentiles, respectively. At the higher LFT and HFT levels, the range of the residual PDF is much smaller and typically with ±5 ppm (LFT) and ±2.5 ppm (HFT).

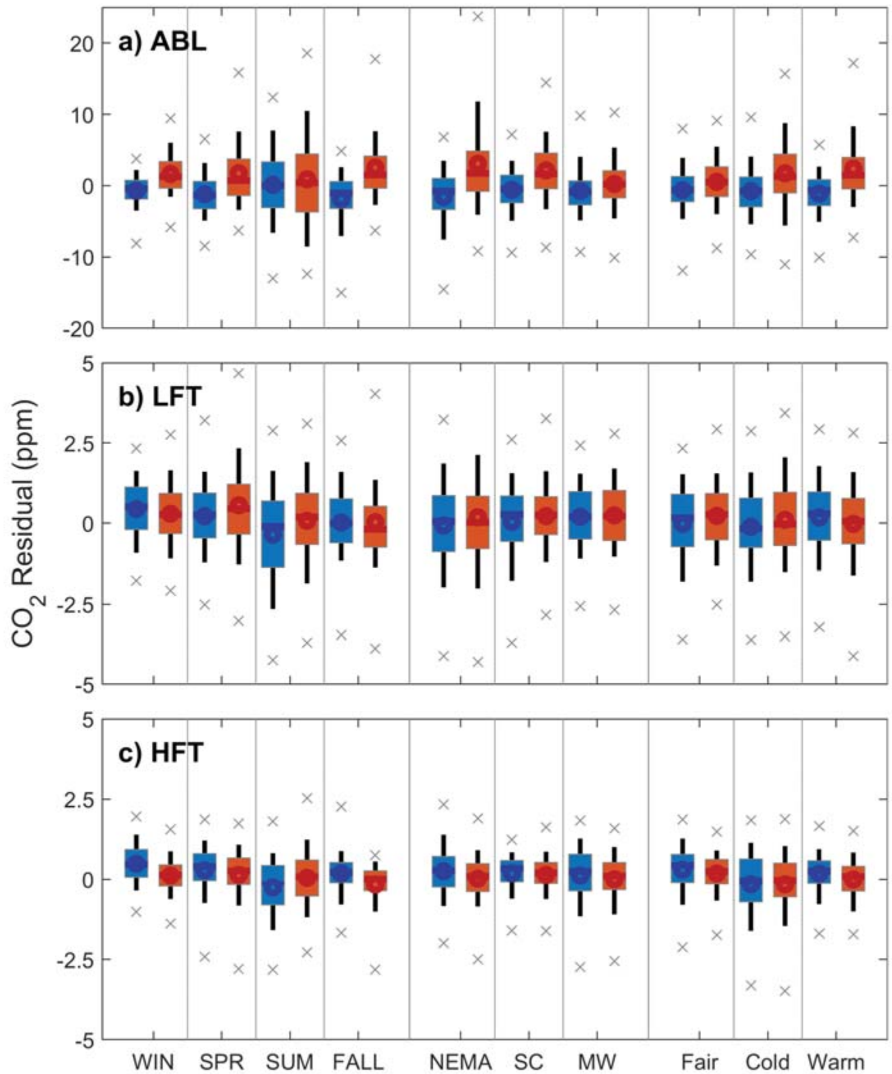

Figure 4. Box and whisker of $[CO_2]$ residual distributions from CT (blue) and WRF (red) for seasons, regions, and airmasses in (a) atmospheric boundary-layer – ABL; (b) lower free troposphere – LFT; (c) higher free troposphere – HFT. The median and mean are indicated by horizontal lines and circles, respectively. The box indicates 25^{th} and 75^{th} percentiles, whiskers 10^{th} and 90^{th} percentiles, and grey crosses indicate the 2.5^{th} and 97.5^{th} percentiles.

A quantile by quantile (Q-Q) comparison of CT and WRF residuals to normal distributions with corresponding means and standard deviations (Supporting Figure S3) further reinforces the notion of non-Gaussian [CO_2] residuals encountered for the entire dataset holding true across seasons, regions, and airmasses. The Q-Q plots also reveal the largest deviations from Gaussian behavior for CT and WRF to be at the tail ends of the residual PDFs, further highlighting the potential of large magnitude residuals to impact summary statistics such as bias or RMSD, which are commonly used to constrain inversion systems.

3.1.1 Regional, seasonal, and airmass dependent bias and RMSD

Past studies of model observation mismatch have often reported on bias and root mean square deviation (RMSD) between model and observations (Figure 5). The median residual and Mean Absolute Deviation (MAD) are reported in Supporting Figure S4. As expected, biases for LFT and HFT are much closer to zero compared to biases in the ABL. There are substantial disagreements between CT and WRF both in magnitude and sign of the bias. For higher atmospheric levels for which effects of local fluxes and mixing are less important – and thus are more likely to reflect background conditions – CT and WRF show closer agreement.

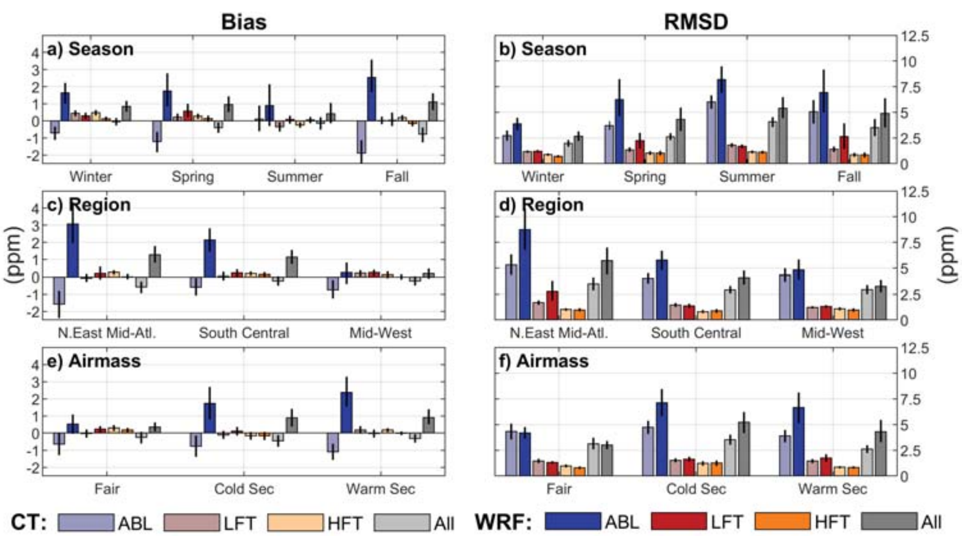

Figure 5. Comparison of CarbonTracker and WRF-Chem bias (a,c,e) and RMSD (b,d,f) for levels ABL, LFT, and HFT and separated by climatological season (a,b), region (c,d), and airmass (e,f). Bootstrapped 95% confidence intervals, using a block bootstrap (see methods) are shown in black.

With respect to climatological seasons (Figure 5 a), WRF and CT show similar behavior in the total magnitude of biases in the ABL, but signs are opposite between CT and WRF, while LFT and HFT biases are comparatively small (typically $< |0.5|$ ppm). Fall and Spring show the worst model performance for both CT (-1.89 $\pm$ 4.70 ppm; -1.22$\pm$3.50 ppm, respectively) and WRF (2.53$\pm$6.46 ppm; 1.75$\pm$5.99 ppm, respectively), followed by Winter and Summer. For Summer, CT has a near-zero bias (0.10$\pm$6.01 ppm), while the bias from WRF remains considerable (0.89$\pm$8.13 ppm). Interestingly, the comparatively large bias for Fall is confined to the ABL, while LFT and HFT biases are virtually absent. This is in contrast to Winter, when model observation mismatch in the ABL also extend to positive biases at LFT and HFT levels.

Seasonal Root Mean Square Deviation (RMSD, Figure 5 b) for CT and WRF increase from Winter to Summer and then decrease slightly during Fall, which is consistent with the frequency of occurrence for cloud convection.

Overall, the median difference (Supporting Figure S4) is much smaller than the bias for ABL, indicating that the heavy tails of the residual distribution contribute considerably to the overall bias. For LFT and HFT, median residual and bias are similar to each other.

Comparing CT and WRF residuals by study region (Figure 5 c), we find that both CT and WRF struggle in particular to accurately represent ABL [CO_2] (biases in excess of $\pm$1 ppm) in the North East Mid-Atlantic region, which has the most complex terrain of the three study regions and also exhibits complex atmospheric flow patterns. In contrast, Mid-West and South Central regions exhibit comparable biases for CT of -0.74 $\pm$ 4.26 ppm and -0.60 $\pm$ 3.99 ppm, while WRF has a high bias of 2.14 $\pm$ 5.37 ppm in the South central and a near zero bias (0.23 $\pm$ 4.82 ppm) in the Mid-West ABL. Different from seasonal RMSD patterns, regional RMSD is comparable in magnitude between CT and WRF, except for NEMA where RMSDs in ABL and LFT are $\sim$60% larger for WRF compared to CT. Generally speaking median residuals exhibit a similar behavior, but with a smaller magnitude ($<$ 1.5 ppm for all cases).

3.1.2 Comparison to previous studies

[CO_2] uncertainties over North America have been addressed in previous studies either through comparison of models with concentration tower observations or through model-model comparison.

A previous effort to characterize uncertainties stemming from biospheric carbon fluxes and atmospheric transport using perturbed WRF-Chem ensembles over North America during summer 2016 (Chen, Zhang, Lauvaux, et al., 2019) found that near surface [CO_2] uncertainties arising from fluxes (~ 6 ppm) exceed transport uncertainty (~ 4 ppm) during the daytime, while background uncertainty was less important (~ 1 ppm). In the free troposphere, the importance of flux and transport uncertainty were both reduced to ~ 1 ppm respectively (with transport uncertainty exceeding flux uncertainty), while background uncertainty remained unchanged. These uncertainties are comparable in magnitude to standard deviations of summertime model observation residuals for WRF found in this study (ABL: 8.13 ppm, LFT: 1.67 ppm; Table S1).

Chen, Zhang, Lauvaux, et al. (2019) identified the Mid-West and Mid-Atlantic as regions of largest model uncertainty due to terrestrial carbon fluxes, and noted that strong horizontal and vertical CO_2 gradients in this region also give rise to larger uncertainties due to transport. Results from our study also show large [CO_2] residuals in the NEMA region, but smaller errors in the Midwest, albeit for all seasons taken together.

Our result that model observation mismatches were largest in NEMA is supported by Chen, Zhang, Zhang, et al. (2019), who compared ACT to CT-NRT v2017 and CAMS for Summer 2016 and Winter 2017 and found negative biases for CT-NRT in the Mid-Atlantic for Summer 2016. CT-NRT's Summer 2016 ABL bias averaged across all regions was approximately -1.5 ppm while CT data used in this study had near zero bias and WRF had a positive bias of ~ 1 ppm. A comparison to CAMS (Copernicus Atmosphere Monitoring Service) showed that CAMS biases were much larger in magnitude compared to the biases found in this work. Chen, Zhang, Zhang, et al. (2019) also identified the NEMA as a region of high bias and particularly during Summer. Given the fact that NEMA is downwind of MW, which is the region of largest uncertainty in terrestrial carbon fluxes (Chen, Zhang, Lauvaux, et al., 2019; Feng, Lauvaux, Davis, et al., 2019), model data mismatches in this region are likely to result from both flux and transport uncertainty. RMSDs in this work are also comparable in magnitude to RMSDs calculated using a WRF-model ensemble of approximately 4.5 ppm for daily values and 4 ppm for 7–10 day averaging (Feng, Lauvaux, Keller, et al., 2019), who also identified the biosphere as the major source of ABL model uncertainty (~ 3 ppm). This uncertainty was invariant to averaging at less than seasonal timescales, while transport uncertainty diminished when averaged over time (~ 2 ppm and 1 ppm for averaging windows of 1 and 10 days), becoming less important than uncertainties from boundary inflow and fossil fuels. Given that ACT's insitu [CO_2] observations reflect airmass history, flux error is likely a large portion of RMSDs encountered in this work.

A tower-based comparison of WRF-Chem and Carbontracker/TM5 using CT2009 fluxes during the growing season of 2006 (Díaz-Isaac et al., 2014) highlighted the impacts of modeled near surface dynamics on ABL [CO_2]. While CarbonTracker underestimated CO_2 drawdown during summer, WRF had a tendency to overestimate drawdown, while using the same set of surface fluxes. Additionally, the authors found that WRF exhibited shallower ABLs with small within-ABL vertical gradients, indicating more well mixed conditions in the ABL compared to TM5/Carbontracker, whereas TM5 /Carbontracker showed stronger vertical mixing between ABL and free troposphere. Our results (Figure 5) show a tendency in CT to have opposite biases between ABL and LFT, which may be indicative of excess vertical mixing in CT. WRF, in contrast, has a more consistent positive bias at all levels.

Model resolution is also an important factor for model performance. A comparison of [CO_2] surface observations to the CAMS CO_2 forecasting system showed a 1.8-2.5 ppm reduction of RMSD (corresponding to 33%), when reducing horizontal model resolution from 80 km to 9 km (Agustí-Panareda et al., 2019). This was attributed to both better representation of modeled wind fields (i.e. transport) and spatial variability in surface carbon fluxes. While the WRF-Chem resolutions used in this studies had a 27 km resolution and surface fluxes were at $1^\circ \times 1^\circ$ resolution, RMSDs of order 5 ppm encountered for ACT were comparable to CAMS RMSDs at 9 km. At the same time, WRF RMSDs were larger than those of CT at the coarser 1-degree resolution, conflicting with results found by Agustí-Panareda et al. (2019). One potential explanation for this discrepancy is the fact that while neither WRF nor CT are capable of directly resolving convective cells, WRF has a sufficiently high resolution to resolve features of warm and cold fronts. Consequently, small errors in frontal location and other synoptic features can lead to large errors in modeled [CO_2] in WRF, while CT does not have the same small-scale variability and thus shows lower total bias but a less realistic distribution of [CO_2] (Figures 1 a–c and Supporting Figures S1-S2). This hypothesis is consistent with the fact that Summer, which has the most active cloud convection, shows small bias in WRF but the largest RMSD (Figure 5). Additionally, posterior carbon fluxes have been optimized for CT and not for WRF. The differing behavior between CT and WRF and the effect of flux optimization are further discussed in section 3.2. It remains to be seen whether a further reduction of WRF resolution below 27 km, which would allow for convection resolving simulations, would increase model accuracy or would further exacerbate errors due to location errors of synoptic structures, which do not appear in the coarser CT.

The CarbonTracker inversion system (Peters et al., 2007) uses RMSD between observations and atmospheric model to estimate its assumptions for model-data mismatch (MDM) that constrain the inversion system (specifically: MDM = 0.85-0.95×RMSD). CT's choice of using seasonally, regionally, and vertical level specific MDM values appears to be justified, based on our results, that residuals strong vary between region, season, and level (Figures 3–5). At the same time, other inversion systems such as CarbonTracker-Lagrange (CT-L, Hu et al., 2019) do not specify seasonally differing MDMs. Given CT-L's regional focus and finer resolution, seasonally varying MDMs appear to be advantageous given our findings of seasonally varying model residuals.

Note that the previous studies discussed here did not perform a weather aware analysis in the sense that they did not separate model observation comparisons by airmass or weather conditions. In fact, when comparing aircraft observations to models, there are likely issues of representativeness, as for example NOAA/GML Global Greenhouse Gas Reference Network profiles (Sweeney et al., 2015) are collected using small aircraft, which are limited to operating in fair weather conditions.

3.1.3 Interpretation of large residuals

Given the importance of characterizing model-observation-mismatch for atmospheric inversion results and given the fact that our model residual statistics are heavily influenced by the long tails of the [CO_2] residual PDF, we proceed to investigate what conditions are most conducive to the occurrence of large magnitude residuals. To do so, we chose to focus on the ABL [CO_2] residuals in the tails ($< 5^{th}$ and $> 95^{th}$ percentiles) of the distribution. We also chose to concentrate on WRF, which due to its higher resolution is more capable of resolving frontal structures. We find that large magnitude residuals were not randomly distributed across all flight days, but rather concentrated on specific days for which model observation residuals tended to be large. For example, the 10 days with the largest fraction of large positive [CO_2] residuals contributed to 72% of all positive large magnitude residuals. Additionally, 9 out of 10 days were associated with research flights that included a frontal crossing and 6 out of 10 days were for the NEMA region. At the negative residual tail end, we found that 10 days contributed to 68% of all large magnitude residuals. These days with large negative residuals were highly concentrated during Summer (8 out of 10) and specifically the MW region (5 days during Summer). Unlike the positive residual days, weather did not appear to play a major role during negative residual days, which may be due to the fact carbon fluxes in MW are underestimated for the MW agricultural belt, such that transport errors associated with synoptic systems do not play a considerable role.

The fact that large magnitude positive residuals are concentrated during frontal conditions, highlights the fact that CO_2 transport and associated model errors are highly dependent on synoptic scale conditions. It is likely that comparatively small errors in modeled frontal location, which arise despite WRF being nudged to ERA-5 analysis, combined with observed large cross frontal [CO_2] differences (Pal et al., 2020) can result in large [CO_2] residuals. Also, ACT observations revealed characteristic bands of elevated [CO_2] ahead of the cold front, which the WRF model may not be able to adequately reproduce. Given the importance of synoptic weather systems to mid-latitude carbon transport (e.g. Parazoo et al., 2008, 2011) as well as the large associated model residuals, weather aware specification of prior model observation mismatch could be beneficial for inversion systems and particularly regional inversions. The interplay between season and airmass on model residuals is further discussed in section 3.2.

The large contribution of MW summer to negative residuals (i.e. overestimation of modeled [CO_2] within the ABL) coincides with the fact that the U.S. Midwest is dominated by high intensity agriculture and particularly corn, which makes this region a large continental carbon sink during the agricultural growing season, leading to CO_2 depletion within the ABL. Consequently, underestimation of terrestrial carbon fluxes is a likely source of this model data mismatch for this region. At the same time, tall ABLs during summer and associated entrainment of free-tropospheric air counteract CO_2 in the ABL, but models such as WRF can have considerable random errors in ABL heights that are variable between regions (e.g. Díaz-Isaac et al., 2018). The covariance between terrestrial carbon fluxes and ABL heights (also referred to as rectifier effect, Denning et al., 1995) makes it difficult to attribute model observation differences into flux and model errors. The impact of ABL heights and specifically differences in simulated ABL mixing between models (e.g. Díaz-Isaac et al., 2014) on [CO_2] residuals is further discussed in section 3.2.3

3.1.4 Implications for inversion systems

The [CO_2] model-observation residuals encountered in this study do not follow Gaussian distributions (Figures $2+3$). Given that this study cannot separate between transport and flux errors, it is theoretically possible that the total residuals, which are non-Gaussian, are the result of normally distributed flux and transport errors which are superimposed onto each other. At the same time, we encounter non-Gaussian residuals at all vertical levels, including LFT and HFT, where the influence of surface fluxes is smaller compared to the surface, implying that transport model errors, taken in isolation, are also non-Gaussian. Atmospheric inversion systems require the specification of model-data mismatch errors in order to constrain the flux optimization. The CarbonTracker (Peters et al., 2007) and most other operational inversions require mismatches to be normally distributed and transport errors to be unbiased. While we find that the overall bias of the model-data mismatch is comparatively small, its non-Gaussian nature found in this study has the potential to impact inversion results. The heavy tails of the [CO_2] residual contribution, have outsized impacts on RMSD and standard deviation. The resulting larger RMSDs and standard deviations, which are used to prescribe Gaussian errors, in consequence, reduce the sensitivity of inversion systems to observations. Given that we find that a large fraction of large magnitude model observation mismatches stem from a small number of days, and (in the case of positive residuals) from days with frontal activity, it appears that specifying weather aware model data mismatches in regional inversion systems could increase the sensitivity of the inversion to observations and thus improve flux estimates.

3.2 Comparison of model residuals between CT and WRF

A comparison of the joint residual statistics (Figure 6) reveals their differing behavior in the two modeling systems when disaggregated according to seasons, regions, and airmasses. Given that CT and WRF use identical carbon surface fluxes and that WRF uses [CO_2] from CT as lateral boundary condition, we infer that differences between CT and WRF are due to tracer transport differences.

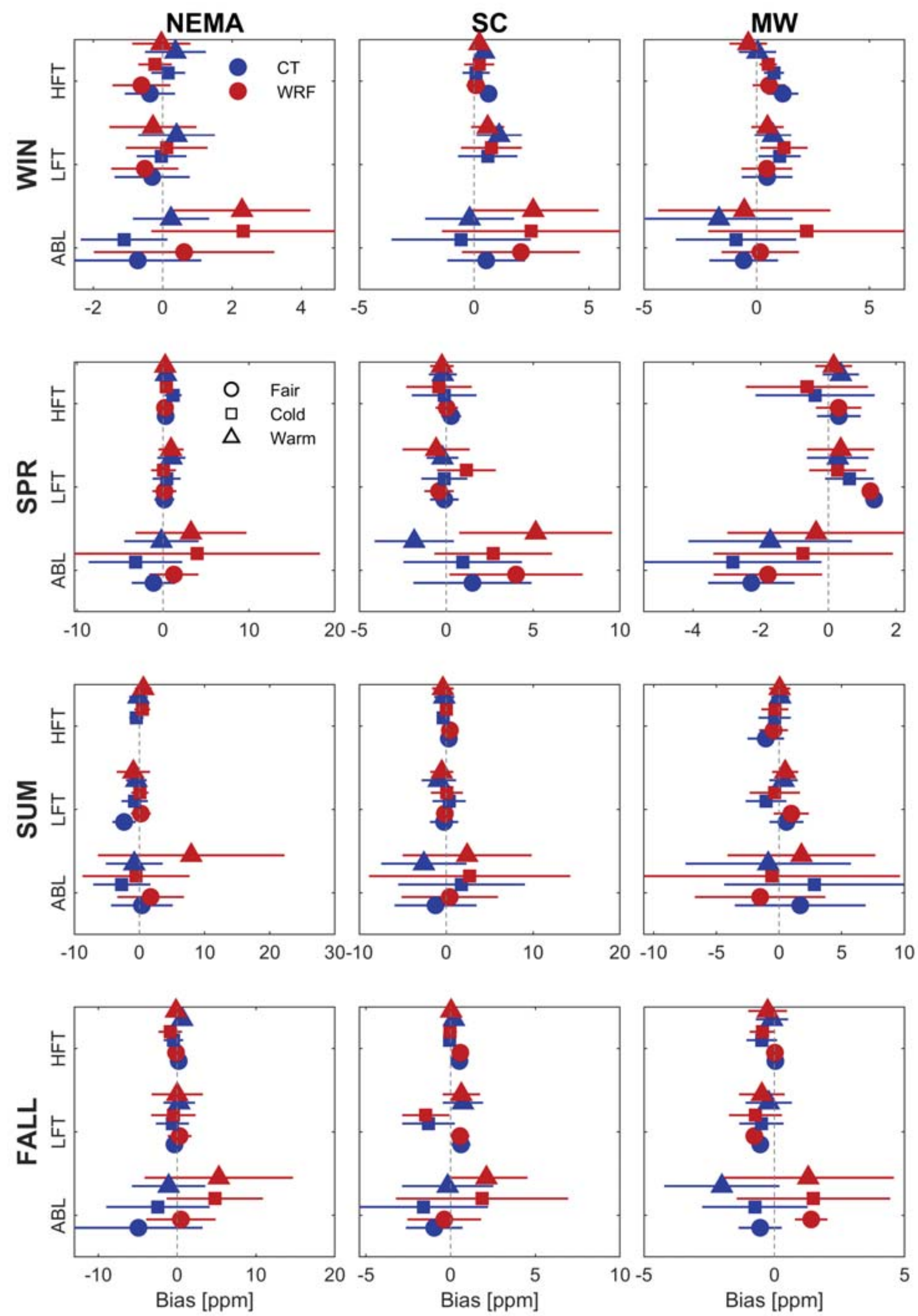

Figure 6. Comparison of CarbonTracker (as described in methods) and WRF-Chem CO_2 residual as function of climatological season (rows), region (columns) and airmass (symbols) for the three observation levels.

3.2.1 Residuals in the free troposphere

Free tropospheric [CO_2] residuals are small compared to ABL residuals for all seasons (Figures 6 and Supporting Figure S5 showing only FT and HFT), regions, and airmass conditions and mean differences between CT and WRF are smaller than their respective standard deviations. The magnitude of HFT bias is of order 0.5 ppm for both CT and WRF, while the standard deviation of residuals is of order 1 ppm. In the vast majority of cases CT and WRF bias differences are less than 0.5 ppm and show the same sign. A similar picture emerges for LFT, but with slightly larger magnitude biases and standard deviations (|0.8| ppm and 1.5–2 ppm, for mean and standard deviation, respectively). Similarly, the difference between CT and WRF also tends to be less than 0.5 ppm and to be of similar sign (Supporting Figures S6–S12). The fact that model-model mismatches are of similar magnitude to model observation mismatch, highlights the difficulties in separating the two. It is notable that WRF tends to have lower free tropospheric [CO_2] than CT, except during Summer where the reverse appears to be true for NEMA and MW, but not necessarily SC, where onshore flow of homogeneous high CO_2 air from the Gulf of Mexico occurs. Given the fact that CT and WRF employ the same carbon fluxes and WRF uses CT atmospheric mole fraction for lateral boundary conditions (i.e. background), it appears reasonable to infer that transport uncertainty between CT and WRF in the free troposphere is of order 0.5 ppm. However, given the comparatively large volume of the free troposphere compared to the atmospheric boundary-layer, even small model errors represent large quantities of carbon and will affect column averaged [CO_2] (often referred to as XCO_2), which is thought to be less sensitive to ABL dynamics and surface flux heterogeneity (e.g. Keppel-Aleks et al., 2011). Also, it has been estimated that a difference of 0.5 ppm between to boundary-condition products over North America produces an offset of $0.8\,\mathrm{PgC\,y^{-1}}$ in the North American terrestrial carbon flux, which is similar to the magnitude to the actual North American sink of around $0.5–1.0\,\mathrm{PgC\,y^{-1}}$ (Gourdji et al., 2012). This further highlights the need to further reduce uncertainties in global products. Additionally, given that the free troposphere is not in direct exchange with the surface, free tropospheric [CO_2] biases can be integrated over continental scales, making attribution of flux errors to specific regions and processes difficult.

3.2.2 Interpretation of free tropospheric differences

CarbonTracker XCO_2 (CT2015) was also shown to have good agreement with aircraft observations using the NOAA aircore network with spatial XCO_2 gradients mostly reflecting large-scale circulation (Lan et al., 2017). Therefore, comparatively small free tropospheric biases are in line with our expectations. Sweeney et al. (2015) showed that vertically homogeneous oceanic background air becomes increasingly less homogeneous with residence time over land, in response to terrestrial carbon fluxes and upward mixing of the flux signal. Since, WRF and CT employ identical fluxes, differences between CT and WRF are either due to differences in vertical mixing or airmass history. Due to the fact that ABL volume and mass are small compared to the free troposphere, vertical mixing differences are difficult to diagnose using ACT data. However, since HFT and LFT [CO_2] are lower in WRF during winter, when the terrestrial biosphere acts as a carbon source and higher during summer, when there is carbon uptake, as well as the fact the CT has been documented to have strong vertical mixing (Díaz-Isaac et al., 2014; Schuh et al., 2019), the differences between WRF and CT with respect to ABL to FT [CO_2] are consistent with different vertical mixing strengths between models. These findings are also consistent with Butler et al. (2020), who found model data mismatches to result from model transport differences below 850 hPa. Cloud convection associated with frontal lifting causes convective mass flux and presents a potentially important avenue for vertical transport of CO_2. While CarbonTracker and the underlying TM5 chemical transport model operate on an approximately 4-times coarser horizontal resolution than WRF, CarbonTracker includes parameterized convective mass fluxes taken from the parent ECMWF (European Centre for Medium-Range Weather Forecasts) model. WRF in contrast, with its finer horizontal and vertical resolution, resolves a larger portion of vertical motion, but does not presently have explicitly coupled convective tracer mass-flux associated with clouds. This omission may cause underestimation of vertical transport, which is consistent with the observed opposite sign of ABL to FT CO_2 residuals between CarbonTracker and WRF found predominantly during Winter and Spring.

3.2.3 Residuals in the atmospheric boundary layer

A different picture emerges for ABL [CO_2] residuals (Figure 6). We encounter larger differences between CT and WRF. In the ABL, CT exhibits a low bias and WRF a high bias (Figure 3 a) for most seasons and regions. Exceptions include Summer in MW, for which CT has high bias in cold and fair conditions, while WRF shows the opposite behavior. The differences in mean residual between models are generally smaller than one standard deviation, highlighting the large temporal and spatial variability of model observation residuals. In general, we find the largest variation in [CO_2] residuals as indicated by their standard deviation during Summer conditions, which have the most active cloud convection and biosphere fluxes. Winter, Spring, and Fall exhibit much less variation, except for the NEMA region. In NEMA the standard deviations of residuals remain large during Spring and Fall, which may be due to topographic effects, long continental upwind trajectories, and regional fossil fuel emissions. Fair weather conditions on average are not only associated with the lowest magnitude in bias, but also show the smallest differences between CT and WRF across seasons and regions.

This work investigates total ABL [CO_2] biases as the difference between modeled and observed [CO_2] which consist of flux errors and transport uncertainty. Despite prescribed carbon fluxes being optimized to continental [CO_2] observations, ABL [CO_2] bias magnitudes for specific regions and seasons are approximately 1–3 ppm in Winter, 1–10 ppm during Summer and 1–5 ppm during Spring and Fall, highlighting the remaining uncertainties associated with biospheric carbon fluxes and atmospheric transport.

Because we use posterior carbon fluxes from CT, one can expect CT to show smaller magnitude biases compared to WRF which has a different atmospheric transport of CO_2 – care should be taken to not interpret bias differences between WRF and CT as differences in model quality. Instead, model-model differences between CT and WRF should be seen to reflect transport uncertainty. Model-model differences follow a similar seasonal pattern compared to bias but can reach slightly larger magnitudes (Winter: 1–4 ppm; Spring: 1–8 ppm; Summer: 1–10 ppm; Fall 1–8 ppm). We find larger model-model differences for warm and cold airmasses associated with synoptic systems compared to fair weather conditions. Also, biases are generally smaller in magnitude for fair weather, highlighting the role of dynamics processes on model performance. Model-model differences encountered in our work are larger in magnitude compared to values found by Chen, Zhang, Lauvaux, et al. (2019). A recent study using 45 different combinations of physical parameterizations in WRF (Díaz-Isaac et al., 2018) revealed ABL CO_2 transport uncertainties of 3–4 ppm. Since model-observation differences are in the same range as model-model differences, we can infer that transport uncertainty is a large contributor total ABL [CO_2] biases and must be resolved before reaching conclusions about flux errors alone.

In addition to substantial vertical [CO_2] gradients resulting from ABL enrichment or depletion of [CO_2] due to surface fluxes Pal et al. (2020) encountered large horizontal cross frontal [CO_2] differences during Summer 2016, which arise from differences in airmass history as well as modification of surface fluxes in response to cloud shading and reduced ABL mixing (e.g. Chan et al., 2004; Pal et al., 2020). Given the importance of cross frontal CO_2 differences for atmospheric CO_2 transport and potentially inversion system performance, it is important for atmospheric models to accurately represent these cross-frontal [CO_2] differences. Especially during Summer, when differences are largest with 5-30 ppm (Pal et al., 2020), we find differences in warm and cold sector bias in both models to exceed 5 ppm, such that modeled cross-frontal [CO_2] differences can differ considerably from observations. Smaller bias differences are found for the other seasons and regions, except for Winter in MW. However, given that Summer is a season with high convective activity and large terrestrial biogenic carbon fluxes, misrepresentation of cross-frontal gradients may have substantial impact on modeled atmospheric carbon fluxes and thus atmospheric inversions. This finding further highlights the potential need for weather-aware inversion approaches.

3.2.4 Potential sources of mismatch

The ABL is in direct contact with both the surface and the free troposphere, thus making accurate prediction of ABL [CO_2] a particularly challenging problem. Despite using posterior biospheric CO_2 fluxes from CT, considerable uncertainty in surface carbon fluxes remains an issue. Additionally, CT is optimized to continental scale CO_2 observations and large variation of bias exists between regions, seasons, and airmass conditions. Besides surface fluxes, ABL growth and resulting entrainment of free tropospheric air into the ABL as well as convection lead to CO_2 exchange between ABL and LFT. Given the importance of vertical mixing for inversion accuracy (Stephens et al., 2007; Peylin et al., 2013; Schuh et al., 2019) we proceed to investigate potential impacts of ABL depth (in conjunction with ABL to LFT [CO_2] differences) on model-model bias differences. We find that CT tends to exhibit deeper ABLs for all seasons except Fall (Figure 7), which would be consistent with, CT's demonstrated low bias for Winter and Spring (when ABLs are enriched in CO_2 compared to LFT) as well as the high bias during Summer (when ABLs are depleted in CO_2). However, a more complicated picture emerges, when taking into account observed vertical [CO_2] differences (Supporting Figure S13).

One caveat is the fact that this comparison uses ABL depths directly provided from CT and WRF model output. For CT this means that ABL depths are calculated based on the Richardson number, while WRF ABL depths are diagnosed in the turbulence parameterization. However, despite these differences in ABL definition, we believe that using the ABL definition native to the modeling system should accurately reflect the model's vertical ABL mixing. Based on ABL depth differences between CT and WRF ranging from -20% to 35% and typical ABL to LFT [CO_2] differences of less than 10 ppm magnitude, we estimate the maximum impact of ABL depth differences between CT and WRF to be less than 3 ppm. Consequently, while ABL [CO_2] bias differences between CT and WRF during Winter are explainable by differences in entrainment of free tropospheric air at the ABL-top, model-model residual differences between CT and WRF within the ABL are considerably larger than 3 ppm for all seasons except Winter and can thus not be explained by entrainment alone. This result leaves cloud convection associated with frontal lifting and horizontal advection differences as the likely main source for the differing behavior between CT and WRF.

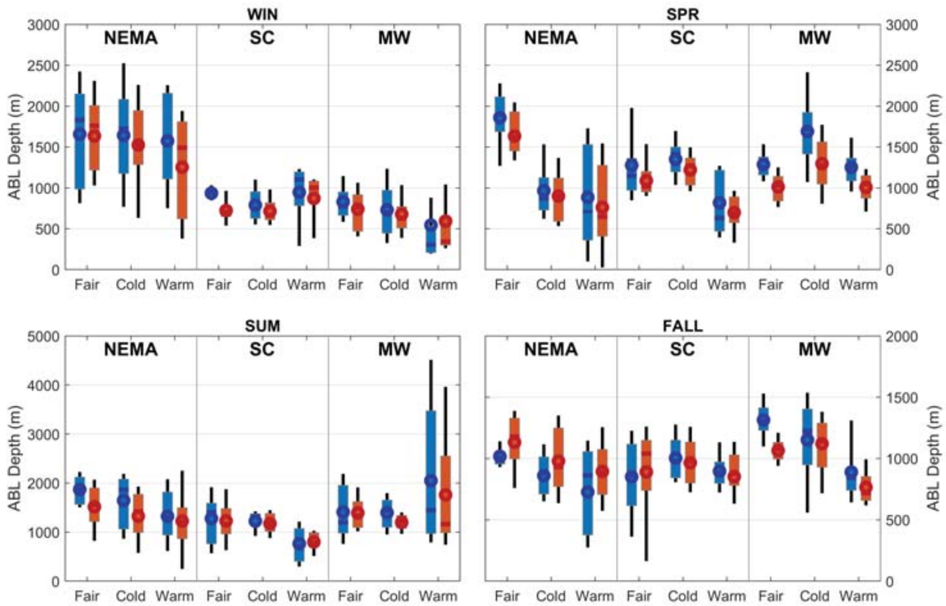

Figure 7. Comparison of CarbonTracker (as described in methods, red) and WRF-Chem (blue) diagnosed ABL heights separated by season, region, and airmass. The boxplot indicates 10^{th}, 25^{th}, 75^{th}, and 90^{th} percentiles of the distribution. The median and mean are indicated by horizontal lines and circles, respectively.

Differing [CO_2] residuals between CT and WRF highlight the importance of CO_2 transport differences in frontal systems. At the same time, considerable biases remain for CT and WRF shows larger magnitude biases than CT. We hypothesize that WRF, due to its higher resolution, is capable of reproducing frontal location and structure, while TM5 which underlies CT is less capable of doing so. CT's fluxes are optimized without taking into account transport uncertainty differences associated with frontal systems and using a model resolution that does not fully resolve synoptic scale weather. Consequently, terrestrial carbon fluxes optimized with the CarbonTracker system and applied to WRF, then lead to considerably higher [CO_2] biases in warm airmasses compared to cold airmasses in WRF, while biases in warm and cold sectors for CT, which has a coarser resolution, are more consistent.

Given the importance of midlatitude synoptic scale systems to North American meridional carbon transport, our findings support the notion that inversion systems can be improved by considering the effects of frontal passage through, for example, warm and cold sector specific prescribed model-data mismatches. At the same time, increasing prescribed model-data mismatches near fronts without addressing model biases would deemphasize observations near frontal structures in inversion systems. This would potentially reduce changes to prior fluxes in vicinity of synoptic systems, which may be especially problematic, because frontal systems present a complex environment, where surface flux priors from land-surface models such as CASA may be highly uncertain.

3.3 Spatial structure of model observation mismatch

Spatial analysis of model-observation mismatch through experimental variograms (Figure 8), confirms the previously reported findings. CT and WRF show similar structural behavior for LFT and HFT, while substantial differences emerge within the atmospheric boundary layer. We determine the spatial extent of mismatch correlations (variogram range) to be between 300 and 600 km (LFT: 267 and 309 km for CT and WRF respectively; HFT: 405 and 576 km). The corresponding variances (variogram sill) are for LFT 1.11 and 1.43 ppm^2 for CT and WRF, respectively as well as 0.48 and 0.62 ppm^2 for HFT. In the ABL, the range is estimated 356 km for CT and 693 km for WRF, while corresponding sills are 13.10 and 36.60 ppm^2. Note, however, that these values are highly uncertain as we find generally large variability of model-data mismatches, as indicated by the shading in Figure 8, compared to the average variogram. Also, we have comparatively little data that extended beyond 300 km as indicated by the drop in variance, due to the inherent limitations of airborne data collection. Therefore, larger magnitude values for range as encountered in the ABL and WRF particularly are associated with larger uncertainties in the fitting of the experimental variogram.

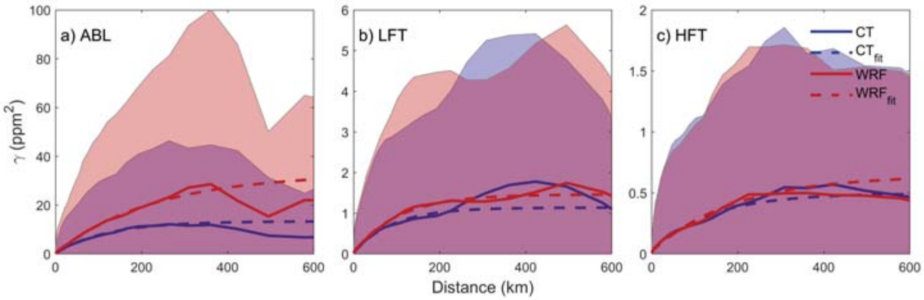

Figure 8. Experimental variogram for [CO_2] residuals for CT and WRF at levels (a) ABL; (b) LFT; and (c) HFT. The dashed lines indicate an exponential variogram fit. Shaded areas show the standard-deviation within each bin of the experimental variogram.

Overall, the mean spatial variance (γ) is small compared to the variability of model data residuals (shading in Figure 8). Unfortunately, despite ACT's more than 400,000 observations, we were not able to differentiate variogram statistics for season, region, and airmass. We hypothesize that this is due to lack of observations at large distances that preclude robust calculations of range and sill.

Recently, Lauvaux et al. (2019) investigated spatial error structures of in situ [CO_2] from tower observations and found characteristic length scales (L) of order 100–150 km during using a simple exponential ($e^{-x/L}$). Since the range of experimental variograms is assumed to be $3\times L$, we find our airborne observations comparable to the values given by Lauvaux et al. (2019). Characteristic length-scales of order 100 km imply that [CO_2] observations at the NOAA GML tall tower network (Andrews et al., 2014) are independent of each other, while sufficient averaging lengths should be applied to satellite XCO_2 measurements.

3.4 Additional considerations

Our results show distinct [CO_2] biases when observations are segregated by air mass. Consequently, model evaluations, as commonly done, that average across different synoptic conditions are likely to hide canceling biases. Also, many observational systems, such as satellites (e.g. OCO-2, Crowell et al., 2019) and the NOAA aircraft profiling efforts (Sweeney et al., 2015) selectively sample fair weather conditions, which are were found to be less biased. Resulting evaluations of model-data mismatch may thus underestimate the magnitude of transport model bias.

Biases related to air mass are likely linked to systematic differences in atmospheric transport and the systematic differences in representation of weather system [CO_2], found in this work, may propagate to global meridional transport of [CO_2]. They therefore may significantly affect global [CO_2] inversion estimates as illustrated by Schuh et al. (2019) and Barnes et al. (2016). Additional numerical studies and model-data comparisons should be undertaken to quantify this link.

The importance of simulated transport on model-data mismatch is further highlighted by the fact that CT and WRF biases are of opposite sign, despite common carbon surface fluxes. Similarly, we find differences between CT and WRF with respect to modeled cross frontal [CO_2] differences, especially during Summer when WRF over-predicts differences while CT tends to under-predict. The exact cause of this difference between models, which will affect CO_2 transport in synoptic systems (Pal et al., 2020) is currently unclear. Our results show that ABL depth alone cannot account for reported differences. Potential causes may include resolved vertical transport and parameterized cloud mass flux (Parazoo et al., 2008).

Magnitudes of model-data residuals strongly depend on airmass history and specifically travel time over land (Sweeney et al., 2015; Lan et al., 2017) during which air parcels are subject to CO_2 exchange with the biosphere. For example, warm sector airmasses originating from the south have less fetch over land compared to cold sector airmasses from the north. Therefore, southern air originating from the Gulf of Mexico provides a homogeneous CO_2 background and thus less deviation from oceanic backgrounds, while northern airmasses that traveled through areas of large biospheric carbon fluxes such as the Mid-West agricultural belt or boreal forests have much more varied [CO_2]. This highlights the importance of airmass history and transport error for model-observation mismatch. While our work points to transport error differences as one source of the model-data mismatch difference between warm and cold airmasses, a true segregation of transport from flux errors will likely require calibrated transport ensembles (Díaz-Isaac et al., 2019; Feng, Lauvaux, Keller, et al., 2019; Feng, Lauvaux, Davis, et al., 2019).

Considering this work as a naive and uncalibrated 2-member model-ensemble, we find seasonally varying model-model differences of 1–10 ppm. Within this range larger differences pertain to warm and cold airmasses, while smaller differences pertain to fair weather conditions. Unfortunately, model-model differences are in the same range as comparisons with ACT observations, such that attribution of transport errors from our work appears to be not possible, thus necessitating more targeted modeling studies.

In comparison to ABL [CO_2] residuals, residuals in the free troposphere were much lower (< 0.5 ppm in HFT) and differences between CT and WRF were small, implying that transport model errors were less important. Therefore, CO_2 observations in the higher free troposphere may in many cases serve as continental background for greenhouse gas measurements (e.g. Baier et al., 2020).

4 Conclusions

We use more than 400,000 CO_2 dry mole fraction observations collected during four flight campaigns spanning all four seasons and three regions (Northeast Mid-Atlantic, Mid-West, South-Central) in the Eastern U.S. to investigate model-observation mismatches for the WRF-Chem regional model and the global CarbonTracker system. A particular focus of this investigation and the ACT-America project in general, were synoptically active conditions, which present a major component of mid-latitude CO_2 transport and thus have the potential to greatly impact CO_2 inversion results.

Using identical carbon surface fluxes, we found that both models were capable of reproducing the [CO_2] dynamics over the Eastern U.S. At the same time, model-model mismatches and model observation mismatches were found to be strongly related to season and airmass, with synoptically active conditions and seasons to exhibit higher bias than fair weather conditions.

While errors in CT posterior fluxes likely play a considerable role in model-observation mismatch, we also qualitatively identified CO_2 transport as a major component, because the CT exhibited negative bias, while WRF had positive bias, despite common fluxes. However, it was not possible to quantify the magnitude of transport error, which was found to be due to horizontal transport rather than boundary-layer depth errors alone. While the two models used in this study could be considered a naive 2-member ensemble, further studies using carefully assembled model ensembles are needed to characterize transport uncertainty. Better quantification of transport uncertainty and improvements to transport models has the potential to improve inversion efforts as currently observations may be overly discounted in inversion products.

Comparing the lower resolution and global CT system with the WRF regional model, we find that while CT was capable of reproducing the principal [CO_2] dynamics associated with synoptic scale systems, WRF's higher resolution showed a clearer distinction between [CO_2] residuals in warm and cold airmasses. Given the stark cross frontal [CO_2] differences and the overall importance of weather systems for CO_2 transport, there is a likely benefit to making transport errors in inversion systems weather aware. This idea also highlights the potential of regional inversion systems to improve posterior carbon flux estimates. At the same time, caution should be taken because residual distributions were highly non-Gaussian and long-tailed and the higher resolution WRF-model had heavier tails than CT, such that the assumption of Gaussian errors in regional inversion systems lead to a further discounting of observational evidence due to overestimation of transport errors.

In contrast to considerable model biases in the atmospheric boundary-layer we only found small biases in the free troposphere and only small differences between models, highlighting the fact that upper tropospheric measurements of CO_2 may be suitable for characterizing continental CO_2 background conditions, which would improve our ability to investigate near surface.

In summary, our work demonstrated the utility of using ACT airborne [CO_2] measurements to investigate CO_2 model-observation mismatch across seasons, regions, and airmass conditions and provide a pathway for similar investigations using targeted model ensembles and to identify the processes responsible for model-observation mismatch.

Acknowledgments

Observational data is available from the ACT public data repository hosted by Oak Ridge National Lab. This work uses the *ACT-America: L3 Merged In Situ Atmospheric Trace Gases and Flask Data, Eastern USA* (Davis et al., 2018). WRF simulation for ACT-America (Feng et al., 2020) are available at The Pennsylvania State University Data Commons. CarbonTracker CT2017 and CT-NRT.v2019-2 results are provided by NOAA ESRL, Boulder, Colorado, USA from the website at http://carbontracker.noaa.gov. The use of NOAA's Observation Package (OBSPACK) for CT2017 and CT-NRT-2019.2 downloaded from the ObsPack Data Portal at www.esrl.noaa.gov/gmd/ccgg/obspack/ is acknowledged. The Atmospheric Carbon and Transport-America (ACT) project was sponsored by the National Aeronautics and Space Administration (NASA) under awards NNX15AG76G and NNX15AJ06G. T. Lauvaux was also supported by the French research program Make Our Planet Great Again (Project CIUDAD - CNRS). We thank NASA's Airborne Sciences program, NASA Headquarters and staff, in particular, Kenneth W. Jucks and Jennifer R. Olson for their support of our mission. We would like to acknowledge the contributions of ACT collaborators, in particular, NASA project managers, scientists, and engineers and our colleagues at NOAA, Colorado State University for their excellent cooperation during the field campaign. Thanks are also due to the flight crews and aircraft facility groups from Wallops Flight Facility, Langley Research Center, and Duncan Aviation for their outstanding work supporting these flights and measurements. We also thank Hannah Halliday and John B. Novack for their contributions in data collection.

References

Agustí-Panareda, A., Diamantakis, M., Massart, S., Chevallier, F., Muñoz-Sabater, J., Barré, J., ... Wunch, D. (2019). Modelling CO_2 weather – why horizontal resolution matters. *Atmospheric Chemistry and Physics*, *19*(11), 7347–7376. doi: 10.5194/acp-19-7347-2019

Andres, R. J., Gregg, J. S., Losey, L., Marland, G., & Boden, T. A. (2011). Monthly, global emissions of carbon dioxide from fossil fuel consumption. *Tellus B: Chemical and Physical Meteorology*, *63*(3), 309–327. doi: 10.1111/j.1600-0889.2011.00530.x

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., ... Tans, P. P. (2014). CO_2, CO, and CH_4 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. doi: 10.5194/amt-7-647-2014

Baier, B. C., Sweeney, C., Choi, Y., Davis, K. J., DiGangi, J. P., Feng, S., ... Weibring, P. (2020). Multispecies assessment of factors influencing regional CO_2 and CH_4 enhancements during the winter 2017 ACT-America campaign. *Journal of Geophysical Research: Atmospheres*, *125*(2), e2019JD031339. doi: 10.1029/2019JD031339

Baker, D. F., Doney, S. C., & Schimel, D. S. (2006). Variational data assimilation for atmospheric CO_2. *Tellus B: Chemical and Physical Meteorology*, *58*(5), 359–365. doi: 10.1111/j.1600-0889.2006.00218.x

Barnes, E. A., Parazoo, N., Orbe, C., & Denning, A. S. (2016). Isentropic transport and the seasonal cycle amplitude of CO_2. *Journal of Geophysical Research: Atmospheres*, *121*(13), 8106–8124. doi: 10.1002/2016JD025109

Basu, S., Baker, D. F., Chevallier, F., Patra, P. K., Liu, J., & Miller, J. B. (2018). The impact of transport model differences on CO_2 surface flux estimates from OCO-2 retrievals of column average CO_2. *Atmospheric Chemistry and Physics*, *18*(10), 7189–7215. doi: 10.5194/acp-18-7189-2018

Blasing, T. J., Broniak, C. T., & Marland, G. (2005). The annual cycle of fossil-fuel carbon dioxide emissions in the United States. *Tellus B: Chemical and Physical Meteorology*, *57*(2), 107–115. doi: 10.3402/tellusb.v57i2.16779

Boden, T., Andres, R., & Marland, G. (2016). *Global, Regional, and National Fossil-Fuel CO2 Emissions (1751 - 2013) (V. 2016).* Environmental System Science Data Infrastructure for a Virtual Ecosystem; Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). doi: 10.3334/CDIAC/00001_V2016

Bousquet, P., Ciais, P., Peylin, P., Ramonet, M., & Monfray, P. (1999). Inverse modeling of annual atmospheric CO2 sources and sinks: 1. Method and control inversion. *Journal of Geophysical Research: Atmospheres*, *104*(D21), 26161–26178. doi: 10.1029/1999JD900342

Butler, M. P., Lauvaux, T., Feng, S., Liu, J., Bowman, K. W., & Davis, K. J. (2020). Atmospheric simulations of total column CO_2 mole fractions from global to mesoscale within the Carbon Monitoring System Flux Inversion Framework. *Atmosphere*, *11*(8), 787. doi: 10.3390/atmos11080787

Chan, D., Yuen, C. W., Higuchi, K., Shashkov, A., Liu, J., Chen, J., & Worthy, D. (2004). On the CO_2 exchange between the atmosphere and the biosphere: The role of synoptic and mesoscale processes. *Tellus B: Chemical and Physical Meteorology*, *56*(3), 194–212. doi: 10.3402/tellusb.v56i3.16424

Chen, H. W., Zhang, F., Lauvaux, T., Davis, K. J., Feng, S., Butler, M., & Alley, R. B. (2019). Characterization of regional-scale CO_2 transport uncertainties in an ensemble with flow-dependent transport errors. *Geophysical Research Letters*, *46*(7), 4049–4058. doi: 10.1029/2018GL081341

Chen, H. W., Zhang, L. N., Zhang, F., Davis, K. J., Lauvaux, T., Pal, S., … DiGangi, J. P. (2019). Evaluation of regional CO_2 mole fractions in the ECMWF CAMS real-time atmospheric analysis and NOAA CarbonTracker Near-Real-Time reanalysis with airborne observations from ACT-America field campaigns. *Journal of Geophysical Research: Atmospheres*, *124*(14), 8119–8133. doi: 10.1029/2018JD029992

Chevallier, F., Feng, L., Bösch, H., Palmer, P. I., & Rayner, P. J. (2010). On the impact of transport model errors for the estimation of CO_2 surface fluxes from GOSAT observations. *Geophysical Research Letters*, *37*(21). doi: 10.1029/2010GL044652

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., … Thornton, P. E. (2013). Carbon and other biogeochemical cycles. In *Climate Change*

2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 465–570). Cambridge, UK; New York, USA: Cambridge University Press.

Comission, E. (2019). *Emissions Database for Global Atmospheric Research (EDGAR).*

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., … Jones, D. B. A. (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. doi: 10.5194/acp-19-9797-2019

Davis, K., Obland, M., Lin, B., Lauvuax, T., O'Dell, C., Meadows, B., … Nehrir, A. (2018). *ACT-America: L3 Merged In Situ Atmospheric Trace Gases and Flask Data, Eastern USA.* ORNL Distributed Active Archive Center. doi: 10.3334/ORNLDAAC/1593

Denning, A. S., Fung, I. Y., & Randall, D. (1995). Latitudinal gradient of atmospheric CO_2 due to seasonal exchange with land biota. *Nature*, *376*(6537), 240. doi: 10.1038/376240a0

Díaz-Isaac, L. I., Lauvaux, T., Bocquet, M., & Davis, K. J. (2019). Calibration of a multi-physics ensemble for estimating the uncertainty of a greenhouse gas atmospheric transport model. *Atmospheric Chemistry and Physics*, *19*(8), 5695–5718. doi: 10.5194/acp-19-5695-2019

Díaz-Isaac, L. I., Lauvaux, T., & Davis, K. J. (2018). Impact of physical parameterizations and initial conditions on simulated atmospheric transport and CO_2 mole fractions in the US Midwest. *Atmospheric Chemistry and Physics*, *18*(20), 14813–14835. doi: 10.5194/acp-18-14813-2018

Díaz-Isaac, L. I., Lauvaux, T., Davis, K. J., Miles, N. L., Richardson, S. J., Jacobson, A. R., & Andrews, A. E. (2014). Model-data comparison of MCI field campaign atmospheric CO_2 mole fractions. *Journal of Geophysical Research: Atmospheres*, *119*(17), 10536–10551. doi: 10.1002/2014JD021593

Fast, J. D., Gustafson, W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., … Peckham, S. E. (2006). Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. *Journal of Geophysical Research: Atmospheres*, *111*(D21). doi: 10.1029/2005JD006721

Feng, S., Lauvaux, T., Barkley, Z., Davis, K., Butler, M., Deng, A., ... Stauffer, D. (2020). *Full WRF-Chem output in support of the NASA Atmospheric Carbon and Transport (ACT)-America project (7/1/2016 – 7/31/2019).* Penn State Data Commons. doi: 10.26208/RQF5-Q142

Feng, S., Lauvaux, T., Davis, K. J., Keller, K., Zhou, Y., Williams, C., ... Baker, I. (2019). Seasonal characteristics of model uncertainties from biogenic fluxes, transport, and large-scale boundary inflow in atmospheric CO_2 simulations over North America. *Journal of Geophysical Research: Atmospheres*, *124*(24), 14325–14346. doi: 10.1029/2019JD031165

Feng, S., Lauvaux, T., Keller, K., Davis, K. J., Rayner, P., Oda, T., & Gurney, K. R. (2019). A road map for improving the treatment of uncertainties in high-resolution regional carbon flux inverse estimates. *Geophysical Research Letters*, *46*(22), 13431–13469. doi: 10.1029/2019GL082987

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *J. Climate*, *27*(2), 511–526. doi: 10.1175/JCLI-D-12-00579.1

Gaubert, B., Stephens, B. B., Basu, S., Chevallier, F., Deng, F., Kort, E. A., ... Yin, Y. (2019). Global atmospheric CO_2 inverse models converging on neutral tropical land exchange, but disagreeing on fossil fuel and atmospheric growth rate. *Biogeosciences*, *16*(1), 117–134. doi: 10.5194/bg-16-117-2019

Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P., Vermeulen, A. T., ... Santaguida, R. (2007). Comparing atmospheric transport models for future regional inversions over Europe – Part 1: Mapping the atmospheric CO_2 signals. *Atmospheric Chemistry and Physics*, *7*(13), 3461–3479. doi: 10.5194/acp-7-3461-2007

Gerbig, C., Körner, S., & Lin, J. C. (2008). Vertical mixing in atmospheric tracer transport models: Error characterization and propagation. *Atmospheric Chemistry and Physics*, *8*(3), 591–602. doi: 10.5194/acp-8-591-2008

Gerken, T. (2019, May). *TobGerken/ACT_ManeuverFlags V1.0.* Zenodo. doi: 10.5281/zenodo.2677933

Giglio, L., Loboda, T., Roy, D. P., Quayle, B., & Justice, C. O. (2009). An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote*

Sensing of Environment, *113*(2), 408–420. doi: 10.1016/j.rse.2008.10.006

Giglio, L., Randerson, J. T., & van der Werf, G. R. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *Journal of Geophysical Research: Biogeosciences*, *118*(1), 317–328. doi: 10.1002/jgrg.20042

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., ... Michalak, A. M. (2012). North American CO_2 exchange: Inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion. *Biogeosciences*, *9*(1), 457–475. doi: 10.5194/bg-9-457-2012

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, *39*(37), 6957–6975. doi: 10.1016/j.atmosenv.2005.04.027

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., ... Yuen, C.-W. (2002). Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models. *Nature*, *415*(6872), 626. doi: 10.1038/415626a

Hu, L., Andrews, A. E., Thoning, K. W., Sweeney, C., Miller, J. B., Michalak, A. M., ... van der Velde, I. R. (2019). Enhanced North American carbon uptake associated with El Niño. *Science Advances*, *5*(6), eaaw0076. doi: 10.1126/sciadv.aaw0076

Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., ... Cook, R. (2012). North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison. *Ecological Modelling*, *232*, 144–157. doi: 10.1016/j.ecolmodel.2012.02.004

Jacobson, A. R., Fletcher, S. E. M., Gruber, N., Sarmiento, J. L., & Gloor, M. (2007). A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes. *Global Biogeochemical Cycles*, *21*(1). doi: 10.1029/2005GB002556

Keppel-Aleks, G., Wennberg, P. O., & Schneider, T. (2011). Sources of variations in total column carbon dioxide. *Atmospheric Chemistry and Physics*, *11*(8), 3581–3593. doi: 10.5194/acp-11-3581-2011

Keppel-Aleks, G., Wennberg, P. O., Washenfelder, R. A., Wunch, D., Schneider, T., Toon, G. C., ... Wofsy, S. C. (2012). The imprint of surface fluxes and transport on variations in total column carbon dioxide. *Biogeosciences*, *9*(3), 875–891. doi: 10.5194/bg-9-875-2012

Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., ... Bergamaschi, P. (2005). The two-way nested global chemistry-transport zoom model TM5: Algorithm and applications. *Atmospheric Chemistry and Physics*, *5*(2), 417–432. doi: 10.5194/acp-5-417-2005

Lan, X., Tans, P., Sweeney, C., Andrews, A., Jacobson, A., Crotwell, M., ... Wolter, S. (2017). Gradients of column CO_2 across North America from the NOAA Global Greenhouse Gas Reference Network. *Atmospheric Chemistry and Physics*, *17*(24), 15151–15165. doi: 10.5194/acp-17-15151-2017

Lauvaux, T., & Davis, K. J. (2014). Planetary boundary layer errors in mesoscale inversions of column-integrated CO_2 measurements. *Journal of Geophysical Research: Atmospheres*, *119*(2), 490–508. doi: 10.1002/2013JD020175

Lauvaux, T., Díaz-Isaac, L. I., Bocquet, M., & Bousserez, N. (2019). Diagnosing spatial error structures in CO_2 mole fractions and XCO_2 column mole fractions from atmospheric transport. *Atmospheric Chemistry and Physics*, *19*(18), 12007–12024. doi: 10.5194/acp-19-12007-2019

Lauvaux, T., Schuh, A. E., Bocquet, M., Wu, L., Richardson, S., Miles, N., & Davis, K. J. (2012). Network design for mesoscale inversions of CO_2 budget of the corn belt: Exploring uncertaintie sources and sinks. *Tellus B: Chemical and Physical Meteorology*, *64*(1), 17980. doi: 10.3402/tellusb.v64i0.17980

Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., ... Davis, K. J. (2012). Constraining the CO_2 budget of the corn belt: Exploring uncertainties from the assumptions in a mesoscale inverse system. *Atmospheric Chemistry and Physics*, *12*(1), 337–354. doi: 10.5194/acp-12-337-2012

Masarie, K. A., Peters, W., Jacobson, A. R., & Tans, P. P. (2014). ObsPack: A framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements. *Earth System Science Data*, *6*(2), 375–384. doi: 10.5194/essd-6-375-2014

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, *58*(8), 1246–1266. doi: 10.2113/gsecongeo.58.8.1246

Merrill, J. T., & Moody, J. L. (1996). Synoptic meteorology and transport during the North Atlantic Regional Experiment (NARE) intensive: Overview. *Journal of Geophysical Research: Atmospheres*, *101*(D22), 28903–28921. doi: 10.1029/96JD00097

Myhre, G., Shindell, D. T., Breon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., … Zhang, H. (2013). Anthropogenic and Natural Radiative Forcing. In *Climate Change 2013 - The Physical Science Basis* (pp. 659–740). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107415324.018

Nassar, R., Napier-Linton, L., Gurney, K. R., Andres, R. J., Oda, T., Vogel, F. R., & Deng, F. (2013). Improving the temporal and spatial distribution of CO_2 emissions from global fossil fuel emission data sets. *Journal of Geophysical Research: Atmospheres*, *118*(2), 917–933. doi: 10.1029/2012JD018196

Oda, T., & Maksyutov, S. (2011). A very high-resolution (1 km x 1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmospheric Chemistry and Physics*, *11*(2), 543–556. doi: 10.5194/acp-11-543-2011

Olsen, S. C., & Randerson, J. T. (2004). Differences between surface and column atmospheric CO2 and implications for carbon cycle research. *Journal of Geophysical Research: Atmospheres*, *109*(D2). doi: 10.1029/2003JD003968

Pal, S., Davis, K. J., Lauvaux, T., Browell, E. V., Gaudet, B. J., Stauffer, D. R., … Zhang, F. (2020). Observations of greenhouse gas changes across summer frontal boundaries in the eastern United States. *J. Geophys. Res. Atmos.*, *125*, e2019JD030526. doi: 10.1029/2019JD030526

Parazoo, N. C., Denning, A. S., Berry, J. A., Wolf, A., Randall, D. A., Kawa, S. R., … Doney, S. C. (2011). Moist synoptic transport of CO2 along the mid-latitude storm track. *Geophysical Research Letters*, *38*(9). doi: 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_2 in North America, South America and Europe. *Atmospheric Chemistry and Physics*, *8*(23), 7239–7254. doi: 10.5194/acp-8-7239-2008

Parazoo, N. C., Denning, A. S., Kawa, S. R., Pawson, S., & Lokupitiya, R. (2012). CO_2 flux estimation errors associated with moist atmospheric processes. *Atmospheric Chemistry and Physics*, *12*(14), 6405–6416. doi: 10.5194/acp-12-6405

-2012

Patton, A., Politis, D. N., & White, H. (2009). Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White. *Econometric Reviews*, *28*(4), 372–375. doi: 10.1080/07474930802459016

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., ... Tans, P. P. (2007). An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *PNAS*, *104*(48), 18925–18930. doi: 10.1073/pnas.0708986104

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., ... Zhang, X. (2013). Global atmospheric carbon budget: Results from an ensemble of atmospheric CO_2 inversions. *Biogeosciences*, *10*(10), 6699–6720. doi: 10.5194/bg-10-6699-2013

Peylin, P., Rayner, P. J., Bousquet, P., Carouge, C., Hourdin, F., Heinrich, P., ... AEROCARB contributors (2005). Daily CO_2 flux estimates over Europe from continuous atmospheric measurements: 1. Inverse methodology. *Atmospheric Chemistry and Physics*, *5*(12), 3173–3186. doi: 10.5194/acp-5-3173-2005

Politis, D. N., & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, *23*(1), 53–70. doi: 10.1081/ETC-120028836

Potter, C. S., Klooster, S. A., Myneni, R., Genovese, V., Tan, P.-N., & Kumar, V. (2003). Continental-scale comparisons of terrestrial carbon sinks estimated from satellite data and ecosystem modeling 1982–1998. *Global and Planetary Change*, *39*(3), 201–213. doi: 10.1016/j.gloplacha.2003.07.001

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., & Klooster, S. A. (1993). Terrestrial ecosystem production: A process model based on global satellite and surface data. *Global Biogeochemical Cycles*, *7*(4), 811–841. doi: 10.1029/93GB02725

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., ... Duda, M. G. (2017). The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, *98*(8), 1717–1737. doi: 10.1175/BAMS-D-15-00308.1

Rasmussen, L. A. (1991). Piecewise integral splines of low degree. *Computers & Geosciences*, *17*(9), 1255–1263. doi: 10.1016/0098-3004(91)90027-B

Schuh, A. E., Jacobson, A. R., Basu, S., Weir, B., Baker, D., Bowman, K., ... Palmer, P. I. (2019). Quantifying the impact of atmospheric transport uncertainty on CO_2 surface flux estimates. *Global Biogeochemical Cycles*, *33*(4), 484–500. doi: 10.1029/2018GB006086

Schuh, A. E., Lauvaux, T., West, T. O., Denning, A. S., Davis, K. J., Miles, N., ... Ogle, S. (2013). Evaluating atmospheric CO_2 inversions at multiple scales over a highly inventoried agricultural landscape. *Glob Change Biol*, *19*(5), 1424–1439. doi: 10.1111/gcb.12141

Schwanghart, W. (2013). *Experimental (Semi-) Variogram Version 1.4.0.* https://www.mathworks.com/matlabcentral/fileexchange/20355-experimental-semi-variogram.

Skamarock, W. C., Klemp, J., Dudhia, J., Gill, D., Barker, D., Wang, W., ... Duda, M. (2008). *A Description of the Advanced Research WRF Version 3* (Tech. Rep.). UCAR/NCAR. doi: 10.5065/D68S4MVH

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., ... Denning, A. S. (2007). Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO_2. *Science*, *316*(5832), 1732–1735. doi: 10.1126/science.1137004

Sweeney, C., Karion, A., Wolter, S., Newberger, T., Guenther, D., Higgs, J. A., ... Tans, P. P. (2015). Seasonal climatology of CO_2 across North America from aircraft measurements in the NOAA/ESRL Global Greenhouse Gas Reference Network. *Journal of Geophysical Research: Atmospheres*, *120*(10), 5155–5190. doi: 10.1002/2014JD022591

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., ... de Baar, H. J. W. (2009). Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(8), 554–577. doi: 10.1016/j.dsr2.2008.12.009

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., & Arellano Jr., A. F. (2006). Interannual variability in global biomass burning emissions from 1997 to 2004. *Atmospheric Chemistry and Physics*, *6*(11), 3423–3441. doi: 10.5194/acp-6-3423-2006

Wei, Y., Shresha, R., Pal, S., Gerken, T., McNelis, J., Singh, D., ... Davis, K. J. (in review). The ACT-America Datasets: Description, Management and Delivery. *Earth and Space Sciences*.